UNIVERSITÀ DEGLI STUDI DI PADOVA
SEDE AMMINISTRATIVA: UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE FARMACEUTICHE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI
INDIRIZZO: SCIENZE FARMACEUTICHE
XXI CICLO

THE DOUBLE FACE OF NUCLEIC ACIDS

DIRETTORE DELLA SCUOLA: PROF. MAURIZIO CASARIN

RELATORE: PROF.SSA BARBARA GATTO

CORRELATORE: YVES POMMIER, M.D., PH.D.

DOTTORANDA: JESSICA MARINELLO

31 GENNAIO 2009

UNIVERSITY OF STUDY OF PADOVA
ADMINISTRATIVE OFFICE: UNIVERSITY OF STUDY OF PADOVA
DEPARTMENT OF PHARMACEUTICAL SCIENCES

PH.D. SCHOOL IN MOLECULAR SCIENCES
ORIENTATION: PHARMACEUTICAL SCIENCES
XXI CYCLE

THE DOUBLE FACE OF NUCLEIC ACIDS

SCHOOL DIRECTOR: PROF. MAURIZIO CASARIN

SUPERVISOR: PROF. BARBARA GATTO

SUPPORTING SUPERVISOR: YVES POMMIER, M.D., PH.D.

PH.D. STUDENT: JESSICA MARINELLO

31ST JANUARY 2009

Alla mia famiglia

INDEX

ABSTRACT

Nucleic acids have been the object of intense and thorough observations for more than a century, leading to the understanding of their structure and function. Their role as genetic information carriers is well established but there are evidences that they are also involved in a series of other less known processes. RNA for example is a single strand biopolymer that, dependently from the primary sequence, can assume complex three-dimensional foldings that allow binding to diverse target molecules or the involvement in catalysis. Being closely related to several cellular events, they are also useful therapeutic targets.

This work is essentially divided in two parts, aimed at studying two different and complementary aspects of nucleic acids. While the first project goes into the understanding of the "RNA world" and of the applicability of extraordinary RNA conformational variability in a diagnostic and therapeutic system, the second topic is addressed to study the ability of two compounds to interfere with DNA integration, a genetic process that allows the viral DNA to be inserted into the host genome. Therefore nucleic acids demonstrate their "double face" meaning their ability to be leading actors in diagnosis and therapy but also indispensable supporting actors, being the main targets for several drugs.

The first part of the PhD program was focused on the use of SELEX to discover a new diagnostic or therapeutic tool. A protocol for the selection of RNA aptamers binding folic acid was developed. In particular, after 12 cycles of SELEX using affinity chromatography, it was possible to select an enriched pool of single stranded RNA oligonucleotides with affinity for the target. Subcloning and sequencing of the above-mentioned pool allowed the subsequent analysis of the oligonucleotide three-dimensional structure and the determination of the minimum binding sequence to folic acid. Computer tools and RNase protection assay were associated in the experimental protocol to obtain the most probable outcome. A clear region of ligand interaction on RNA was identified; this minimum sequence was chemically synthesized and Isothermal Titration Calorimetry (ITC) used for the determination of the binding constant of folic acid, that resulted in the nanomolar range. The approval from FDA and EMEA of Macugen®, a pegylated aptamer selective for the binding to VEGF and used in the treatment of age-related macular degeneration, is a proof that SELEX can be used to develop good drugs and diagnostics. RNA aptamers against folic acid could now find their way for diagnostic or therapeutic application. Therefore further studies must be addressed to allow their applicability to the desired purpose.

The second part of the PhD program was carried out in the Laboratory of Molecular Pharmacology of the National Cancer Institute (National Institute of Health - Bethesda - Maryland - USA) and it takes into consideration DNA integration as target of anti-HIV drugs. HIV-1 integrase is one of the viral enzymes encoded from the POL gene, that catalyses the insertion of the viral DNA into host chromosomes. Using synthetic oligonucleotides mimicking the terminal portion of the U5 viral LTR, it was analyzed the effect of two important drugs (raltegravir, the first integrase inhibitor approved by FDA last year, and elvitegravir, in advanced stage of human clinical trials) on the recombinant wild type integrase and on a series of resistant mutants. The study was addressed to compare raltegravir and elvitegravir on the same *in vitro* system, with the aim to understand how different aminoacidic point mutations of the protein sequence are responsible in the onset of drug resistance.

RIASSUNTO

Dal giorno della loro scoperta più di un secolo fa, gli acidi nucleici sono stati analizzati e studiati approfonditamente in tutti gli aspetti, cercando di svelare gli ancora numerosi segreti nascosti nella loro struttura per comprenderne appieno le funzioni. Il ruolo più assodato ed essi associato è sicuramente quello di trasportatori dell'informazione genetica ma numerose evidenze sperimentali confermano il coinvolgimento in una serie di altri processi meno noti. Basta pensare a come l'RNA, grazie alle complesse conformazioni tridimensionali che può assumere dipendentemente dalla sua sequenza primaria, risulti coinvolto nel legame a svariate molecole e risulti ancor più straordinariamente responsabile di processi catalitici. Essendo quindi correlati a numerosi eventi cellulari, gli acidi nucleici sono utili bersagli terapeutici.

Il presente lavoro è suddiviso essenzialmente in due parti ed è finalizzato alla comprensione di due aspetti differenti ma complementari legati agli acidi nucleici. Il primo progetto infatti si pone l'obiettivo di conoscere il vasto "mondo dell'RNA" e di applicare la straordinaria variabilità conformazionale di questo biopolimero ad un sistema terapeutico o diagnostico. La seconda parte della ricerca è invece indirizzata all'analisi dell'effetto di composti sul processo di integrazione del DNA, un meccanismo genetico che permette al DNA virale di essere correttamente inserito nel genoma dell'ospite. Gli acidi nucleici dimostrano quindi la loro "doppia faccia" intesa come la loro capacità di essere i protagonisti in sistemi di diagnosi o di terapia ma anche indispensabili attori co-protagonisti, essendo essi stessi i bersagli principali di numerosi farmaci.

La prima metà del ciclo di dottorato di ricerca si è concentrata sull'uso della tecnologia SELEX. Inizialmente si è sviluppato un protocollo per la selezione di aptameri contro l'acido folico. Un totale di 12 cicli di SELEX mediante cromatografia d'affinità, hanno condotto alla selezione di un pool di molecole di RNA a singolo filamento, aventi un'elevata predisposizione a legare il target d'interesse. Il subclonaggio e il sequenziamento del suddetto insieme di molecole ha poi permesso di approfondire l'analisi delle caratteristiche tridimensionali degli oligonucleotidi e della minima sequenza responsabile del legame all'acido folico. A questo scopo, è stata condotta un'analisi informatizzata elementare ed è stato sviluppato un protocollo per l'RNase protection assay. L'associazione dei risultati ottenuti con le due diverse metodiche ha permesso di definire una chiara regione di interazione tra RNA e ligando. Questa minima sequenza è stata quindi sintetizzata chimicamente e sottoposta, mediante Isothermal Titration Calorimetry (ITC), alla determinazione della costante di legame, che è risultata essere nell'intervallo nanomolare.

L'approvazione da parte di EMEA e FDA di Macugen®, un aptamero peghilato selettivo per il legame a VEGF e usato nel trattamento della degenerazione maculare senile, è una prova di come la SELEX possa essere un utile strumento per lo sviluppo di farmaci e diagnostici. Gli aptameri ad RNA contro l'acido folico selezionati nel presente lavoro possono ora seguire la via dello sviluppo per l'applicazione desiderata e ulteriori studi saranno condotti a questo scopo.

La seconda metà del ciclo di dottorato di ricerca è stata invece svolta nel Laboratorio di Farmacologia Molecolare del National Cancer Institute (National Institute of Health - Bethesda - Maryland - USA) e ha preso in considerazione il processo di integrazione del DNA come bersaglio di farmaci anti-HIV. L'integrasi dell'HIV-1 è un enzima virale codificato dal gene POL e catalizza l'inserzione del DNA virale nei cromosomi della cellula ospite. Oligonucleotidi sintetici, con sequenza corrispondente alla porzione U5 terminale della LTR virale, sono stati utilizzati allo scopo di analizzare l'effetto di due importanti farmaci (raltegravir, il primo farmaco inibitore dell'integrasi approvato dall'FDA lo scorso anno, e elvitegravir, un composto in stadio avanzato di ricerca clinica) sull'enzima nativo ricombinante e su una serie di mutanti resistenti ai farmaci in questione. Lo studio è stato essenzialmente indirizzato a confrontare raltegravir e elvitegravir sullo stesso sistema *in vitro*, cercando di capire come differenti mutazioni aminoacidiche della proteina siano responsabili dell'insorgenza di resistenza al trattamento.

FIRST PART

SELEX FOR FOLIC ACID:

HOW RNA OLIGONUCLEOTIDES

BECOME DIAGNOSTIC AND THERAPEUTIC TOOLS

"The purpose of research is to explore the unknown.
The desire for new knowledge calls forth the answers to new questions"

Walter Gilbert

Nobel Prize in Chemistry
8th December 1980

1. INTRODUCTION

1.1 THE SELEX METHODOLOGY

The principal function of nucleic acids has always been associated to their informational role that allows genetic information to be passed from organism to organism. However, several evidences demonstrated that they have other functions, participating in binding to certain target molecules and even being involved in catalysis. The essential dogma is that where nucleic acids are known to bind to some targets with high affinity and specificity, the binding depends on the exact sequences of nucleotides that comprise the DNA or RNA ligand. tRNA is the paradigm of these highly specialized structures, formed by a single-stranded RNA that adopts a spatial conformation making it a selective carrier for specific amino acids (*Figure 1*).

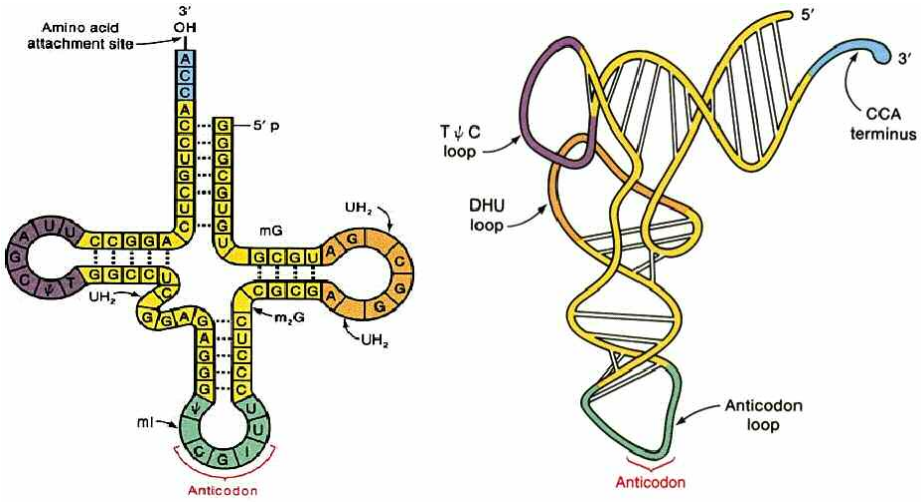

Figure 1: *the folding of tRNA is encoded in its sequence and the folding determines its molecular interaction.*

The SELEX methodology (Systematic Evolution of Ligands by Exponential enrichment), applied for the first time in 1990 by Gold [1] and Szostak [2] independently, was evolved from the unique insight that nucleic acids have sufficient capacity for forming a variety of two- and three-dimensional structures and sufficient chemical versatility available within their monomers to act as ligands with virtually any compound. This powerful technique allows the screening of large combinatorial library of nucleic acids with particular recognition properties. In fact, in a mixture of nucleic acids with several possible sequences and structures, there is for sure a wide range of binding affinities for a given target. Incubation of

the mixture with the target permits binding of the oligonucleotides with the higher affinity constant. After partitioning of unbound and bound molecules, dissociation of nucleic acid-target pairs and amplification of the selected RNA or DNA pool, a second nucleic acid mixture is generated, enriched for the higher binding affinity candidates. Additional rounds of selection progressively favor the best ligands until the resulting nucleic acid mixture is predominantly composed of only a few sequences. Cycle of selection and amplification are repeated until a desired goal is achieved and it usually corresponds to no significant improvement in binding strength with the repetition of the cycle. Finally, cloning and sequencing of the selected clones reveal the sequence and the structure of the specific selected ligands. In summary, SELEX adopts the principles of evolution, i.e. variation, screening and replication of the best molecule possessing a specific activity and it is comparable with combinatorial chemistry techniques.

The oligonucleotides resulting from the selection process are referred to as “aptamers”, derived from the Latin word “*aptus*”, meaning “to fit”. Aptamers are extraordinary tools. It is enough to think that they bind with affinity, selectivity and specificity comparable to monoclonal antibodies, having *Kd* values in the low nanomolar to picomolar range [3]. Some major strengths make them unique. For example they can be obtained without the use of animals because SELEX is a technology totally performed in a reaction tube. Their handling is easy because nucleic acids are easily renaturable upon denaturation, and they can be easily conjugated to several reporters and polymers. Furthermore they can be obtained by highly standardized and automated chemical synthesis, with an appreciable decrease of production and purification costs. Finally their synthesis is highly reproducible from batch to batch and they are not subject to the variability typical of biotechnological products. Interestingly, they are poor antigens due to their small size and similarity to endogenous molecules. Therefore they are considered low or nonimmunogenic and nontoxic molecules.

The initial mixture necessary for the selection includes single stranded DNAs or RNAs having a central randomized sequence portion as well as two conserved sequences at both ends, essential for the oligonucleotide amplification by PCR. The combinatorial random region can be nowadays easily synthesized asking to the synthesizer to use a statistical mixture of the four bases at the randomized positions. This implies a theoretical 4^n number of different molecules, where *n* is the length of the randomized sequence. In practice, the complexity of a typical library is limited to 10^{15} individual sequences.

A variety of nucleic acid primary, secondary and tertiary structures are known to exist and are commonly involved in non-Watson-Crick type interactions. Noteworthy are hairpin loops, symmetric and asymmetric bulges, pseudoknots (*Figure 2*). These non-canonical

intramolecular interactions are responsible for the interaction with the target and they can be formed in a nucleic acid sequence of no more than 30 nucleotides. Therefore it is preferred to perform the *in vitro* selection using mixtures with a randomized segment from 20 to 50 nucleotides in length.

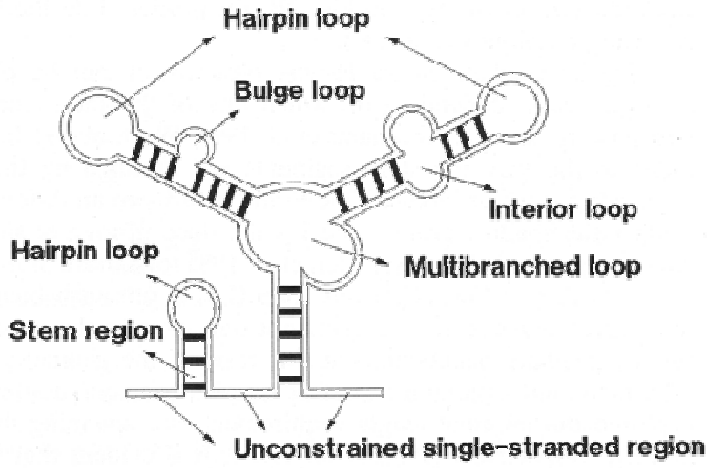

Figure 2: *schematic representation of the classical secondary structure of RNA. Hairpins, bulges and interior loops can be detected. Another important structural motif is formed upon base pairing of nucleotides from a loop with a complementary region outside that loop. The resulting structure is called a pseudoknot (element not shown in the figure).*

The composition of the random library is essential for the success of SELEX. In fact DNA is more stable than RNA and can be easily synthesized, but, on the other hand, RNA molecules have more complex folding motifs and are more suitable for *in vivo* applications. To overcome the primary limitation of the use of RNA aptamers, that is their nuclease sensitivity, the stability of these molecules can be improved by the introduction of a chemical modification on the 2' of the ribose ring. 2'-amino or 2'-fluoro groups are usually introduced in pyrimidine rings along the sequence because the pyrimidine-specific nucleases are the most abundant in biological fluids (*Figure 3*). Furthermore 2'-O-methyl nucleotides are also protected against minor endonucleases. All these modifications protect the RNA from degradation, considerably increasing the half life. They can be introduced directly into the initial combinatorial library because these backbone modifications are compatible with the enzymes used in the SELEX process (such as reverse transcriptase or DNA or RNA polymerases). However they can also be introduced later after the selection. Additional protection from exonucleases can be provided through terminal capping with small molecules such as an amine linker, a phosphate group or an inverted thymidine residue. Noteworthy are also oligonucleotides modified with phosphorothioate linkages, LNA (Locked Nucleic Acid), and Spiegelmers, L-oligonucleotides that, being composed of unnatural monomers, are not

good substrates for nucleases and are stable for more than 60 hours in biological fluids [4]. Improved pharmacokinetics and a considerable stability to renal and plasma clearance have been accomplished by conjugating polyethylene glycol (PEG) to the aptamer, or by covalently attaching dialkylglycerol (DAG) and then incorporating this conjugate into liposomes.

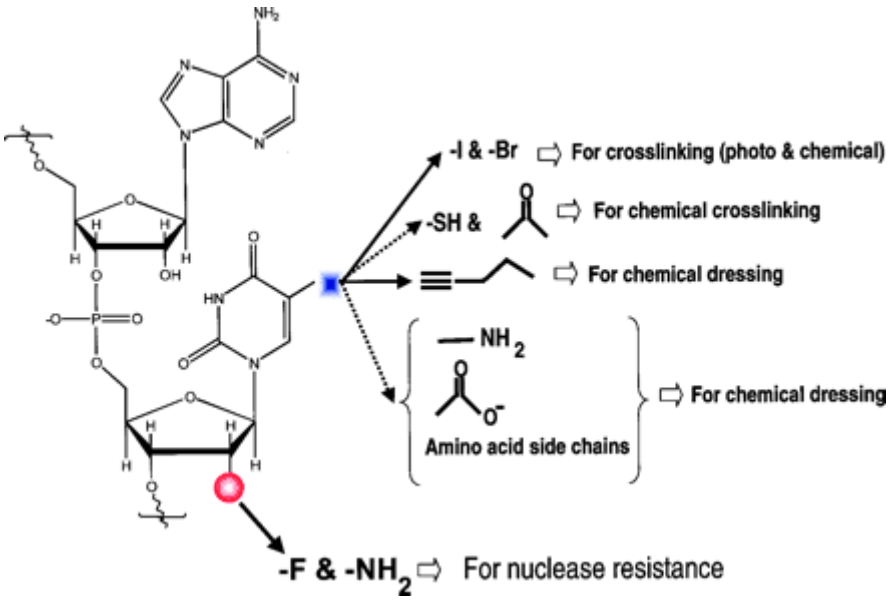

Figure 3: *possible modifications on the aptamer strand. 2' ribose substitutions confer nuclease stability, whereas modifications at the C5 position of the pyrimidines could be used to generate covalent cross-links with targets. From [5].*

SELEX literature is now extremely broad and varied. Since their discovery, aptamers have been generated against a wide variety of targets ranging from small molecules, peptides, amino acids and proteins, including cell membrane proteins. Binding a wide range of molecules, they have applications as therapeutics, diagnostics and research tools [6] (for additional information see paragraph 1.3).

SELEX experiments take from weeks to months to be completed but in 2001 the process has been automated [7, 8]. Furthermore, Golden [9] proposed a novel methodology called PhotoSELEX. It is based on the incorporation of a modified nucleotide activated by absorption of light in place of a native base. Therefore the aptamers can form a photo-induced covalent bond with the target molecule, resulting in greater sensitivity and specificity than conventional aptamers. Instead, the cell-based SELEX, called Cell-SELEX, allows the selection of a panel of target cell-specific aptamers (*Figure 4*). The technique is a promising tool especially for the identification of unique molecular features of cancer cells. At the same time, it generates probes that can recognize such unique features with very high affinity and

specificity, leading to the formulation of new biomarker-based assays that provide information for accurate disease diagnosis and prognosis. [10].

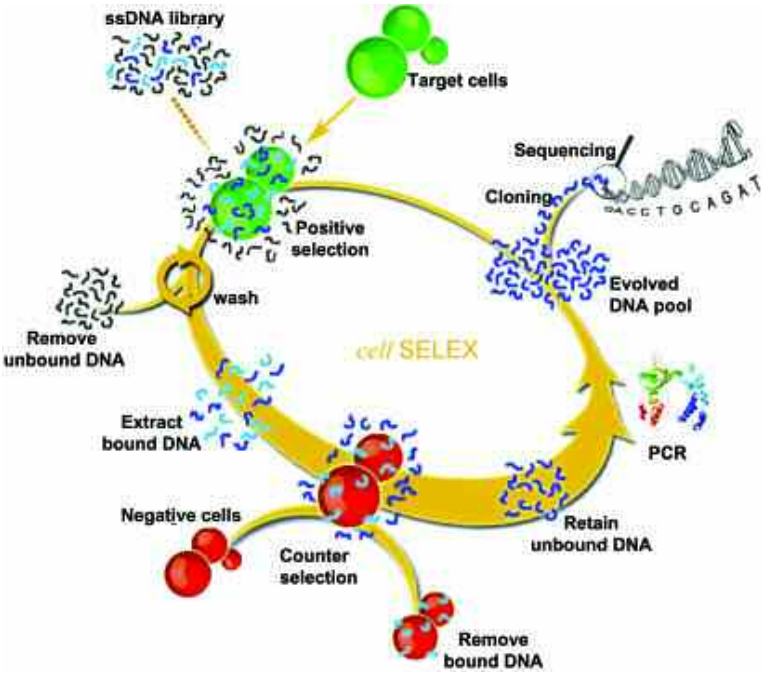

Figure 4: *representation of a cell-based aptamer selection. From [10].*

1.2 THE PROCESS INTO DETAILS

The SELEX process begins, as previously described, with a random sequence library obtained from combinatorial chemical synthesis of single stranded DNAs. In a typical DNA SELEX, the library is directly incubated with the target of interest in a buffer of choice (referred as Binding Buffer, BB) at a given temperature. The BB usually contains magnesium, that permits the stabilization of the secondary structure and is therefore essential for the correct folding of the aptamers [11]. Instead, in a RNA SELEX, the DNA library is previously transcribed and the obtained RNA mixture incubated with the target of interest. During this step, a small fraction of individual sequences interact with the target. These sequences are then separated from all the others by means of any one of the physical separation techniques. Nitrocellulose filter portioning is often used with protein targets, while small targets are preferably immobilized on a solid support, generating an affinity matrix from which the unbound species can be easily removed by washing. At this point, the bound sequences are isolated and directly amplified by PCR during a DNA SELEX, while in a RNA SELEX the isolated sequences are reverse transcribed and then amplified. The amplification allows to obtain an enriched library to be used for the next selection cycle. After several rounds the

affinity saturation is achieved. The required total number of cycles is dependent on the imposed degree of stringency, that governs the enrichment efficiency of high-affinity binders, as well as on the nature of the target, but in general it goes from 8 to 15 cycles. The enriched library is finally cloned and sequenced to obtain the sequence information of each member (*Figure 5*). The fixed sequence regions of the obtained aptamers are then truncated to eliminate nucleotide stretches that are not important for direct interaction with the target or for folding into the structure that permits target binding, and the sequence is produced by chemical synthesis. Assays for the determination of the minimum target binding sequence and of the affinity constant are then performed and the aptamers developed according to the desired employment.

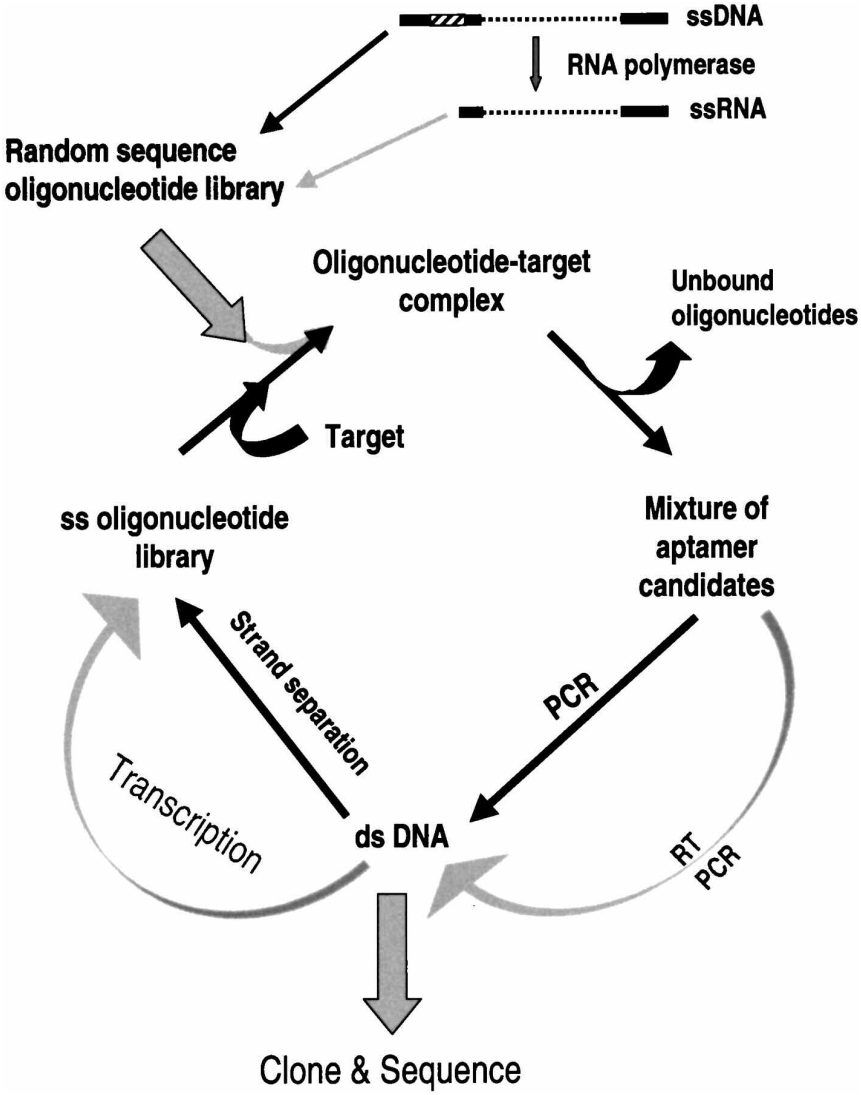

Figure 5: *scheme indicating the key steps in a SELEX process. From [5].*

1.3 APTAMERS APPLICATIONS

Aptamers as Detection Reagents. Aptamers are binding molecules that rival antibodies and therefore, soon after their discovery, they demonstrated their potential and versatility in the field of diagnostics, partially replacing monoclonal antibodies, which are still the ubiquitous reagents used in these assays. Illustrative is an aptamer chip that can assess approximately 50 different analytes in patient samples [12]. But several other examples are present in literature. *Fredriksson* [13] developed a proximity-dependent DNA ligation assay, reaching a sensitivity 1000-fold higher than a conventional ELISA (zeptomolar range). The system is based on two aptamers binding different epitopes of the cytokine platelet-derived growth factor (PDGF). When they bind the target and they are consequently in proximity, they can be ligated enzymatically and the resulting molecule can be amplified by two primers, each specific for one of the aptamer sequences. In common use are also the aptamer beacons, aptamers that generate fluorescent signals upon structural rearrangement. Usually a fluorophore is attached to the DNA aptamer but, in the absence of the target, it is quenched by a complementary DNA strand carrying a quencher dye. When the aptamer bind to the target, the quencher-labeled DNA is released and the step can be monitored by real-time PCR [14, 15].

Gold and co-workers at Somalogic Inc. developed an aptamer-based microarray using so-called photoaptamers [16, 17]. Aptamers bearing 5'-iodo- or 5'-bromo-substituted uridinetriphosphate (UTP) or 2'-deoxyuridinetriphosphate (dUTP) can be cross-linked to a target protein using UV photoexcitation. The technique is convenient because aptamers are small in size allowing high-density aptamer coupling to the chip surface. In addition, the UV inducted covalent bonds allow harsh washing conditions to remove nonspecifically bound proteins from the array surface. Finally, the diversity between analytes (proteins or peptides) and capturing reagents (nucleic acids) allows the use of a generic protein stain to visualize the molecular recognition.

Several studies demonstrated as aptamers are appropriate tools in ELISA-like assays, in flow cytometry and in microscopy studies. For example, a fluorescein isothiocyanate (FITC)-conjugated aptamer has been used as tumor marker to selectively visualize the paths and the branching of the neoangiogenetic, pathologic microvasculature in tissue section of rat brain glioblastoma [18] (*Figure 6*).

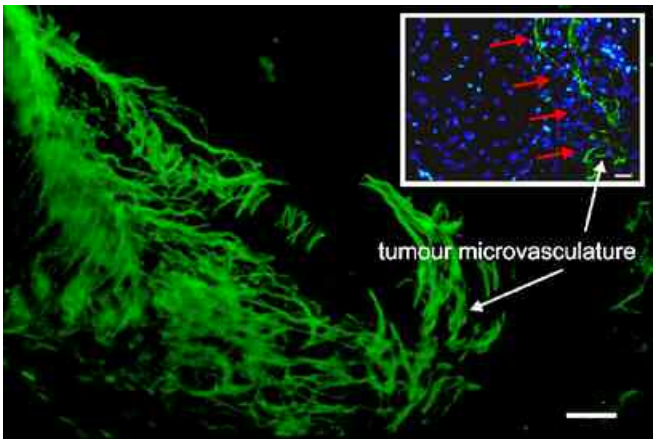

Figure 6: *selective staining of the activated tumor microvasculature on tissue sections of rat brain glioblastoma using a FITC-conjugated aptamer (green). In blue, the DAPI staining of cellular nuclei. From [3].*

Aptamers in Analytical Application. Aptamers, for their potential as molecular recognition tools, can be incorporated into analytical devices and used as immobilized ligands in separation technologies. Aptamers can be easily modified at 3' or 5' end for direct immobilization on carrier material for affinity chromatography. Biotin, primary amino or other functional groups can be introduced allowing their binding to the matrix. Therefore it is possible to purify a desired protein from a complex cell medium using aptamers as capture ligands, with the advantage to maintain the protein function, compared to antibody-based affinity purifications which require denaturing elution conditions such as low pH or detergents. Noteworthy is also their use as affinity probes in capillary electrophoresis-based quantitative assays of proteins, as substrate for protein capture, and analysis in MALDI Mass Spectrometry, as biocomponents in biosensors. Their exceptional specificity allows them to discriminate between closely related isoforms or different conformational states of the same target molecule.

Aptamers for Target Validation. The possibility to have molecular and cellular strategies to determine the function of proteins, to determine the relation between a specific protein and a disease and to evaluate if these disease-causing agents are possible therapeutic targets, is nowadays the priority. Technologies addressing the process of target validation are several. Genetic inactivation at the genomic or transcriptional level is used in the so called *loss-of-function* technologies. Between these approaches, noteworthy are gene knockout, antisense oligonucleotides, ribozymes and RNA interference. However they provide only limited information about the correlation between protein and disease process, because abolishing a protein more or less completely from its functional environment incurs the risk of causing pleiotropic effects due to the impact on other important factors in the system. Antibodies are

also used as target validation technologies but the long development time and the *in vivo* production make difficult their application. In addition they are limited to the extracellular environment. For all these reasons, aptamers are excellent alternative for target validation. Their labeling is useful for localization studies and the low molecular weight allows the accessibility to intracellular target (*Figure 7*). In fact they can be conjugated with vector systems that enable high-level expression in both the cell nucleus and cytoplasm, or they can be directly delivered using electroporation or transfection reagents such as liposomes. They inhibit the target by competitive and noncompetitive mechanism, resulting in the possibility to study inhibitory mechanisms or protein conformations in high-order complexes or protein networks, without modification of the proteomic status of the model system. To date, several aptamers against intra- and extracellular target have been selected [19]. Noteworthy is also the use of aptamers in High-Throughput Screening (HTS). It is based on the fact that small molecules with the same binding site and similar inhibition profiles are able to dislodge aptamers from their position. In fact small molecules and aptamers are able to inhibits protein in a similar way, binding in cavities and clefts. Therefore in HTS, libraries of small molecules are tested on their ability to compete with the aptamers [20].

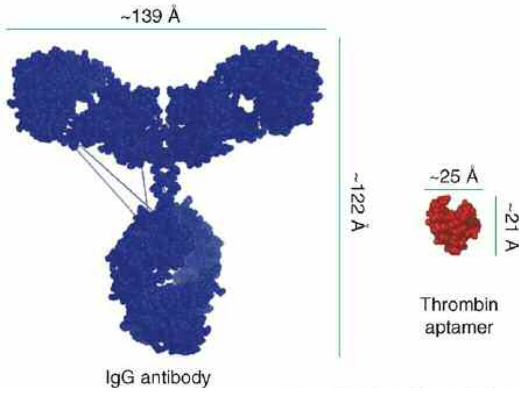

Figure 7: *antibody-aptamer size comparison. From [4].*

Aptamers as Therapeutics. The most successful therapeutic application of an aptamer is the adaptation of an antivascular endothelial growth factor (anti-VEGF) aptamer for the treatment of age-related macular degeneration (AMD). AMD is a leading cause of irreversible, severe loss of vision among old people in the developed world. It is characterized by cumulative damage or genetic defects of the retinal pigment epithelium, leading to gradual loss of the protective cells ("dry" AMD) or abnormal growth of blood vessels into the retina ("wet" AMD). Growth factors play an important role in the development and growth of neovascularization and vascular leakage in the "wet" form. VEGF is a potent permeability factor of about 35-45 kDa that mediates the complex cascade of ocular angiogenesis. At least

four isoforms (121, 165, 189 and 205 amino acids) derive from alternative splicing of the mRNA, but the most pathogenic species is $VEGF_{165}$. The vascular factor stimulates endothelial cells by binding to tyrosine kinase receptors and determining cell proliferation, migration and survival. In December 2004, the FDA approved pegaptanib sodium (Macugen®) as the only therapy for the treatment of all subtypes of wet AMD. Pegaptanib is a 28-base RNA oligonucleotide that terminates in a pentylamino linker, to which two branched 20-kDa monomethoxy polyethylene glycol (PEG) units are covalently attached via the two amino groups on a lysine residue (*Figure 8*). It is intravitreally administrated and it binds to extracellular $VEGF_{165}$ isoform and prevents the vascular factor from stimulating the receptor on the surface of the endothelial cells. Therefore it can be considered a selective VEGF antagonist. The approval of this aptamer as drug is an important milestone that witnesses the innovation and versatility of these compounds. In addition, Macugen® shows that pegylation does not reduce protein affinity, since the aptamer maintains its high affinity and specificity also *in vivo [21]*.

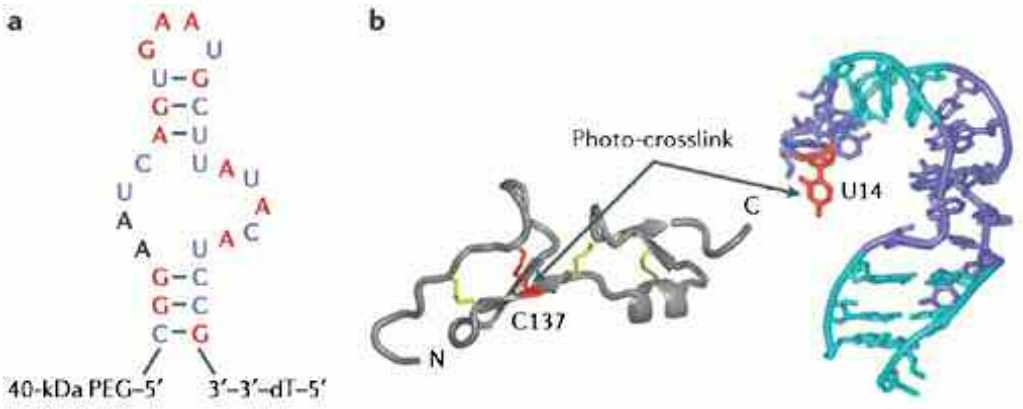

Figure 8: a) *Sequence and predicted secondary structure of pegaptanib. In red the 2'-O-methylated purines and in blue the 2'-fluorine-modified pyrimidines.* ***b)*** *Interaction between the cysteine 137 in the heparin-binding domain of* $VEGF_{165}$ *and the uridine 14 of the aptamer. From [22].*

Anticoagulant therapy is also moving in the direction of evolving new strategies and some aptamers demonstrated to be promising drug candidates [4]. ARC183 has passed Phase I clinical trials as antithrombin aptamer used for transient anticoagulation during coronary artery bypass graft surgery. REG1 is instead an aptamer-antidote pair. In particular, a modified RNA aptamer conjugated to a cholesterol moiety is able to bind the coagulation factor IXa. The antidote for this aptamer is a 17 base pair, 2'-O-methyl RNA oligonucleotide that binds to the aptamer, disrupts its structure and renders it inactive (*Figure 9*). This mechanism is advantageous because it allows more control over and better timing of the reversal of anticoagulant activity.

Several aptamers have been selected for viral coats and they might be promising antiviral agents. RNA aptamer specific for the conserved coreceptor region of gp120 of HIV-1, as well as for Rev, Gag, Tat, reverse transcriptase and integrase are known, but nowadays there is no clinical employment.

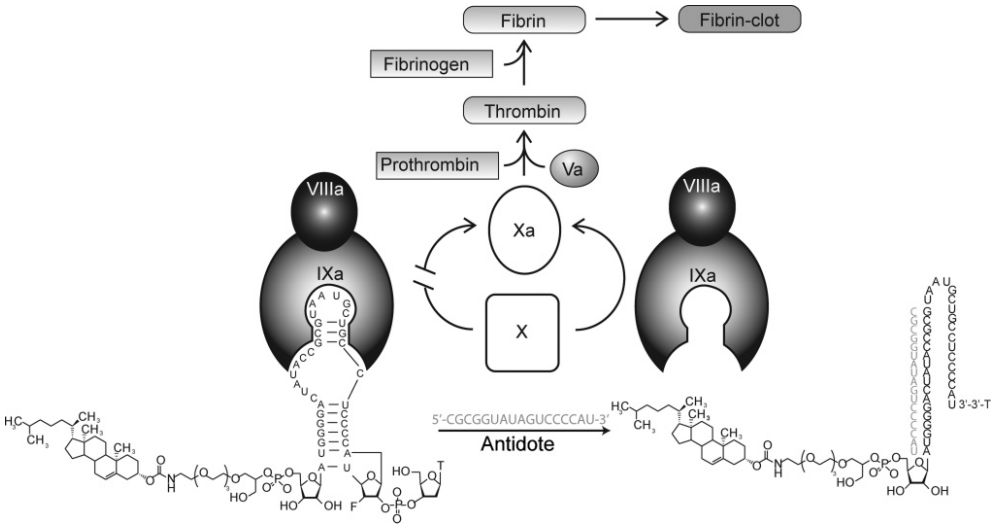

Figure 9: *mechanism of the aptamer-antidote pair. The aptamer binds to activated factor IXa and prevents the cleavage of factor X. In the presence of the antidote, the aptamer is released from factor IXa. From [23].*

Catalytic Oligonucleotides. Nowadays it is well known that proteins are not the only biopolymers capable for catalyzing chemical transformation. In nature in fact there are RNA sequences able to catalyze the hydrolysis and formation of phosphodiesters. It was in 1990 that Cech and Altman [24, 25] discovered the catalytic nucleic acids, hence deserving the Nobel prize. Catalytic RNAs, known as ribozymes, can be found in bacteria, in viruses, in plants, in low eukaryotes and in vertebrates, man included, and they can be divided in two categories dependently on the mechanism of action: large and small catalytic RNAs. The first group includes group I and group II introns and RNase P [26, 27], with sizes that vary between few hundred and few thousand nucleotides. Their action results in the formation of a free 3'-hydroxyl and a 5'-phosphate group (*Figure 10*). Small catalytic RNAs have a limited size, up to 155 nucleotides, and the category includes the hammerhead, the hairpin, the hepatitis delta virus (HDV) and the varkud satellite (VS) RNA ribozymes. They catalyze self-cleaving reactions (also known as intramolecular or in *cis* catalysis), resulting in a 2',3'-cyclic phosphate and a 5'-hydroxyl and involving the 2'-hydroxyl group of the ribose [28, 29] (*Figure 11*).

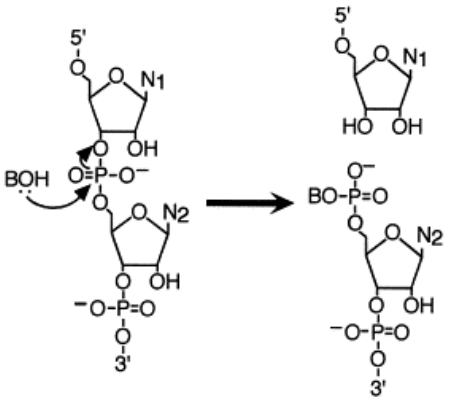

Figure 10: *mechanism of action of large catalytic RNAs. The nucleophilic BOH is the 3'-hydroxyl of the cofactor which is a guanosine for group I introns, a nucleotide in group II introns and a water molecule in the RNase P-catalyzed reaction. From [30].*

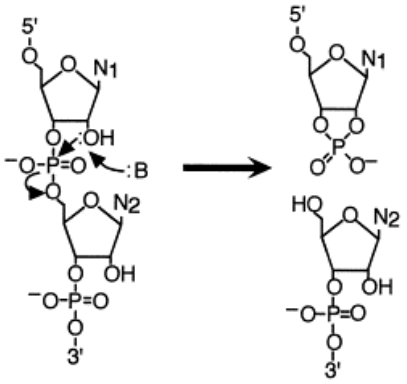

Figure 11: *mechanism of action of small catalytic RNAs. The reaction is most likely initiated by the activation of the 2' hydroxyl on the ribose located at the scissile bond and it results in a product with a 2',3'-cyclic phosphate. From [30].*

Ribozymes can not be considered real enzymes because they are modified after reaction. The only exception is RNase P, that cleaves the 5' end of the tRNA precursor. However all the ribozymes can be modified, permitting to work in *trans* and to split the substrate from the catalytic part. They operate as DNA and RNA endonucleases, ligase, kinase, phosphatase, RNA capping.

The *in vitro* selection has permitted during the years the discovery of several catalytic nuclei acids. Like proteins, ribozymes can be designed to be allosterically regulated. The resulting molecules, called aptazymes, contain separate catalytic and effector-binding domains that interact in a ligand-dependent manner. Aptamers are prone to adaptative binding upon interaction with a specific ligand and therefore they are suitable to append the recognition domain to the ribozyme. The technology, called "allosteric selection" is based on the binding of a random region on the ribozyme molecule and consequently the selection for the desired effector [31].

1.4 FOLIC ACID

Folic acid (pteroylglutamic acid, Vitamin B9 or Vitamin M) gets its name from the Latin word "*folium*", meaning "leaf". It is the oxidized and most active form of the vitamin and it is present in vitamin preparations and food fortifications. There is in fact a distinction between food folate and folic acid, essentially in the bioavailability. For example, the food folate is only about half as available as folic acid consumed on an empty stomach, because it occurs in natural foods as polyglutamate, a form less absorbed than free folate. It is present in spinach, dried beans and peas, turnip greens, sunflower seeds and other fruits and vegetables.

It is structurally composed by three molecules: 6-methylpterin, p-aminobenzoic acid (PABA) and glutamic acid (*Figure 12*). The glutamic acid does not participate in the coenzyme functions of folic acid. Instead, folic acid in the interior of the cell may contain a "chain" of glutamic acids, which serves as a negatively charged portion to keep the coenzyme inside cells or bound to the appropriate enzymes. Contrary to that, the pteridine portion and the p-aminobenzoic acid participate directly in the metabolic reactions of folate.

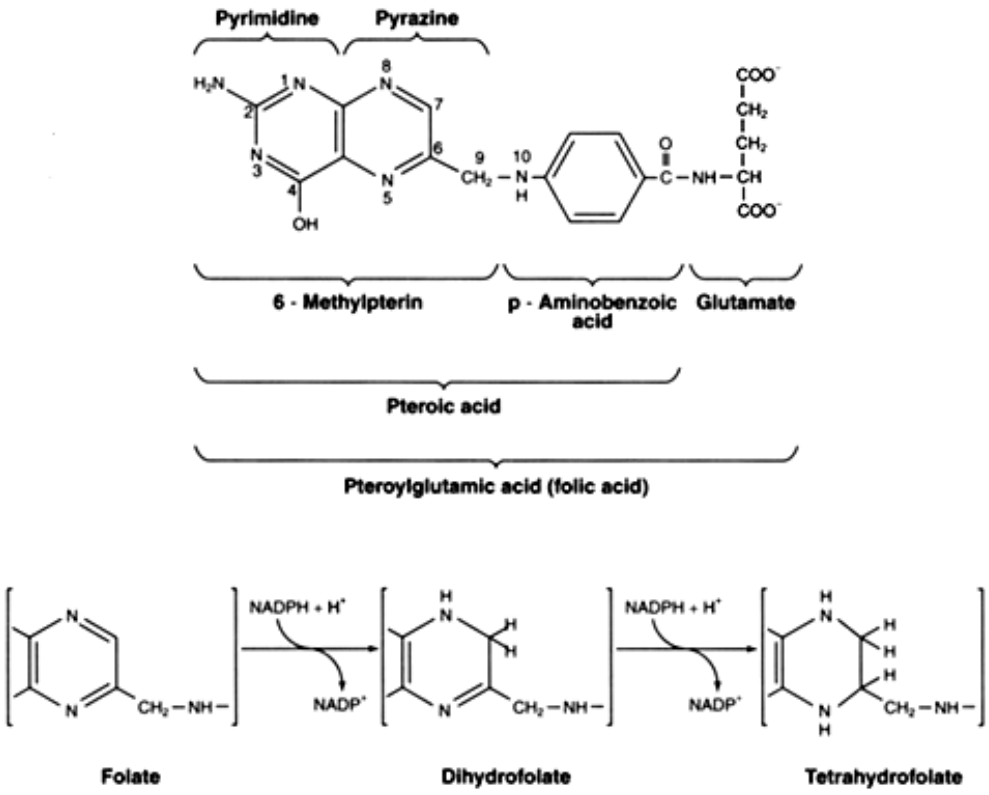

Figure 12: *chemical structure of folic acid. It is formed by the union of three molecules: 6-methylpterin, p-aminobenzoic acid and glutamic acid. The reduction of folate to tetrahydrofolate by dihydrofolate reductase, generates the central compound in the single-carbon metabolism.*

The main function of folic acid is to be a coenzyme in single-carbon transfers in the metabolism of nucleic and amino acids and thus fills an important function in purine and pyrimidine metabolism. The key compound in transfer reactions is the tetrahydrofolate. The formation of tetrahydrofolate begins when folate is reduced to dihydrofolate (FH_2), which is then reduced to tetrahydrofolate (FH_4), by dihydrofolate reductase (DHFR) in presence of NADPH cofactor (*Figure 12*). FH_4 is then transformed in methylene tetrahydrofolate by the addition of a methylene group from a carbon donor as formaldehyde, serine or glicine. The methylene group can also be reduced giving methyl tetrahydrofolate or oxidated giving formil terahydrofolate. All these folate derivatives are direct substrates in the single-carbon transfer reaction above mentioned.

Folic acid provides the methyl essential for amino acid methionine metabolism (*Figure 13*). In particular, vitamin B12 is the only acceptor of methyl tetrahydrofolate and it consequently transfer the methyl group of folate derivative to homocysteine by action of homocysteine methyltransferase. This process allows the formation of an other important carrier of methyl group, the S-adenosyl-methionine (SAM). It is in fact involved in methylation of DNA and RNA and in the conversion of epinephrine to norepinefrine.

A deficiency of B12 can determine accumulation of unusable methyl tetrahydrofolate mimicking a folate deficiency. However, both B6, B9 and B12 deficiency determine hyperhomocysteinemia associated to a number of diseases.

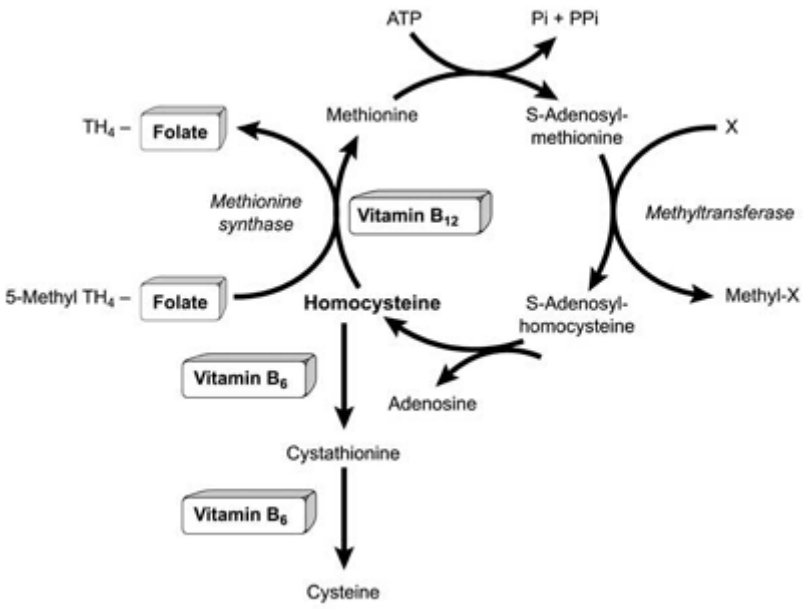

Figure 13: *metabolism of homocysteine. Homocysteine may be methylated to form methionine by a folate-dependent reaction that is catalyzed by methionine synthase, a vitamin B12-dependent enzyme. Alternatively, homocysteine may be metabolized to cysteine in reactions catalyzed by two vitamin B6-dependent enzymes.*

Folate is involved also in the methylation of deoxyuridine monophosphate (dUMP) to generate thymidylate (dTMP), which is needed for DNA synthesis. Consequently deficiency at this level determines megaloblastic anemia, a deficit in the generation of mature red blood cells.

Folic acid is necessary for the production and maintenance of new cells, especially during periods of rapid cell division and growth such as infancy and pregnancy. It is fundamental during the first stage of embryonal development: its adequate intake during the periconceptional period, the time just before and just after a woman becomes pregnant, helps protect against a number of congenital malformations including tube defects [32]. The primordial central nervous system begins as a plate of cells early in embryonic life, which folds on itself to form a tube. The closure of the tube occurs in the interval of days 21-28 postconception. Failure of closure results in neural tube defects as anencephaly, in which the cerebral cortex and overlying bony calvarium fail to develop, and spina bifida, in which the spinal cord is dysplastic and the overlying spinal column is absent. The risk of these defects is significantly reduced when supplemental folic acid is consumed prior to and during the first month following conception. In many countries a fortification program has been introduced and folic acid added to several foods. For example in the United States, in 1996 the FDA required the addition of folic acid to breads, cereals, flours, corn meals, pastas, rice and other grain products, and nowadays fortified foods are the major source of folic acid in the American diet. The only concern is about the ability of folic acid to hide a vitamin B_{12} deficiency. If the vitamin deficiency is not treated, it determines permanent nerve damages. Therefore, the intake of folic acid should not exceed 1 mg per day to prevent folic acid from masking symptoms of vitamin B_{12} deficiency.

A low level of folic acid increases the blood level of homocysteine, a risk factor for heart disease and stroke. It is not completely clear if supplementation with folic acid can lower the risk of developing coronary heart disease and further studies are necessary in this direction. Folic acid seems to be correlated also to depression and to memory and mental agility.

As folic acid is important for rapidly dividing cells and tissues, drugs that interfere with the folate metabolism are used as anticancer drug. Methotrexate is an antineoplastic compound which acts as an antimetabolite of folic acid. It and its active metabolites compete for the folate binding site of the enzyme dihydrofolate reductase. The competitive inhibition leads to blockage of tetrahydrofolate synthesis, depletion of nucleotide precursors and inhibition of DNA, RNA and protein synthesis. It also inhibits thymidylate synthase and the transport of reduced folates into the cell. The main toxic effect of metrotexate is related to its inability to distinguish between normal and cancer cells.

Determination of folic acid's plasmatic level is essential in pregnancy and pathological conditions. The plasma concentration reflects a person's recent intake of folic acid in the diet. Its amount in red blood cells is instead related to the period when the cell was generated, as much as two months earlier and therefore it is a more accurate way to measure the body's level of folate. CMIA (Chemiluminescent Magnetic Immunoassay), FPIA (Fluorescence Polarization Immunoassay) and RIA (Radio Immunoassay) are common techniques used for this purpose. All of them involve the use of monoclonal antibodies, resulting sometimes in abnormal and false values due to interaction between detection system and human antibodies, and therefore justifying the necessity to develop a new strategy for the determination of folic acid in patients.

2. AIM OF THE WORK

The SELEX methodology is an effective tool for the identification of nucleic acids with specific recognition properties for a given target. The approval from FDA and EMEA of Macugen®, a pegylated aptamer selective for the binding to VEGF and used in the treatment of age-related macular degeneration, is a proof that SELEX can be used to develop good drugs and diagnostics. Therefore the technique was chosen with the aim to identify a sequence able to bind with high affinity and specificity folic acid. The process was developed using a RNA library for the more complex folding motifs that RNA can adopt.

The isolation of an aptamer specific for folic acid could lead, after modifications increasing the stability of the RNA oligonucleotide and conjugation with a desired reporter, to obtain an efficient diagnostic tool. The control of folic acid's plasmatic level is important in pregnancy and pathological conditions. Different methodologies are nowadays used for this purpose but all of them are based on the use of monoclonal antibodies, resulting in interaction between detection system and human antibodies. Aptamers could overcome these problems as they have sensitivity comparable to antibodies but they lack the non-specific recognition problems associated with the use of other biotechnological products. In additions aptamers have an advantage on the shelf-life and they can be obtained by highly standardized and automated chemical synthesis, with an appreciable decrease of production and purification costs, and high reproducibility from batch to batch.

Receptors for folic acid are known to be overexpressed in cancer cells. It is in fact an essential factor in cells and tissues that rapidly divide, for its involvement in the biosynthesis of purines and pyrimidines. The availability of an aptamer binding folic acid can be exploited in site-specific drug directioning. Indeed, the conjugation of a drug used in cancer therapy with the aptamer could lead to the specific accumulation of the drug in proximity of cells overexpressing folic acid receptors, therefore cancer cells.

The knowledge of a specific RNA sequence able to bind a factor with a crucial rule in many conditions, can be adapted to many purposes, highlighting the importance of the present work.

3. EXPERIMENTAL PROCEDURES

3.1 OLIGONUCLEOTIDES

The random pool, 122 bases long, 85 of which are randomized, was synthesized by Metabion International AG (Martinsried, Germany) following the classical chemistry. The sequence is hereafter reported. The highlighted sequence corresponds to the reverse transcription promoter. In green boxes the fixed regions of the oligonucleotides.

5'-GGAGAAAGGGCCGCGGG- -N_{85}- **-AGAA**GAGGTGACGGTACCAG-3'

For DNA amplification by PCR, the following primers were used:

- **C_{1KPNI}** : 5'-CTGGTACCGTCACCTC-3' this primer is 16 bases long and it is complementary to the 3' end of the random pool.
- **C_2** : 5'-TCTAATACGACTCACTATAGGAGAAAGGGCAGCGG-3' this primer is 35 bases long and it is analogous to the 5' end of the random pool. The underlined region corresponds to the T7 RNA polymerase promoter needed for the transcription.

Other primers were used for the insertion of BamHI and EcoRI restriction endonuclease sites, allowing the subcloning of selected sequences inside the plasmid of interest. There primers determined the removal during subcloning of undesired sequences as the T7 RNA polymerase promoter. They are:

- **FA1**: 5'-AGG-CTT-GGA-TCC-CGT-CAC-CTC-TTC-T-3' this primer is 25 bases long and it is used to replace $C_{1KPNI.}$ It presents the cleavage site for BamHI (underlined sequence).
- **FA2**: 5'-AGG-AGA-AAG-GGC-CGC-3' this primer is 15 bases long, it does not contain cleavage sites but it is essential for the removal of T7 RNA polymerase promoter.
- **FA3**: 5'-TAC-GTA-GAA-TTC-GGA-GAA-AGG-GCC-GC-3' this primer is 26 bases long and it replaces FA2 at the 5' end. It contains the cleavage site for EcoRI (underlined sequence).

All the primers were synthesized by Metabion International AG (Martinsried, Germany).

3.2 PCR AMPLIFICATION

A mix containing from 5 to 15 ng of DNA depending on the SELEX cycle, 1X Taq polymerase reaction buffer, 25 μM final concentration of each primer, 200 μM final concentration of dNTPs and 5 units of *Taq* polymerase (Amersham Biosciences, GE Healthcare Bio-Sciences Corp., NJ, USA), in a total volume of 100 μl, was prepared. The mix was incubated in a thermal cycler using the following parameters:

Cycle	*Temperature*	*Time*
1	94 °C	5 minutes
2	94 °C	30 seconds
3	53 °C	1 minute
4	72 °C	30 seconds
5	72 °C	7 minutes
6	4 °C	hold

Cycles from 2 to 4 were repeated for a total of 17 times. The dsDNA was then purified with QIAquick PCR Purification Kit (Qiagen Inc., CA, USA) to remove unincorporated oligonucleotides and then acetate/ethanol precipitated.

3.3 IN VITRO TRANSCRIPTION

In vitro transcription was used to obtain RNA from the dsDNA. The DNA library pool was transcribed into an RNA library pool by incubating overnight at 37 °C in the presence of 1X transcription buffer, 10 mM DTT, 10 mM NTPs, 160 units of RNase inhibitor (Promega, WI, USA), 150 units of T7 RNA Polymerase (Stratagene, CA, USA) and 3000 Ci/mMol of ^{32}P[α-UTP] (Perkin Elmer, MA, USA). The DNA template was subsequently degraded by incubating the reaction mixture with 3 units of DNase at 37 °C for 1 hour. The unincorporated nucleotides were removed with G-25 columns following the manufacturer's instructions. After acetate/ethanol precipitation, the RNA was resuspended and quantified through absorption at 260 nm.

3.4 REVERSE TRANSCRIPTION

The RNA obtained by the selection was reverse transcribed by incubating for 50 minutes at 42 °C the RNA itself with 2.5 µM primer C_{1KPNI}, 1X reverse transcription buffer, 10 mM DTT, 1 mM dNTPs, 400 units of SuperScriptII Reverse Transcriptase (Invitrogen, CA, USA). The enzyme was then inactivated heating to 70 °C for 15 minutes.

3.5 AFFINITY CHROMATOGRAPHY MATRIX PREPARATION

Epoxy-activated Sepharose 6B (Amersham Biosciences, GE Healthcare Bio-Sciences Corp., NJ, USA) was hydrated and equilibrated with a 0.1 M solution of $NaHCO_3$, pH 10. The folic acid coupling was performed by incubation of the desired concentration of folic acid solution with the matrix, in the dark and under stirring, at 30 °C for 24 hours (*Figure 14*).

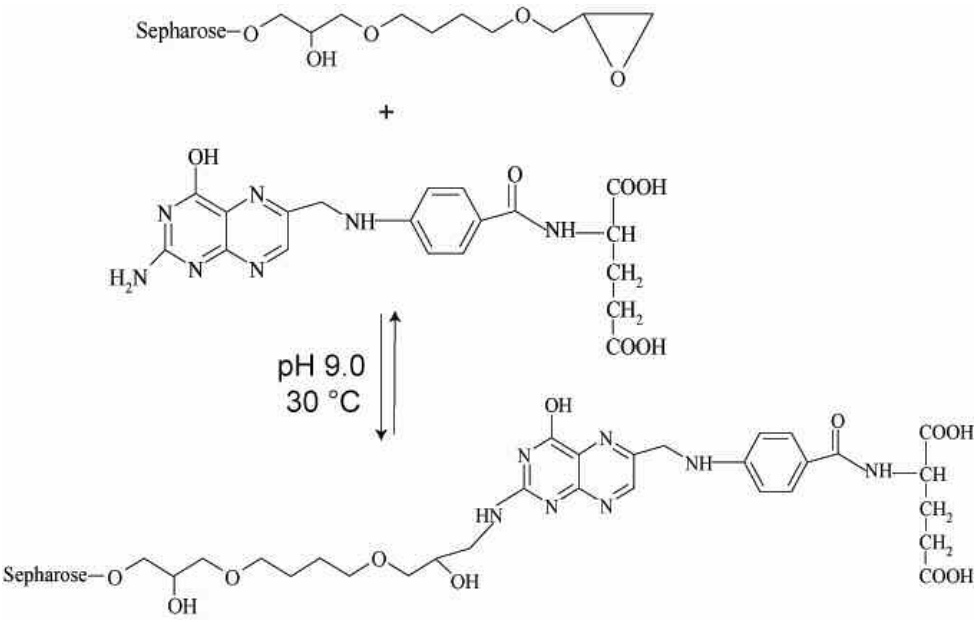

Figure 14: *schematic illustration of the chemistry used for folic acid coupling to the Sepharose matrix.*

The coupled medium was then incubated overnight at 30 °C with 1 M monoethanolamine (pH 8) to block any remaining active groups. Finally, the medium was washed with a solution containing 0.5 M NaCl in 0.1 M $NaHCO_3$ (pH 8.3), followed by a second solution containing 0.5 M NaCl in $CH_3COO^-Na^+$ (pH 4), and a third solution of sterile water (pH 7). The wash was repeated twice and the matrix was resuspended in 10 mM Tris and 5 mM EDTA and stored at 4 °C, protected from light. The folic acid affinity chromatography matrix was prepared by packing the coupled medium (1000 µl, 300 µl, or 200 µl depending on the

SELEX cycle) into a column (Poly-prep BioRad, CA, USA). The packed column was washed with Binding Buffer (BB: 250 mM NaCl,50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$) and equilibrated prior to the selection process, preventing the formation of aspecific binding (as electrostatic interactions). The magnesium allows a higher stability of the RNA structures.

3.6 IN VITRO SELECTION OF FOLIC ACID-BINDING APTAMERS

Prior to incubation with the folic acid-modified affinity column, the RNA pool was denatured at 95 °C for 3 minutes and allowed to renature at room temperature for 30 minutes. To eliminate RNA molecules that non-specifically bind to the column matrix, the initial pool was first incubated with a not functionalized column (this step was called preselection and was used also after the fifth Selex cycle). The flow-through fraction from this incubation was subsequently transferred to a folic acid-modified affinity column and incubated for 30 minutes. Following the incubation period, the affinity column was washed with binding buffer (the amount of BB changed dependently on the cycle) to remove unbound RNA. Bound RNA was eluted by competition with a more concentrated solution of folic acid. The eluted RNA was recovered by ethanol precipitation. Reverse transcription and PCR amplification were then performed. The DNA library was finally transcribed into a RNA library pool for the subsequent selection cycle. A total of 12 selection cycles were carried out during the *in vitro* selection process.

3.7 SUBCLONING OF SELECTED SEQUENCES

The DNA pool from cycle 11 was amplified by PCR using primers that inserted BamHI and EcoRI restriction sites. The amplified pool was cloned into pBluescriptIIKS(+) plasmid (Fermentas International Inc., Ontario, Canada), using the restriction sites present in the fixed regions of the aptamer sequence and the plasmid construct. In particular, a determined amount of aptamers (1-10 μg) and of plasmid (1-5 μg) were treated separately with 10 units of BamHI and EcoRI (Promega, WI, USA) in presence of Buffer H (900 mM Tris-HCl pH 7.5, 500 mM NaCl, 100 mM $MgCl_2$) and BSA 1X. The mixtures were incubated overnight (aptamer DNA) or 1 hour (plasmid vector) at 37 °C and the obtained cleaved products were purified using *QIAquick Gel Extraction Kit Protocol* following manufacturer's instructions. For the ligase reaction, about 100 ng of cleaved plasmid and 600 ng of cleaved aptamer DNA were mixed

with 6 units of T4 Ligase (Promega, WI, USA) in buffer 300 mM Tris-HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP and incubated 3 hours at 22 °C.

Competent *Escherichia coli* DH5α cells (Invitrogen, CA, USA) were transformed by heat shock treatment with the obtained recombinant plasmid. Cells were incubated on ice for 30 minutes in presence of the plasmid. The mixture was heated for 45 seconds at 37 °C and placed on ice again for 2 minutes. 900 µl of Luria-Bertani (LB) broth medium were added to the transformation reaction and incubated at 37 °C for 1 hour with shaking at 225-250 rpm before being spread on LB-agar plates containing ampicillin (100 mg/ml) and incubated at 37 °C overnight.

The day after some colonies were picked up and amplified in 5 ml of LB broth medium with ampicillin (100 µg/ml final) for 6 hours at 37 °C with shaking. From this bacterial solution was then extracted the plasmid DNA using the Qiagen QIAprep Spin Miniprep Kit (Valencia, CA), according to the manufacturer's instructions. Subcloning was confirmed by colony PCR and positive colonies were sequenced by BMR Genomics S.r.l..

3.8 SEQUENCE ANALYSIS

The resulting sequences were aligned using MultAlin sequence alignment program (http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html). Secondary structures' predictions were obtained using Kinefold program (http://kinefold.curie.fr/).

3.9 RNASE PROTECTION ASSAY

The DNA of the desired aptamer family is transcribed and the obtained RNA dephosphorylated in 5' end with Calf Intestinal Alkalyne Phosphatase (CIAP) and labeled using T4 polynucleotyde kinase (Ambion, Applied Biosystems, TX, USA) in presence of [γ-^{32}P]ATP (3000Ci/mmol, Perkin Elmer, MA, USA) following manufacturer's instructions. After PAGE-purification of the radiolabeled RNA, the RNAse protection assay was used to determine the minimum binding sequence to folic acid. 0.2 µg of RNA were resuspended in *structural buffer* (1X solution: 10 mM Tris pH 7, 0.1 M KCl, 10 mM $MgCl_2$), heated to 95 °C in water, and allowed to assume the correct folding with the slow cooling of the water to room temperature. The folic acid is then added to the RNA solution (ratio RNA:folic acid 1:10000)

and the mixture incubated for 30 minutes at room temperature in the dark. 4 µg of tRNA were then added and the mixture divided in four aliquots, ready for the RNase treatment. The first aliquot was not treated; the second was treated directly with the RNases stock solution (i.e. 1 µg RNase A, 1U RNase T1 and 0.1U RNase V1); the third was treated with 1/10 of the RNase concentration used in the second aliquote; the fourth was treated with 1/100 of the RNase concentration used in the second aliquote. RNase A, V1 and T1 were used for this analysis (Ambion, Applied Biosystems, TX, USA). After 10 minutes, the activity of the RNases was blocked adding gel loading dye with formammide. The samples were denatured at 95 °C for 5 minutes and loaded in 10% polyacrylamide denaturing sequencing gels. Dried gels were visualized using photographic films (Hyperfilm MP Amersham, GE Healthcare Bio-Sciences Corp., NJ, USA).

The same technique was used to determine the possible secondary structure of the aptamer. The procedure was the same but there was no incubation of the RNA with the folic acid.

Ladders of the sequence were obtained by alkaline hydrolysis and by identification of all the present guanines after treatment of the denatured RNA with RNase T1. Alkaline hydrolysis procedure: 0.1 µg of RNA were diluted in *hydrolysis buffer* (1X solution: 50 mM sodium carbonate pH 9.2, 1 mM EDTA). 3 µg of tRNA were then added and the mix was divided in three aliquotes and placed at 95°C. After the desired period of time (4.5, 5.5, 6.5 or 7.5 minutes) the tubes were placed on ice to block the reaction and gel loading dye with formammide was added. Samples were ready to be loaded in the sequencing gel. RNase T1 treatment: 0.2 µg of RNA were diluted in *sequence buffer* (1X solution: 20 mM sodium citrate pH 5, 1 mM EDTA and 7 M urea). 3 µg of tRNA were then added and the mix placed at 50°C for 5 minutes to allow denaturation. The solution was divided in four aliquots and these were treated with the endonucleolitic enzyme as previously described for the determination of aptamer structure.

3.10 APTAMER BINDING CONSTANT DETERMINATION

The constant of binding of folic acid to RNA aptamer was determined by Isothermal Titration Calorimetry (ITC, MicroCal Inc., Northampton, MA, USA). A syringe containing a solution of folic acid (the ligand) was titrated into a cell containing a solution of RNA oligonucleotide at constant temperature. The heat released or absorbed after each injection was monitored over time by the system and represented by a peak associated to heat change.

The quantity of heat is in fact directly proportional to the amount of binding. As the system reached saturation, the heat signal diminished until only heats of dilution were observed. The binding curve was obtained from a plot of the heats from each injection against the ratio of folic acid and RNA in the cell.

Going into details about the system upon ITC is based, a cell feedback network (CFB) is used to differentially measure and compensate for heat released or absorbed between the sample and reference cell. A thermoelectric device measures the temperature difference between the two cells and a second device measures the temperature difference between the cells and the jacket. The temperature difference between sample and reference cell (ΔT1) is kept at a constant value by the addition or removal of heat to the sample cell using the CFB system. The integral of the power required to maintain a constant temperature over time is a measure of total heat resulting from the process being studied (*Figure 15* and *Figure 16*).

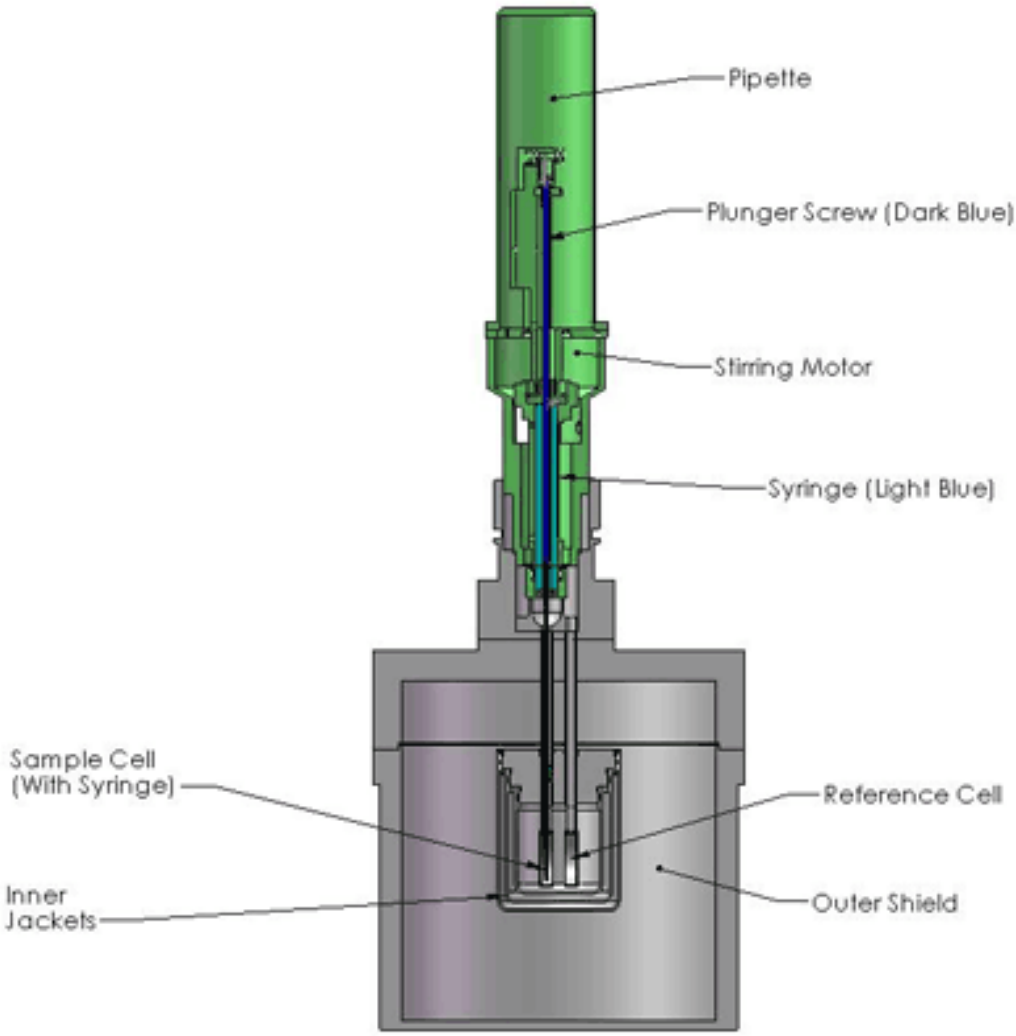

Figure 15: *representation of the ITC cells and syringe. During the experiment the syringe rotates and the conformation of its end provides continuous mixing in the cell. The computer-controlled plunger injects precise volumes of ligand.*

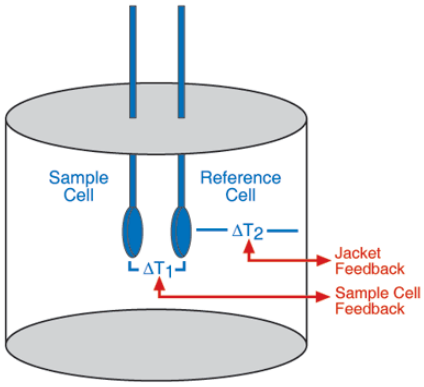

Figure 16: *cell feedback network system (CFB). A thermoelectric device measures the temperature difference between the two cells and a second device measures the temperature difference between the cells and the jacket.*

ITC experiments were performed as hereafter described. A 50 bases oligonucleotide carrying the region involved with the binding to folic acid was synthesized by Metabion International AG (Martinsried, Germany). The sequence is hereafter reported:

5'- CCGCGGCACAGUGUGUGAUGAUAUGU
UACGCACUCAUAGGGCUGUUGUAG - 3'

A 10 µM solution of RNA was prepared by dissolving it in *structural buffer* (1X solution: 10 mM Tris pH 7, 0.1 M KCl, 10 mM $MgCl_2$), and $NaHCO_3$ pH 10 (0.5 mM final concentration) was added to it. Folic acid was solubilized in the same buffer at the final concentration of 150 µM. The RNA was heated to 95 °C in water and slowly renatured with the cooling of water to room temperature. The RNA and folic acid solutions were degassed and loaded in the cell and in the syringe. Total volume RNA in the cell: ~1.5 ml. Total volume folic acid in the syringe: ~280 µl. The ITC experiments were carried out at 25 °C. The injection syringe rotated at ~400 rpm and the time interval between the injections was 240 seconds. The duration of each injection was 30 seconds.

Experiments using the same conditions were also performed with a portion of the Trans-activation Responsive Element (TAR) in place of RNA aptamer to check the selectivity of the binding of folic acid. The used TAR sequence is hereafter reported:

5'- GGCAGAUCUGAGCCUGGGAGCUCUCUGCC- 3'

Raw data were corrected by subtraction of the heat arising from the dilution of ligand into buffer. For this purpose, experiments were carried out in absence of RNA, with the simple solution of structural buffer with $NaHCO_3$ in the cell.

Curve fitting of the experimental ITC data were performed using Origin 7.0 software adapted for ITC data analysis.

4. RESULTS

4.1 SELECTION OF FOLIC ACID-BINDING RNA APTAMERS

The SELEX process started with a random sequence library obtained from combinatorial chemical synthesis of DNA. Each member of the library was a linear oligomer of a unique sequence. The presence of random region leads to the presence of 10^{15} different DNA molecules that guarantee the success of finding molecules with the ability to interact specifically with the target of interest. The ssDNA pool presented 85 randomised positions, flanked by two conserved regions for PCR amplification (see *Experimental Procedures* section). The selection was accomplished by affinity chromatography, and in particular immobilizing the folic acid to the epoxy-activated Sepharose 6B. The DNA pool was transcribed and the obtained radiolabeled RNA was loaded in the column. After incubation for 30 minutes, the column was washed to eliminate the unbound oligonucleotides. On the contrary, all the bound RNA molecules were eluted by competition adding a more concentrated folic acid solution to the column. These last fractions were precipitated and purified. Reverse transcription and PCR amplification permitted to start again with a new cycle of selection. The yield of each cycle was determined as the percentage of radioactivity of fractions containing the oligonucleotides of interest upon the total radioactivity of the initial loaded pool.

During the process, some expedients were adopted to enhance the stringency of the selection and therefore increase the specificity of the selected pool. In particular the number and the volume of washes with Binding Buffer was increased, the concentration of folic acid in the matrix reduced and at the same time the amount of RNA loaded in the column increased. Before cycle 1 and 6, precolumn cycles were performed in order to avoid the unspecific binding of the aptamers to the resin. To do so, the pool was loaded in the column in which the Sepharose matrix was not functionalized with folic acid: the first two fractions eluted, which contained the aptamers with the lowest affinity for the resin, were used for the selection with folic acid. *Table 1* summarizes all the conditions adopted for each cycle.

In *Figure 17* it is possible to see the profile of the yields with the SELEX progress. The first four cycles were characterized by an almost null value. This is an expectable outcome. In fact initially there are not many sequences able to bind the target. It is only after three or four cycles of amplification of the few best molecules that the yield begins to increase. In this case the exponential enrichment started with the fifth cycle and the following preselection allowed

to remove all the non-specific binding molecules and to further increase the yield of the sixth step. The decreasing outline of seventh and eighth cycles was probably due to the preparation of new chromatographic matrix and buffer solutions. This demonstrates that even minor changes in used conditions can affect the selection process.

The SELEX was terminated after the twelfth cycle: in fact at this point the obtained yield was similar to the yield of the eleventh cycle (17,4% compare to 17,7%) meaning that the twelfth cycle produced no further increase in pool affinity. The enriched pool coming from the eleventh cycle was cloned.

Table 1: *scheme of the SELEX cycles.*

Cycle Number	*Amount of Loaded RNA (nanomoles)*	*Folic Acid Concentration in the Matrix (mM)*	*Wash Fractions n° x V (μl)*	*Cycle Yield %*
Preselection				
1	12	8	10 x 1000	≤ 0.5%
2	0,7	8	10 x 100	≤ 0.5%
3	0,7	1,7	10 x 300	0,6%
4	0,95	1,7	10 x 300	≤ 0.5%
5	1,9	1,7	10 x 600	1,75%
Preselection				
6	1,5	1,7	10 x 600	12,2%
7	1,9	1,7	15 x 600	10,5%
8	1,74	1,7	22 x 600	4,19%
9	2,46	1,7	22 x 600	10,8%
10	2,47	1,5	22 x 600	12,7%
11	4,2	1,5	22 x 600	17,7%
12	4,96	1,5	22 x 600	17,4%

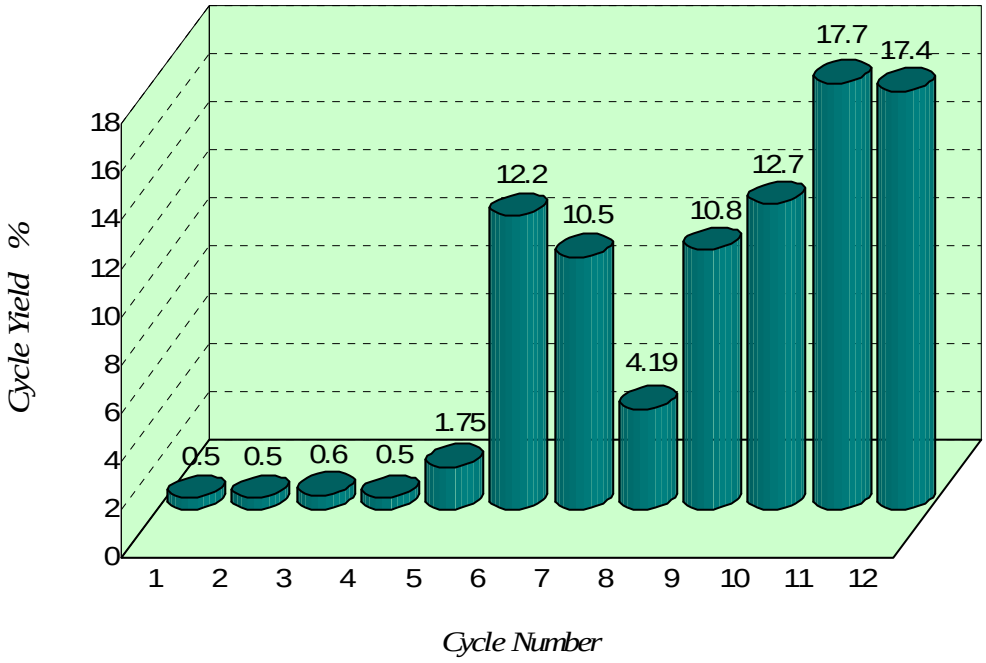

Figure 17: *graphical representation of the yield of the in vitro selection of RNA aptamers against folic acid.*

4.2 SUBCLONING AND SEQUENCING OF THE SELECTED APTAMERS

After the selection, the pool of aptamers was PCR amplified, subcloned and sequenced. For this purpose the pBluescriptIIKS(+) plasmid vector was chosen. The plasmid vector carries the gene encoding resistance to ampicillin and contains a multiple cloning site (MCS) that includes more than 20 restriction endonuclease sites. The sites are designed to be unique within the vector sequence so that cutting the vector with a restriction endonuclease and cloning foreign DNA into this site do not disrupt other critical features of the vector. The sequences that directly flank the MCS are useful for manipulation or analysis of inserted DNA. In fact for these sequences there are commercially available complementary oligonucleotides that can be used for priming PCR or DNA sequencing reactions, and therefore such primers are tools for amplification or sequencing of any DNA fragment inserted into the MCS.

Aptamer pool coming from the eleventh cycle was amplified using primers that permitted the insertion of EcoRI and BamHI restriction endonuclease sites. Afterwards both plasmid vector and amplified DNA were treated with the same endonucleolitic enzymes, generating sticky ends ready to be ligated one on the other by T4 Ligase. The recombinant plasmids were then inoculated in *Escherichia coli* DH5α cells and these spread on LB-agar plates containing ampicillin and incubated at 37° C overnight. Single colonies were picked up and each one

amplified in 5 ml of LB medium with ampicillin. Plasmid DNAs were then extracted and purified from bacterial solutions and a careful analysis by colony PCR was performed to identify which clones carried the recombinant plasmid. In particular, taking advantage of FA1 and FA3 primers, all the extracted plasmids were amplified. Annealing of primers and PCR process is possible only in plasmids containing the aptamer sequence, thus resulting in the appearing of a band around 150 bp in agarose gel. Out of all the picked colonies, 48 demonstrated to contain the desired length of aptamer sequence (*Figure 18*).

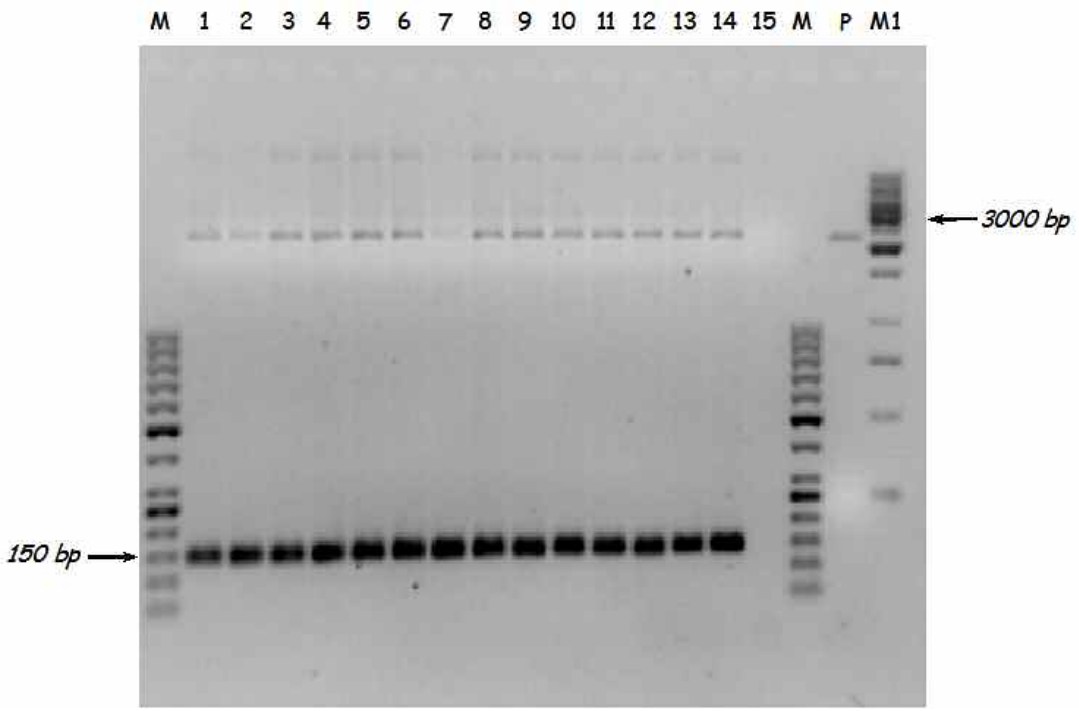

Figure 18: *the plasmid DNAs extracted from colonies were amplified by PCR using FA1 and FA3 primers. The reaction permits to amplify only the aptamer sequence if correctly inserted in the vector. This 1.5% agarose gel shows the appearance of bands around 150 bp in 14 out of the 15 here analyzed colonies. Consequently only the 15th bacterial clone does not contain the aptamer sequence. M: 50 bp ladder. M1: 1 Kb ladder. P: circular plasmid amplified as control.*

The plasmid DNAs that demonstrated to contain the sequence of interest were sequenced using the *M13 forward* primer that binds a complementary sequence placed about 85 bases upstream the restriction site for BamHI. The analysis led to obtain the antisense sequence of the original selected RNA. Therefore all the sequences were transformed in the complementary strands and analyzed with computer tools to see if a common consensus motif could be found. Usually only a small part (approximately from 15 to 25 bases) along the entire random region is responsible for the target binding and this portion could be highly conserved in all the selected aptamers. The determination of the minimum binding sequence

is therefore the main aim of all the SELEX processes because it represents the minimum component to develop for aptamer application in the desired system.

It is important to notice that the computational investigation was not exhaustive and performed essentially to ease and direct the following molecular biology steps necessary for the characterization of the selected aptamers.

The MultAlin program was used for this purpose. It examines the analogy between sequences, giving a percentage of the homology of them and highlighting the conserved feature in the aligned RNA. A consensus bigger than 90% is reported with red bases and a consensus bigger than 50% with blue bases. All the other non conserved bases are black. Sequence alignments were carried out for all molecules only on the randomized portion, since the annealing sequences were present in every molecule of the initial pool and likely were not responsible for the binding specificity of the oligonucleotides, although they could somehow give a contribution to the folding (*Figure 19*).

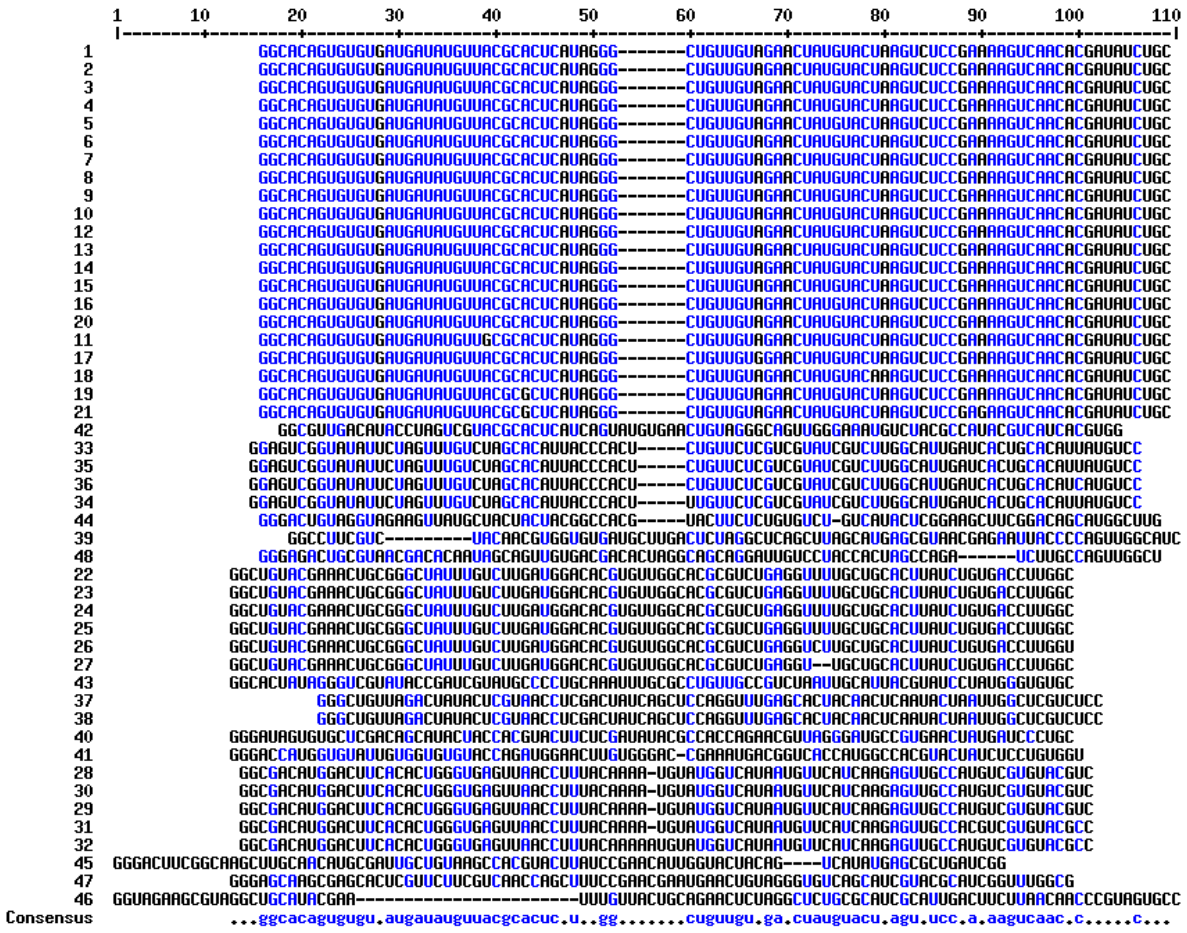

Figure 19: *analysis with MultAlin of all the RNA sequences obtained from sequencing. In blue is reported a consensus bigger than 50% between bases in that position.*

The first analysis revealed the presence not of a real consensus motif between all the sequences but it demonstrated the presence of completely identical sequences repeated many times. Therefore it was decided to classify all the sequences in families, each family containing all the clones carrying the same sequence. Inside the same family, only limited nucleotide substitutions, deletions or insertions were permitted. In this way, the 48 clones were divided in a total of 15 families. Family **A** contained 21 sets of almost identical sequences, family **B** 6 sets, family **C** 5 sets, family **D** 4 sets, family **E** 2 sets and all the remaining families contained on the contrary only one unique sequence. The sequence of each obtained family of aptamers is hereafter reported (5' to 3').

*Family **A**, clones with sequence:*

GGCACAGUGUGUGAUGAUAUGUUACGCACUCAUAGGGCUGUUGUAGAACUAUGUACUAAGUCU
CCGAAAAGUCAACACGAUAUCUGC

*Family **B**, clones with sequence:*

GGCUGUACGAAACUGCGGGCUAUUUGUCUUGAUGGACACGUGUUGGCACGCGUCUGAGGUUUU
GCUGCACUUAUCUGUGACCUUGGC

*Family **C**, clones with sequence:*

GGCGACAUGGACUUCACACUGGGUGAGUUAACCUUUACAAAAUGUAUGGUCAUAAUGUUCAUC
AAGAGUUGCCAUGUCGUGUACGUC

*Family **D**, clones with sequence:*

GGAGUCGGUAUAUUCUAGUUUGUCUAGCACAUUACCCACUCUGUUCUCGUCGUAUCGUCUUGGC
AUUGAUCACUGCACAUUAUGUCC

*Family **E**, clones with sequence:*

GGGCUGUUAGACUAUACUCGUAACCUCGACUAUCAGCUCCAGGUUUGAGCACUACAACUCAAUA
CUAAUUGGCUCGUCUCC

*Family **F**, clone with sequence:*

GGCCUUCGUCUACAACGUGGUGUGAUGCUUGACUCUAGGCUCAGCUUAGCAUGAGCGUAACGA
GAAUUACCCCAGUUGGCAUC

*Family **G**, clone with sequence:*

GGGAUAGUGUGCUCGACAGCAUACUACCACGUACUUCUCGAUAUACGCCACCAGAACGUUAGGG
AUGCCGUGAACUAUGAUCCCUGC

*Family **H**, clone with sequence:*

GGGACCAUGGUGUAUUGUGGUGUGUACCAGAUGGAACUUGUGGGACCGAAAUGACGGUCACCA
UGGCCACGUACUAUCUCCUGUGGU

*Family **I**, clone with sequence:*

GGCGUUGACAUACCUAGUCGUACGCACUCAUCAGUAUGUGAACUGUAGGGCAGUUGGGAAAUG
UCUACGCCAUACGUCAUCACGUGG

*Family **L**, clone with sequence:*

GGCACUAUAGGGUCGUAUACCGAUCGUAUGCCCCUGCAAAUUUGCGCCUGUUGCCGUCUAAUUG
CAUUACGUAUCCUAUGGGUGUGC

*Family **M**, clone with sequence:*

GGGACUGUAGGUAGAAGUUAUGCUACUACUACGGCCACGUACUUCUCUGUGUCUGUCAUACUC
GGAAGCUUCGGACAGCAUGGCUUG

*Family **N**, clone with sequence:*

GGGACUUCGGCAAGCUUGCAACAUGCGAUUGCUGUAAGCCACGUACUUAUCCGAACAUUGGUAC
UACAGUCAUAUGAGCGCUGAUCGG

*Family **O**, clone with sequence:*

GGUAGAAGCGUAGGCUGCAUACGAAUUUGUUACUGCAGAACUCUAGGCUCUGCGCAUCGCAUU
GACUUCUUAACAACCCGUAGUGCC

*Family **P**, clone with sequence:*

GGGAGCAAGCGAGCACUCGUUCUUCGUCAACCAGCUUUCCGAACGAAUGAACUGUAGGGUGUCA
GCAUCGUACGCAUCGGUUUGGCG

*Family **Q**, clone with sequence:*

GGGAGACUGCGUAACGACACAAUAGCAGUUGUGACGACACUAGGCAGCAGGAUUGUCCUACCAC
UAGCCAGAUCUUGCCAGUUGGCU

The first evidence of the subdivision in families was the fact that family **A** was characterized by the highest presence of clones having that specific sequence. The percentage of homology inside this family (*Figure 20*), as for all the others (data not shown), was analyzed using MultAlin. Almost all the clones of family **A** carry the same sequence except some that present one or two single point mutations.

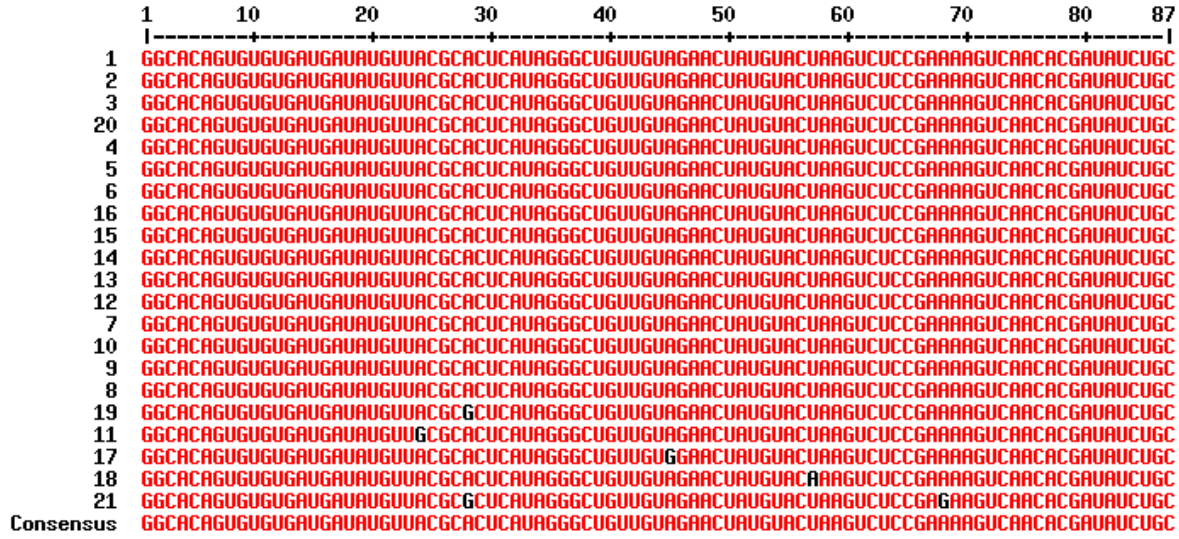

Figure 20: *analysis with Multalin of all the members of family A. In red is reported a consensus bigger than 90% between bases in that position. In black the single point substitutions carried by some clones.*

Afterwards, it was decided to take a single sequence for each family and perform an alignment study with the selected 15 sequences. The percentage of homology calculated with

this strategy was again not enough to determine a real consensus between families (*Figure 21*). The homology in fact did not involve long tracts of the oligonucleotides.

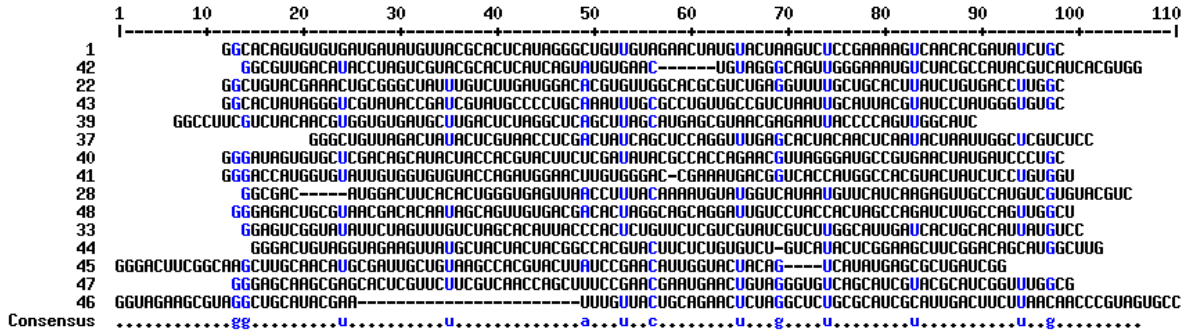

Figure 21: *alignment with MultAlin of a single member of each family. In blue is reported a consensus bigger than 50% between bases in that position. In black all the not conserved bases.*

Evident was the presence of different lengths of random region. The initial pool used for the SELEX process presented 85 randomized bases. After sequencing it was possible to identify clones having from 81 to 88 randomized bases, resulting from mistakes during the numerous rounds of amplifications.

These results allowed to conclude that, even if no single consensus sequence conserved among the entire population of the selected clones was discovered, 21 clones on a total of 48 were carrying the same sequence. These were clones belonging to family **A** and presumably they carried the sequence with highest affinity for folic acid, being more amplified during the column selection. Therefore the following step was performed only with the RNA sequence of family **A**.

4.3 THREE-DIMENSIONAL FOLDING AND MINIMUM BINDING SEQUENCE DETERMINATION

The alignment of the selected sequences using computer tools was not able to identify a common region in the RNA primary sequence presumably involved in the binding to folic acid. Therefore an RNase protection protocol was developed to determine at the same time the three-dimensional folding of the oligonucleotides and the sequence responsible for the interaction with target. Family **A** was considered the most important group and it was chosen for this type of analysis. The direct action of RNase V1, RNase T1 and RNase A was used to determine the possible folding of aptamers. The action of the same three nucleases on the oligonucleotide preincubated with folic acid was used to understand the binding site. In theory

after the interaction of the two elements, a small portion of the RNA is protected by folic acid and not subjected to cleavage by the enzymes.

The plasmid DNA of interest was restricted with EcoRI and BamHI and the aptamer sequence separated and purified from the vector, and finally amplified by PCR using primers FA1 and C2 that allow the insertion in the appropriate position of the T7 RNA polymerase promoter, previously removed during subcloning step. The obtained DNA was transcribed and, after removal of the phosphate in 5' end resulting from the transcription, radiolabeled with [γ-^{32}P]ATP. After PAGE purification, the RNA was ready to be tested on the RNase protection assay.

RNase V1, RNase T1 and RNase A were selected for their cutting specificity. RNase V1 cuts with non-sequence specificity double stranded RNA regions and tertiary structure as pseudoknots. RNase T1 cuts specifically guanines in single stranded RNA regions. RNase A is specific for cytosines and uridines in single stranded RNA regions.

Alkaline hydrolysis allowed to obtain a marker of all the bases of the RNA. In fact the presence of the 2' OH of the RNA ribose ring permits in particular conditions the removal of a single nucleotide at a time, giving a ladder essential for the determination of the sequence.

The sequence of two clones belonging to family **A** was analyzed using this technique. Sequence 4 represents the general sequence of clones belonging to family **A**. Sequence 21 is also part of this family but it carries small differences and in particular two adenines are substituted with guanines (see black bases in *Figure 20*). The obtained results for each considered member of family **A** are hereafter reported.

Sequence 21. In the same sequencing gel, samples treated in presence and in absence of folic acid were loaded. In this way it was possible to determine at the same time both the target binding site and the three-dimensional folding. The identification of the region responsible for the interaction with folic acid is straightforward. In fact, the inhibition effect on endonucleolitic cleavage, due to the presence of folic acid bound to the aptamer, results in the disappearance of bands in sequencing gels. Several gels were run for different times, allowing a good separation and analysis of all the regions of the aptamer sequence. Here are reported only illustrative images. Instead the text summarizes all the obtained results.

After incubation with folic acid the signal corresponding to action of RNase T1 on bases G21, G23, G35, G49 and G57 clearly disappeared (compare lane g with lane l, *Figure 22*), showing the interaction of the target with these guanines. The cleavage by RNase V1 can demonstrate both interaction with folic acid and double stranded secondary structure. U24, G25, U26 and G27 were cleaved by the enzyme confirming their involvement in a stem

region (lanes p, q, t and u, *Figure 22*). They seemed to be not protected by the presence of folic acid (compare lane q with lane u, *Figure 22*). Contrary to that, G13, G15, G16, A18, C19, A20, G21 and U22 were protected by its presence (compare lane p and lane t, *Figure 22*). Finally, it was possible to detect also a partial protection of G54 and U55 (compare lane q and lane u, *Figure 22*). The action of RNase A on cytosines and uridines allowed to define a clear interaction between C45 and U47 of RNA and target (compare lane xx with lane zz, *Figure 22*). In addition C11, C12, C14, C17, C19, U22 and U24 were protected from cleavage (compare lane w with lane yy or lane x with lane z, *Figure 22*). The action of RNase A on G15 and G23 was a non-specific cutting effect and therefore it was not considered for the analysis (lane yy, *Figure 22*).

Some bases were cleaved by more than one nuclease. For example G57 was cleaved by both RNase V1 and T1, and C19 by both RNase V1 and A. This could be due to the presence of pseudoknots or loops in proximity of these bases, determining their participation in both single and double stranded secondary structures.

In *Figure 24* is reported the sequence of RNA obtained from transcription of clone 21. The yellow highlighted bases showed interaction with folic acid in the RNase protection assay. The two guanines in red are instead the single point mutations carried by this aptamer compare to the general sequence of family **A**.

Sequence 4. The pattern obtained with this aptamer is overall similar to that obtained with sequence 21. The action of RNase T1 was completely blocked in position G49 and partially reduced in position G57 by the presence of folic acid (compare lane g and lane l, *Figure 23*). After treatment with RNase V1, G13, C14, G15, G16, C17, C19, G21, U22, G23, C40, C45, G54 and U55 also demonstrated to be protected and therefore to interact with the target (compare lane p and lane t, *Figure 23*). RNase V1 was able to cut also in positions U24, G25, U26 and G27, confirming the presence of this portion in a stem secondary structure (lanes p and t, *Figure 23*). Finally C45 and U47 were protected from cleavage by RNase A in presence of folic acid (compare lane xx and lane zz, *Figure 23*). This experiment showed as the presence of the target increased cleavage by nucleases in positions between G57 and C101 (for example compare lane p with lane t or lane xx with lane zz, *Figure 23*). This can be ascribed to the fact that, as a portion of the oligonucleotide is protected, nucleases direct their activity on the remaining free part, determining a shift of their effect on the accessible sites.

Figure 25 reports a summary of the obtained results for sequence 4. The yellow highlighted bases showed interaction with folic acid in the RNase protection assay.

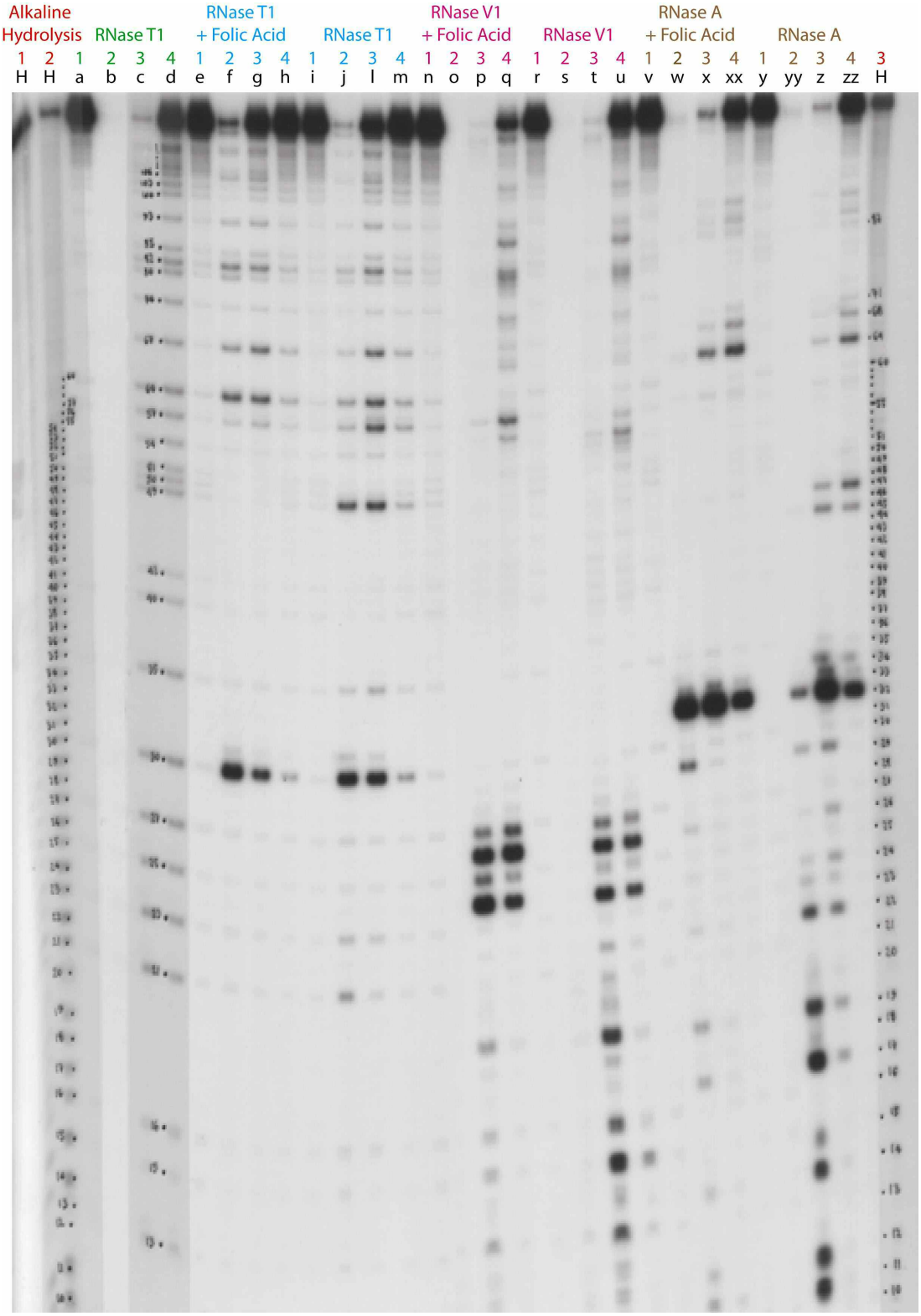

Figure 22: *RNase protection assay on sequence 21. 10% polyacrylamide sequencing gel, run for 1 hour and a half, 75 W. The samples are treated as described in Experimental Procedures section. Alkaline hydrolysis time (lanes* ***1H***, ***2H***, ***3H*** *in* ***red****): 7.5 minutes. In* ***green****: denatured RNA treated with RNase T1 for the determination of all the guanines present along the sequence. In* ***light blue****: folded RNA treated with RNase T1. In* ***purple****: folded RNA treated with RNase V1. In* ***brown****: folded RNA treated with RNase A. Samples treated with RNase:* ***1)*** *samples not treated;* ***2)*** *samples treated with endonucleases' stock solution;* ***3)*** *samples treated with a 1/10 dilution of endonucleases' stock solution;* ***4)*** *samples treated with a 1/100 dilution of endonucleases' stock solution. When present, folic acid is in a ratio 10000:1 with RNA.*

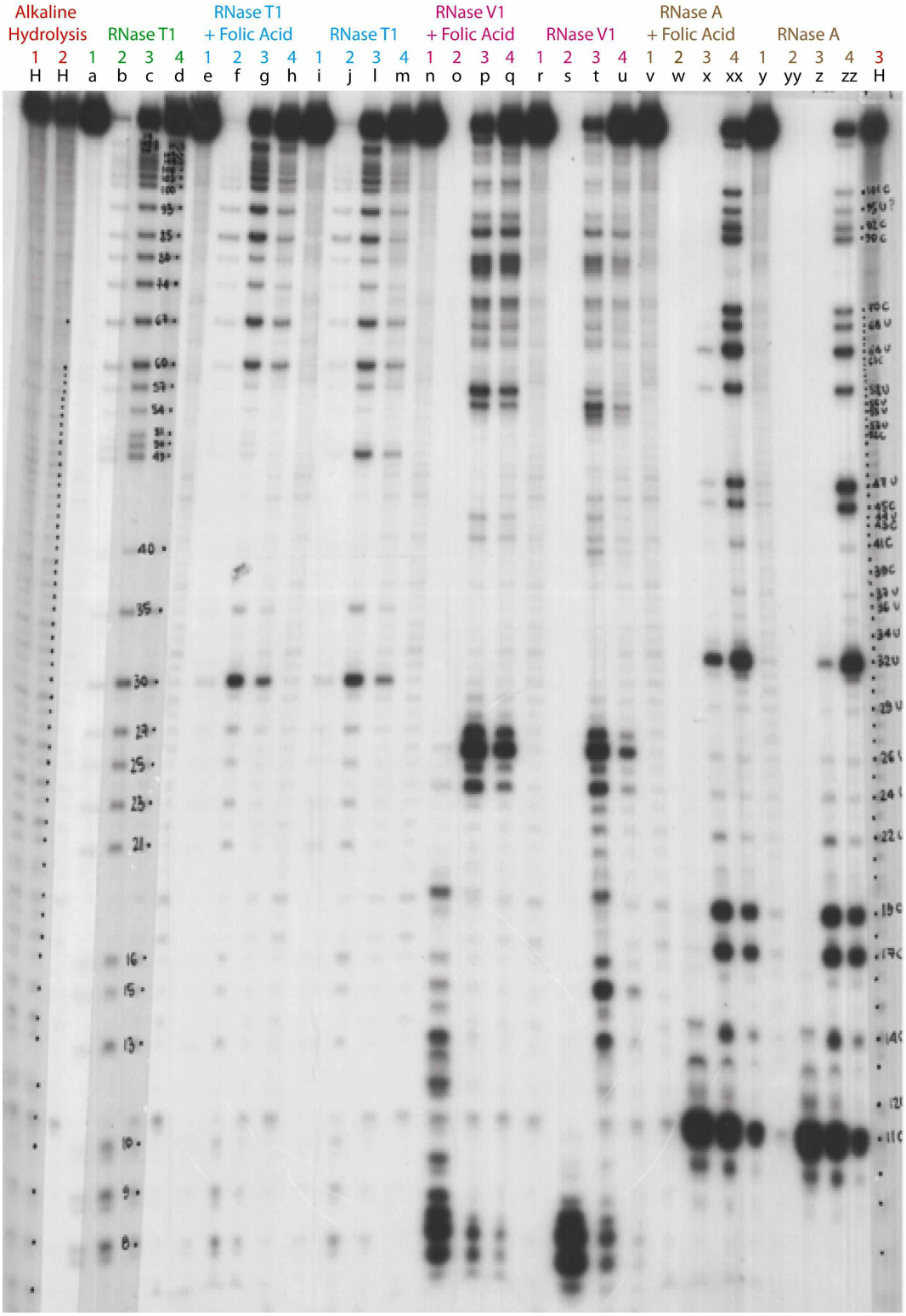

Figure 23: *RNase protection assay on sequence 4. 10% polyacrylamide sequencing gel, run for 1 hour and a half, 75 W. The samples are treated as described in Experimental Procedures section. Alkaline hydrolysis time:* ***1H*** *4.5 minutes****;*** ***2H*** *5.5 minutes****;*** ***3H*** *6.5 minutes. In* ***green****: denatured RNA treated with RNase T1 for the determination of all the guanines present along the sequence. In* ***light blue****: folded RNA treated with RNase T1. In* ***purple****: folded RNA treated with RNase V1. In* ***brown****: folded RNA treated with RNase A. Samples treated with RNase:* ***1)*** *samples not treated;* ***2)*** *samples treated with endonucleases' stock solution;* ***3)*** *samples treated with a 1/10 dilution of endonucleases' stock solution;* ***4)*** *samples treated with a 1/100 dilution of endonucleases' stock solution. When present, folic acid is in a ratio 10000:1 with RNA.*

Essentially, it is possible to summarize all the obtained data as hereafter reported. All the analyzed sequences demonstrated interaction with folic acid in the region between residue 11 and residue 24. In addition, other residues resulted implicated in the binding. Noteworthy are residues G35, G40, C45, U47, G49, G54, U55 and G57. Finally, the region including at least residues from 24 to 27 was folded as double strand. The interaction was mainly close to the 5' end of the aptamer, without involvement of the 3' end. The presence of single point mutations in sequence 21 did not affect the binding property of the oligonucleotide because these guanines were not involved directly in the binding.

5' - GGAGAAAGGG$_{10}$CCGC[GGCACA$_{20}$GUGUGUGAUG$_{30}$AUAUGUUACG$_{40}$C GCUCAUAGG$_{50}$GCUGUUGUAG$_{60}$AACUAUGUAC$_{70}$UAAGUCUCCG$_{80}$AGAAG UCAAC$_{90}$ACGAUAUCUG$_{100}$C]AGAAGAGGU$_{110}$GACGGGAUCC$_{120}$AAGCCU - 3'

Figure 24: *RNA sequence 21. Highlighted in yellow the bases that showed interaction with folic acid in the RNase protection assay (Figure 22). Single point mutations are reported in red. Between blue square brackets the aptamer region. Outside, the portion of RNA coming from transcription of primers.*

5' - GGAGAAAGGG$_{10}$CCGC[GGCACA$_{20}$GUGUGUGAUG$_{30}$AUAUGUUACG$_{40}$C ACUCAUAGG$_{50}$GCUGUUGUAG$_{60}$AACUAUGUAC$_{70}$UAAGUCUCCG$_{80}$AAAAG UCAAC$_{90}$ACGAUAUCUG$_{100}$C]AGAAGAGGU$_{110}$GACGGGAUCC$_{120}$AAGCCU - 3'

Figure 25: *RNA sequence 4. Highlighted in yellow the bases that showed interaction with folic acid in the RNase protection assay (Figure 23). Between blue square brackets the aptamer region. Outside, the portion of RNA coming from transcription of primers.*

Sequencing gels, together with the use of Kinefold program (http://kinefold.curie.fr/), a computer tool for secondary structure prediction, allowed to define the more favored structure, in terms of stabilization energy (ΔG value in Kcal/mol). We assembled all the obtained data overlapping the computer structural information and the electrophoretic information to obtain the more accurate outcome. The chosen parameters to predict with Kinefold the possible folding were essentially "renaturation folding", that mimics the renaturation that was used in our *in vitro* system, and "pseudoknots allowed", giving the possibility to the RNA to fold in tertiary structures.

The computer program determined a set of possible foldings, with correlated energy values, for both sequences 4 and 21. Therefore the purpose of the next step was to define, helped by the results obtained with RNase protection assay, which were the real aptamer structures in our system. *Figure 26* shows the more convincing example for both sequences. It

is possible to notice that the structure is conserved and the presence of two guanines in sequence 21 does not change the general folding. In particular, guanine as well as adenine in position 42 are able to pair uridine and form a stem region. The pairing between a guanine and a uridine is called “wobble GU”, a non Watson-Crick pairing with a good thermodynamic stability and a greater flexibility and affinity for metal ions [33, 34]. Bases U24, G25, U26 and G27 are present in double stranded structure as determined by electrophoretic analysis. In addition, also the region between residues G13 and C17 and residues G54 and U55 are folded in stem conformation.

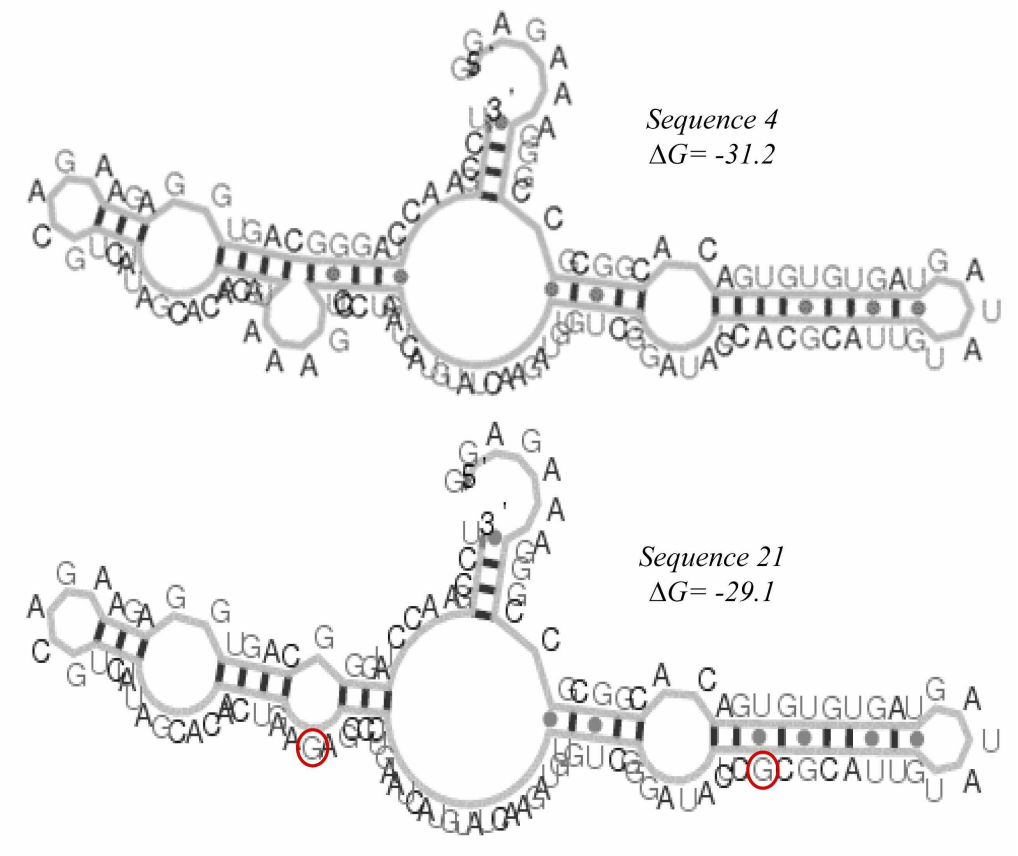

Figure 26: *prediction with Kinefold program of secondary structures adopted by sequence 4 and sequence 21. The ΔG values (Kcal/mol) associated to each folding are also reported. The two single point mutations carried by sequence 21 are encircled in red.*

Taken together, these results allow to define a specific aptamer region responsible for the interaction with folic acid. *Figure 27* correlates data obtained with RNase protection assay with the secondary structure of the oligonucleotides predicted with Kinefold. The minimum area is formed by a stem of five bases, an asymmetric loop with two bases in a side and five bases on the other side, and an other stem of about six bases in length.

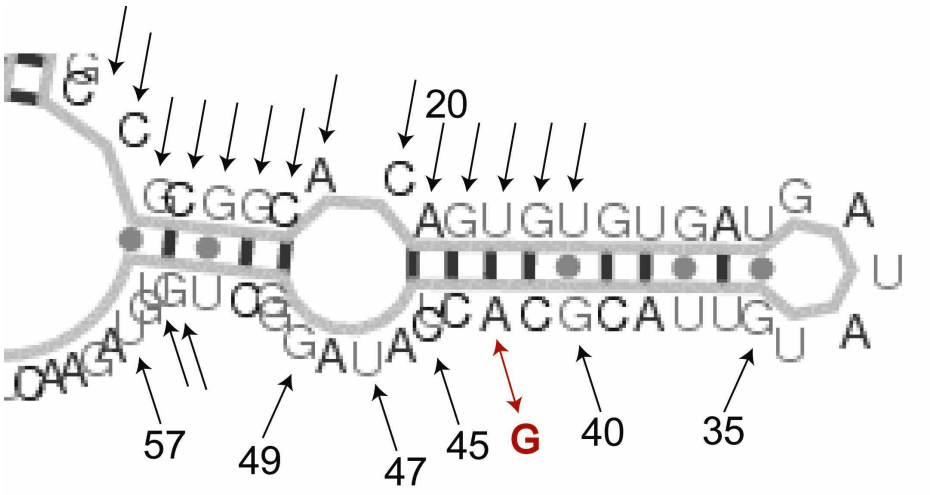

Figure 27: *correlation between the region of interaction identified with RNase protection assay and the secondary structure predicted by Kinefold.*

4.4 DETERMINATION OF APTAMER AFFINITY CONSTANT

The results obtained with RNase protection assay, in association with computer outcomes, were promising. In fact it was possible to define an unambiguous aptamer region involved in the binding to folic acid. An RNA sequence of 50 bases, including the just mentioned area of interest, was synthesized and used for the determination of the binding constant between oligonucleotide and target. For this purpose the Isothermal Titration Calorimetry (ITC) (MicroCal, GE Healthcare) was utilized. This thermodynamic technology allows to measure biomolecular interactions, determining all binding parameters in a single experiment. When substances bind, heat is either generated or absorbed. ITC directly detects the heat released or absorbed during a biomolecular binding event, determining the binding constant (K_B), the reaction stoichiometry (n), the enthalpy (ΔH) and the entropy (ΔS), providing a complete thermodynamic profile of the molecular interaction.

Kinefold was used to make sure that the 50 bases RNA oligonucleotide, used for the determination of the binding constant, was folded as the full length aptamer used in RNase protection assay. Afterwards, experiments were performed as reported in *Experimental Procedures* section and *Figure 28* shows the obtained thermodynamic profile. The best-fit parameters associated to the curve are reported on the top graph of the just mentioned figure.

"n" allows to define the stoichiometry of the reaction. A value of 0.5 leads to conclude that a single molecule of folic acid binds two molecules of RNA. Considering that the constant of binding (K_B) corresponds to 1/Kd, the dissociation constant of the complex is 602

± 46 nM. The enthalpy variation and the progress of the curve allow to define that the reaction leading to the formation of the complex is exothermic.

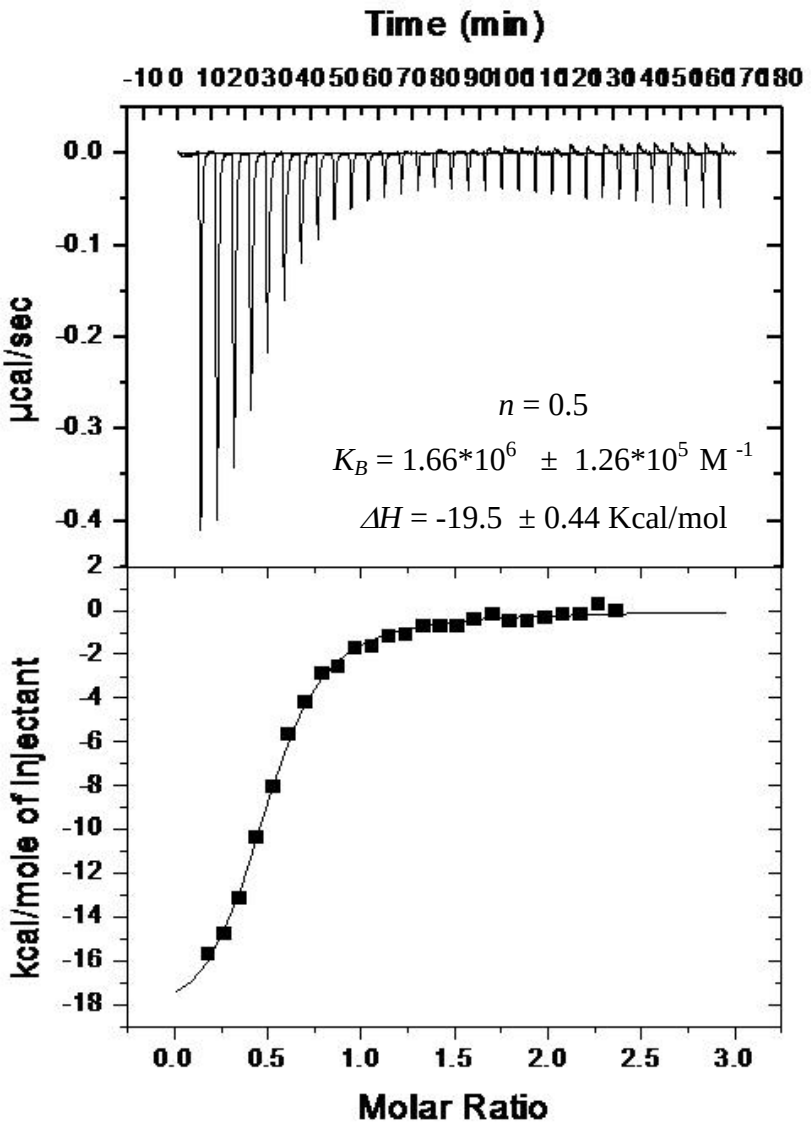

Figure 28: *ITC curve of the binding of folic acid to aptamer. Top graph shows raw data in terms of µcal/second plotted against time in minutes, after the integration baseline has been subtracted. The graph reports the best-fit parameters associated to the curve. n: stoichiometry of the reaction; K_B: constant of binding; ΔH: enthalpy variation. The bottom graph shows normalized integration data in terms of kcal/mole of injectant plotted against molar ratio. 10 µM RNA solution was loaded into a 1.5 ml sample cell and a 150 µM folic acid solution was loaded into the 280 µl injection syringe. 8 µl was injected in each time at the stirring rate of ~400 rpm.*

To demonstrate that the obtained binding constant is real and not related to an aspecific interaction of folic acid with a general RNA, ITC experiments were performed in presence of Trans-activation Responsive Region (TAR) in place of the aptamer RNA. TAR is a short RNA structure located at the 5' end of all nascent HIV-1 transcripts. Its functions and three-dimensional folding are well known and this is the reason why this RNA was chosen to

perform the analysis. A 29 bases portion of TAR was chemically synthesized and loaded in the cell for the titration with folic acid. From the thermodynamic profile obtained with ITC for this reaction was not possible to determine a constant of binding, proving the inability of folic acid to bind a folded RNA with a sequence non specific for folic acid (*Figure 29*).

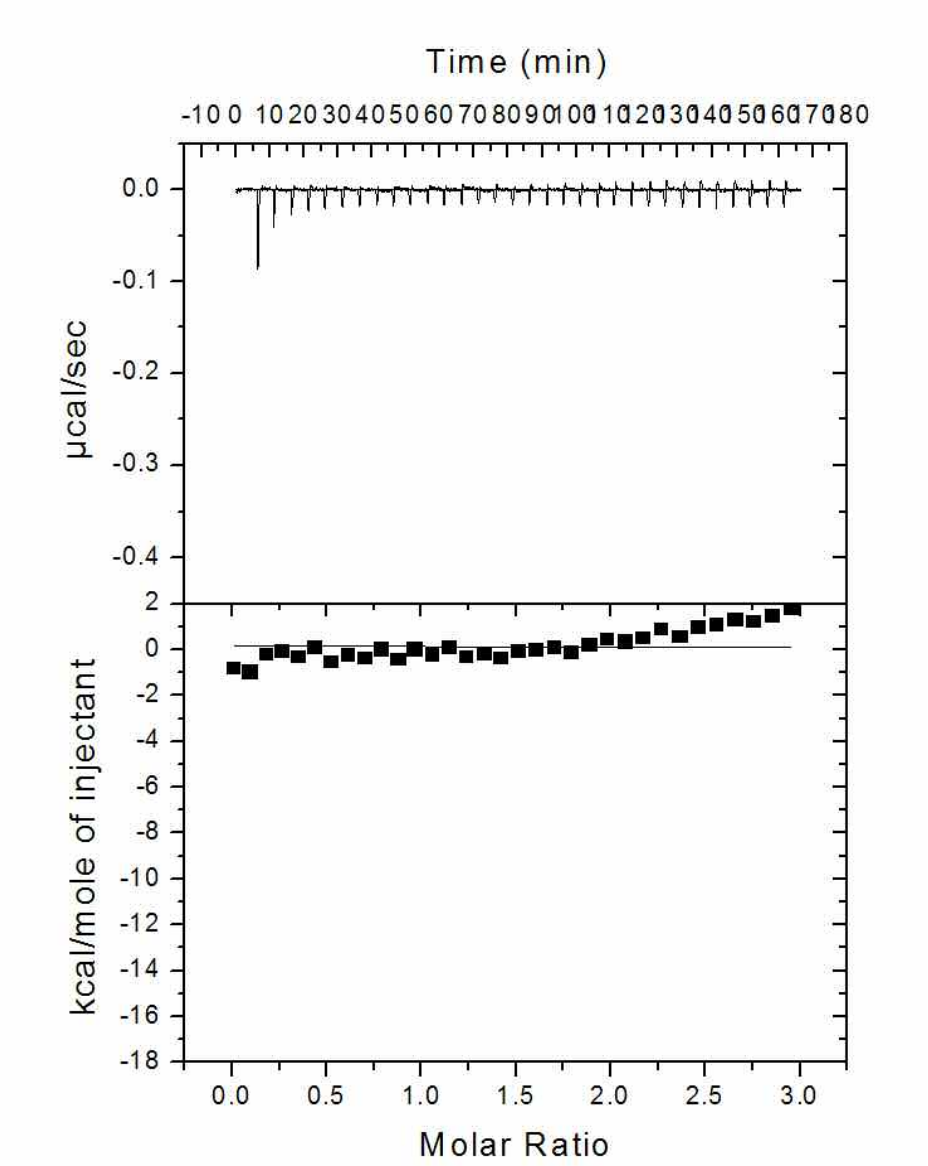

Figure 29: *ITC curve of the binding of folic acid to TAR. Top graph shows raw data in terms of µcal/second plotted against time in minutes, after the integration baseline has been subtracted. The bottom graph shows normalized integration data in terms of kcal/mole of injectant plotted against molar ratio. 10 µM RNA TAR solution was loaded into a 1.5 ml sample cell and a 150 µM folic acid solution was loaded into the 280 µl injection syringe. 8 µl was injected in each time at the stirring rate of ~400 rpm.*

5. CONCLUSIONS

The described project was directed to the application of the *in vitro* selection methodology in the discovery of biopolymer binding folic acid to be developed as diagnostic and therapeutic tool.

The selection by affinity chromatography allowed to identify a pool of aptamers binding folic acid, a molecule involved in many physiological and pathological processes. In particular after 12 cycles of selection and amplification, a satisfying enrichment was obtained and therefore the pool subcloned and sequenced. A basic computer analysis did not determine a real consensus between all the sequences but suggested a subdivision of the clones in families. Some sequences were in fact repeated many times and the subdivision allowed to group together all the clones carrying the same aptamer sequence, and only limited nucleotide substitutions, deletion or insertion were permitted for each family. Family **A** was characterized by the highest presence of clones and presumably these are the sequences with highest affinity for folic acid, being more amplified during the column selection. The sequence of two clones of family **A** were analyzed by RNase protection assay to identify the three-dimensional folding and the minimum region involved in the interaction with folic acid. Kinefold program was used to confirm the data obtained by electrophoresys. The minimum binding sequence was defined as a stem of five bases, an asymmetric loop with two bases in a side and five bases on the other side, and another stem of about six bases in length. An RNA oligonucleotide of 50 bases in length including the just mentioned area of interaction, was chemically synthesized and used for Isothermal Tritration Calorimetry assays, obtaining a thermodynamic profile of the reaction that leads to the formation of the complex between aptamer and folic acid. A dissociation constant value of about 600 ± 46 nM confirmed the success of the performed SELEX. The stoichiometry of the reaction corresponded to 0.5, meaning that two molecules of RNA bound a single molecule of folic acid. The reaction was exothermic, with a ΔH value of -19.5 Kcal/mol. The inability of folic acid to bind a general RNA sequence (as TAR) confirmed the selectivity of the selected aptamer.

In conclusion it is possible to say that, an RNA aptamer able to bind folic acid with a really high affinity and selectivity was obtained. Additional studies will now be addressed to adapt the oligonucleotide to the desired application. Backbone substitutions are necessary to increase the stability to nucleases. Conjugation with a reporter could lead to the development of a useful aptamer beacon for the determination of the folic acid blood concentration. Additionally, conjugation with anticancer drugs could be used for the site-specific targeting of

the considered compounds. The possibility to adapt the selected aptamer to several purposes, justify the importance of the present work.

SECOND PART

RALTEGRAVIR AND ELVITEGRAVIR:

HOW DNA INTEGRATION IS A TARGET FOR DRUG

"The mammalian cell can be compared to a great metropolitan city, containing a central storehouse of information, power plants that generate energy, manufacturing districts, ports of entry and exit and an elaborate transportation system. Viruses, like daily commuters entering a city from the suburbs, must find their way into the cell, move to specific intracellular locations to carry out essential tasks and then, after replication, find their way back out of the cell again"

Stephen P. Goff

"Host Factors Exploited by Retroviruses"

Nature Reviews - Volume 5 - April 2007

6. INTRODUCTION

6.1 HUMAN IMMUNODEFICIENCY VIRUS

The Human Immunodeficiency Virus (HIV) is the causative agent of the Acquired Immune Deficiency Syndrome (AIDS), which, if untreated, is a fatal disease. About 33 million people worldwide are infected with the virus, with a rate of 2.5 million new infections every year [36]. Over past 25 years, this disease has become a leading cause of mortality worldwide and the main cause of death in sub-Saharan Africa.

There are two types of HIV, each evolved from a different Simian Immunodeficiency Virus (SIV): HIV-1 and HIV-2.

HIV-2 infections are mainly restricted to West-Central Africa although they have been characterized also in India and South Korea. The present work is addressed to the study of the most common type 1 and therefore all the information hereafter reported are related only to this particular family.

HIV-1 is much more prevalent and it is responsible for the AIDS pandemic. Its progenitor is the Simian Immunodeficiency Virus (SIV), which infected chimpanzee communities in southern Cameroon and was then transmitted to humans. Probably three transmissions took place independently, originating the three large groups into which HIV-1 is nowadays classified: M (*Major*), O (*Outlier*) and N (*Non Major-Non Outlier*). Phylogenetic analysis demonstrated the relation between M and N groups with strains found in chimpanzees, while for group O there are suggestions that probably it originated from gorillas, in which the closest relatives of this group have been identified. In particular the most convincing hypothesis proposes that chimpanzees were the original reservoir of SIVs. Distinct chimpanzee communities in southern Cameroon transmitted divergent SIVcpz to humans, giving rise to HIV-1 group M and N and transmitted HIV-1 group O-like viruses either to gorillas and humans independently, or to gorillas that then transmitted the virus to humans [37] (*Figure 30*).

Group M is the predominant, including the most of the HIV infections worldwide and it has been divided into subtypes denoted with letters (A, B, C, D, F, G, H, J and K) and sub-subtypes denoted with numbers (A1, A2, A3, A4, F1 and F2). There are also evidences that intersubtype recombinants are taking part in the pandemic. They result from recombination of subtypes within a dually infected person, and the recombinant form is then passed to other people. These variations are defined as Circulating Recombinant Forms (CRFs), if identified

in three or more people with no direct epidemiologic linkage, or Unique Recombinant Forms (URFs), if recovered from only a single person. In Europe, North-America and Oceania, subtype B of group M is the most prevalent one [38]. General characteristics of the subtypes are reported in *Figure 31*.

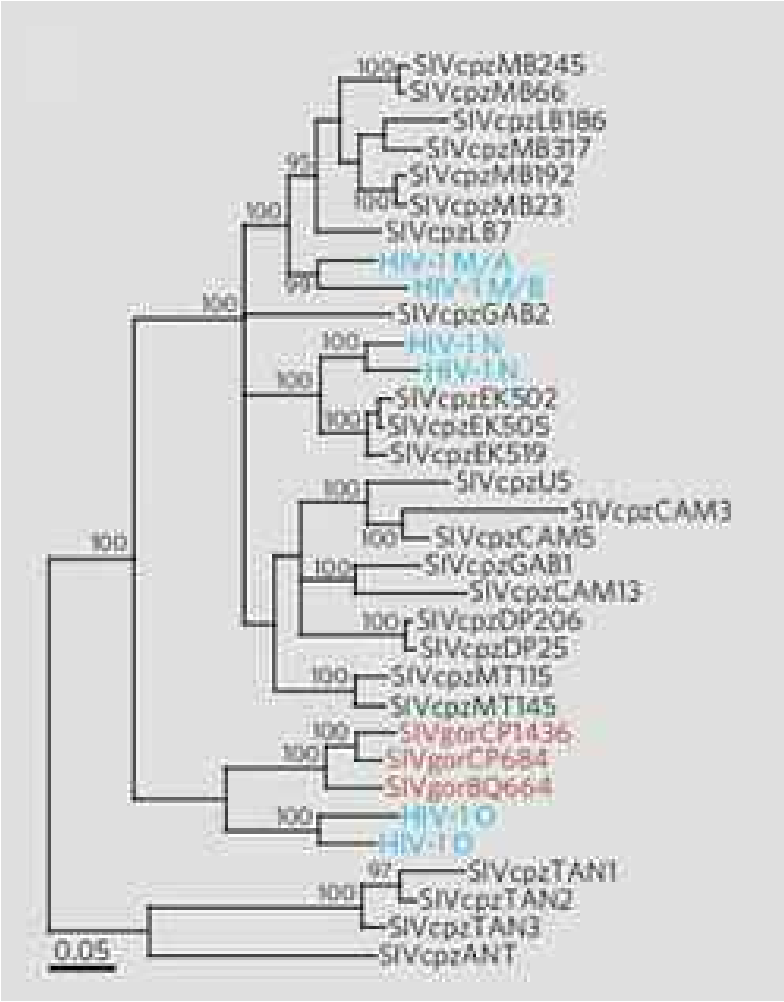

Figure 30: *evolutionary relationships of SIVgor, the three groups of HIV-1 and SIVcpz. From [37].*

The epidemiology of HIV-1 is really complex and the classification is ever-changing, trying to follow the rapid evolution of the virus. In fact it has a big genomic diversity, due for example to its high replication rate and to the recombination, that, as described before, generates diversity by exchanging long fragments between different genomes. Noteworthy is also the lack of proofreading activity by the reverse transcriptase (RT), that leads to a mutation rate of approximately 3.4×10^{-5} mutations per base pair per replication cycle. The length of the viral genome is about 10^4 base pairs and every day there is the production of 10^{10} new virions. Consequently millions of viral variants are produced in a single day within any infected person [38].

Phylogenetic Classifications of HIV-1.

Classification	Definition	Examples
Subtypes or clades	Genetically related HIV-1 strains that are essentially phylogenetically equidistant, generating a starlike, rather than a treelike, phylogeny	Subtypes A, B, C, D, F, G, H, J, and K are currently known; A through D are highly prevalent, others have low prevalence and limited geographic distributions
Sub-subtypes	Distinct lineages within a subtype; genetic distance between sub-subtypes is smaller than that between subtypes	Subtypes A and F are subdivided into sub-subtypes A1 through A4 and F1 and F2, respectively; mostly these circulate in Central and West Africa
Intersubtype recombinant forms	Mosaic strains with segments from two or more subtypes alternating across the genome	Common in mixed-subtype epidemics; thought to result from infection of a person with more than one HIV-1 subtype
Circulating recombinant forms	Specific recombinant forms that are spreading in a population; new forms are defined when three people without direct epidemiologic linkage are found to be infected; the assigned name reflects sequence of discovery and subtype composition, with "cpx" indicating forms containing three or more subtypes	Currently, 43 forms are described; CRF01_AE and CRF02_AG are found principally in Southeast Asia and West Africa, respectively; others have more limited distributions
Unique recombinant forms	Intersubtype recombinant forms recovered from only a single person	Hundreds of forms have been described on the basis of partial or complete genome sequences; their potential for epidemic spread is unknown
Geographically distinct lineages	Lineages, often country-specific, that are distinguishable phylogenetically; unlike sub-subtypes, they are not phylogenetically equidistant within subtypes	Thai B, Indian C, West vs. East African D, and Former Soviet Union A (FSU-A)

Features of the HIV-1 Pandemic, According to Subtype or Circulating Recombinant Form (CRF).*

Subtype or CRF	Location	Global Prevalence	Tropism and Replication	Disease Progression	Response to Therapy
Subtype					
A	East and Central Africa, Central Asia, Eastern Europe	12.3%	Mostly uses CCR5, even in late infection[40]	NA	No significant difference as compared with C and D[41]
B	Americas, Western Europe, East Asia, Oceania	10.2%	Uses CCR5 early, with increasing use of CXCR4 in late infection[28]	HLA-B7 associated with poor CTL response and increased viremia[42,43]; HLA-B57 associated with slow progression[42]; B strain in Brazil associated with slow progression[44]	NA
C	India, Eastern and Southern Africa	49.9%	Mostly uses CCR5, even in late infection[28]; increased vaginal shedding[30] and mother-to-child transmission[29,45]	HLA-B57 associated with slow progression[42]	No significant difference as compared with A and D[41]; differential pathways to resistance[46-49]
D	East Africa	2.5%	Uses CXCR4 in early infection[27]	Progression more rapid than A in Uganda, Kenya, and Tanzania[37-39]	NA
G	West Africa	6.3%	NA	NA	NA
F, H, J, and K	Various	Each <1.0%	NA	NA	NA
CRF					
CRF01_AE	Southeast Asia	4.7%	May have higher initial viral load than B but subtype may be a confounder[50]	Possibly accelerated progression as compared with B[36]	NA
CRF02_AG	West Africa	4.8%	Higher rate of replication in vitro than B[51]	NA	NA
Other	Various	Each <0.1%	NA	NA	NA

Figure 31: *the top table reports the phylogenetic classifications of HIV-1; the bottom table reports the features of the HIV-1 pandemic, according to subtype or Circulating Recombinant Form (CRF). NA: not available. CTL: cytotoxic T lymphocyte. From [38].*

HIV-1 is a lentivirus belonging to the retrovirus family. It has a cone shaped capsid with a diameter of 40-60 nm at the wide end and about 20 nm at the narrow end. Generally, there is a single capsid per virion, but literature reports also the presence of virions with two or more capsids. It is constructed from a lattice of capsid protein (CA) and its most important function is the protection of the genome. Capsid structure must guarantee the stability to survive in the extracellular environment, but also the ability to alter the conformation permitting the release of the genome into the host cell at the appropriate time. A layer of matrix protein (MA) lies between the capsid and the envelope. There are two proteins, essential for the recognition and internalization steps, associated to the envelope at the surface of the virion. A transmembrane protein (TM) called gp41 (glycoprotein with a molecular weight of 41 kD), that presents the C-terminal bound to the matrix protein inside the virion, and a heavily glycosilated surface protein (SU) called gp120 (glycoprotein with a molecular weight of 120 kD). A trimer of gp41-gp120 is called spike and it is possible to see about 14 spikes at the surface of each virion (*Figure 32*).

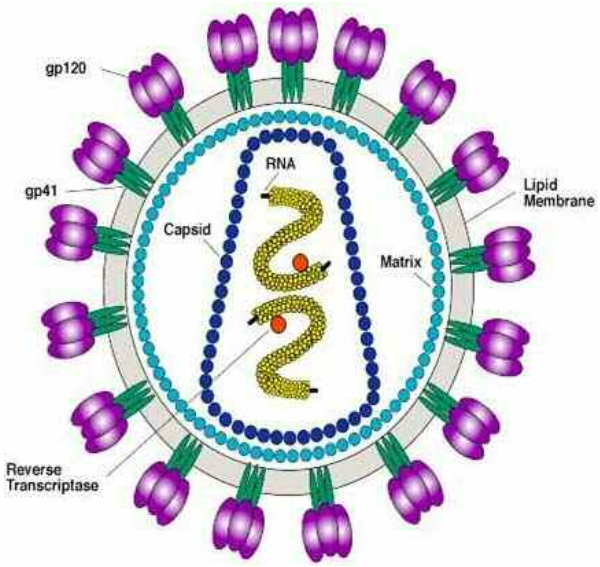

Figure 32: *virion structure.*

From http://www.stanford.edu/group/virus/retro/2005gongishmail/HIV.html.

HIV-1 contains two copies of single stranded RNA present as a dimer formed by base pairing between complementary sequences. The length of the genome is about 9.3 kb and it encodes the *gag*, *pol* and *env* genes together with auxiliary genes important for gene expression, transport of components within the cell and modification of the host's immune response. There are three reading frames in the genome and all of them are used. There are extensive overlappings, as for example the overlapping of *vpu* in frame 2 with *env* in frame 3. Some sequences are also split, with functional sequences being formed when the transcripts

are spliced (e.g. for *tat* and *rev*). At the ends (termini) of many linear virus genomes there are repeated sequences that include promoters, enhancer, origins of replication and other elements involved in the control of events in virus replication. These sequences are known as terminal repeat (*Figure 33*).

It's possible to find proteins bound non-covalently to the nucleic acids. Usually these proteins are encoded by the virus genome. The most abundant is the nucleocapsid protein (NC). NC is rich in basic amino acids (29% of the residues are lysine, arginine or histidine) and have two zinc fingers.

Noteworthy is also the presence of a molecule of transfer RNA ($tRNA^{lys-3}$) of the host cell, bound of each copy of the viral RNA through base pairing in a specific area called Primer Binding Site (PBS). This tRNA, together with other molecules of the host cell, is packaged in the virion during assembly. It is the primer for the synthesis of the (-)DNA by reverse transcriptase.

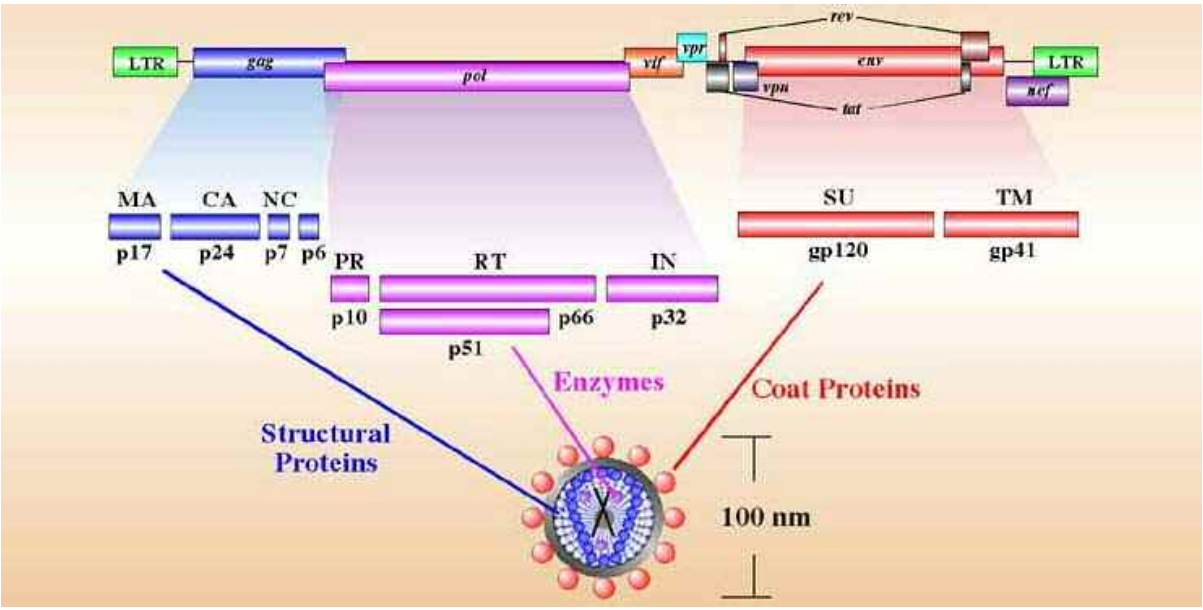

Figure 33: *genomic organization of HIV-1.*

From http://www.stanford.edu/group/virus/retro/2005gongishmail/HIV.html

6.2 HIV-1 REPLICATION

Viral Attachment and Entry. HIV-1 infects cells of the haematopoietic system, (T lymphocytes, monocytes, macrophages and dendritic cells), as well as of the central nervous system (astrocytes and microglial cells). The cell receptor responsible for the recognition is the CD4, a 58 kDA monomeric glycoprotein, which contains four immunoglobulin-like

domains in the extracellular portion. The interaction between receptor CD4 and envelope protein gp120 is the initial and crucial step for HIV infection. In fact it determines a conformational change in the surface spike that allows exposure and stabilization of the co-receptor binding site of gp120. Molecules of the host membrane that act as co-receptors are chemokine receptors with seven transmembrane domains and they fall into two major classes, determined by the arrangement of cysteine residues near the N terminus: C-C and C-X-C, where C is a cysteine and X is any amino acid. Therefore the chemokine receptors are named CCR and CXCR, respectively. CCR5 and CXCR4 seem to represent the most relevant co-receptors for HIV-1. Most of the viral strains use CCR5 and are known as R5 strains. Strains that use the other co-receptor are called X4 strains. There are also some strains that can use either co-receptors (R5X4 strains). HIV-1 transmission generally results from R5 viruses, even if the host is exposed to both variants, but after several years also X4 viruses can be detected in infected individuals and the shift from one co-receptor to the other is strongly associated with more rapid $CD4^+$ T cell decline and enhanced progression to AIDS. The mechanisms of conversion is not known yet [39]. Individuals who are homozygous for a mutation in the CCR5 gene and express no receptor on their cells are highly resistant to HIV-1 infection, while heterozygous people have increased resistance to infection.

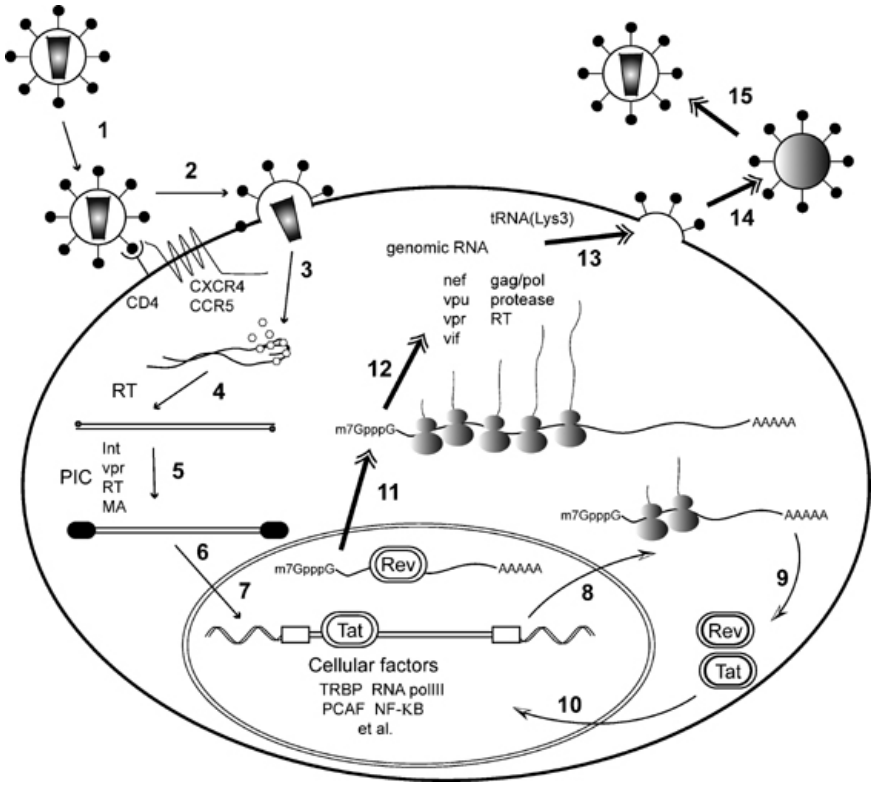

Figure 34: *HIV life cycle.* ***(1)*** *Adsorption to CD4 receptor and CXCR4 or CCR5 co-receptor.* ***(2)*** *Fusion.* ***(3)*** *Uncoating of viral genomic RNA.* ***(4)*** *Reverse transcription (RT, reverse transcriptase).* ***(5)*** *Formation of pre-integration complex (PIC).* ***(6)*** *Nuclear import of PIC.* ***(7)*** *Integration of proviral DNA into host genome.* ***(8)*** *Transcription of early multiply spliced mRNAs.* ***(9)*** *Translation of early regulatory proteins, Tat and Rev.* ***(10)***

Nuclear import of Tat and Rev. Tat increases transcription of viral mRNAs. ***(11)*** *Rev mediates export of singly spliced and unspliced viral mRNAs.* ***(12)*** *Translation of viral structural proteins.* ***(13)*** *Assembly at the plasma membrane of viral genomic RNA, proteins, and cellular factors.* ***(14)*** *Viral budding.* ***(15)*** *Viral maturation. From [40].*

Binding of gp120 to receptor and co-receptor triggers structural changes, allowing re-arrangement of gp41, which proceeds to fuse the membranes of the virion and the cell.

At this point the content of the virion is released into the cytoplasm of the host cell and develop into the Reverse Transcription Complex (RTC) (*Figure 34* summarizes all the HIV life cycle).

Reverse Transcription. The single stranded viral RNA must be converted to double stranded DNA by the viral enzyme reverse transcriptase, to enable HIV to be integrated into the host DNA and to use the cellular genetic machinery to make new viruses. Reverse transcription is initiated in the cytoplasm in the context of a large complex, the RTC, that associates rapidly with microtubules soon after formation. Viral proteins, including matrix, Vpr, integrase, reverse transcriptase, genomic RNA, and possibly host cell proteins, are thought to constitute part of the RTC.

The primer for synthesis of the (-) DNA is the tRNA bound to the genome, while a 19 nucleotide long polypurine tract (PPT) near the viral 3' end is the primer site for the synthesis of the (+)DNA. The DNA that results from transcription is longer than the RNA genome, because each of the termini has now the sequence U3-R-U5, called Long Terminal Repeat (LTR), derived from the acquisition of a U3 sequence at the 5' end and of a U5 sequence at the 3' end (*Figure 35*).

Integration. Integration is the mechanism used by the virus to invade the genome of the host cell. It allows the rapid viral genome's transcription and consequently the initiation of progeny production. At the same time it can also permit, through latency, the evasion of the immune surveillance. Integrase (IN) is the viral enzyme catalyzing this stage and its action is mainly divided in two steps: initially IN removes a terminal conserved dinucleotide from each end of the reverse transcribed cDNA, in a reaction called 3'-processing (3'-P). Secondarily, the enzyme catalyzes a transesterification reaction called strand transfer (ST), in which staggered cleavage of opposite target DNA strands provides the energy to join the newly recessed viral 3' ends to them. Host cell DNA repair enzymes are then necessary for the final gap repair, resulting in a full integrated genome. Integration occurs within a large nucleoprotein structure, called Pre-Integration Complex (PIC). It consists of the viral cDNA,

viral proteins retained from the initial reverse transcription complex and host cell proteins. While 3'-P occurs at cytoplasmic level, for the second step is necessary the translocation of all the components to the nucleus. HIV can productively infect both dividing and resting cells, as the PICs are able to enter also intact nuclei through a not known mechanism.

Detailed information about the integration step are reported on paragraph 6.3.

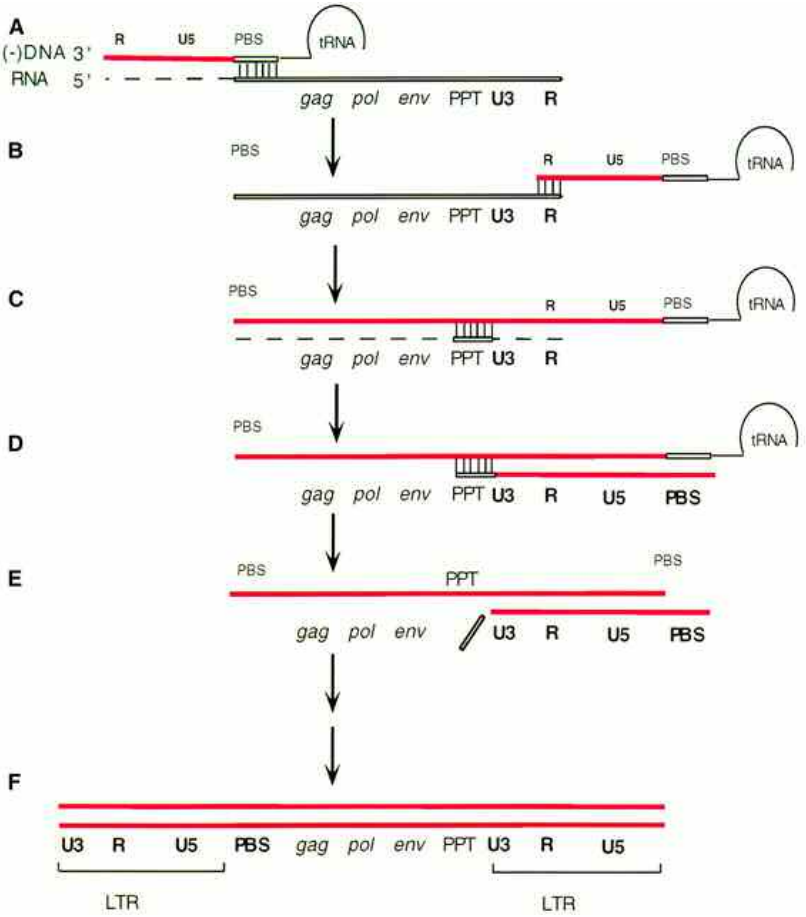

Figure 35: *reverse transcription of HIV-1 genome.* ***A)*** *Synthesis of (-)DNA is initiated using a cellular tRNA annealed to the Primer Binding Site (PBS). The RNA strand of the obtained duplex is degraded by RNase H.* ***B)*** *The (-)DNA strand is transferred to the 3' end of the viral genome and the transfer is mediated by identical repeated (R) sequences.* ***C)*** *The (-)DNA synthesis resumes and RNase H digests the RNA template but not the Polypurine Tract (PPT).* ***D)*** *PPT is used as a primer for second strand DNA synthesis.* ***E)*** *RNase H removes tRNA and PPT.* ***F)*** *Completion of strands' synthesis results in a linear DNA duplex with Long Terminal Repeats (LTR) at both ends. From [41].*

Transcription. Once the viral DNA is integrated into the host genome, the virus uses the transcription machinery of the infected cell to generate RNAs that will function as mRNAs or as genomes for progeny virions. The two integrated LTRs of the provirus are functionally different, even presenting identical sequences. Cell transcription factors bind to promoter and enhancer sequences in the upstream LTR, and transcription initiates at the U3-R junction by

cell RNA polymerase II. Transcription is then terminated in the downstream LTR, at the R-U5 junction after polyadenylation signal. Many of the genome-length transcripts are spliced and mainly three size classes of virus transcript can be detected in infected cells: the largest is of entire genome length (9.3 kb), one derives from a single splicing event (around 4.5 kb) and the final class, of about 2 kb, is the result of multiple splicing. All together, the splicing events result in more than 30 mRNA species. Each transcript is capped and polyadenylated and afterwards exported from the nucleus [42].

Translation and Post-Translational Modifications. The genic expression can be divided in two stages: early and late gene expression.

1) Early gene expression. Initially the primary transcripts are multiply spliced and their cytoplasmic translation leads to the formation of Nef, Tat e Rev proteins.

Nef protein (Negative Regulatory Factor) is myristylated just after the synthesis and it performs a number of roles in the cytoplasm as the reduction of the expression at the cell surface of CD4 and MHC class I and II proteins, change that can shield HIV-infected cells from immune surveillance.

Tat protein (Transactivator of Transcription) enhances the transcription after binding to the TAR region (Transactivation Response Element) at the 5' end of the nascent transcripts. In particular it has a nuclear localization signal that directs it to the nucleus and, as other cell kinases (for example CycT and Cdk9), binds to the structured TAR element. The consequent phosphorylation of the cytoplasmic tail of RNA Polimerase II, results in the release of stalled enzyme and allows its rapid elongation along the proviral DNA, with increased rate of transcription. In absence of Tat, transcripts are incomplete but they are enough to allow the synthesis of a small amount of Tat, which then assists the synthesis of genome-length RNA.

Rev protein (Regulator of Expression of Virion Proteins) has a nuclear localization signal and the ability to bind the RRE element (Rev Response Element), present only in the unspliced and singly spliced transcripts. Its accumulation in the nucleus determines the shift from early to late protein synthesis. Genome-length and singly spliced transcripts are transported to the cytoplasm only after the binding of multiple copies of Rev.

2) Late gene expression. In a more advanced stage of infection, the translation moves from early to late expression, also thanks to proteins as Rev. The translation of the genome-length mRNAs determines the production of Gag and Gag-Pol polyproteins, resulting from the *gag* (group-specific antigen) and *pol* (polymerase) genes. *Gag* gene encodes for internal structural proteins as MA, CA, p2, NC, p1 and p6, while *pol* gene assures the synthesis of protease, reverse transcriptase and integrase. HIV-1 requires greater quantities of Gag proteins

than Pol proteins and a particular mechanism permits to have about 95% of ribosomes that terminate the synthesis after crossing the stop codon at the end of the *gag* gene, while a 5% of the ribosomes, after frameshift, continue translation of the *pol* region, yielding the Gag-Pol polyprotein. In particular, the ribosome traverses the sequence UUUUUUA, known as "slippery sequence", at the junction of the NC and p1 domains of *gag*. The sequence, together with a secondary structure positioned downstream, determines the slipping from reading frame 1, that encodes for Gag proteins, to 3, that encodes for Pol proteins, leading to the production of the Gag-Pol polyprotein. The polyproteins are then myristylated.

While Gag and Gag-Pol are translated on free ribosomes, Env proteins are synthesized in the rough endoplasmic reticulum from singly spliced mRNAs and then transported to the plasma membrane via the Golgi complex. The *env* (envelope) gene encodes for envelope proteins. During the route from RE and plasma membrane, the proteins are heavily glycosylated and cleaved by a host protease, giving origin to SU and TM molecules (gp120 and gp41).

Also Vif (Virion Infectivity Factor), Vpr (Viral Protein R) and Vpu (Viral Protein U) proteins are translated from singly spliced transcripts. Vif plays an important role in ensuring the exclusion from progeny virions of cell enzymes (APOBEC3F and APOBEC3G) that could interfere with replication in the next host cell. These molecules can in fact induce lethal mutations by deaminating deoxycytidine to deoxyuridine during reverse transcription. The resulting nonsense or missense mutations change the coding capacity of the mRNA. Vpu is, as Env proteins, translated in the rough endoplasmic reticulum and is a membrane-associated protein with at least two different biological functions: the degradation of CD4 in the endoplasmic reticulum and the enhancement of virion release from the plasma membrane of HIV-1 infected cells. Vpr has been shown to play multiple functions as control of accuracy of RT, nuclear import of the viral DNA as a component of the PIC, cell cycle progression and regulation of apoptosis.

Assembly and Release of Virions. The assembly of HIV-1 occurs on the inner surface of the plasma membrane. Gag and Gag-Pol polyproteins attach the cell membrane by the myristyl group on the N-terminal. The RNA dimer, genome of the new virion, results from base pairing between complementary sequences in the 5' end of each RNA. The basic NC domains with their zinc fingers bind to the virus RNA and mediate the formation of the genome dimer. p6 domain contains a proline-threonine-alanine-proline sequence and is responsible for the release of the budding virion from the host cell.

The immature virion acquires its envelope by budding from the cell surface. The copies of Gag and Gag-Pol are radially arranged, with the N-teminal facing outward and the C-terminal inward.

Maturation. As the virion buds from the host cell, the maturation process occurs. Protease is competent for this step. The enzyme cleaves newly synthesized polyproteins at the appropriate places to create the mature protein components of an infectious virion. Without effective protease, virions remain uninfectious. The active form of the enzyme is a homodimer and each monomer contains 99 amino acids and is identical in conformation. The secondary structure includes one α-helix and two antiparallel β-sheets. The active site forms at the dimer interface and it is characterized by the highly conserved catalytic triad Asp-Thr-Gly (respectively in position 25, 26 and 27), common for the aspartyl protease superfamily. The two Asp 25 act as the catalytic residues. As integrase, also protease uses a water molecule as nucleophile to hydrolyze the scissile peptide bond.

6.3 HIV-1 INTEGRASE

Integration is an important step during viral infection. First of all, it allows the provirus access to the cellular machinery for viral gene expression. Furthermore, it permits the maintaining of the chronic infection, ensuring that the parental and daughter cells will carry the viral genes.

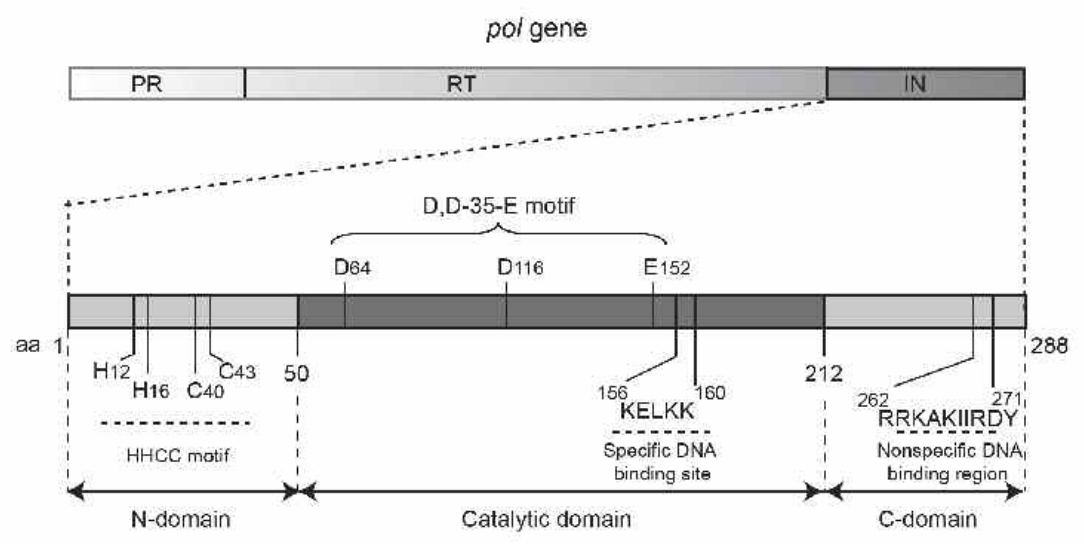

Figure 36: *pol gene and domains of HIV-1 integrase. PR: protease; RT: reverse transcriptase; IN : integrase. From [43].*

HIV-1 integrase is a 32 kDa protein encoded within the *pol* gene together with protease and reverse transcriptase, other two essential viral enzymes. In particular the 3' end of the gene encodes for IN, while protease is at the 5' end and RT in the centre (*Figure 36*). The gene is translated as the *Pol polyprotein* and it is a task of protease to cleave and generate IN. There are approximately 40-100 integrase molecules packaged within each HIV particle [44]. Integrase belongs to the family of the polynucleotidyl transferases/esterases, that includes between all also transposases and RNase H [45].

Integrase is constituted of three structural domains: the amino-terminal domain (NTD), the catalytic core domain (CCD) and the carboxy-terminal domain (CTD). X-ray diffraction or solution NMR determined the atomic structure of each of these separate domains (*Figure 37*) and of the association of the core domain with NTD or the CTD separately. Besides this, the bigger limitation for the full understanding of the enzyme's catalytic activity and of the binding mode of known integrase inhibitors, is the lack of the crystal structure of full-length IN and of its complex with DNA substrate.

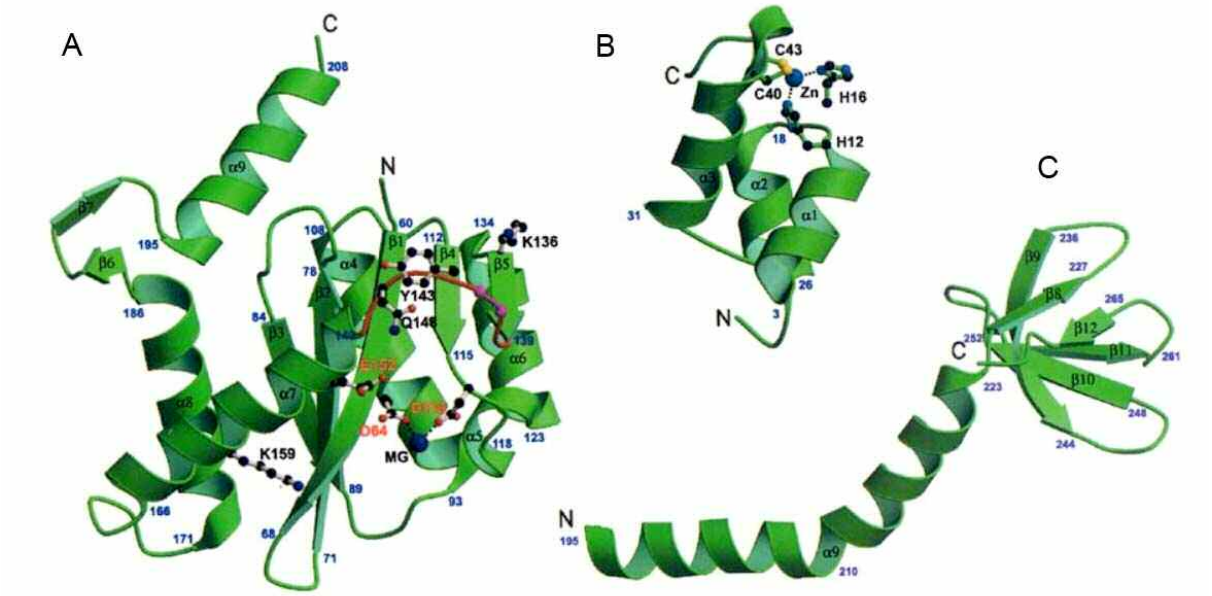

Figure 37: *the secondary structures of the three domains of HIV-1 integrase.* ***(A)*** *Catalytic core (PDB code 1QS4). Flexible loop shown in red. In magenta the residues which are missing in the model. In red is also reported the DDE motif.* ***(B)*** *N-terminal domain (PDB code 1K6Y). In ball-and-stick the His_2Cys_2 domain that binds Zn^{2+}.* ***(C)*** *C-terminal domain (PDB code 1EX4). Here is reported also the α-helix which joins the C-terminus to the core domain. Blue numbers mark the amino acid positions. From [46].*

The NTD, from residue 1 to 50, is a highly conserved region among retroviruses' integrases and contains an HHCC motif (His 12, His 16, Cys 40, Cys 43) able to bind one Zn^{2+}. The structure of this portion determined by X-ray [47] is characterized by three α-helices, while NMR analysis shows the insertion of an additional α-helix between the first

and the second α-helices of the crystallographic model [48, 49]. This disagreement results in a completely different dimerization interface that can be justified as multiple arrangements of the multimers, as it is possible to see in *Figure 38*. The Zn^{2+} ion stabilizes the lower region of the NTD, while the upper region is stabilized by a hydrophobic core.

The exact role of NTD has not been completely understood. Mutations of cysteines or histidines of the zinc finger motif lead to impaired catalytic activity [50]. Other studies showed that zinc promotes and stabilizes multimerization [51] and it seems likely that this is the function that permits efficient integration.

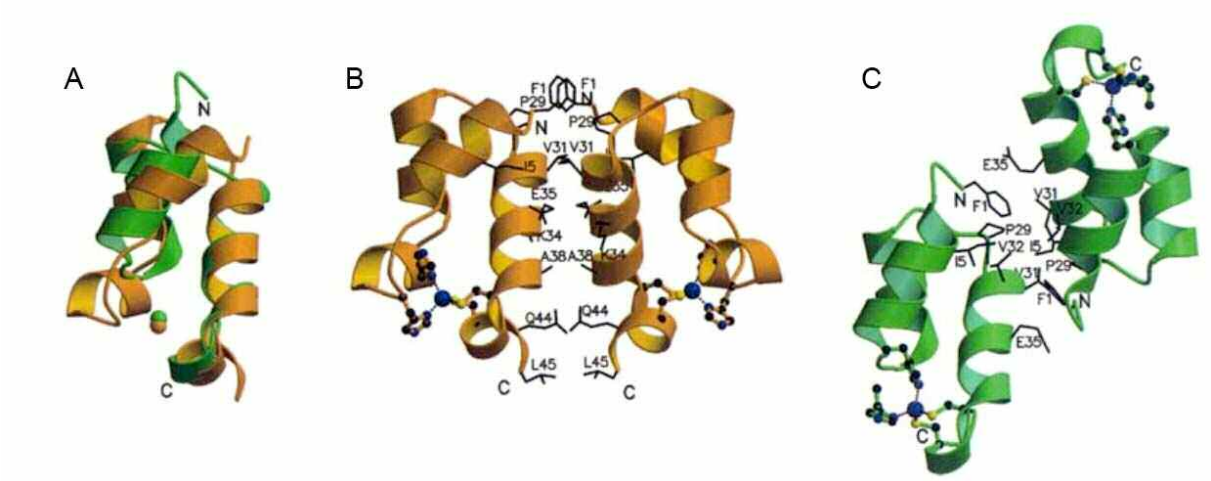

Figure 38: *Structures of the N-terminal domain.* ***(A)*** *Superimposition of the structure determined by X-ray (green, PDB code 1K6Y) and NMR (gold, PDB code 1WJA). It is possible to identify differences in the region between residues 9 and 19. The NMR structure results in the insertion of an α-helix between the first and the second α-helices of the X-ray model.* ***(B)*** *Dimerization interface in the NMR structure.* ***(C)*** *Dimerization interface in the X-ray structure. In ball-and-stick the four residues which chelate Zn^{2+}. In bonds the residues which are at the dimerization interface. Zn^{2+} cation is shown as a ball and dashed lines show the ion coordination. From [46].*

The CTD (residues 212-288) is the least conserved domain compared with NTD and CCD. It is required for integrase activity because its deletion abolishes both 3'-processing and strand transfer [52]. This domain binds DNA nonspecifically but, as for NTD, its real significance remains to be determined. Solution NMR studies determined its structure as five antiparallel β-strands, with an overall SH3-like fold (a fold similar to that of the Src Homology 3 domain) (*Figure 37C*). CTD is connected to the core by an α-helix. Different data are reported for the dimer structure meaning that the interaction can vary considerably and the preferred mode of dimerization is not at all clear [53].

The catalytic core domain, from residue 50 to 212, contains the canonical three-amino-acid motif presents in all the polynucleotidyl transferases. The catalytic triad is called DDE motif, from the aspartic acid in position 64, that in position 116 and the glutamic acid in

position 152. These acidic residues are highly conserved in all integrases and retrotransposases and their mutation completely abolishes the enzymatic activity.

CCD needs the presence of the terminal domains to perform both 3'-processing and strand transfer, while it is able alone to carry out disintegration, the reversal of the joining reaction described only *in vitro* and never demonstrated *in vivo*.

Between the three domains, the core was the first to be solved by X-ray in 1994 by *Dyda* and co-workers [54], using the substitution F185K that permitted to overcome the poor solubility but maintained the catalytic activity although failed to replicate. The insolubility of integrase is still nowadays a big issue. Often in literature F185K mutation is associated with a cysteine to serine change at position 280, that renders the integrase more soluble (>10 mg/ml) [55], or the phenylalanine in 185 is substituted with a histidine, obtaining a soluble enzyme still able to replicate. It is important to underline that, thanks to a good purification's strategy, all the results obtained and presented in the present PhD dissertation were performed using WT enzyme, without substitutions along the sequence to increase the solubility. Thus, while for *in vitro* studies the solubility problem is overcome, for crystallization purpose it still remains a big matter.

Nowadays there are numerous structures of the CCD in Protein Data Bank but its complex with the substrate has yet to be reported, not allowing the complete understanding of the interaction.

The core consists of an α/β structure, containing five β-sheet strands together with six α-helices (*Figure 37A*). An important loop, between residues 140 and 150, is adjacent to the active site. In all the analyzed crystals, this loop is disordered and shows different conformations. This evidence leads to the conclusion that this flexible area adopts a relevant conformation only in the presence of the DNA substrate. It is likely that the two glycins in position 140 and 149 act as a hinge allowing the opening and the closure of the loop, movement that can determine first the entrance and then the stabilization of the DNA in a particular position.

Superimposition of the crystal structures of the three domains shows that probably NTD lies in the centre between CCD and CTD, close to the extended α-helix joining these last two [47] (*Figure 39*).

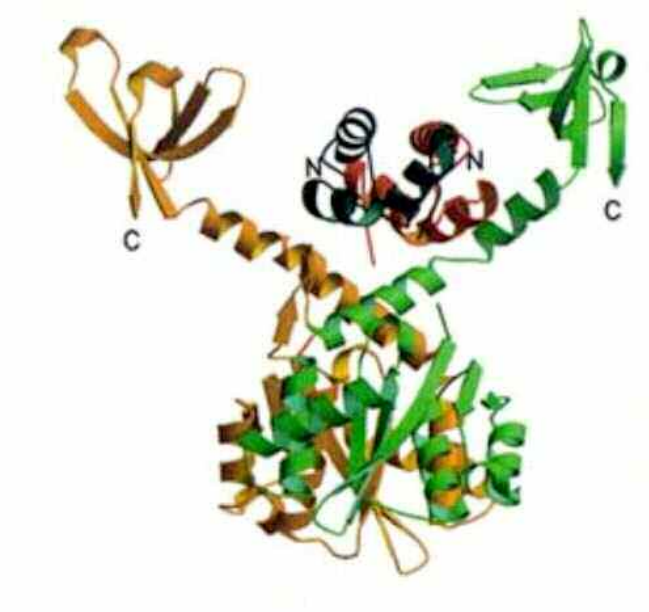

Figure 39: *structures of N-terminal domain plus core (PDB code 1K6Y) and C-terminal domain plus core (PDB code 1EX4) are superimposed to give a possible model of full length integrase. The model shows no steric overlap of the protein residues.*

Integrase requires a metal cofactor, Mg^{2+} or Mn^{2+}, for the catalytic activity. Manganese is present at very low intracellular concentration ($<10^{-7}$ M) while magnesium is more abundant ($<10^{-3}$ M). DNA binding and sequence specificity are differentially influenced by the type of metal ion used. Integrase produces nonspecific nicks in presence of Mn^{2+} and it is capable to use various nucleophiles, as glycerol or methanol, to carry out 3'-processing in presence of this ion [56]. When Mg^{2+} is the cofactor, IN uses water as nucleophile and, in *in vitro* experiments, the preference for Mg^{2+} increases with the increasing of the length of the DNA substrate. Finally the specific photocrosslinking between protein and DNA is increased in presence of magnesium compared to manganese [57].

All these evidences suggest that in presence of manganese the active site is less "tight" and permit the use of a wide selection of nucleophiles, reducing consequently the DNA-binding specificity of the 3'-processing reaction [58]. These reasons lead to the assumption that Mg^{2+} is the relevant cofactor *in vivo*.

There are doubts about the presence of one or two metal ions bound in the active site of the enzyme. In fact crystal structures of HIV-1 integrase reveal the presence of a single ion that forms a coordination complex between the two aspartic acids in position 64 and 116. However it seems likely that a second ion could be coordinated between D116 and E152 once integrase binds its DNA substrate [59]. Argument for this is the ASV integrase crystal structure that shows the presence of the second metal and the general two-metal structure for polynucleotide transferases [60, 61].

All three domains of IN interact with DNA. The core domain is obviously in close contact with both viral and host DNA, and in particular with the last six base pairs of viral LTR ends and the site of integration on target DNA. The NTD seems to be in proximity of the target DNA and in particular in 5' position to the site of integration. This domain is not involved in the substrate specificity. The CTD interacts with terminal bases of the viral LTRs, bases located farther then those interacting with the core. This is a non sequence-specific binding and it seems to be responsible for the correct positioning and orientation of the viral DNA. Therefore, the absence of the CTD determines defective catalytic activity presumably because the DNA substrate is not able to correctly fit in the active site.

Essential feature for integration is the presence of a conserved terminal sequence in the viral LTR region. Additions or substitutions of the minimum sequence CAGT lead to an inefficient enzymatic activity. Noteworthy is also the fact that, substitutions that determine strengthening of the hydrogen bonds between the plus and minus strands on the vicinity of the CA dinucleotide, decrease 3'-processing activity. Contrary to that, substitutions leading to weaker or disrupted base pairing increase activity.

Crosslinking studies demonstrated the interactions of specific amino acids with the viral DNA. Essential is the interaction mediated by Q148 on the terminal bases [58, 62]. This residue lies in the disordered loop and it is difficult to determine its exact position during catalysis. As previously described, the loop has been proposed to be involved in a conformational change that occur upon DNA binding [54] and therefore its function is probably to stabilize the 5' end of the viral DNA. Mutations of the glutamine 148 strongly impair the enzymatic activity. Along the loop, also the tyrosine in position 143 showed to have contact with the viral DNA [58].

6.4 BIOCHEMICAL BASIS OF THE INTEGRATION REACTION

The integration reaction could be essentially divided in two main steps: 3'- processing and strand transfer.

3'-processing. After entering the cytoplasm of the host cell, the reverse transcriptase is in charge of transcribe the viral RNA genome in a double stranded DNA. As soon as this synthesis is completed, within 5-6 hours postinfection, it is integrase's turn. In particular, it is at the cytoplasmic level that integrase catalyzes the removal of a GT dinucleotides from both 3' ends of the LTRs. The processing reaction is an endonucleolytic cleavage that occurs

immediately 3’ of a conserved CA sequence, and results in exposure of a new hydroxyl group at the 3’ end of the viral DNA. The CA is invariable in all retroviruses, while the removed terminal dinucleotide is not. 3’-processing is permitted by the nucleophilic attack of a water molecule on the internucleotide phosphodiester bond and the reaction leaves an overhanging 5’-AC on the complementary strand.

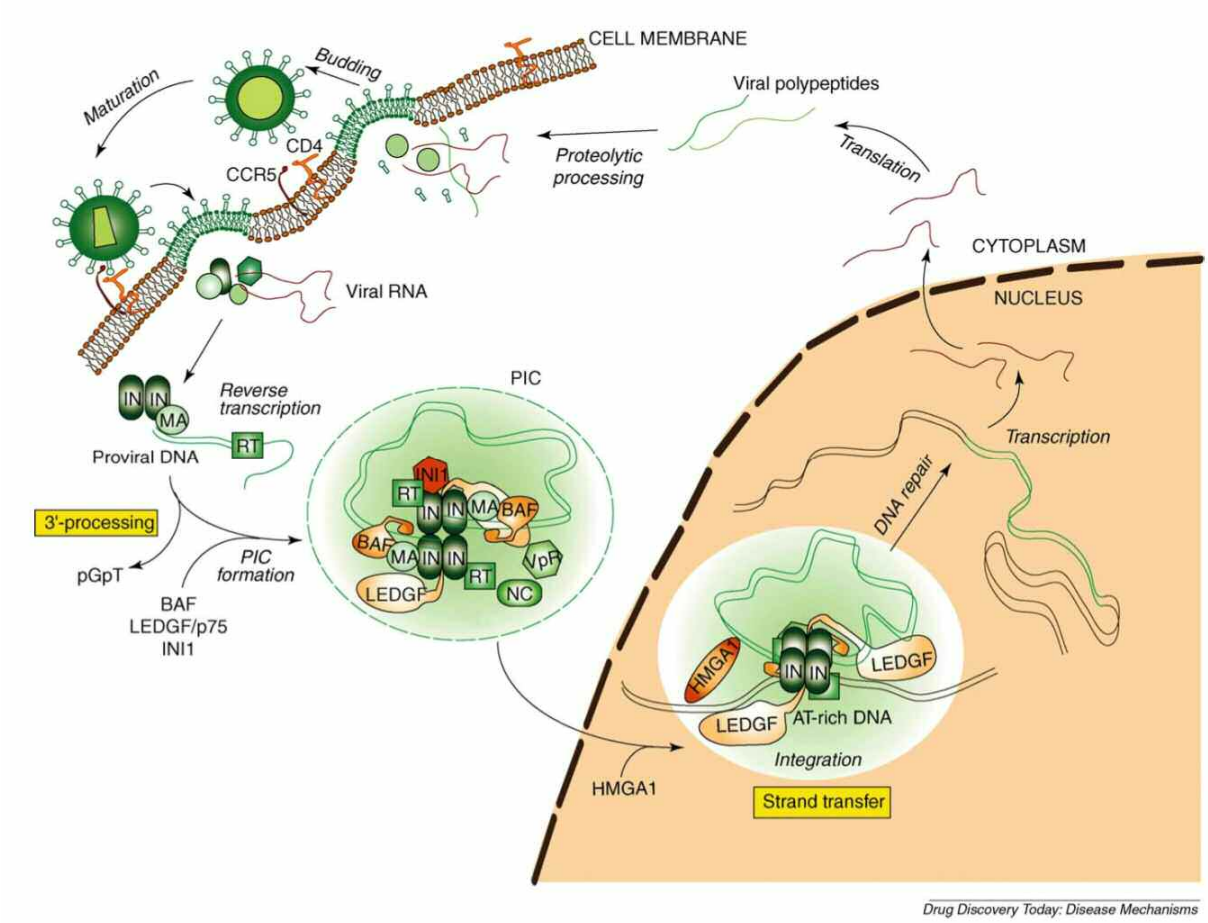

Figure 40: *HIV-1 life cycle. The integration step starts with 3’-processing which consists in an endonucleolytic cleavage at the 3’ end of the proviral DNA. The preintegration complex (PIC), containing viral and cellular co-factors, is then traslocated into the nucleus and the processed proviral DNA ends are inserted into the host DNA, in a step called strand transfer. Finally the reparation of gaps between viral and chromosomal DNA leads to the integrated proviral DNA. From [63].*

After 3’-processing, integrase remains bound to the viral DNA in the so called pre-integration complex (PIC) (*Figure 40*). This is a large nucleoprotein complex that bridges both ends of the viral DNA. It contains viral proteins, probably retained from the initial reverse transcription complex, but also cellular proteins [64]. Between the viral co-factors, it is possible to detect the RT, the matrix (MA), the nucleocapsid (NC) and the viral protein R (Vpr). Between the cellular co-factors packaged within the PIC there are: the barrier to auto-integration factor (BAF), that prevents autointegration and stimulates integration in

chromosomes; the integrase interactor 1 protein (INI1), the first integrase-binding protein discovered; the high mobility group protein A1 (HMGA1), that stimulates integrase activity by bridging and compacting the viral cDNA; the lens epithelium-derived growth factor (LEDGF-p75), whose exact role is under investigation but it seems to be involved above all in the chromosome tethering (for additional information about this factor see paragraph 6.7). To this list it is possible to add other recently discovered interacting proteins as the Polycomb group embryonic ectoderm development protein (EED), the hepatoma-derived growth factor related protein 2 (HRP2), the heat-shock protein 60 (HSP60) and the p300 acetyltransferase [63]. In summary the essential functions of these factors seems to be the nuclear import and trafficking, and the engagement with chromatin.

Once inside the nucleus, the generated 3'-OH DNA ends of viral DNA become the reactive intermediates required for the second step of integration, the strand transfer.

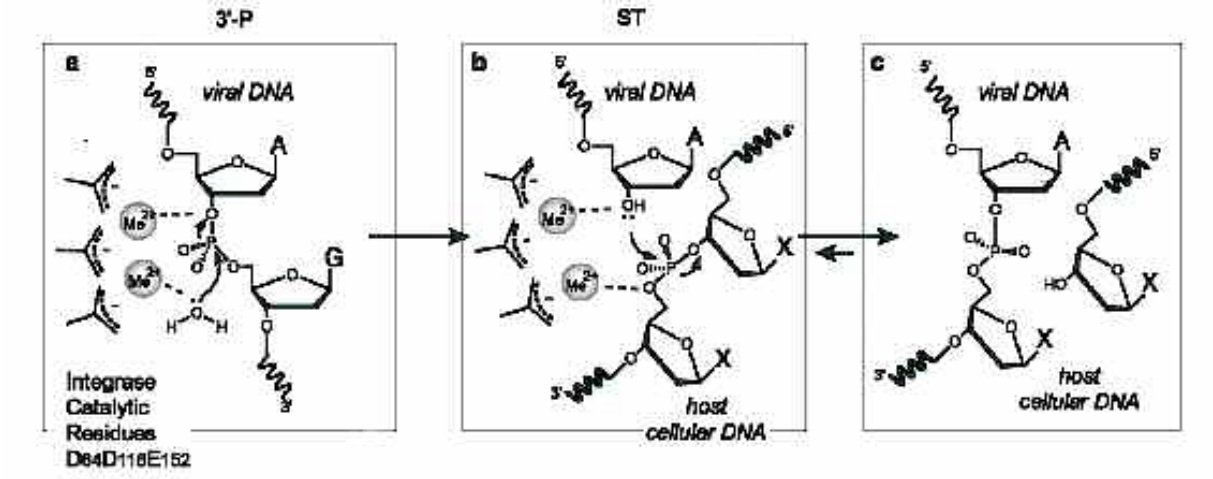

Figure 41: *biochemical steps of retroviral integration.* ***(a)*** *IN binds the LTR of viral DNA and catalyzes the nucleophilic attack by a water molecule of the phosphodiester backbone (3'-P: 3'-processing).* ***(b)*** *The complex is translocated to the nucleus where IN promotes the nucleophilic attack of the host DNA by the 3'-hydroxy viral nucleophilic end (ST: strand transfer). From [43].*

Strand transfer. This step is essentially the insertion of the viral DNA into the host genome. It is a concerted cleavage and ligation reaction. The two 3'-OH ends of the viral DNA attack phosphodiester bonds on opposite strands of the target DNA, separated by five bases. Therefore the joining takes place on the same face of the double helix, flanking the major grove. This transesterification reaction does not require exogenous energy source, because it is obtained by the breaking of the chromosome DNA phosphodiester bond. 3'-processing and strand transfer can be considered the same chemical reaction. In the first step there is a nucleophilic attack by a hydroxyl group of a water molecule to the phosphate

backbone; during strand transfer the nucleophile is the free hydroxyl group of the processed adenosine, and it attacks the phosphodiester bond of host DNA strands (*Figure 41*).

Integration is complete after the ligation of the 5' end of the viral DNA. In particular, trimming of the dinucleotide 5' overhang of the viral DNA and extension of the 3'-OH of the target DNA, result in the repaired gap and a fully integrated genome. Cellular enzymes seem to be responsible of these last steps, although the real identity of these factors remains uncertain.

Genomic integration of HIV-1 tends to occur in active transcription units without favoring the promoter regions. In these areas no hotspots for joining have been observed *in vitro* but it seems unlikely that the insertion is sequence-independent. LEDGF/p75 could be an important factor involved in the site-specific directioning.

Finally, several evidences report that the dimer of integrase is responsible for the 3'-processing reaction, while a tetramer (dimer of dimer) is required for the insertion of the two viral ends into the host genome [65] (*Figure 42*).

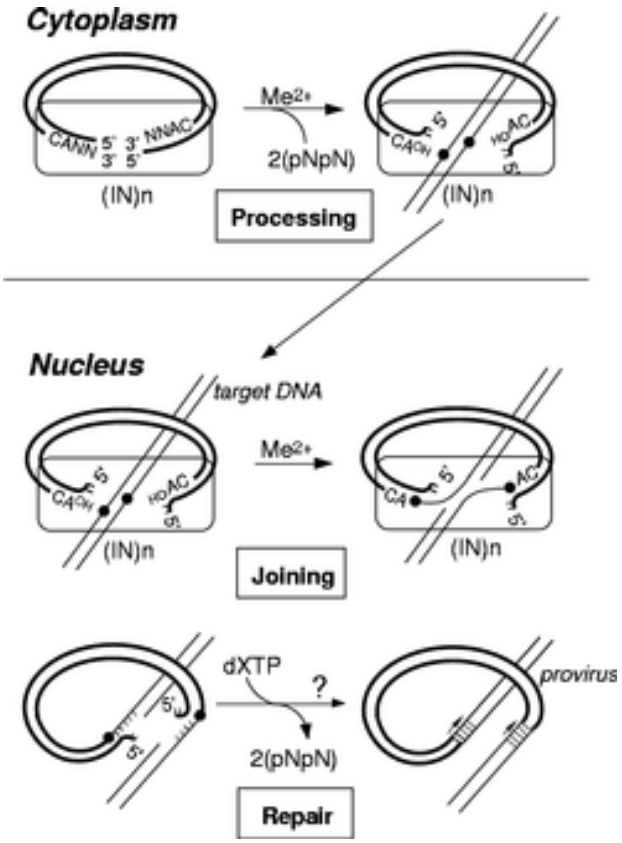

Figure 42: *the steps of retroviral DNA integration reaction. In the first step, the GT dinucleotide is removed from the 3' ends of the viral DNA, immediately following a conserved CA dinucleotide. In the transfer step, these new 3' ends are jointed to host target DNA in a concerted cleavage-ligation reaction. (IN)n indicates a multimer of integrase, minimally a dimer. Repair takes place in vivo with the help of cellular factors and requires new DNA synthesis to fill in the gaps in host DNA. From http://www.fccc.edu/research/reports/current/skalka.html.*

6.5 THERAPEUTIC AGENTS IN AIDS TREATMENT AND VIRAL DRUG RESISTANCE

Zidovudine (azidothymidine, AZT), the first antiretroviral agent, was approved by FDA in 1987 [66], five years after the discovery of the HIV. It was given as monotherapy and in high dosages, resulting in many treatment failures due to acquired drug resistance and in severe adverse effects. For four years monotherapy remained the standard of care but after the approval of didanosine (1991), stavudine (1994) and lamivudine (1995), the idea of a dual antiretroviral therapy was pursued. Also the dual therapy led to development of resistant viruses, giving improvement in quality of life correlated to decreased viral load and increased CD4 count only during the initial part of the therapy and later ending in a viral rebound.

Once that protease inhibitors were developed and saquinavir approved by FDA (1995), the concept of highly active antiretroviral therapy (HAART) took off and it remains still nowadays a standard treatment strategy for HIV-1 infected patients. HAART regimen consists of two nucleoside reverse transcriptase inhibitors (NRTIs) associated to a protease inhibitor (PI) or a non-nucleoside reverse transcriptase inhibitor (NNRTI). In fact anti-HIV drug combinations are much more effective than mono- and dual therapies because they reduce the emergence of drug-resistant viruses, as viral fitness is decreased after appearing of multiple mutations necessary to determine drug resistance.

HAART brought to improvements as better quality of life, slowed disease progression, increased survival, decreased opportunistic infections, increased CD4 counts and decreased viral loads. Despite that, the very high pill burdens, the complex dosing schedules and the specific food requirements often lead to loss of drug adherence with treatment failure and multidrug resistance. This is why additional therapeutic approaches are necessary and why research is working with the aim to discover novel inhibitors of HIV-1 targeting other stages of the viral life cycle.

A brief description of all the categories of antiretroviral agents is hereafter reported. Instead, integrase inhibitors are described in details in paragraph 6.6.

FDA-approved therapies for the treatment of HIV – October 2007.

Approval date	Generic name	Brand name	Manufacturer
Nucleoside reverse transcriptase inhibitors (NRTIs)			
1987	Zidovudine (AZT)	Retrovir	GlaxoSmithKline
1991	Didanosine (ddI)	Videx	Bristol-Myers Squibb
1992	Zalcitabine (ddC)	Hivid	Roche Pharmaceuticals
1994	Stavudine (d4T)	Zerit	Bristol-Myers Squibb
1995	Lamivudine (3TC)	Epivir	GlaxoSmithKline
1997	Lamivudine + zidovudine	Combivir	GlaxoSmithKline
1998	Abacavir	Ziagen	GlaxoSmithKline
2000	Abacavir + lamivudine + zidovudine	Trizivir	GlaxoSmithKline
2000	Didanosine (ddI)	Videx EC	Bristol-Myers Squibb
2001	Tenofovir disoproxil fumarate	Viread	Gilead Sciences
2003	Emtricitabine (FTC)	Emtriva	Gilead Sciences
2004	Abacavir + lamivudine	Epzicom	GlaxoSmithKline
2004	Emtricitabine + tenofovir	Truvada	Gilead Sciences
Non-nucleoside reverse transcriptase inhibitors (NNRTIs)			
1996	Nevirapine	Viramune	Boehringer Ingelheim
1997	Delavirdine (DLV)	Rescriptor	Pfizer
1998	Efavirenz	Sustiva	Bristol-Myers Squibb
Protease inhibitors			
1995	Saquinavir	Invirase	Roche Pharmaceuticals
1996	Ritonavir	Norvir	Abbott Laboratories
1996	Indinavir (IDV)	Crixivan	Merck
1997	Nelfinavir	Viracept	Pfizer
1997	Saquinavir mesylate	Fortovase	Roche Pharmaceuticals
1999	Amprenavir	Agenerase	GlaxoSmithKline
2000	Lopinavir + ritonavir	Kaletra	Abbott Laboratories
2003	Atazanavir	Reyataz	Bristol-Myers Squibb
2003	Fosamprenavir	Lexiva	GlaxoSmithKline
2005	Tipranavir	Aptivus	Boehringer Ingelheim
2006	Darunavir	Prezista	Tibotec, Inc.
Fusion inhibitors			
2003	Enfuvirtide (T-20)	Fuzeon	Roche Pharmaceuticals & Trimeris
Multiclass combinations			
2006	Efavirenz + emtricitabine + tenofovir disoproxil furmarate	Atripla	Bristol-Myers Squibb and Gilead Sciences
Entry inhibitors			
2007	Maraviroc	Selzentry	Pfizer
Integrase inhibitors			
2007	Raltegravir	Isentress	Merck

Figure 43: *FDA-approved therapies for the treatment of HIV-1. From [67].*

Nucleoside/Nucleotide Reverse Transcriptase Inhibitors. NRTIs, nucleoside analogue reverse transcriptase inhibitors, were the first class of drugs approved by FDA. These compounds are structurally similar to nucleosides and they act by interfering with the synthesis of virus nucleic acids. Before starting their antiviral effect, they need to entry the host cell and be triphosphorylated by cellular kinases to become nucleotide analogues. NRTIs

are lacking in a 3'-OH group at the sugar moiety and they determine chain termination during DNA synthesis catalyzed by RT. In fact when incorporated into a growing strand, their structure prevents the formation of a 3'-5' phosphodiester bond between the NRTI and incoming 5'-nucleoside triphosphates.

To date, there are eight FDA approved NRTIs: zidovudine (AZT, Retrovir®), didanosine (ddI, Videx®), zalcitabine (ddC, Hivid®), stavudine (d4T, Zerit®), lamivudine (3TC, Epivir®), abacavir (ABC, Ziagen®), tenofovir disoproxil fumarate (TDF, Viread®) and emtricitabine (FTC, Emtriva®) (*Figure 44*).

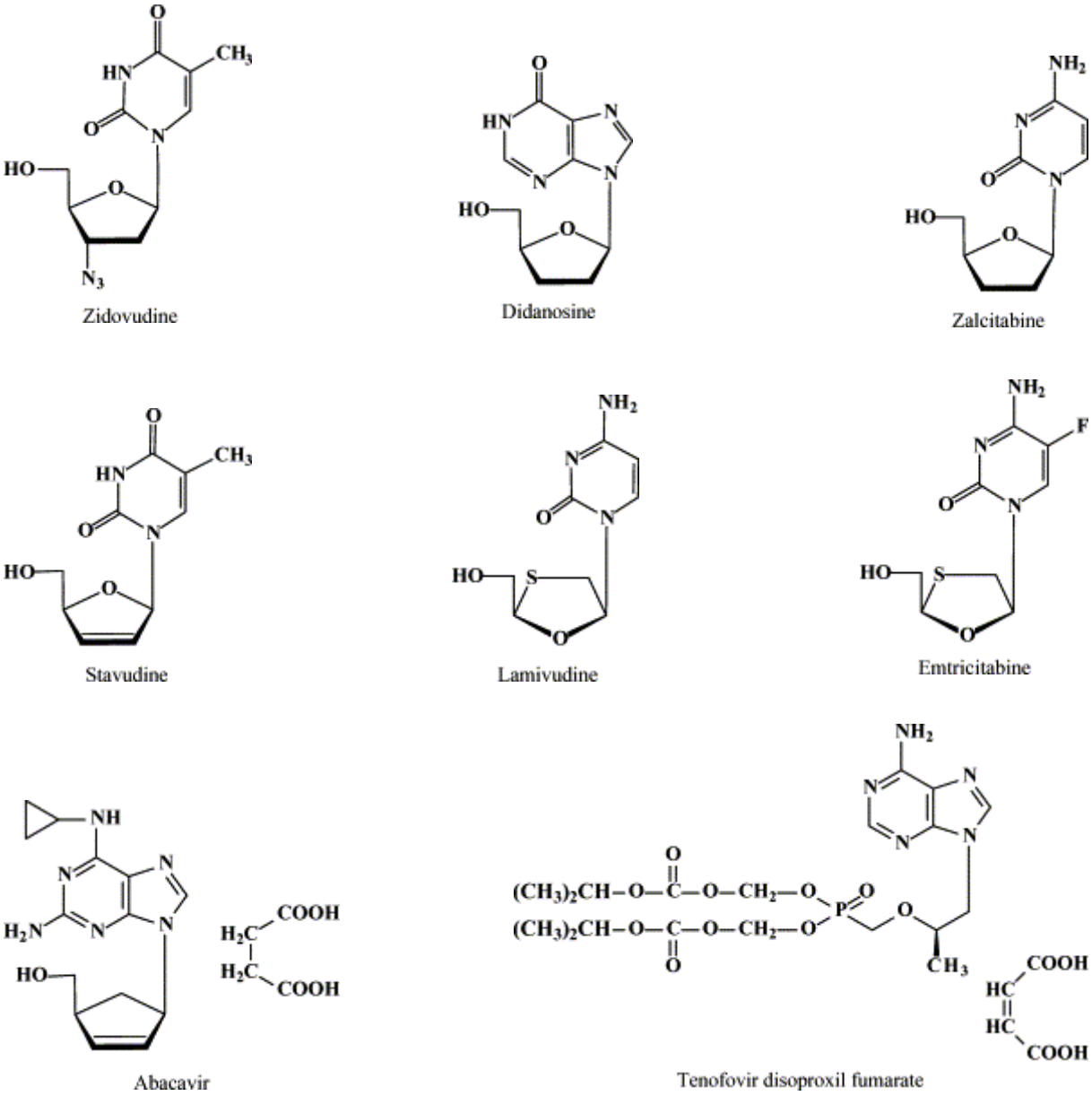

Figure 44: *chemical structures of Nucleoside and Nucleotide Reverse Transcriptase Inhibitors.*

Tenofovir disoproxil fumarate is the only nucleotide analogue of this category. Unlike nucleoside analogues, it already contains a phosphate group and only requires two

phosphorylation steps [68]. Consequently, the problem of the rate-limiting step, corresponding to the addition of the first phosphate group, is overcome. It is also active in both resting and actively dividing lymphoid cells and macrophages whereas the other compounds of the category have no activity on resting cells [69]. Tenofovir has become one of the preferred agents in both treatment-experienced and treatment-naïve patients.

Emtricitabine is the newest between NRTIs (2003) and differs from lamivudine (approved eight years earlier) only in a fluorinated ring [70]. Emtricitabine shows to be eleven times more active against HIV-1 *in vitro* than lamivudine, but there is no evidences of an increased clinical effect *in vivo*.

In HAART, the idea in choosing two NRTIs is linked on the synergism of analogs. Between the purine family there are guanosine- (abacavir) and adenosine-analogs (didanosine and tenofovir), while between the pyrimidine family there are thymidine- (zidovudine and stavudine) and cytosine-analogs (lamivudine, emtricitabine and zalcitabine). Some combinations are not recommended due to overlapping toxicities or antagonistic effect (as stavudine with zalcitabine or zidovudine and didanosine with zalcitabine or stavudine). Triple combinations are not recommended for treatment failure and rapid onset of resistance. An exception is the combination abacavir-lamivudine-zidovudine (Trizivir®), used when an NNRTI or a PI can not or should not be used.

All the above-mentioned drugs are linked to the emergence of resistant HIV-1 strains. Resistance to NRTIs is mediated by two mechanisms: the first is the reversal of chain termination with removal of NRTIs from the 3' end of the chain, an ATP-dependent pyrophosphorolysis [71]; the second is an increased discrimination between native deoxyribonucleotides and NRTIs and therefore the prevention of NRTI incorporation into nascent chain.

Primary and secondary mutations in RT enzyme correlated to resistance have been shown to increase fitness over the WT virus. After interruption of therapy, reversion of most of these mutations is slower than substitutions in other enzymes as in protease, implying some compensations by secondary mutations on viral fitness.

Non-Nucleoside Reverse Transcriptase Inhibitors. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) target different sites in the enzyme to those targeted by the nucleoside analogues and they do not require intracellular metabolism to be active. They are highly selective for the binding to HIV-1 RT, being not able to inhibit RT of other lentivirus as HIV-2 and SIV [72]. These compounds bind the enzyme and form a hydrophobic pocket proximal to the polymerase active site [73]. Therefore they are defined non-competitive inhibitors. The

interaction with the nonsubstrate-binding site proximal to the catalytic residues changes the spatial conformation of the substrate-binding site and reduces the polymerase activity of RT. This secondary pocket exists only in presence of NNRTIs and it is not visible in the unliganded enzyme [74].

Three NNRTIs have been approved by FDA so far: nevirapine (NVP, Viramune®), delavirdine (DLV, Rescriptor®) and efavirenz (EFV, Sustiva®) (*Figure 45*).

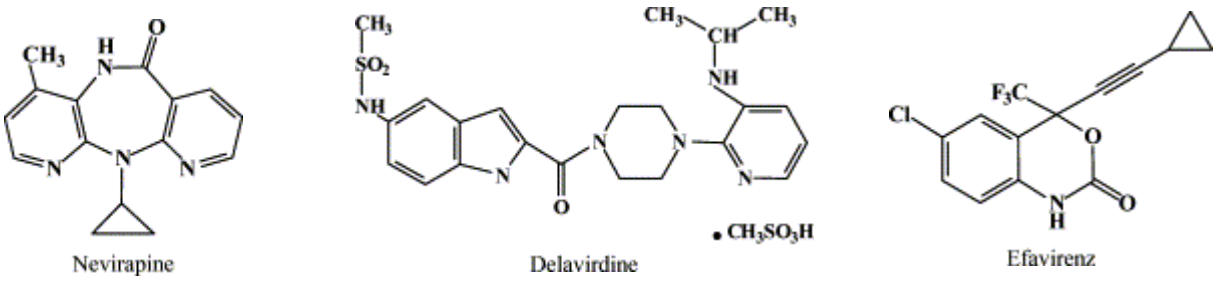

Figure 45: *chemical structures of Non Nucleoside Reverse Transcriptase Inhibitors.*

Efavirenz is the first choice based on many clinical studies but is contraindicated in pregnancy, determining neural tube defects. Nevirapine replaces efavirenz in patients experiencing central nervous system adverse effects or in patients intolerant to PIs.

The use of this class of drugs is correlated to emergence of high-level resistance, consistent with single amino acid substitutions located in the non catalytic site. These mutations can be selected after only a single dose of NNRTI. These modifications have limited effect on replicative viral fitness and persist in the absence of drug pressure. This can be justified by the fact that the NNRTI-binding pocket has only a structural role in the catalytic activity of RT and changes that maintain the general hydrophobicity and architecture will confer resistance without negative impacts on enzyme function.

Protease Inhibitors. Maturation of a retrovirus virion involves the cleavage by protease of the Gag and Gag-Pol polyprotein precursors to form the structural proteins and the enzymes of the new viral particle. All the PIs approved before Tipranavir (2005) are peptides fitting into the active site of the HIV-1 protease and mimicking the cleavage sites. This process results in fewer virions budded from HIV-1 infected cells.

For clinical use there are ten PIs approved by FDA: saquinavir (SQV, Invirase®, Fortovase®), ritonavir (RTV, Norvir®), indinavir (IDV, Crixivan®), nelfinavir (NFV, Viracept®), amprenavir (APV, Agenerase®), lopinavir (LPV), atazanavir (ATZ, Reyataz®),

fosamprenavir (Lexiva®), tipranavir (TPV, Aptivus®) and darunavir (TMC114, Prezista®) (*Figure 46*).

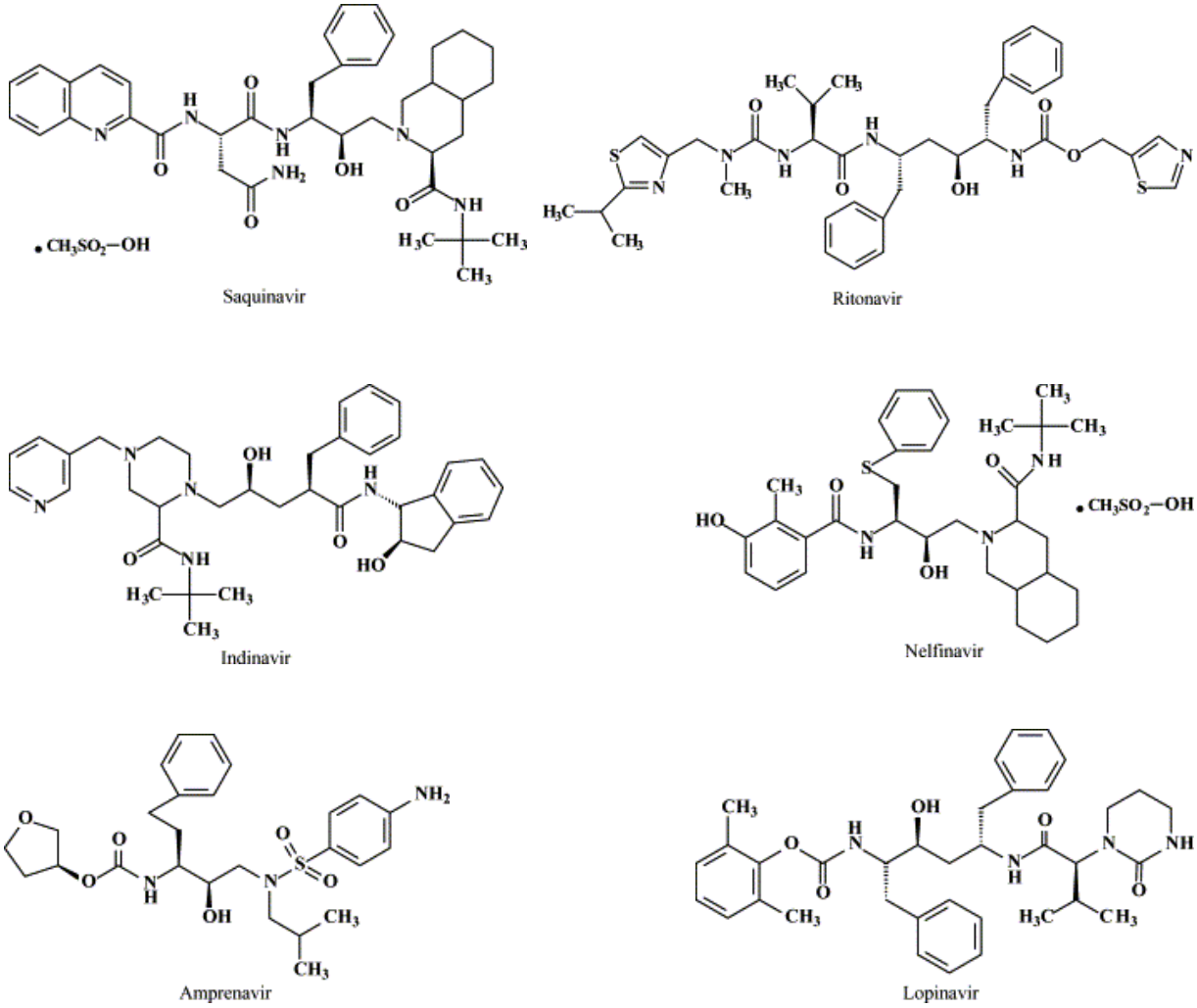

Figure 46: *chemical structures of some Protease Inhibitors.*

The combination lopinavir-ritonavir is the first-line agent used during HAART, followed by fosamprenavir or atazanavir boosted with ritonavir or as an nonboosted regimen.

Tipranavir is the first nonpeptide protease inhibitor and it was approved by FDA in 2005 and viruses carrying multiple protease inhibitor mutations show to be still sensitive to this drug. *In vivo* studies demonstrated that from 16 to 20 protease gene mutations are needed to confer resistance to tipranavir. Probably its molecular flexibility allows it to fit into the active pocket of the enzyme in viruses that have become resistant to other protease inhibitors. This compound is approved only for use in association with ritonavir. It shows a lot of drug-drug interactions and therefore it is contraindicated in combination with many classes of compounds and used only in salvage therapy.

Ritonavir is the most powerful inhibitor of this class and the concept to combine it with an other compound is one of the mainstays of therapy. In this way it is possible to have prolonged blood levels, increased potency, decreased dosage and reduced administration. Nelfinavir is not used as first-line agent because it is not really potent, but it is a useful tool in pregnant women. Indinavir and saquinavir are old protease inhibitors used only as alternative options.

In general it is preferred the use of NNRTIs instead of PIs in a patient's initial regimen because of adverse drug reactions and drug-drug interactions correlated to protease inhibitor therapy.

Initially it was expected infrequent and uncommon resistance to PIs during treatment because of the vital role of the enzyme in the life cycle of HIV-1 and its relatively small size. Unfortunately at least 20 amino acid substitutions have been directly associated with the use of this class of drugs. Resistance to PIs evolves in three steps. It starts with the appearance of primary resistance mutations in the protease gene that cluster near the active site of the enzyme at position located at the substrate/inhibitor-binding site. These substitutions usually have deleterious effects on both catalytic activity and viral replicative fitness [75]. During the second step of resistance's evolution there is the selection of secondary compensatory protease mutations with the purpose to repair the enzymatic function and rescue viral fitness. In fact these mutations partially compensate the impairment on HIV replication, being located outside of the substrate-binding region of protease and allowing a conformational adaptation to the primary changes in the active site [76]. Finally the last step is the selection of mutations in the major cleavage sites of the Gag and Gag-Pol polyprotein precursors, providing better peptide substrate for the mutated protease and compensating for the resistance-associated loss of viral fitness [77]. These substitutions appear to restore protein processing and to increase production of HIV-1 protease enzyme.

Interesting is the evidence that some viruses carrying a mutated protease display defects in the processing of the RT enzyme [78]. This reduction in the levels of RT in the virions results in a reduction in viral fitness. It has also been shown that mutations in RT developed after treatment with AZT can partially rescue the replicative defect of a PI-resistant virus.

Primary protease mutations are lost more rapidly than secondary compensatory mutations or primary mutations in RT, suggesting that primary PI resistance substitutions impose the greatest impairment on viral replicative capacity [79].

All the inhibitors of this category share relatively similar chemical structures and cross-resistance is commonly observed.

Fusion/Entry inhibitors. This new class of drugs prevents the entry of HIV virus into the CD4 cell. The entrance of the virus is fine regulated by a series of processes. The viral envelope glycoprotein complex (gp41 and gp120), the host cell CD4 receptors and the chemokine host cell receptors (CCR5 and CXCR4) are directly involved in this step. In particular the binding of gp120 to the CD4 receptor causes structural changes within the glycoprotein and determine the possibility of the gp120 itself to bind the CCR5 or the CXCR4. Once gp120 is bound to the chemochine receptor, the two helical region domains HR1 and HR2 of gp41 assemble into a six-helix bundle or hairpin structure that allows the passage of the HIV viral capsid into the host CD4 cell.

In March 2003, FDA approved enfuvirtide (T-20, Fuzeon®) (*Figure 47*), the first drug of this category. It is a synthetic 36-amino acids peptide structural analog of the HR2 domain of the gp41 and it binds to the HR1 region. It prevents the formation of the six-helix structure and the infection of the cell (*Figure 48*). It has much lower activity against HIV-2, which has a trans-membrane glycoprotein only distantly related to that of HIV-1. A disadvantage of enfuvirtide is that it has to be given by subcutaneous injection. Today it is recommended in treatment experienced patients who have had multiple treatment failures.

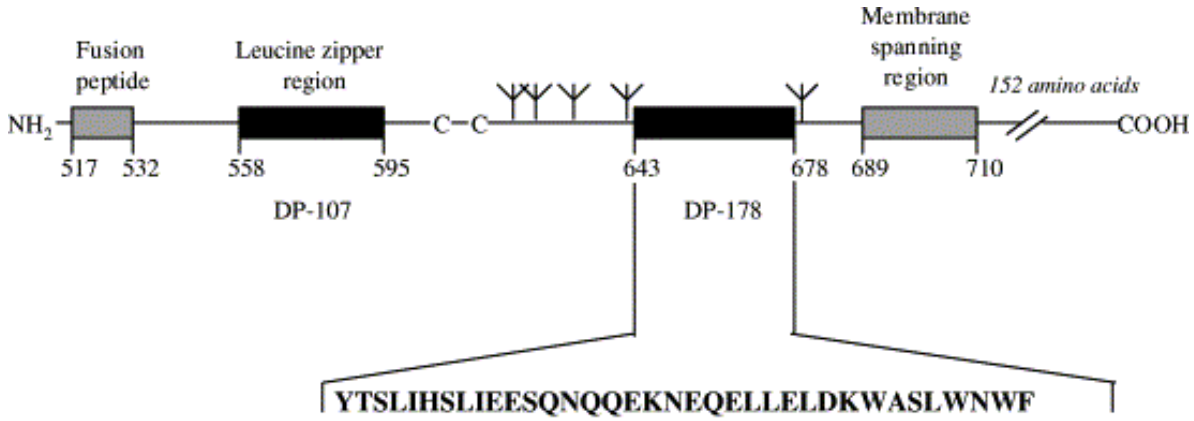

Figure 47: *structure of enfuvirtide.*

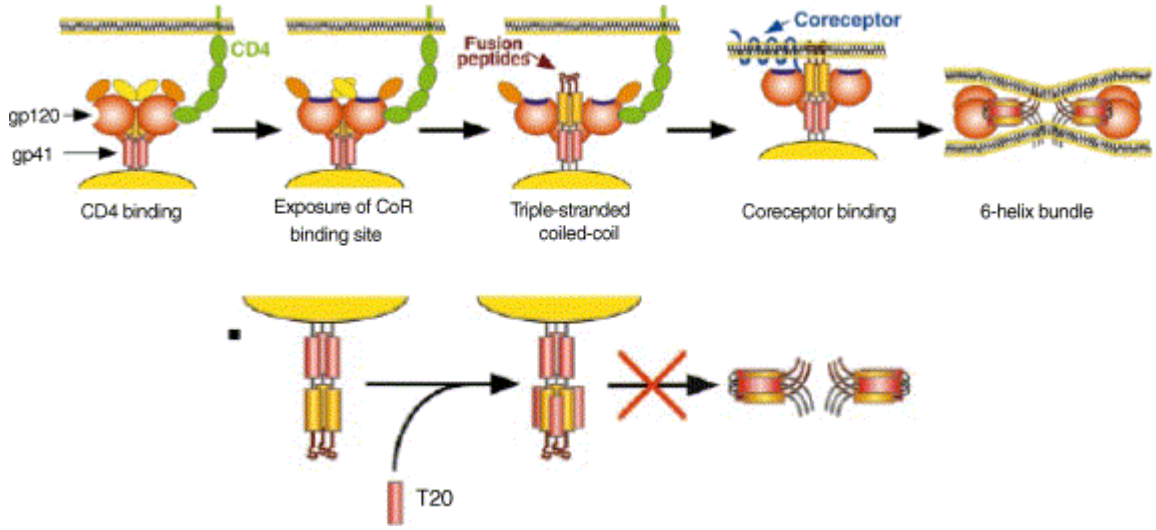

Figure 48: *model for the inhibition of fusion by enfuvirtide. After binding of CD4 to the gp120, there is exposure of a conserved region in gp120 that allows the binding to coreceptor. This brings the viral envelope in close proximity to the target membrane. The formation of the six-helix bundle of gp41 leads to membrane fusion. Enfuvirtide binds and prevents the transition to the six-helix bundle and membrane fusion. From [80].*

A very attractive treatment option is the use of CCR5 and CXCR4 antagonists. The CCR5 receptor is the primary mode of entry for HIV, but some viruses may use CXCR4 and others may use both. Patients can survive if CCR5 receptor is blocked because it is dispensable for human health. Unlike that, CXCR4 plays a critical role for homing, differentiation and function of hematopoietic cells and its absence is not tolerated over a long period of time.

Maraviroc is a CCR5 antagonist developed by Pfizer and recently approved for the treatment of antiretroviral-experienced patients who do not have detectable CXCR4-utilizing virus (*Figure 49*).

A problem of this class of drugs is the selective pressure between two receptors when one is inhibited. If CCR5 is blocked, the use of CXCR4 by HIV can be upregulated and can lead to an accelerated disease progression.

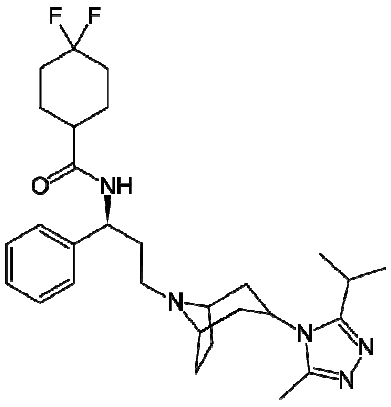

Figure 49: *chemical structure of maraviroc.*

Different evolutionary pathways will be required for the emergence of resistance to entry inhibitors (EIs), because HIV-1 *env* genetic diversity is great unlike reverse transcriptase and protease genes. Most drug resistance mutations associated with CCR5 antagonists have been identified in the V3 loop of the gp120. Nonetheless other mutations outside the V3 loop, spread throughout the gp120 and gp41, have been identified without understanding the connection between these substitution and the onset of resistance.

There is no certain ideas yet about the relation between EIs resistance and replicative fitness. In general, the fitness of a virus is strongly influenced by the *env* gene because the ability of a virus to bind a CCR5 and enter a cell plays an important role in the replication efficiency, and entry remains the bottleneck to "weed out" less fit variants. Consequently, resistance to EIs could be associated with increased entry efficiency and increased fitness. However some viruses could select for mutational pathways that confer drug resistance unrelated to the general process of entry and result in decrease fitness.

6.6 HIV-1 INTEGRASE INHIBITORS

The perpetual advent of antiretroviral-resistant HIV-1 strains has brought to the necessity to discover new validated pharmacological targets. The rationale for focusing the attention on integrase has been clear for many years. This enzyme is essential for retroviral replication but there are no evidences of a host-cell equivalent. This means that integrase inhibitors (INIs) should not interfere with normal cellular processes and therefore have a high therapeutic index. However the mechanistic and structural similarities between recombinases, RNases and integrases determine sometimes cross-reactivity.

To conclude that integrase is the target of an antiviral compound, four essential criteria need to be met. First, in time-of-drug-addition experiments, drug has to be effective between 4 and 16 hours following infection, consistent with the integration phase following reverse transcription and before maturation [81]. Second, treated cells must show decrease integration into host chromosomes and accumulation of 2-long terminal repeat circles derived from the accumulation of viral cDNA and its circularization by cellular enzymes [82]. Third, drug-resistant viruses must be characterized by the presence of integrase mutations [83]. Fourth, in biochemical assays performed using recombinant integrase carrying drug-resistant mutations, the drug should be inactive [84].

Crystallographic studies and consequently structure-based inhibitor design approaches are difficult and challenging, due to the low solubility of IN and its propensity to form aggregates

in solution. Still nowadays, 30 years after the discovery of the enzyme, only one example of a co-crystal structure of IN with an inhibitor has been determined [85].

The first large screening on the capacity of known compounds to inhibit IN was published in 1993 [86]. The original idea was to test whether DNA topoisomerase inhibition was correlated to integrase inhibition, because the two enzymes seemed to have analogies as the production of 5' overhangs in DNA, the ability to recombine cleaved double-stranded DNA and the possibility of a covalently linked intermediate. The inhibition of the two above-mentioned enzymes appeared to be not strongly correlated, consistent with the fact that some compounds resulted inhibitors and some others inactive. Therefore the study was expanded screening a broad spectrum of DNA-binding compounds to determine whether integrase inhibition was correlated with DNA binding, but also in this case there was no strong correlation.

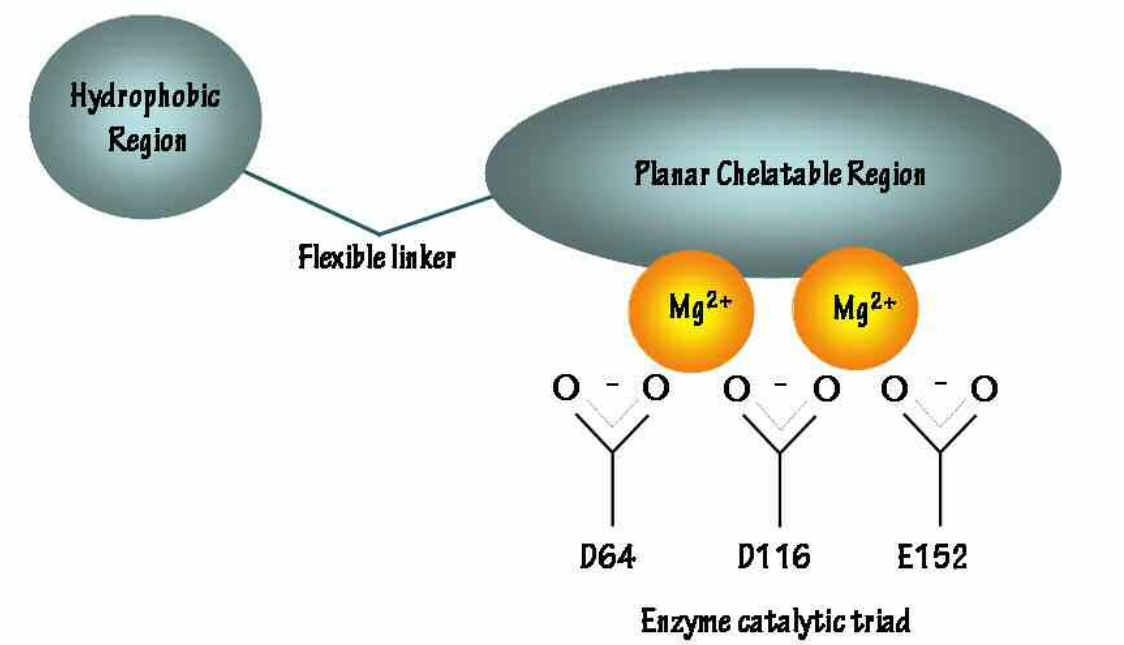

Figure 50: *minimal pharmacophore for the interaction with the catalytic active site of HIV-1 integrase.*

It was only in 2000 that Hazuda and coworkers [81] demonstrated the evidence that the diketo acid (DKA) function is the key motif responsible for the inhibition of the strand transfer step during integration. Since that day, a plethora of IN inhibitors has been discovered and β-diketo acids and their related compounds represent still nowadays the most convincing and biologically validated inhibitors of the viral enzyme.

The minimal pharmacophore-elements for the interaction with the catalytic active site of IN are a planar heteroatom two-metal chelation region, a flexible linker and a hydrophobic aromatic substituent (*Figure 50*). Every portion of this template has a specific effect. The

chelation region, usually constituted by a similar diketo acid function, binds the divalent metal cofactor(s) coordinated at the active site by the catalytically essential DDE motif. The aril group specifically interacts with an hydrophobic surface or pocket of the enzyme. The linker controls the right distance between the other two elements, allowing the proper position for the inhibitor effect.

Some of the earliest examples of a diketo acid with integrase inhibitory activity are L-731,988 and L-708,906 [81] (*Figure 51*) with strand transfer IC_{50} on recombinant IN corresponding to 0.050 μM and 0.100 μM respectively. Several modifications of the general scaffold and in particular the introduction of an heterocyclic system linked to the diketo acid moiety, brought to the development by Shionogi & Co of S-1360 (*Figure 51*), the first IN inhibitor to enter human clinical trials. It has a triazole directly connected to the diketo functionality. In *in vitro* assays, S-1360 showed to be very active and able to inhibit HIV-1 IN with an IC_{50} value of 0.02 μM. It also demonstrated potent antiviral activity against a variety of clinical HIV isolates. Unfortunately it failed efficacy tests due to *in vivo* metabolism and rapid body clearance via a non-cytochrome P450 pathway. In fact it is substrate for the aldo-keto reductase family of enzymes, carbonyl reducing enzymes in human liver. Therefore the conversion into an inactive metabolite, followed by a rapid clearance, is responsible of the lack of potent antiretroviral activity in infected patients and it is no longer being developed. It is interesting to note that the diketo functionality responsible for the specific inhibition of IN, is also the limitation for therapeutic efficacy in humans.

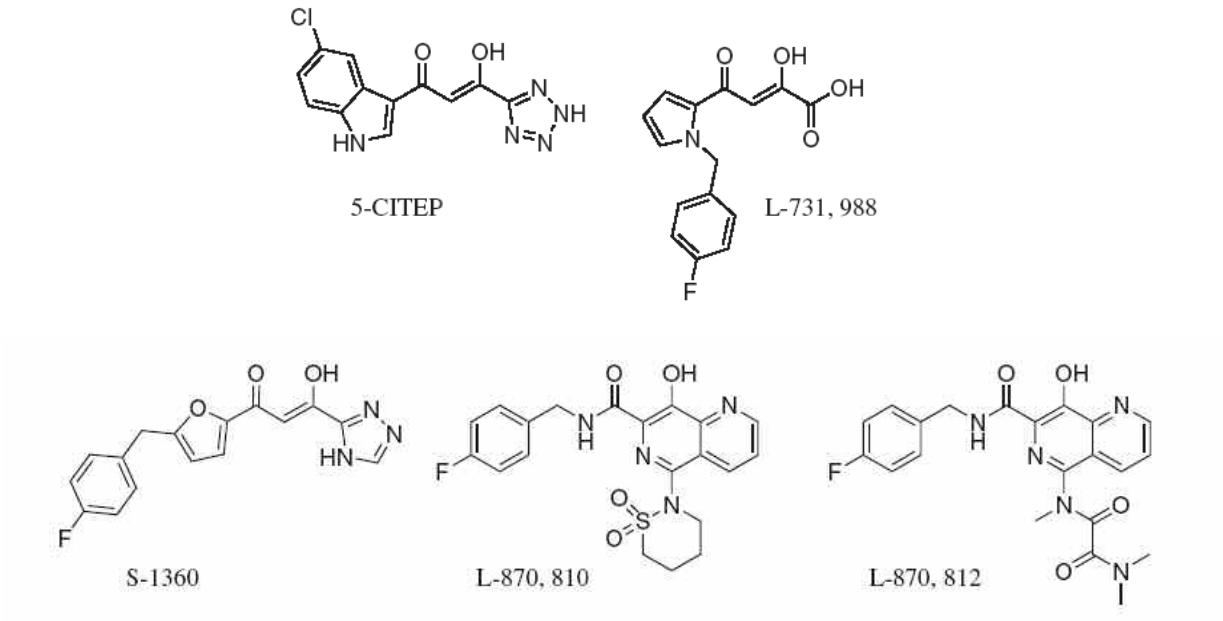

Figure 51: *chemical structure of some Integrase Inhibitors.*

Noteworthy was the study and the development by Shionogi of 5CITEP (*Figure 51*), the first, and still nowadays the only, integrase inhibitor co-crystallized with the enzyme [85]. The reported structure of the complex provided a platform for the design of derived compounds. In the crystal, 5CITEP is located centrally in the active site between the three acidic residues required for catalysis and it does not displace the bound magnesium ion, which remains complexed to the two aspartates. It is bound to the protein surface by a variety of hydrogen bonds between the tetrazolium ring and amino acids as Lys156 and Lys159. In addition the N1 atom of the indole interacts by hydrogen bond with Gln148. The inhibitor seems to mimic the DNA in the binding site with the distance between indolic and tetrazolic rings similar to that between two nucleotides. Unfortunately, the described structure displays only one of the two possible metal ions in the catalytic cavity. Some authors attributed the position of the inhibitor only to physical entrapment during crystallization, but biochemical data confirmed some observed interactions and thus the validity of the model proposed by *Goldgur* [87].

Savarino [88] extracted from the structure of *Goldgur [85]* the 3D coordinates of 5CITEP and transposed them into a two-metal model of IN CCD. The result of the docking is reported in *Figure 52*. The residue Q148 is at hydrogen bonding distance from the inhibitor, and it appears an additional contact with an other residue of the flexible loop, Y143.

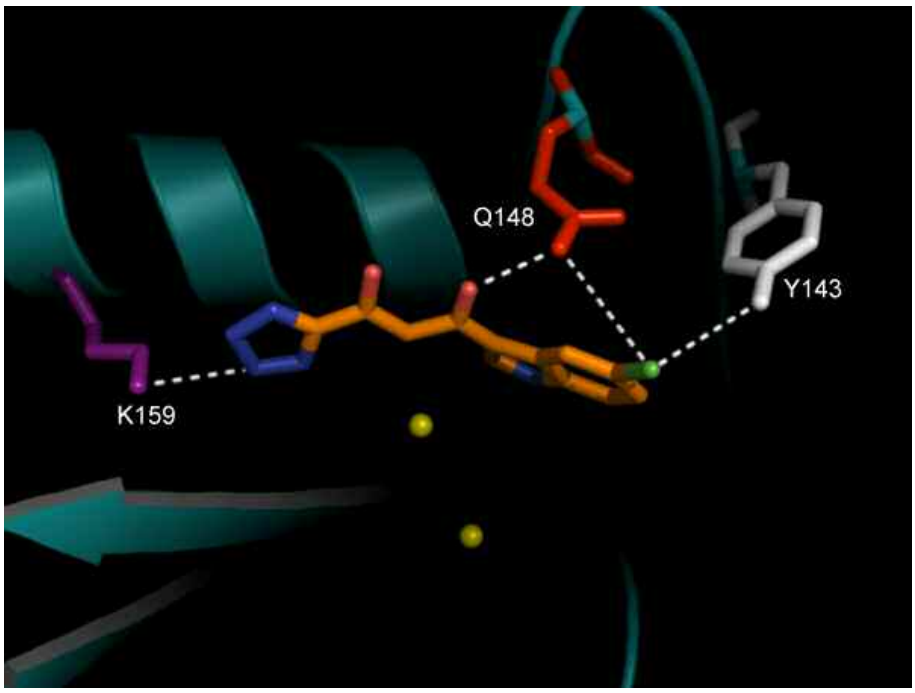

Figure 52: *possible interaction of 5CITEP with HIV-1 integrase. The amino acids involved in the interaction with the inhibitor are also DNA-interacting residues. Metal ions are shown as yellow spheres. Possible hydrogen bonds are shown as dashed lines. From [88].*

In early 2002, research at Merck laboratories conducted to the discovery of the second generation scaffold, with two important compounds belonging to this class: L-870,810 and L-870,812 (*Figure 51*). They result from the transfer of the diketoacid pharmacophore into a naphthyridine carboxamide core. In particular the naphthyridine nitrogen heteroatoms are used either as metal binding motif or as part of an internal hydrogen bond to maintain the co-planarity of the aromatic system and amido group. Clinical evaluation of L-870,810 showed success on short-term monotherapy in both naïve and treatment experienced HIV-1 infected patients. Unfortunately this compound was halted due to an observed long-term dosing toxicity in the liver and kidneys of dogs [89] and in addition, its high affinity for serum protein binding resulted in a lower effective plasma drug concentration.

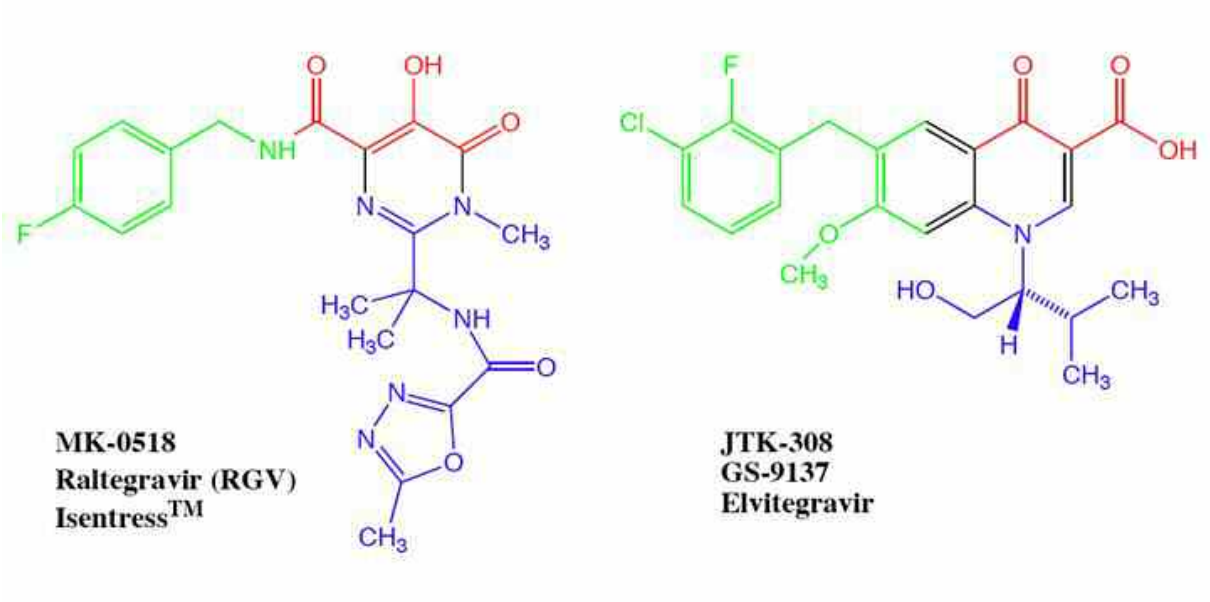

Figure 53: *chemical structures of raltegravir and elvitegravir. In red the simil diketo acid portions, in green the planar hydrophobic regions and in blue the additional chains important to improve physicochemical properties.*

Nowadays, the more advanced compounds are essentially two: GS-9137 and MK-0518 (*Figure 53*). GS-9137, known with the name of elvitegravir, is a quinolone carboxylic acid currently developed by Gilead Sciences and undergoing phase III of clinical trials. It has been discovered from researchers at Japan Tobacco through modification of antibiotic quinolones [90]. Its 4-quinolone-3-carboxylic acid scaffold provides an alternative for the diketoacid moiety. It shows *in vitro* a potent antiviral activity against HIV with a protein-binding adjusted IC_{50} of 16 nM and activity also against NRTI-, NNRTI- and protease inhibitor-resistant laboratory strains. *In vivo* elvitegravir showed to be orally bioavailable and well tolerated at all studied doses in healthy volunteers. It demonstrated an acceptable pharmacokinetic profile and no adverse drug interactions were reported when in combination with other clinically used HIV-1 inhibitors as emtricitabine and tenofovir. In a phase II study,

antiretroviral treatment-experienced patients using a 125 mg dose of elvitegravir (boosted with ritonavir) along with an active optimized background regimen showed >2-$\log_{10}$ declines in their viral loads that were durable through week 24 [91].

MK-0518, known as raltegravir and commercialized by Merck with the name of Isentress®, was the first IN inhibitor to be approved by FDA in October 2007 for the treatment of HIV-1 infection in combination with other antiretroviral agents. It is a pyrimidinone carboxamide fortuitously discovered from the evolution of DKA in the HCV polymerase program [92]. In particular the IRBM (Research Institute in Molecular Biology) in Rome developed the drug from the initial finding that the dihydroxypyrimidine carboxamide reported in *Figure 54* was a potent, reversible and selective HIV-integrase strand transfer inhibitor showing nanomolar activity in the enzymatic assay (IC_{50} = 0.085 μM) while being completely inactive on the HCV polymerase [93]. Substitutions at the 2-position of the core resulted to be well tolerated with respect to the IN inhibitory potency, indicating that this substituent is not involved in strong interactions with the enzyme active site, and allowed to improve physicochemical properties of the compound. Raltegravir is the result of the effort to optimize the initial structure.

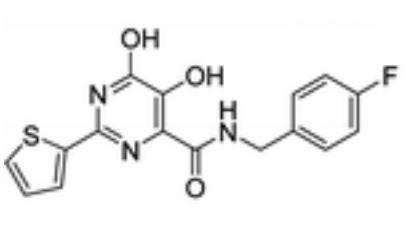

Figure 54: *raltegravir is the result of the optimization of this dihydroxypyrimidine, that demonstrated to be a potent, reversible and selective integrase inhibitor during a HCV polymerase program in Merck Research Laboratories.*

The major clearance mechanism for raltegravir is UGT1A1 mediated glucuronidation of the molecule in liver [94] and it is neither a substrate nor an inhibitor of the cytochrome P450 enzymes [95]. Thus, it is expected to have a reduced potential drug interaction profile with currently used antiretroviral agents. The absorption of raltegravir is rapid, with the acquisition of peak plasma concentration occurring with a median of 0.5 - 1.3 h. The terminal half-life is approximately 7 - 12 h. The pharmacokinetic profile supports for a twice-daily dosing schedule.

Two main clinical studies, BENCHMRK-1 and BENCHMRK-2, were designed to evaluate the safety and efficacy of MK-0518 given orally at a dosage of 400 mg twice daily versus placebo, in addition to optimized background therapy, in triple-class antiretroviral therapy-resistant HIV-positive patients. The results for both studies were similar, with a 77%

of patients achieving an HIV RNA level of <400 copies/ml with MK-0518 versus 41% of patients with placebo. Also the CD4 cell count confirmed the therapeutic efficacy, with value of 86 cells/mm^3 for the MK-0518 treated group versus 40 cells/mm^3 for the placebo group [96].

6.7 AN EMERGING ANTI-RETROVIRAL STRATEGY

The first integrase inhibitor was approved only about one year ago but nowadays literature reports already many evidences of the rapid onset of resistance after treatment with this class of drug. All the described and known compounds are in fact able to modify the enzymatic activity through an interaction with the catalytic site of the IN core domain. The binding of compounds in this specific area inevitably leads to the appearance of amino acidic point mutations in the surroundings of the DDE motif. This is a good reason to start studying a different approach for the inhibition of viral integrase, aimed to overcome the ineffectiveness of known compounds on resistant viral strain.

Throughout the viral life cycle, integrase interacts with a wide range of different viral or host-cell proteins. The disruption of these key interactions, indispensable for various functions as nuclear import, site-specific DNA targeting or virion assembly, could be a novel therapeutic approach for design and development of antiretroviral agents.

Currently one of the most studied interaction is between IN and LEDGF/p75 (Lens Epithelium-Derived Growth Factor). LEDGF/p75 belongs to the hepatoma-derived growth-factor family and it is an ubiquitous nuclear protein, tightly associated with chromatin throughout the cell cycle, not essential for cell or organism survival. Together with the LEDGF/p52 splice variant, it is a product of the PSIP1 human gene. The smaller p52 isoform lacks the domain responsible for the interaction with integrase. The protein has numerous functions and properties in cells. Above all, it increases cell survival following stress-induced apoptosis, through the induction of stress-related gene expression.

The chromatin association of LEDGF/p75 is mediated by three conserved sequence elements in the N-terminal domain of the protein : the PWWP domain (i.e. a Pro-Trp-Trp-Pro domain), the nuclear localization signal (NLS) and a dual copy of the AT-hook DNA binding motif [97]. Instead of this, the domain that mediates the interaction with integrase is located in the c-terminal portion and it is called IBD, integrase binding domain (*Figure 55*).

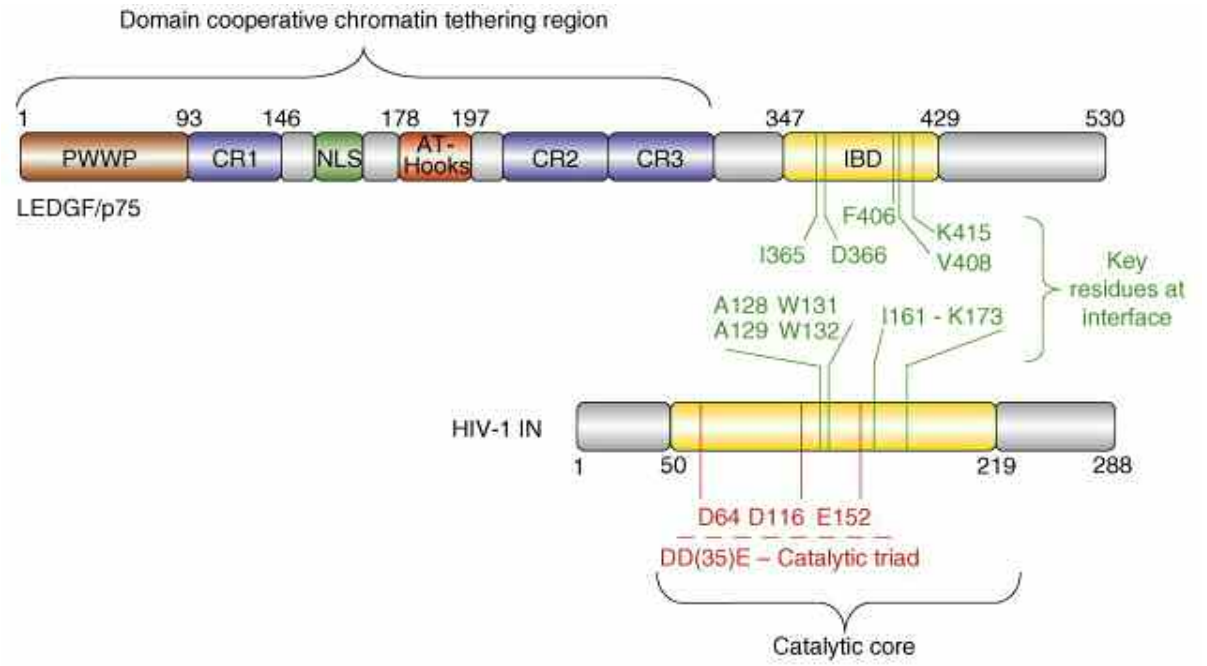

Figure 55: *organization of integrase and LEDGF/p75 domains. In green, the amino acids at the interface contact region. LEDGF/p75 domains: in brown the Pro-Trp-Trp-Pro domain (PWWP domain), that mediate protein-protein and protein-DNA interactions; in blue the Charged Regions (CR), which contain many polar amino acids; in green the Nuclear Localization Signal (NLS), in red, the two AT-hook motifs that bind AT-rich DNA regions; in yellow, the Integrase Binding Domain (IBD). From [98].*

Cherepanov and co-workers obtained for the first time the coprecipitation of the protein with HIV-1 IN [99] and cocrystallized the two components together allowing the understanding of the basis for the recognition [100]. The IBD forms a right handed α-helical hairpin domain possessing structural similarities to HEAT repeat domains, and it burrows into a cleft created by the IN dimer interface (*Figure 56*). In particular, two LEDGF/p75 IBD domains bind at nearly identical positions at either end of the IN CCD dimer, interacting with residues 166-171 from one IN monomer (between α-helices 4 and 5 of IN) and with hydrophobic areas of α-helices 1 and 3 of the second IN monomer.

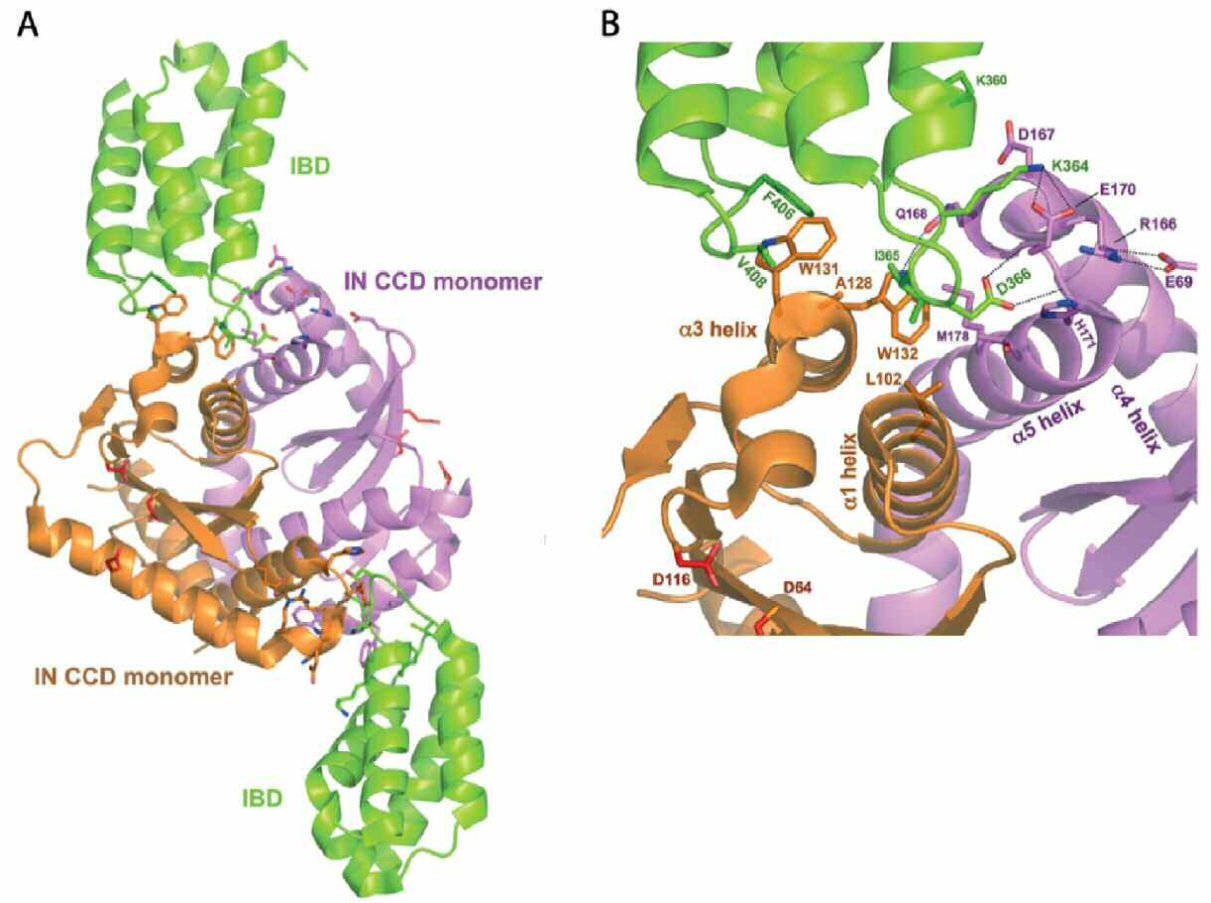

Figure 56: *structure of IN-LEDGF/p75 complex.* ***(A)*** *Structure of the dimer.* ***(B)*** *Enlarged image of the molecular interface (PDB code 2B4J). IBD of LEDGF/p75 is in green, IN CCD monomers in orange and pink. DDE motifs are in red. From [101].*

Numerous literature data showed as LEDGF-p75 is important for lentiviral integration. However it is not strictly essential because LEDGF-depleted cells maintain a level of integration around 10% [102]. A model to explain its function in the HIV-1 integration was proposed by *Engelman* and *Cherepanov* [97] after accurate evaluation of all the performed and reported studies. They proposed that the factor is the key that retroviruses use to do not leave integration entirely to chance and consequently to have a mechanism for selecting suitable target loci. The chromatin-associated factor acts as a receptor for incoming PIC particles, encouraging the IN strand transfer activity in the nearby genomic locus. If the factor is reduced or ablated, a larger fraction of PICs will rely on a slower, cofactor-independent pathway, extending the time for integration and increasing the opportunity for the cell to destroy them by ubiquitination and proteosomal degradation (*Figure 57*).

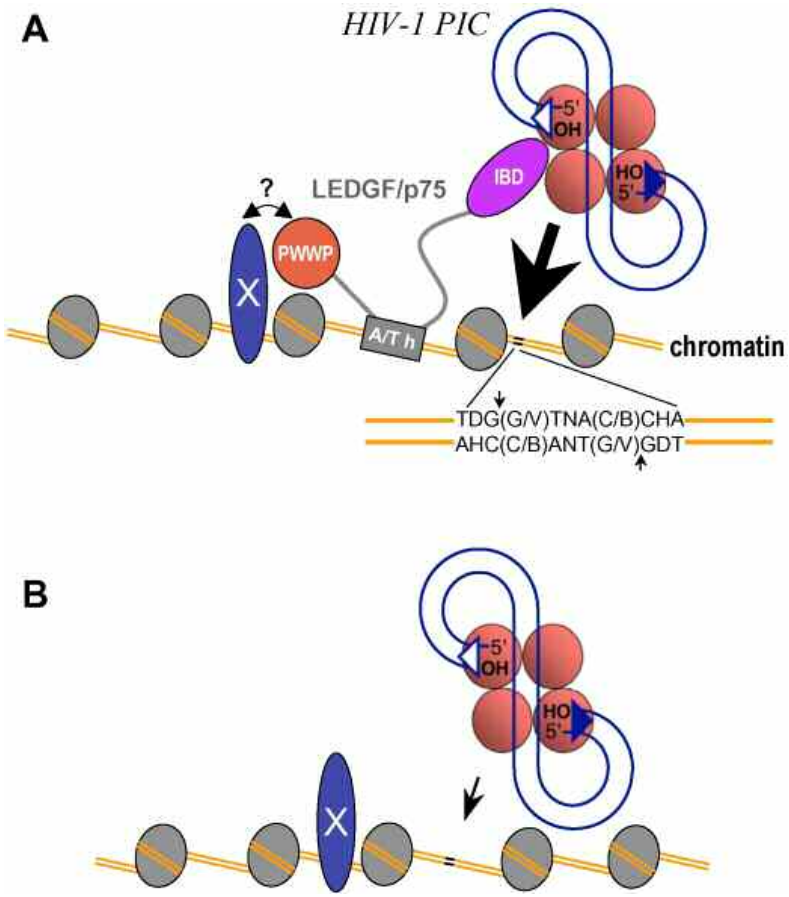

Figure 57: *model for site-directed integration. Histone proteins are shown as gray ovals and chromosomal DNA as orange lines.* ***(A)*** *LEDGF/p75 engages chromatin directly via NLS or AT-hook motifs, or indirectly by interaction of PWWP with histone proteins or unknown chromatin factors (blue ovals X). Therefore integration is encouraged at a nearby position.* ***(B)*** *PIC can engage the DNA also in absence of LEDGF/p75, but the level of integration is 10-fold reduced under this condition. From [97].*

On can define LEDGF/p75 as a chromatin docking factor, with ability to determine IN trafficking in cells, to allow an efficient viral replication and to influence the genome-wide integration distribution of the retrovirus. Therefore the possibility to disrupt the p75-IN interaction would produce therapeutically beneficial results, although drug development on a protein-protein interaction is more challenging than enzyme active sites, because they are much larger, flatter, with high affinity and there are not available antagonist molecules used as template. Research is trying to obtain a small molecule with sufficient binding affinity to occupy the pocket and preclude LEDGF/p75 from binding [103]. In addition, peptides derived from the face of the IBD that interacts with IN CCD have been reported to block integration in a noncompetitive manner.

Nowadays, this seems to be one of the best approach to evolve for the targeting of the integration step but, as it is reported on *Figure 55*, numerous groups are working on several other targets.

Targets and related therapies			
Target	**Strategic approaches to target**	**Expected outcome of intervention at target**	**Who is working on the target**
BAF[a]	DNA binding Interaction with LAP2α[b] Interaction with MA[c]	Stimulation of autointegration	Craigie Engelman
HMGA1[d]	DNA binding	Inhibition of chromosomal integration	Bushman Leis
INI1[e]	DNA Binding Interaction with IN[f]	Inhibition of chromosomal integration Inhibition of virus assembly	Kalpana
LEDGF/p75[g]	DNA binding Interaction with IN	Inhibition of nuclear import and chromosomal integration	Bushman Debyser Engelman Poeschla
HRP2[h]	Interaction with IN	Inhibition of nuclear import and chromosomal integration	Engelman
p300 acetyltransferase	Interaction with IN IN acetylation	Impaired proviral integration	Giacca
HSP60[i]	Interaction with IN	Impaired IN folding	Litvak
Polycomb group EED[j]	Interaction with IN	Inhibition of nuclear import and chromosomal integration	Boulanger
MA	Interaction with IN Interaction with BAF	Inhibition of nuclear import	Bukrinsky Bushman Wilson
Vpr[k]	Interactions with UNG[l] Vpr-mediated cell-cycle arrest Interactions with nucleoporins and importins Aida	Inhibition of nuclear import Stimulation of viral mutation rates	Benichou Burkinsky Green Zeichner
NC[m]	DNA and RNA binding Interaction with RT[n]	Inhibition of reverse transcription Inhibition of integration	Darlix Gorelick
RT	RNA binding Interaction with NC Interaction with IN	Inhibition of reverse transcription Inhibition of integration	Chow Darlix Prasad Roques

Figure 58: *targets and related therapies. [a]Barrier to Autointegration Factor; [b]Lamina-Associated Polypeptide 2α; [c]Matrix; [d]High Mobility Group Protein A1; [e]Integrase Interactor 1; [f]Integrase; [g]Lens Epithelium-Derived Growth Factor/p75; [h]Hepatoma-Derived Growth Factor Related Protein 2; [i]Heat Shock Protein 60; [j]Polycomb Group Embryonic Ectoderm Development Protein; [k]Viral Protein R; [l]Uracil DNA Glycosylase; [m]Nucleocapsid; [n]Reverse Transcriptase. From [63].*

6.8 HIV VACCINATION: AN ILLUSION?

25 years of efforts have already been made trying to develop an AIDS vaccine. In the far 1984, the U.S. Health Secretary Margaret Heckler proclaimed that, in two years from then, a preventive HIV vaccine would have been available [104]. But still nowadays the development of this therapeutic strategy does not look very promising. Difficulties are linked to numerous factors. HIV has an enormous sequence diversity, with high capacity of mutation and adaptation thanks to an error-prone reverse transcriptase, a high propensity for recombination,

that occurs when one person is co-infected with two separate strains of the virus, and a high replication rate. In addition, the virus has evolved strategies to impede the immune attack. For example surface envelope viral proteins avoid antibody recognition, and Nef protein is able to down-regulate molecules responsible for the T cell recognition of infected cells, as the major histocompatibility complex (MHC). Besides that, integration of the viral DNA into the host genome permits to establish a latent reservoir of infected lymphocytes, an irreversible process that ends only after cell death. In this way the infection is maintained and the virus is immunologically silent until the cell will be activated. Finally, a further problem for vaccine development is the immune impairment the first period after infection. In fact HIV targets $CD4^+$ T lymphocytes, determining a massive loss of memory $CD4^+$ T cells and therefore resulting in a defective immune response.

Together, all these factors are responsible for the failure of this strategy. A recently tested vaccine, developed with a collaboration of Merck and National Institute of Health, demonstrated to do not protect against infection and to do not contain virus replication in already infected persons, but it apparently also increased the susceptibility to infection of those presenting antibodies to the adenovirus vector used to deliver the vaccine [105].

The future in this field is the exploring of alternative approaches to vaccination, studying nonclassical routes to antibody-mediated protection. This is a difficult way but, given the current state of knowledge, research community has no intention to give up.

7. AIM OF THE WORK

Since the first cases reported in 1981 [106-108], acquired immunodeficiency syndrome (AIDS) is still nowadays an incurable disease, also if nevertheless largely manageable. 33 million people worldwide are infected with HIV, the causative agent for AIDS, and every year 2.5 million of new infection and 2.1 million of death are correlated to this virus. An extraordinary scientific and therapeutic progress in this field led to the advent of the highly active antiretroviral therapy (HAART) that consists of a protease inhibitor (PI) or a non-nucleoside reverse transcriptase inhibitor (NNRTI) in combination with two nucleoside reverse transcriptase inhibitors (NRTI). However, prolonged therapy usually leads to drug resistance. One way to overcome such resistances and to reduce therapeutic side effects is to identify novel targets for therapy. Recently, three different classes of drugs have been approved as treatments for HIV infection: fusion inhibitors with the injectable peptide enfuvirtide (Fuzeon®); chemokine co-receptor (CCR) inhibitors with maraviroc (Celsentri®); integrase inhibitors with raltegravir (Isentress®).

Integrase is a good target for a therapeutic approach because it is absolutely required for viral replication and it has no cellular equivalent. The US Food and Drug Administration recently approved the first integrase inhibitor (raltegravir, Merck & Co.) for the treatment of HIV-1 infection, 14 years only after the discovery of the first integrase inhibitors in 1993 [86]. Raltegravir is administered to treatment-experienced adult patients who have evidence of viral replication and HIV-1 strains resistant to multiple antiretroviral agents (http://www.fda.gov/bbs/topics/NEWS/2007/NEW01726.html). Elvitegravir (co-developed by Gilead Sciences Inc. and Japan Tobacco Inc.) is the other IN inhibitor in advanced stage of human clinical trials. Despite major clinical activity of raltegravir in the treatment of multidrug refractory patients, ongoing studies have already evidenced mutations in the IN gene associated with virological failure in patients receiving raltegravir [Merck Protocol 005] [67, 109]. HIV is in fact a viral pathogen characterized by a high mutational rate and the high levels of virion propagation ensure the selection of mutant viral strains displaying an increased fitness under therapeutic pressure.

The aim of the present work was the expression and the *in vitro* study of recombinant WT integrase and of selected IN mutants reported to be involved in drug resistance, the evaluation of the impact of these single amino acid substitutions on IN biochemical catalytic activities and the comparison of raltegravir and elvitegravir to determine their relative potency and their selectivity for strand transfer, 3'-processing and disintegration against WT integrase. The

potential reduction in sensitivity of the selected IN mutants to both drugs was evaluated and a cross-resistance pattern was determined.

8. EXPERIMENTAL PROCEDURES

8.1 OLIGONUCLEOTIDE SYNTHESIS, PURIFICATION AND LABELING

Oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, IA) or from Midland Certified Reagent Company, Inc. (Midland, Texas).

All oligonucleotides were purified on denaturing 20% polyacrylamide gels. A high concentrated solution of the oligonucleotide was loaded in the gel and run for about 2 hours at 70 Watts. The band corresponding to the DNA of interest was then visualized with UV irradiation and was cut. The oligonucleotide was eluted from the band using the Elutrap Electroeluition System (Whatman - Schleicher&Schuell). In particular this device consists of a sample chamber limited at each end by an inert membrane with a dense matrix through which buffer ions and molecules less than 3-5 kD can pass under the influence of an electric field. On one side there is also a microporous membrane that acts as a prefilter that prevents acrylamide and other particulates from entering the purified sample. Together, the membranes form a "trap" into which the sample migrates. The eluted oligonucleotide is precipitated overnight at -20 °C in presence of 0.3M sodium acetate pH 5.2 and 3-volume of 100% ethanol.

Single-stranded oligonucleotides were 5'-labeled using T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) with [γ-^{32}P]ATP (Perkin-Elmer Life and Analytical Sciences, Boston, MA). A 200 nM oligonucleotide solution was incubated at 37 °C for 30 minutes in presence of 1X T4 Polynucleotide Kinase Reaction Buffer, 100 μCi of [γ-^{32}P]ATP and 10 units of enzyme in a final volume of 50 μl. Unincorporated nucleotides were removed by Mini Quick Spin Oligo Columns (Roche). The duplex DNA was annealed by addition of an equal concentration of the complementary strand, heating to 95°C for 4 minutes, and slow cooling to room temperature. Sequences of the oligonucleotides used are indicated in *Table 2*.

Table 2: *oligonucleotide sequences used in assays.*

Oligo	Size	Sequence
a	21-mer	5'-GTG TGG AAA ATC TCT AGC AGT-3'
b	21-mer	5'-ACT GCT AGA GAT TTT CCA CAC-3'
c	19-mer	5'-GTG TGG AAA ATC TCT AGC A-3'
d	15-mer	5'-GAA AGC GAC CGC GCC-3'
e	34-mer	5'-GTG TGG AAA ATC TCT AGC AGG GGC TAT GGC GTC C-3'
f	30-mer	5'-GGA CGC CAT AGC CCC GGC GCG GTC GCT TTC-3'

8.2 DRUGS

Raltegravir (RGV) and elvitegravir (EVG) were obtained from the NIH Chemical Genomics Center (National Human Genome Research Institute, NIH, Bethesda, MD). Raltegravir (MK-0518) was purified directly from pharmaceutically available tablet formulation (both extraction and HPLC purification was performed). Elvitegravir (JTK-303) was synthesized in a similar manner to known procedures and purified via preparative HPLC. Purity according to LCMS analysis was >95%. ^{1}HNMR (DMSO-d6, 400M Hz) (δ) ppm: 0.69 (d, J=6.8 Hz, 3H), 1.10 (d, J=6.4, Hz, 3H), 2.25-2.39 (m, 1H), 3.70-3.76 (m, 1H), 3.88-3.97 (m, 1H), 3.99 (s, 3H), 4.05 (s, 2H) 4.78-4.84 (m, 1H), 5.13-5.18 (m, 1H), 7.10-7.20 (m, 2H), 7.38-7.44 (m, 2H), 7.96 (s, 1H), 8.81 (s, 1H). LRMS calculated for $C_{23}H_{23}ClFNO_5$ (M + H) 448.1; found 448.1.

8.3 MUTAGENESIS

IN mutants were created using the Stratagene QuikChange Site-Directed Mutagenesis Kit (La Jolla, CA).

Two complimentary oligonucleotide primers containing the desired mutation, flanked by unmodified nucleotide sequence, were designed for each mutant. A mix containing 10 ng of the plasmid DNA coding for the integrase (the HIV-1 integrase gene has been cloned into a pET-15b ampicillin-resistant vector – Novagen, Milwaukee, WI), 1X Reaction Buffer, 125 ng of each primer, dNTP mix and 2.5 units of *Pfu Turbo* DNA polymerase, in a total volume of

50 µl, was prepared. The mix was incubated in a thermal cycler using the following parameters:

Cycle	*Temperature*	*Time*
1	95 °C	30 seconds
2	95 °C	30 seconds
3	55 °C	1 minute
4	68 °C	10 minutes
5	68 °C	7 minutes
6	4 °C	hold

Cycles from 2 to 4 were repeated for a total of 12 times. Following temperature cycling, 10 units of the *Dpn I* restriction enzyme was added directly to the amplification reaction and this was gently and thoroughly mixed by pipetting the solution up and down several times. The reaction was then spinned down for 1 minute and incubated at 37 °C for 1 hour to digest the parental supercoiled dsDNA.

XL1-Blue Supercompetent Cells were transformed by the amplified DNA using a heat shock treatment. Cells were incubated on ice for 30 minutes in presence of the plasmid. The mixture was then heated for 45 seconds at 42 °C and placed on ice again for 2 minutes. 500 µl of NZY^{+} broth were added to the transformation reaction and incubated at 37 °C for 1 hour with shaking at 225-250 rpm before being spread on LB-agar plates containing ampicillin (100 mg/ml; KD Medical, Columbia, MD) and incubated at 37 °C overnight.

The day after some colonies were picked up and amplified in 5 ml of Luria-Bertani (LB) broth medium (KD Medical, Columbia, MD) with ampicillin (100 µg/ml final) for 6 hours at 37 °C with shaking. From this bacterial solution was then extracted the plasmid DNA using the Qiagen QIAprep Spin Miniprep Kit (Valencia, CA), according to the manufacturer's instructions. The presence of desired mutations and the integrity of the remainder of the IN sequence were verified by DNA sequencing. The plasmid concentration was determined by UV absorption.

With the extracted plasmid DNA, BL21 Star^{TM} (DE3)pLysS were transformed using the same protocol described for XL1-Blue Supercompetent Cells. As already described, cells were then spread on an LB-Agar plate containing ampicillin and grown overnight at 37 °C to select a single clone. A single colony was grown in 5 ml LB supplemented with ampicillin. The obtained transformed bacteria were conserved at -80 °C in presence of 50% glycerol.

8.4 INTEGRASE EXPRESSION AND PURIFICATION

Cells from the glycerol stock were spread on LB-agar plates containing ampicillin and let grow at 37 °C overnight. The colonies were then harvested, expanded in 1000 ml of fresh LB-ampicillin broth and incubated at 37 °C with shaking at 190 rpm, until an absorbance of 0.5 at 600 nm was reached. The expression of integrase was then induced adding isopropyl-beta-D-thiogalactopyranoside (IPTG; Sigma) to a final concentration of 0.4 mM and incubating for a further three hours. Cells were then centrifuged at 3000 rpm for 10 minutes.

The pellet was resuspended in ice cold Lysis Buffer (see *Table 3* for buffers' composition) (15 ml for 1000 ml of culture) and lysozyme (0.2 mg/ml final concentration) was added to the solution. The lysate was incubated on ice for 30 minutes with occasional stirring until the suspension became viscous and then it was sonicated on ice for approximately 1 minute. The lysate was centrifuged for 20 minutes at 30000 g and 4 °C to pellet the cellular debris. During this time, 1 ml of Chelating Sepharose^{TM} Fast Flow (Amersham Biosciences - Uppsala Sweden) was loaded into a 10 ml Poly-Prep Chromatography Columns (BioRad - Hercules, CA), followed by approximately 10 ml of water. After complete elution of water, 10 ml of a solution 0.4 M of $NiSO_4$ was added to the matrix in the column, followed by 10 ml of Eluition Buffer containing 5 mM imidazole. At that point the supernatant derived from the centrifugation of the bacteria was loaded in the column to permit the binding of the protein to the matrix. The column was then washed from all the impurities with Eluition Buffer at increasing concentration of imidazole (20 mM, 60 mM and 250 mM). The protein was eluted adding 750 mM imidazole Eluition Buffer and the first 3 ml fraction was recovered and dialyzed overnight in Slide-A-LyzerR Dialysis Cassette 10000 MWCO (Thermo Scientific, Pierce Protein Research Products - Rockford, IL). A small amount of each fraction derived from the purification process was loaded in a SDS-PAGE (16% Tricine Gel, Invitrogen) to determine the purity of the separated protein and run at 100 Volts for 1 hour and a half. Protein Ladder: MultiMark Multi-Colored Standard (Invitrogen).

The following day an additional 30 minutes dialysis was performed in presence of DTT to a final concentration of 5 mM in the same Dialysis Buffer and then the recovered protein solution was quantified by determination of the Absorbance at 280 nm, using an extinction coefficient of 50460 mol^{-1} cm^{-1} l.

***Table 3**: buffers' compositions for protein purification.*

Lysis Buffer	
PIPES pH 6.8	20 mM
NaCl	1 M
Imidazole	5 mM
Glycerol	10%

Eluition Buffers	
PIPES pH 6.8	20 mM
NaCl	0.5 mM
Imidazole	5 mM, 20 mM, 60 mM, 250 mM or 750 mM[1]
Glycerol	10%
Glycerol	10%
β-mercaptoethanol[2]	2 mM

[1]The 750 mM imidazole Eluition Buffer contains also 50 μM $ZnCl_2$.

[2] β-mercaptoethanol must be fresh added to the solutions at the moment of the use

Dialysis Buffer	
PIPES pH 6.8	25 mM
NaCl	750 mM
EDTA	0.1 mM
Glycerol	50%

8.5 INTEGRASE REACTIONS

IN reactions were carried out by mixing 20 nM DNA with 400 nM IN (unless otherwise indicated) in a buffer containing 20 mM MOPS, pH 7.2, 7.5 mM $MgCl_2$ or $MnCl_2$, 14 mM 2-mercaptoethanol, and the drug of interest or 10% DMSO. Reactions were incubated at 37°C for 1 h and quenched by addition of an equal volume of gel loading dye (formamide containing 1% SDS, 0.25% bromophenol blue and xylene cyanol). Reaction products were separated in 16% polyacrylamide denaturing sequencing gels. Dried gels were visualized

using the Typhoon 8600 (GE Healthcare, Piscataway, NJ). Densitometry analyses were performed using ImageQuant software from GE Healthcare.

8.6 SEQUENCING GELS

A solution 16% (w/v) acrylamide/bisacrylamide (19:1), containing 7 M Urea and 1X TBE, was prepared. To 50 ml of this solution, 300 µl of 10% ammonium persulfate (APS - Sigma) and 30 µl of N,N,N',N'-Tetramethylethylenediamine (TEMED - Sigma) were added. The resulting solution was then immediately poured into the assembled glass plates and left to polymerize for a minimum of 30 minutes in presence of a selected gel comb.

8.7 CYTOPATHIC ASSAY IN MT-2 INFECTED CELLS

This assay directly evaluates the antiviral activity and the toxicity of drugs in cells infected with HIV-1 using the Cell Counting Kit-8 (CCK-8, Dojindo Molecular Technologies Inc., Gaithersburg, MD). A 96-well plate was prepared with 50 µL of compounds' solutions at the desired concentrations. In a 15 ml falcon tube the desired amount of MT-2 cells (amount per well: 3×10^3 cells in 50 µL of RPMI-based culture medium with 10% fetal calf serum with 50 U ml-1 penicillin and 50 mg ml-1 streptomycin) were infected with the equivalent of 100 $TCID_{50}$ of HIV-1 ($TCID_{50}$: 50% tissue-culture infective dose, defined as the quantity of a cytopathogenic agent that will produce a cytopathic effect in 50% of the cultures inoculated) and instantly plated. Half of the plate was used as control to evaluate drugs' toxicity and infected cells were added to the same drugs' concentrations. The plate was then incubated for 7 days. After that period of time, 10 µL of the CCK-8 solution were added to each well and the plate was incubated again for 2 hours. The absorbance at 450 nm was measured using a microplate reader and the viability curve depending on the drug concentration was determined from the obtained values. In this assay, the cytopathic effect of HIV-1 requires several rounds of infection and therefore is susceptible to inhibitors that act on the late stages of viral replication. The concentration of drug required to inhibit HIV-1 induced cytotoxicity by 50% was calculated from the plot of drug concentration versus the percent viable cells compared to the untreated HIV-1 infected control. The concentration of drug that induced 50% cytotoxicity in uninfected cells was calculated from the plot of drug concentration versus the percent viable cells in uninfected cells.

9. RESULTS

9.1 EXPRESSION OF PROTEINS AND STUDY OF THEIR IN VITRO ENZYMATIC ACTIVITY

As already described, one of the aims of the work is the *in vitro* study and characterization of HIV-1 WT integrase and of a series of drug-resistant integrase mutants. To reach this goal we started with the expression of all the proteins of interest as described in the *Experimental Procedures* section.

HIV-1 Wild-Type Integrase. After expression in *E. coli*, the purity of the obtained protein was checked by SDS-PAGE. In particular, a small amount of each fraction derived from affinity chromatography purification was loaded in a protein gel and run for about 1 hour and a half at 100 Volts (*Figure 59*). The electrophoresis enabled to identify fractions containing the protein and to dialyze them obtaining a 50% glycerol stock to keep at -80 °C.

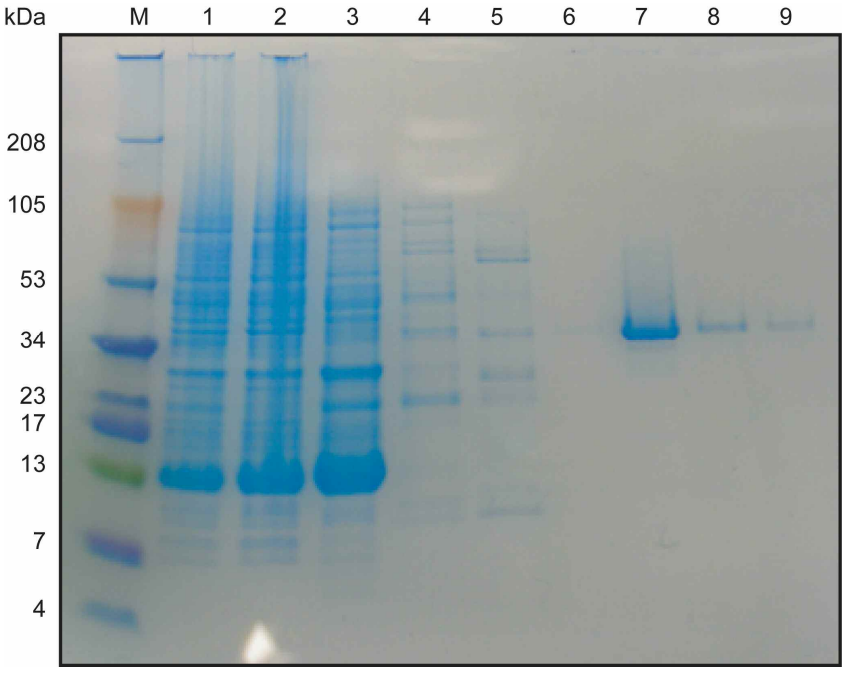

Figure 59: *purity of HIV-1 WT integrase after affinity chromatography. 16% Tricine Gel run at 100 Volts for 1 hour and a half. M, protein ladder; lane 1, solution before purification; lane 2, flow through; lane 3, 20 mM imidazole Eluition Buffer; lane 4, 60 mM imidazole Eluition Buffer; lane 5, 250 mM imidazole Eluition Buffer; lane 6, 750 mM imidazole Eluition Buffer, first five drops; lane 7, 750 mM imidazole Eluition Buffer, first 1.5 ml aliquote; lane 8, 750 mM imidazole Eluition Buffer, second 1.5 ml aliquote; lane 9, 750 mM imidazole Eluition*

Buffer, third 1.5 ml aliquote. Aliquots 7 and 8 were selected and dialyzed overnight. It is in fact possible to see integrase WT in these lanes as a clear band around 34 kDa.

The catalytic activity of the expressed enzyme was then tested using an *in vitro* assay developed in our laboratory [110, 111]. In fact the integration reaction can be reproduced and monitored using the recombinant HIV-1 integrase and a ^{32}P radiolabeled DNA substrate (*Figure 60*). This substrate consists of a single DNA duplex mimicking the last 21 nucleotides of the HIV-1 U5 LTR. The cleavage reaction (3'-P) is observed as accumulation of a 19-mer DNA band on a denaturing sequencing gel. Another identical duplex serves as a target DNA, and the ST reaction products (STPs) are observed as bands migrating higher than the substrate 21-mer, as the integration can occur at different positions on the acceptor DNA molecule and lead to several bigger reaction products (*Figure 60 panel A*).

It is also possible to monitor specifically the ST without any interference of the 3'-P using a "precleaved" 19-mer 5' end labeled oligonucleotide, annealed to its 21-mer complementary strand. In this case the final GT dinucleotide has already been removed and the enzyme can perform directly the insertion of the substrate in an analogous one without the necessity to do the cleavage of the top strand 3' end (*Figure 60 panel B*).

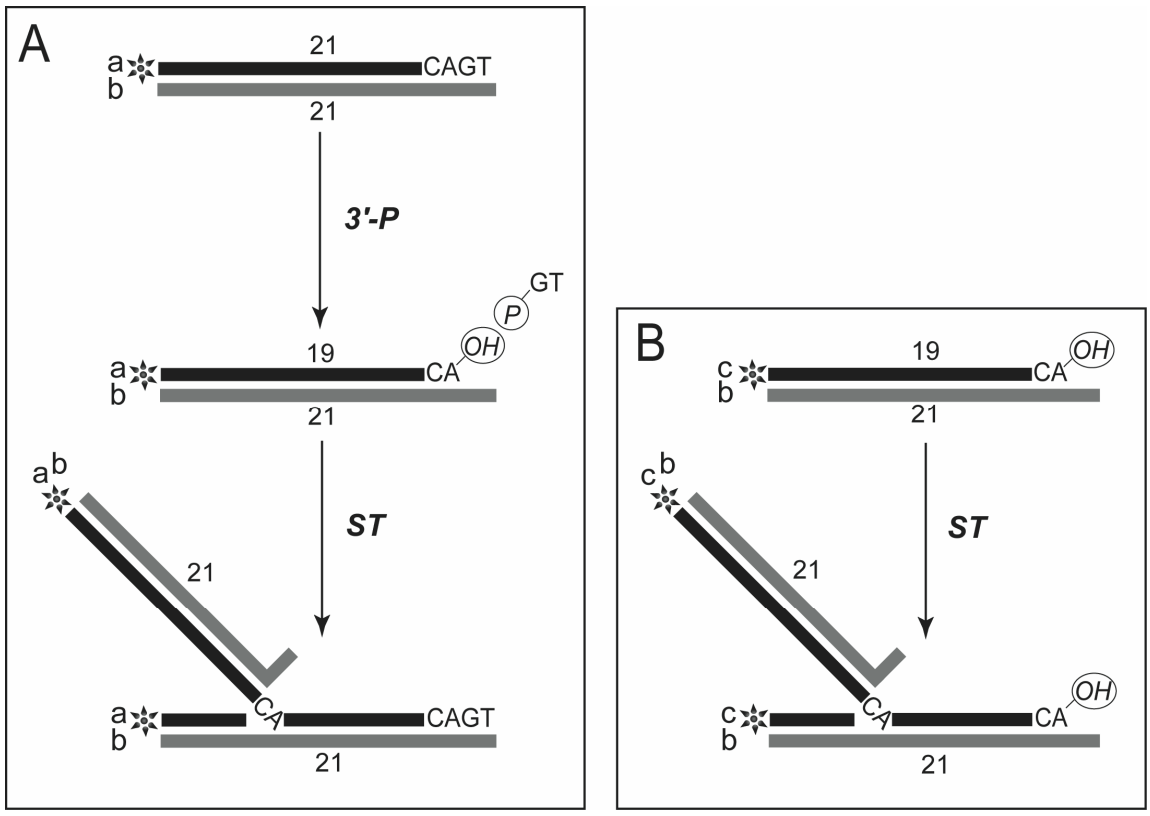

Figure 60: *schematic representation of the two integrase reactions and assays used to test the enzymes in vitro.* ***(A)*** *Schematic representation of the 3'-processing (3'-P) and strand transfer (ST) reactions using full-length*

DNA substrate. ***(B)*** *Schematic representation of the ST assay using precleaved substrate. Letters refer to Table 2 and numbers refer to the total number of bases. Asterisks indicate the 5' end labeling with* ^{32}P.

The reaction was performed as described in the *Experimental Procedures* section, testing different concentrations of enzyme.

In all integrase structures reported to date and solved in the presence of divalent metal ions, only one physiologically relevant metal ion can be located in the core domain. In particular, the divalent metal Mg^{2+} was shown to be coordinated with water molecules and oxygens of the residues D64 and D116 [85, 112]. Because a second metal has been observed in an ASV integrase crystal structure and because of the two-metal structure for polynucleotide transferases, it has been proposed that a second metal (Mg^{2+} or Mn^{2+}) can be coordinated between D116 and E152 once HIV-1 integrase binds its DNA substrate. Although Mg^{2+} has been used to crystallize the protein and it is the most common divalent metal in the biological systems, it has never been shown that it is the ion utilized *in vivo* by the integrase, also if it is likely. In literature it is possible to find studies performed using either one or the other, but also both together. Therefore it was decided to perform the analysis of the enzyme activity in presence of both magnesium and manganese. Typical gel images are shown in *Figure 61*. HIV-1 WT integrase shows to be actively performing 3'-processing and strand transfer in presence of both metals and in all tested enzyme's concentrations. In the past literature, it has been shown experimentally that when divalent manganese replaces magnesium in the active site of a magnesium-utilizing enzyme, the catalytic activity of that enzyme often is maintained. Unlike this, magnesium is less often a catalytically competent replacement for divalent manganese. Several lines of evidence suggest that DNA binding and sequence specificity are differentially affected by the type of metal ion used [58]. In addition, in the presence of manganese, integrase is capable of using alternate nucleophiles to carry out the 3'-processing reaction, including alcohols such as glycerol or methanol. In magnesium, water is the predominant nucleophile used by the protein. Specific photocrosslinking is also dramatically reduced in the presence of manganese as compared to magnesium. These facts suggest that the active site of the enzyme may be considerably less "tight" in the presence of manganese, allowing a wider selection of nucleophiles, and potentially reducing the DNA-binding specificity of the 3'-processing reaction. Data also report a different ability of diketo acids compounds to interfere with the enzyme activity depending on the metal present inside the catalytic site of the enzyme [113]. These arguments suggest that many of the effects of HIV-1 integrase under manganese conditions may not mimic the effects that are important in the presence of magnesium, likely to be the essential

cation *in vivo*. This is why in the second part of the work, it was decided to study the effect of the drugs only in presence of magnesium.

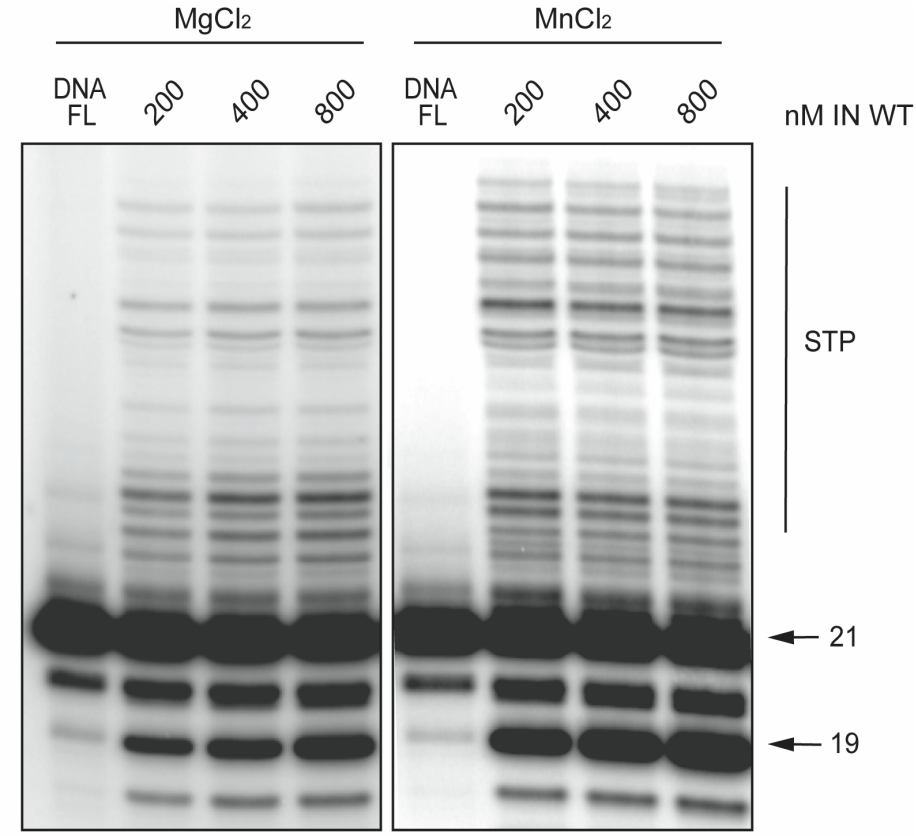

Figure 61: *activity of HIV-1 integrase. Gel image of a typical experiment using full length DNA substrate (FL). Bands labeled "21", "19" and "STP" correspond to DNA substrate, 3'-processing and strand transfer products, respectively. Three different concentrations of enzyme (200, 400 and 800 nM) were tested in presence of both magnesium and manganese.*

HIV-1 Integrase Mutants. As evidenced in numerous ongoing studies, some mutations in the IN gene are associated with therapeutic failure in patients receiving raltegravir [Merck Protocol 005]. Also *in vitro*, replication of HIV in the presence of increasing concentrations of STIs selects for amino acid changes within the integrase coding region, consequently leading to resistant viruses. It was decided to perform a deep study of the information available in literature regarding the *in vivo* and *in vitro* onset of drug resistance. A table that summarizes all the collected data for the two most advanced compounds is reported below (*Table 4*).

RALTEGRAVIR

D 10	N 17	R 20	H 51	G 59	T 66	L 68	V 72	L 74	E 92	Q 95	T 97	H 114	F 121	T 124	A 128	E 138	G 140	Y 143	P 145	Q 146	S 147	Q 148	V 151	S 153	M 154	N 155	K 156	E 157	K 160	G 163	V 165	V 201	I 204	T 206	S 230	D 232	R 263	Ref
								M	Q		A					K A	S A	H				H R K	I			H				R K						N		[109] ¶
									K								S					H				H		Q										[114]
								M	Q									C R				H K R				H												[115] §
							I	M			A					A	S A	C H R				H R	I			H S	N		N	R	I	I		S	R N			[116] *
	S			E					Q						A							K R			I	H				R			T					[117]

¶ *Hazuda D.J. et al.* associate raltegravir resistance with two genetic pathways. These can be essentially summarize as:

- Q148H/K/R plus E138K, G140S/A;
- N155H plus L74M, E92Q, G163R.
-

§ *Cooper D.A. et al.* report the genetic profile of patients with virological failure in BENCHMRK studies. Other mutations were identified in addition to the above-mentioned, but these have not yet been tested in phenotypic assays: S24N, M50I, G70D, G163R, I203M, L234I.

* *Mascolini M.* summarizes the most important information presented during the XVII International HIV Drug Resistance Workshop (Spain, June 2008). In particular the report considers data from:

- *Hatano H, Lampiris H, Huang W, et al.* Virological and immunological outcomes in a cohort of patients failing integrase inhibitors. XVII International HIV Drug Resistance Workshop. June 10-14, 2008, Sitges, Spain. Abstract 10.;
- *Katlama C, Caby F, Andrade RM, et al.* Virological evolution in HIV-treatment-experienced patients with raltegravir-based salvage regimens. XVII International HIV Drug Resistance Workshop. June 10-14, 2008, Sitges, Spain. Abstract 11.
- Miller MD, Danovich RM, Ke Y, et al. Longitudinal analysis of resistance to the HIV-1 integrase inhibitor raltegravir: results from P005, a phase II study in treatment- experienced patients. XVII International HIV Drug Resistance Workshop. June 10-14, 2008, Sitges, Spain. Abstract 6.
- *Da Silva D, Pellegrin I, Anies G, et al.* Mutations patterns in the HIV-1 integrase related to virological failure on raltegravir-containing regimens. XVII International HIV Drug Resistance Workshop. June 10-14, 2008, Sitges, Spain. Abstract 12.
- *Ceccherini-Silberstein F, Armenia D, D'Arrigo R, et al.* Virological response and resistance in multi-experienced patients treated with raltegravir. XVII International HIV Drug Resistance Workshop. June 10-14, 2008, Sitges, Spain. Abstract 18.

ELVITEGRAVIR

D 10	N 17	R 20	H 51	G 59	T 66	L 68	V 72	L 74	E 92	Q 95	T 97	H 114	F 121	T 124	A 128	E 138	G 140	Y 143	P 145	Q 146	S 147	Q 148	V 151	S 153	M 154	N 155	K 156	E 157	K 160	G 163	V 165	V 201	I 204	T 206	S 230	D 232	R 263	Ref
					I				Q				Y											Y														K	[118]
					I A K				K							K					G	R K H				H												[118]
			Y		I				Q	K						K				P	G							Q									[119] ╬
					I A K		A		Q V					A	T				S	S L		K R	I														[120]
E		K			I A			M	Q			Y			T	K						R												R			[121] ╩
						V I			Q																												[122]
					I A K		A		Q V					A	T				S	L		K R	I														[117]

╬ *Shimura et al.* associate elvitegravir resistance with two genetic pathways. These can be essentially summarize as:
- T66I plus Q95K, E138K, Q146P, S147G;
- E92Q plus H51Y, S147G, E157Q.

╩ *Goethals et al*. divide the above-mentioned mutations in two groups:
- Primary mutations, that confer high changes in elvitegravir susceptibility (T66I/A, E92Q and Q148R);
- Secondary mutations observed accompanying one or more primary mutations but that resulted in no change in elvitegravir susceptibility (D10E, R20K, L74M, H114Y, A128T, E138K and S230R).

Table 4: *collection of literature data about onset of mutations on integrase sequence, secondary to raltegravir (previous page) and elvitegravir (this page) treatment.*

Mutations in *Table 4* have been identified with different procedures: genetic sequencing of viruses from patients in whom the drug is failing; correlation studies between genotype at baseline and virologic response in patients exposed to the drug; *in vitro* passage experiments or validation of contribution to resistance by using site-directed mutagenesis; susceptibility testing of laboratory or clinical isolates.

Noteworthy until last year, there were essentially two main works published on the subject. *Malet et al.* [114] reported a complete analysis of raltegravir induced mutations. Starting from the study of the IN gene sequence of nine patients who received raltegravir and suffered virological failure while on regimen, they detected some mutations and they looked at the *in vitro* effect of the drug on these mutated enzymes compare to WT. They found at least four genetic profiles (E92Q, G140S+Q148H, N155H and E157Q) associated with *in vivo* treatment failure and resistance to raltegravir (*Figure 62*). The same mutations showed *in vitro* a strong impairment of the activity in the absence of raltegravir and a strong drug resistance. *Shimura et al.* [119] performed instead an *in vitro* study selecting elvitegravir-resistant viral variants using a dose escalation method. Two different experiments were conducted. The first was initiated with 0.5 nM EVG and led to the appearance of five amino acid substitutions (Q146P, T66I, S147G, Q95K and E138K). The second was initiated with 0.1 nM EVG and led to four amino acid mutations (E92Q, H51Y, S147G and E157Q) (*Figure 63*). They generated infectious HIV-1 clones containing these IN substitutions and they determined the susceptibility to the drug of the resulting selected variants. Thus they classified mutations into different groups: T66I and E92Q that confer a significant reduced susceptibility to EVG (37-and 36-fold reduced); Q146P and S147G that confer more moderate reduction (11-fold); H51Y, Q95K and E157Q that confer a small reduction in EVG susceptibility (less than 6-fold). T66I and E92Q were therefore considered primary mutations providing the highest change in susceptibility and the other mutations, defined as secondary mutations, enhanced the level of resistance to EVG when combined with the primary mutations.

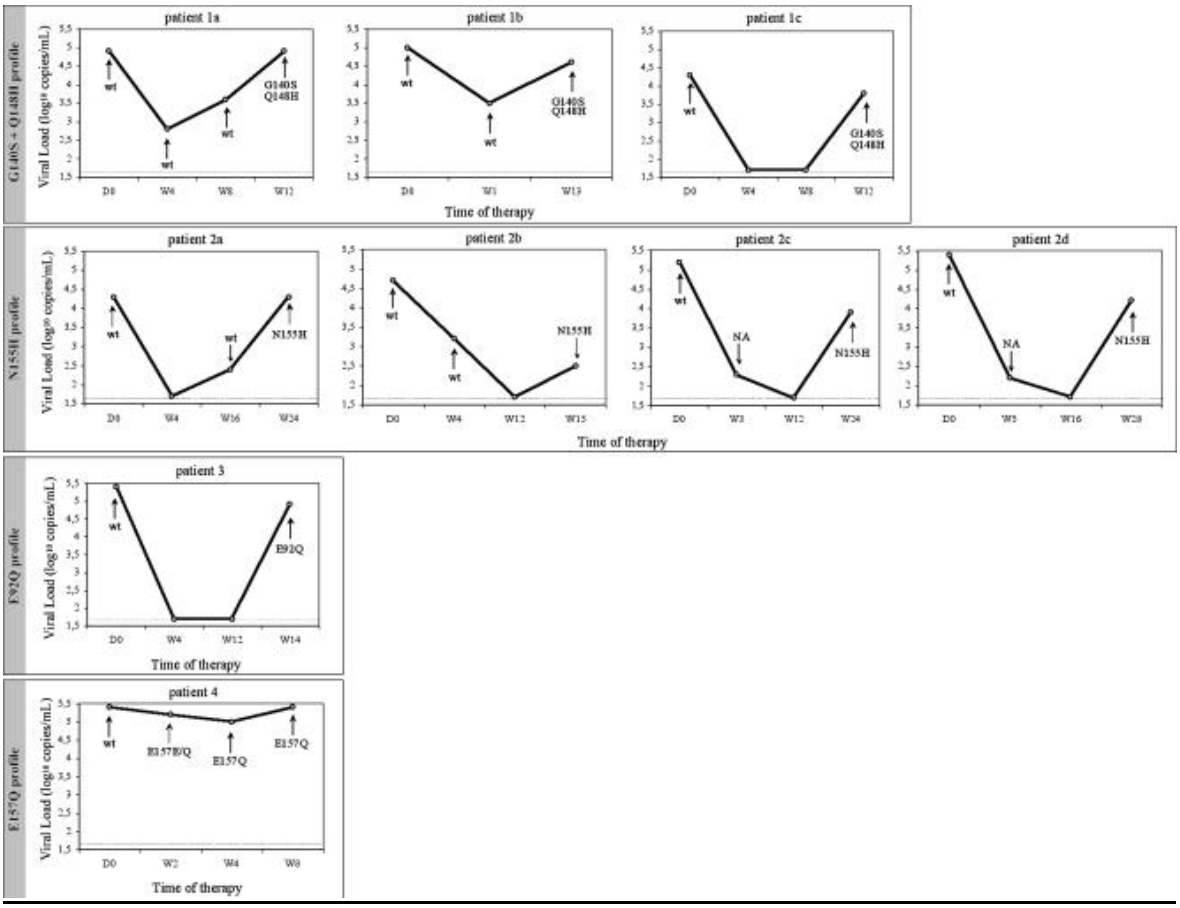

Figure 62: *kinetics of HIV RNA copy numbers in plasma for nine patients failing raltegravir therapy. Dotted lines at 1.60 log_{10} copies/ml (40 copies/ml) show the detection limit of the viral load assay. wt, wild type; NA, nonamplifiable. From Malet et al. [114].*

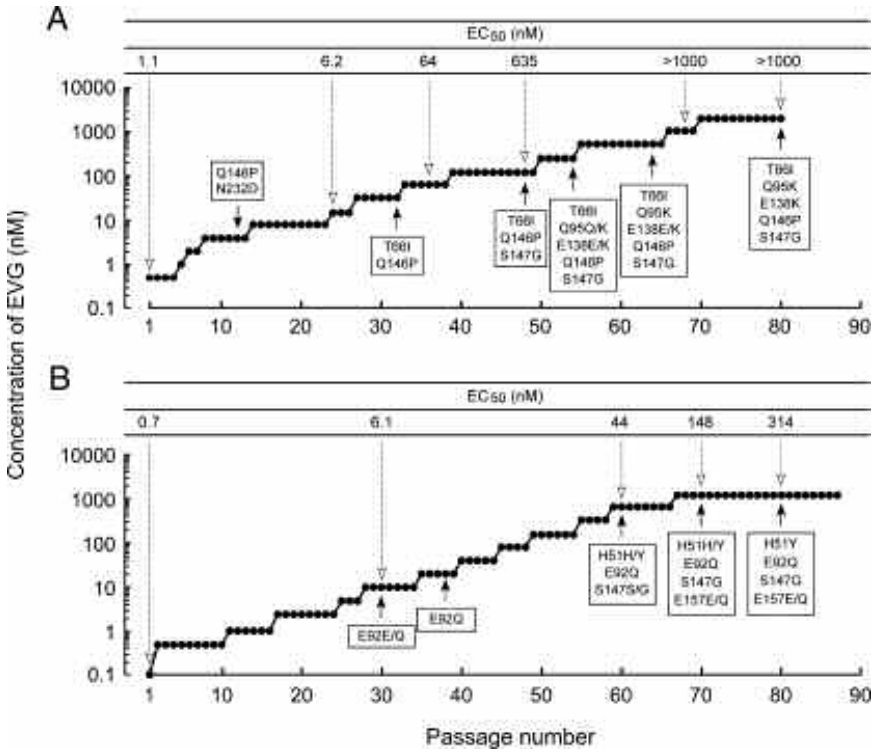

Figure 63: *induction of EVG-resistant HIV-1. Data from MT-2 cells are shown. The initial concentration of EVG were 0.5 nM* ***(A)*** *and 0.1 nM* ***(B)****. Results are from two identical but independent experiments. At the*

indicated passage number (black arrowheads), proviral DNA extracted from infected MT-2 cells was sequenced. Amino acid substitutions are shown. The EC_{50} values of HIV-1 variants selected by EVG at the indicated passage number (white arrowheads) were determined using MAGI assay ***(A)*** *or the production of p24 in MT-2 cells* ***(B)****. From Shimura et al. [119].*

Some positions along the core domain which correlate with drug resistance were selected and the site-directed mutagenesis technique was used to produce a pool of enzymes to be tested. The selected mutations are: T66I, L74M, E92Q, F121Y, Q148K, S153Y and N155H.

Figure 64 panel A shows the schematic distribution of the chosen mutations along the catalytic core domain (CCD) of integrase. Examination of the spatial arrangement of these mutations in the crystal structure of the CCD of IN (*Figure 64 panel B*) shows their distribution around and their relative proximity to the triad of catalytic acidic residues of IN (DDE motif shown in red with divalent metal in orange).

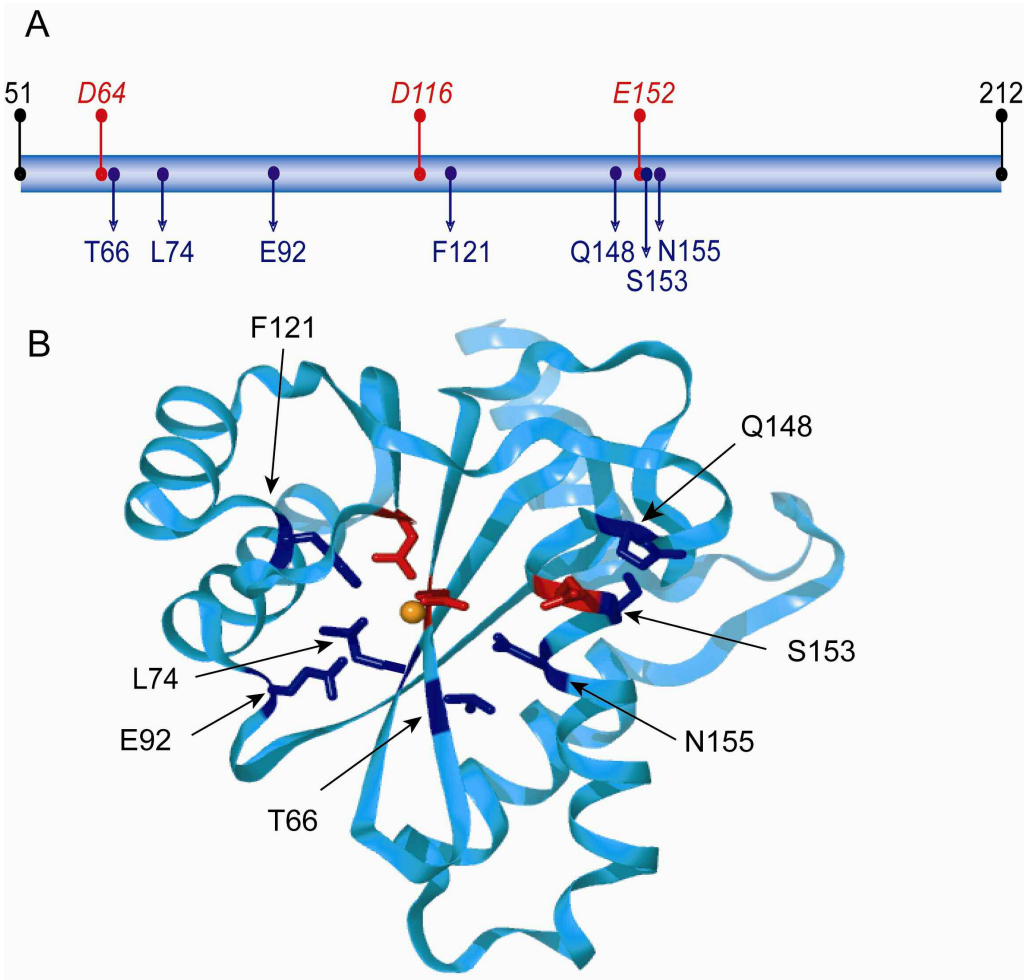

Figure 64: *HIV-1 integrase core domain residues associated with drug-resistance.* ***(A)*** *Linear representation of the HIV-1 integrase catalytic core domain (CCD). Catalytic acidic residues (DDE motif) are shown in red. The seven residues mutated in drug-resistant viruses and selected for comparison with wild-type HIV-1 integrase are shown in dark blue.* ***(B)*** *Three-dimensional representation of the HIV-1 integrase CCD. The catalytic residues*

(red) and the seven selected mutations (dark blue) are shown as stick representation in the crystal structure of the CCD (PDB code 1BI4) [112].

After expression and purification of these mutant enzymes as reported in the *Experimental Procedures* section, the concentrations of the final stock solutions were determined by absorbance at 280 nm. To confirm purity and concentration, SDS-PAGE was used. *Figure 65* shows a protein gel in which the same amount, based on the spectrophotometric data, of each enzyme was loaded. In all the stock solutions, it was possible to detect the presence of small quantities of impurity. The concentration of protein of interest was similar for all the considered batches.

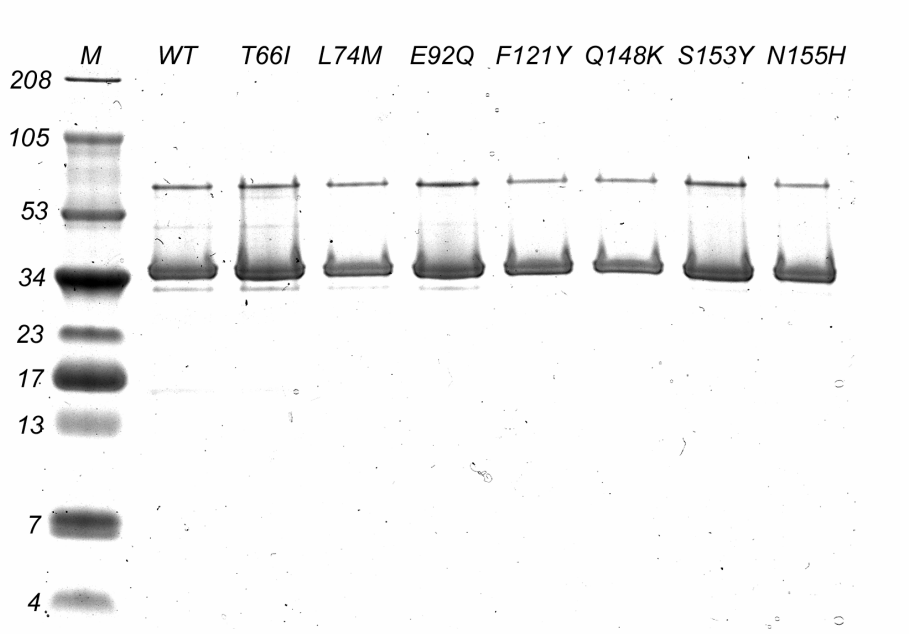

Figure 65: *purity and concentration of all the expressed enzymes. 16% Tricine Gel run at 100 Volts for 1 hour and a half. M, protein ladder (molecular weight in KDa reported on the right). The name of the loaded enzyme for each lane is reported on the top. 50 picomoles are loaded in each well.*

With the obtained enzymes it was possible to start studying the effect of considered mutations on the catalytic activity of integrase. Each mutant was evaluated using the above-described *in vitro* assay. The reaction was also always performed in presence of WT IN, reporting the total activity of this enzyme as 100%. A typical gel showing this type of analysis is shown in *Figure 66*, in which the activity of F121Y IN is compared to WT IN using the dual ST and 3'-P assay (full length substrate). The same assay was performed for all the other mutant enzymes and also in presence of precleaved substrate to avoid interference from 3'-P defect.

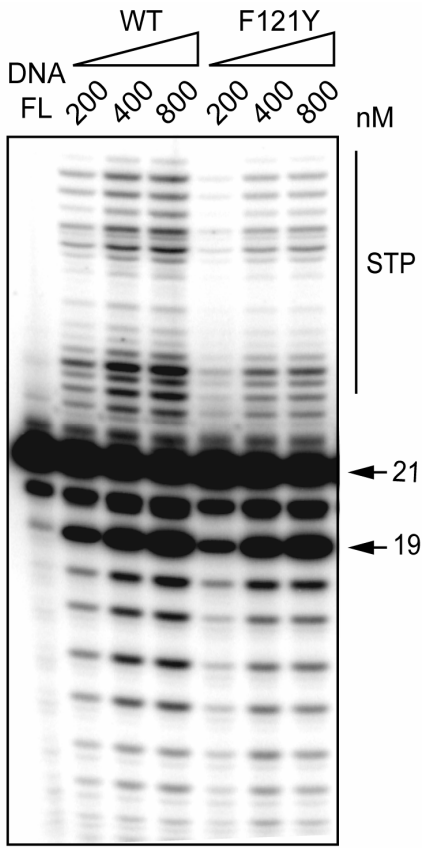

Figure 66: *catalytic activities of HIV-1 integrase mutants compared to wild-type using full length DNA substrate (DNA FL). Gel image of a typical reaction and showing selective defect for ST vs 3'-P for the F121Y mutant. Bands labeled "21", "19" and "STP" correspond to DNA substrate, 3'-processing and strand transfer products, respectively. Three different concentrations of enzymes (200, 400 and 800 nM) were tested.*

After quantitative analysis and normalization to WT IN, it was possible to determine the catalytic profile of each enzyme. Histograms on *Figure 67* show and compare the ability to carry out the 3'-processing and the strand transfer steps for all of them. The first step catalyzed by IN (3'-P) was only marginally impaired in most mutants (*Figure 67 left panel*). In particular, only T66I and S153Y IN lost partially their 3'-P activity. Q148K IN, on the other hand, was almost completely unable to catalyze 3'-P, showing an activity lower than 10% of WT.

Conversely, the second integration step (ST) was defective for most of the mutant enzymes (*Figure 67 central panel*, ST-FL). The ranking order for the most defective mutants for ST in the dual 3'-P and ST assay was Q148K, S153Y, T66I, F121Y, N155H, E92Q. The L74M mutant was the only one whose ST and 3'-P activities were close to WT IN. Noteworthy is the fact that mutations leading to ST deficiency appeared to affect the efficiency of ST but had no detectable impact on ST integration patterns. Panel right in *Figure 67* shows further analyses of ST deficiencies for the mutants performed with the precleaved substrate (ST-PC). The obtained profile was similar to that obtained with full length substrate.

These experiments demonstrate that only one of the mutants, Q148K, is severely defective for both 3'-P and ST. Two mutants (T66I and S153Y) are partially defective for both 3'-P and ST. Three mutants are selectively defective for ST (F121Y, E92Q and N155H). Only one mutant (L74M) appeared normal for both 3'-P and ST.

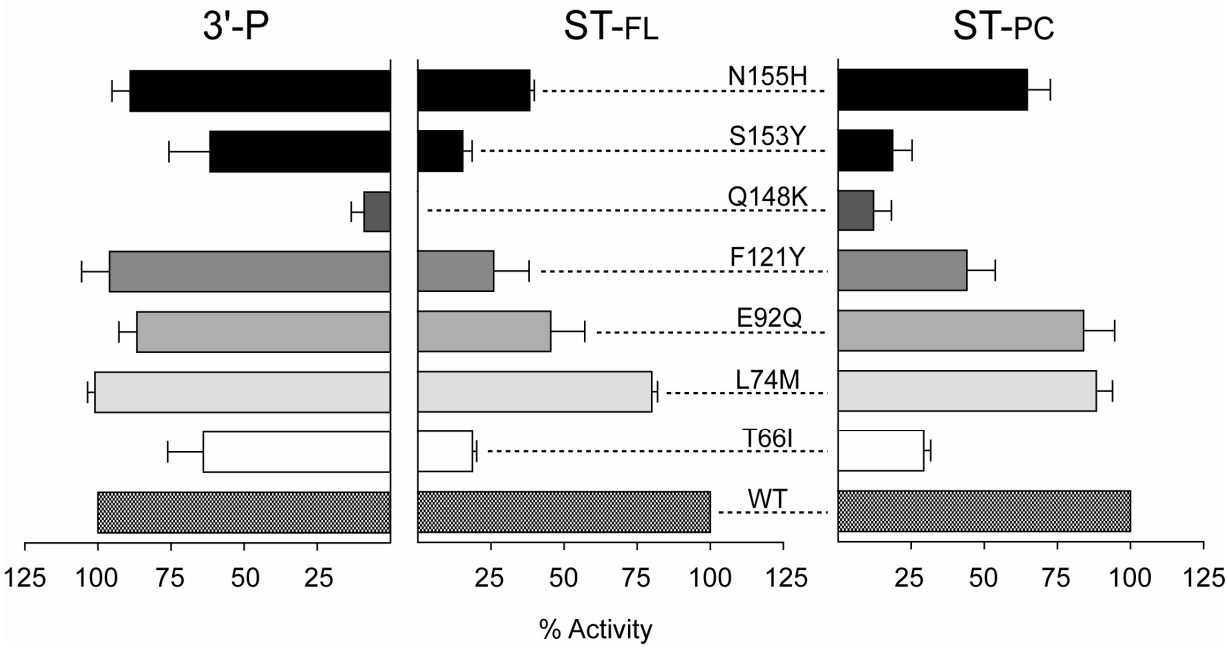

Figure 67: *defective catalytic activities of the drug-resistant HIV-1 integrase mutants compared to wild-type HIV-1 integrase. Comparative activities in assay performed with the full-length substrate (left and central panel: 3'-P and ST activities respectively). Strand transfer defect of the integrase mutants using the precleaved DNA substrate (right panel).*

When analyzing the defective activity of Q148K IN (*Figure 68 panel A*), it was possible to detect the prevalence of a 20-mer product corresponding to 3'-terminal mononucleotide cleavage. Such product was also produced by WT IN, although to a lesser extent than the canonical 19-mer 3'-processing product. Densitometry tracing for the Q148K mutant (*Figure 68 panel B*) showed this non-processive reaction product (which does not lead to ST) resulting from excision of the last base at the 3' end of the DNA. This 3'-mononucleotide cleavage reaction was also seen with WT IN (compare middle panels in *Figure 68B*, and gel images in *Figures 68A*) and the other mutants. Additional experiments with a topoisomerase I DNA substrate that does not bear sequence similarity with the HIV-U5-LTR showed lack of 3'-cleavage for the topoisomerase I substrate (data not shown) indicating that the 3'-mononucleotide cleavage reaction was dependent on the HIV-LTR sequence.

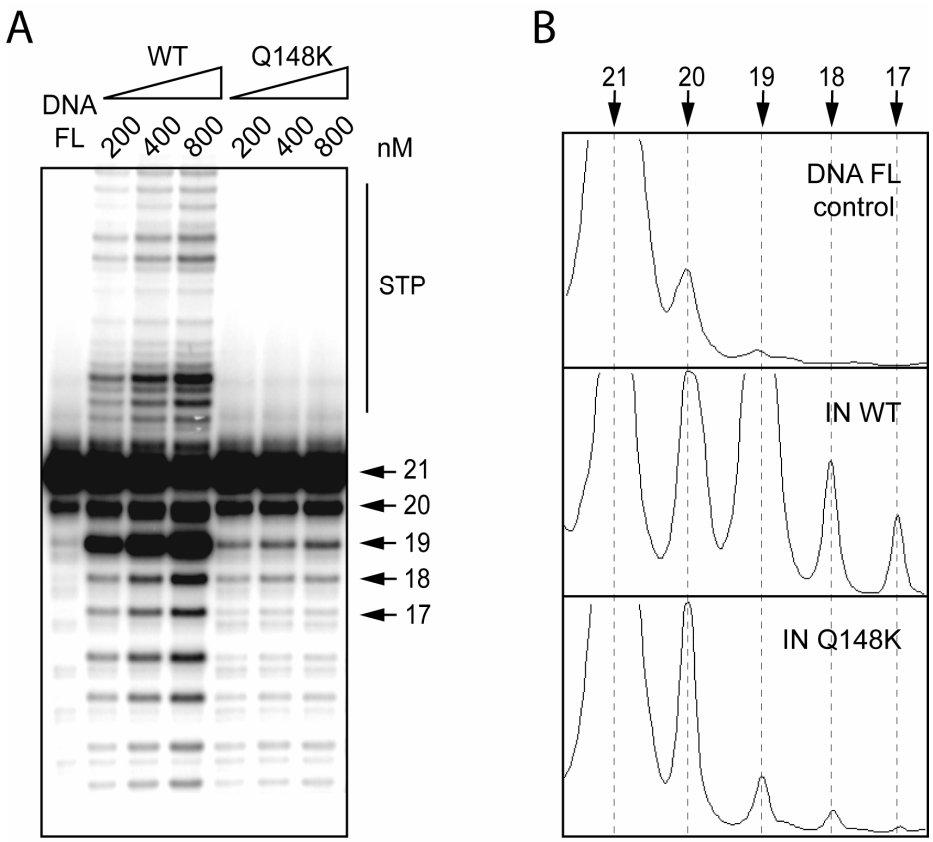

Figure 68: (A) *Gel image of typical 3'-P and ST reactions using three different concentrations of Q148K, and showing defective ST and 3'-P activities of the Q148K mutant HIV-1 integrase. Increasing enzyme concentrations are indicated above gel.* ***(B)*** *Defective 3'-P of the Q148K mutant shown as densitometry analysis of the gel presented in panel A (lanes 1, 4 and 7). DNA bands are indicated on top and on dotted lines and correspond to annotation in panel A.*

9.2 Comparison of Effects of Raltegravir and Elvitegravir on ST, 3'-P and Disintegration Reactions Mediated by Wild-Type Integrase

Since the discovery of integrase by *Grandgenett* [123, 124] a wide variety of inhibitors of the enzyme have been discovered. In November 2007 FDA approved raltegravir, the first drug with this mechanism of action. Currently another IN inhibitor, elvitegravir, is being investigated in clinical studies of HIV-1 infected patients. We decided to compare side by side *in vitro* these two novel compounds, highlighting advantages and disadvantages of one on the other.

We started by investigating their influence on WT IN activity with the gel-based assay. Representative gels in *Figure 69* demonstrate concentration-dependent inhibition of IN-mediated ST and 3'-P activities by both drugs. In particular, in *Figure 69 panel A*, it is possible to see how increasing concentrations of both drugs were able to inhibit the formation

of the 19-mer product, resulting from the processing step, and the formation of bands at higher molecular weight (STP), resulting from the transfer step. This assay was performed using a full length substrate (*Figure 69 panel A*) and a precleaved substrate (*Figure 69 panel B*) to see the direct effect of both drugs on ST without interference of 3'-P.

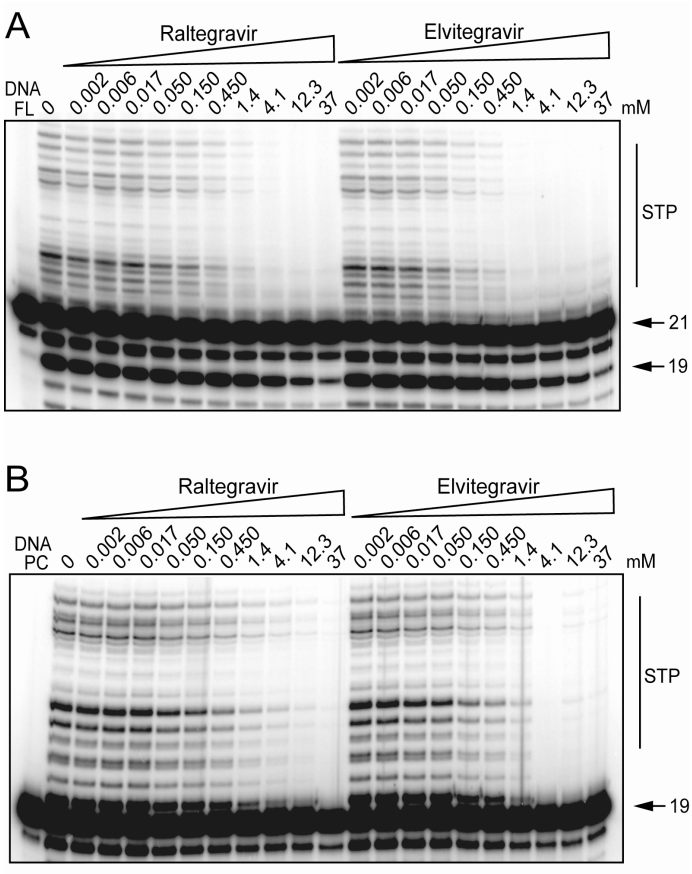

Figure 69: *comparative inhibition of wild-type HIV-1 integrase by raltegravir and elvitegravir.* ***(A)*** *Gel image of a typical dose-response experiment using full length DNA substrate (FL). Bands labeled "21", "19" and "STP" correspond to DNA substrate, 3'-processing and strand transfer products, respectively.* ***(B)*** *Gel image of a typical dose-response experiment using precleaved DNA substrate (PC).*

Densitometric analyses of independent experiments were carried out as described and shown in *Figure 69*, and permitted to obtain inhibition curves as in *Figure 70*. Important information were detectable directly from graph in *Figure 70* or from the IC_{50} values reported in *Table 5* determined using concentration ranges of 0.002-37 μM. First of all, it was clear the more intense interference of elvitegravir compared to raltegravir in inhibiting the catalytic activity on WT IN (compare open symbols to close symbols). In fact elvitegravir was able to inhibit ST activity at a concentration three-fold lower than raltegravir in both assays (28 nM vs 87 nM in full length substrate assay; 54 nM vs 190 nM in precleaved substrate assay). Elvitegravir was also slightly more potent as a 3'-P inhibitor compare to raltegravir (8 μM vs 12.8 μM).

These data demonstrated also that the inhibitory effect of the drugs on ST was not due to an inhibition of 3'-P. This conclusion was consistent with the fact that the inhibition curves for the two different steps did not overlap because concentrations required to inhibit 3'-P were at least two orders of magnitude higher than those required to inhibit ST. The more than 100-fold ratios between IC_{50} for ST and 3'-P indicated the high specificity of both compounds to interfere with ST under conditions where 3'-P remained fully effective (ratio 3'-P/ST: 288 for elvitegravir; 147 for raltegravir).

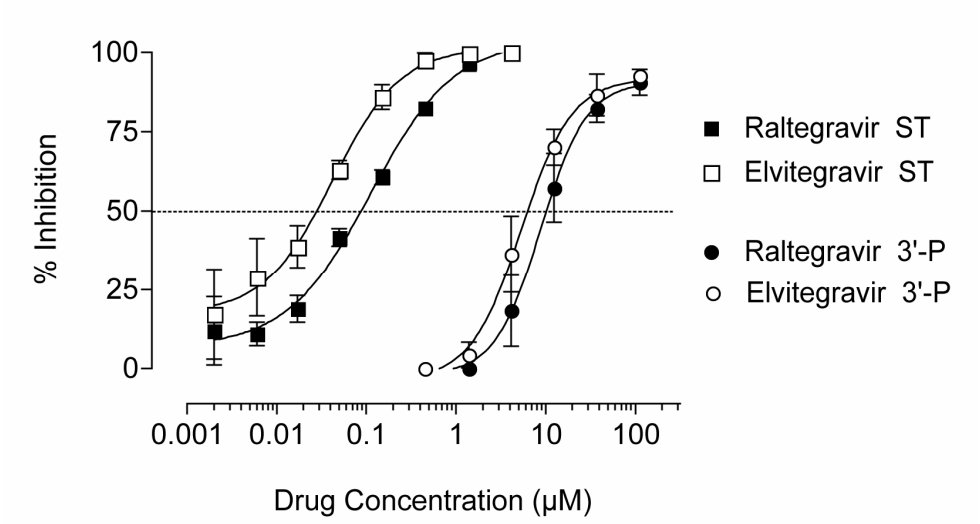

Figure 70: *quantitation from densitometric analyses of gels performed as in Figure 11 panel A. Error bars indicate SD for at least four independent experiments. IC_{50}s obtained from this graph are summarized in Table 5.*

Table 5 reports that the ST IC_{50} values derived using full length substrate were lower than those obtained with precleaved. One explanation could be that the FL substrate mimicked better the natural viral DNA therefore the enzyme could more easily recognize and bind to this substrate, leading to a greater inhibition by the drugs.

Table 5: *raltegravir and elvitegravir IC_{50} value (μM) against IN wild-type activities.*

	Raltegravir	**Elvitegravir**
3'-P	12.79 ± 6	8.06 ± 4.27
ST (FL)	0.087 ± 0.008	0.028 ± 0.006
ST (PC)	0.190 ± 0.056	0.054 ± 0.013
Disintegration	> 111	> 37

Abbreviations: FL: full-length substrate; PC: precleaved substrate (see *Figure 60A* and *60B*, respectively).

We next determined whether raltegravir and elvitegravir were able to inhibit the disintegration reaction. This third catalytic activity has been described only *in vitro* and there are no evidences of this step *in vivo*. It is an apparent reversal of the joining reaction in which a Y-shaped substrate representing the insertion of the viral DNA ends in a target DNA is resolved intramolecularly into the viral and target components [125]. Schematic representation of the assay is reported in *Figure 71 panel A.* In *Figure 71 panel B,* raltegravir shows its inability to inhibit this process at concentrations up to 111 μM. Similarly, elvitegravir is ineffective in disintegration reactions and only a minor inhibition is detectable above 37 μM.

In summary, elvitegravir and raltegravir are low nanomolar inhibitors of recombinant IN *in vitro*. They are selective for the ST step of the reaction and elvitegravir is more potent than raltegravir. Together, all the findings suggest that both compounds are highly specific for an intermediate step in the integrase catalytic cycle, which immediately follows 3'-P. Such selectivity is remarkable considering that all three reactions (3'-P, ST and disintegration) utilize the same catalytic site with the three catalytic acidic residues of IN. The specificity for the ST step indicates that both drugs can selectively recognize a specific configuration of the IN-DNA complex and bind at its interface, forming a quaternary complex with the catalytic metal(s) in the enzyme active site together with the 3'-processed donor (viral) DNA. Thus, raltegravir and elvitegravir are candidate interfacial inhibitors [126, 127].

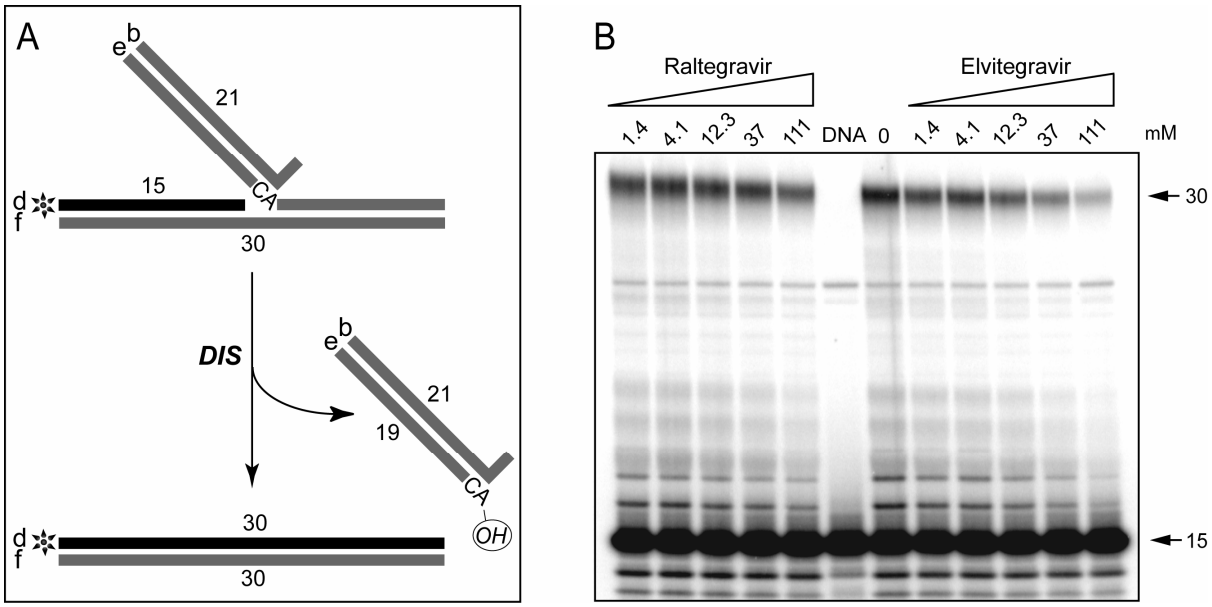

Figure 71: *disintegration reaction* ***(A)*** *Schematic representation of the disintegration reaction. Minuscule letters refer to Table 2 (Experimental Procedures section) and asterisks indicate the 5' end labeling with* ^{32}P. ***(B)*** *Weak or lack of inhibition of wild-type HIV-1 integrase-mediated disintegration by raltegravir and elvitegravir. Bands labeled "15" and "30" correspond to disintegration substrate and product, respectively.*

9.3 CROSS-RESISTANCE OF THE INTEGRASE MUTANTS TO RALTEGRAVIR AND ELVITEGRAVIR

The following step was the examination of the drug resistance of the expressed IN mutants. In fact pharmaceutical research is currently trying to overcome resistance to raltegravir that has already been evidenced in patients. Therefore, to consider elvitegravir as a better drug, it would have to at least partially overcome the resistance profile given by raltegravir.

We decided to compare the effect of the two compounds side by side on all the eight recombinant enzymes, using the precleaved substrate since both drugs are selective ST inhibitors. Gels of *Figure 72*, *panels from A to C*, show three representative experiments comparing the activity of elvitegravir on ST in all the enzymes. Only two of the mutants, L74M and F121Y, preserved the same sensibility to the compound as WT IN. The other mutants showed various degree of resistance and they can be divided in high resistant enzymes (Q148K, T66I and S153Y) and medium resistant enzymes (E92Q and N155H). The same analysis was performed with raltegravir. All the results are summarized in *panel D* and *E* (*Figure 72*) in which are reported independent duplicate ST IC_{50} values for both compounds. *Panel F* reports the fold increase in ST IC_{50} values normalized to WT integrase and shows comparable resistance profiles for the two drugs. Important information detectable from this graph is the greater resistance of all the mutants to elvitegravir compare to raltegravir. In fact T66I, Q148K, S153Y and N155H were at least 2-fold more sensitive to raltegravir. A detailed study of the resistance profiles suggest that mutant enzymes can be divided in four groups:

- Enzymes not resistant (L74M and F121Y).
- Enzymes equally resistant to both drugs (E92Q).
- Enzymes more resistant to elvitegravir than raltegravir (T66I, Q148K and N155H).
- Enzymes selectively more resistant to elvitegravir than raltegravir (S153Y).

Q148K IN displayed the most resistant profile with a 77-fold increase for elvitegravir and 27-fold for raltegravir. The next most resistant integrase was T66I with a 11- and 4.6-fold increase for elvitegravir and raltegravir, respectively.

The qualitative similarity in drug resistance profile across the seven mutants is consistent with the similarities of the two drugs with respect to ST selectivity vs 3'-P and disintegration. Thus, these results suggest that both drugs interact in a similar way with IN and that they bind

to a common region in the IN catalytic site. The region is probably outlined by the drug resistance mutations (see *Figure 64B*).

Nevertheless, based on these results with recombinant IN enzymes bearing resistance mutations, it seems unlikely that elvitegravir will be a rationale choice to overcome resistance to raltegravir.

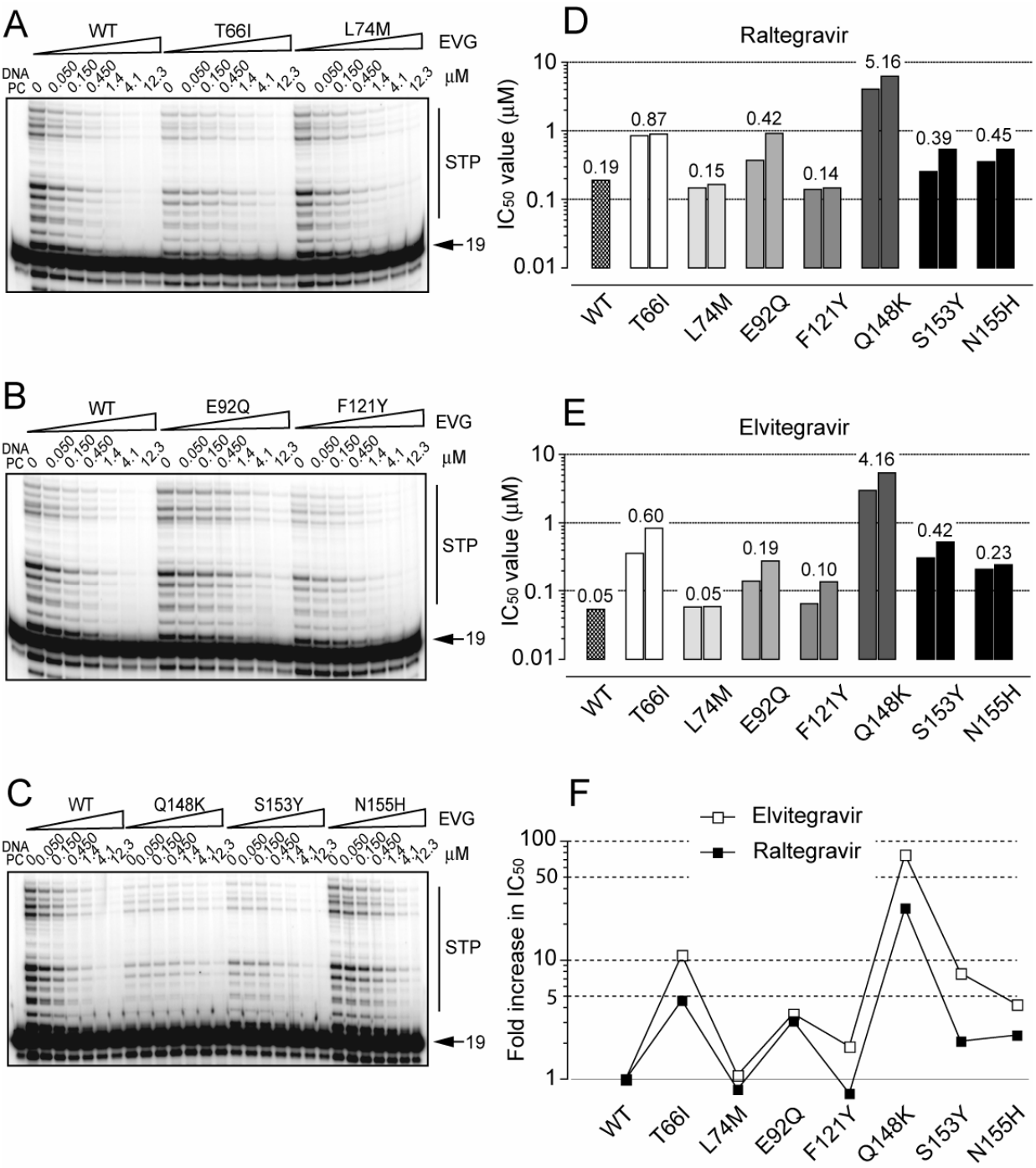

Figure 72: *comparative impact of the HIV-1 integrase mutations on integrase strand transfer inhibition by raltegravir and elvitegravir.* ***(A–C)*** *Representative gel images of elvitegravir (EVG) concentration-response experiments using precleaved DNA substrate (PC). Wild-type (WT) HIV-1 integrase and the mutants are indicated above gel pictures.* ***(D–E)*** *Comparison of drug IC_{50} value derived from gels performed as in panels A-C. Values for WT enzyme were derived from the mean of at least 10 experiments. Values for mutant HIV-1*

integrase are reported directly on the graph for two independent experiments. IC_{50} values (µM; mean of at least 2 independent experiments) are indicated above bars. ***(F)*** *Graphical representation of the patterns of cross-resistance for raltegravir and elvitegravir. Values were derived from experiments performed as described in panels A-E.*

9.4 EFFECT OF CONSIDERED MUTATIONS ON 3'-P INHIBITION BY RALTEGRAVIR AND ELVITEGRAVIR

To evaluate the drug effects on 3'-P in the seven IN mutants, we performed the same drug dose-response reactions using the full-length substrate (see *Figure 60A*). For all the WT and mutant enzymes (except E92Q), both drugs gave a similar weak inhibition of 3'-P. *Figure 73A* shows a representative experiment demonstrating increased inhibitory effect of elvitegravir on 3'-P mediated by the E92Q mutant. Similar experiments were performed with raltegravir, and results for both elvitegravir and raltegravir are summarized in *Figure 73B-C*. These results demonstrate that the E92Q mutation renders IN more resistant to both raltegravir and elvitegravir with regard to ST but sensitize the enzyme to both drugs with respect to 3'-P.

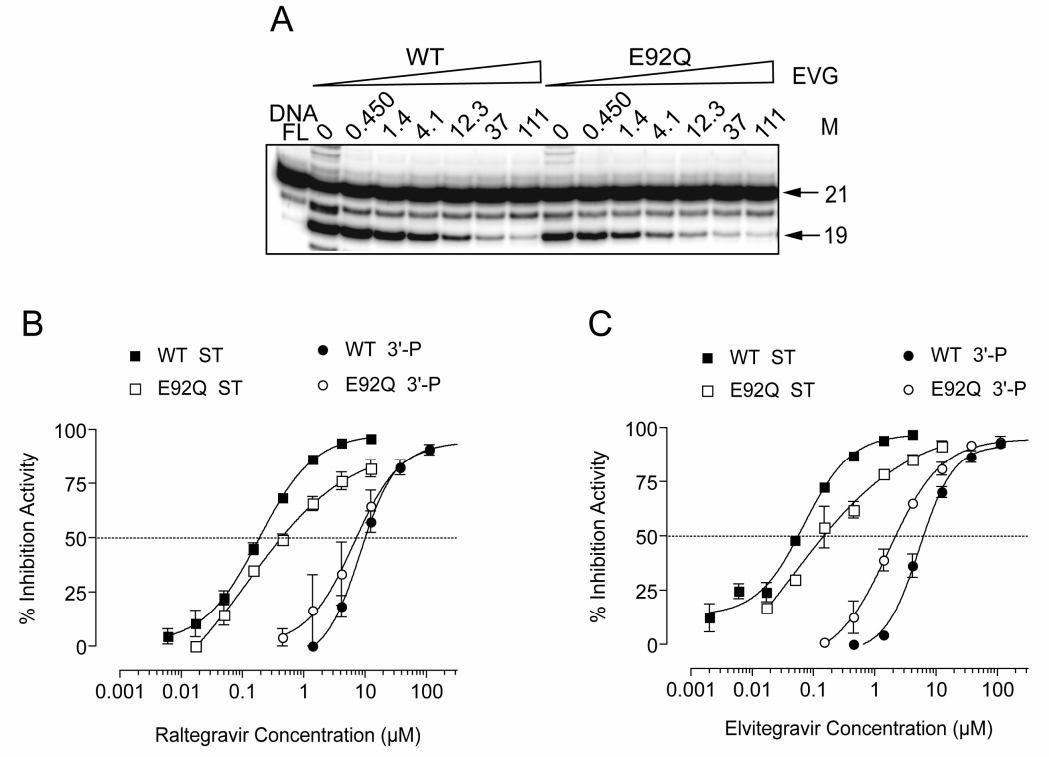

Figure 73: *different effect of the E92Q integrase mutation on inhibition of 3'-processing vs. strand transfer by raltegravir and elvitegravir.* ***(A)*** *Gel image of a representative dose-response experiment using full-length DNA substrate (FL) in the presence of elvitegravir.* ***(B-C)*** *Quantitation by densitometryic analyses of experiments performed as in panel A for raltegravir* ***(B)*** *and elvitegravir* ***(C)****.*

9.5 EFFECT OF THE ORDER OF ADDITION ON DRUG RESISTANCE

Because the previous results had been obtained by adding the drugs to IN followed by the immediate addition of the DNA substrate, we next wished to determine whether preincubation of the drugs with IN or addition of the drugs to the preassembled IN-DNA complexes would affect resistance. We designed two experimental protocols (*Figure 74 panel A*). In protocol 1, IN-DNA complexes were pre-assembled in the absence of drug for 15 minutes on ice to prevent catalysis. After this preincubation period, the drugs were added and reactions were initiated by placing the samples at 37°C. Reactions were carried out for 1 hour. In protocol 2, the drugs were mixed with IN for 15 minutes. Then the DNA was added and reactions were continued for an additional hour. *Figure 74 panel B* shows a representative experiments performed with the E92Q mutant. It demonstrates the absence of detectable difference for strand transfer IC_{50} values determined in the two different conditions for raltegravir. Similar results were obtained with the T66I mutant and WT IN, and with elvitegravir (data not shown). Thus, our results demonstrate that both drugs inhibit IN independently of the order of addition and that preincubating the drugs with IN does not affect IN resistance. These findings are also consistent with the binding of raltegravir and elvitegravir at the IN-DNA interface rather than to IN alone and confirm the fact that both drugs need a quaternary complex (enzyme, viral DNA, metal and drug itself) to be active [126, 127].

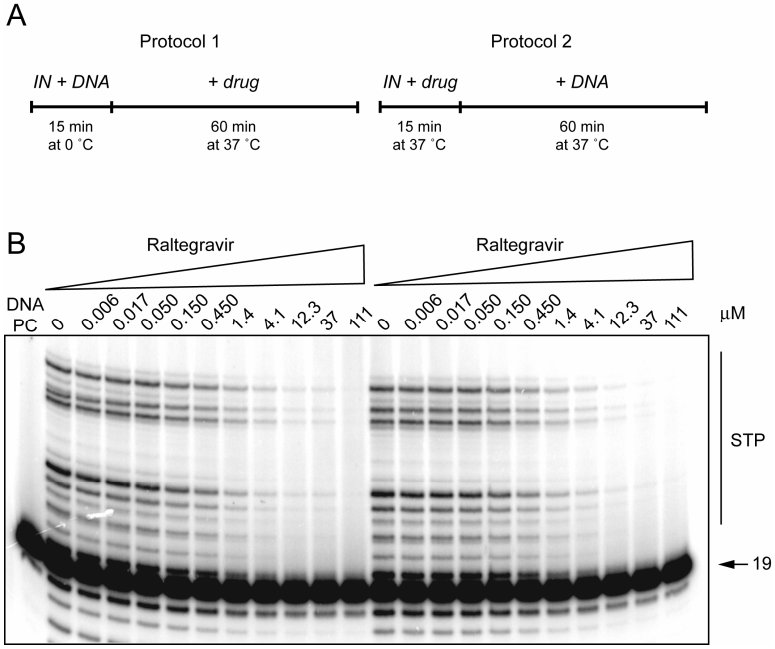

Figure 74: *comparison of the inhibition of HIV-1 integrase by raltegravir and Elvitegravir using different orders of addition.* ***(A)*** *Schematic representation of the two protocols used.* ***(B)*** *Gel image of the inhibition of HIV-1 integrase E92Q by raltegravir using both protocols.*

9.6 RELATION BETWEEN DRUG RESISTANCE AND CATALYTIC ACTIVITY FOR THE HIV-1 INTEGRASE MUTANTS

The fact that the seven mutations analyzed in the present study conferred a range of activity deficiencies (see *Figure 67*) and drug resistances (see *Figure 72*) enabled us to examine the relationship between these two parameters. The relationship is represented graphically in *Figure 75*. *Panel A* shows an overall linear correlation between loss of ST activity of the mutants and resistance to elvitegravir. The three most resistant enzymes (Q148K, T66I and S153Y) were also most defective in ST activity. A similar trend was observed for raltegravir (*Figure 75 panel B*). The correlation between resistance mutations and catalytic defects is consistent with the specific interaction of the drugs with the IN catalytic site.

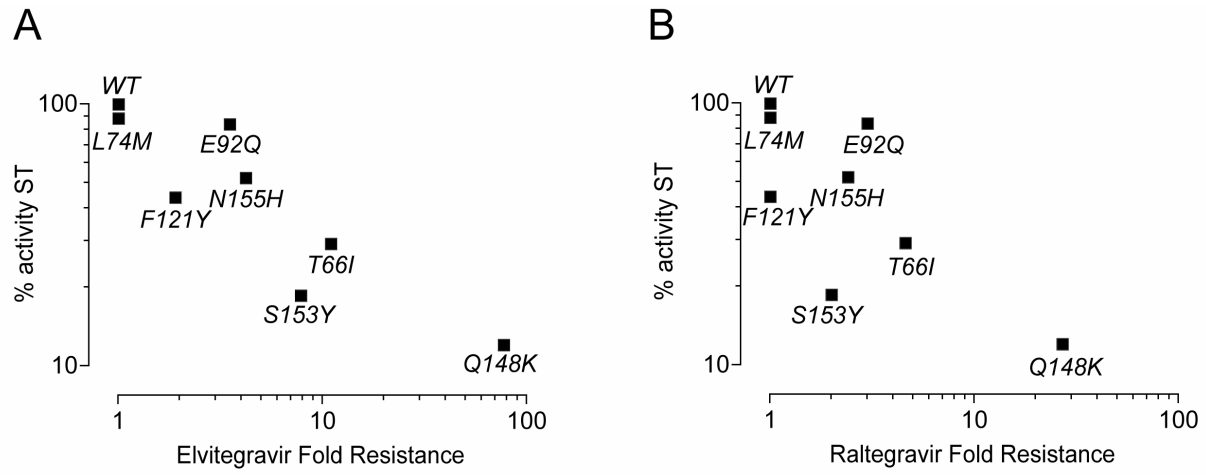

Figure 75: *correlation between drug resistance and defective catalytic activity for the HIV-1 integrase mutants.* ***(A)*** *Elvitegravir.* ***(B)*** *Raltegravir.*

Even if it may be difficult to correlate this *in vitro* observed effect with what could happen *in vivo*, it can be interesting to highlight the fact that during antiretroviral therapy, the selection of drug resistant mutants reflects a complex interaction between the effect of a given mutation on drug activity and the effect of this particular mutation on the viral replication potential ("viral fitness") [128]. There seems to exist two phases in the evolution of viral fitness. The first characterized by the selection of viruses with reduced drug susceptibility but which also display an impaired replication potential. The second with the generation and selection of additional compensatory mutations that restore replication [129, 130].

Therefore it could be interesting to understand if the insertion of additional mutations in the expressed single mutants can overcome also *in vitro* the defect on ST activity.

Our results show that mutation L74M does not confer loss of activity and drug resistance. This mutation could therefore be secondary to compensate the effect of other substitutions on the viral replication potential. Hazuda et al. [109] reported the analysis of the Merck Protocol 005, a Phase II study in HIV-1 patients failing therapy with PI, NNRTI and NRTI resistance. After treatment with raltegravir, integrase of patients with virologic failure was analyzed by sequencing. Raltegravir failure was then associated with integrase mutations in two distinct genetic pathways defined by a signature mutation at either Q148 (H, R or K) or N155 (H) and one or more secondary mutations unique to each pathway. The N155H pathway was correlated to the appearance of L74M. Therefore we expressed and tested the recombinant N155H+L74M double mutant to understand if the insertion of the secondary mutation can compensate the *in vitro* activity defect of the N155H single mutant. The resulting double mutant enzyme did not recover the activity defect and instead showed a decreased ST capacity.

9.7 Effect of Raltegravir and Elvitegravir on HIV-1 Cytopathic Effect (CPE)

To complete the study it was decided to evaluate the drug susceptibility in HIV-1 infected MT-2 cells, a line of leukemic T cells. Cells were infected with 100 $TCID_{50}$ of HIV-1 and seeded into 96-well plates in the presence of various concentrations of the test compounds (see *Experimental Procedures* section). In parallel uninfected cells were plated in the presence of the same compounds' concentrations, testing the drug cytotoxicity. *Figure 76* shows the relation between cell viability and drugs' concentration in infected and uninfected MT-2 cells. Both drugs were able to protect cells from HIV-1 cytopathic effect, giving a 80% protection at nanomolar concentrations (~10 nM). Elvitegravir exhibited some cytotoxicity at micromolar concentrations in the absence of virus (*Figure 76 panel B*). This cytotoxic effect was not present in the experimens performed with raltegravir (*panel A*). Therefore, raltegravir appears to be a more selective drug than elvitegravir, interacting with integrase probably with a more pronounced specificity. Elvitegravir probably interacts with secondary targets, resulting in an increased and undesired toxicity.

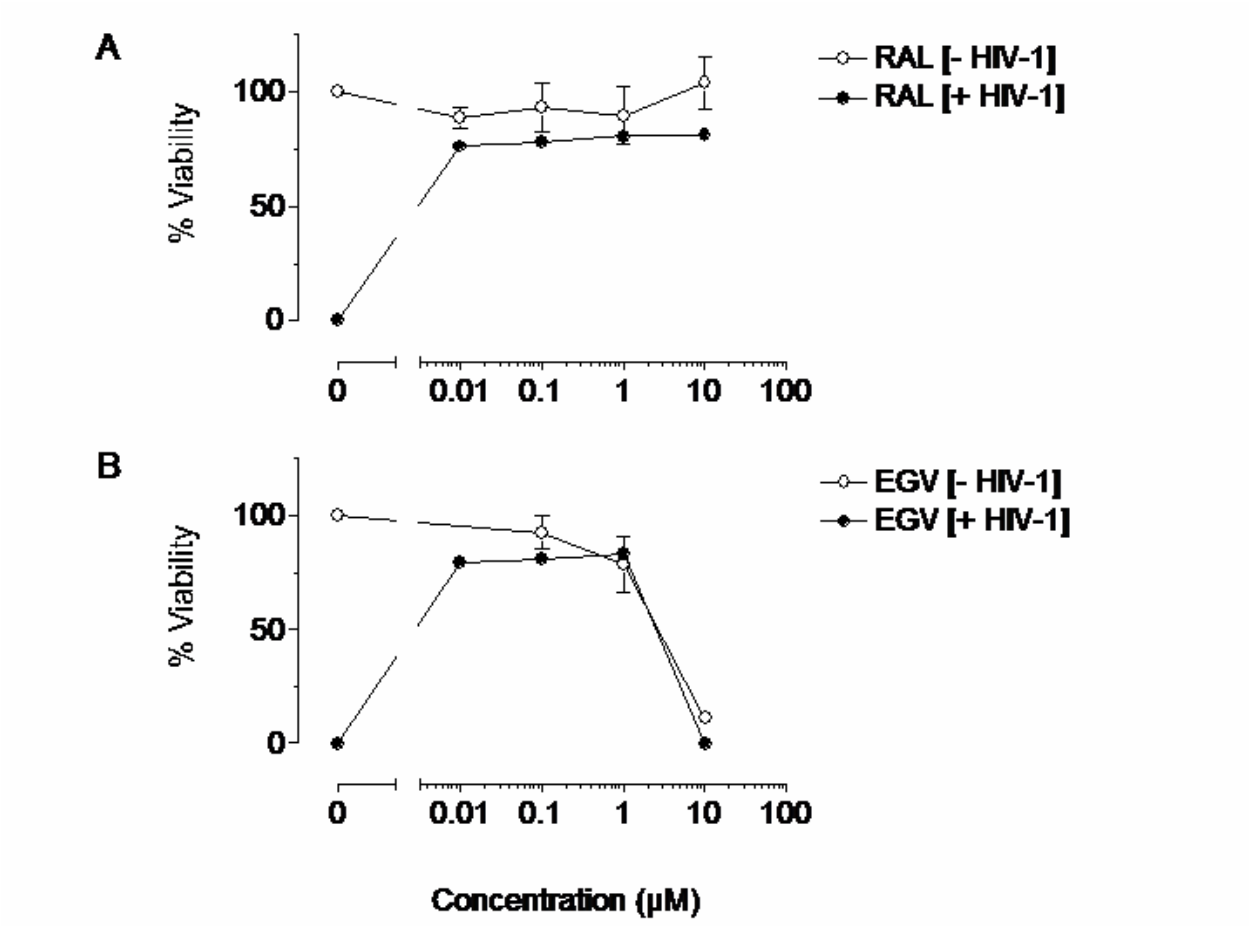

Figure 76: *relation between cell viability and drugs' concentration in infected and not infected MT-2 cells.* ***(A)*** *Experiments performed in presence of raltegravir.* ***(B)*** *Experiments performed in presence of elvitegravir. White track: not infected cells; black track: infected cells.*

10. WORK IN PROGRESS

10.1 STRATEGY TO UNCOUPLE DONOR AND ACCEPTOR IN THE IN VITRO REACTION

As described in the *Results* section, the viral integration step can be studied *in vitro* using an oligonucleotide mimicking the last 21 nucleotides of the HIV-1 U5 LTR. Recombinant integrase removes from this DNA duplex substrate the GT dinucleotide immediately 3' from a conserved CA sequence. Consequently the cleaved substrate is used as a donor for the strand transfer step and it is inserted in another identical duplex that serves as an acceptor (target) DNA. Therefore the DNA duplex substrate is used at the same time as a donor and as an acceptor. This procedure does not permit to study *in vitro* the integration in a different DNA target. It could be interesting in fact analyze if the enzyme has the ability to transfer the viral DNA into specific sequences of the host DNA. To achieve this goal it is necessary to uncouple the figure of the DNA donor from that of the DNA acceptor.

Phosphorothioate oligonucleotides are characterized by the substitution of a non-bridging oxygen at a phosphate group for sulfur atom (*Figure 77*). The sulfurization of the internucleotide bond gives a range of interesting characteristics. For example it dramatically reduces the action of endo- and exonuclease, it increases the hydrophobicity and consequently the potential for crossing the lipid bilayer, it confers chirality on the phosphorus. Their versatility has made them attractive to enzymologists and molecular biologists.

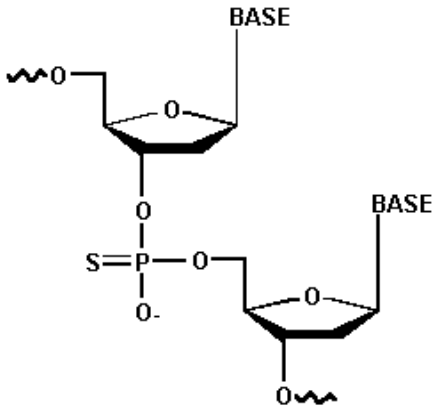

Figure 77: *phosphorothioate oligonucleotide. A non-bridging oxygen is substituted with a sulfur.*

Inspired by the fact that integrase has an endonuclease-similar effect on viral and host DNA and that phosphorothioate oligonucleotides are weakly cleaved by endo- and

exonuclease, we decided to determine if precleaved phosphorothioate substrates could selectively be used as donors. After binding of the enzyme to the precleaved phosphorothioate substrate, integrase should selectively transfer this DNA into a different normal phosphodiester DNA target (acceptor), therefore losing the autointegration effect (defined as integration of donor into donor itself).

Phosphorothioate substrates were designed for this purpose, avoiding the incorporation of thioate linkages at the first three sites at the 5' end because these oligonucleotides are not easily labeled by T4 polynucleotide kinase.

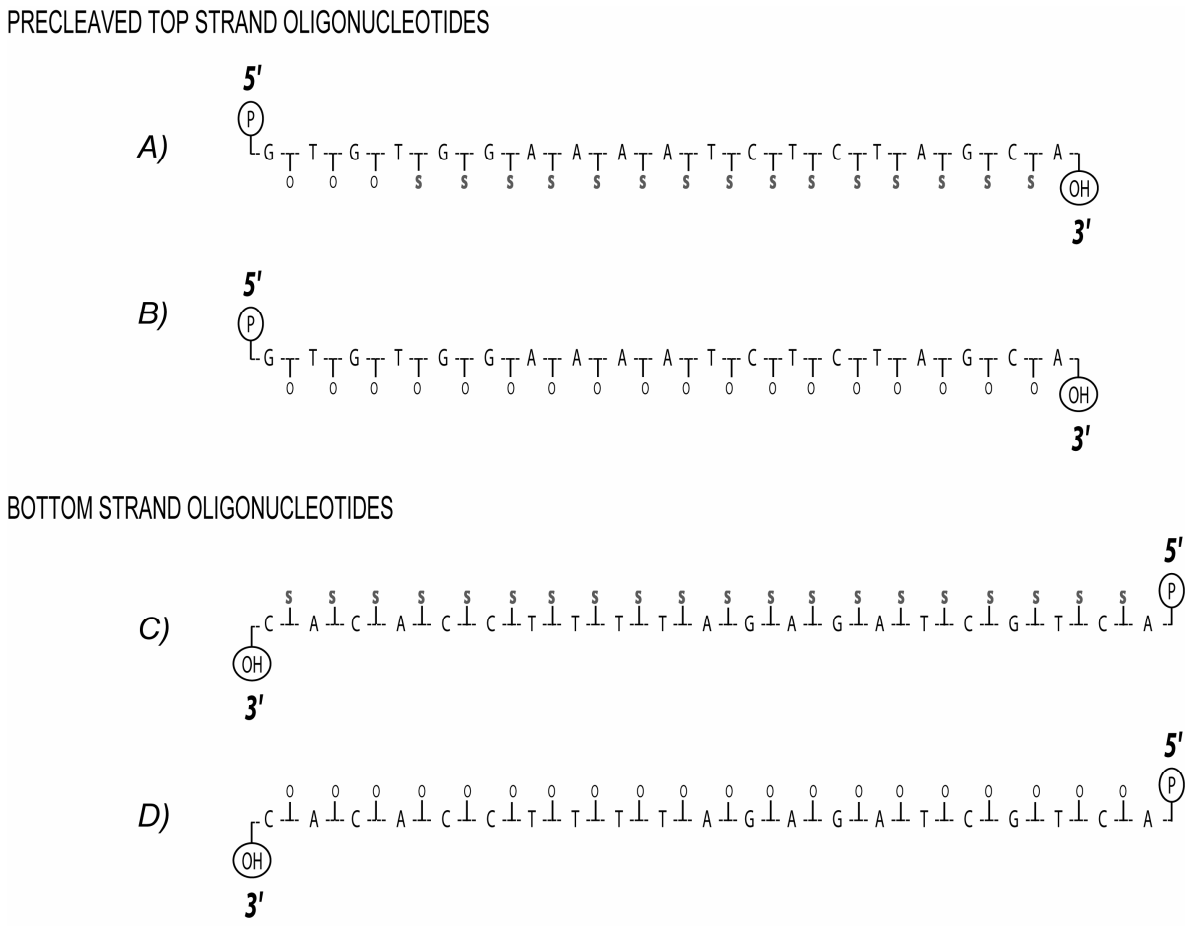

Figure 78: *phosphorothioate and phosphodiester oligonucleotides used to uncouple the donor substrate from the acceptor substrate.*

Initially we tested a DNA duplex derived from the annealing of oligonucleotide *A* with *C* or with *D* in *Figure 78.* As it is possible to see in the gel in *Figure 79 panel A*, WT integrase was unable to utilize these substrates as donors. Lanes 1 and 4 are representative of autointegration, while lanes 2 and 5 and lanes 3 and 6 represent integration in a phosphodiester acceptor and a partial phosphorothioate acceptor, respectively. Both in magnesium and in manganese it was not possible to detect strand transfer activity. The

recombinant enzyme is probably not able to correctly recognize and bind to this substrate. Published data shows that the chirality of the phosphorus in phosphorothioate oligonucleotides can perturbs the minor (S enantiomer) or the major (R enantiomer) groove dimension. In addition, the sulfur-containing phosphate backbone of substituted duplexes are observed to be less electronegative than the regular backbone, which consequently affects hydration pattern and residence time of water molecules in the first hydration layer of sulfur atoms. The role of electrostatics in the molecular interactions of proteins has been well established in a number of studies, and these have revealed that charge distribution plays important role in molecular recognition. Therefore, all of these reasons could potentially be responsible for the loss of activity in our assay.

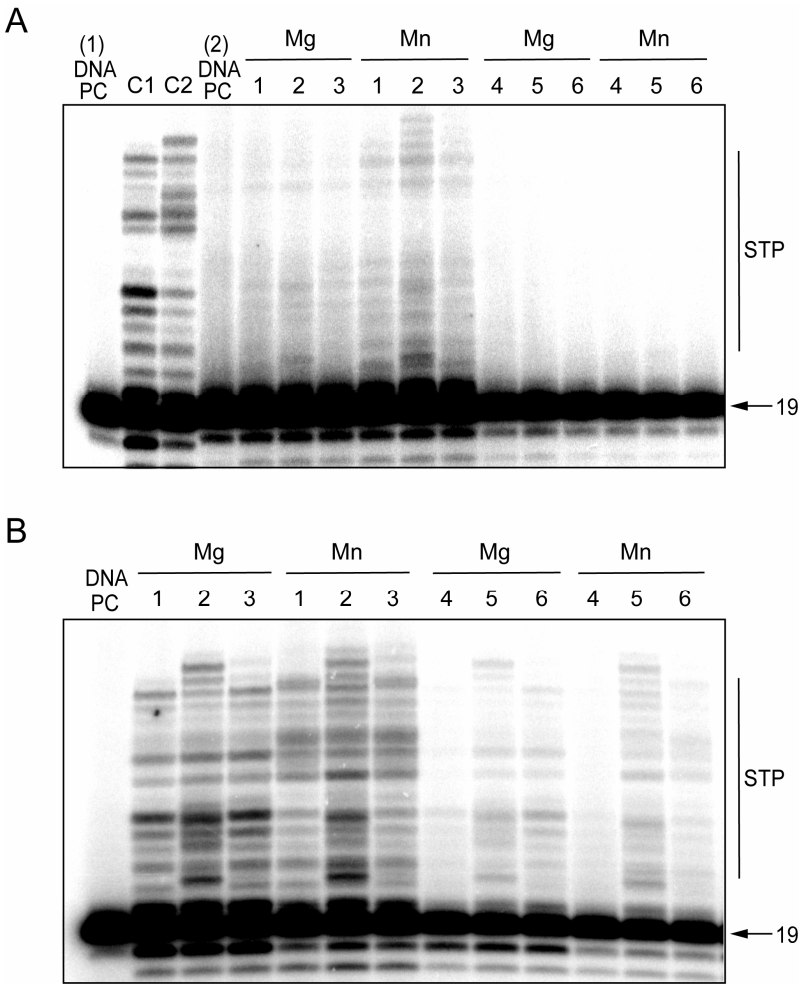

Figure 79: *gel images of experiments performed with phosphorothioate oligonucleotides as substrate for the in vitro integration assay.* ***(A)*** *Donor substrate: precleaved phosphorothioate top strand annealed with a phosphodiester bottom strand (lanes 1, 2 and 3) or with a phosphorothioate bottom strand (lanes 4, 5 and 6) (see Figure 78 for oligonucleotides' description). (1)DNA PC: precleaved phosphodiester substrate; C1: normal reaction as described in Figure 60 panel B in presence of magnesium; C2: normal reaction as described in Figure 60 panel B in presence of manganese; (2)DNA PC: precleaved phosphorothioate substrate; lanes 1 and 4: no acceptor substrate; lanes 2 and 5: full length phosphodiester acceptor substrate; lanes 3 and 6: full length mixed (phoshorothioate top strand annealed with a phosphodiester bottom strand) acceptor substrate.* ***(B)***

Donor substrate: precleaved phosphodiester top strand annealed with a phosphodiester bottom strand (lanes 1, 2 and 3) or with a phosphorothioate bottom strand (lanes 4, 5 and 6) (see Figure 78 for oligonucleotides' description). DNA PC: precleaved phosphodiester substrate; lanes 1 and 4: no acceptor substrate; lanes 2 and 5: full length phosphodiester acceptor substrate; lanes 3 and 6: full length mixed (phoshorothioate top strand annealed with a phosphodiester bottom strand) acceptor substrate. Bands labeled 19 and STP correspond to DNA substrate and strand transfer products, respectively.

Therefore we decided to study different substrates. Only the combination of a phosphodiester top strand and phosphorothioate bottom strand donor (oligonucleotide *B* annealed with *C* in *Figure 78*) resulted in a good reduction of autointegration (gel *Figure 79 panel B lane 4*) and in a sufficient integration in a phosphodiester acceptor (*Figure 79 panel B lane 5*). However, in experiments performed with longer DNA acceptors carrying random sequences, this weak integration signal was completely lost (data not shown), meaning that *in vitro* the enzyme has difficulties to recognize a non LTR sequence.
These are only preliminary results and additional experiments are warranted to define a good protocol to uncouple donor and acceptor substrates.

11. CONCLUSIONS

The present work was aimed to the *in vitro* study of WT integrase and of a series of mutants reported in literature as resistant to therapy. In particular the attention was addressed to the three reactions catalyzed by the enzyme: 3'-processing, strand transfer and disintegration. Eight recombinant integrases were expressed and purified. Their activity was analyzed with the use of oligonucleotides that mimic the terminal portion of the U5 viral LTR. Experiments demonstrated that 3'-P was only marginally impaired in most mutants compare to WT enzyme. The only mutation that markedly impacted 3'-P was Q148K. Conversely, strand transfer step was defective for most of the mutants. It was possible to define different profiles. Mutations T66I and S153Y determined partially defective 3'-P and ST; mutations F121Y, E92Q and N155H determined a selective defection on ST; Q148K IN was severely defective for both steps while L74M IN was the only one whose activities were close to WT IN.

The effect of two compounds on WT IN and on all the expressed mutants was then evaluated. Raltegravir, the first integrase inhibitor to be approved by FDA, and elvitegravir, which is undergoing clinical trials, were compared side by side.

Both compounds demonstrated a potent and selective inhibition of strand transfer at low nanomolar concentration. The drug concentrations required to inhibit 3'-P were at least 2 orders of magnitude higher than those required to inhibit ST. In addition, elvitegravir was approximately 2-3 fold more potent than raltegravir against IN-mediated ST. Instead disintegration step was not inhibited by both compounds. These findings suggest that both raltegravir and elvitegravir are highly specific for an intermediate step in the integrase catalytic cycle, which immediately follows 3'-P. The specificity for the ST step indicates that the drugs can selectively recognize a specific configuration of the IN-DNA complex and that they probably bind at the interface of the IN-viral DNA processed complex by forming a quaternary complex, with the catalytic metal in the enzyme active site. The fact that preincubation of the drugs with IN did not increase the drug inhibitory activities is also consistent with their binding at the IN-DNA interface rather than to IN alone. Thus, raltegravir and elvitegravir are candidate interfacial inhibitors.

The resistance profile for both drugs was then evaluated. The data showed that overall resistance to elvitegravir is greater than to raltegravir. The qualitative similarity in drug resistance profile across the seven mutants for raltegravir and elvitegravir is consistent with similarities of the drugs with respect to ST selectivity vs 3'-P and disintegration. Thus, these

results suggest that they interact in a similar way with IN and that they bind to a common region in the IN catalytic site, region that is probably outlined by the drug resistance mutations considered in this work. Experiments allowed also dividing the enzymes in three groups: the first includes mutations that confer no selectivity to either drug (L74M and E92Q); the second includes mutations that confer approximately 2-fold less resistance to raltegravir than elvitegravir (T66I, F121Y, Q148K and N155H); the third includes the S153Y mutation which appears more selectively resistant to elvitegravir than raltegravir. Based on these results with recombinant IN enzymes bearing resistance mutations, it seems unlikely that elvitegravir will be a rationale choice to overcome resistance to raltegravir.

Finally, the relationship between catalytic activity and drug resistance was examined. There was an overall linear correlation between loss of ST activity of the mutants and resistance to the considered drugs. The three most resistant enzymes (Q148K, T66I and S153Y) were also the most defective in ST activity. This is consistent with the specific interaction of the drugs with the IN catalytic site.

The data have been published in:

Marinello J, Marchand C, Mott BT, Bain A, Thomas CJ and Pommier Y. Comparison of Raltegravir and Elvitegravir on HIV-1 Integrase Catalytic Reactions and on a Series of Drug-Resistant Integrase Mutants. Biochemistry, Vol. 47, No. 36, 2008.

13. ABBREVIATION LIST

3'-P	3'-Processing
AIDS	Acquired Immune Deficiency Syndrome
AMD	Age-Related Macular Degeneration
APS	Ammonium Persulfate
ASV	Avian Sarcoma Virus
AZT	Azidothymidine (Zidovudine)
BAF	Barrier to Autointegration Factor
BB	Binding Buffer
CA	Capsid Protein
CCD	Catalytic Core Domain
CCK-8	Cell Counting Kit-8
CCR5	Chemokine co-receptor
CFB	Cell Feedback Network
CIAP	Calf Intestinal Alkalyne Phosphatase
CMIA	Chemiluminescent Magnetic Immunoassay
CR	Charged Region
CRF	Circulating Recombinant Form
CTD	Carboxy Terminal Domain
CXCR4	Chemokine co-receptor
DAG	Dialkylglycerol
DDE	Asp-Asp-Glu Motif
DHFR	Dihydrofolate Reductase
DKA	Diketo Acid
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
dTMP	Deoxythymidine Monophosphate
DTT	Dithiothreitol
dUMP	Deoxyuracil Monophosphate
dUTP	Deoxyuracil Triphosphate
EDTA	Ethylene Diamine Tetracetic Acid
EED	Embryonic Ectoderm Development Protein

EI	Entry Inhibitor
ELISA	Enzyme Linked Immuno Sorbent Assay
EMEA	European Medicines Evaluation Agency
env	Envelope Gene
EVG	Elvitegravir
FDA	Food and Drug Administration
FH_2	Dihydrofolate
FH_4	Tetrahydrofolate
FITC	Fluorescein Isothiocyanate
FPIA	Fluorescence Polarization Immunoassay
gag	Group-Specific Antigen Gene
gp	glycoprotein
HAART	Highly Active Antiretroviral Therapy
HCV	Hepatitis C Virus
HDV	Hepatitis Delta Virus
HHCC	His-His-Cys-Cys Motif
HIV	Human Immunodeficiency Virus
HMGA1	High Mobility Group Protein A1
HRP2	Hepatoma-Derived Growth Factor Related Protein 2
HSP60	Heat Shock Protein 60
HTS	High-Throughput Screening
IBD	Integrase Binding Domain
IN	Integrase
INI	Integrase Inhibitor
INI1	Integrase Interactor 1 Protein
IPTG	Isopropil-Beta-D-Thiogalactopyranoside
IRBM	Research Institute on Molecular Biology
ITC	Isothermal Tritration Calorimetry
K_B	Binding Constant
Kd	Dissociation Constant
LB	Luria-Bertani Medium
LEDGF	Lens Epithelium-Derived Growth Factor
LNA	Locked Nucleic Acid
LTR	Long Terminal Repeat
MA	Matrix Protein

MALDI	Matrix-Assisted Laser Desorption/Ionization
MCS	Multiple Cloning Site
MHC	Major Histocompatibility Complex
MOPS	3-(N-Morpholino)Propane Sulfonic Acid
mRNA	Messenger RNA
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NC	Nucleocapsid
Nef	Negative Regulatory Factor
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance
NNRTI	Non-Nucleoside Reverse Transcriptase Inhibitor
NRTI	Nucleoside/Nucleotide Reverse Transcriptase Inhibitor
NTD	Amino-Terminal Domain
PABA	Para-Aminobenzoic Acid
PBS	Primer Binding Site
PCR	Polymerase Chain Reaction
PDGF	Platelet-Derived Growth Factor
PEG	Polyethylene Glycol
PI	Protease Inhibitor
PIC	Pre-Integration Complex
Pol	Polymerase Gene
PPT	Polypurine Tract
PWWP	Pro-Trp-Trp-Pro Domain
Rev	Regulator of Expression of Virion Proteins
RGV	Raltegravir
RIA	Radio Immunoassay
RNA	Ribonucleic Acid
RNase	Ribonuclease Enzyme
RRE	Rev Response Element
RTC	Reverse Transcription Complex
SAM	S-Adenosyl-Methionine
SDS	Sodium Dodecyl Sulphate
SELEX	Systematic Evolution of Ligands by EXponential Enrichment
SIV	Simian Immunodeficiency Virus
ST	Strand Transfer

STP	Strand Transfer Reaction Product
SU	Surface Protein
TAR	Trans-Activation Responsive Element
Tat	Transactivator of Transcription
$TCID_{50}$	50% Tissue-Culture Infective Dose
TEMED	N,N,N',N'-Tetramethylethylenediamine
TM	Transmembrane Protein
tRNA	Transfer RNA
URF	Unique Recombinant Form
UTP	Uridine Triphosphate
UV	Ultraviolet light
VEGF	Vascular Endothelial Growth Factor
Vif	Virion Infectivity Factor
Vpr	Viral Protein R
Vpu	Viral Protein U
VS	Varkud Satellite
WT	Wild-Type
ΔG	Gibbs Free Energy Variation
ΔH	Enthalpy Variation

13. REFERENCES

1. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-10.
2. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. **346**(6287): p. 818-22.
3. Proske, D., et al., *Aptamers--basic research, drug development, and clinical applications.* Appl Microbiol Biotechnol, 2005. **69**(4): p. 367-74.
4. Lee, J.F., G.M. Stovall, and A.D. Ellington, *Aptamer therapeutics advance.* Curr Opin Chem Biol, 2006. **10**(3): p. 282-9.
5. Jayasena, S.D., *Aptamers: an emerging class of molecules that rival antibodies in diagnostics.* Clin Chem, 1999. **45**(9): p. 1628-50.
6. Tombelli, S., M. Minunni, and M. Mascini, *Analytical applications of aptamers.* Biosens Bioelectron, 2005. **20**(12): p. 2424-34.
7. Cox, J.C. and A.D. Ellington, *Automated selection of anti-protein aptamers.* Bioorg Med Chem, 2001. **9**(10): p. 2525-31.
8. Cox, J.C., et al., *Automated acquisition of aptamer sequences.* Comb Chem High Throughput Screen, 2002. **5**(4): p. 289-99.
9. Golden, M.C., et al., *Diagnostic potential of PhotoSELEX-evolved ssDNA aptamers.* J Biotechnol, 2000. **81**(2-3): p. 167-78.
10. Shangguan, D., et al., *Aptamers evolved from live cells as effective molecular probes for cancer study.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11838-43.
11. Russell, R., et al., *Rapid compaction during RNA folding.* Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4266-71.
12. Heilig, J., *Abstract Book 2nd Annual Nucleic Acid World Summit Boston.* 2004.
13. Fredriksson, S., et al., *Protein detection using proximity-dependent DNA ligation assays.* Nat Biotechnol, 2002. **20**(5): p. 473-7.
14. Hamaguchi, N., A. Ellington, and M. Stanton, *Aptamer beacons for the direct detection of proteins.* Anal Biochem, 2001. **294**(2): p. 126-31.
15. Yamamoto, R., T. Baba, and P.K. Kumar, *Molecular beacon aptamer fluoresces in the presence of Tat protein of HIV-1.* Genes Cells, 2000. **5**(5): p. 389-96.
16. Bock, C., et al., *Photoaptamer arrays applied to multiplexed proteomic analysis.* Proteomics, 2004. **4**(3): p. 609-18.
17. Petach, H. and L. Gold, *Dimensionality is the issue: use of photoaptamers in protein microarrays.* Curr Opin Biotechnol, 2002. **13**(4): p. 309-14.
18. Blank, M., et al., *Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. selective targeting of endothelial regulatory protein pigpen.* J Biol Chem, 2001. **276**(19): p. 16464-8.
19. Marro, M.L., et al., *Identification of potent and selective RNA antagonists of the IFN-gamma-inducible CXCL10 chemokine.* Biochemistry, 2005. **44**(23): p. 8449-60.
20. Blank, M. and M. Blind, *Aptamers as tools for target validation.* Curr Opin Chem Biol, 2005. **9**(4): p. 336-42.
21. Zhou, B. and B. Wang, *Pegaptanib for the treatment of age-related macular degeneration.* Exp Eye Res, 2006. **83**(3): p. 615-9.
22. Ng, E.W., et al., *Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease.* Nat Rev Drug Discov, 2006. **5**(2): p. 123-32.
23. Famulok, M., J.S. Hartig, and G. Mayer, *Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy.* Chem Rev, 2007. **107**(9): p. 3715-43.

24. Cech, T.R., *Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena.* Biosci Rep, 1990. **10**(3): p. 239-61.
25. Altman, S., *Nobel lecture. Enzymatic cleavage of RNA by RNA.* Biosci Rep, 1990. **10**(4): p. 317-37.
26. Saldanha, R., et al., *Group I and group II introns.* Faseb J, 1993. **7**(1): p. 15-24.
27. Frank, D.N. and N.R. Pace, *Ribonuclease P: unity and diversity in a tRNA processing ribozyme.* Annu Rev Biochem, 1998. **67**: p. 153-80.
28. Lilley, D.M., *RNA folding and catalysis.* Genetica, 1999. **106**(1-2): p. 95-102.
29. Scott, W.G., *Biophysical and biochemical investigations of RNA catalysis in the hammerhead ribozyme.* Q Rev Biophys, 1999. **32**(3): p. 241-84.
30. Tanner, N.K., *Ribozymes: the characteristics and properties of catalytic RNAs.* FEMS Microbiol Rev, 1999. **23**(3): p. 257-75.
31. Koizumi, M., et al., *Allosteric ribozymes sensitive to the second messengers cAMP and cGMP.* Nucleic Acids Symp Ser, 1999(42): p. 275-6.
32. Pitkin, R.M., *Folate and neural tube defects.* Am J Clin Nutr, 2007. **85**(1): p. 285S-288S.
33. Westhof, E. and V. Fritsch, *RNA folding: beyond Watson-Crick pairs.* Structure, 2000. **8**(3): p. R55-65.
34. Chandrasekhar, K. and R. Malathhi, *Non-Watson Crick base pairs might stabilize RNA structural motifs in ribozymes -- a comparative study of group-I intron structures.* J Biosci, 2003. **28**(5): p. 547-55.
35. Gilbert, W., *Origin of life: the RNA world.* Nature, 1986. **319**: p. 618.
36. Pace, P. and M. Rowley, *Integrase inhibitors for the treatment of HIV infection.* Curr Opin Drug Discov Devel, 2008. **11**(4): p. 471-9.
37. Van Heuverswyn, F., et al., *Human immunodeficiency viruses: SIV infection in wild gorillas.* Nature, 2006. **444**(7116): p. 164.
38. Taylor, B.S., et al., *The challenge of HIV-1 subtype diversity.* N Engl J Med, 2008. **358**(15): p. 1590-602.
39. Cohen, M.S., et al., *The spread, treatment, and prevention of HIV-1: evolution of a global pandemic.* J Clin Invest, 2008. **118**(4): p. 1244-54.
40. Scherer, L., J.J. Rossi, and M.S. Weinberg, *Progress and prospects: RNA-based therapies for treatment of HIV infection.* Gene Ther, 2007. **14**(14): p. 1057-64.
41. Sarafianos, S.G., et al., *Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA.* Embo J, 2001. **20**(6): p. 1449-61.
42. Wu, Y., *HIV-1 gene expression: lessons from provirus and non-integrated DNA.* Retrovirology, 2004. **1**: p. 13.
43. Semenova, E.A., C. Marchand, and Y. Pommier, *HIV-1 integrase inhibitors: update and perspectives.* Adv Pharmacol, 2008. **56**: p. 199-228.
44. Pommier, Y., A.A. Johnson, and C. Marchand, *Integrase inhibitors to treat HIV/AIDS.* Nat Rev Drug Discov, 2005. **4**(3): p. 236-48.
45. Rice, P.A. and T.A. Baker, *Comparative architecture of transposase and integrase complexes.* Nat Struct Biol, 2001. **8**(5): p. 302-7.
46. Chiu, T.K. and D.R. Davies, *Structure and function of HIV-1 integrase.* Curr Top Med Chem, 2004. **4**(9): p. 965-77.
47. Wang, J.Y., et al., *Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein.* Embo J, 2001. **20**(24): p. 7333-43.
48. Cai, M., et al., *Solution structure of the N-terminal zinc binding domain of HIV-1 integrase.* Nat Struct Biol, 1997. **4**(7): p. 567-77.
49. Eijkelenboom, A.P., et al., *The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc.* Curr Biol, 1997. **7**(10): p. 739-46.

50. Engelman, A. and R. Craigie, *Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro.* J Virol, 1992. **66**(11): p. 6361-9.
51. Zheng, R., T.M. Jenkins, and R. Craigie, *Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13659-64.
52. Asante-Appiah, E. and A.M. Skalka, *Molecular mechanisms in retrovirus DNA integration.* Antiviral Res, 1997. **36**(3): p. 139-56.
53. Chen, J.C., et al., *Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding.* Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8233-8.
54. Dyda, F., et al., *Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.* Science, 1994. **266**(5193): p. 1981-6.
55. Jenkins, T.M., et al., *A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization.* J Biol Chem, 1996. **271**(13): p. 7712-8.
56. Katzman, M. and M. Sudol, *Nonspecific alcoholysis, a novel endonuclease activity of human immunodeficiency virus type 1 and other retroviral integrases.* J Virol, 1996. **70**(4): p. 2598-604.
57. Engelman, A. and R. Craigie, *Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity.* J Virol, 1995. **69**(9): p. 5908-11.
58. Esposito, D. and R. Craigie, *Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction.* Embo J, 1998. **17**(19): p. 5832-43.
59. Grobler, J.A., et al., *Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6661-6.
60. Bujacz, G., et al., *Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity.* J Biol Chem, 1997. **272**(29): p. 18161-8.
61. Beese, L.S. and T.A. Steitz, *Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism.* Embo J, 1991. **10**(1): p. 25-33.
62. Johnson, A.A., et al., *Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop.* J Biol Chem, 2006. **281**(1): p. 461-7.
63. Marchand, C., et al., *Mechanisms and inhibition of HIV integration.* Drug Discovery Today:Disease Mechanisms, 2006. **3**(2): p. 253-260.
64. Van Maele, B., et al., *Cellular co-factors of HIV-1 integration.* Trends Biochem Sci, 2006. **31**(2): p. 98-105.
65. Guiot, E., et al., *Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity.* J Biol Chem, 2006. **281**(32): p. 22707-19.
66. Young, F.E., *The role of the FDA in the effort against AIDS.* Public Health Rep, 1988. **103**(3): p. 242-5.
67. Evering, T.H. and M. Markowitz, *Raltegravir: an integrase inhibitor for HIV-1.* Expert Opin Investig Drugs, 2008. **17**(3): p. 413-22.
68. Grim, S.A. and F. Romanelli, *Tenofovir disoproxil fumarate.* Ann Pharmacother, 2003. **37**(6): p. 849-59.
69. Kearney, B.P., J.F. Flaherty, and J. Shah, *Tenofovir disoproxil fumarate: clinical pharmacology and pharmacokinetics.* Clin Pharmacokinet, 2004. **43**(9): p. 595-612.
70. Modrzejewski, K.A. and R.A. Herman, *Emtricitabine: a once-daily nucleoside reverse transcriptase inhibitor.* Ann Pharmacother, 2004. **38**(6): p. 1006-14.

71. Arion, D., et al., *Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase.* Biochemistry, 1998. **37**(45): p. 15908-17.
72. Quinones-Mateu, M.E., et al., *Viral drug resistance and fitness.* Adv Pharmacol, 2008. **56**: p. 257-96.
73. Kohlstaedt, L.A., et al., *Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor.* Science, 1992. **256**(5065): p. 1783-90.
74. Hsiou, Y., et al., *Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms.* Structure, 1996. **4**(7): p. 853-60.
75. Nijhuis, M., S. Deeks, and C. Boucher, *Implications of antiretroviral resistance on viral fitness.* Curr Opin Infect Dis, 2001. **14**(1): p. 23-8.
76. Nijhuis, M., et al., *Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy.* Aids, 1999. **13**(17): p. 2349-59.
77. Zhang, Y.M., et al., *Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites.* J Virol, 1997. **71**(9): p. 6662-70.
78. de la Carriere, L.C., et al., *Effects of human immunodeficiency virus type 1 resistance to protease inhibitors on reverse transcriptase processing, activity, and drug sensitivity.* J Virol, 1999. **73**(4): p. 3455-9.
79. Birk, M., V. Svedhem, and A. Sonnerborg, *Kinetics of HIV-1 RNA and resistance-associated mutations after cessation of antiretroviral combination therapy.* Aids, 2001. **15**(11): p. 1359-68.
80. De Clercq, E., *Antiviral drugs in current clinical use.* Journal of Clinical Virology, 2004. **30**: p. 115-133.
81. Hazuda, D.J., et al., *Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells.* Science, 2000. **287**(5453): p. 646-50.
82. Svarovskaia, E.S., et al., *Azido-containing diketo acid derivatives inhibit human immunodeficiency virus type 1 integrase in vivo and influence the frequency of deletions at two-long-terminal-repeat-circle junctions.* J Virol, 2004. **78**(7): p. 3210-22.
83. Fikkert, V., et al., *Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations.* J Virol, 2003. **77**(21): p. 11459-70.
84. King, P.J. and W.E. Robinson, Jr., *Resistance to the anti-human immunodeficiency virus type 1 compound L-chicoric acid results from a single mutation at amino acid 140 of integrase.* J Virol, 1998. **72**(10): p. 8420-4.
85. Goldgur, Y., et al., *Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design.* Proc Natl Acad Sci U S A, 1999. **96**(23): p. 13040-3.
86. Fesen, M.R., et al., *Inhibitors of human immunodeficiency virus integrase.* Proc Natl Acad Sci U S A, 1993. **90**(6): p. 2399-403.
87. Johnson, A.A., et al., *Probing HIV-1 integrase inhibitor binding sites with position-specific integrase-DNA cross-linking assays.* Mol Pharmacol, 2007. **71**(3): p. 893-901.
88. Savarino, A., *In-Silico docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate.* Retrovirology, 2007. **4**: p. 21.
89. Little, S., et al., *Protocol 004 study team. Antiviral effect of L-870810, a novel HIV-1 integrase inhibitor, in HIV-1 infected patients.* 12th Conference on Retroviruses and Opportunistic Infections, 2005. **Boston - MA**.

90. Sato, M., et al., *Novel HIV-1 integrase inhibitors derived from quinolone antibiotics.* J Med Chem, 2006. **49**(5): p. 1506-8.
91. Zolopa, A.R., et al., *The HIV integrase inhibitor GS-9137 demonstrates potent antiretroviral activity in treatment-experienced patients.* 14th Conference on Retroviruses and Opportunistic Infections, 2007. **Los Angeles - CA**: p. [Abstract 143LB].
92. Summa, V., et al., *HCV NS5b RNA-dependent RNA polymerase inhibitors: from alpha,gamma-diketoacids to 4,5-dihydroxypyrimidine- or 3-methyl-5-hydroxypyrimidinonecarboxylic acids. Design and synthesis.* J Med Chem, 2004. **47**(22): p. 5336-9.
93. Summa, V., et al., *4,5-dihydroxypyrimidine carboxamides and N-alkyl-5-hydroxypyrimidinone carboxamides are potent, selective HIV integrase inhibitors with good pharmacokinetic profiles in preclinical species.* J Med Chem, 2006. **49**(23): p. 6646-9.
94. Kassahun, K., et al., *Metabolism and disposition in humans of raltegravir (MK-0518), an anti-AIDS drug targeting the human immunodeficiency virus 1 integrase enzyme.* Drug Metab Dispos, 2007. **35**(9): p. 1657-63.
95. Markowitz, M., et al., *Antiretroviral activity, pharmacokinetics, and tolerability of MK-0518, a novel inhibitor of HIV-1 integrase, dosed as monotherapy for 10 days in treatment-naive HIV-1-infected individuals.* J Acquir Immune Defic Syndr, 2006. **43**(5): p. 509-15.
96. Steigbigel, R.T., et al., *Raltegravir with optimized background therapy for resistant HIV-1 infection.* N Engl J Med, 2008. **359**(4): p. 339-54.
97. Engelman, A. and P. Cherepanov, *The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication.* PLoS Pathog, 2008. **4**(3): p. e1000046.
98. Al-Mawsawi, L.Q. and N. Neamati, *Blocking interactions between HIV-1 integrase and cellular cofactors: an emerging anti-retroviral strategy.* Trends Pharmacol Sci, 2007. **28**(10): p. 526-35.
99. Cherepanov, P., et al., *HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells.* J Biol Chem, 2003. **278**(1): p. 372-81.
100. Cherepanov, P., et al., *Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75.* Proc Natl Acad Sci U S A, 2005. **102**(48): p. 17308-13.
101. Poeschla, E.M., *Integrase, LEDGF/p75 and HIV replication.* Cell Mol Life Sci, 2008. **65**(9): p. 1403-24.
102. Shun, M.C., et al., *LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration.* Genes Dev, 2007. **21**(14): p. 1767-78.
103. Molteni, V., et al., *Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 4): p. 536-44.
104. Walker, B.D. and D.R. Burton, *Toward an AIDS vaccine.* Science, 2008. **320**(5877): p. 760-4.
105. Steinbrook, R., *One step forward, two steps back--will there ever be an AIDS vaccine?* N Engl J Med, 2007. **357**(26): p. 2653-5.
106. Gottlieb, M.S., et al., *Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency.* N Engl J Med, 1981. **305**(24): p. 1425-31.
107. Masur, H., et al., *An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction.* N Engl J Med, 1981. **305**(24): p. 1431-8.

108. Siegal, F.P., et al., *Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesions.* N Engl J Med, 1981. **305**(24): p. 1439-44.
109. Hazuda, D.J., et al., *Resistance to the HIV-integrase inhibitor raltegravir: analysis of protocol 005, a Phase II study in patients with triple-class resistant HIV-1 infection.* Antiviral Therapy, 2007. **12**: p. Abstract 8.
110. Marchand, C., N. Neamati, and Y. Pommier, *In vitro human immunodeficiency virus type 1 integrase assays.* Methods Enzymol, 2001. **340**: p. 624-33.
111. Craigie, R., et al., *A rapid in vitro assay for HIV DNA integration.* Nucleic Acids Res, 1991. **19**(10): p. 2729-34.
112. Maignan, S., et al., *Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases.* J Mol Biol, 1998. **282**(2): p. 359-68.
113. Marchand, C., et al., *Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S).* Mol Pharmacol, 2003. **64**(3): p. 600-9.
114. Malet, I., et al., *Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro.* Antimicrob Agents Chemother, 2008. **52**(4): p. 1351-8.
115. Cooper, D.A., et al., *Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection.* New England Journal Medicine, 2008. **359**(4): p. 355-365.
116. Mascolini, M., *Early clues from raltegravir failure in clinical practice; diverse patterns of resistance mutations.* XVII International HIV Drug Resistance Workshop - Sitges - Spain, 2008. **Brief Report**.
117. Kobayashi, M., et al., *Selection of diverse and clinically relevant integrase inhibitor-resistant human immunodeficiency virus type 1 mutants.* Antiviral Research, 2008. **80**: p. 213-222.
118. Jones, G., et al., *Resistance profile of HIV-1 mutants in vitro selected by the HIV-1 integrase inhibitor, GS-9137 (JTK-303).* 14th Conference on Retroviruses and Opportunistic Infections - Los Angeles, 2007. **[Abstract]**.
119. Shimura, K., et al., *Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137).* J Virol, 2008. **82**(2): p. 764-74.
120. Garvey, E.P., et al., *The naphthyridinone GSK364735 is a novel, potent human immunodeficiency virus type 1 integrase inhibitor and antiretroviral.* Antimicrobial Agents and Chemotherapy, 2007. **52**(3): p. 901-908.
121. Goethals, O., et al., *Resistance mutations in HIV-1 integrase selected with elvitegravir confer reduced susceptibility to a wide range of integrase inhibitors.* J. Virol, 2008.
122. Goodman, D., et al., *Integrase inhibitor resistance involves complex interactions among primary and secondary resistance mutations: a novel mutation L68V/I associates with E92Q and increases resistance.* XVII International HIV Drug Resistance Workshop - Sitges - Spain, 2008. **[Abstract 13]**.
123. Schiff, R.D. and D.P. Grandgenett, *Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the beta polypeptide of the avian retrovirus DNA polymerase.* J Virol, 1978. **28**(1): p. 279-91.
124. Grandgenett, D.P., A.C. Vora, and R.D. Schiff, *A 32,000-dalton nucleic acid-binding protein from avian retravirus cores possesses DNA endonuclease activity.* Virology, 1978. **89**(1): p. 119-32.
125. Chow, S.A., et al., *Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.* Science, 1992. **255**(5045): p. 723-6.
126. Pommier, Y. and C. Marchand, *Interfacial inhibitors of protein-nucleic acid interactions.* Curr Med Chem Anticancer Agents, 2005. **5**(4): p. 421-9.

127. Pommier, Y. and J. Cherfils, *Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery.* Trends Pharmacol Sci, 2005. **26**(3): p. 138-45.
128. Nicastri, E., et al., *Replication capacity, biological phenotype, and drug resistance of HIV strains isolated from patients failing antiretroviral therapy.* J Med Virol, 2003. **69**(1): p. 1-6.
129. Mammano, F., C. Petit, and F. Clavel, *Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients.* J Virol, 1998. **72**(9): p. 7632-7.
130. Martinez-Picado, J., et al., *Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1.* J Virol, 1999. **73**(5): p. 3744-52.

ACKNOWLEDGEMENTS

I want thank all people that ran with me through these three years and taught me how to carry out good scientific research. In particular my supervisor Barbara Gatto and Manlio Palumbo. I'm also infinitely grateful to Yves Pommier for the opportunity to spend more than one year in his laboratory and to experience the work in a big research institute as NIH. Finally thank you to Christophe Marchand for the constant support and to the other members of LMP who helped me feeling at home in a place so far away from Italy.