Computational Analysis of Human Immunodeficiency Virus (HIV) Type-1 Reverse Transcriptase Crystallographic Models Based on Significant Conserved Residues Found in Highly Active Antiretroviral Therapy (HAART)-Treated Patients

Stefano Alcaro*,[1], Anna Artese[1], Francesca Ceccherini-Silberstein[2], Vitaliano Chiarella[3], Salvatore Dimonte[2], Francesco Ortuso[1] and Carlo Federico Perno[2]

[1]*Laboratorio di Chimica Farmaceutica Computazionale - Dipartimento di Scienze Farmacobiologiche Università "Magna Græcia" di Catanzaro, Campus Universitario, Viale Europa, 88100 Catanzaro, Italy*

[2]*Dipartimento di Medicina Sperimentale e Scienze Biochimiche, Università di Roma "Tor Vergata", Via Montpellier, 1, 00133, Roma, Italy*

[3]*Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, Via E. Fermi, 40, 00044 Frascati (Roma), Italy*

Abstract: Reverse transcription of the viral single-stranded (+) RNA genome into double-stranded DNA is an essential step in the human immunodeficiency virus' (HIV) life-cycle. Although several viral proteins are involved in the regulation and/or efficiency of reverse transcription, the process of retroviral DNA synthesis is entirely dependent on the enzymatic activities of the retroviral reverse transcriptase enzyme (RT). Due to its crucial role in the HIV life-cycle, RT is a primary target for anti-HIV drug development. Nonetheless, drug resistance is the major problem affecting the clinical efficacy of antiretroviral agents. Incomplete pharmacological pressure represents the logical cause and not the consequence of different mutation pathways in RT associated with approved inhibitors resistance.

In this review we have analyzed RT Protein Data Bank (PDB) models using our innovative computational approach "GRID Based Pharmacophore Model" (GBPM). This method was applied to clinically relevant RT conserved residues found in a large cohort of HAART treated patients. The PDB entries have been selected among the unbound and the complexed models with DNA and/or inhibitors. Such an approach has revealed itself useful to highlight the mutation effects in the drug-RT recognition as well as in the heterodimer stabilization of the enzyme. Most of the clinical and biochemical evidences already reported in the literature have been rationalized at molecular level *via* the GBPM computational approach. A definite future application of this method will be the identification of conserved regions of critical macromolecules, such as the HIV-1 RT, to be targeted for the development of innovative therapeutic agents.

Keywords: Reverse transcriptase, crystallographic models, GBPM, HAART, conserved residues, interaction energy.

INTRODUCTION

Reverse transcriptase (RT) is an essential enzyme for the HIV life-cycle. In the replication cycle, RT converts viral genomic RNA into a double stranded linear DNA. It is an asymmetric heterodimer composed of a 560 amino-acid 66-kDa subunit (p66) and a 440 amino-acid 51-kDa subunit (p51). HIV-1 protease provides for the cleavage of the C-terminal RNase H domain of the p66 subunit, thus generating the p51 polypeptide. Located in the p66 subunit at spatially distinct regions, the p66/p51 HIV-1 RT heterodimer contains one DNA polymerization active site and one RNase H active site. Even if the p51 subunit presents the same amino acid sequence of p66, it is not directly involved in catalytic enzyme activities, but is pivotal for loading the p66 subunit on the template/primer [1].

The p66 subunit consists of the *fingers* (residues 1-85 and 118-155), *palm* (residues 86-117 and 156-237), and *thumb* (residues 238-318) subdomains. Placed in the *palm* subdomain, the three aspartic acid residues Asp110, Asp185 and Asp186 define the polymerase active site. Other regions include the *connection* (residues 319-426) and the C-terminal

RNase H domains (residues 427-560). The RNase H active site is identified by four conserved acidic amino acids (Asp443, Glu478, Asp498 and Asp549). The *connection* domain has the crucial role of linking the polymerase and RNase H domains, but is also implicated in nucleic acid substrate interactions and RT inter-subunit interactions. The p51 subunit contains the same polymerase and *connection* domains as the p66 subunit, but their folding is different from those of the p66 subunit. The p66 subunit is able to accommodate a nucleic acid template strand due to its "open" catalytically-competent conformation. On the other hand the p51 subunit shows a compact folded conformation so that the active site residues in this subunit are buried and therefore non-functional [2]. The interactions between the two subunits are mediated by their *connection* subdomains; other contacts between the *thumb* subdomain of p51 and RNase H subdomain of p66 are also relevant [3]. The RNase H domain of HIV-1 RT has been crystallized in free form, since isolation makes it inactive; the presence of p51 restores its activity, indicating that additional RT residues play a significant role in RNase function [4].

Crystallographic studies of HIV-1 RT with duplex DNA and an RNA/DNA hybrid allowed to identify several enzymatic regions considered essential for efficient nucleic acid polymerization and hydrolysis as well as translocation step. The DNA duplex binds to the RNase H active site along a groove far away from the polymerase active site. The main

contacts of the nucleic acid occur with the p66 subunit of the RT heterodimer and in the minor groove take place direct interactions with bases. There are van der Waals contacts with Pro157 and Met184 and with Ile94, and hydrogen bonds with Tyr183 [5]. The catalytic triad is placed near to the 3' terminus. The side chains of these residues are carboxyl groups able to induce the DNA nucleophilic attack during the polymerization step. The primer strand interacts with the loop between *palm* and *thumb* subdomains (Met230 and Gly231: the "primer grip"), and both strands establish contacts with residues in helices H and I. Part of the catalytic site, the YMMD motif (Tyr183-Asp186) results crucial in polymerase activity, while single substitutions at one or more of these residues are related to a drastic decrease of viral replication. Several mutagenesis studies showed that some residues of the primer grip region (residues 226-235) have a pivotal role both in catalytic activity and in heterodimer association [6].

Although p51 subunit lacks a direct RT catalytic activity, it plays an essential role in dimer stabilization and enzyme functions. Its *thumb* subdomain is responsible for guaranteeing RT activities through crucial interactions with the RNAse H domain. Site-directed mutagenesis analysis confirmed such a data, showing that Asn255Asp, Asn265Asp and Cys280Ser mutations are related to an incorrect polymerization process, a reduced DNA affinity and a decreased RNase H activity, respectively [7].

The biologically relevant form of RT is represented by the heterodimer; the monomeric subunits only show low catalytic activity. A structural analysis demonstrated that p66 and p51 subunits mainly interact *via* three contacts, with a large hydrophobic surface character [8]. These three major interactions occur at the dimer interface, including the *fingers* subdomain of p51 with the *palm* of p66, the *connection* subdomains of both subunits and the *thumb* subdomain of p51 with the RNase H domain of p66. Specific mutagenesis studies showed that mutating several residues in these regions dramatically reduced the interactions between p51 and p66 subunits, thereby decreasing the heterodimer stability [9]. Another critical region essential for proper RT dimerization contains a hepta leucine-repeat motif from leucine 282 to leucine 310 [10]. This area presents an important role in protein-protein interactions required for dimerization [11]. After site-directed mutagenesis of these leucine residues, Leu289Lys of p66 was unable to dimerize with neither itself nor the wild-type (Leu289Lys of p51).

The loop between β-strands 7 and 8 in HIV-1 RT is highly conserved in amino acid sequence [12]. This sequence, "SINNET" (residues 134-139), is relatively conserved in drug-naïve HIV-1 clones/isolates with the exception of Ile135, which is sometimes substituted by threonine or valine. Some structural analyses indicated the presence of a key inter-subunit interaction between the side-chain of Asn136 in p51 and the backbone of residues Pro95 and His96 at the carboxy-terminal end of the β5-strand in p66 [13]. Alanine substitution of Asn136 and of other four amino acids in the β7-β8 loop destabilized the heterodimeric RT, reduced the DNA binding ability of the enzyme and significantly decreased viral infectivity [9, 14].

Several drugs have been studied with the aim to block RT activities such as nucleoside and nucleotide inhibitors (NRTIs) and non-nucleoside inhibitors (NNRTIs), whose RT binding site is specific.

The NRTIs are analogues of the natural nucleosides. In order for them to be activated, they require phosphorylation in the host's cells and, by a competitive mechanism with its binding site, behave as chain terminators. To exert such a mechanism, these inhibitors must have a free hydroxyl group in 3' position [15]. Subsequently they are converted in triphosphate derivatives by a cellular thymidine kinase, which activates them to block the natural dNTPs binding to DNA [16]. In contrast, the nucleotide-RTIs, such as adefovir and tenofovir, already possess a phosphonate group and, therefore, only need two phosphorylation steps to be converted to the active metabolites [17].

The polymerase activity of the catalytic triad can be significantly influenced by conformational changes that occur in an allosteric site known as NNRTI binding pocket (NNRTI-BP). This site is located between the β6-β10-β9 and β12-β13-β14 sheets in the *palm* subdomain of the p66 subunit. It shows a largely hydrophobic nature with substantial aromatic character (Tyr181, Tyr188, Phe227, Trp229, and Tyr232), but it is also surrounded by several hydrophilic residues (Lys101, Lys103, Ser105, Asp192, and Glu224 of the p66 subunit and Glu138 of the β7-β8 loop of the p51 subunit).

Residues Leu100, Lys101, Lys103, Val179, and Tyr181 of the p66 subunit and Glu138 of the p51 subunit define a putative solvent accessible entrance to the NNRTI-BP placed at the heterodimer interface [18]. Structural studies of HIV-1 RT demonstrate that the binding of an NNRTI causes both short-range and long-range distortions of the enzyme. The short-range distortions consist of some conformational changes of the amino acids and/or structural elements that form the NNRTI-BP, such as the re-orientation of the Tyr181 and Tyr188 side chains and the displacement of the β12-β13-β14 sheet. The long-range distortions involve a displacement of the p66 *connection*, the RNase H domain, and the p51 subunit with respect to the polymerase active site [18]. There are two interesting aspects of the NNRTI-BP: a) its location close to the subunit-subunit interface b) the pocket entrance composed by residues from the p66 (Leu100, Lys101, Lys103, Val179 and Tyr181) and the p51 (Glu138) subunits that also form part of the RT dimer interface.

NNRTIs can be classified into more than 30 structurally different groups of compounds. They can be divided into first- or second-generation compounds [19]. The first generation NNRTIs, such as nevirapine, delavirdine, TIBO and loviride were mainly discovered by random screening and are associated with the rapid development of drug resistance mutations. Molecular modeling strategies, rationale-based drug synthesis and biological and pharmacokinetic evaluations allowed to identify other first generation NNRTIs, which include efavirenz, the quinoxaline talviraline and the imidazole capravirine. These NNRTIs show an high potency, and in general are more active against a broader spectrum of drug-resistant strains of HIV-1 [20]. Finally, second generation NNRTIs, such as TMC-125 (etravirine), TMC-278, IDX899, RDEA806, have been designed to be active against the most relevant resistance mutations that develop after

failure of efavirenz or nevirapine treatment and are drugs with high genetic barrier to resistance [21]. In particular etravirine is an highly flexible compound active against HIV-1 strains resistant to current NNRTIs and has confirmed its efficacy in controlled trials with treatment-experienced patients (DUET studies) [22].

The identification of NNRTI common pharmacophore features is critical for the interaction with RT NNRTI-BP. Some structural data is available [23] and highlights an aromatic ring capable of π-stacking interactions, NH-C=O or NH-C=S groups able to participate in hydrogen bonding, and one or more hydrocarbon-rich regions involved in hydrophobic contacts.

HIV-1 RT possesses the intrinsic property to lack proofreading function, in contrast to other DNA polymerases. New viral variants are always generated due to RT error-prone nature, together with the high rate of virus production sustained by HIV-1 infection *in vivo*.

Drug pressure is responsible to dramatically increase such an intrinsic variability, with the final development of mutations, especially in the *pol* region encoding both RT and protease (PR) enzymes. Drug resistance is the major problem affecting the clinical efficacy of antiretroviral agents. Resistance is the consequence of mutations that emerge in the viral proteins targeted by antiretroviral agents and viral replication is poorly controlled due to the transmission of drug-resistant strains [24]. The combination of antiretroviral drugs used for the treatment of HIV infection is known as highly active antiretroviral therapy (HAART) and its regimens generally comprise three antiretroviral drugs, usually two NRTIs and either PR inhibitors or a NNRTI drug. Resistance to nucleoside and nucleotide analogues is based on two distinct mechanisms:

a)   incorporation impairment of the analogue into DNA;

b)   removal of the analogue from the prematurely terminated DNA chain.

The incorporation of a nucleoside analogue into drug-sensitive viruses results in the termination of the viral DNA chain. Mutations in drug-resistant viruses prevent the incorporation of the nucleoside analogue into the growing viral DNA chain. Removal of the nucleoside analogue from the terminated DNA chain is associated with a group of mutations commonly defined "thymidine analogue mutations" (TAMs). These mutations frequently emerge after the failure of drug combinations including thymidine analogues, such as zidovudine and stavudine. However they can promote resistance to almost all nucleoside analogues. These substitutions promote resistance by ATP- or pyrophosphate-mediated removal of nucleoside analogues from the 3' end of the terminated strand [24]. Resistance associated to the failure of treatment with NNRTIs is caused by several mutations located in the hydrophobic pocket (NNRTI-BP) targeted by these compounds. These mutations are drug-dependent and reduce the affinity of the drug, allowing DNA polymerization to proceed normally [25].

Although HIV-1, due to its high replication and mutation rate, shows resistance to the currently available drugs, several regions are highly conserved even under pharmacological pressure. Several HIV-1 RT residues were found to be highly conserved, highlighting their crucial role in protein stability, heterodimerization step, template-primer binding and catalytic activity [12]. Moreover the RT has been recently crystallized in the presence of several antiretroviral drugs and a large number of HIV-1 RT structural models are deposited into the Protein Data Bank (PDB) [26].

Recently we have proposed a new computational method [27] for designing pharmacophore models starting from detailed macromolecular structures, such as those obtained by crystallographic measurements. The method, called GBPM (GRID Based Pharmacophore Model), is based on the well known computational approach [28] developed by Prof. Goodford in the early '80 and implemented in the GRID software [29]. The GBPM has been successfully applied in studying supramolecular complexes of different chemical nature [27].

In this review we have analyzed RT crystallographic models by our innovative computational approach GBPM and applied it to clinically relevant enzyme conserved residues found in a large cohort of HAART treated patients. Such an approach, in agreement with clinical results, could allow to design new pharmacophoric models helpful in the drug discovery of more effective anti-RT inhibitors with more favorable resistance profiles.

THE GBPM METHOD

The GBPM method was developed with the aim to generate pharmacophore models useful for QSAR and virtual screening experiments by means of an unbiased computational protocol. The protein-protein interactions were the first case studies considered by this method, but other unpublished applications have been successfully tested in our laboratory and will be the subject of future publications. In this manuscript we report a new GBPM experience carried out to link structural information of the HIV-1 RT available from the PDB with mutational analysis based on clinical data. The GBPM identification of conserved regions in relevant macromolecules, such as the HIV-1 RT, can be another future study aimed at rational drug design of novel therapeutic agents.

The GRID-based pharmacophore model is created in a 6-step procedure, as shown in Fig. (**1**) [27]. The first one performs the PDB file pre-treatment producing three different model structures: the complex (subunits α+β), the receptor (subunit α) and the ligand (subunit β).

The second step calculates the GRID molecular interaction fields (MIF) [28] with a certain probe onto the three targets above reported.

In the third step an energy comparison of the MIFs is performed by the GRID GRAB utility [28], generating maps with focused information on the interaction areas. The fourth step is related to the identification of most relevant interaction points. With the aim to get a suitable model, these operations should be repeated using at least three different probes: a generic hydrophobic (DRY), an hydrogen bond acceptor (O) and an hydrogen bond donor (N1).

In the fifth step the information obtained from the different probes are unified into a preliminary pharmacophore model.

We carried out the GBPM analysis up to the fifth step of the procedure, in order to compare it with the clinical data and to highlight the most involved residues in the recognition areas.
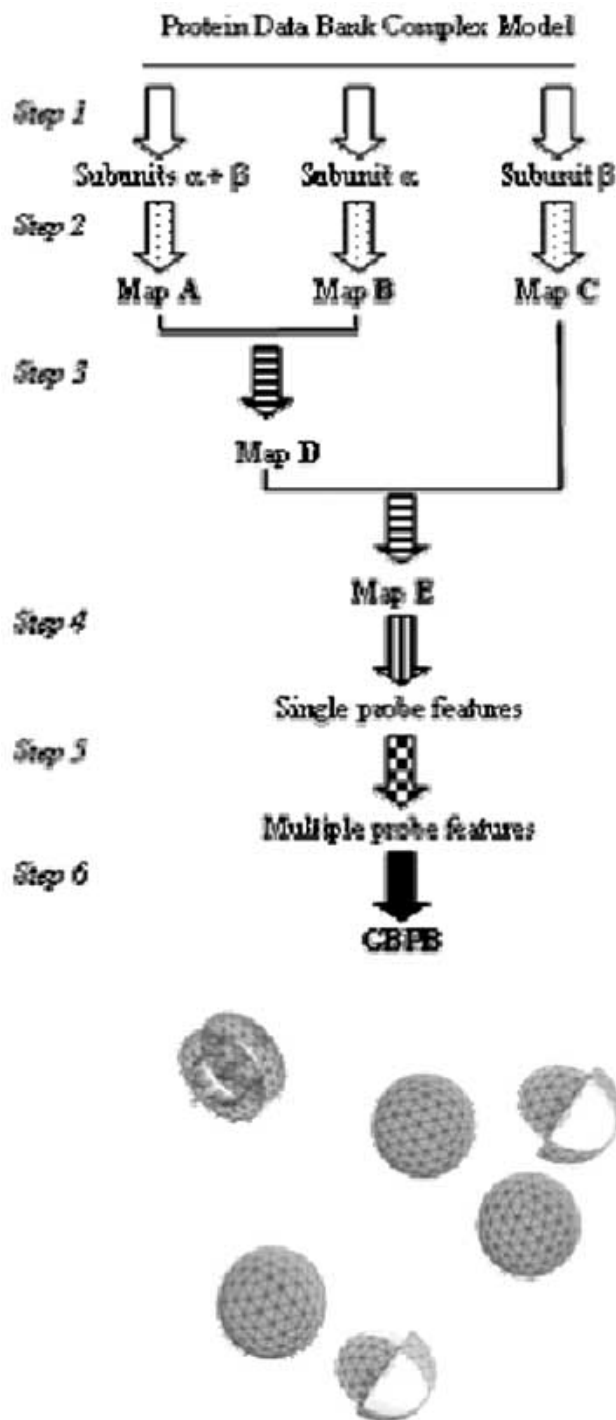

Fig. (1). Flow chart of the six steps of the GBPM method [29].

CLINICAL SEQUENCES ANALYSIS

The analysis included a large number of HIV-1 RT sequences, ranging between a minimum of 221 sequences (with full-length ORF, 560 amino acids) to a maximum of 5576 sequences (with partial RT coverage), derived from unique isolates of HAART treated patients. For the analysis of residues 1-320, the sequences were extracted from the public Stanford Database [30], whereas for the analysis of regions including *thumb*, *connection* and RT/RNase-H subdomains (321-560 residues) the sequences were retrieved both from our laboratory (Tor Vergata University) and Stanford database. The number of the available sequences, analyzed at each RT position, showed these distributions: Pro1-Thr240 amino acids, minimum 4437 maximum 5576 sequences; Val241-Glu300 amino acids, minimum 2551 maximum 4459 sequences; Leu301-Asp320 amino acids, minimum 927 maximum 2099 sequences; Pro321-Gln480 amino acids, minimum 517 maximum 960 sequences; Ala481-Leu560 amino acids, minimum 211 maximum 510 sequences. Consensus B was used as a reference strain for the definition of mutations [31]. All RT sequences were obtained from HIV-1 B infected patients exposed to at least 1-7 NRTIs (among zidovudine, didanosine, zalcitabine, lamivudine, stavudine, abacavir, tenofovir), and 0-3 NNRTIs (among nevirapine, delavirdine, efavirenz).

Sequences having a mixture of wild-type and mutated residues at single positions were considered to have a mutation at that position. When the mixture contained two different mutations, both were considered and reported.

PDB MODELS SELECTION

All RT crystallographic models were downloaded from Protein Data Bank internet site [26]. In order to carry out the superimposition of different PDB models, we used the Maestro program [32] and we calculated the root mean square deviations (RMSD) using both the backbone and the α-carbon atoms. The crystallographic structures were selected considering the resolution of the models and classified on the basis of nucleic acid presence. We chose all PDB models with a maximum resolution value of 3.0 Å and we excluded those crystals that, compared with RT wild type, showed a large number of deletions and mutations.

The structural analysis was carried out using 96 HIV-1 RT crystallographic models, as reported in Table **1**.

The PDB models with the codes 1bqm, 1bqn, 1j5o, 1uwb, 2i5y, 1t03, 1n5y, 1rtd, 1hvu and 1hmv were excluded due to their poor resolution.

Starting from all the available models, we selected 31 PDB structures with a number of deletions lower than 15 on the p66 subunit of which some were co-crystallized with different ligands (Table **2**).

Crystals 1hnv and 1tvr were complexed with the same NNRTI **8** (TIBO) and presented the same resolution and number of deletions. Thus the model chosen (1hnv) had the lowest number of mutations. With respect to the PDB models that were co-crystallized with **10** (nevirapine), the choice of 1s1x was justified since it had a better resolution than 1jlb.

Similarly we have proceeded for the models with **5** (NNRTI HBY), by selecting the best in resolution, i.e. 1hqu and discarding 2ic3.

Among the identified models, some PDB were complexed with a class of NNRTIs (ligands **15**, **7** and **6** Table **2**)

Table 1. **Summary of the HIV-1 RT PDB Models. PDB is the Protein Data Bank Code, Res is the Resolution in Å, Del and Mut are Respectively the Number of Deletions and Mutations in the p66 Subunit with Respect to the Wild-Type Sequence**

PDB	Res	Del	Mut	PDB	Res	Del	Mut	PDB	Res	Del	Mut
1tkz	2.81	43	11	1rt2	2.55	18	11	1c1c	2.50	24	11
1tl3	2.80	43	11	1klm	2.65	23	11	1n6q	3.00	2	15
1tkt	2.60	44	11	3hvt	2.90	5	13	1fko	2.90	18	12
1tl1	2.90	43	11	1bqm	3.10	4	12	1c0t	2.70	70	11
1tkx	2.85	25	11	1bqn	3.30	2	14	1c0u	2.52	22	11
1s1t	2.40	26	12	1tvr	3.00	2	15	1hni	2.80	2	14
1s1u	3.00	35	12	1uwb	3.20	2	14	1dtq	2.80	36	11
1s1v	2.60	26	13	1qe1	2.85	4	15	1dtt	3.00	20	11
1s1w	2.70	35	11	1rev	2.60	36	11	1hnv	3.00	2	14
1s1x	2.80	14	11	2i5j	3.15	8	14	1rt3	3.00	39	11
1jlq	3.00	27	11	1rtj	2.35	17	11	1rt4	2.90	24	11
1jla	2.50	13	11	2b5j	2.90	8	14	1rt6	2.80	22	11
1jlb	3.00	10	11	2ban	2.95	8	14	1rt7	3.00	25	11
1jlc	3.00	25	12	1rth	2.20	17	11	1rt5	2.90	24	11
1jle	2.80	34	11	1rti	3.00	17	11	2be2	2.43	8	14
1jlf	2.60	22	11	1vrt	2.20	35	11	2iaj	2.50	9	15
1jlg	2.60	29	12	1vru	2.40	34	11	1rt1	2.55	22	11
1lwe	2.81	17	11	2ic3	3.00	2	14	1t03	3.10	2	15
1lwc	2.62	53	11	1sv5	2.90	8	15	1n5y	3.10	2	15
1lw2	3.00	41	12	1tv6	2.80	22	14	1eet	2.73	18	14
1lwf	2.80	17	11	2b6a	2.65	2	14	1rtd	3.20	6	15
1jkh	2.50	9	11	1dlo	2.70	4	14	2opp	2.55	41	11
1lw0	2.80	47	11	1t05	3.00	6	15	2opr	2.90	21	11
1s6p	2.90	8	14	1r0a	2.80	2	15	2ops	2.30	24	11
1s6q	3.00	8	14	2hmi	2.80	2	14	1hqe	2.70	6	15
1s9g	2.80	18	14	2hnd	2.50	37	11	2opq	2.80	44	11
1suq	3.00	8	14	2hny	2.50	37	11	1hvu	4.75	6	14
1ep4	2.50	38	11	2hnz	3.00	43	11	1hmv	3.20	24	11
1fkp	2.90	19	12	1hpz	3.00	4	14	1ikv	3.00	4	14
1s9e	2.60	8	14	1hqu	2.70	4	15	1ikw	3.00	3	14
1fk9	2.50	41	11	1hys	3.00	7	14	1ikx	2.80	3	14
1c1b	2.50	24	11	1j5o	3.50	2	15	1iky	3.00	3	15

possessing the same structural scaffold (JANSSEN-R) but different moieties, so we considered the three models (2b5j, 1sv5 and 1s9e) with the best resolution, as examples of this class of drugs.

As far as the unbound enzyme is concerned we included in our study the models with codes 1dlo, 1hqe and 1qe1 for a better resolution relative to the model 1hpz.
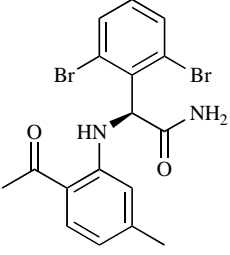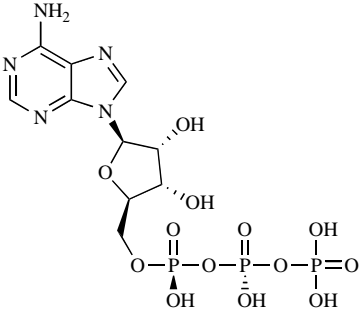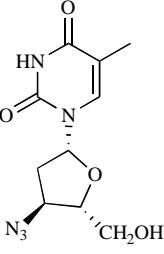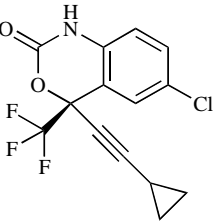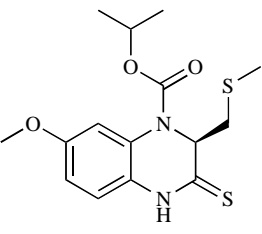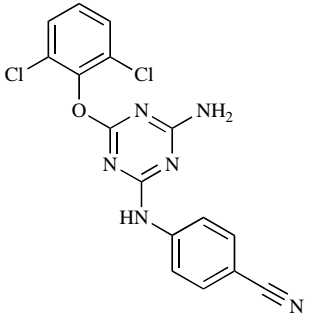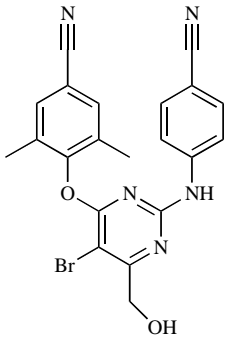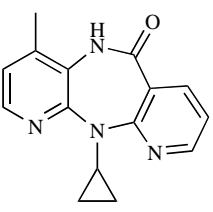
After the PDB selection step driven by the resolution, the number of deletions and mutations, we have divided our models into three groups characterized by similar RT folding properties (Table **3**). They are respectively related to unbound enzyme models (first group), RT complexes with nucleic acid template (second group) and models cocrystallized with NNRTI drugs (third group).
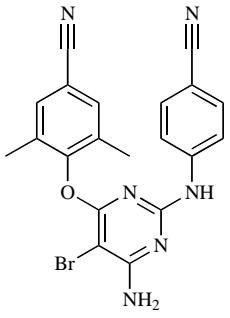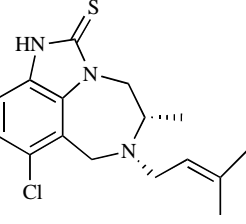
In order to find out the most significant subunit interface interactions for p66-p51, only the protein residues of 20 PDB models were considered in this analysis.

With the aim to observe RT conformational changes as a result of ligand binding, the RMSD analysis was performed with all PDB superimposed models. The maximum value

Table 2.   Chemical Structures of Ligands of the Co-Crystallized HIV-1 RT Complexes Used for the GBPM Analysis

	Ligand		Ligand
	1*		2
	3†		4*
	5*		6*
	7*		8*
	9*		10*

	Ligand		Ligand
	11*		12*
	13†		14*
	15*		16*

*and †indicate, respectively, NNRTI and NRTI ligands.

Table 3.   Crystallographic Models Used for the GBPM Analysis. PDB is the Protein Data Bank Code, Res is the Resolution in Å, Ligand is the Drug Co-Crystallized with the Enzyme, Template is the Nucleic Acid Present in the Model

PDB	Res	Ligand	Template	Case study		
				D	T	N
1hnv	3.00	8	-	*		*
1hni	2.80	1	-	*		*
1jla	2.50	11	-	*		*
2hmi	2.80	-	DNA	*	*	
1n6q	3.00	3	DNA	*	*	
1hys	3.00	-	DNA/RNA	*	*	
1r0a	2.80	-	DNA	*	*	
1t05	3.00	13	DNA	*	*	
1s9e	2.60	6	-	*		*
2be2	2.43	12	-	*		*
2b5j	2.90	15	-	*		*
1ikx	2.80	16	-	*		*
1hqe	2.70	-	-	*		
1jkh	2.50	4	-	*		

(Table 3). Contd…..

PDB	Res	Ligand	Template	Case study		
				D	**T**	**N**
1sv5	2.90	7	-	*		*
1suq	3.00	9	-	*		*
1s1x	2.80	10	-	*		*
1dlo	2.70	-	-	*		
1qe1	2.80	-	-	*		
2iaj	2.50	2	-	*		
1hqu	2.70	5	-	*		*
2b6a	2.65	14	-	*		*
1ikw	3.00	4	-			*

The case study indicates the RT area of GBPM application; D, T and N represent, respectively, the p66-p51 dimer interface, the RT-template region and the NNRTI binding pocket.

was equal to 2.03 Å computed on RT α-carbons, 2.01 Å on the backbone and 2.03 Å on α-carbons in RT-DNA bound models. Superimpositions of RT and DNA for the 5 PDB DNA-complexes highlighted few differences in the folding properties of the enzyme models.

For the second group the GBPM [27] method was applied to RT-template complexes to analyze HIV-1 RT conserved areas mainly involved in interactions with the nucleic acid.

In order to analyze the residues of the NNRTI-BP responsible for the interactions with the drug, the method was applied to 13 crystallographic models where RT was complexed with different inhibitors.

With the aim to select the most involved residues in the recognition step, we applied a geometric criterion using a maximum distance equal to 10 Å from the three analyzed groups' guest (drug, DNA or HIV-1 RT subunit). They were then compared to the identified conserved amino acids for statistical analysis.

The selection and classification of the crystallographic models for the computational study was performed taking into account two different aspects.

The first one was related to the completeness of the available crystallographic models, i.e. number of mutations and deletions with respect to the wild type (WT) sequence.

The second aspect considered was the conformational effect onto the RT. In reality, due to the great heterogeneity of the available co-crystallized PDB models, the enzyme is conformationally sensitive to the presence of nucleic acids and ligands, i.e. inhibitors. So we decided to divide our training set into three sections: a) the unbound enzyme models; b) the RT structures with nucleic acid templates; c) co-crystals of RT and antiretroviral non-nucleoside agents (NNRTI).

In this study GBPM analysis was applied the to the three identified groups of HIV-1 reverse transcriptase models.

In the first analysis, since both subunits of the enzyme are proteins, p66 was considered the host (subunit α) and p51 the guest (subunit β) and vice versa (Fig. **2A**).

In the second case, (Fig. **2B**), in order to find out the RT residues that mostly interact with the nucleic acid, RT-

template complex was used as complex (subunits α+β), the unbound enzyme was the host (subunit α) and the nucleic acid the guest (subunit β).

In the third analysis (Fig. **3**), GBPM procedure was applied to the RT PDB models where the enzyme was complexing different NNRTIs. This study identified the RT residues of the NNRTI binding pocket (NNRTI-BP) able to well recognize the inhibitor.

In order to evaluate the reliability of our approach, a statistical analysis was performed to estimate the degree of conservation for the identified residues with respect to the studied target and the reproducibility of the clinical data.

In the following paragraphs, the three main RT crystallographic groups previously classified will be discussed in detail.

COMPUTATIONAL ANALYSIS

The GBPM method has been described in a recent publication [27]. In the GRID calculations the lone pairs, the tautomeric hydrogen atoms and torsion angles, relative to the sp^3 oxygen atoms and the amide atoms, have been allowed to be settled on the basis of the probe influence, while the coordinates of all the other atoms have been considered rigid (directive MOVE = 0). Default values have been used for the other parameters.

All calculations were performed by a Linux cluster of 8 Intel Xeon dual-processors at 3.2 GHz with 2Gb of RAM. Graphic manipulations and RMSD analyses were carried out by the Maestro Graphical User Interface ver 4.1.012 for Linux operating systems [32]. All 3D figures were obtained using PyMOL graphics and modelling package ver 0.98 [33].

RESULTS AND DISCUSSION

RT Sequence of Clinical Samples

The present study analyzed for the first time the conserved amino acid positions of the complete RT enzyme sequence, by considering all *palm*, *fingers*, *thumb*, *connection* and RNase-H subdomains. The sequences were obtained
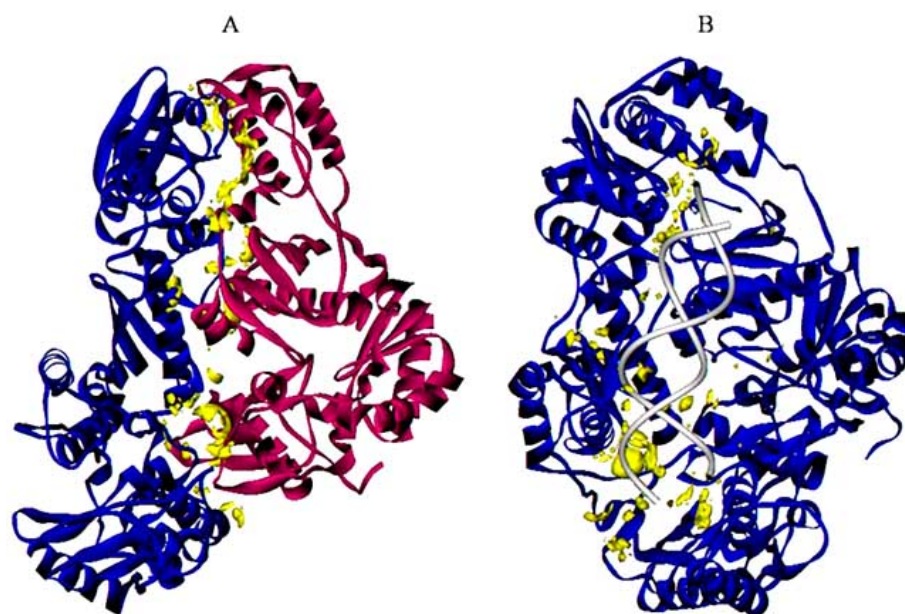
Fig. (2). Examples of GBPM generated maps (solid yellow surfaces) using the DRY probes at: **(A)** the p66 (blue cartoon) / p51 (magenta cartoon) dimerization interface and **(B)** the RT-nucleic acid (gray model) protein (blue cartoon) contact area, respectively obtained with PDB models 2iaj and 2hmi.
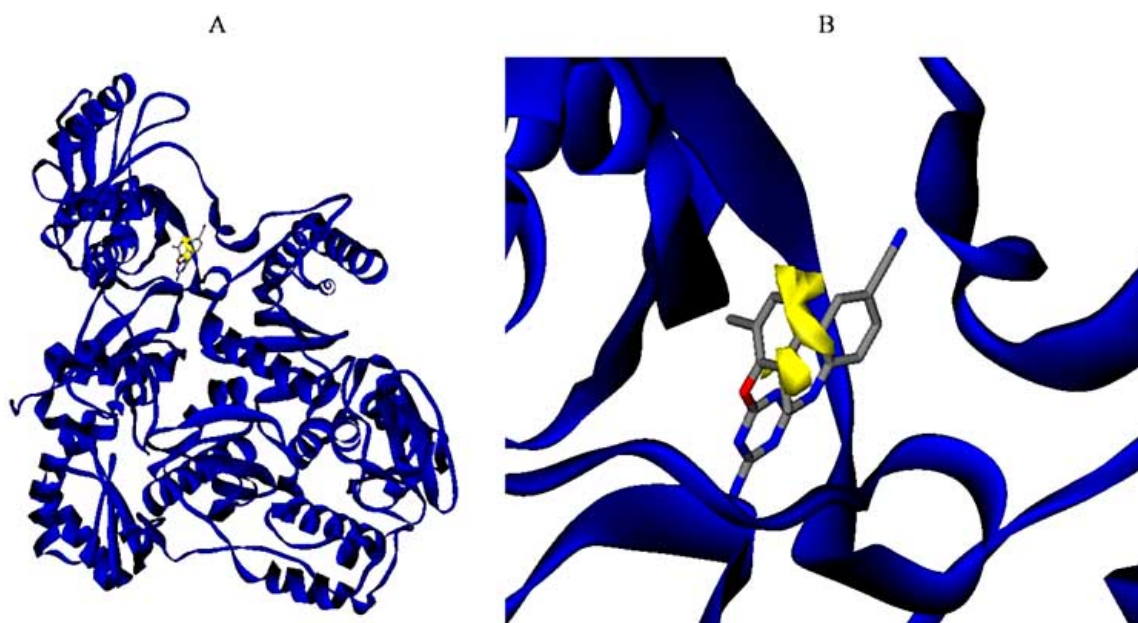
Fig. (3). Representation of GBPM method applied to an RT-NNRTI complexed model. DRY MIFs are displayed in yellow using the PDB model 1s9e. **(A)** RT complexed to the drug displaying all the enzyme; **(B)** The NNRTI binding pocket with the drug (JANSSEN-R129385) and the yellow DRY maps fitting with the aromatic portions.

from a large number of HIV-1 B infected patients (>5500 individuals), all treated with at least one RT inhibitor.

One of our previous studies already characterized the conservation of RT N-terminal domains *in vivo*, analyzing the first 320 amino acids of 2236 RT plasma-derived samples from a well-defined Italian cohort of 1704 HIV-1 infected individuals (457 drug-naïve and 1247 drug-treated) [12]. In naïve patients, 233 out of these 320 residues (73%) were conserved (<1% variability). In patients treated with RT inhibitors, despite the significant drug pressure and the appearance of mutations primarily associated with resistance,

202 amino acids (63%) remained highly conserved, and appeared mostly distributed in regions of variable length [12]. Similarly, another recent study examined approximately 550 RT C-terminal sequences (residues 298-560, in the *connection* and RNase-H subdomains), in both drug-naïve and NRTI-experienced patients infected with different HIV-1 subtypes [34]. In this study, the sequences were retrieved from both Brazilian clinical viral isolates and public HIV sequence databases. It has been shown that one third of invariable sites was common to both groups of patients. Moreover, the 61% of conservation in the sequences of NRTI-experienced patients was observed.

Amino acids not mutated or mutated with a prevalence ≤1% were defined as conserved. HIV-1 RT depictions (Fig. **4**) were colored according to mutation frequency rate, so that blue corresponds to highly conserved residues (variability in ≤1% of patients), cyan to residues mutated only in 1-5% of patients, green to amino acids mutated in 5-10% of patients, yellow to residues mutated in 10-25% of patients, and red to amino acids that appear mutated in >25% of patients.

The current analysis of sequences, retrieved both from the public Stanford Database and our laboratory, showed that 338 out of 560 RT residues (60.4%) were conserved (≤1% variability). Considering the first 320 amino acids of RT, the conservation is about 63.4%; such a value exactly corresponds to our previous observation [12]. The remaining C-terminal residues (321-560) highlight a conservation of 56.5%. The RNase-H subdomain (codons 441-560) shows a lower level of conservation (54.2%) in comparison to the whole protein and to other regions such as the *connection* subdomain (codons 298-440, 59.2% of conservation), or the *thumb* subdomain (codons 238-318, 62.96% of conservation).

Several invariant residues clustered into large defined regions containing functionally important residues, scattered throughout different areas of the viral protein (Fig. **4**). The ten longest invariant areas, comprising between 7 and 20 consecutive invariant amino acids, were: I (Leu12-Pro19), II (Gln91-Pro97), III (Phe124-Ser134), IV (Arg143-Pro150), V (Trp229-Asp237), VI (Trp252-Tyr271), VII (Leu349-Ala355), VIII (Trp406-Glu415), IX (Val417-Glu430), X (Val493-Ala502).

These areas are located in the following RT domains: I, III and IV in the *fingers* subdomain, II in the *palm*, V and VI in the *thumb*, VII, VIII and IX in the *connection* and finally the area X in the RNase-H domain.

The areas VIII and IX are separated by only one residue, Phe416, which shows a low variability (<5%) suggesting that this large area of 25 amino acids should have an important function. Interestingly, the areas I, II, III, IV, V and VI were found invariant already in our previous work [12], and, similarly the areas VIII and IX were found invariant in the Brazilian study [34].

In addition to the ten large highly conserved described areas, other invariant residues were scattered throughout the sequence either individually or in small stretches of four to six amino acids: Thr7-Val10, Lys22-Leu26, Asn54-Pro59, Arg78-Lys82, Thr84-Glu89, Leu109-Ala114 (including the catalytic Asp110), Gly152-Pro157, Ser191-Ile195, Trp239-Gln242, Glu305-Leu310, Pro313-Gly316, His361-Asp364, Tyr441-Ala445, Thr472-Gln475, Gly541-Glu546.

Heterodimer Interface Area Analysis

GBPM method was applied onto 20 HIV-1 RT crystallographic models not complexed to inhibitors (Table **3**). The aim was to identify residues mainly involved in heterodimer stabilization. We always adopted the DRY, N1 and O probes mimicking the hydrophobic, the hydrogen bond donor and acceptor, respectively.

The first conserved RT area (Leu12-Pro19), whose functional role is still uncharacterized, was well recognized in both p66 and p51 subunits by all used probes (Supplementary Material Fig. **S1A**).

DRY MIFs highlighted the aromatic residue Trp88 (Fig. **5A**), located onto the β5a strand. Such a residue was already found to be relevant in hydrophobic contacts at the dimer interface, emphasizing its pivotal role in primer/template binding [35]. Trp88 is able to establish interactions with Arg143 of p51, which is strongly recognized by O probe when p51 subunit was the host and considered important for the stability of the entire protein (data not shown).

Also the second conserved RT area (Gln91-Pro97) was identified by our computational study. By analyzing the O probe MIFs (Supplementary Material Fig. **S1B**), two residues of p66, Pro95 and His96, mainly involved in interactions with the β7-β8 loop (Pro133-Pro140), were found to be crucial in stabilizing the RT heterodimeric form. This observation can be explained by the ability of these residues to produce hydrogen bonds (p66 His96/p51 Asn136) [13].

The area defined by RT residues Phe124-Ser134 (area III), located in the *fingers* subdomain, was well recognized when p51 was the host (Supplementary Material Fig. **S2A**). Notably the invariant Phe130 is involved in protein stability. In fact, although more exposed, all Phe130 side-chain atoms establish van der Waals contacts with carbon atoms of other hydrophobic side chains of the p66 subunit, contributing to the overall stability of RT [36].

The fourth conserved area, placed in the *palm* subdomain, was identified using p66 as host. In particular Ser156, a functional residue that is conserved but buried, resulted well defined by O and N1 probes, emphasizing its indirect role in RT function for determining the proper conformation of the β8-αE loop in the p66 subunit (Supplementary Material Fig. **S2B**) [36].

Among all conserved amino acids identified by the DRY probe, Gln182 and Tyr183 were highlighted (Supplementary Material Fig. **S3**), confirming their productive role in RT polymerase activity [6, 37]. Tyr183, which is part of the Tyr183-Met184-Asp185-Asp186 (YMDD) motif, was found to be pivotal in the catalytic process, due to its hydrophobic interactions with the primer strand and its hydrogen bond with the nucleic acid cytosine O2 atom. The mutation of Tyr183, well indicated by DRY probe, with non aromatic residues induces a remarkable decrease of the polymerase activity, confirming its contribution in hydrophobic interactions [38].

The heterodimerization is a two-step process. The first one involves a concentration-dependent association of the two subunits, followed by a slow isomerization/maturation step that involves p51 *fingers* and *thumb* interactions with the *palm* and RNase H of p66, respectively. Specifically, residues Pro52-Pro55 and Ile135-Pro140 of the p51 *fingers* appear to interact with residues Gln85-His96 of the p66 *palm*, while residues Cys280-Leu295 of the p51 *thumb* subdomain appear to contact residues Val536-Asn545 of RNase H. In the first step of RT dimerization p51 and p66 hydrophobic interactions have been found to be crucial. This well conserved motif is believed to include residues Trp401-Trp410 of p66 and Pro392-Trp401 of p51. The typical
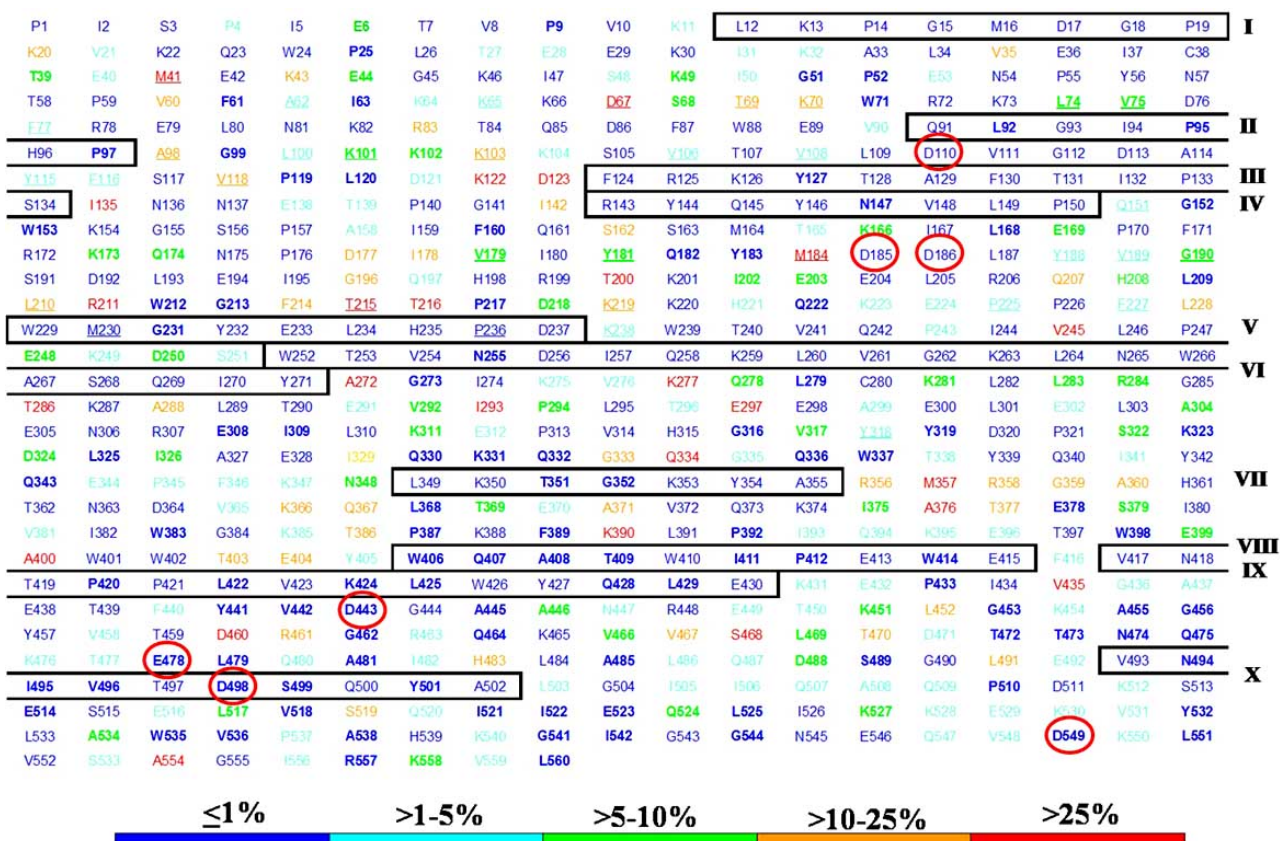
Fig. (4). Conserved regions of HIV-1 RT in infected patients. The sequence of the 560 amino acids of HIV-1 RT of clade B consensus (shown as a reference) is colored according to the frequency rate of mutations observed in a minimum of 221 to a maximum of 5631 plasma samples obtained from RTI-treated patients. The long invariant regions are boxed (I to X).
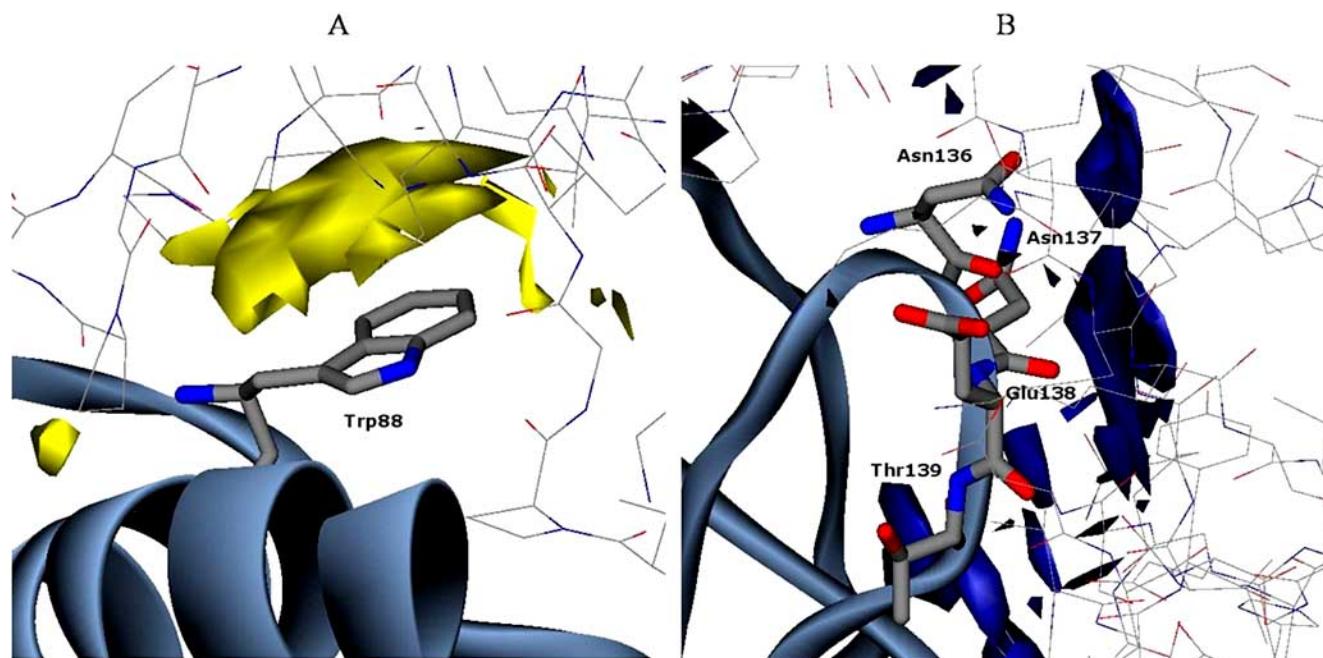
Fig. (5). MIFs obtained with PDB model 1hnv: (**A**) Interaction areas of p66 Trp88 mapped by DRY probe (yellow area) in presence of the p51 subunit. The p66 subunit is represented as light blue solid ribbon and p51 residues are displayed in wire. Trp88 is shown as CPK colored sticks. (**B**) Interaction areas of p51 β7-β8 loop mapped by N1 probe (blue areas) in presence of the p66 subunit. The p51 subunit is reported as light blue solid ribbon and p66 residues are displayed in wire. β7-β8 loop is represented as CPK colored sticks.

connection subdomain has six tryptophan residues and a tyrosine between amino acid residues Trp398-Trp414 [9].

The conserved residues Pro52-Pro55 of p51 subunit resulted well identified by all probes (Supplementary Material Fig. **S4A**), according to their pivotal role in interactions to the p66 *thumb* subdomain. In fact, the opposite highly conserved sequence Gln85-Gly93, located in the template grip of the *palm* subdomain, was well recognized by all used probes (Supplementary Material Fig. **S4B**). This is in agreement with their essential role in interactions with the DNA sugar-phosphate backbone of the template strand [12] and in contacts with the p51 *fingers* [9].

Due to its ability to donate/accept hydrogen bonds, β7-β8 loop of p51 was well mapped, in particular by using the N1 probe, that is able to well recognize residues Asn136, Asn137, Glu138, Thr139 and Pro140. These p51 residues are able to make contacts with p66, as confirmed by site-specific mutagenesis studies [13]. Asn136 of p51 is able to establish hydrogen bonds with His96 and Pro95 of p66 and this interaction was detected by the opposite probe properties when each subunit was considered as host. As shown in Fig. (**5B**), Asn136 was well recognized by N1 probe using p51 as host. This can be considered as a typical complementary example between subunits and opposite probes, because it is related to the same previously reported interaction found with the O probe using p66 as host (Supplementary Material Fig. **S1B**).

The β7-β8 loop of p51 was found to interact with other residues of p66, such as Tyr181, Gln182 and Tyr183, essential for RT polymerase activity and for the formation of the NNRTI binding pocket [39, 40].

No matter is the host or the guest, DRY MIFs always identified a large hydrophobic area defined "tryptophan rich motif", consisting of six tryptophan residues placed at codons 398, 401, 402, 406, 410 and 414, crucial in RT dimerization [9]. GBPM recognized five of six tryptophan residues, including the two pivotal Trp401 and Trp414 of p66 subunit, that, if mutated, cause the loss of RT activity (Fig. **6**).

The perfect mapping result obtained with Trp401 of p51 is displayed in the Supplementary Material Fig. **S5**.

A second area involved in heterodimer formation, crucial in protein-protein interactions, was represented by a repeating motif of eight leucine residues (282-310). Leu289 of p51 resulted well recognized by DRY probe (Supplementary Material Fig. **S6**) through an hydrophobic interaction with Phe61 of p66 and was found to be essential for the stability of the single subunits [41].

Located on the RNase H domain of the enzyme, the two conserved glutamic residues at position 432 and 438 were identified by the N1 probe (Supplementary Material Fig. **S7A**). Their mutation impaired the *in vivo* p66 processing. Glu438 especially has been reported to play a major role in proper heterodimerization and function of virion-associated RT [42]. Polymerase and RNase H activities are also assured by two pivotal proline residues of the p51 subunit, Pro25 and Pro52 [2, 4], well highlighted by the hydrophobic probe (Supplementary Material Figs. **S7B** and **S4A**).

The largest conserved area defined by RT amino acids Trp252-Tyr271 (area VI) was fully identified by all adopted

probes using p51 as host. Part of the *thumb* subdomain, the αH region (Val254 to Ser268) is in contact with the RNase H domain of p66, suggesting a remarkable contribution of this region to the heterodimer stabilization [12]. GBPM analysis agreed as to the important role played by conserved residues Asn255, Gln258, Lys259, Val261, Gly262 and Asn265, well recognized by all probes. In particular O maps put in evidence the highly conserved residue Asn255 [3]. As reported in Supplementary Material Fig. **S8A**, Asn255 of p51, mapped by carbonylic oxygen, was close to Glu432 of p66, well identified by N1 (Supplementary Material Fig. **S7A**), confirming the map complementarities.

Always located on the *thumb* subdomain of p51, residues Cys280, Gly285, Lys287, Leu289 and Thr290 were well recognized by two probes (Supplementary Material Fig. **S8B**), in agreement with their ability to establish favourable interactions with residues Val536, Pro537, His539 and Gly541-Asn545 of RNase H region [9].

Visual inspection of N1 and O MIFs with most of the analyzed PDB models indicated that Lys331 is statistically significant as a "single residue" in pattern interaction (data not shown). This amino acid contributes to RT heterodimerization through hydrogen bonds with the backbone of residues Trp402, Tyr405 and Gln407 of p66, according to the recent mutagenesis analysis showing that mutations at this codon can negatively influence the dimerization step [43].

Also the slightly variable Lys395 and Glu396, located into the *connection* subdomain, resulted essential for primer-template binding due to hydrogen bonding interactions. Both residues are referred to the RNase H primer grip and lie either in the p51 or p66 subunit of RT. *In vivo* and *in vitro* studies of the RNase H primer grip indicate that these residues are important for the proper binding and positioning of the nucleic acid to RT. Lys395, well recognized by the N1 probe, resulted crucial to form hydrogen bonds and other interactions to the primer strand; in fact its substitutions strongly reduced the viral replicative capability with respect to the WT.

According to mutagenesis data, Glu396 was well mapped by N1 and O probes (Supplementary Material Fig. **S9**), suggesting that single Glu396Ala mutation causes HIV-1 polymerase failure [43].

Mutational analysis of HIV-1 reverse transcriptase revealed two adjacent residues of p51 subunit located at positions 419 and 420 as crucial for RT folding and activity. It was observed that the double mutant 419/420 was associated to the loss of RNase H activity, because these mutations are positioned close to the critical "hinge" region, connecting the putative DNA polymerase and the RNase H domains [44]. In agreement with these experimental data, both highly conserved residues resulted well identified by DRY and N1 probes (Supplementary Material Fig. **S10**), confirming their pivotal role in the interactions to the "tryptophan rich motif" of p66 subunit and to the hydrophobic amino acids Leu503 and Trp535 placed at the dimer interface.

The large conserved region located into the RNase H domain (area X Val493-Ala502) was well identified by all used probes. In particular amino acids Asp498, Ser499,
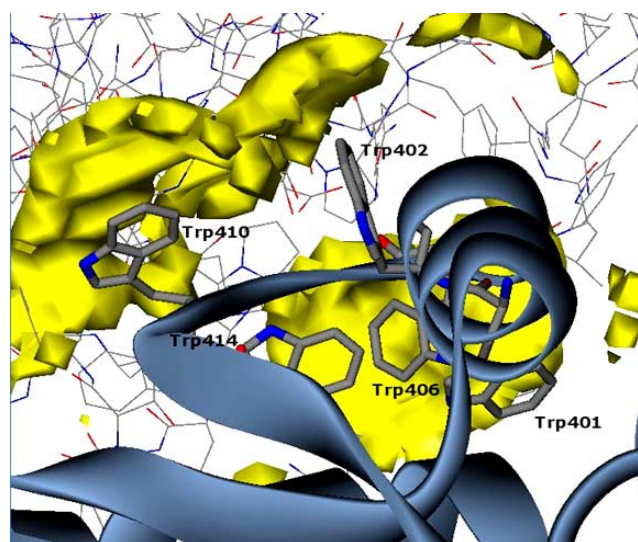
Fig. (6). Interaction areas of p66 "tryptophan rich motif" (CPK colored sticks) well recognized by DRY probe (yellow areas) using the PDB model 1hnv. The p66 subunit is represented as light blue solid ribbon and p51 residues are displayed in wire.

His539, Asn545 and Asp549, known to comprise the RNase H catalytic site, were highlighted [34].

Part of the RNase primer grip, the slightly variable Ile505 resulted well recognized by DRY probe, according to its pivotal role in DNA primer strand positioning near the active site and setting the proper trajectory for the RNA/DNA hybrid incoming (Supplementary Material Fig. **S11A**) [45].

Moreover Gln500 and His539, involved respectively in ribose and phosphate contacts to the RNA primer [46], were well recognized by DRY probe (Supplementary Material Fig. **S11B**). It was demonstrated that His539Asn mutation is responsible for the reduction of RNase H activity and for the increase in zidovudine (**2**) resistance with respect to the WT due to a limited RNA degradation. Consequently there was an increase of the time for **2**-monophosphate to be excised from the terminated primer and for polymerization to resume on an intact template [47].

RT-DNA Complexes Analysis

In order to evaluate the most involved interaction areas between RT and nucleic acid templates, the GBPM method was applied to five complexed RT-template PDB models (Table **4**) using the same probes of the first analysis and taking into account the hydrophilic nature of the nucleic acid.

In this analysis GBPM maps identified several isolated residues mostly involved in interactions with the nucleic acid, such as Trp24, Pro25, Phe61, Ile63, Asp76, Arg78, Ser105-Ser117, Tyr183, Asp185, Asp186, Cys280, Gly285, Leu289, His361 and Arg448.

The DNA recognition showed the p66 conserved areas II, IV, V, VI, VII and X; in the p51 subunit the only identified conserved region was the eighth.

Some of these amino acids were well recognized within the minor groove of DNA, as reported in Fig. (**7A**).
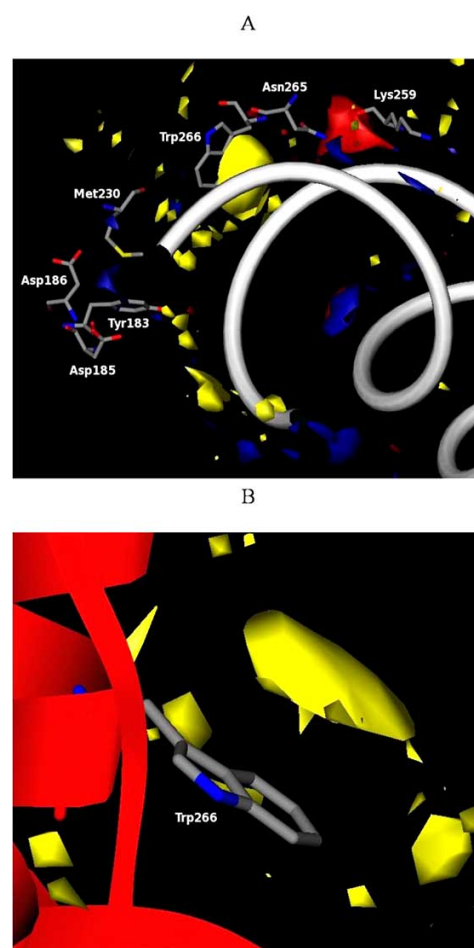
Fig. (7). (**A**) Minor groove DNA recognition with some highly conserved residues mapped with three different probes (PDB model 2hmi). In yellow are reported DRY maps, in blue N1 maps and in red O maps and the most important residues are shown in CPK colored sticks. (**B**) Insights of the p66 Trp266 (CPK colored stick model) well mapped by DRY probe (yellow area). The p66 subunit is represented as red flat ribbon.
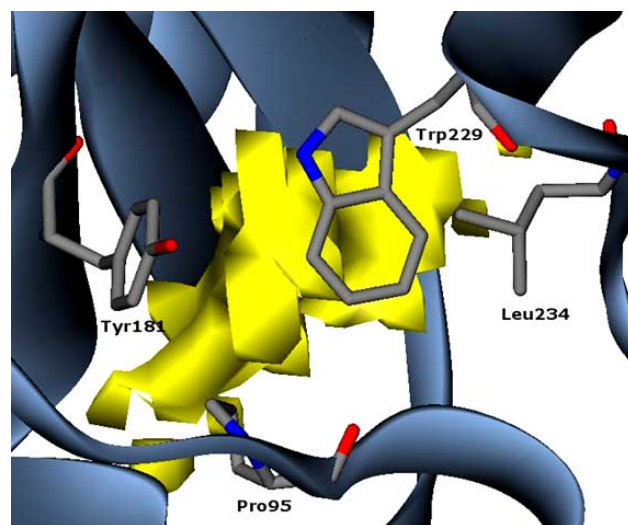
Fig. (8). Interaction areas of Pro95, Tyr181, Trp229 and Leu234 of p66 well recognized by DRY probe (yellow areas) using the PDB model 1s9e. The p66 subunit is reported as light blue solid ribbon and the identified residues are represented as CPK colored sticks.

Located on the *fingers* subdomain of p51, the aromatic amino acids Trp24 and Phe61, involved in RT stabilization, were well defined by DRY probe (Supplementary Material Fig. **S12A**). This provides evidence for their role in hydrophobic interactions and in fact selective mutation of these residues to a glycine dramatically alters the stability of the RT-heterodimer [48].

Placed on p66 *fingers* subdomain, amino acids Asp76 and Arg78 were well recognized using the N1 probe (Supplementary Material Fig. **S12B**), according to their important role in polymerase activity [12].

DRY mapping also allowed to identify the conserved Trp229 (area V), shared by the NNRTI binding pocket (NNRTI-BP) and by the primer grip region (Supplementary Material Fig. **S13**). Mutagenesis studies confirmed the crucial role of Trp229 aromatic ring for RT activity, correct folding of the protein and NNRTI-BP formation [49].

Well defined by the hydrogen bonds donor probe N1 (data not shown), residues Gln91-Ile94 were recognized within the template grip region. Their importance in positioning the template strand near the polymerase active site has been reported in the literature [50]. In particular, located in the *palm* subdomain of the p66 subunit, the highly conserved Ile94 was well recognized by all used probes, suggesting its relevance in interacting with the 2'-OH of RNA strand [51].

In the area from Gly152 to Pro157, the conserved residues 152-155 were well recognized by N1 probe (Supplementary Material Fig. **S14**). Gly152-Gly155 have an essential role in DNA interactions and RNA hydrogen bonding network [51]. In particular Trp153 is located next to the conserved Leu149-Pro150-Gln151-Gly152 (LPQG) motif, at the juncture of the *palm* and *fingers* subdomains. Ala substitutions at this codon substantially hindered the polymerase function of the enzyme, impacting the geometry of the dNTP binding pocket [52]. Pro157, that constitutes the N-terminus of the αE of the *palm* subdomain, was found to be pivotal in maintaining the structural integrity of β8-αE connecting loop and the appropriate conformation of the template grip [53].

Asp110, Tyr183, Asp185 and Asp186 were well defined by the DRY and N1 probes (Supplementary Material Fig. **S15**), according to the crucial interactions of these amino acids with DNA phosphates and nucleotide bases. In fact the aspartic residues side chains establish hydrogen bonds with the 3' hydroxyl group, facilitating the nucleophilic attack during the polymerization step. On the other hand Tyr183 was found to be pivotal in the catalytic process, due to its ability to establish hydrophobic interactions with the primer strand [38].

YMDD motif conformation is highly influenced by the structure of the primer grip region, formed by β12-β13 sheets in the p66 *palm* subdomain, which is in close proximity to nucleotides at the 3' terminus. This structural element is involved in template translocation. Mutagenesis experiments demonstrated that substitution of alanine residues in p66 β12-β13 hairpin (Trp229, Met230, Gly231 and Tyr232) alters both DNA polymerase and RNase H activities [6]. GBPM maps highlighted the importance of the nucleic acid interactions with Met230 and Gly231 (Supplementary Material Fig. **S16A**). Other studies showed that mutations of Tyr232, well recognized by the N1 probe, induce an incorrect RNase H cleavage due to the crucial role of this aromatic amino acid in π-π interactions with the template primer and in DNA-dependent DNA polymerase function (Supplementary Material Fig. **S16B**) [54].

Located on the polymerase domain of the *thumb* subdomain (area VI), residues Gln258, Lys259, Gly262 and Lys263 were well recognized in all maps (Supplementary Material Fig. **S17**). It has been shown by alanine-scanning mutagenesis that these amino acids, that are part of the helix clamp, resulted critical for template-primer affinity [55].

Actually cross-linking of DNA to a Gln258Cys mutated RT in both subunits decreased the stability of the complex and caused the dissociation of the p51 subunit from the cross-linked p66-DNA complex; this event is due to the disruption of the interactions between p51 Gln258 side chain (*thumb* subdomain) and the RNase H subdomain [56].

Further site-specific mutagenesis studies demonstrated that some amino acids of the sixth area, such as Asn255 and Trp266, highly conserved in treated patients [12], resulted essential for enzyme stability and catalytic activity [57]. These residues, located in p66 conserved regions, were found to interact with DNA terminal sequence in the minor groove by means of hydrophobic contacts. This finding is in agreement with our structural data, thus validating the reliability of the GBPM method and underlining Trp266 relevance in interactions with the primer-template. This residue, whose pivotal role was confirmed by mutagenesis analyses, was well indicated by DRY probe (Fig. **7B**) and is part of the Minor Groove Binding Track (MGBT) region, crucial to establish contacts with the DNA in the minor groove. The loss of these interactions is associated with a remarkable decreasing in DNA binding affinity and in frame shift process [57].

Located at the *connection* subdomain, two highly conserved residues of the p51 subunit, His361 and Thr362, can interact with the phosphate backbone of the DNA primer strand near to RNase H cleavage site. A mutagenesis analysis demonstrated that most of the mutations in the *connection* domain have relatively small effects on the virus titer; however the His361Ala and Thr362Ala substitutions decreased the titer with respect to the WT virus. Moreover the His361Ala mutation also had a significant effect on RNase H cleavage, due to the histidine interactions with the nucleic acid through the p66 subunit [7]. In agreement with these data, both His and Thr analyzed residues were well recognized by N1 probe, confirming their ability to establish hydrogen bonds with the template (Supplementary Material Fig. **S18**).

The RNase H domain is folded into a mixed 5-stranded β-sheet flanked by 4 asymmetric α-helices; the strong structure conservation suggests that this region is very stable. The well recognized and crucial Asp498 residue is located in the active site and is implicated in catalysis. Replacement of this residue generates an unstable enzyme [44].

In this same subdomain, Arg448 and Gln500, responsible to establish crucial interactions with base, sugar and phos-

phates of RNA primer [34], were well recognized in all obtained maps.

RNase H primer grip residues Thr473-Lys476, Tyr501 and Ile505 were identified by all used probes (data not shown), in agreement with some mutagenesis data regarding their essential role in polymerization-independent RNase H activity, dimerization and DNA strand transfer [58].

NNRTI Binding Pocket Analysis

GBPM approach was finally applied to the 13 selected RT models co-crystallized with non-nucleoside drugs (Table **5**). In all analyzed models, the most involved residues in NNRTI-BP recognition resulted: Ile94-Pro97 (area II), Gly99-Leu109, Val179, Ile180, Tyr181, Gln182, Tyr183, Asp186-Val189, Gly190, Ser191, Pro225, Phe227, Trp229-Trp237 (area V), Tyr318-Pro321 of the p66 subunit and the β7-β8 loop of the p51 subunit. Most of these amino acids are reported in a recent manuscript showing the binding pocket for the new NNRTI CP-94,707 [59] and the RT residues mainly involved in drug stabilization.

Previous conservation studies highlighted the crucial role of residues Pro95, Tyr181, Phe227, Trp229, Leu234 (Fig. **8**) and Tyr318 (Supplementary Material Fig. **S19**) in catalytic activity after pharmacological treatment with NNRTIs [12].

The first RT conserved area was well defined by our computational approach with the DRY recognition of heterocyclic rings and hydrophobic side chains [12]. A site-specific mutagenesis analysis showed that the substitution Pro95Ala involves a drastic decreasing of RT RNA-dependent DNA polymerase function due to the lack of the hydrogen bonding between Pro95 of p66 and Asn137 of p51, with consequent destabilization of the dimerization effect [13].

Located in the β5b-β6 loop of the *palm* subdomain, the conserved residues Pro97 (Supplementary Material Fig. **S20A**), Gly99 and Ser105 resulted well indicated by DRY, O, and N1 probes (data not shown), respectively, confirming their essential role in polymerase activity and in hydrophobic contacts with NNRTIs [60, 61].

As reported in literature [18], Leu100Ile and Val106Ala mutations cause a decreasing of side-chains length and a distortion of the binding pocket region with a complete loss of protein-drug contacts. In agreement with these data, the highly conserved Leu100 and the slightly variable Val106 were well recognized by the DRY probe, emphasizing their crucial role in hydrophobic interactions with NNRTIs (Supplementary Material Fig. **S20B**).

Also the variable Lys101 and Lys102 are involved in direct and indirect contacts with non-nucleoside drugs. Actually these inhibitors, through water molecules, are able to establish hydrogen bonds with the Lys101carbonylic group [62], highlighted in our GBPM models by N1 and O probes (Supplementary Material Fig. **S21A**), and to the carbonylic and amide groups of Lys102 [63], which were all well recognized by N1 probe (Supplementary Material Fig. **S21B**).

In an external region of NNRTI-BP is located the highly variable Lys103, responsible to stabilize NNRTI into the pocket through an hydrogen bonding. N1 and O maps agreed with this observation (Supplementary Material Fig. **S21A**),

even in presence of Lys103Asn mutation, that induces an hydrogen bond between the phenol hydroxyl of Tyr188 and the carboxamide side-chain of Asn103. Consequently this hampers the reorientation of Tyr181 and Tyr188 towards the polymerase active site, thereby increasing the stabilization of the closed-pocket form of this mutant enzyme [18, 64].

Non-nucleoside drugs, in spite of their structural heterogeneity, have in common a remarkable lipophilic nature and interact with residues Tyr181, Tyr188, Pro225, Phe227, Trp229 and Pro236 through hydrophobic contacts. NNRTI binding to RT causes the side chains of both Tyr181 and Tyr188 to rotate away from their positions by twisting about 30° in the hydrophobic core, thereby creating a space for accommodating the ligand [18]. The hydrophobic interactions stabilize the enzyme-inhibitor complex which takes place on the aromatic side chains of Tyr181 and Tyr188 and the π-electron donor groups of the NNRTIs. In agreement with such an observation, these crucial residues were well recognized by DRY mapping (Supplementary Material Fig. **S22A**). Moreover, as reported in literature, the single mutations Tyr181Cys and Tyr188Cys were associated to resistance to first generation NNRTIs, and to nevirapine and efavirenz, respectively, due to the loss of the aromatic ring [20, 65].

Part of the YMDD motif, the highly conserved Tyr183 was well defined using DRY probe, according to its pivotal role in the interactions to the drugs. In fact some mutagenesis studies showed that Tyr183Ser mutation resulted in a 99% reduction of the polymerase activity, and a subsequent analysis demonstrated that this mutation caused a 77-fold decrease in the affinity for dNTP substrates (Supplementary Material Fig. **S22B**) [66]. More recent data have shown that Tyr183Phe mutation increases the fidelity of HIV-1 RT [38, 67], and that Tyr183Ala mutant is active on DNA templates but not on RNA templates [68].

Also the variable Val179 is part of NNRTI-BP and is placed at dimerization interface. This residue is involved in an hydrogen bond with Lys103 and facilitates the incoming NNRTI into the hydrophobic pocket [69]. Analyzing N1 maps, Val179 and Gly190, located in binding pocket β-sheet [70], were well recognized, remarking their relevance in interactions with the drugs (Supplementary Material Fig. **S23A**). Amino acid substitutions at RT codon 190 are frequently selected during HAART regimens and mutagenesis data confirmed that an high-level resistance to nevirapine and moderate resistance to both stavudine and zidovudine are associated with Gly190Ser/Ala/Glu mutations [71].

Since the NNRTI-binding pocket of HIV-1 RT includes amino acids Leu100-Asp110, Ile180-Gly190 and Lys220-Thr240, mutations in these regions can substantially decrease susceptibility to all NNRTIs [72]. In particular N1 and DRY mappings defined the highly conserved and hydrophobic residues Ile180 and Val189, highlighting their importance in favourable contacts within the NNRTI binding pocket (Supplementary Material Fig. **S23B**) [12, 60].

Pro225 is one of the amino acids forming the entrance rim of the hydrophobic NNRTI-BP, and it is highly conserved in the RTs of HIV-1, HIV-2 and simian immunodeficiency virus (SIV) [73]. This amino acid was well recog-

nized by DRY probe (Supplementary Material Fig. **S24A**) according to its essential role in lipophilic interactions to the drugs. Indeed Pro225His mutation in the recombinant HIV-1 resulted in a 2±6-fold decreases of sensitivity to all NNRTIs tested, except for BHAP U-90152, where a 2±5-fold increase in sensitivity was observed [72].

Reverse transcriptase residues from Glu224 to His235 are known to be involved in the primer terminus positioning ("primer grip") [74]. The aromatic and conserved amino acids Phe227 and Trp229, well identified by DRY probe, are essential in NNRTI-BP due to their ability to establish π-stacking interactions to the drugs [59]. Mutagenesis studies confirmed the crucial role of Trp229 aromatic ring for RT activity, correct folding of the protein and NNRTI-BP formation [48]. In the primer grip area, also Leu234 resulted well defined by DRY probe (Supplementary Material Fig. **S24B**).

This residue is known to be essential in RT dimerization step as much as the substitution Leu234Ala inhibits p66/p51 association and induces long range conformational changes, consequently altering the NNRTI-BP structure [8].

Several studies showed that His235 and Pro236, through their heterocyclic rings, result crucial to establish, respectively, hydrophobic interactions and hydrogen bonds to NNRTIs and to stabilize the open conformation of the binding pocket [59, 62]. In agreement with these data, Pro236 resulted well recognized by N1 probe (Supplementary Material Fig. **S19**).

N1 molecular interaction field highlighted the slightly conserved residue Tyr318, known to interact with NNRTI-BP through the phenolic hydroxyl group of its side-chain. Actually recent mutagenesis studies confirmed the pivotal role of this amino acid, whose substitution with a non-aromatic residue causes the lack of RT activity and the development of NNRTI associated resistance [75].

RT p51 subunit is known to be involved in heterodimer stabilization and in the formation of the hydrophobic pocket at which the NNRTIs are targeted. Also in NNRTI recognition the β7-β8 loop was well indicated by all used probes, according to its functional and structural roles in NNRTI-BP stabilization. In particular Glu138, essential for RT sensitivity and resistance to the inhibitory effect of the TSAO derivatives, was well defined by O probe (data not shown) due to its capability to establish hydrogen bonds with the drug [76].

CONCLUSIONS

In summary, in this review we have reproduced most of the mutational observations using the new GBPM approach. This computational method is outstanding since it was successfully applied to the RT-template, NNRTI recognition and heterodimer stabilization.

In order to obtain a statistical analysis, the most involved residues in the recognition step were selected and compared with respect to the identified conserved amino acids. In the heterodimer analysis we have identified among the conserved reference residues 93% (154 of 166) and 90% (118 of 130), respectively using p66 and p51 as host. In the DNA

complexes, GBPM allowed to identify 55 conserved residues out of a possible 74 (74%). Furthermore, the same degree of conservation was also obtained in NNRTI recognition. Finally, the majority of invariant amino acids clustered into defined regions comprising small stretches of four to six amino acids as well as in large areas containing from seven to twenty-nine consecutive residues. This finding suggests that participation of consecutive HIV-1 RT residues in structural domains is strictly required for cooperative functions and sustainable enzyme activity. Asides from confirming the conservation of amino acids that are already known to be important for catalytic activity, stability of the heterodimer interface, and/or primer/template binding, other new invariable residues are now identified and mapped onto the three-dimensional structure of the enzyme. These new results could assist in the rational design of new HIV-1 inhibitors with alternative mechanisms of action and more favorable resistance profiles.

ABBREVIATIONS

AZT	=	Zidovudine
dNTP	=	deoxy-nucleoside triphosphate
GBPM	=	Grid Based Pharmacophore Model
HAART	=	Highly Active Antiretroviral Therapy
HIV	=	Human Immunodeficiency Virus
MGBT	=	Minor Groove Binding Track
MIF	=	Molecular Interaction Fields
NNRTI-BP	=	Non-nucleoside Reverse Transcriptase inhibitor-Binding Pocket
NNRTIs	=	Non-nucleoside Reverse Transcriptase inhibitors
NRTIs	=	Nucleoside Reverse Transcriptase inhibitors
ORF	=	Open Reading Frame
PDB	=	Protein Data Bank
PR	=	Protease
RMSD	=	Root Mean Square Deviation
RT	=	Reverse Transcriptase
SIV	=	Simian Immunodeficiency Virus
TAMs	=	Thymidine Analogue Mutations

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article. In this section 24 additional Figures are reported.

REFERENCES

[1] Gotte, M.; Li, X.; Wainberg, M.A. HIV-1 reverse transcription: a brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Arch. Biochem. Biophys.*, **1999**, *365*, 199-210.

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

[3] Wang, J.; Smerdon, S.J.; Jager, J.; Kohlstaedt, L.A.; Rice, P.A.; Friedman, J.M. Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. *Proc. Natl. Acad. Sci. U.S.A.*, **1994**, *91*, 7242-7246.

[4] Camarasa, M.J.; Velàzquez, S.; San-Fèlix, A.; Pèrez-Pèrez, M.J.; Gago, F. Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: a single mode of inhibition for the three HIV enzymes? *Antiviral. Res*., **2006**, *71*, 260-267.

[5] Huang, H.; Chopra, R.; Verdine, G.L.; Harrison, S.C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*, **1998**, *282*, 1669-1675.

[6] Ding, J.; Das, K.; Hsiou, Y.; Sarafianos, S.G.; Clark, A.D. Jr; Jacobo-Molina, A.; Tantillo, C.; Hughes, S.H.; Arnold, E. Structure and Functional Implications of the Polymerase Active Site Region in a Complex of HIV-1 RT with a Double-stranded DNA Template-primer and an Antibody Fab Fragment at 2.8 Å Resolution. *J. Mol. Biol.*, **1998**, *284*, 1095-1111.

[7] Julias, J.G.; McWilliams, M.J.; Sarafianos, S.G.; Alvord, W.G.; Arnold, E.; Hughes, S.H. Mutation of amino acids in the connection domain of human immunodeficiency virus type 1 reverse transcriptase that contact the template-primer affects RNase H activity. *J. Virol.*, **2003**, *15*, 8548-8554.

[8] Tachedjian, G.; Aronson, H.E.G.; Goff, S.P. Analysis of mutations and suppressors affecting interactions between the subunits of the HIV type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.*, **2000**, *97*, 6334-6339.

[9] Mulky, A.; Sarafianos, S.G.; Jia, Y.; Arnold, E.; Kappes, J.C. Identification of amino acid residues in the human immunodeficiency virus type-1 reverse transcriptase tryptophan-repeat motif that are required for subunit interaction using infectious virions. Intrinsic tryptophan fluorescence of bovine liver adenosine kinase, characterization of ligand binding sites and conformational changes. *J. Mol. Biol.*, **2005**, *349*, 673-684.

[10] Baillon, J.G.; Nashed, N.T.; Kumar, A.; Wilson, S.H.; Jerina, D.M. A leucine zipper-like motif may mediate HIV reverse transcriptase subunit binding. *New Biol.*, **1991**, *3*, 1015-1019.

[11] Becerra, S.P.; Kumar, A.; Lewis, M.S.; Widen, S.G.; Abbotts, J.; Karawya, E.M.; Hughes, S.H.; Shiloach, J.; Wilson, S.H. Protein-protein interactions of HIV-1 reverse transcriptase: implication of central and C-terminal regions in subunit binding. *Biochemistry*, **1991**, *30*, 11707-11719.

[12] Ceccherini-Silberstein, F.; Gago, F.; Santoro, M.; Gori, C.; Svicher, V.; Rodriguez-Barrios, F.; D'Arrigo, R.; Ciccozzi, M.; Bertoli, A.; D'Arminio Monforte, A.; Balzarini, J.; Antinori, A.; Perno, C.F. High Sequence Conservation of Human Immunodeficiency Virus Type-1 Reverse Transcriptase under Drug Pressure despite the Continuous Appearance of Mutations. *J. Virol.*, **2005**, *79*, 10718-10729.

[13] Auwerx, J.; Van Nieuwenhove, J.; Rodrìguez-Barrios, F.; de Castro, S.; Velàzquez, S.; Ceccherini-Silberstein, F.; De Clercq, E.; Camarasa, M.J.; Perno, C.F.; Gago, F.; Balzarini, J. The N137 and P140 amino acids in the p51 and the P95 amino acid in the p66 subunit of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase are instrumental to maintain catalytic activity and to design new classes of anti-HIV-1 drugs. *FEBS Lett.*, **2005**, *579*, 2294-2300.

[14] Pandey, P.K.; Kaushik, N.; Singh, K.; Sharma, B.; Upadhyay, A.K.; Kumar, S.; Harris, D.; Pandey, V.N. Insertion of a small peptide of six amino acids into the beta7-beta8 loop of the p51 subunit of HIV-1 reverse transcriptase perturbs the heterodimer and affects its activities. *BMC Biochem.*, **2002**, *3*, 1-18.

[15] Hoffmann, C.; Rockstroh, J.K.; Kamps, B.S. *HIV medicine*, Flying Publisher: Paris, **2006**.

[16] Lewis, W. Nucleoside reverse transcriptase inhibitors, mitochondrial DNA and AIDS therapy. *Antivir. Ther.*, **2005**, *10*, Suppl 2, M13-27.

[17] Margot, N.A.; Isaacson, E.; McGowan, I.; Cheng, A.; Miller, M.D. Extended treatment with tenofovir disoproxil fumarate in treatment-experienced HIV-1-infected patients: genotypic, phenotypic, and rebound analyses. *J. Acquired Immune Defic. Syndr.*, **2003**, *33*, 15-21.

[18] Sluis-Cremer, N.; Temiz, N.A.; Bahar, I. Conformational changes in HIV-1 reverse transcriptase induced by non-nucleoside reverse transcriptase inhibitor binding. *Curr. HIV Res.*, **2004**, *2*, 323-332.

[19] De Clercq, E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents*, **2009**, *33*, 307-320.

[20] Campiani, G.; Ramunno, A.; Maga, G.; Nacci, V.; Fattorusso, C.; Catalanotti, B.; Morelli, E.; Novellino, E. Non-nucleoside HIV-1 reverse transcriptase (RT) inhibitors: past, present, and future perspectives. *Curr. Pharm. Des.*, **2002**, *8*, 615-657.

[21] XVII International AIDS conference 2008: Abstracts THAB0103, THAB0402 and THAB0403.

[22] Vingerhoets, J.; Peeters, M.; Azijn, H.; Tambuyzer, L.; Hoogstoel, A.; Nijs, S.; de Béthune, M.P.; Picchio, G. An update of the list of NNRTI mutations associated with decreased virological response to etravirine: multivariate analyses on the pooled DUET-1 and DUET-2 clinical trial data [abstract 24]. *Antivir. Ther.*, **2008**, *13*, Suppl 3:A26.

[23] Parniak, M.A.; Sluis-Cremer, N. Inhibitors of HIV-1 reverse transcriptase. *Adv. Pharmacol.*, **2000**, *49*, 67-109.

[24] Clavel, F.; Hance, A.J. HIV drug resistance. *New Engl. J. Med.*, **2004**, *350*, 1023-1035.

[25] Boyer, P.L.; Currens, M.J.; McMahon, J.B.; Boyd, M.R.; Hughes, S.H. Analysis of nonnucleoside drug-resistant variants of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.*, **1993**, *67*, 2412-2420.

[26] http://www.rcsb.org (108 RT available models).

[27] Ortuso, F.; Langer, T.; Alcaro, S. GBPM: GRID-based pharmacophore model: concept and application studies to protein-protein recognition. *Bioinformatics*, **2006**, *22*, 1449-1455.

[28] Goodford, P.J. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.*, **1985**, *28*, 849-857.

[29] Molecular Discovery Ltd., www.moleculardiscovery.com

[30] http://hivdb.stanford.edu/index.html

[31] Los Alamos DB, http://www.hiv.lanl.gov/content/index

[32] (a) MacroModel ver 7.2, Schrödinger Inc., Portland, OR, 1998-2001 (b) Mohamadi, F.; Richards, N.G.J.; Guida, W.C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W.C. MacroModel - An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.*, **1990**, *11*, 440-467.

[33] DeLano, W.L. The PyMOL Molecular Graphics System (**2002**) DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org

[34] Santos, A.F.; Lengruber, R.B.; Soares, E.A.; Jere, A.; Sprinz, E.; Martinez, A.M.; Silveira, J.; Sion, F.S.; Pathak, V.K.; Soares, M.A. Conservation patterns of HIV-1 RT connection and RNase H domains: identification of new mutations in NRTI-treated patients. *PLoS ONE*, **2008**, *3*, e1781.

[35] Mellors, J.W.; Bazmi, H.Z.; Schinazi, R.F.; Roy, B.M.; Hsiou, Y.; Arnold, E.; Weir, J.; Mayers, D.L. Novel mutations in reverse transcriptase of human immunodeficiency virus type 1 reduce susceptibility to foscarnet in laboratory and clinical isolates. *Antimicrob. Agents Chemother.*, **1995**, *39*, 1087-1092.

[36] Wrobel, J.A.; Chao, S.F.; Conrad, M.J.; Merker, J.D.; Swanstrom, R.; Pielak, G.J.; Hutchison III, C.A. A genetic approach for identifying critical residues in the fingers and palm subdomains of HIV-1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.*, **1998**, *95*, 638-645.

[37]　Fisher, T.S.; Darden, T.; Prasad, V.R. Mutations proximal to the minor groove-binding track of human immunodeficiency virus type 1 reverse transcriptase differentially affect utilization of RNA versus DNA as template. *J. Virol.*, **2003**, *77*, 5837-5845.

[38]　Harris, D.; Yadav, P.N.; Pandey, V.N. Loss of polymerase activity due to Tyr to Phe substitution in the YMDD motif of human immunodeficiency virus type-1 reverse transcriptase is compensated by Met to Val substitution within the same motif. *Biochemistry*, **1998**, *37*, 9630-9640.

[39]　Mulky, A.; Vu, B.C.; Conway, J.A.; Hughes, S.H.; Kappes, J.C. Analysis of amino acids in the beta7-beta8 loop of human immunodeficiency virus type 1 reverse transcriptase for their role in virus replication. *J. Mol. Biol.*, **2007**, *365*, 1368-1378.

[40]　Pandey, P.K.; Kaushik, N.; Talele, T.T.; Yadav, P.N.; Pandey, V.N. The beta7-beta8 loop of the p51 subunit in the heterodimeric (p66/p51) human immunodeficiency virus type 1 reverse tran-scriptase is essential for the catalytic function of the p66 subunit. *Biochemistry*, **2001**, *40*, 9505-9512.

[41]　Goel, R.; Beard, W.A.; Kumar, A.; Casas-Finet, J.R.; Strub, M.P.; Stahl, S.J.; Lewis, M.S.; Bebenek, K.; Becerra, S.P.; Kunkel, T.A. Structure/function studies of HIV-1(1) reverse transcriptase: di-merization-defective mutant L289K. *Biochemistry*, **1993**, *32*, 13012-1318.

[42]　Navarro, J.M.; Damier, L.; Boretto, J.; Priet, S.; Canard, B.; Quérat, G.; Sire, J. Glutamic residue 438 within the Protease-sensitive sub-domain of HIV-1 Reverse Transcriptase is critical for heterodimer processing in viral particles. *Virology*, **2001**, *290*, 300-308.

[43]　Tachedjian, G.; Aronson, H.E.; De Los Santos, M.; Seehra, J.; McCoy, J.M.; Goff, S.P. Role of residues in the tryptophan repeat motif for HIV-1 reverse transcriptase dimerization. *J. Mol. Biol.*, **2003**, *326*, 381-396.

[44]　Hizi, A.; Shaharaban, M. Functional analysis of novel selective mutants of the reverse transcriptase of human immunodeficiency virus type 1. *J. Biol. Chem.*, **1992**, *267*, 18255-18258.

[45]　Klarmann, G.J.; Hawkins, M.E.; Le Grice, S.F.J. Uncovering the Complexities of Retroviral Ribonuclease H Reveals its Potential as a Therapeutic Target. *AIDS Rev.*, **2002**, *4*, 183-194.

[46]　Yi-Brunozzi, H.Y.; Le Grice, S.F.J. Uncovering the complexities of retroviral Ribonuclease H reveals its potential as a therapeutic tar-get. *J. Biol. Chem.*, **2005**, *280*, 20154-20162.

[47]　Nikolenko, G.N.; Palmer, S.; Maldarelli, F.; Mellors, J.W.; Coffin, J.M.; Pathak, V.K. Mechanism for nucleoside analog-mediated ab-rogation of HIV-1 replication: balance between RNase H activity and nucleotide excision. *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, *102*, 2093-2098.

[48]　Depollier, J.; Hourdou, M.L.; Aldrian-Herrada, G.; Rothwell, P.; Restle, T.; Divita, G. Insight into the mechanism of a peptide in-hibitor of HIV reverse transcriptase dimerization. *Biochemistry*, **2005**, *44*, 1909-1918.

[49]　Pelemans, H.; Esnouf, R.; De Clercq, E.; Balzarini, J. Mutational analysis of trp-229 of human immunodeficiency virus type 1 re-verse transcriptase (RT) identifies this amino acid residue as a prime target for the rational design of new non-nucleoside RT in-hibitors. *Mol. Pharmacol.*, **2000**, *57*, 954-960.

[50]　Ding, J.; Hughes, S.H.; Arnold, E. Protein-nucleic acid interactions and DNA conformation in a complex of human immunodeficiency virus type 1 reverse transcriptase with a double-stranded DNA template-primer. *Biopolymers*, **1997**, *44*, 125-138.

[51]　Sarafianos, G.S.; Das, K.; Tantillo, C.; Clark Jr, A.D.; Ding, J.; Whitcomb, J.M.; Boyer, P.L.; Hughes, H.S.; Arnold, E. Crystal structure of HIV-1 reverse transcriptase in complex with a poly-purine tract RNA: DNA. *EMBO J.*, **2001**, *20*, 1449-1461.

[52]　Sharma, B.; Kaushik, N.; Singh, K.; Kumar, S.; Pandey, V.N. Sub-stitution of conserved hydrophobic residues in motifs B and C of HIV-1 RT alters the geometry of its catalytic pocket. *Biochemistry*, **2002**, *41*, 15685-15697.

[53]　Klarmann, G.J.; Smith, R.A.; Schinazi, R.F.; North, T.W.; Preston, B.D. Site-specific incorporation of nucleoside analogs by HIV-1 Reverse Transcriptase and the template grip mutant P157S. Template interactions influence substrate recognition at the polymerase active site. *J. Biol. Chem.*, **2000**, *275*, 359-366.

[54]　Ghosh, M.; Williams, J.; Powell, M.D.; Levin, G.J.; Le Grice, S.F.J. Mutating a conserved motif of the HIV-1 Reverse Tran-

[55]　scriptase palm subdomain alters primer utilization. *Biochemistry*, **1997**, *36*, 5758-5768.

[55]　Beard, W.A.; Stahl, S.J.; Kim, H.R.; Bebenek, K.; Kumar, A.; Strub, M.P.; Becerra, S.P.; Kunkel, T.A.; Wilson, S.H. Structure/function studies of human immunodeficiency virus type 1 reverse transcriptase. Alanine scanning mutagenesis of an alpha-helix in the thumb subdomain. *J. Biol.Chem.*, **1994**, *269*, 28091-28097.

[56]　Sarafianos, S.G.; Clark Jr., A.D.; Tuske, S.; Squire, C.J.; Das, K.; Sheng, D.; Ilankumaran, P.; Ramesha, A.R.; Kroth, H.; Sayer, J.M.; Jerina, D.M.; Boyer, P.L.; Hughes, S.H.; Arnold, E. Trapping HIV-1 Reverse Transcriptase before and after translocation on DNA. *J. Biol. Chem.*, **2003**, *278*, 16280-16288.

[57]　Beard, W.A.; Bebenek, K.; Darden, T.A.; Li, L.; Prasad, R.; Kunkel, T.A.; Wilson, S.H. Vertical-scanning mutagenesis of a critical tryptophan in the minor groove binding track of HIV-1 re-verse transcriptase. Molecular nature of polymerase-nucleic acid interactions. *J. Biol. Chem.*, **1998**, *273*, 30435-30442.

[58]　Rausch, J.W.; Lener, D.; Miller, J.T.; Julias, J.G.; Hughes, S.H.; Le Grice, S.F.J. Altering the RNase H primer grip of human immu-nodeficiency virus Reverse Transcriptase modifies cleavage speci-ficity. *Biochemistry*, **2002**, *41*, 4856-4865.

[59]　Pata, D.J.; Stirtan, W.G.; Goldstein, S.W.; Steitz, T.A. Structure of HIV-1 reverse transcriptase bound to an inhibitor active against mutant reverse transcriptases resistant to other non-nucleoside in-hibitors. *Proc. Natl. Acad. Sci. U.S.A.*, **2004**, *101*, 10548-10553.

[60]　Boyer, P.L.; Ferris, A.L.; Clark, P.; Whitmer, J.; Frank, P.; Tanti-llo, C.; Arnold, E.; Hughes, S.H. Mutational analysis of the fingers and palm subdomains of human immunodeficiency virus type-1 (HIV-1) reverse transcriptase. *J. Mol. Biol.*, **1994**, *243*, 472-483.

[61]　Das, K.; Ding, J.; Hsiou, Y.; Clark Jr, A.D.; Moereels, H.; Koy-mans, L.; Andries, K.; Pauwels, R.; Janssen, P.A.J.; Boyer, P.L.; Clark, P.; Smith Jr, R.H.; Kroeger Smith, M.B.; Michejda, C.J.; Hughes, S.H.; Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.*, **1996**, *264*, 1085-1100.

[62]　Ren, J.; Nichols, C.E.; Stamp, A.; Chamberlain, P.P.; Ferris, R.; Weaver, K.L.; Short, S.A.; Stammers, D.K. Structural insights into mechanisms of non-nucleoside drug resistance for HIV-1 reverse transcriptase mutated at codons 101 or 138. *FEBS J.*, **2006**, *273*, 3850-3860.

[63]　Esnouf, R.M.; Ren, J.; Hopkins, A.L.; Ross, C.K.; Jones, E.Y.; Stammers, D.K.; Stuart, D.I. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis (het-eroaryl) piperazine (BHAP) U-90152 explain resistance mutations for this non-nucleoside inhibitor. *Proc. Natl. Acad. Sci. U.S.A.*, **1997**, *94*, 3984-3989.

[64]　Rodrìguez-Barrios, F.; Balzarini, J.; Gago, F. The molecular basis of resilience to the effect of the Lys103Asn mutation in non-nucleoside HIV-1 reverse transcriptase inhibitors studied by tar-geted molecular dynamics simulations. *J. Am. Chem. Soc.*, **2005**, *127*, 7570-7578.

[65]　Smerdon, S.J.; Jager, J.; Wang, J.; Kohlstaedt, L.A.; Chirino, A.J.; Friedman, J.M.; Rice, P.A.; Steitz, T.A. Structure of the binding site for non-nucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.*, **1994**, *91*, 3911-3915.

[66]　Lowe, D.M.; Parmar, V.; Kemp, S.D.; Larder, B.A. Mutational analysis of two conserved sequence motifs in HIV-1 reverse tran-scriptase. *FEBS Lett.*, **1991**, *282*, 231-234.

[67]　Bakhanashvili, M.; Avidan, O.; Hizi, A. Mutational studies of human immunodeficiency virus type 1 reverse transcriptase: the in-volvement of residues 183 and 184 in the fidelity of DNA synthe-sis. *FEBS Lett.*, **1996**, *391*, 257-262.

[68]　Harris, D.; Kaushik, N.; Pandey, P.K.; Yadav, P.N.S.; Pandey, V.N. Functional analysis of amino acid residues constituting the dNTP binding pocket of HIV-1 Reverse Transcriptase. *J. Biol. Chem.*, **1998**, *273*, 33624-33634.

[69]　Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. *Biophys. J.*, **2003**, *84*, 3547-3563.

[70] Bacolla, A.; Shih, C.K.; Rose, J.M.; Piras, G.; Warren, T.; Grygon, C.A.; Ingraham, R.; Cousins, R.B.; Greenwood, D.; Richman, D.; Chengli, Y.C.; Griffin, J. Amino acid substitutions in HIV-1 reverse transcriptase with corresponding residues from HIV-2. Effect on kinetic constants and inhibition by non-nucleoside analogs. *J. Biol. Chem.,* **1993**, *268*, 16571-16577.

[71] Paolucci, S.; Baldanti, F.; Campanini, G.; Cancio, R.; Belfiore, A.; Maga, G.; Gerna, G. NNRTI-selected mutations at codon 190 of human immunodeficiency virus type 1 reverse transcriptase decrease susceptibility to stavudine and zidovudine. *Antiviral. Res.,* **2007**, *76*, 99-103.

[72] Huang, W.; Gamarnik, A.; Limoli, K.; Petropoulos, C.J.; Whitcomb, J.M. Amino acid substitutions at position 190 of human immunodeficiency virus type 1 reverse transcriptase increase susceptibility to delavirdine and impair virus replication. *J. Virol.,* **2003**, *77*, 1512-1523.

[73] Pelemans, H.; Esnouf, R.M.; Parniak, M.A.; Vandamme, A.M.; De Clercq, E.; Balzarini, J. A proline-to-histidine substitution at position 225 of human immunodeficiency virus type 1 (HIV-1) reverse

[74] transcriptase (RT) sensitizes HIV-1 RT to BHAP U-90152. *J. Gen. Virol.,* **1998**, *79*, 1347-1352.

Palaniappan, C.; Wisniewski, M.; Jacques, P.S.; Le Grice, S.F.; Fay, P.J.; Bambara, R.A. Mutations within the primer grip region of HIV-1 reverse transcriptase result in loss of RNase H function. *J. Biol. Chem.,* **1997**, *272*, 11157-11164.

[75] Pelemans, H.; Esnouf, R.M.; Jonckheere, H.; De Clercq, E.; Balzarini, J. Mutational analysis of Tyr-318 within the non-nucleoside reverse transcriptase inhibitor binding pocket of human immunodeficiency virus type I reverse transcriptase. *J. Biol. Chem.,* **1998**, *273*, 34234-34239.

[76] Jonckheere, H.; Taymans, J.M.; Balzarini, J.; Velazquez, S.; Camaras, M.J.; Desmyter, J.; De Clercq, E.; Annè, J. Resistance of HIV-1 reverse transcriptase against [2',5'-bis-O-(tert-butyldimethylsilyl)-3'-spiro-5"-(4"-amino-1",2"- oxathiole-2",2"-dioxide)] (TSAO) derivatives is determined by the mutation Glu138-->Lys on the p51 subunit. *J. Biol. Chem.,* **1994**, *269*, 25255-25258.