Università degli Studi di Cagliari

DOTTORATO DI RICERCA

Sviluppo e sperimentazione dei farmaci antivirali

Ciclo XXIII

TITOLO TESI

A Computational Approach for

Identification and Development of Novel Inhibitors Targeting Viral Polymerases

BIO/19

Presentata da:	Shailendra Asthana
Coordinatore Dottorato:	Prof.ssa. Alessandra Pani
Tutor:	Prof. Paolo La Colla
	Dr. Attilio V. Vargiu
	Prof. Paolo Ruggerone

Esame finale anno accademico 2009-2010

University of Cagliari

Department of Biomedical Science and Technology

Section of General Microbiology and Virology

& Microbial Biotechnologies

Research Doctorate in

Development and Evaluation of Antiviral Drugs

Coordinator of the Doctorate

Prof. Alessandra Pani

A Computational Approach for Identification and Development of Novel Inhibitors Targeting Viral Polymerases

Supervisor:

Prof. Paolo La Colla

Co-Supervisor:

Dr. Attilio V. Vargiu

Prof. Paolo Ruggerone

Candidate:

Shailendra Asthana

XXIII Cycle 2008-2010

Acknowledgments

A journey is easier when you travel together. Interdependence is certainly more valuable than independence. This thesis is the result of three years of work whereby I have been accompanied and supported by many people. This is an opportunity to thanks all of those who have been with me during these three years.

The first persons I would like to thank are my advisors Prof. Paolo La Colla, Prof. Paolo Ruggerone and Dr. Attillio V. Vargiu. I started in Prof. La Colla's lab in June, 2005 in a project and since then continued in his lab. I have always been influenced by his integral view on research and his mission for providing 'only high-quality work and not less' have made a deep impression on me.

Taking this as an great opportunity, I would like to extend my sincere thanks to Prof. Paolo Ruggerone for accepting me in his lab. I got immense encouragement and inspiration during my stay in his lab. I continued in his lab as a student jointly advised by him and Dr. Attillio Vargiu. Dr. Vargiu's helpful discussions and support have played an important role in the quality of my research and its successful completion.
I

I would like to thank my friend and colleague of three years, Dr. Saumya Shukla. We came together to Italy, joined PhD together and worked on the same project. Ours is an experience of growing up together. We have struggled, achieved and now celebrating the completion of our PhD together. Thanks for being with me through all ups and downs.

Successful completion of a thesis requires loads of work from many many people, some contribute to it professionally and others by creating an atmosphere suitable for research and high efficiency.During these three years, I made some very good friends. I take this as an opportunity to thank all of them to make my life easier and to make me feel at home. Cagliari have become a home away from home. I would like to extend my thanks to all my friends for their kind support and encouragement .

Finally, I would like to thank my parents for all their efforts and support all through my life. I feel a deep sense of gratitude for my father and mother who formed part of my vision and taught me the good things that really matter in life. I would like to dedicate this work to my parents.

Summary

Positive strand RNA viruses, which include hepatitis C virus (HCV), human immunodeficiency virus (HIV), and Bovine Viral Diarrhea Virus (BVDV), are known to create havoc for humans and animal health alike. Although vaccines have helped to control several of the most important viral pathogens, there is currently little prospect of an effective vaccine for either HCV or HIV. These pathogens infect ~170 million and ~40 million people worldwide, respectively, hastening the need for effective antiviral drugs. Likewise BVDV infects domesticated livestock causing significant economic losses worldwide. The development of new, effective antiviral compounds for combating these debilitating human (HIV and HCV) and animal pathogen (BVDV) is therefore of paramount importance, and is the focus of this thesis.

Herein, polymerases of three positive strand RNA viruses, viz HCV, BVDV and HIV have been targeted with the goal of improving the efficacy of antivirals against wide range of resistant mutations. Lack of effective therapies for these viral infections as most of the established treatments are not always effective or well tolerated, highlights an urgent need for further refinement and development of antiviral drugs. It is not only the specific need that has inspired this work but also the idea to test and develop protocols that might enable a more rational structure-based drug design to be performed by keeping a tradeoff among rapidity, accuracy, and efficacy.

Traditional methods for general drug discovery typically include evaluating random compound libraries for activity in relevant cell-free or cell-based assays. Success in antiviral development has emerged from the discovery of more focused libraries that provide clues about structure-activity relationships. Combining these with more recent approaches including structural biology and computational modeling can work efficiently to hasten discovery of active molecules. The ability to design drugs interfering with the progression of infection of virus comes with i)- the knowledge of pathological, cellular and molecular mechanism involved in the disease; and ii)- the identification of macromolecule (i.e possible drug target) involved in pathological pathways, their 3D structures and their functions. The biological activity of drug molecules is dependent on the three-dimensional arrangement of its functional groups, which specifically bind to their target. Consequently, the structural information of the target protein is essential in drug development.

Proteins are dynamic molecules and often undergo conformational change upon ligand binding. The flexible loop regions and in general the flexibility of the structure have a critical functional role in enzymes, but those features and their connection with the functionality of protein are hard to retrieve from x-ray, NMR techniques and cryo-EM techniques.

Being aware of the importance of the relationship structure-function and structure-activity at large, i.e., including dynamics and interactions with solvent, in our work we are trying to address some of the relevant problems of drug development; basic key determinants in protein-ligand stability, mechanism of inhibition, why and how, flexibility and collective motion of the protein is essential part in improvement of rational drug design, how mutation renders the protein resistant again potent drugs; the effect of resistance mutation on the flexibility and stability of protein, what is the mechanism of drug resistance, change in energetics consequences, affecting the conformation in wild and mutant systems. Various biophysical techniques of the computational arsenal we have applied have provided huge wealth of information related to protein dynamics and protein-ligand recognition. These methods have grown in their effectiveness not only by offering a deeper understanding of the basic science, the biological events and molecular interactions that define a target for therapeutic intervention, but also because of advances in algorithms, representations, and mathematical procedures for studying such processes.

This work represents the application of several computational techniques, such as docking, molecular dynamics, algorithms to calculate free energy of binding of ligands into the binding pocket (ex MM-PBSA) and algorithms to study rare events (for ex. binding and unbinding of ligand from the binding site, Metadynamics) to explore, at microscopic level, the key pattern of interaction between protein and ligand, to understand the effect of mutations, to get an insight of the full docking and undocking path and to calculate binding energetics.

Contents:

        - Structure of Receptors (x-ray)
        - MD Simulation protocol
            - Parametrization
            - Dynamics
            - Clustering
        - Structure of Complexes
            - Docking
            - Analysis of structures and dynamics
            - Area Analysis
    - Results and Discussion
        - Biological
            - Molecular characterization of drug-resistant mutants
        - Computational
            - Clustering
            - Extensive docking
            - Analysis of the dynamics trajectory
                - Secondary structural change
                - Overall pertubative change in four models
                - Dynamic Behavior of binding site
                    - Mobility of the loop encircling the cavity
                    - The Linker
                    - Role of mutation and its effect
                    - Role of I261
                    - The Electrostatic map of binding site
                - Inhibitor binding mode
                    - Wild type complex
                    - Mutant complex
            - Hypothetical Mechanism of Drug resistance
                - Area Analysis
                - Model
                - Chunnel analysis
    - Conclusion
    - References
- **The Energetic aspect**
    - Abstract
    - Introduction
    - Result and Discussion
        - Unbinding of the 227G: Use of Metadynamics
        - Binding Free energies: Use of MM-PBSA
        - Hydration of the compound
    - Conclusion
    - Material and Methods

- References

III)- DABO snuggles its way to HIV-RT Inhibition

- Abstract
- Epidemiology
- Result and discussion
  - Docking
  - MD Simulations
    - Interactions pattern
    - RT residues important for binding of DABO compounds
    - Comparison of DABO with similar compounds (X-ray)
    - The 5-6 position triggers
    - Protein Flexibility
    - Mechanism of inhibition
    - Conclusion
    - Work in progress
    - References

IV)-Conclusion

List of Figures :

1)Introduction:

Chapter1)-

red stick representation) to the final and stable one (yellow) enclosed by L2 and motif I. Blue and green conformations are intermediate ones along the flipping process.

Figure6: Root-mean-square deviation (RMSD) of Four system (APO, APOm, COM and COMm) of backbone atoms as a function of time. B)-RMSD value per residue averaged over last 20ns with respect to starting structure.

Figure7: Root mean square fluctuation (RMSF) of BVDV RdRp, extracted from B-factors (X-ray) or calculated from MD simulations (APO, APOm, COM and COMm).

Figure8: Minimum distance between the loop L2-L4 and L1-L4 along the equilibrium trajectory: A)-The graph (left) indicates the minimum distance values that were sampled by loop L2-L4 throughout the trajectory B) Average distances (and standard deviations) between the loops L_4, L_2, L_3 and L_1.

Figure9: Time evolution of protein (APO and APOm) or complex (COM and COMm) Cα-RMSDs, calculated with respect to starting structures A) Binding Site B)- Linker region C) Carbon alpha distribution of linker

Figure10: Superimposition of the x-ray with APO (A) and APOm(B) taking snapshots at time corresponding to peaks and dips in the rmsd of linker (as denoted by arrows and dots in figure fig6c. The movement of loops which corresponds to binding pocket during the MD simulation. A)- periodic movement of loops focusing linker in APO while a transient movement observed in (B)APOm .

Figure11: Intra-protein H-bonds network formed between residues of the fingertip and surroundings.

Figure12: Electrostatic potential maps of the 227G binding pockets of RdRp: (A) COM(I261); (B) COMm (M261) surface (positive in blue, negative in red, and neutral in white).

Figure13: Superimposition of the four loops L1, L2, L3 and L4 from the four systems APO, APOm, COM and COMm: red, blue, green and cyan color indicate APO, APOm, COM and COMm system respectively.

Figure14: Interaction map of COM-227G and COMm-227G with key residues of four loops represented L1, L2, L3 and L4 in magenta color and motif-I is blue in color.

Figure15: Top view of COM (a) and COMm (b) representative conformations compared with the X-ray structure of BVDV RdRp (c) . For the sake of clarity the views are slightly rotated from the the axis entering the RNA template gate, and different in each picture

Figure16:Area of the region enclosed by loops L1, L2 and L4 lining the entrance gate.

Figure17:Model illustrating the putative mechanism of BVDV RdRp resistance against 227G. A cartoon representation of BVDV RdRp is shown with the thumb (blue), finger (green) and palm (pink) domain.

Figure18: chunnel analysis- We find only the three known tunnels with chunnel analysis. Left pannel correspond to clusters obtained in COM while right panel correspond to clusters of COMm.

Chapter2)-

Figure1:(a) Ball and stick representation of 227G with oxygen atoms in red, nitrogen atoms in blue,carbon atoms in cyan, sulfur atom in yellow, and hydrogen atoms in white. (b) Cartoon representation of BVDV RdRp. Different domains of RdRp are differently colored: thumb in green, finger in blue, palm in red, and the unique N-terminal in magenta.

Figure2: Evolution of the center of mass of 227G in (a) COMPLEX and (b) COMPLEXm during the metadynamics simulations of the undocking process.

Figure3: Free energy surfaces as a function of dcom and nhph for the unbinding of 227G from COMPLEX. Isosurfaces are drawn one per 1 kcal/mol. Insets show Ligplot of snapshots of 227G in the minima extracted from the metadynamics runs.

Figure4: Same as Figure 3 for the unbinding of 227G from COMPLEXm.

Figure5: Ligand-residue interaction energy (in kcal/mol) for COMPLEX (cyan) and COMPLEXm (blue).

Figure6: Histogram of number of water molecules around the drug during the simulation in COMPLEX and COMPLEXm system represented by green and cyan, respectively.

Chapter3)-

Figure1:Cartoon representation of HIV-RT. Two subunits P66(colored according to the subdomains) and P51(white) are depicted here. Finger is represented in blue, Palm in red, Thumb in green, the connection in magenta and RNaseH in yellow. The template is also represented here with the two strands being colored in orange and cyan.

Figure2: Docked conformation of DABO MC1220 superimposed on the Crystallographic structure of HEPT compound(1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine).(1RT1[51]) MC1220(depicted as red ball and stick) binds into an orientation similar to that of HEPT(represented as blue ball and stick) compounds. B) Docked structures of the DABO compound into the NNRTI binding cavity. The orientation acquired by the four DABO compounds were nearly identical. MC1220 is depicted in mauve, MC1332 in blue, MC1195 in red and MC1346 in green color.

Figure3:Pictorial representation of ligand-residue interaction spectrum of key residues of NNRTI binding site with DABO compounds. The residues are represented in stick while the HB are depicted in magenta color spirals. B) Ligand-residue interaction energies in kcal/mol for the residues that contribute most to the ligand-surroundings. The red color bar denote the electrostatic contribution, Blue correspond to the Van Der Waals energies and Black denotes the total interaction energies.

Figure4: The chemical structure of the two Rt Inhibitors studied, A)-HEPT, B)-MKC-442. Images taken from Hopkins A.L. Et al 1996; J.Me. Chem.

Figure5:Conformational Flexibilty. RMSF of all C^{α} atoms from their time-averaged positions. Each subunit is indicated by black lines, along with the subdomains of the p66 subunit (F, fingers; P, palm; T, thumb; C, connection; R, RNH).

Figure6:Minimum distance between the thumb and finger subdomains along the trajectory, A)Normalized histogram calculated for each system. The red and purple dotted line indicate the distance between the thumb and finger domain in 1DLO(without substrate) and 1RTD (with substrate) respectively. B) Cartoon representation of the polymerase region, illustrating the distance calculated. Here 1DLO, represented in red, is superimposed with MC1220 bound structure, in black, 1332-bound structure, in yellow, 1195-bound structure, in blue, and 1346 bound structure in green color.

Figure7:Minimum distances between two beta sheets of the NNRTI binding site which harbors two important conserved structural features of HIVRT, the Primer grip and catalytic triad(YMDD motif). A) Normalized histogram of the distances are repersented. The red and purple dotted line indicate the minimum distances between the beta sheets in 1DLO(without substrate) and 1RTD (with substrate) respectively. B) Illustration of the distances calculated. Superimposition of beta sheets of 1DLO(transparent pink) with the MC1220 bound structure(black)

List of Tables :

Introduction:

I)-

Background and Computational Approaches

Introduction

"Prediction is the attempt to use existing knowledge to foretell an event before it happens. In particular, modeling approaches will be crucial for moving biology from a descriptive to a predictive science. Pharmaceutical companies identify molecular interventions that they predict will lead to therapies at the organism level, suggesting that computational biology can play a key role in the pharmaceutical industry."

The history of antiviral drug development has taken numerous circuitous routes from the discovery and development of the first US Food and Drug Administration (FDA) approved antiviral agent, 5-iodo-2 -deoxyuridine (Idoxuridine) by William H. Prusoff [1,2], to the development of simplified but powerful triple combinations for HIV such as Atripla and Trizivir [3,4,5]. Many challenges are encountered during the development of antiviral agents, including adverse events and the development of drug resistant viruses, which necessitate chemists, biologists, and pharmacologists to develop improved, more potent, and less toxic medicines with "high genetic barrier." Although there are major differences among viruses, specific virological, pharmacological and biophysical approaches are trying to identify common aspects across many viral diseases to develop either broad spectrum and specific antivirals.

Viral infections, directly or indirectly, effects the huge population of human being in the world. Positive strand RNA viruses, which include hepatitis C virus (HCV), human immunodeficiency virus (HIV), and Bovine Viral Diarrhea Virus (BVDV), are known to create havoc for humans and animal health alike.

Why HCV: Since the discovery of HCV in 1989, it has become the frequent cause of acute and chronic hepatitis, liver cirrhosis and hepatocellular carcinoma worldwide, and has created a significant burden to healthcare systems due to mortality, morbidity and treatment costs [6,7,8,9]. To give the numbers, hepatitis C (officially recognized as the "silent epidemic") infected approximately 180 million people worldwide, and it is the leading reason for liver transplantation in the United States[10,11]. Unfortunately, more than 20 years after discovery of HCV [56] therapeutic options remain limited.

Why BVDV: Likewise, Bovine viral diarrhea virus (BVDV), is a major viral pathogen in cattle and other ruminants[12]. Infection of cattle with BVDV can result in a wide assortment of disease manifestations including resorption, mummification, or abortion of the dead fetus while those fetuses who survive early infection may be malformed or blind, may have skeletal defects, respiratory problems, underdeveloped brain or weak immune system[13]. This results in a high mortality rate in cattle as well as cause huge economic losses to cattle producers throughout the world due to decreased performance, loss of milk production, reproductive wastage, and increased risk of morbidity and mortality[14]. It is the most costly viral disease in US cattle herds, with losses estimated at $10–$88 per head or $2 billion dollars per year[66] and thus there is an urgent need for highly effective and selective inhibitors to combat the deadly pathogen.

Why HIV: Projected as the third leading cause of death, by the year 2030, Human immunodeficiency virus reverse transcriptase (RT) inhibition has become a major cause of concern worldwide [15][16]. Since its discovery in 1983, the HIV epidemic has fuelled the need for the development of effective antiviral therapy.

According to the Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO) 2009 AIDS epidemic update, approximately 33.3 million people globally are living with HIV infection, 2.3 million of whom are children (Figure 2,3). One in 155 adults and adolescents are infected with human immunodeficiency virus (HIV) globally, with an astonishing rate of one in 21 in Sub-Saharan Africa.[17][18] In the USA alone, more than half a million people were living with HIV in 2007, and 44,000 new cases were reported in the same year.[19][20] The gravity of this infectious disease has led to a need for the development an approval of a range of pharmaceutical compound that interfere with various stages of the HIV-1 retro viral life cycle.

In fact, vaccination against viral infection came out as a powerful weapon, as it have led, or might led, to the eradication of important viral pathogens, such as smallpox, polio, measles, mumps, rubella, etc. However, other viral diseases, particularly HIV and HCV, have so far proved to be intractable for the vaccine approach. Although, studies have come up with new antiviral drugs against HCV and HIV (listed in Tables 1 and 2,

approved by FDA), there is yet considerable room for the improvement, as these compounds are not always efficacious or well tolerated. The emergence of viral resistance to drugs and drug related side effects are among limitations of development of antivirals. Initially the +ve strand viruses, which are susceptible to variety of antivirals, due to the propping up of resistance mutations were able to skirt every drug from the biotech arsenal. The capacity of the viral pathogen for rapid evolution is one of the main reason for the need of further refinement and development of antiviral drug design.

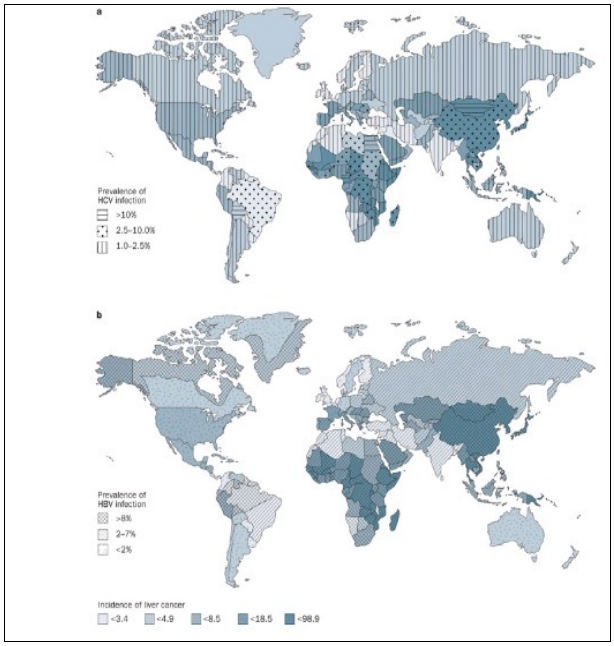


Figure1: Global variations in age adjusted incidence rates of liver cancer, prevalence of chronic HCV infection and chronic HBV infection. Adapted from: Yang, J. D. & Roberts, L. R. (2010) Hepatocellular carcinoma: a global view Nat, Rev. Gastroenterol. Hepatol. doi:10.1038/nrgastro.2010.100

Antiviral drugs could, in principle, be targeted at any of the stages of viral replication. Yet for many years viral diseases have been considered as troublesome for selective antiviral chemotherapy. The reason is that the replication cycle of the virus was assumed to be too closely interwoven with normal cell metabolism, so that any attempt to suppress virus reproduction would be doomed to kill (or severely harm) the uninfected cell as well. With the elucidation of virus specific events as target for chemotherapeutic attack and the advent of a number of specific antiviral agents, it has become increasingly clear that a selective chemotherapy of virus infection can be achieved and that virus reproduction can be suppressed without deleterious effect on the host [21] For example, RNA Dependant RNA Polymerases (RdRps) are now considered to be one of the most interesting targets for drugs since polymerase activity is essential for viral replication, and human host cells are devoid of such RdRp [22]. Moreover, success stories of inhibitors against the HIV reverse transcriptase, the hepatitis B virus polymerase and the HCV RdRp have validated viral polymerases as therapeutic targets [6,59].

In today's scenario, viral polymerase have become the most attractive target for the development of antiretroviral compounds. Having said this, the antiviral compounds against viral polymerases encompass of two classes of agents that inhibit polymerase by two different mechanism: nucleoside inhibitors (NIs), which mimic the endogenous substrates and bind competitively at a catalytic site, and non-nucleoside inhibitors (NNIs), which are a diverse group of compounds that bind to a allosteric cavity, inducing conformational changes that inhibits the function of RT and RdRp.[23,24]. These NIs and NNIs have become the cornerstone for successful treatment of viral infection.

Despite of being used as effective antiviral therapies, side effects of NIs are real and should not to be discounted. NIs generally (but not always) act with greater specificity for the viral polymerase, compared to mammalian DNA polymerases [24]. However, there is a separate enzyme (polymerase-gamma) inside the cell that replicates mitochondrial DNA and that can be affected by NIs. NIs can deplete or impair the function of this enzyme under certain circumstances. Another drawback of NIs is that they require three phosphorylation steps, catalyzed by cellular kinases, to be converted into active triphosphate metabolites. The active metabolites then act as competitive inhibitors or alternative substrates with respect to the normal substrates (either dATP, dGTP, dCTP or dTTP) and lead to the termination of chain elongation. Thus, their activation and efficacy depend on the metabolic state of the infected host cells. Given the properties of existing drugs, for new NIs, it is become increasingly difficult to comply with the demands for higher activity (potency), lower toxicity (side effects) and the more favorable resistance profile required for approval as antiviral drugs. These drawbacks of NIs are not affecting in NNIs and hence the latter are being used as powerful weapons to combat viral infection and are the focus of this thesis.

The above mentioned statistics about the debilitating human (HCV and HIV) and animal (BVDV) pathogen, as well as the success stories of effectiveness of viral polymerase as a target for development of antiviral therapies, have prompted me to target polymerases of three positive strand RNA viruses, viz, Hepatitis C Virus (HCV), Bovine Viral diarrhea Virus (BVDV) and Human Immunodeficiency Virus (HIV) in this thesis . Herein, I have extended my investigation to study the molecular recognition processes involving many different class of NNIs belonging to Benzimidazole, Imidazoquinoline, Pyridoxoquinoline, Phenanthroline and DABO against the former mentioned three protein systems, with the aim to offer hints for the design of more effective and potent NNIs effective against wide range of resistant mutations. Herein, Plethora of biophysical techniques has been applied on the three protein systems to answer fundamental questions of drug-protein molecular recognition, drug resistance, drug inhibition, thereby providing clues for improvement of antivirals.

In the last twenty-five years the advances in biological sciences, including many experimental techniques such as quantitative structure activity relationship, high-throughput screening (mass spectrometry) for identification of new lead compounds, replicon, and cell based assays to predict toxicity and resistance aspects, pharmacokinetic and pharmacodynamic aspects as radioactive enzymatic assays and fluorescence based assays. This plethora of methods are used to identify, evaluate and screen antiviral compounds against specific targets by providing their CC_{50}, EC_{50} and IC_{50} values. Another ground-breaking step in the antiviral development was the genome projects [26,27], which resolve the genome of organisms and led to an abundant amount of information about the amino acid sequences of proteins in pathogens. The wealth of knowledge about genetic sequences and sequences of amino acid, is invaluable for the modern taxonomic identification. [28,29]

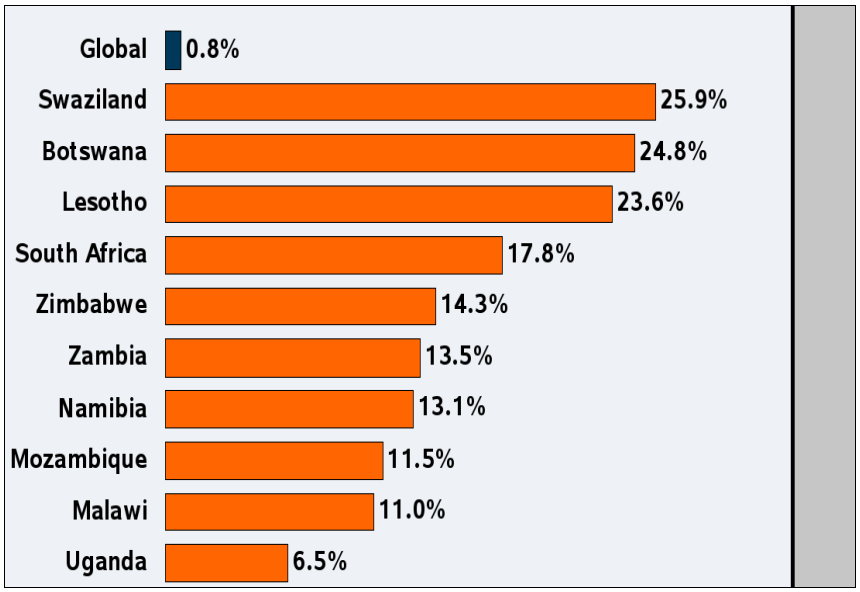


Figure2: Top 10 Countries, Adult HIV Prevalence Rate, 2009. Note: Data are estimate. Prevalence rates include adults aged 15-49. Adapted from : Kaiser Family Foundation, www.GlobalHealthFacts.org, based on UNAIDS, Report on the Global AIDS Epidemic, 2010.

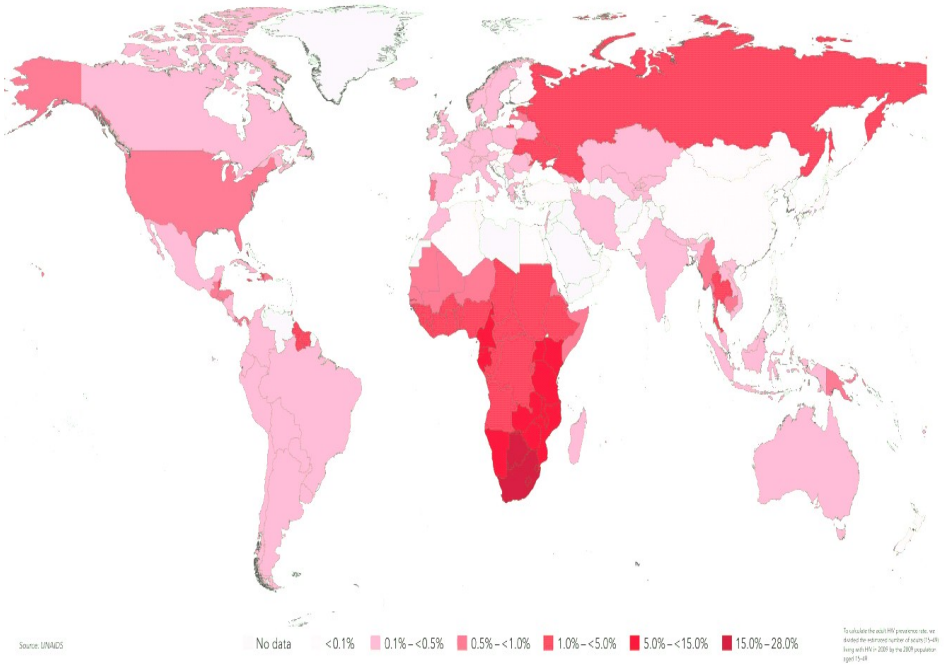


Figure3: A global View of HIV infection, 33.3 million people [31.4-35.3] living with HIV, 2009. Adapted From: Joint United Nations Programme on HIV/AIDS(UNAIDS) and World Health Organization (WHO) 2009 AIDS epidemic update

However, the amino acid sequence of proteins yields only limited information about the structure and function of biological macromolecules. Processes such as molecular recognition with respect to protein-ligand interaction, catalytic function of enzymes, gating mechanisms in ion channels, conformational change required by the protein to accommodate the ligands into binding site, mechanisms of inhibition and drug resistance, change in the protein dynamics on introduction of resistance mutations and understanding the dynamic behavior of proteins at different timescales, from fast internal motions to slow conformational changes, or even protein folding processes the assembly, and signal transduction or immune response pathways cannot be explained by these sequences.

To bridge this gap the development in the field of computational biochemistry is tremendously contributing by improving our understanding of complex bio-molecular systems such as proteins-ligand complexes, nucleic acids and bi-layer membranes. To gain insight into such processes, the three-dimensional structure of a folded amino acid sequences, a functional protein, is necessary. The structure determination of proteins or nucleic acids is based on experimental techniques such as x-ray crystallography [28,29] or nuclear magnetic resonance (NMR)[30,31], which have made remarkable progress in solving high-resolution structures over the past years. Such structures are archived and accessible via the RSCB Protein Data Bank. Furthermore, theoretical approaches for structure prediction from amino acid sequences are under constant development and have been monitored by CASP (Critical Assessment of Structure Prediction) over the last 13 years.

However, the view of a static protein or nucleic acid as the one extracted from x-ray cannot account for all the crucial features of dynamics and interaction processes, including interactions with the solvent. A particular snapshot of a protein structure is a point in configuration space. Accordingly, a fluctuating structure of a protein is described by a region on a high-dimensional complex free energy landscape, which is dynamically explored by the system.[32] Such protein structure dynamics is the key to obtain insight into the function of bio-molecules. Note that an apparently *simple* process, such as the binding of a small molecule to its protein target, is governed by not easily determinable or measurable entropic and enthalpic factors. The mobility of both ligand and receptor, the effect of the protein environment on the charge distribution over the ligand, [33,34,35] and their interactions with the surrounding water molecules, complicate the quantitative description of the process. A structure derived by classical X-ray crystallography is just an average and thus lacks dynamics. Moreover, the free energy minimum of the structure derived with the help of such a protein crystal is governed by the crystallization conditions and possibly changed by the typically low temperature usage. However, experimental methods to probe the structural dynamics with atomic detail are available, even though with intrinsic limitations. Since then several biochemical and biophysical techniques have came up and have been used to answer some intriguing and fundamental questions of protein-ligand dynamics such as the conformational change required by the protein to accommodate the ligands into binding site, mechanisms of inhibition and drug resistance, change in the protein dynamics on introduction of resistance mutations and understanding the dynamic behavior of proteins at different timescales, from fast internal motions to slow conformational changes, or even protein folding processes. All these development have brought a renaissance in the field of rational drug designing.

Molecular docking is a widely-used computational tool for the study of molecular recognition, which aims to predict the binding mode and binding affinity of a complex formed by two or more constituent molecules with known structures. An important type of molecular docking is protein-ligand docking because of its therapeutic applications in modern structure-based drug design. Some advancements in docking include to a certain extent protein flexibility in the docking process, extensive exploration of the ligand conformation within the binding site, refinement and stability evaluation of the final complexes, and estimation of the binding free energies. The most popular docking programs include AutoDock[36] FlexX[37] GOLD[38] and GLIDE[39] among others. In our work we used the recent advances of protein flexibility, ligand sampling, and scoring functions the three important aspects in protein-ligand docking. We tried to cover the case where both the structure of the protein receptor and the binding site are known as well as, where there is lack of information of binding site.

The appropriate experimental techniques or computational methods to use will depend on the characteristics of the system itself and the information available. Thus, the outcomes of the docking procedures require a validation and a refinement via other techniques. Not surprisingly, molecular dynamics (MD) simulations have played a dominant role in these attempts to improve docking procedures. Indeed, fast and inexpensive

docking protocols can be combined with accurate but more costly MD techniques to predict more reliable protein–ligand complexes. The strength of this combination lies in their complementary strengths and weaknesses. One the one hand, docking techniques are used to explore the vast conformational space of ligands in a short time, allowing the scrutiny of large libraries of drug-like compounds at a reasonable cost. The major drawbacks are the lack, or poor flexibility of the protein, which is not permitted to adjust its conformation upon ligand binding, and the absence of a unique and widely applicable scoring function, necessary to generate a reliable ranking of the final complexes. On the other hand, MD simulations can treat both ligand and protein in a flexible way, allowing for an induced fit of the receptor-binding site around the newly introduced ligand. In addition, the effect of explicit water molecules can be studied directly, and very accurate binding free energies can be obtained. However, the main problems with MD simulations are that they are time-consuming and that the system can get trapped in local minima. Therefore, the combination of the two techniques in a protocol where docking is used for the fast screening of large libraries, generation of numerous binding poses and MD simulations are then applied to explore conformations of the protein receptor, optimize the structures of the final complexes, and calculate accurate energies, is a logical approach to improving the drug-design process.

However, the combination of the different techniques is a delicate issue in itself. For example, MD simulations might be very helpful also in the preparation phase of the docking and not only as an ex-post validation procedure. In fact, proteins in solution are mobile molecules, thus, they do not exist in a single conformation, but in a manifold of different conformational states separated by low-and higher-energy barriers. It is widely accepted that flexible loop regions have a critical functional role in enzymes. Lack of consideration of binding site flexibility has led to failures in predicting protein functions and in successful docking ligands with protein receptors. The distribution and stability of each conformational state of the protein will depend on the physicochemical properties of the environment and the protein itself (e.g., free or ligand-bound).[40] Moreover, not all these conformations will be equally able to bind productively with a given ligand. Some will be more likely to accommodate the ligand molecule within the binding site without having to undergo large changes, while others will be less likely, or even incapable, of accommodating the ligand due, for example, to loop conformations that block the access to the binding site[40]. The presence of the ligand itself is expected to affect the structure of the binding site and the dynamic equilibrium between different conformational states of the protein.[41] During a binding event, the protein conformer most likely to accommodate the ligand will be depleted from solution to form a ligand-bound complex, and other conformers will then adjust to fill the vacated conformational space, driving the binding process forward[42]. Therefore, an ensemble of receptor conformations and not a single structure is expected to provide a better representation of the system. Such Multiple state of protein can be captured via clustering. The clustering of protein with low RMSD provides different conformations of proteins states. Docking against several structures of the protein increases the chances of finding a receptor in the right conformational state to accommodate a particular ligand. However, it also reduces the selectivity of the docking process, as a wider variety of ligands will be able to fit in this more relaxed representation of the protein. It is important, therefore, to use accurate scoring functions during the final screening process to maximize selection of the most active ligands.

Some of the most successful docking schemes are based on a combination of the two approaches in a multistage process, where fast methods are used to screen a large number of ligands and accurate methods to refine the docking geometry and binding energy for the most promising candidates. The efficiency of the search and optimization methods used to find the global minimum of the ligand-receptor conformation energy is important for the success of the first group of methods and fundamental for the second group in which, given the greater complexity of the Hamiltonian, each energy evaluation is computationally expensive. However, another limit of the approach is present. In principle, a straightforward MD simulation[43], when based on a reasonably accurate force field and including solvation effects, should be able to find the docked geometry and the binding affinity. Unfortunately, most of the time this turns out to be not feasible, since the time spans that can be simulated are much shorter than the time necessary for the real binding process. For this reason, in the field of docking, MD is used in connection with some other method to sample the conformational space efficiently. Such methods include parallel tempering[44] stochastic tunneling[45] taboo search[46] multicanonical MD[47] umbrella sampling/weighted histogram analysis method[48] force probe MD[49-51] and molecular dynamics docking[52]. When the docked geometry is known, a different class of methods based on MD or MC can be used to predict the binding affinities. These use free energy perturbation, thermodynamic integration,[53-57] MM/PBSA,[58] and their variations.

The sampling method (metadynamics) [59] used in the present thesis has proved to be useful in various fields, including physics[60], biophysics[61-63], and statistical mechanics[59]. This method have been successfully adapted to docking, where it is able not only to find the docked geometry and predict the binding affinity (ΔG binding) but also to explore the whole docking process from the solution to the docking cavity, including barriers and intermediate minima. Docking with metadynamics [64,65] involve guessing the ligand poses using standard docking programs, relaxing the ligand/target complex by MD, and sorting out the poses by their binding energy using metadynamics. These steps have been able to identify the correct poses in few benchmarks (four in total) even in challenging cases (in two of the benchmark the ligands bind to a surface and in the others protein flexibility is important for docking) [66,67]. In view of these early results, combined strategies are promising docking protocols that might be further tested. Our work should be seen as a proof of concept of this combination of computational strategies.

The binding free energy of the process can also be quantitatively predicted. A challenge for the future is to obtain the correct docking pose without any prior knowledge of a putative binding site. The theoretical prediction of binding affinities is one of the most important problems in computational biochemistry. It complements experimental analysis and adds molecular insight to the macroscopic properties measured therein. It serves as a cornerstone in disease research and rational drug design where accurate scoring functions remain a challenge. It is no wonder, then, that computational models aimed at the prediction of binding affinities have been highly sought after for over half a century and are the subject of frequent reviews [68,69]. The theory underlying binding affinities has been well described by many, yet the complexity and accuracy of its application has varied. The most rigorous methods involve alchemical or structural transformations such as free energy perturbation and thermodynamic integration[70,71]. End-point free energy methods, such as the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) model, have received much attention and widespread application in recent literature. These methods benefit from computational efficiency as only the initial and final states of the system are evaluated. The MM-PBSA approach are being extensively used as 1) it does not contain any parameters that vary for different ligand receptor systems and it involves a set of physically well defined terms and 2) the solvent is treated implicitly and the electrostatic components are obtained from a dielectric continuum model with a dielectric constant for the solute and the solvent.

Motivation of the Work

The major part of my thesis focuses on ligands (newly synthesized and biologically potent), which bind to BVDV RdRp, targeting non-nucleoside analogs. Despite the latest methodological advances and the pharmacological relevance of NNIs drugs [8-24]. The interaction of NNIs with RdRp of BVDV has been less characterized. Only few recent theoretical studies have appeared in the literature (including some from our group) concerning the structure, the thermodynamics, the kinetics and the molecular recognition of RdRp binders. This contrasts with the relatively large number of studies on NNIs ligands on HCV RdRp. More importantly, many issues are still not fully understood, such as the contributions of structurally and functionally conserved regions of the RdRp, flexibility of the specific domain, identification of binding site, mechanism of inhibition and mechanism of drug resistance etc. Thus, there is still a large variety of questions unanswered at microscopic level, specially in case of BVDV RdRp. This information is essential for structural biology and drug design, as the latter relies on the assumption that beneficial effects of drugs origin from the interaction with its target.

The main objective of this Thesis is to address the above-mentioned challenges targeting BVDV RdRp, HCV RdRp and HIV-RT. Our study is in tune with the spirit of previous investigations performed in different group. Subsequently, I extended my investigation to the molecular recognition events involving several different class of compounds found active against BVDV, HCV and HIV. The molecules belong to Benzimidazole, Imidazoquinoline, Pyridoxoquinoline, Phenanthroline and DABO are used as a ligand to understand protein-ligand recognition. The focus has been on the calculation of free energies, the evaluation of enthalpic vs entropic contributions and the role of the solvent, the estimation of the weights of various interactions. The obtained results have yielded insights on similarities and differences in the molecular recognition process between the different kinds of RdRp ligands.

Organization of the thesis:

This thesis is organized as follows. Chapter 2 includes details on the main ACTORS of our study which include Polymerases of BVDV, HCV and HIV viruses and the promising Antiviral compounds. Thereafter, in Chapter 3 an extensive description of the methods used to solve the biological problem is provided. The rest of the thesis has the following chapters:

Chapter 4: Point mutation I261M affects the dynamics of BVDV and its interaction with benzimidazole antiviral 227G.

Key-abstract

Bovine viral diarrhea virus (BVDV) is a *Pestivirus* of the *Flaviviridae* family and represents a major viral pathogen in cattle and other ruminants. Infection with BVDV can result in a wide assortment of disease manifestations including resorption, mummification, or abortion of the dead fetus. Extensive study of BVDV is required not only because it causes heavy agronomic losses but also it is considered to be a valuable surrogate model for the study of the hepatitis C virus (HCV). Recently the point mutation I261M on the thumb domain was shown to confer resistance to BDVD against 227G and other benzimidazole compounds.

Here we investigated the role of this mutation by using a non conventional approach, which is not based on binding free energy calculations on structures of the mutated complex which are taken a priori similar to those of the wild one. Namely, we firstly performed MD simulations on the wild and mutated BVDV RdRp proteins in aqueous solution. Then, we selected representative equilibrium conformations by performing a cluster analysis, and ran docking calculations of 277G on representative of the 5 most populated clusters of each protein. Finally, high-score poses were subjected to MD simulations to assess structural and dynamical differences between wild and mutated 227G-protein adducts.

Interestingly, the mutation affects the structure and the dynamics of the protein, particularly in the region of binding of the ligand, and this results in different binding sites of 227G onto the two proteins. Moreover, while 227G closes the entrance for the RNA primer in the case of the wild protein, in presence of the mutation a gate and a channel leading to the catalytic site are still present. These results could furnish a possible molecular explanation of the resistance mechanism by mutation I261M.

In a related paper of us, we reported mutagenesis experiments aimed at confirming that BVDV RdRp is the target of our lead benzimidazole compound 227G. Here, we performed a multidisciplinary computational study to furnish a molecular-level explanation of effect of the resistant mutation. Our results are fully compatible with the mechanism of action furnished in the previous paper. Namely, we see that the mutation I261M alters the binding mode of the ligand, leaving the channels for the entrance of the template RNA, of NTP and the exit of double strand RNA all open. The dynamics of the protein is also altered to a less extent as compared to the wild type.

Results:

Experimental:

1) Site directed mutagenesis was done to identify the resistant mutation of 227G.

2) The mutated protein retains activity.

Computational:

1. Extensive docking studies were performed to identify the best orientation of 227G in COMPLEXm (227G in mutated BVDV). The effect of mutations was not calculated as in previous published works (i.e. in a static manner), but from a dynamical point of view.

2. Long MD simulations (total time ~120 ns) were performed to identify and to compare the effect of mutation on the structure of the APO and HOLO proteins with respect to the wild type one. The simulated systems are (30ns in each case): APO (apo-protein WT), APOm (apo-protein mutated), COMPLEX (WT

with 227G) and COMPLEXm (mutated protein with 227G).

3. Concerning the APO proteins, our simulations reveal:

    1. Significant structural changes in the region of binding, namely the linker region turns away from the binding pocket. This will have important consequences for the binding and the stability of the compound.

    2. The overall dynamics as well as the "functional channels" of the protein are similar, which is consistent with experiments showing that the mutant retains enzymatic activity.

4. Concerning the complexes:

    1. At opposite with COMPLEX, due to the movement of the linker (probably consequence of the presence of the long side chain of M261) in COMPLEXm, 227G is not stable in the initial docking pose, but moves and reorients itself many times until finds a stable position.
    2.In this position the drug is more solvent exposed, and accordingly to this its binding free energy is less. Moreover, in this new conformation the interaction between the drug and the linker and another loop (L3), which leaves the channel for the RNA template entrance opened (as well as the other channels). Thus, in principle the enzyme is still able to work, which still is consistent with experiments and can furnish a molecular-level rationale thereon.

    3. To further strengthen this hypothesis, we verified that also the dynamics of the APO and HOLO proteins are very similar in the mutant, while this is not true in the wild type, where the binding of the drug has a drastic effect.

    4. Concluding, only in the wild type 227G is able to alter structure and dynamics, so altering the functionality.

Chapter5 : Inhibition of viral RNA polymerase investigated by computer simulations

Key-Abstract:
The drug resistant of various RdRp inhibitors has been studied using a new computational protocol, that is, extensive molecular docking on cluster basis, leading to valuable insight into the resistance mechanisms and structure –resistance corrections of the RdRp inhibitors associated with mutation lying in the strategic location inside the motif-I. By using the cluster basis molecular docking method, the calculated mutation caused shifts of the binding free energies linearly correlate very well with those derived from the corresponding experimental data, suggesting that the newly proposed protocol may be used as generalized approach to predict drug resistance associated with mutation I261M. Because it is essentially important for understanding the structure-resistance correlation and for structure based drug design to develop an effective computational protocol for drug resistance prediction, the reasonable and computationally efficient protocol for drug resistance prediction should be valuable for future structure-based design and discovery of anti-resistance drugs in various therapeutic areas.

Result :

1. We develop a new protocol to identify and to validate a proper binding mode using molecular docking, molecular dynamics, clustering, MM-PBSA and Metadynamics techniques.
2. We tried to explore the rare events, as to pull out the drug from their binding site "escape mechanism" through a new algorithm "metaD". During escape or dissociation of drug from the cavity, we observed some local transient binding site, which was barely noticed by x-ray and other techniques. The trapping of the drug into these local transient binding pockets could be a reason of increase of residence time of drug into binding cavity.
3. We compare the escape mechanism of 227G from wild-227G complex and Mutated-227G complex, and we observed there is a very low barrier ~5kcal/mol was required for the 227G to come out from the binding site, which is ~12kcal/mol in case of wild-227G complex, could be a loss of entropy in wild-227G complex.

Chapter6: Combining Docking, Molecular Dynamics to Predict Binding Modes and Affinities for Non-nucleoside Inhibitors to HIV-1 Reverse Transcriptase

Abstract:

The reverse transcriptase of human immunodeficiency virus (HIV) catalyzes a series of reactions to convert the single-stranded RNA genome of HIV into double-stranded DNA for host-cell integration, This task requires the reverse transcriptase from the viral genome. Docking, scoring, molecular dynamics (MD), methods are used here to predict binding modes and affinities for a set of 4 non-nucleoside inhibitors to HIV-1 reverse transcriptase. The location of each drug outside the binding pocket was determined by an automated docking program, and steering into the binding pocket followed a route that is likely to represent the actual entrance pathway. The comparison of the dynamics of DABO with crystal structure of structurally similar compounds has provide information on possible molecular mechanisms of ligand binding, specificity and regulation of RT. The present calculations provide a validation of the combination of docking, MD as a powerful tool in structure-based drug design, and the methodology is easily scalable for attaining a higher throughput of compounds. This coherent picture strongly suggests that attempts to understand through the structure-based drug design may be considerably more successful if dynamic structural aspects of the type studied here are considered,particularly in those region which is more flexible but functionally important.

Results:

1)-To validate, how the binding of MC1220 at an allosteric site affects the conformational dynamics of RT such that it inhibits DNA polymerization.
2)-A comparative analysis have been performed to understand the triggering role of “X” and “Y” position in DABO series.
3)- The probable mechanism of inhibition was gel well with the hypothesis of inhibition of DABOs.
4)- MD simulation has been done in order to under stand the key determinants and binding pattern of DABO and its analogs.

Work not presented here.....

Chapter1: Different binding sites of the benzimidazole compound 227G on HCV and BVDV RdRps revealed by MD simulations.

Key-abstract

The virally encoded RdRp has emerged as a prime target in search for specific HCV and BVDV antiviral. Benzimidazole class of compound has been extensively pursued as potent Non-Nucleoside Inhibitors (NNI’s) against both classes of viruses. For the first time, Screening efforts in our lab have come out with potent NNI belonging to benzimidazole class of compound which is active against both HCV and BVDV RdRp. Our tested compound was confirmed to strongly inhibit BVDV RdRp activity in a dose-dependent manner and they also showed inhibition on HCV1b-NS5B.

While, the HCV RdRp structure in complex with the related analogs showed that these inhibitors bind at the surface of the thumb, the resistant mutation and newly identified binding site for the same analog were found to lye in the finger domain of BVDV RdRp. Thus, the molecular mechanism of inhibition of the inhibitor found to be different in the otherwise closely related polymerases.

Molecular dynamics simulations helped us to gain a deeper insight into the interaction pattern, mechanism of inhibition of inhibitor onto the two RdRps. The multidisciplinary approach allowed us to (i) critically assess the dogma of antiviral therapy; (ii) to screen efficiently the possible information valuable for designing effective and selective inhibitors by identifying at a molecular level the interaction pattern of the compound with its viral targets; (iii) to assess the use of BVDV as surrogate for the development of antivirals against HCV.

Results:

Experimental

1. The benzimidazole compound 227G is active against HCV and BVDV RdRps with low CC_{50}, EC_{50} and IC_{50} values. First time a benzimidazole is active against both RdRps.
2. Crystal structures of HCV in complex with indole compounds have been reported in the literature (other's works), while here, we have performed mutagenesis studies on BVDV showing RdRp to be the target of 227G.

Computational

1. Site identification: 227G, in HCV bind on thumb domain while in BVDV, a **novel** binding site was identified in the finger domain.
2. In HCV the binding mode of 227G turns out to be very similar to those of indole compounds. The binding of 227G disrupts the connection between the fingertip and the thumb domains, leading the enzyme to assume an open (and inactive) conformation. This is likely to be the mechanism of action by 227G.
3. In BVDV the binding site is different, as well is the mechanism of action. Indeed the drug binds to the finger domain and by interacting with the four loops from finger and thumb region completely closes the entrance of the channel to the RNA template. Performing different analyses has validated this finding.

Conclusion:

1. The benzimidazole derivative 227G is reported to be active against both HCV and BVDV at low micro-molar range.
2. Different Binding site :Despite, 227G, being active against both RdRp, its binding site on both the HCV and BVDV RdRp is different. In HCV RdRp it binds in a thumb domain while in BVDV RdRp it was found to bind in the finger domain of RdRp.
3. Mechanism of inhibition: We hypothesized the mechanism of inhibition of 227G in BVDV and HCV RdRp. In HCV RdRp, 227G binds in a thumb domain, disrupting the connection between the thumb and finger domain and thereby stop polymerization while in BVDV RdRp it blocks the entrance to the template channel and ultimately the inhibits te polymerisation.
4. The inventory of key interaction pattern between 227G and HCV and BVDV RdRp were obatined with an aim of providing possible hints at improving potency of 227G.

Chapter 2: Quinoline Tricyclic Derivatives Design, Synthesis and Preliminary *In vitro* and *In silico* Antiviral Activity Against Flaviviridae Family of Three New Classes of Virus-Encoded RNA-Dependent RNA Polymerase (RdRp) Inhibitors.

Key-abstract:

In this preliminary study three new classes of RNA-dependent RNA polymerase (RdRp) of Flaviviridae inhibitors, the linear N-tricyclic systems derived by condensation of the quinoline nucleus with 1,2,3-triazole, imidazole or pyrazine (obtaining triazolo[4,5-g]quinolines, imidazo[4,5-g]quinolines and pyrido[2,3-g]quinoxalines respectively), has been discovered. In particular the activity of the title compounds were evaluated in cell culture systems against YFV (as viruses representative of Flaviviruses) and BVDV (Pestiviruses) . Furthermore the cytotoxicity was evaluated in parallel cell-based assays. 1:1 Mixture of bis-triazoloquinolines (1m), imidazoquinolines (2e,h) and pyridoquinoxalines (4h,j and 5n) showed anti-BVDV activity in the range 1-5 μM in cell-based assays. Mutation experiments versus both NS5b and NS3 enzymes of resistant strains of BVDV of two selected compounds (2h and 5m) confirmed the inhibition of the enzyme target. Overall, by these experiments the imidazo[4,5-g]quinoline (2h) emerged as a potent BVDV polymerase inhibitor endowed with IC_{50} = of 0.06 μM. Herein, we have utilized several biophysical computational approaches, ranging from Docking, standard molecular dynamics and metadynamics , in order to predict the binding site, to identify key determinants of ligand binding, the energetics of unbinding as well as the escape mechanism of the lead compounds PS999 and PS1036 in BVDV RdRp. MM-PBSA techniques were used to calculate the binding free energies of 227G in both HCV and BVDV RdRp. Calculated binding free energies were then converted to

computational IC_{50} which were then compared with the experimentally available IC_{50} values.

Biological Results:

1. Compounds belonging to triazolo[4,5-*g*]quinolines, imidazo[4,5-*g*]quinolines and pyrido[2,3-*g*]quinoxalines were discovered and their activity were evaluated against YFV and BVDV viruses.
2. Mutation experiments confirmed BVDV RdRp as the target for the compounds

Computational results

1. Identified Binding cavity for leads belonging to pyridoxoquinoxaline(PS1036) as well as imidazoquinoline(PS999) derivative in BVDV RdRp. The choice for the study of the leads was based on the knowledge of the resistant mutations against these compound.
2. MD simulations on the complexes obtained from docking were performed. Key interaction responsible for the binding of compounds in the putative binding cavity were determined .
3. Metadynamics runs were performed to identify alternate binding poses other than that obtained from docking. Energetics of unbinding was estimated and pathway for the dissociation of the compounds was computed.
4. Binding energy of the stable poses obtained from metadynamics were calculated. The calculated IC_{50} value coincided well with the experimental IC_{50} values.

Conclusion

1. Conglomeration of resistant mutation of different classes of compounds(in this work, pyridoxoquinoxalines and imidazoquinoline) in the finger domain of BVDV RdRp indicated that the binding site of NNI may lye in the finger domain. We performed extensive computational studies and identified stable poses of the compounds in the finger domain of BVDV RdRp. Interstingly, 227G, a benzimidazole compound,(chapter 1) was found to bind in the same cavity. Thus we proposed that finger domain can be a “Hot-Spot” for NNI binding.

2. We proposed a similar hypothesis for mechanism of inhibition as in case of 227G. Since the compound target same binding site hence they can stop polymerisation by blocking the entrance to the template channel.

3. Extracted key binding features of compounds in binding pocket can provide basis for the development of more efficient antivirals.

Chapter-3: Identification of binding cavity for novel Non-Nucleodside inhibitors of HCV RdRp: A molecular docking study

Key-Abstract:

The virus encoded RdRp has emerged as a prime target in the search for specific HCV antivirals. Identification and successes of NNI, which are non-competitive diverse small molecules, against other polymerase, have encouraged them to be used as effective antivirals against HCV RdRp. Screening studies have came out with compounds belonging to Pyridoquinoxaline(PS1097) Imidazoquinolines(PS1126) and Phnanthroline (PS1101), effective against HCV RdRp at a very low micromolar range. Information regarding the binding cavity of these compounds are still lacking. Herein, we have utilized docking procedure to investigate binding sites, binding modes as well as binding affinity of these compounds in HCV RdRp. To rule out any bias, each of the three reported HCV NS5B NNI binding site represented by (A) Benzofurans (B) N,N-disubstituted phenylalanine and (C) benzothiadiazine inhibitors was examined for the binding of the compounds. We then used all atom standard molecular dynamics (MD) simulations to investigate the most probable binding site for the compounds in both complexes; dissociation was observed in two out of the three cavities under investigation. Therefore, Identifying the binding cavities and orientation of our compounds can help us to build a microscopically well-funded picture to elucidate the mode of action of compounds and thereby providing clues for rational drug design.

Results:

1. A blind docking procedure was applied in all four well known binding sites reported for different classes of compounds in HCV RdRp, to obtained the most prominent binding pocket for compound PS1097 (Pyridoxiquinoline) and compound PS1126 (Imidazoquinoline).
2. We selected the most stable docking pose for each binding site to do MD simulation for better characterization of their binding affinity as well as to refine the docking poses whether respective docked poses were stable in their respective binding pockets or not.
3. We performed 20ns of MD simulation of PS1097 at site C and PS1126 at site D, where we found these compounds are most stable with having Hbonds and HpH contacts.
4. We calculated the binding free energy of these compounds. From the binding free energy we obtained a IC_{50} value in very good agreement with the experimental one.
5. We compared the relevance of our findings, a new binding mode in two different binding sites with previously reported x-ray data targeted these binding sites, and we found a good match in form of interaction map.

Conclusion

1. We identified probable binding cavity for NNIs belonging to three different classes Pyridoxoquinoxaline , Imidazoquinoxaline and phenonthrolines. This was an encouraging performance, given that all molecular modeling studies were performed in the absence of any available crystal structure of the protein in complex with these classes of inhibitors.
2. Key binding pattern of the compounds in the cavity were identified with an aim of further optimisation of the lead compounds in order to increase their potency.
3. Calculated binding free energies were found to be in agreement with the experimental data.
4. Furthermore, we were successful in throwing light on the probable mechanism of inhibition of the three compounds reported in this paper.

Thesis in Flow-chart

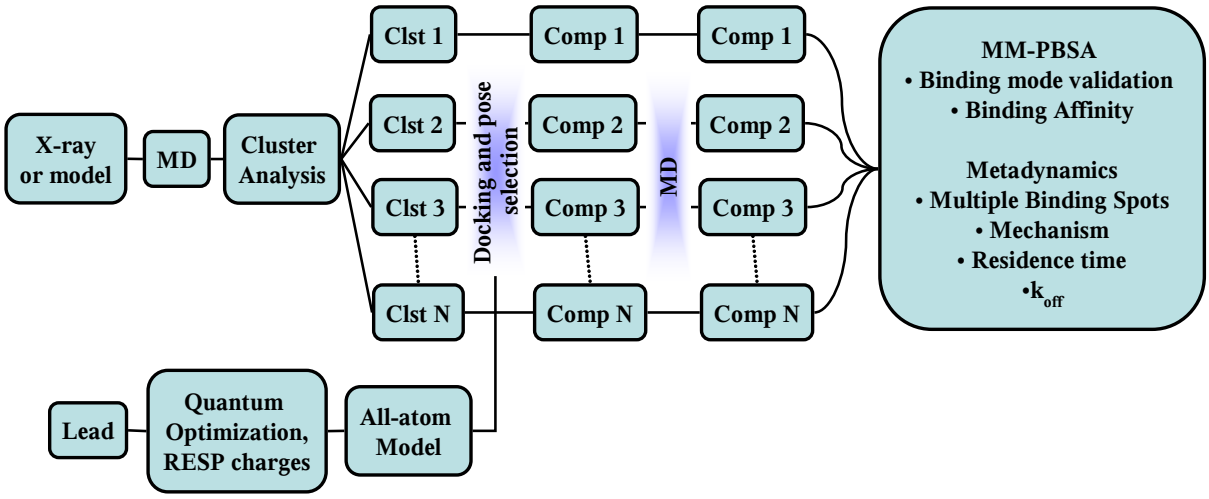
Antiviral Drug Development

Biological System (Polymerases)

BVDV(1S48) **HCV**(1NB4, 2BRK) **HIV** (1RT1,1DLO)

Class	Inhibitors	Class	Inhibitors	Class	Inhibitors
Benzi	$227G^{0.002}$	**Benzi**	$227G^{0.4}$	**DABOs**	$MC1220^{0.004}$
	$ASCC170^{3.0}$				$MC1332^{0.008}$
	$ASCC169^{12}$	**Imido**	$PS1126^{2.0}$		$MC1346^{0.005}$
Imido	$PS999^{0.6}$	**Pyrido**	$PS1097^{0.6}$		$MC1195^{0.05}$
Pyrido	$PS1036^{1.0}$	**Phenan**	$PS1101^{1.0}$		

Receptors and Antivirals; IC_{50} values in case if BVDV and HCV, while EC_{50} in Case of HIV-RT are in superscript form. The X-ray structures mainly used are written as superscript on receptors.



Computational methodology used to understand protein-ligand recognition.

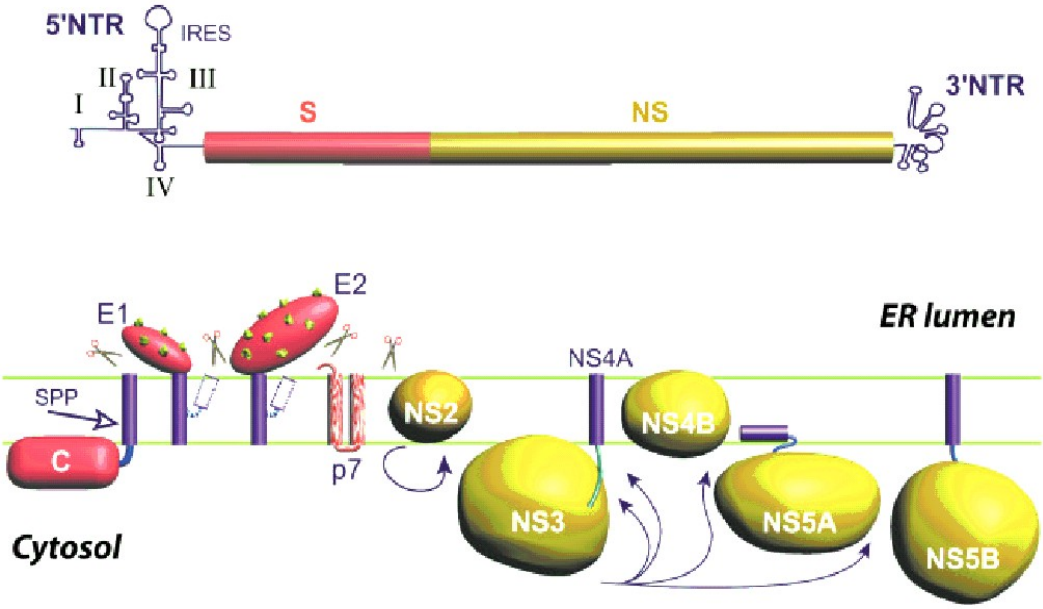
The Actors: Polymerases and Antivirals

Polymerases (HCV, BVDV and HIV)

"Infections by RNA viruses continue to exist as significant public health problems worldwide. So there is an urgent need for safer and more efficacious treatment option against infections caused by RNA viruses, researchers have devoted significant efforts over the last two decades to discovering and developing new antiviral agents" by Todd Apple.

Genomic organization:

Hepatitis C virus is a positive-sense single-stranded RNA virus with a genome of initiation of viral RNA translation. The HCV genomic RNA encodes a polyprotein of 3010–3011 amino acids which undergoes co-translational and post-translational proteolytic processing in the cytoplasm or in the endoplasmic reticulum of the infected cell to give rise to four structural and six non-structural (NS) proteins [72]. Figure4 summarizes the information concerning the identification and the function of individual gene products. The structural proteins consist of the capsid or core protein C, two envelope glycoproteins E1 and E2, and small hydrophobic polypeptide p7. The non-structural proteins are NS2, NS3, NS4A and NS4B, NS5A and NS5B. NS2 has a zinc-stimulated protease activity that cleaves the NS2/NS3 junction [73]. NS3, a 70 kDa protein, has been subject of intensive study because it is a multifunctional molecule with a trypsin-like serine protease catalytic domain at the N-terminal 181 residues [74], and a nucleoside triphosphatase (NTPase)/RNA about 9.6 kb [75]. It consists of a single open-reading frame (ORF) flanked by 5' and 3' non-translated regions (NTRs) [76]. The 5' NTR contains the internal ribosome-entry site (IRES). NS4A is a small protein (8 kDa) that is an NS3 protease cofactor [77]. NS4B is a hydrophobic, 27 kDa protein, and it has been suggested that it may function as an anchor to secure part of the HCV replication apparatus to the endoplasmic reticulum (ER) membrane [78]. Two cytoplasmic phosphoproteins, p56 and p58 (56 kDa and 58 kDa, respectively) are produced from the NS5A region of the HCV genome [79,80]. Both proteins are phosphorylated at serine residues in the region between amino acids 2200 and 2250 and in the C-terminal region of NS5A. Although the function of NS5A in viral replication is unknown, it is possible that NS5A might be involved in the resistance of HCV to a-interferon therapy [81]. The purified enzyme NS5B exhibits a highly processive primer-independent RNA-dependent RNA polymerase (RdRp) activity capable of copying in vitro transcribed full-length genomic HCV RNA [82].



Amino acids	Protein	Functional
1-191	C	nucleocapsid
192-383	E1	Envelope glycoprotein
384-746	E2	Envelope glycoprotein
747-809	P7	virioporin
810-1026	NS2	Zn-activated NS2/3 auto-protease
1027-1657	NS3	Ser protease (aa 1-180)
		RNA helicase (aa 181-631)
1658-1711	NS4A	NS3 Ser protease co-factor (aa 21-34)
1712-1972	NS4B	Induces the formation of intracellular membrane ves.
1973-2420	NS5A	alpha-interferon resistance
2420-3010	NS5B	RNA-dependent-RNA-polymerase

Figure4: HCV genome organization (top) and polyprotein processing (bottom). The 5 untranslated region (UTR) consists of four highly structured domains and contains the internal ribosome entry site (IRES). The 3 UTR consists of stable stem-loop structures and an internal poly(U)/polypyrimidine tract. The central 9.6-kb ORF codes for a polyprotein of slightly more than 3000 aa depending on the HCV genotype. S and NS correspond to regions coding for structural and nonstructural proteins, respectively. The polyprotein processing and the location of the 10 HCV proteins relative to the ER membrane are schematically represented. Scissors indicate ER signal peptidase cleavage sites; cyclic arrow, autocatalytic cleavage of the NS2-NS3 junction; black arrows, NS3/NS4A proteinase complex cleavage sites; intramembrane arrow, cleavage by the signal peptide peptidase. The transmembrane domains of E1 and E2 are shown after signal-peptidase cleavage and reorientation of the respective C-terminus hydrophobic stretches (dotted rectangles). Green spots denote glycosylation sites of the E1 and E2 envelope proteins. Adapted from : Hepatology. 2004 Jan;39(1):5-19.Structural biology of hepatitis C virus.Penin F, Dubuisson J, Rey FA, Moradpour D, Pawlotsky JM.

BVDV Genome

Pestiviruses have a positive sense single stranded RNA genome [ss (+) RNA]. The BVDV genome has a size of approximately 12'500 nucleotides, the sequence of which is known since 1988. Some cp strains integrate small variable segments of viral nucleic acid or the host cell genome in certain places of their genome (in NS2 or between NS2 and NS3). Others show duplicates of specific protein coding regions (Npro, NS3), consequently their genome size rises to about 16.5 kb. The genomic RNA has one open reading frame of about 4000 codons, which encompasses most of the viral genome. Translation of the BVDV genome yields one precursor poly protein, which is cleaved co- and postranslationally by viral and host cell encoded proteases (”processing”). Most of the virally encoded cleaving is catalysed by a serine protease domain within NS3 and generates the non-structural proteins NS3 to NS5B, whereas the structural proteins are believed to be cleaved by cellular proteases.(Figure4)

HIV Genome

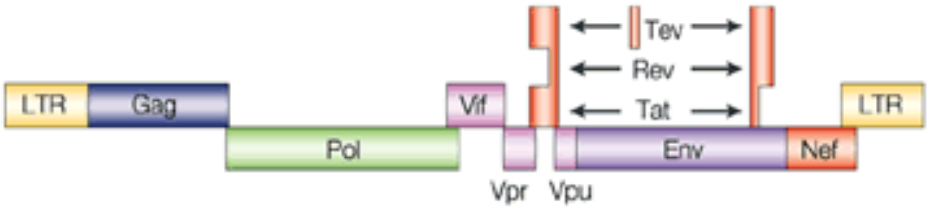


Figure5: Figure depicts HIV genome organization.

Human Immunodeficiency Virus type 1 (HIV-1) is a single stranded +RNA lentivirus from the Retroviridae family. Genomic organization is listed as below (Figure 5)

GAG The genomic region encoding the capsid proteins (group specific antigens). The precursor is the p55 myristoylated protein, which is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCapsid), and p6 proteins, by the viral protease. Gag associates with the plasma membrane, where virus assembly takes place. The 55-kDa Gag precursor is called assemblin to indicate its role in viral assembly.

POL The genomic region encoding the viral enzymes protease, reverse transcriptase, and integrase. These enzymes are produced as a Gag-Pol precursor polyprotein, which is processed by the viral protease; the Gag-Pol precursor is produced by ribosome frameshifting near the 3' end of *gag*.

ENV Viral glycoproteins produced as a precursor (gp160), which is processed to give a noncovalent complex of the external glycoprotein gp120 and the transmembrane glycoprotein gp41. The *env* gene does not actually code for gp120 and gp41, but for a precursor to both, gp160. During HIV reproduction, the host cell's own enzymes cleave gp160 into gp120 and gp41. The host cell protease that cleaves Env into gp120 and gp41 is Furin.

TAT Transactivator of HIV gene expression. Tat vastly increases the level of transcription of the HIV dsRNA. Before Tat is present, a small number of RNA transcripts will be made, which allow the Tat protein to be produced. Tat then binds to cellular factors and mediates their phosphorylation, resulting in increased transcription of all HIV genes, providing a positive feedback cycle.

REV The second necessary regulatory factor for HIV expression. A 19-kD phosphoprotein, localized primarily in the nucleolus/nucleus, Rev acts by binding to RRE and promoting the nuclear export, stabilization, and utilization of the viral mRNAs containing RRE. Rev is considered the most functionally conserved regulatory

protein of lentiviruses. Rev cycles rapidly between the nucleus and the cytoplasm.

VIF Viral infectivity factor, a basic protein typically 23 kD. Promotes the infectivity but not the production of viral particles. In the absence of Vif, the produced viral particles are defective, while the cell-to-cell transmission of virus is not affected significantly. Found in almost all lentiviruses, Vif is a cytoplasmic protein, existing in both a soluble cytosolic form and a membrane-associated form. The latter form of Vif is a peripheral membrane protein that is tightly associated with the cytoplasmic side of cellular membranes.

VPR Vpr (viral protein R) is a 96-amino acid (14-kD) protein, which is incorporated into the virion. It interacts with the p6 Gag part of the Pr55 Gag precursor. Vpr detected in the cell is localized to the nucleus. Proposed functions for Vpr include the targeting the nuclear import of preintegration complexes, cell growth arrest, transactivation of cellular genes, and induction of cellular differentiation. In HIV-2, SIV-SMM, SIV-RCM, SIV-MND-2, and SIV-DRL the Vpx gene is apparently the result of a Vpr gene duplication event, possibly by recombination.

VPU Vpu (viral protein U) is unique to HIV-1, SIVcpz (the closest SIV relative of HIV-1), SIV-GSN, SIV-MUS, SIV-MON and SIV-DEN. There is no similar gene in HIV-2, SIV-SMM, or other SIVs. Vpu is a 16-kd (81-amino acid) type I integral membrane protein with at least two different biological functions: (a) degradation of CD4 in the endoplasmic reticulum, and (b) enhancement of virion release from the plasma membrane of HIV-1-infected cells. Env and Vpu are expressed from a bicistronic mRNA. Vpu probably possesses an N-terminal hydrophobic membrane anchor and a hydrophilic moiety.

NEF A multifunctional 27-kd myristoylated protein produced by an ORF located at the 3' end of the primate lentiviruses. Other forms of Nef are known, including nonmyristoylated variants. Nef is predominantly cytoplasmic and associated with the plasma membrane via the myristoyl residue linked to the conserved second amino acid (Gly).

VPX A virion protein of 12 kD found in HIV-2, SIV-SMM, SIV-RCM, SIV-MND-2, and SIV-DRL and not in HIV-1 or other SIVs. This accessory gene is a homolog of HIV-1 vpr, and viruses with vpx carry both vpr and vpx. Vpx function in relation to Vpr is not fully elucidated; both are incorporated into virions at levels comparable to Gag proteins through interactions with Gag p6. Vpx is necessary for efficient replication of SIV-SMM in PBMCs.Figure5 depicts HIV genome organization.

Domain Architecture (HCV, BVDV and HIV)

Viral polymerase

It has been already mentioned in the introduction that Polymerase (RdRp in BVDV and HCV, RT in HIV) are considered as one of the most attractive target for the development of effective antivirals, and is a focus of this thesis. Herein this section, we are providing details on the structural organization of the protein with special reference to the different domains. Information regarding the catalytic site and special structural features are also provided here. An in depth understanding of the structural organization of protein is fundamental for a rational approach to designing of new antivirals.

Structure of polymerase subdomains: HCV

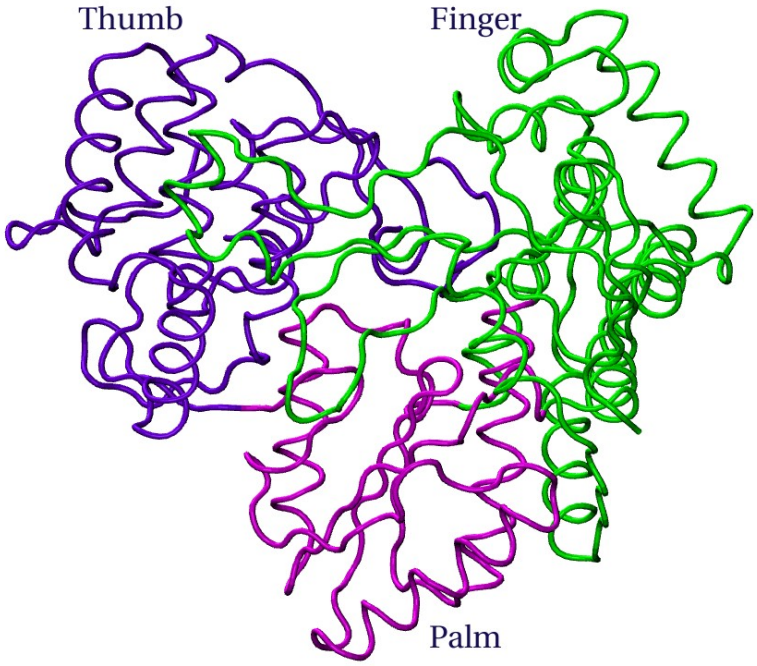


Figure6: Cartoon representation of HCV RdRp. Different domains of RdRp viz thumb in blue, finger in green, and palm in magenta.

A major advance in the understanding of the NS5B polymerase was provided by the resolution of the three-dimensional structures of several truncated forms of the apoenzyme and of complexes with nucleotides or RNA templates [83-88]. The NS5B has the canonical 'right hand' shape, with the characteristic fingers, palm and thumb sub-domains (Figure 6). Similarly to other RdRps, the HCV polymerase has a compact shape due to the presence of two extended loops, the fingertips, that connect the fingers and thumb domains and completely encircle the active site cavity, to which the RNA template and the nucleoside triphosphate (NTP) substrates have access through two positively charged tunnels [85].

The **palm** sub-domain (residue 188 to 227, 287 to 370) constitutes the catalytic site of the molecule and contains the residues responsible for the nucleotidyl transfer reaction within the D-(X)4-D and the GDD motif. The two-metal-ion **catalytic site** is conserved essentially in all known viral polymerases.[89-91] The magnesium ions assist phosphodiester bond formation by polarizing the hydroxyl group at the 3′ terminus of the growing RNA chain, facilitating nucleophilic attack upon the dNTP substrate in addition to stabilizing the transition state in which the phosphorus of the dNTP is linked to five oxygen atoms. The residues essential for metal binding are Asp220 and the carbonyl of the Thr221 peptide backbone, as well as Asp318 and Asp319.

Residues 1 to 187 and 228 to 286 constitute the **finger** sub-domain. The unique shape of the HCV NS5B RNA-dependent RNA polymerase (RdRp) arises from extended packing interactions between the fingers and thumb sub-domains. This inter domain linkage is provided by insertions within the fingers sub-domain. These insertions form two loops that pack against the thumb sub-domain (Figure6) The longer loop is located near the N-terminus and consists of residues 11–45; the shorter includes residues 139–160. These loops regulate the coordinated movements of the fingers and thumb during the polymerase reaction cycle.[83]

The **thumb** sub-domain of NS5B includes residues (371–528). Two structural elements peculiar to the NS5B structure are a **β-hairpin** (β-loop), protruding from the thumb into the active site, and a **C- terminal** region, located immediately before the transmembrane domain, that folds from the surface of the thumb towards the active site and establishes a series of hydrophobic interactions with a shallow pocket comprised between the palm and thumb sub-domains[89]. These elements (β-hairpin and C-terminus) are probably involved in positioning the 3′ terminus of template RNA and are considered crucial for template selection.[89]

BVDV

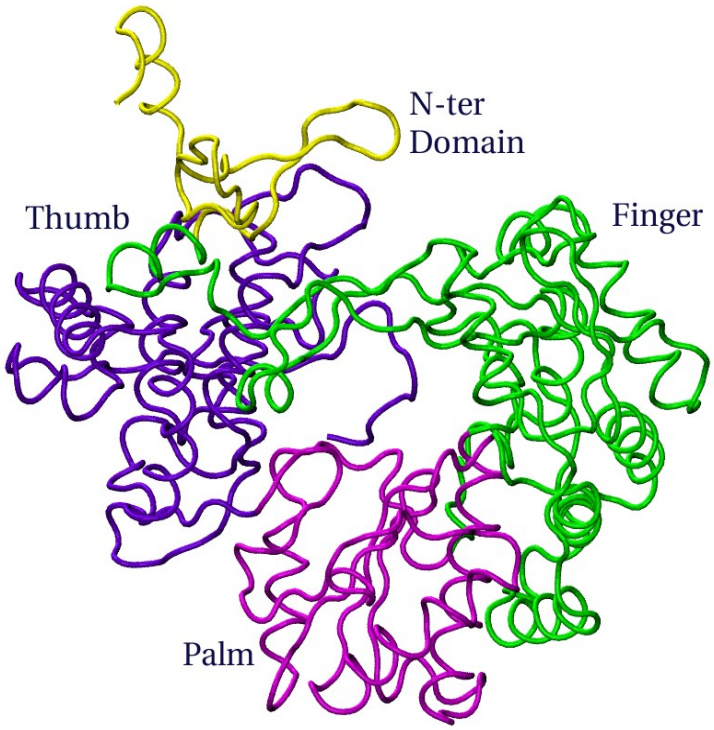


Figure7: Cartoon representation of BVDV RdRp. Different domains of RdRp viz thumb in blue, finger in green, palm in magenta and unique N-terminal is yellow in color.

The BVDV polymerase core (residues 139–679) has a roughly spherical shape. Similar to the HCV RdRp, BVDV RdRp comprises of three sub-domain, that is, finger, palm and thumb (Figure 7). In addition, there is an N-terminal domain (residues 71–138), which is considered to be unique in BVDV RdRp.

The **finger** domain (residues 139–313 and 351–410) consists of 12 α-helices and 11 β-strands (ref1). In BVDV RdRp, as in other viral RdRps, the N terminus of the finger domain, together with a long insert in the fingers domain (residues 260–288), forms a fingertip region that associates with the thumb domain [92]. Among RdRps the conformation of the polypeptide chain in the fingertip region differs. For example, whereas BVDV polymerase has a three-stranded fingertip region, HCV [93-97] and calicivirus polymerases [98] have a four-stranded, and ϕ6 polymerase a six-stranded fingertip region [99]. The finger and thumb domains are linked through the fingertip region, thus the conformational change induced by template binding into the central

channel may be limited [94,95]. The remainder of the finger domain is comprised of a β-strand-rich region (β-fingers) and an α-helix-rich region (α-fingers) close to the palm domain. Although there is little sequence identity, the topology of the fingers domain, excluding the finger-tips, is similar to that of other viral polymerases, making possible the structural alignment of the BVDV, HCV, calicivirus, and ϕ6 polymerase fingers domains. As expected, BVDV polymerase has greater structural similarity to HCV polymerase, indicating a closer evolutionary link between theses two viruses than with the other viruses.[choi-2004]

The **palm** domain is the catalytic domain and shows the greatest structural conservation among all known polymerases. The palm domain of BVDV polymerase (residues 314–350, 411–500) consists of a four-stranded β-sheet flanked by two α-helices on one side and an additional α-helix on the other side of the β-sheet, nested between the palm and thumb domains. Compared to HCV polymerase, BVDV polymerase has a small α-helix (residues 433–439) inserted immediately before the first β-strand [84] of the β-sheet. A comparative analysis of RNA polymerase sequences in positive-strand RNA viruses has identified eight conserved sequence motifs, I-VIII [100,101]. Five of these motifs are in the palm domain, including the Gly-Asp-Asp (GDD) motif VI, which is essential for catalytic activity [102].

The **thumb** domain, consisting of the C-terminal region of the polypeptide chain, is the most diverse feature among the known polymerase structures. The thumb domain of BVDV polymerase (residues 501–679) contains eight α-helices and five β-strands and is larger than the thumb domain of other viral RdRps Although there is some structural similarity between the HCV and BVDV thumb domains, the overall topology is rather different. The BVDV polymerase thumb domain possesses a structural element (“the β-thumb region”) in which two β-strands and their connecting loops form a layer that occludes the active site cavity. The β-thumb region interacts with the fingers and palm domains through a long C-terminal loop (residues 670–679), tucked between the palm and thumb domains. Together with a loop in the thumb domain, the β-thumb reduces the volume of the template channel. A similar protrusion into the active site, formed by a β-hairpin, is present in the HCV polymerase structure, although the protrusion originates from a different part of the thumb domain .

N-Terminal Domain.
At its N terminus the BVDV RdRp is 130 residues longer than HCV RdRp. The function of this N-terminal domain is not known, although up to 90 residues can be truncated from BVDV polymerase without loss of polymerase activity in vitro [102]. It has been suggested that the N-terminal region in other polymerases is required for binding proteins in the replication complex [103,104]. The ordered part of the N-terminal region (residues 92–138) folds into a separate domain. The N-terminal domain is situated over the thumb domain, interacting with the fingertip region and thumb domain through a β-hairpin motif. The β-hairpin motif points toward the template channel and partially occludes the channel entrance, possibly explaining why some N-terminal truncated proteins have greater polymerase activity than wild type [102]. The entrance to the template-binding channel created by the hairpin motif of the N-terminal domain is highly positively charged. An equivalent positively charged surface at the entrance to the channel was found in ϕ6 polymerase, where it was proposed to act as a “plough” that unwinds the double-stranded RNA, thus aiding the function of a helicase [99]. Similarly in BVDV polymerase, this positively charged surface might be used to open up RNA secondary structural hairpins before the ss-RNA template entering the active center.

HIV-RT

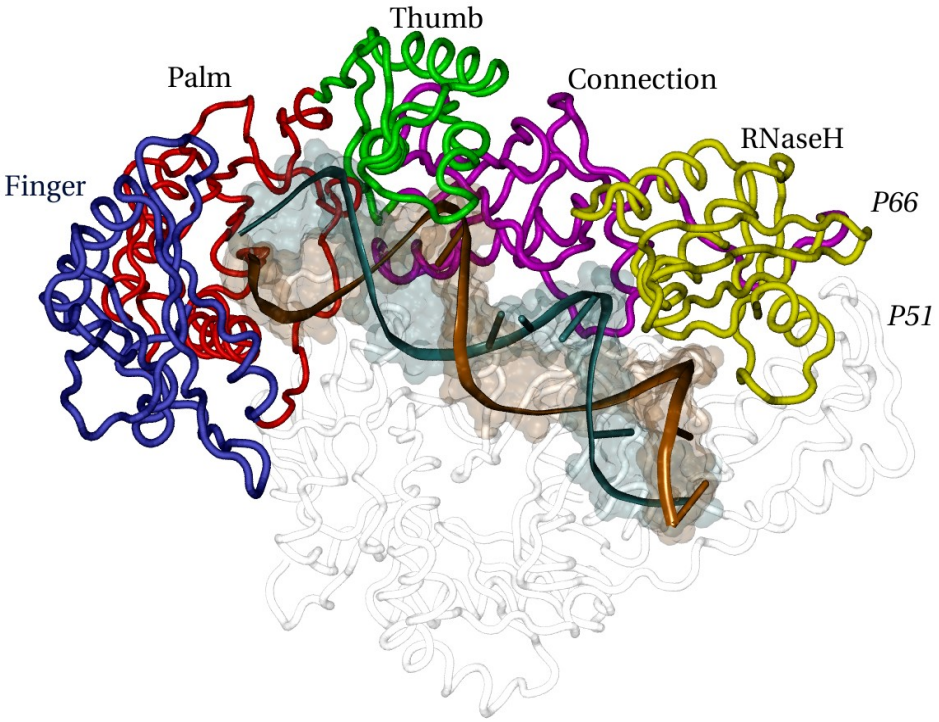


Figure8: Cartoon representation of HIV-RT. Two subunits P66 (colored according to the subdomains) and P51(white) are depicted here. Finger is represented in blue, Palm in red, Thumb in green, the connection in magenta and RNaseH in yellow. The template is also represented here with the two strands being colored in orange and cyan.

HIV-1 RT catalyzes the transcription of the single-stranded RNA viral genome into a double-stranded DNA form, which can be integrated into the human genome as the provirus. A wealth of crystallographic studies has been performed on RT, yielding ~ 60 high-resolution structures of the protein in a variety of states (as reviewed in Ref. [105] and documented in Ref. [106]). These include the “apo” form (no substrate and no inhibitor, e.g., Protein Data Bank (PDB) code 1DLO [107] and 1HMV[108], substrate-bound forms (binary complexes of protein with nucleic acid substrate, and the ternary complex of protein with nucleic acid and nucleoside triphosphate, e.g., PDB codes 2HMI[109] and 1RTD[110], and NNRTI-bound forms (co-crystallized NNRTIs, but no substrate, e.g., PDB codes 1BQM[111] and 1EP4[112]). RT is a 1000-amino-acid heterodimer of p66 and p51 subunits, each of them composed of “fingers,” “thumb,” “palm,” and “connection” subdomains (see Figure 8). The spatial arrangement of these subdomains is very different between the two subunits and it is thought that the p51 subunit mainly plays a structural role, with polymerization occurring at the p66 subunit.[105] Together, the fingers, palm, and thumb subdomains of p66 resemble a right hand and form a “clamp” that holds the double-stranded template–primer in position. Notably, the palm subdomain contains the “catalytic triad” of three aspartates (Asp110, Asp185, and Asp186), which are essential for the addition of nucleotide to the growing primer strand, and the so-called “primer grip,” which is though to be required for correct positioning of the 3′ end of the primer.[113] The p66 subunit contains an additional subdomain, known as the RNase H (RNH) domain, which is responsible for the other enzymatic activity of the protein degradation of the RNA strand during polymerization.

Antivirals, a small journey

HCV/BVDV

To date, the majority of studied antiviral drugs targets viral polymerases as the primary mechanism of action. Both biochemical and cell-based replicon assays have been employed in the identification and optimization of novel HCV and BVDV RdRp inhibitors. RdRp inhibitors reported can be divided into two groups: (i) nucleoside analogues, and (ii) non-nucleoside inhibitors (NNIs). Nucleoside analogues prevent replication of HCV RNA, blocking the elongation of new viral RNA strands, they generally target the polymerase active site in a competitive manner and tipically exhibit broader spectrum activity. Non-nucleoside counterparts inhibit the HCV polymerase enzyme itself and they have much greater specificity, acting by either interfering directly with the active site or binding to the allosteric site and preventing the initiation process.

Nucleoside inhibitors

Nucleoside inhibitors, whether chain terminators or non-chain terminators, can be effective in inhibiting the virus replication. Upon entry into the cells, nucleoside analogues are first converted to nucleotide triphosphates (NTP). The unnatural nucleoside inhibitors can serve as competitive substrates for the polymerase and can be incorporated into the nascent chain by the viral polymerase. This incorporation can lead to premature termination of the elongation process. The incorporated nucleotides may cause base mismatch in subsequent rounds of replication, resulting in accumulated mutations in the viral genome and then the so-called error catastrophe. Ribavirin-is a nucleoside analogue and it is the only one currently in use in HCV therapy, though its precise mechanism of action remains unclear and ribavirin used as monotherapy has no durable antiviral efficacy. Several sugar-modified and base-modified nucleoside analogues have been reported to inhibit the RdRp enzymatic activity and block HCV replication in the replicon systems. The nucleoside analogues, inhibiting NS5B polymerase activity, have mostly 2′-methylribose structures and act as chain terminator [114,115]. Chain terminators have proven to be effective against viral DNA polymerases, and a number of these compounds have been approved for treatment of HIV and hepatitis B. Nucleoside inhibitors are less prone to drug resistance induced by mutations than non-nucleoside inhibitors because they bind to the enzyme's active site, which is more conserved and less genetically variable than the allosteric (non-catalytic) sites bound by non-nucleosides. Anyway mutation may not be the only resistance mechanism. Researcher, led by Matthias Gotte at McGill University in Montreal, just published evidence for a new mechanism of viral resistance to nucleoside inhibitors: pyrophosphorolytic excision.The RNA polymerase, in the presence of cellular pyrophosphatase, can literary cut the nucleoside analogue drug off the chain, allowing the chain to resume growth. How much this happens in HCV is not yet known.

Several compounds, based on dioxolane triphosphate and nucleoside phosphate, have been reported [116,117]. β-D-2′-methyribofuranosyl guanosine is the most prominent, since it can be phosphorylated *in vivo* and orally administered. The only nucleoside analogue shown to have efficacy in a clinical study is NM 283 (Figure9, left panel) (Valopicitabine; University of Cagliari/Idenix/Novartis) [118]. A component of NM 283 is the orally administrable 3'-O-valine ester of 2′-C-methyl-cytidine, NM 107 (Figure9 right panel). NM 107, once phosphorylated intracellularly to its 5' triphosphate metabolite, is initially known to be effective against BVDV. Later, a replicon cell lines-based assay revealed it to be effective against HCV. NM 283 is effective against replicons originating from different HCV strains. Replicons resistant to 2'-C-methylcytidine contain the S282T mutation in the viral polymerase and show a reduced fitness. NM 283, combined with pegylated IFN, exhibited improved efficacy and adequate tolerability compared to current therapies for chronic hepatitis C patients.

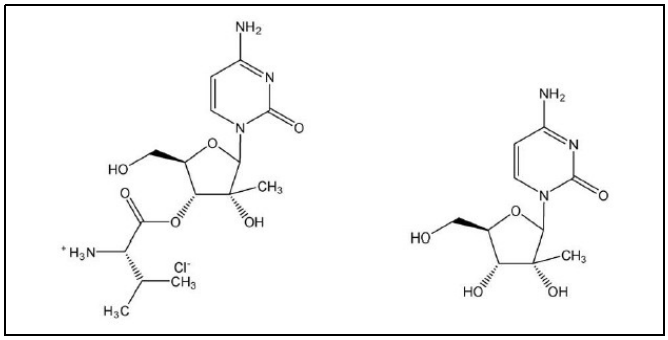


Figure9: The structures of NM 283 (left) and NM 107 (right)

Apart from NM283 other 2′-methyl nucleosides that selectively inhibit of HCV replication are 2′-O-methyl-cytidine, 2′-C-methyl-adeno-sine, 2′-C-methyl-guanosine, and 7-deaza-2′-C-methyl-adeno-sine. These are clinically less advanced, but some of them have been extensively characterized in vitro as well as in preclinical models of HCV infection [119-121]. In particular, the high inhibitory potency and reduced cellular toxicity of 7-deaza-2′-C-methyl-adenosine, combined with a very promising pharmacokinetic profile in preclinical animal species, makes this compound an attractive candidate for clinical development [122].

Lastly, RG7128 (Roche/Pharmasset), a 2′-modified nucleo-side analogue, has recently entered early clinical trials. RG7128 is an oral prodrug of 2′-deoxy-2′-fluoro-2′-C-methyl-cytidine, a pyrimidine nucleoside analogue that was shown to act as a non-obligate chain terminator of the NS5B polymerase activity [123] and to efficiently inhibit the replication of HCV RNA in hepatoma cell lines[Table 1] [124].

A 4′-modified nucleoside, 4′-azido-cytidine, was recently reported as a potent competitive inhibitor of NS5B-dependent RNA synthesis and hepatitis C virus replication in cell culture [125]. This compound was shown to inhibit RNA synthesis in a competitive manner and to be a moderately potent inhibitor of the replicon system. R1626 will enter a phase II trial at the two lower doses in combination with PEG-IFN with or without ribavirin.

Non-nucleoside inhibitors

Non-nucleoside inhibitors are the most diverse of the known HCV NS5B inhibitors. They act mostly as allosteric inhibitors and interfere with the conformational transition during the initiation of RNA synthesis [126]. In fact, besides the active site, the X-ray derived co-crystal structures of compounds bound to NS5B revealed distinct allosteric regulatory sites that are located distant to the active site and are targets for developing antiviral agents. Accordingly, compounds that interact either with the RdRp active site or the allosteric site could potentially interfere with substrate binding and/or conformational changes, thus effectively inhibiting initiation.

Drug Name / Category	**Drug Name / Category**	**Pharmaceutical Company**	**Clinical Phase**
RG7128 (Polymerase Inhibitor)	RG7227 (ITMN-191) (Danoprevir) Protease Inhibitor	Genentech in collaboration with Pharmasset & InterMune	Phase I
Telaprevir Protease Inhibitor	VX-222 Polymerase Inhibitor	Vertex	Phase II
IDX375	Polymerase Inhibitor	Idenix	Phase I
ABT-072	Polymerase Inhibitor	Abbott	Phase I
MK-3281	Polymerase Inhibitor	Merck	Phase I
PSI-7851	Polymerase Inhibitor	Pharmasset	Phase I
ABT-333	Polymerase Inhibitor	Abbott	Phase I
VX-916	HCV Polymerase	Vertex	Phase I
Filibuvir (PF-00868554)	HCV Polymerase Inhibitor	Pfizer	Phase I
RG7128	Polymerase Inhibitor	Pharmasset/Genentech	Phase I
VX-222	Polymerase Inhibitor	Vertex	Phase II
IDX184	Polymerase Inhibitor	Idenix	Phase II
ANA598	Polymerase Inhibitor	Anadys Pharmaceuticals	Phase II
GS 9190	Polymerase Inhibitor	Gilead	Phase II
VX-759	Polymerase Inhibitor	Vertex	Phase II
PSI-7977	Polymerase Inhibitor	Pharmasset	Phase IIa

Table.1: Approved antiviral drugs against HCV. Adapted from: HCV Advocate; July 24, 2010;Hepatitis C treatment in current clinical development; Alan Franciscus.

Several classes of structurally distinct NNIs of the HCV RdRp have been identified and disclosed; these included but are not limited to benzothiadiazines, benzimidazoles/diamides, disubstituted phenylalanine/thiophene amides and substituted pyranones. Benzimidazole derivatives were the first non-nucleoside inhibitors that demonstrated to be active against the HCV polymerase by Japan Tobacco[127]. One representative compound showed an EC_{50} value in the submicromolar range. Investigation of the mechanism of action indicated that the compound did not compete with incorporation of NTP. Moreover, mutations, conferring resistance to these compounds, were mapped to proline residue 495, which is located on the surface of the polymerase thumb domain and away from the active site, suggesting that the compounds acted as allosteric inhibitors, by blocking the activity of the polymerase prior to the elongation step. The first NNIs of HCV, that entered clinical trials, were JTK-109 and JTK-003 (Japan Tobacco). These benzimidazole derivatives act as allosteric inhibitors and block the polymerase before elongation. Resistance of HCV replicons to this class of compounds maps to P495. Co-crystallization studies of the HCV polymerase, with a related analogue, suggests that these compounds bind on the surface of the thumb domain at a position that is in the 'closed' enzyme occupied by one of the fingertips. The binding of the compound to this cavity forces the enzyme in the 'open' configuration, annihilating its functionality. Boeringer Ingelheim reported a series of benzimidazole-containing heterocycles, by extending the original derivatives to topologically related scaffolds and incorporating an amide moiety inside the molecule. Interestingly, subsequent substitution of the benzimidazole with a pyrazolopyrimidine moiety led to active compounds with an EC_{50} less than 1 μM.

Thiophene derivatives are reversible allosteric inhibitors of the enzyme. Replicons resistant to the compounds contain mutation at position 419 and 423 at the base of the thumb, which is different from the position where the JT-compounds bind. Binding of the thiophene analogues to this allosteric site also results in a conformational change of the enzyme that may explain the antiviral activity.

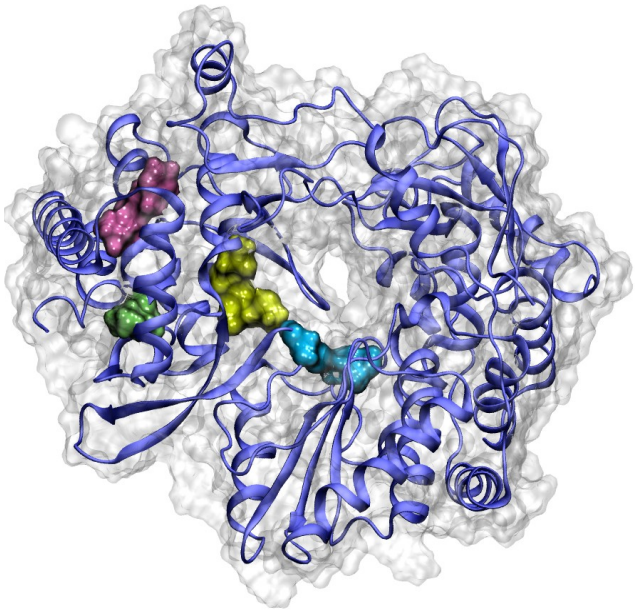


Figure10. Ribbon representation of the overall structure of HCV NS5B polymerase in blue along with transparent surface view in white. Solid surface representations of four allosteric binding sites are also shown. The dark red colored surface corresponds to the thumb allosteric binding pocket of indole group "siteA"; the dark green colored surface corresponds to the 2nd thumb allosteric binding site "site B", the dark yellow colored surface corresponds to the allosteric binding site situated in the palm domain "site C" and the dark skyblue colored surface correspond to 2nd palm domain binding site "site D".

In the literature several Benzothiadiazines were reported that target the viral polymerase. Although this class of compounds induce a variety of mutation in the enzyme, only methionine 414 is believed to be part of the inhibitor binding site. This residue is located in the inner surface of the thumb domain, close to the catalytic site of the enzyme. The allosteric binding site of the Benzothiadiazine is thus again different from that of Thiophene and Benzimidazole derivatives. Cross-resistance studies and synergistic inhibition of HCV polymerase, by combination of a Benzimidazole and a Benzothiadiazine, further confirmed that these two structurally distinct classes of inhibitors had non-overlapping binding sites and thus acted with different mode of action. Heterocyclic Benzo-1,2,4-thiadiazine derivatives are specific inhibitors of RNA synthesis in HCV replicon systems [131]. GlaxoSmithKline (formerly SmithKline Beecham) disclosed a novel class of Benzothiadiazine derivatives, from which a representative compound displayed good potency both in biochemical assay, with an EC_{50} value of 80 nM and HCV replicon assay with an EC_{50} value of 500 nM. Moreover, the Benzothiadiazine derivatives were shown to be highly selective for the HCV RdRp, failing to inhibit other viral and mammalian polymerases. Treatment with the compound and IFN-α resulted in a highly synergistic effect in the replicon system. Further study demonstrated that the compounds are non-competitive for NTP incorporation and act to arrest *de novo* initiation of RNA synthesis prior to the elongation phase, possibly through interacting with the functionally critical NS5B polymerase active site. Substitution of the quinolinone moiety with pyrrolone group led to compounds with greater potency.

Researcher at Shire Biochem disclosed two classes of HCV polymerase inhibitors, including phenylalanine and thiophene carboxylate derivatives. Further studies demonstrated that the inhibitors bound to a allosteric site in the thumb domain, disticnt from the binding site of benzimidazole. In addition, other promising non-nucleoside inhibitors are in different phases of clinical trials. R803, a small molecule HCV RdRp inhibitor, developed by Rigel Pharmaceuticals, was found to be active in the replicon system, with EC_{50} below to 10 μM. R803 entered into a multi-dose phase I/II clinical study for chronic HCV infection, but recently the study has been cancelled. HCV-086, another orally available small molecule inhibitor of NS5B polymerase, was co-developed by Viropharma and Wyeth. Results from a phase 1b study demonstrated that HCV-086 possessed favourable pharmacokinetics and was generally safe and well tolerated. However, the overall antiviral activity of HCV-086 did not warrant further development. The follow-on compound HCV-796, a novel non-nucleoside HCV polymerase inhibitor, is being evaluated in ongoing clinical trials in combination with PEG-IFN. HCV-796 is well tolerated with no dose limiting toxicities and it displayed dose-dependent antiviral activity across multiple genotypes. The most significant and sustained reduction in viral load from baseline was observed on patient infected with genotype 1b HCV. Anyway, in August Viro Pharma and its development partner Wyeth announced discontinuation of phase 2 dosing of the polymerase inhibitor HCV-796, because of elevated liver enzymes in some patients. During 2007, companies, including seven biotechs, abandoned or suspended at least eight antiviral drugs in clinical development. The disappointments are not new: drugs specifically targeting the hepatitis C virus have been in the works since the early 1990s, but no compound of this kind has yet advanced to phase 3 trials. Four of the failed drugs were inhibitors of NS5B RNA polymerase. GlaxoSmithKline, in London, and XTL Biopharmaceuticals, of Valley Cottage, New York, discontinued their polymerase inhibitors in phase 1. These failures are a serious setback for the field, but they do not indicate a general drawback of the entire class of compounds. In fact, each drug has its own problem and has its own characteristics. At least three polymerase inhibitors remain in clinical development. A part from the molecules that are known to have entered clinical trials, several other NNIs of HCV are in advanced preclinical development and early results for the newer polymerase inhibitors look promising. A series of α,γ-diketoacid compounds, as inhibitor of HCV polymerase, has been identified. Further optimization led to identification of a potent HCV NS5B polymerase inhibitor, with an EC_{50} of 45nM. Compounds of dihydroxypyrimidine carboxylic acid class are believed to chelate the two catalytic Mg^{2+} ions in the active site as diketoacid compounds. However, no data of *in vivo* activity for the pyrophosphate mimics are available and the high ionic nature of these compounds may raise concerns, such as low bioavailability and toxicity. The binding sites for (2s)-2-[(2,4-dichloro-benzoyl)-(3-trifluoromethylbenzyl)-amino]-3-phenyl-propionic acid and (2s)-2-[(5-Benzofuran-2-ylthiophen-2-ylmethyl)-(2,4-dichlorobenzoyl)-amino]-3-phenylpropionic acid are located in the thumb subdomain 35 Å far from the active site. A benzimidazole inhibitor and N,N disubstituted phenylalanine inhibitor also induced an allosteric interaction [106]. There are 3 classes of inhibitors that bind to a pocket in the thumb domain. Phene 2-carboxylic acids, phenylalanine derivatives and cyclopentyl dihydropyran-2-ones all bind to methionine-423 [132]. Other non-nucleoside inhibitors, known to bind to methionine 423 in the palm and thumb domain of the active site cleft, include benzothiadiazine [131] and benzimidazole 5-carboxamide. Benzimidazole 5-carboxamide inhibits the initiation phase and acts in a non-

competitive manner with respect to NPT incorporation [132].

Natural and synthetic coumarins have multiple biological activities. For instance, they have been claimed to be useful as anticoagulant, antibacterial, anti-inflammatory, anticancer and anti-HIV agents. Very few are reported to be able to counteract HCV. Only some coumarin derivatives like Osthole (7-methoxy-8-prenylcoumarin) is capable of inhibiting HCV replication and/or proliferation. In particular, Osthole is able to counteract the progression of hepatitis C into hepatocarcinoma. Since the great variability of pharmacological response of coumarins is connected to large modification of their substituents, Mazzei and co-workers (Department of Pharmaceutical Sciences, Genova, Italy) found interesting to synthesize some coumarin Mannich bases in order to examine their ability to inhibit HCV replication. This idea proved to be a winning strategy: in fact, a number of synthesized compounds showed moderate, but significant activity against Flaviviridae family, paving the way for new coumarins to be used in the hepatitis C infection [133].

Non-nucleoside active against BVDV

VP32947 [134](Baginski et al., 2000), mizoribine [135](Yanagida et al., 2004), [136]BPIP (Paeshuyse et al., 2006), Acridones [137](Tabarrini etal., 2006), AG110 [138](Paeshuyse et al., 2007), SC-560[139] (Okamoto et al., 2009), iminosugar derivatives [140](Chang et al., 2009), LZ37 [141](Paeshuyse et al., 2009), BIT225 [142](Luscombe et al., 2010). γ-carboline [143](Sako et al.,2008; Aoyama et al., 2009; Mohammed T.A. Salim et al. 2010).[145-147]

HIV-RT

In a journey of 25 years, HIV-1, the retrovirus responsible for the acquired immunodeficiency syndrome (AIDS), has gone from being an "inherently untreatable" infectious agent to one eminently susceptible to a range of approved therapies. HIV epidemic fueled development of new antiviral drug classes, based on advances in the understanding of the viral life cycle, have transformed what used to be a rapid and lethal infection into a chronic condition that can be controlled for many years through combination of therapies with different classes of antiviral drugs — known as highly active antiretroviral therapy (HAART).

It all started in 1985, two years after the identification of human immunodeficiency virus (HIV) [120] and one year after the initial evidence about its etiological link to AIDS was reported [148,149]. Samuel Broder's group at the National Cancer Institute together with collaborators from Burroughs-Welcome company, identified 3-azidothymidine (AZT,zidovudine) as the first nucleoside inhibitor with in vitro anti-HIV activity.

The discovery of the anti-HIV activity of AZT was a landmark in the development of antivirals, providing the first proof of concept that the replication of HIV could be controlled by chemotherapy and thereby establishing the foundation of antiretroviral drug discovery research [150]. Thus, AZT became the first nucleoside HIV reverse transcriptase inhibitor (NRTI). Over the course of 25 years that followed after this seminal discovery, seven nucleosides have been approved by the United States Food and Drug Administration(FDA) for the treatment of HIV infection starting with the approval of AZT in 1987 and followed by didanosine (ddI), zalcitabine (ddC), stavudine (d4T), lamivudine (3TC), abacavir (ABC), tenofovir disoproxil fumarate (Table 1) Despite the approval of NRTIs as effective therapies for HIV-1/AIDS, side-effects of these antiretrovirals were found to be real and certainly not to be discounted.

Members of the of NRTIs were eventually joined by nonnucleoside RT inhibitors (NNRTIs), which were discovered in 1990, and interact with an allosteric binding site on HIV-1 reverse transcriptase that becomes exposed upon ligand binding[151]. NNRTIs are a key part of typical HAART regimes for treatment-naive patients (two NRTIs and one NNRTI), owing to their potency, favorable safety profile and ease of dosing. However, the relatively rapid emergence of resistance, resulting from mutations of residues that surround the NNRTI binding site (in particular K103N and Y181C), is a serious limitation. In addition to the three NNRTIs — nevirapine, delavirdine and efavirenz — that have been approved for the treatment of HIV, a few more are in

clinical development[152], including rilpivirine, etravirine and dapivirine. Their mechanism of action is similar to that of the approved NNRTIs in that they interact with a specific binding site of the reverse transcriptase, thereby blocking the enzyme's activity. It has recently been demonstrated that NNRTI binding to the polymerase domain of the reverse transcriptase interferes with RNase H activity, and that mutations in the NNRTI binding site (K103N, Y181C, Y188L and K103N/Y181C) reduce the potency of RNase H inhibition[152].

Despite progress in the treatment of HIV, there is still considerable room for improvement and expansion of antiviral drugs. Antiretroviral therapy has brought about a substantial decrease in the death rate due to HIV-1 infection, changing it from a rapidly lethal disease into a chronic manageable condition, compatible with very long survival. This has special implications within the classic boundaries of public health around the world, but at the same time in certain regions may also affect a cycle of economic and civil instability in which HIV-1/AIDS is both cause and consequence. Many challenges remain, including 1) the life-long duration of therapy; 2) the ultimate role of pre-exposure prophylaxis (PrEP); 3) the cardiometabolic side-effects or other toxicities of long-term therapy; 4) the emergence of drug-resistance and viral genetic diversity (non-B subtypes); 5) the specter of new cross-species transmissions from established retroviral reservoirs in apes and Old World monkeys; and 6) the continued pace of new HIV-1 infections in many parts of the world. All of these factors make refining current therapies and developing new therapeutic paradigms essential priorities. Fortunately, there are exciting new insights into the biology of HIV-1, its interaction with cellular resistance factors, and novel points of attack for future therapies. The current science will lead to new therapeutic strategies with far-reaching implications in the HIV-1/AIDS pandemic[153,154]

Brand name	Generic name(s)	Manufacturer name	Approval date	Time to approval
Nucleoside reverse transcriptase inhibitors (NRTIs)[a,b]				
Retrovir	Zidovudine, azidothymidine, AZT, ZDV	GlaxoSmithKline (original sponsor Burroughs-Wellcome)	19 March 1987	3.5 months
Videx	Didanosine, dideoxyinosine, ddI	Bristol Myers-Squibb	9 October 1991	6 months
Hivid	Zalcitabine, dideoxycytidine, ddC (no longer marked as of December 31, 2006)	Hoffmann-La Roche	19 June 1992	7.6 months
Zerit	Stavudine, d4T	Bristol Myers-Squibb	24 June 1994	5.9 months
Epivir	Lamivudine, 3TC	GlaxoSmithKline	17 November 1995	4.4 months
Combivir	Lamivudine and zidovudine	GlaxoSmithKline	27 September 1997	3.9 months
Ziagen	Abacavir sulfate, ABC	GlaxoSmithKline	17 December 1998	5.8 months
Videx EC	Enteric coated didanosine, ddI EC	Bristol Myers-Squibb	31 October 2000	9 months
Trizivir	Abacavir, zidovudine, and lamivudine	GlaxoSmithKline	14 November 2000	10.9 months
Viread	Tenofovir disoproxil fumarate, TDF	Gilead Sciences	26 October 2001	5.9 months
Emtriva	Emtricitabine, FTC	Gilead Sciences	02 July 2003	10 months
Epzicom	Abacavir and lamivudine	GlaxoSmithKline	02 August 2004	10 months
Truvada	Tenofovir disoproxil fumarate and emtricitabine	Gilead Sciences	02 August 2004	5 months
Nonnucleoside reverse transcriptase inhibitors (NNRTIs)[c]				
Viramune	Nevirapine, NVP	Boehringer Ingelheim	21 June 1996	3.9 months
Rescriptor	Delavirdine, DLV	Pfizer	4 April 1997	8.7 months
Sustiva	Efavirenz, EFV	Bristol Myers-Squibb	17 September 1998	3.2 months
Intelence	Etravirine	Tibotec Therapeutics	18 Jane 2008	6 months

Table. 2: Approved antiretroviral drugs. Adapted from: Drugs Used in the Treatment of HIV Infection, U.S. FDA, http://www.fda.gov/oashi/aids/virals.html. Drugs are listed in order of FDA approval within each class.

Reference :

1. Prusoff WH (1959) Synthesis and biological activities of iododeoxyuridine, an analog of thymidine. Biochim Biophys Acta 32:295–296
2. Prusoff WH, Chen MS, Fischer PH, Lin TS, Shiau GT, Schinazi RF, Walker J (1979) Antiviral iodinated pyrimidine deoxyribonucleosides: 5-iodo-2 -deoxyuridine; 5-iodo-2-deoxycytidine;5-iodo-5 -amino-2 , 5 -dideoxyuridine. Pharmacol Ther 7:1–34
3. De Clercq E (2007) The design of drugs for HIV and HCV. Nat Rev Drug Discov 6:1001–1018
4. Gallant JE, DeJesus E, Arribas JR, Pozniak AL, Gazzard B, Campo RE, Lu B, McColl D, Chuck S, Enejosa J, Toole JJ, Cheng AK (2006) Tenofovir DF, emtricitabine, and efavirenz vs. zidovu dine, lamivudine, and efavirenz for HIV. N Engl J Med 354:251–260
5. Schinazi RF (1991) Combined therapeutic modalities for viral infections – rationale and clinical potential. In: Chou TC, Rideout DC (eds) Synergism and antagonism in chemotherapy. Academic, Orlando, FL, pp 110–181
6. de Clercq, Erik; Field, Hugh J (5 October 2005). "Antiviral prodrugs — the development of successful prodrug strategies for antiviral chemotherapy". British Journal of Pharmacology (Wiley-Blackwell) 147 (1): pp.1–11. January 2006.
7. Choo, Q. L., G. Kuo, et al. (1989). "Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome." Science 244(4902): 359-362.
8. Saito I Fau - Miyamura, T., A. Miyamura T Fau - Ohbayashi, et al. (1990). "Hepatitis C virus infection is associated with the development of hepatocellular carcinoma." PNAS 87(17): 6547–6549.(0027-8424)
9. Cohen, J. (1999). "The Scientific Challenge of Hepatitis C." Science 285(5424): 26-30.
10. De Francesco, R. and A. Carfí (2007). "Advances in the development of new therapeutic agents targeting the NS3-4A serine protease or the NS5B RNA-dependent RNA polymerase of the hepatitis C virus." Advanced Drug Delivery Reviews 59(12): 1242-1262.
11. Shi, S. T., K. J. Herlihy, et al. (2008). "In Vitro Resistance Study of AG-021541, a Novel Nonnucleoside Inhibitor of the Hepatitis C Virus RNA-Dependent RNA Polymerase." Antimicrob. Agents Chemother. 52(2): 675-683.
12. Ghany MG, Strader DB, Thomas DL, Seeff LB; American Association for the Study of Liver Diseases. Diagnosis, management, and treatment of hepatitis C: an update. Hepatology 49, 1335–1374 (2009).
13. Erica Weiskircher13, Jason Aligo1, Gang Ning2 and Kouacou V Konan1, 2009.Bovine viral diarrhea virus NS4B protein is an integral membrane protein associated with Golgi markers and rearranged host membranes.Virology Journal 2009, 6:185
14. E.J Richey.(2009).http://www.cattlenetwork.com/Cattle-Health--Treatment--Prevention---Control-Of-BVD/2009-06-01/Article.aspx?oid=496856&fid=VN-ANIMAL_HEALTH-BVD-ARTICLES
15. Daniel L. Grooms; M. Daniel Givens; Michael W. Sanderson;Bradley J. White; MS; Dale M. Groteluesche; David R. Smith.BVD Virus Control and Eradication - Academy of Veterinary Consultants
16. S.S. Carroll, J.E. Tomassini, M. Bosserman, K. Getty, M.W. Stahlhut, A.B. Eldrup, B. Bhat, D. Hall, A.L. Simcoe, R. LaFemina, C.A. Rutkowski, B. Wolanski, Z. Yang, G. Migliaccio, R. De Francesco, L.C. Kuo, M. MacCoss, D.B. Olsen, Inhibition of hepatitis C virus RNA replication by 2¢-modified nucleoside analogs, J. Biol. Chem. 278 (14) (2003) 11979–11984.
17. D.B. Olsen, A.B. Eldrup, L. Bartholomew, B. Bhat, M.R. Bosserman, A. Ceccacci, L.F. Colwell, J.F. Fay, O.A. Flores, K.L. Getty, J.A. Grobler, R.L. LaFemina, E.J. Markel, G. Migliaccio, M. Prhavc, M.W. Stahlhut, J.E.Tomassini, M. MacCoss, D.J. Hazuda, S.S. Carroll, A 7-deaza-adenosine analog is a potent and selective inhibitor of hepatitis C virus replication with excellent pharmacokinetic properties, Antimicrob. Agents Chemother.48 (10) (2004) 3944–3953.
18. P.A. Furman, E. Murakami, H. Bao, J. Symons, M.J. Otto, Inhibition of HCV replication by PSI-6130:mechanism of biochemical activation and inhibition, J. Hepatol. 46 (Suppl 1) (2007) S224.
19. World Health Organization. World Health Statistics 2009: Cause of spe cific mortality and morbidity. Available at: http://www.who.int/whosis/ whostat/EN_WHS09_Table2.pdf. Accessed on Dec 6, 2009.
20. Wheeler W, Mahle K, Bodnar U, et al. Antiretroviral drug resistance mutations and subtypes in drug-naïve persons newly diagnosed with HIV-1 infection, United States, Mar 2003–Oct 2006. 14th Conference on Retroviruses and Opportunistic Infections (CROI 2007). 2007 Feb 25–28; Los Angeles, CA.

21. Das K, Sarafianos SG, Arnold E, Hughes SH, William JL, Lane MD. Encyclopedia of Biological Chemistry. Elsevier; New York: 2004. HIV-1 Reverse Transcriptase Structure; pp. 388–392.
22. Erik De Clercq; New developments in anti-HIV chemotherapy; Volume 56, Issues 1-2, 1 March 2001, Pages 3-12
23. Cerutti H, Casas-Mollano JA. On the origin and functions of RNA-mediated silencing: from protists to man; Curr Genet. 2006 Aug;50(2):81-99. Epub 2006 May 12.
24. Cihlar, T., Ray, A., 2010. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. Antiviral Res. 85, 39–58.
25. Martin, J.C., Hitchcock, M.J.M., De Clercq, E., Prusoff, W.H., 2010. Early nucleoside reverse transcriptase inhibitors for the treatment of HIV: a brief history of Stavudine (D4T) and its comparison with other dideoxynucleosides. Antiviral Res. 85, 34–38.
26. Venter, J. et al. The Sequence of the Human Genome. Science 291, 1304–1351 (2001).
27. Collins, F. et al. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945 (2004)
28. Hebert, P., Cywinska, A., Ball, S. & R deWaard, J. Biological identifications through DNA barcodes. Proceedings- Royal Society of London. Biological sciences 270, 313- 321 (2003).
29. Stoeckle, M. Taxonomy, DNA, and the Bar Code of Life. BioScience 53, 796–797(2003).
30. Hoppe, W., Lohmann, W., Markl, H. & Ziegler, H. Biophysik (1982).
31. Jung, J. & Lee, W. Structure-based Functional Discovery of Proteins: Structural Proteomics. Journal of Biochemistry and Molecular Biology 37, 28–34 (2004).
32. Bringer, A. & Nilges, M. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. Quarterly Reviews of Biophysics 26, 49–125 (1993).
33. Nilges, M. Structure calculation from NMR data. Current Opinion in Structural Biology 6, 617–623 (1996).
34. Frauenfelder, H., Sligar, S. & Wolynes, P. The energy landscapes and motions of proteins. Science 254, 1598–1603 (1991).
35. Cho AE, Guallar V, Berne BJ, Friesner R. Importance of accurate charges in molecular docking:Quantum mechanical/molecular mechanical (QM/MM) approach. J Comput Chem 2005;26: 915 – 931.
36. Morris, G. et al. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. Journal of Computational Chemistry 19, 1639–1662 (1998).
37. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. Journal of Molecular Biology 261, 470–489 (1996).
38. Jones, G., Willett, P., Glen, R., Leach, A. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology 267, 727–748 (1997).
39. Friesner,R.A., Banks,J.L., Murphy,R.B., Halgren,T.A., Klicic,J.J., Mainz,D.T., Repasky,M.P., Knoll,E.H., Shelley,M., Perry,J.K. et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem., 47, 1739–1749.
40. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R. Folding and binding cascades: Dynamic landscapes and population shifts. Protein Sci 2000;9:10 – 19.
41. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng 1999;12:713–720.
42. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 2002;11:184–197.
43. Wang, J.; Dixon, R.; Kollman, P. A. Proteins: Struct., Funct., Genet. 1999, 34, 69-81.
44. Merlitz, H.; Wenzel, W. Chem. Phys. Lett. 2002, 362, 271-277.
45. Merlitz, H.; Burghardt, B.; Wenzel, W. Chem. Phys. Lett. 2003, 370, 68-73
46. Cvijovic, D.; Klinowski, J. Science 1995, 267, 664-666.
47. Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Chem. Phys. Lett. 1997, 278, 297-301.
48. Kumar, S.; Payne, P. W.; Vasquez, M. J. Comput. Chem. 1996, 17, 1269-1275.

49. Heymann, B. A.; Grubmuller, H. Biophys. J. 2001, 61, 1295-1313.
50. Heymann, B. A.; Grubmuller, H. Phys. ReV. Lett. 2000, 84, 6126-6129
51. Grubmuller, H.; Heymann, B. A.; Tavan, P. Science 1996, 271, 997-999
52. Mangoni, M.; Roccatano, D.; Di Nola, A. Proteins: Struct., Funct., Genet. 1999, 35, 153-162.
53. Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A.Science 1987, 235, 574-576.
54. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. Science 1987,236, 564-568.
55. Sneddon, S. F.; Tobias, D. J.; Brooks, C. L., III. J. Mol. Biol. 1989, 209, 817-820
56. Straatsma, T. P.; McCammon, J. A. Annu. ReV. Phys. Chem. 1992, 43, 407-435.
57. Kong, X.; Brooks, C. L. B., III. J. Chem. Phys. 1996, 105, 2414-2423.Kong, X.; Brooks, C. L. B., III. J. Chem. Phys. 1996, 105, 2414-2423.
58. Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Biophys. J. 2004, 86, 67-74
59. Laio, A.; Parrinello, M. Proc Natl. Acad. Sci. U.S.A. 2002, 99, 12562-12566.
60. Iannuzzi, M.; Laio, A.; Parrinello, M. Phys. ReV. Lett. 2003, 90, 238302.
61. Ceccarelli, M.; Danelon, C.; Laio, A.; Parrinello, M. Biophys J. 2004, 87, 58-64.
62. Micheletti, C.; Laio, A.; Parrinello, M. Phys. ReV. Lett. 2004, 170601.
63. Kranjc A, Bongarzone S, Rossetti G, Biarnes X, Cavalli A, Bolognesi ML, Roberti M, Legname G, Carloni P: Docking ligands on protein surfaces: the case study of prion protein. J Chem Theory Comput 2009, 5:2565-2573.
64. Masetti M, Cavalli A, Recanatini M, Gervasio FL: Exploring complex protein-ligand recognition mechanisms with coarse metadynamics. J Phys Chem B 2009, 113:4807-4816.
65. Ceccarelli M, Anedda R, Casu M, Ruggerone P: CO escape from myoglobin with metadynamics simulations. Proteins 2008, 3:1231-1236.
66. Vargiu AV, Ruggerone P, Magistrato A, Carloni P: Dissociation of minor groove binders from DNA: insights from metadynamics simulations. Nucleic Acids Res 2008, 36:5910-5921.
67. Ajay, A., and M. A. Murcko. 1995. Computational methods to predict binding free energy in ligand-receptor complexes. J. Med. Chem. 38: 4953–4967.
68. McCammon, J. A. 1998. Theory of biomolecular recognition. Curr. Opin. Struct. Biol. 8:245–249.
69. Simonson, T., G. Archontis, and M. Karplus. 2002. Free energy simulations come of age: protein-ligand recognition. Acct. Chem. Res. 35:430–437.
70. Beveridge, D. L., and F. M. DiCapua. 1989. Free-energy via molecular simulation—applications to chemical and biomolecular systems. Annu. Rev. Biophys. Bio. 18:431–492.
71. Straatsma, T. P., and J. A. McCammon. 1992. Computational alchemy. Annu. Rev. Phys. Chem. 43:407–435.
72. Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohno K. (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc Natl Acad Sci USA.; 88: 5547–51.
73. Love RA, Parge HE, Wickersham JA, Hostomsky Z, Habuka N, Moomaw EW, Adachi T, Hostomska Z. (1996) The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. Cell.; 87: 331–42.
74. Tomei L, Failla C, Santolini E, De Francesco R, La Monica N. (1993) NS3 is a serine protease required for processing of hepatitis C virus polyprotein. J Virol.; 67: 4017–26.
75. Takamizawa A, Mori C, Fuke I, Manabe S, Murakami S, Fujita J, Onishi E, Andoh T, Yoshida I, Okayama H. (1991) Structure and organisation of the hepatitis C virus genome isolated from human carriers. J Virol.; 65: 1105–113.
76. Tanaka T, Kato N, Cho M-J, Shimotohno K. (1995) A novel sequence found at the 3'terminus of hepatitis C virus genome. Biochem Biophys Res Commun.; 215: 744–9.
77. Failla C, Tomei L, De Francesco R. (1994) Both NS3 and NS4A are required for proteolytic processing of hepatitis C virus nonstructural proteins. J Virol.; 68: 3753–60.
78. Lundin M, Monne M, Widell A, Von Heijne G, Persson MA. (2003) Topology of the membrane-associated hepatitis C virus protein NS4B. J Virol.; 77: 5428–38.

79. Kaneko T, Tanji Y, Satoh S, Hijikata M, Asabe S, Kimura K, Shimotohno K. (1994) Production of two phosphoproteins from the NS5A region of the hepatitis C viral genome. Biochem Biophys Res Commun.; 205: 320–6.
80. Neddermann P, Quintavalle M, Di Pietro Ch, Clementi A, Cerretani M, Altamura S, Bartholomew L, De Francesco R. (2004) Reduction of hepatitis C virus NS5A hyperphosphorylation by selective inhibition of cellular kinases activates viral RNA replication in cell culture. J Virol.; 78: 13306–14.
81. Tan S-L, Katze MG. (2001) How hepatitis C virus counteracts the interferon response: The jury is still out on NS5A.Virology.; 284: 1–12.
82. Oh J-W, Ito T, Lai MC. (1999) A recombinant hepatitis C virus RNA-dependent RNA polymerase capable of copying the full-length viral RNA. J Virol.; 73: 7694–702.
83. Adachi T, Ago H, Habuka N, Okuda K, Komatsu M, Ikeda S & Yatsunami K (2002) The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase. Biochemica et Biophysica Acta 1601:38–48.
84. Ago H, Adachi T, Yoshida A, Yamamoto M, Habuka N, Yatsunami K & Miyano M (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Structure 7:1417–1426.
85. Bressanelli S, Tomei L, Rey FA & De Francesco R (2002) Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. Journal of Virology 76:3482–3492.
86. Bressanelli S, Tomei L, Roussel A, Incitti I, Vitale RL, Mathieu M, De Francesco R & Rey FA (1999) Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proceedings of the National Academy of Sciences, USA 96:13034–13039.
87. Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF & Weber PC (1999) Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nature Structural Biology 6:937–943.
88. O'Farrell D, Trowbridge R, Rowlands D & Jager J (2003) Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and de novo initiation. Journal of Molecular Biology 326:1025–1035.
89. Hansen, J. L.; Long, A. M.; Schultz, S. C. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure 1997, 5, 1109-1122.
90. Doublie, S.; Sawaya, M. R.; Ellenberger, T. An open and closed case for all polymerases. Struct. Fold. Des. 1999, 7, R31-R35.
91. O'Farrell, D.; Trowbridge, R.; Rowlands, D.; Jager, J. Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and de-novo initiation. J. Mol. Biol. 2003, 326, 1025-1035.
92. Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. & Miyano, M. (1999) Structure (Cambridge, U.K.) 7, 1417–1426.
93. Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. & Weber, P. C. (1999) Nat. Struct. Biol. 6, 937–942.
94. Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R. L., Mathieu, M., De Francesco, R. & Rey, F. A. (1999) Proc. Natl. Acad. Sci. USA 96, 13034–13039.
95. Bressanelli, S., Tomei, L., Rey, F. A. & De Francesco, R. (2002) J. Virol. 76, 3482–3492.
96. O'Farrell, D., Trowbridge, R., Rowlands, D. & Jager, J. (2003) J. Mol. Biol. 326, 1025–1035.
97. Ng, K. K. S., Cherney, M. M., Vazquez, A. L., Machı A., Alonso, J. M. M., Parra, F. & James, M. N. G. (2002) J. Biol. Chem. 277, 1381–1387.
98. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. (2001) Nature 410, 235–240.
99. Koonin, E. V. (1991) J. Gen. Virol. 72, 2197–2206.
100. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989) EMBO J. 8, 3867–3874.
101. Lai, V. C. H., Kao, C. C., Ferrari, E., Park, J., Uss, A. S., Wright-Minogue, J., Hong, Z. & Lau, J. Y. N. (1999) J. Virol. 73, 10129–10136.
102. Hong, Z., Cameron, C. E., Walker, M. P., Castro, C., Yao, N., Lau, J. Y. N. & Zhong, W. (2001) Virology 285, 6–11.
103. Shirako, Y., Strauss, E. G. & Strauss, J. H. (2000) Virology 276, 148–160.

104. Carroll S.S., Tomassini J.E., Bosserman M., Getty K., Stahlhut M.W., Eldrup A.B., Bhat B., Hall D., Simcoe A.L, LaFemina R., Rutkowski C.A., Wolanski B., Yang Z., Migliaccio G., De Francesco R., Kuo L.C., MacCoss M., Olsen D.B. Inhibition of hepatitis C virus RNA replication by 2'-modified nucleoside analogs. J Biol Chem. Apr 4; 278 (14), 11979-84 (2003).

105. Das K, Sarafianos SG, Arnold E, Hughes SH, William JL, Lane MD. Encyclopedia of Biological Chemistry. Elsevier; New York: 2004. HIV-1 Reverse Transcriptase Structure; pp. 388–392.

106. Lawtrakul L, Beyer A, Hannongbua S, Wolschann P. Quantitative structural rearrangement of HIV-1 reverse transcriptase on binding to non-nucleoside inhibitors. Monatsh Chem. 2004;135:1033–1046.

107. Arnold E, Das K, Ding J, Yadav PN, Hsiou Y, Boyer PL, Hughes SH. Targeting HIV reverse transcriptase for anti-AIDS drug design: structural and biological considerations for chemotherapeutic strategies. Drug Des Discov. 1996;13:29–47.

108. Rodgers DW, Gamblin SJ, Harris BA, Ray S, Culp JS, Hellmig B, Woolf DJ, Debouck C, Harrison SC. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proc Natl Acad Sci USA. 1995;92:1222–6.

109. Ding J, Das K, Hsiou Y, Sarafianos SG, Clark AD, Jr, Jacobo-Molina A, Tantillo C, Hughes SH, 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 A resolution. J Mol Biol. 1998;284:1095–111.

110. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science. 1998;282:1669–75.

111. Hsiou Y, Das K, Ding J, Clark AD, Jr, Kleim JP, Rosner M, Winkler I, Riess G, Hughes SH, Arnold E. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. J Mol Biol. 1998;284:313–23.

112. Ren J, Nichols C, Bird LE, Fujiwara T, Sugimoto H, Stuart DI, Stammers DK. Binding of the second generation non-nucleoside inhibitor S-1153 to HIV-1 reverse transcriptase involves extensive main chain hydrogen bonding. J Biol Chem. 2000;275:14316–20.

113. Das K, Ding J, Hsiou Y, Clark AD, Jr, Moereels H, Koymans L, Andries K, Pauwels R, Janssen PA, Boyer PL, Clark P, Smith RH, Jr, Kroeger Smith MB, Michejda CJ, Hughes SH, 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;264:1085–100.

114. Carroll S.S., Tomassini J.E., Bosserman M., Getty K., Stahlhut M.W., Eldrup A.B., Bhat B., Hall D., Simcoe A.L, LaFemina R., Rutkowski C.A., Wolanski B., Yang Z., Migliaccio G., De Francesco R., Kuo L.C., MacCoss M., Olsen D.B. Inhibition of hepatitis C virus RNA replication by 2'-modified nucleoside analogs. J Biol Chem. Apr 4; 278 (14), 11979-84 (2003).

115. Witkowski J.T., Robins R.K., Sidwell R.W., Simon L.N. Design, synthesis, and broad spectrum antiviral activity of 1- -D-ribofuranosyl-1,2,4-triazole-3-carboxamide and related nucleosides. J Med Chem. Nov;15(11),1150-4 (1972).

116. Deval J.,. Powdrill M.H., D'Abramo C.M., Cellai L., Gottel M. Pyrophosphorolytic Excision of Nonobligate Chain Terminators by Hepatitis C Virus NS5B Polymerase ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Aug p. 2920–2928 (2007)

117. Benzaria S., Bardiot D., Bouisset T., Counor C., Rabeson C., Pierra C., Storer R., Loi A.G., Cadeddu A., Mura M., Musiu C., Liuzzi M., Loddo R., Bergelson S., Bichko V., Bridges E., Cretton-Scott E., Mao J., Sommadossi J.P., Seifer M., Standring D., Tausek M., Gosselin G., La Colla P. 2'-C-methyl branched pyrimidine ribonucleoside analogues: potent inhibitors of RNA virus replication. Antiviral Chem & Chemother., 18, 225-242 (2007).

118. Storer B.E. An evaluation of phase I clinical trial designs in the continuous dose-response setting. Stat Med. Aug 30; 20 (16), 2399-408 (2001).

119. A.B. Eldrup, C.R. Allerson, C.F. Bennett, S. Bera, B. Bhat, N. Bhat, M.R. Bosserman, J. Brooks, C. Burlein, S.S. Carroll, P.D. Cook, K.L. Getty, M. MacCoss, D.R. McMasters, D.B. Olsen, T.P. Prakash, M. Prhavc, Q.Song, J.E. Tomassini, J. Xia, Structure-activity relationship of purine ribonucleosides for inhibition of hepatitis C virus RNA-dependent RNA polymerase, J. Med. Chem. 47 (9) (2004) 2283–2295.

120. A.B. Eldrup, M. Prhavc, J. Brooks, B. Bhat, T.P. Prakash, Q. Song, S. Bera, N. Bhat, P. Dande, P.D. Cook, C.F. Bennett, S.S. Carroll, R.G. Ball, M. Bosserman, C. Burlein, L.F. Colwell, J.F. Fay, O.A. Flores, K. Getty, R.L. LaFemina, J. Leone, M. MacCoss, D.R. McMasters, J.E. Tomassini, D. Von Langen, B. Wolanski, D.B. Olsen, Structure-activity relationship of heterobase-modified 2′-C-methyl ribonucleosides as inhibitors of hepatitis C virus RNA replication, J. Med. Chem. 47 (21) (2004) 5284–5297.

121. S.S. Carroll, J.E. Tomassini, M. Bosserman, K. Getty, M.W. Stahlhut, A.B. Eldrup, B. Bhat, D. Hall, A.L. Simcoe, R. LaFemina, C.A. Rutkowski, B. Wolanski, Z. Yang, G. Migliaccio, R. De Francesco, L.C. Kuo, M. MacCoss, D.B. Olsen, Inhibition of hepatitis C virus RNA replication by 2¢-modified nucleoside analogs, J. Biol. Chem. 278 (14) (2003) 11979–11984.

122. D.B. Olsen, A.B. Eldrup, L. Bartholomew, B. Bhat, M.R. Bosserman, A. Ceccacci, L.F. Colwell, J.F. Fay, O.A. Flores, K.L. Getty, J.A. Grobler, R.L. LaFemina, E.J. Markel, G. Migliaccio, M. Prhavc, M.W. Stahlhut, J.E.Tomassini, M. MacCoss, D.J. Hazuda, S.S. Carroll, A 7-deaza-adenosine analog is a potent and selective inhibitor of hepatitis C virus replication with excellent pharmacokinetic properties, Antimicrob. Agents Chemother. 48 (10) (2004) 3944–3953.

123. S.S. Carroll, M.-E. Davies, L. Handt, K. Koeplinger, R. Zhang, S.W. Ludmerer, M. MacCoss, D.J. Hazuda, D.B. Olsen, Robust suppression of viral replication by a nucleoside polymerase inhibitor in chimpanzees infected with hepatitis C virus, Hepatology 44 (Suppl 1) (2006) 535A.

124. P.A. Furman, E. Murakami, H. Bao, J. Symons, M.J. Otto, Inhibition of HCV replication by PSI-6130:mechanism of biochemical activation and inhibition, J. Hepatol. 46 (Suppl 1) (2007) S224.

125. W.R. Jiang, S. Ali, S. LePogam, C. Daniel, S. Chiu, T. Kretz, I. Najera, P.A. Furman, N. Cammack, J. Symons, Inhibition of HCV replication by PSI-6130: characterization of activity in the HCV replicon system, J. Hepatol. 46 (Suppl 1) (2007) S228.

126. K. Klumpp, V. Leveque, S. Le Pogam, H. Ma, W.R. Jiang, H. Kang, C. Granycome, M. Singer, C. Laxton, J.Q. Hang, K. Sarma, D.B. Smith, D. Heindl, C.J. Hobbs, J.H. Merrett, J. Symons, N. Cammack, J.A. Martin, R. Devos, I. Najera, The novel nucleoside analog R1479 (4′-azidocytidine) is a potent inhibitor of NS5B-dependent RNA synthesis.

127. Dhanak D., Duffy K.J., Johnston V.K., Lin-Goerke J., Darcy M., Shaw A.N., Gu B., Silverman C., Gates A.T., Nonnemacher M.R., Earnshaw D.L., Casper D.J, Kaura A., Baker A., Greenwood C., Gutshall L.L., Maley D., Del Vecchio A., Macarron R., Hofmann G.A., Alnoah Z., Cheng H.Y., Chan G., Khandekar S., Keenan R.M., Sarisky R.T. Identification and biological characterization of heterocyclic inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. J Biol Chem. Oct 11; 277 (41), 38322-7 (2002).

128. Nguyen T.T., Gates A.T., Gutshall L.L., Johnston V.K., Gu B., Duffy K.J., Sarisky R.T. Resistance profile of a hepatitis C virus RNA-dependent RNA polymerase benzothiadiazine inhibitor. Antimicrob Agents Chemother. Nov; 47 (11), 3525-30 (2003).

129. Tomei L., Altamura S., Bartholomew L., Bisbocci M., Bailey C., Bosserman M., Cellucci A., Forte E., Incitti I., Orsetti L., Koch U., De Francesco R., Olsen D.B., Carroll S.S., Migliaccio G. Characterization of the inhibition of hepatitis C virus RNA replication by nonnucleosides. J Virol. Jan; 78 (2), 938-46 (2004).

130. Tomei L., Altamura S., Bartholomew L., Biroccio A., Ceccacci A., Pacini L., Narjes F., Gennari N., Bisbocci M., Incitti I., Orsetti L., Harper S., Stansfield I., Rowley M., De Francesco R., Migliaccio G. Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. J Virol. Dec; 77 (24),13225-31 (2003).

131. Mazzei M., Nieddu E., Miele M., Balbi A., Ferrone M., Fermeglia M., Mazzei M. T., Pricl S., La Colla P., Marongiu F., Ibba C. and Loddo R. Activity of Mannich bases of 7-hydroxycoumarin against Flaviviridae. Bioorg. & Med. Chem., 16, 2591-2605 (2008).

132. Mathers, C. D. & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3, e442.

133. Mathers, C. D. & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3, e442.

134. Baginski, S.G., Pevear, D.C., Seipel, M., Sun, S.C., Benetatos, C.A., Chunduru, S.K., Rice,C.M., Collett, M.S., 2000. Mechanism of action of a pestivirus antiviral compound. Proc. Natl. Acad. Sci. U.S.A. 97, 7981–7986.

135. Yanagida, K., Baba, C., Baba, M., 2004. Inhibition of bovine viral diarrhea virus (BVDV) by mizoribine: synergistic effect of combination with interferon- . Antiviral Res. 64, 195–201.

136. Paeshuyse, J., Leyssen, P., Mabery, E., Boddeker, N., Vrancken, R., Froeyen, M., Ansari,I.H., Dutartre, H., Rozenski, J., Gil, L.H., Letellier, C., Lanford, R., Canard, B.,Koenen, F., Kerkhofs, P., Donis, R.O., Herdewijn, P., Watson, J., De Clercq, E.,Puerstinger, G., Neyts, J., 2006. A novel, highly selective inhibitor of pestivirus replication that targets the viral RNA-dependent RNA polymerase. J. Virol. 80,149–160.

137. Tabarrini, O., Manfroni, G., Fravolini, A., Cecchetti, V., Sabatini, S., De Clercq, E., Rozenski, J., Canard, B., Dutartre, H., Paeshuyse, J., Neyts, J., 2006. Synthesis and anti-BVDV activity of acridones as new potential antiviral agents. J. Med. Chem. 49, 2621–2627.

138. Paeshuyse, J., Chezal, J.M., Froeyen, M., Leyssen, P., Dutartre, H., Vrancken, R., Canard, B., Letellier, C., Li, T., Mittendorfer, H., Koenen, F., Kerkhofs, P., De Clercq, E., Herdewijn, P., Puerstinger, G., Gueiffier, A., Chavignon, O., Teulade, J.C., Neyts, J., 2007. The imidazopyrrolopyridine analogue AG110 is a novel, highly selective inhibitor of pestiviruses that targets the viral RNA-dependent RNA polymerase at a hot spot for inhibition of viral replication. J. Virol. 81, 11046–11053.

139. Okamoto, M., Sakai, M., Goto, Y., Salim, M.T.A., Baba, C., Goto, K., Watashi, K., Shimotohno, K., Baba, M., 2009. Anti-bovine viral diarrhea virus and hepatitis C virus activity of the cyclooxygenase inhibitor SC-560. Antiviral Chem. Chemother. 20, 47–54.

140. Chang, J., Wang, L., Ma, D., Qu, X., Guo, H., Xu, X., Mason, P.M., Bourne, N., Moriarty, R., Gu, B., Guo, J.T., Block, T.M., 2009. Novel imino sugar derivatives demonstrate potent antiviral activity against flaviviruses. Antimicrob. Agents Chemother. 53,1501–1508.

141. Paeshuyse, J., Letellier, C., Froeyen, M., Dutartre, H., Vrancken, R., Canard, B., De Clercq, E., Gueiffier, A., Teulade, J.C., Herdewijn, P., Puerstinger, G., Koenen, F.,Kerkhofs, P., Baraldi, P.G., Neyts, J., 2009. A pyrazolotriazolopyrimidinamine inhibitor of bovine viral diarrhea virus replication that targets the viral RNA dependent RNA polymerase. Antiviral Res. 82, 141–147.

142. Sako, K., Aoyama, H., Sato, S., Hashimoto, Y., Baba, M., 2008. -Carboline derivatives with anti-bovine viral diarrhea virus (BVDV) activity. Bioorg. Med. Chem. 16, 3780–3790.

143. Aoyama, H., Sako, K., Sato, S., Nakamura, M., Miyachi, H., Goto, Y., Olamoto, M., Baba, M., Hashimoto, Y., 2009. Polymethylated -carbolines with potent anti-bovine viral diarrhea virus (BVDV) activity. Heterocycles 77, 779–785.

144. Luscombe, C.A., Huang, Z., Murray, M.G., Miller, M., Wilkinson, J., Ewart, G.D., 2010. A novel Hepatitis C virus p7 ion channel inhibitor BIT225, inhibits bovine viral diarrhea virus in vitro and shows synergism with recombinant interferon- -2b and nucleoside analogues. Antiviral Res. 86, 144–153.

145. Mohammed TA Salim, Mika Okamoto, Shinnosuke Hosoda, Hiroshi Aoyama, Yuichi Hashimoto, Masanori Baba; Anti-bovine viral diarrhoea virus activity of novel diphenylmethane derivatives;Antiviral Chemistry & Chemotherapy, 2010;20:193-200

146. Mohammed T.A. Salim, Yukinori Goto, Takayuki Hamasaki, Mika Okamoto, Hiroshi Aoyama, Yuichi Hashimoto, Simone Musiu, Jan Paeshuyse , Johan Neyts, Matheus Froeyen, Piet Herdewijn, Masanori Baba; Highly potent and selective inhibition of bovine viral diarrhea virus replication by -γ carboline derivatives; Antiviral Research 88 (2010) 263–268.

147. Sun, J.H., Lemm, J.A., O'Boyle, D.R.2., Racela, J., Colonno, R., Gao, M., 2003. Specific inhibition of bovine viral diarrhea virus replicase. J. Virol. 77, 6753–6760.

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

149. Mitsuya, H., Weinhold, K.J., Furman, P.A., St. Clair, M.H., Lehrman, S.N., Gallo, R.C., Bolognesi, D., Barry, D.W., Broder, S., 1985. 3 -Azido-3 -deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. Proc. Natl. Acad. Sci. U.S.A. 82, 7096–7100.

150. Samuel Broder,2010. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. Antiviral Research.85, 1-18.

151. Martin, J., Hitchcock, M.J., De Clercq, E., Prusoff, W., 2010. A brief history of the first generation nucleoside HIV reverse transcriptase inhibitors. Antiviral Res. 85, 34–38.

152. Hang, J. Q. et al. Substrate-dependent inhibition or stimulation of HIV RNase H activity by non-nucleoside reverse transcriptase inhibitors (NNRTIs). Biochem. Biophys. Res. Commun. 352, 341–350 (2007).

153. De Clercq, E. Emerging anti-HIV drugs. Expert Opin.Emerging Drugs 10, 241–274 (2005)
154. Hashimoto, H., Mizutani, K. & Yoshida, A. in WO 00147883 (Japan Tobacco Inc., Published International Patent Application, 2001).

II
Methods

1 METHODS

1.1 DOCKING

Molecular docking is a multidimensional optimization problem, requiring efficient samplaing across the entire space of positional, orientational, and confermational possibilities available to both a receptor and a ligand it recognizes. Recent flexible docking studies and the results from the docking section of the critical assessment of protein.Structure (CASP2) prediction contest [14] suggest that a suitable combination of homology modeling, docking strategies and experimental studies can be used to reliably predict protein-ligand structures.

Protein-ligand interactions, described by the induced-fit model of molecular recognition, involve structural alterations in the receptor binding site and the ligand. Thus, it is critical that the model allows for flexibility in both the ligand and the receptor. The compouter programs AUTODOCK, GOLD, GLIDE, FLEX etc, which we have now integrated into our studies of protein-ligand interactions, allow partial flexibility of the receptor and full flexibility of the ligand . flexibility of the ligand and receptor is computationally cumbersome to include in the docking simulations.If the ligand and the receptor are completly flexible during the simulation, even if we exclude everything else but the receptor binding site, the computational cost of the docking problem becomes so great. Proper ways to accomodate flexibility in the docking simulations without increasing the computational effort too much is one of the most important issues in the development of new metodologies at the moment.Computer-aided ligand-receptor docking has become an important research tool in structural biology.

1.1.1 *Autodock*

AUTODOCK (current version 3.0 and 4.0) docks flexible small moleculae ligand to rigid macromolecular recepor and with some residue flexibility of receptor. In the docking simulations the number of rorsional degrees of freedom in the ligand can be varied, but some angles and bond distances are maintained rigid. AUTODOCK uses Monte-Carlo simulated annealing and/or Lamarckin genetic algorithm minimization scheme combined with a rapid, atomic resolution, grid based method of energy evaluation utilizing the Amber force field. In order to find low

energy confermations of ligands in the receptor binding site.The overall interaction energy between chemical species (i.e between atoms of the ligand and the receptor) is estimated by considering both Lennard-Jones atom-atom potential and electrostatic effects, summed for the individual interactions between atoms. A distance-dependent dielectric constant or special solvation energy grid can be used to account for the solvent effects.

Stoddard and Koshland were pioneering users of AUTODOCK in a clever prediction of a structure of a peotein-protein complex by docking seprate domains. apartm from many uses of AUTODOCK in various ways of protein ligand interaction, it also helped in invention of a drug which is in clinical trials for HIV. McComman et al in 2007 used AUTODOCK in conjunction with the Relaxed Complex Method. Merck Pharmaceutical company used McCommon's et al group protocol of AUTODOCK to design new drugs that target integrase, which lead in October 2007 to the first clinically approved HIV Integrase inhibitor:Isentress.

The docking of the ligands was explored with AUTODOCK and the docking energy produced with the AUTODOCK force field was corrected with the energy needed to solvate the ligands

1.1.2 *Protocol*

Autodock 4 was used for all docking calculations(Morris GM et al journal of computational chemistry; 2009). The AutoDockTools(ADT) package was employed to generate the docking files and analyze the docking results. In the abscence of literature data for the binding of different class of compounds, we performed molecular docking taking carben alpha atom. of the respective resistant mutant of the corresponding site as in case of HCV and the resistant mutation isolated from our lab in case of BVDV, as a grid center, for each compounds. The interaction pattern and key residues of crystallized ligand (table1), which correspond to the binding site A, B, C, and D were used for the identification of proper orientation of our compounds into the different binding sites. Two different grid boxes, one for each binding site, were centered on the average mass center of the ligands. Thus, a grid box of 91 X 91 X 86 points and a grid spacing of 0.375Å was set in order to accommodate the NNIs that bind on to the different sites of NS5B polymerase.

The second grid box (spacing,0.375Å) of 62 X 75 X 75 points was implemented in such a way to accommodate the NNIs into the more buried area of NS5B. The grid maps were generated for each atom probe, describing its interactions with the compounds. Autogrid 4, as implemented in the Autodock software package, was used to generate grid maps. The Lamarckian genetic algorithm (LGA)62 was employed to generate orientations

or conformations of the ligands within the binding site. The global optimization started with a population of 150 randomly positioned individuals, a maximum of 2.5 X 108 energy evaluations, and a maximum of 27,000000 generations, as mentioned the protocol for blind docking. A total of 200 runs was performed, while all the remaining run parameters were maintained at their default settings. A cluster analysis was carried out using 2Å as the root-mean-square deviation tolerance. Docking experiments were also tried using a single grid box comprising the two binding sites. Although Autodock performed quite well in selectively positioning the ligands in the right pockets, the results in term of root-mean-square-deviation (rmsd) values were not fully satisfying. This is likely due to the fact that with only 100 runs such a large region is not sufficiently explored. Nevertheless, a higher number of conformation sampling, may not assured better results in the absence of required no. of runs. So we went for 200 runs although it was computationally demanding.

1.2 ASSESSMENT OF DOCKING

1.2.1 *Redocking*

. Further refinement of docking was performed, aiming to check out the more realistically binding mode conformations of compounds with no experimental data. In order to check for the reliability of the docking protocol, we performed docking studies on NS5B polymerase complexes. The docking results were evaluated through a comparison of the predicted docked positions of the ligand and the experimental ones. As a measure of docking reliability, the rmsd between the positions of heavy atoms of the ligand in the calculated and experimental structures was considered.

The choice of the best conformation was based on the assumption that, although for high-throughput screening protocols, only the first ranked conformations should be considered (that is, the conformer characterized by the lowest estimated free energy of binding)63. In other cases, the lowest energy conformer of the most populated cluster should also be taken into account64. We applied the second protocol for investigation of binding mode of newly identified potent lead of different class of compounds. While the best docked is the lowest energy docked conformations of the first autodock generated cluster.the best cluster is the lowest energy docked conformation of the most populated cluster and best fitted cluster is the lowest energy docked confermation of the cluster showing the lowest rmsd value.In the ideal case, i.e., the error free docking program, the three conformations would have coincided.

The fact that Autodock reproduces the correct conformations (best fitted cluster) by 100 percent reflects the intrinsic lack of accuracy (at least for NS5B) of the implemented scoring function in selecting the right pose, and this is at the basis for the further analysis using molecular dynamics.

Flexibility of the ligand and the receptor is

1.3 MOLECULAR DYNAMICS

Molecular Dynamics (MD) simulation is a technique founded upon the basic principles of classical mechanics that provide a dynamical picture of the individual particles of the system at a microscopic level. Using this technique successive configuration of the molecular system (in the phase space of coordinates and momenta) is generated by integrating Newton's law of motion. The result is a trajectory, which contains the microscopic time evolution of the system in the phase space [**? ?**]. From the trajectory generated, one can compute the dynamical properties such as absorption spectra, rate constants and transport properties. Further, on combining MD with statistical mechanics as a mean of sampling, one can compute equilibrium properties such as average thermodynamics quantities, structure, and free energies along the reaction path seen as a union of all possible states of the system [**? ?**]. For instance, the statistical ensemble average of an observable **A** can be obtained as:

$$\langle A \rangle = \sum_{t=1}^{\tau \to \infty} A(t) \tag{1.1}$$

The assumption made here is called the ergodic hypothesis, i.e given an infinite amount of time, ensemble average of observable **A**, is equivalent to its time average. The main aspect in atomistic MD simulations are:

- An algorithm that samples the phase space
- The choice of the interaction potential, $V(\mathbf{r})$, between the atoms of the system.

Several simulations approaches were developed in the last decades that differs in the method to sample the phase space. The most fundamental form used to describe equation of motion is the Lagrangian form:

$$\frac{\mathbf{d}}{\mathbf{dt}}\left(\frac{\partial \mathbf{L}}{\partial \dot{\mathbf{q}}_\mathbf{j}}\right) - \left(\frac{\partial L}{\partial q_j}\right) = 0 \tag{1.2}$$

where L(q,$\dot{q}$) is Lagrangian defined as the difference between the kinetic and potential energies L=K-V, where q_j are general-

ized coordinates and $\dot{q}_j$ are the associated time derivative. The momentum conjugate to coordinate q_j is given as:

$$\mathbf{p}_j = \frac{\partial L}{\partial \dot{q}_j} \tag{1.3}$$

On substitution with the usual definition of kinetic and potential terms with cartesian coordinates r_i, equation 1.2 becomes:

$$\mathbf{F}_i = m_i \ddot{\mathbf{r}}_i \text{ with } \mathbf{F}_i = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}_i} \tag{1.4}$$

where $V(\mathbf{r})$, the potential, is a function of the atoms positions and F_i represents the total force on atom i. In this equation one assumes that the nuclear motion of constituent particles obeys the laws of classical mechanics. This is an excellent approximation if the distance in the energetic (translational, rotational and vibrational) levels of the involved degrees of freedom is $<< kT$, where k is the Boltzmann constant and T the temperature.

In the Hamiltonian form the equation of motion for the cartesian coordinates is given by:

$$\dot{\mathbf{r}}_i = \frac{p_i}{m_i} \text{ and } \qquad \dot{\mathbf{p}}_i = \frac{\partial V(r)}{\partial r_i} \tag{1.5}$$

1.3.1 *Integration of Newton equations of Motion*

Under the influence of a potential, the motions of atoms are strongly coupled to each other giving rise to many-body problems that cannot be solved analytically [**?**]. Therefore, in MD calculation an iterative numerical procedure is employed to obtain an approximate solution for the equations of motion.

The two important properties of the equations of motion to be noted are:

- They must be time reversibe (t=-t).
- Conservation of total Energy (Hamiltonian) of the system.

For the first point, as the Newton equations are time reversible also the algorithm used is supposed to satisfy the same time reversal symmetry. The algorithms that are not time reversible do not normally preserve the phase space volume, i.e. they do not satisfy the Liouville theorem. For the second point, conservation of Hamiltonian is equivalent to conservation of total energy of the system and provides an important link between MD and statistical mechanics. The energy conservation condition H(p,r) = E, defines a hypersurface in the phase space called the constant energy, imposing a restriction on system to remain on this surface

[**?**]. A good way to check the accuracy of the algorithm is to follow the temporal evolution of an observable A that should be conserved (e.g. the total energy). In general a good algorithm must be such that:

$$\frac{|\, A(t_n) - A(t_0)\,|}{\langle A(t)\rangle} \ll 1, \qquad \text{for}\,(t_n - t_0) \gg \Delta t \tag{1.6}$$

there is no drift in the total energy.
The MD integration of the Newton's equation which have a continuous form, are based on assumption that position, velocities and other dynamical properties can be discretized using the Taylor series expansion:

$$r(t+\delta t) = r(t) + \Delta t v(t) + \frac{1}{2}\Delta t^2 a(t) + \frac{1}{6}\Delta t^3 b(t) + ... \tag{1.7}$$

$$v(t+\delta t) = v(t) + \Delta t a(t) + \frac{1}{2}\Delta t^2 b(t) + \frac{1}{6}\Delta t^3 c(t) + ... \tag{1.8}$$

The choice of the integration method depends on the degree of accuracy of problem at hand. One of the most useful form used is the velocity verlet algorithm [**?**], a variant of verlet algorithm [**?**]. The advantage is using velocity verlet method is that positions, velocities and acceleration are well synchronized that allow to calculate the kinetic energy contribution to the total energy at same time, from which potential energy is determined. The equations are:

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(\Delta t)\Delta t + \mathbf{a}_i\frac{1}{2}\Delta t^2 + \mathbf{O}(\Delta t^3) \tag{1.9}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + [a_i(t) + a_i(t+\Delta t)]\frac{1}{2}\Delta t + \mathbf{O}(\Delta t^3) \tag{1.10}$$

where $\mathbf{a}_i$,$\mathbf{r}_i$,$\mathbf{v}_i$ are respectively the acceleration on the atom i, the atom position and the atom velocity. The algorithm has an accuracy of $O\,(\Delta t^3)$ for the variables and it is reversible in time.

Together with conservation of energy and time-reversibility another important feature of an integrating algorithm is to permit long time steps Δt. It is expected that the numerical Newtonian trajectory will diverge from the "true" Newtonian trajectory. However, it is important that the integrating algorithm maintains a well defined energy tolerance ΔE throughout the simulation time. The error (ΔE) is known to decrease on decreasing the time step Δt. The aim here is to find a balance between using the largest possible time step and maintaining an acceptable ΔE all along the simulation. A large time step would lead to faster exploration, but energy would fluctuate widely with the possibility of the

simulation being catastrophically unstable, on other hand too short time step would lead to computation being needlessly slow. The choice of an integration step is determined by the nature of forces acting on the system. The golden rule is to choose time step ($\Delta t \sim 10^{-15}$s) such that the fastest motion of the system can be integrated accurately. This requirement is a severe restriction, particularly as high frequency motions are relatively of less interest and have minimal effect on the overall behavior of the system. One suggested approach is to freeze out such vibrations by constraining the appropriate bonds to their equilibrium values. Details of this approach is discussed in subsection 1.3.4, or to use multiple time step approach which is discussed in subsection 1.3.2.

1.3.2 *Multiple Time Step Integrator*

One of the approaches to accelerate the integration of equations of motion, is to use "multi-time" step algorithm such as reverse reference system propagation algorithm (r-RESPA)[**?**]. In the algorithm (r-RESPA), the molecular system is classified into number of groups according to how rapidly the forces varies over time. The starting point is the Liouville operator formulation, which can cast the equations for the Hamiltonian system (see equation1.5) in a general form:

$$\dot{\mathbf{x}} = iL\mathbf{x} \tag{1.11}$$

where $\mathbf{x}$ is the phase vector and iL is the Liouville operator. Consider a molecular system containing N atoms (or 3N degrees of freedom) with $\mathbf{x} = \{r_i, p_i\}$ representing a point in the phase space. The Liouville operator in cartesian coordinates is defined as:

$$i\mathbf{L} = \{..., H\} \equiv \sum_{i=1}^{3N} \left[\frac{\partial H}{\partial p_i} \cdot \frac{\partial}{\partial r_i} - \frac{\partial H}{\partial r_i} \cdot \frac{\partial}{\partial p_i} \right] \tag{1.12}$$

On subsituting equation 1.5 into equation 1.12, we get:

$$i\mathbf{L} = \{..., H\} \equiv \sum_{i=1}^{3N} \left[\frac{p_i}{m_i} \cdot \frac{\partial}{\partial r_i} + \mathbf{F}_i \cdot \frac{\partial}{\partial p_i} \right] \tag{1.13}$$

where F_i is the force on i^{th} degree of freedom, and {...,...} is the poisson bracket. The classical time propagator U(t) is unitary and defined as e^{iLT}, and the evolution of system Eq. 1.11 is expressed as:

$$\mathbf{x}(t) = e^{iLt}\mathbf{x}(0) \tag{1.14}$$

The action of operator U(t) on x(o) cannot be determined analytically, however the operator can be decomposed using Trotter theorem, such that the action of U(t) on x(o)for each part can be evaluated analytically. Applying the Trotter theorem we get:

$$\begin{aligned} \mathbf{e}^{i(L_1+L_2)t} &= \left[e^{i(L_1+L_2)t/P}\right]^P = \left[e^{i(L_1+L_2)\Delta t}\right]^P \\ &= \left[e^{iL_1(\frac{\Delta t}{2})}e^{iL_2\Delta t}e^{iL_1(\frac{\Delta t}{2})}\right]^P + O(t^3/P^2) \end{aligned} \quad (1.15)$$

where $\Delta t = t/P$. For finite P, the numerical iteration procedure is accurate to the second order in the time step at long times. From equation 1.15, for the three exponential terms, we define the discrete time propagator (U_1, U_2) as:

$$\begin{aligned} \mathbf{G}(\Delta t) &= U_1(\frac{\Delta t}{2}) + U_2(\Delta t) + U_1(\frac{\Delta t}{2}) + O(t^3/P^2) \\ &= e^{iL_1(\frac{\Delta t}{2})}e^{iL_2\Delta t}e^{iL_1(\frac{\Delta t}{2})} + O(t\Delta t^2) \end{aligned} \quad (1.16)$$

Since the three exponential terms in G Δt are separately unitary, G (Δt) is also unitary i.e $G^{-1}(t) = G^{\dagger}(t) = G(-t)$. Lets us now consider the propagator generated by subdivison as:

$$\begin{aligned} iL_1 &= \sum_{i=1}^{N} \frac{p_i}{m_i} \cdot \frac{\partial}{\partial r_i} \\ iL_2 &= \sum_{i=1}^{N} F_i \cdot \frac{\partial}{\partial p_i} \end{aligned} \quad (1.17)$$

The operator $U_1(\frac{\Delta t}{2})$ becomes a translation operator on the positions: $r_i \rightarrow r_i + \Delta t(\frac{p_i}{m_i})$, and operator $U_2(\Delta t)$ becomes a translational operator of momenta: $p_i \rightarrow p_i + (\frac{\Delta t}{2})F_i(r)$. On combining these two facts to action of operators in equation 1.16 on complete set of positions and momenta, yields the approximate evolution:

$$\begin{aligned} \mathbf{r}_i(\Delta t) &= r_i(0) + \Delta t v_i(0) + \frac{\Delta t^2}{2m_i}F_i(0) \\ \mathbf{v}_i(\Delta t) &= v_i(0) + \frac{\Delta t}{2m_i}[F_i(0) + F_i(\Delta t)] \end{aligned} \quad (1.18)$$

which is the famous velocity verlet **[?]** integrator derived using the operator formulation. The power of the operator based approach is its symplectic property which ensures no drift in the total energy, resistance to increase in time steps and allows generating stable long trajectories.

r-RESPA algorithms have been successfully employed to incorporate motions on more than two time scales. Let us consider a system with three characteristics time scales, a reference force F_i^{ref}, and two corrections F_i^{del} and F_i^{Del}, such that F_i=F_i^{ref}+F_i^{del}+F_i^{Del}. We define their Liouville operators as iL^{ref}, $iL^{(del)}$ and $iL^{(Del)}$ and the corresponding timescales δt, Δt and $\Delta\mathcal{T}$ respectively. The three time step propogator can then be written as:

$$\exp(iL\Delta\mathcal{T}) = \exp\left(iL^{(Del)}\frac{\Delta\mathcal{T}}{2}\right)\left\{\exp\left(iL^{(del)}\frac{\Delta t}{2}\right)\left[\exp(iL_2^{(ref)}) \times \exp(iL_1^{(ref)}\delta t)\exp\left(iL_2^{(ref)}\frac{\delta t}{2}\right)\right]^n \exp\left(iL^{(del)}\frac{\Delta t}{2}\right)\right\}^m \times \exp\left(iL^{Del}\frac{\Delta\mathcal{T}}{2}\right) \quad (1.19)$$

Thus, the correction due to slowest time scale is applied every m×n timesteps, and the intermediate time scale correction is applied every n steps. Such numerical procedure lead to considerable saving in the CPU time to perform a MD simulation.

1.3.3 *The interaction potential*

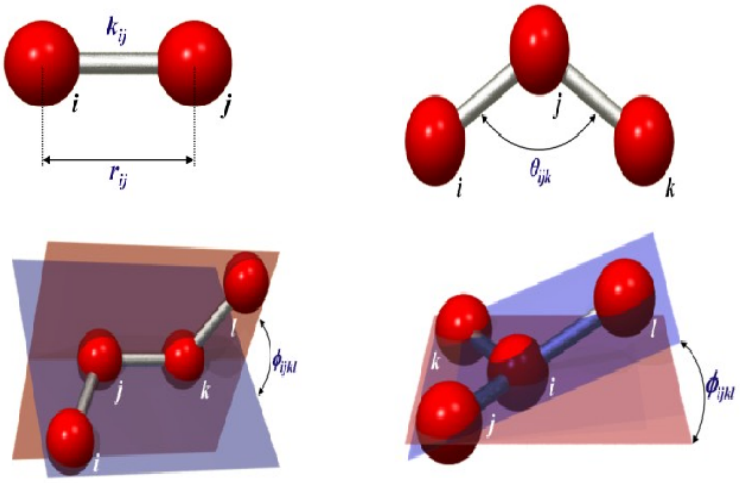


Figure 1: force field

The potential function $V(\mathbf{r})$ from which the forces used in MD are derived depends on the atomic coordinates $\mathbf{r}_i$.

$V(\mathbf{r})$ used in this thesis has the following expression:

$$
\begin{aligned}
V(r_1, r_2, \ldots, r_N) &= \sum_{bonds} \frac{1}{2} K_d (d - d_0)^2 \\
&+ \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \\
&+ \sum_{improper\,dihedrals} \frac{1}{2} K_\xi (\xi - \xi_0)^2 \\
&+ \sum_{dihedrals} K_\phi \left[1 + \cos(n\phi - \delta)\right] \\
&+ \sum_{ij\,LJ} \left[\left(\frac{C_{ij}^{12}}{r_{ij}^{12}} - \frac{C_{ij}^{6}}{r_{ij}^{6}} \right) \right] \\
&+ \sum_{ij\,coulomb} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}
\end{aligned}
\tag{1.20}
$$

The first two terms (two and three body interactions respectively) represent the bonds and angles potentials, that are approximated by harmonic functions. The third and fourth term describe four body interactions. Improper dihedral terms are typically described by an harmonic function. Instead proper dihedrals are described by periodic functions (i.e. cosine functions) of a given periodicity n. The last two terms are a Lennard-Jones (LJ) potential and a coulomb potential between pair (ij) of atoms. The LJ terms reproduce the Van der Walls interactions, while the coulomb potential terms reproduce the electrostatic properties of a protein. These interaction are modelled using the two-body approximation which doesnot explicitly account for the polarization effects, but on a average. The parameters used in this kind of potentials are typically obtained from quantum chemical calculations and experimental data (e.g. crystallographic data, spectroscopic data, etc). Among the popular sets of parameters (force fields) for MD simulations of proteins we can cite for example AMBER, GROMOS, CHARMM and OPLS. They all use the potential function expression given above for all the atoms of the simulated system except for the GROMOS(and CHARMM19 force field) force field in which a united atom description is used for non-polar hydrogens.

In MD simulations the description of the solvent (water for most of the biologically interesting systems) can be explicit or implicit. In the first case solvent molecules with a full atomistic force field description are added in the simulation box at the experimental density. In the implicit solvent description the solvent is treated as a dielectric medium in which the system is embedded. This is clearly a more approximated description but it is also computationally much more efficient since in many practical cases the solvent constitutes the majority of the atoms. In this

thesis we used for all the MD simulations the AMBER-95 and AMBER03 force field with an explicit description of the solvent given by the TIP3P[**?**] water model.

1.3.4 *Constraints for Hydrogen*

Constraints are used in MD to fix bonds to their equilibrium value. This allows increasing the simulation time step Δt. Constraining the bond lenght does not alter significantly the statistics as these are quantum degrees of freedom being mostly in their ground state at the normal simulation temperature. Using the bonds constraints it is possible to use $\Delta t \sim 2fs$ [**?**] (2-4 times larger than the one that can be used without constraints). A common method to introduce constraints is the algorithm SHAKE [**?**], in which after each time step the atoms positions iteratively are modified in order to satisfy the constraint.

SHAKE may have convergence problems when applied to large planar groups and its implementation could hinder the efficiency of computing. To improve these aspects the LINCS algorithm was recently introduced. For water molecules it is also possible to use an analytic solution of SHAKE called SETTLE.

1.3.5 *Boundary conditions*

To simulate a finite size system, boundary conditions are needed to avoid artifacts near the border of the simulation box. Typically periodic boundary conditions (PBC) are used. In this scheme short range non bonded interactions are calculated using the minimal image convention (only the nearest replica is considered). Typically a cut-off radius (R_c) is used for LJ interactions of the order of 10 Å. To avoid interactions between a particle and its periodic image each box side must be larger than $2R_c$.

The coulomb energy is instead treated considering the full periodicity of the system. For a periodic lattice made by N particles it is given by:

$$E = \frac{1}{8\pi\epsilon_0} \sum_{|n|=0}^{\infty}{}^{\star} \left[\sum_{i=0}^{N} \sum_{j=0}^{N} \frac{q_i q_j}{|r_{ij} + n|} \right] \tag{1.21}$$

where n indicates the periodic images, i,j the particles and the symbol $\star$ indicates that the summation does not contain the term with $i = j$ if $n = 0$.

The periodicity of the system speeds up the evaluation of the coulombic term. Although convenient, effective, and simple to apply, certain subtle problems arise for long range forces (electrostatics), whose spatial range may extend beyond the boundaries

of the container into surrounding images, present a challenge. Long range forces can only be correctly calculated by summing over all the periodic replicas of the original system. However, the associated computational effort is considerable. Fortunately, methods have been developed to treat this problem. Specifically, the Ewald summation technique, developed originally to treat Coulomb interactions and later extended to treat general interactions of the form $\frac{1}{r_n}$ for $n \leqslant 3$ has proved enormously successful. The basic idea behind the technique is to divide the relevant part of the potential into a short range and a long range contribution. For the Coulomb potential, $1/r$, for example, this can be achieved via the identity

$$\frac{1}{r} = \frac{erf(\alpha r)}{r} + \frac{erfc(\alpha r)}{r} \tag{1.22}$$

where erf(x) and erfc(x) are the error function and complementary error function, respectively (erf(x) + erfc(x)= 1). The variable, R, is a convergence parameter, which can be optimized for each system studied. The short range term, erfc(αr)/r, is treated as an ordinary short range interaction, i.e., using a spherical cutoff to truncate the interaction at large spatial distances where the potential is small. The long range term, erf(αr)/r, is Fourier transformed into reciprocal space, where it takes the short-ranged form, exp(-$g^2/4\alpha^2$), and can be evaluated accurately by summing over only a small number of reciprocal space vectors of the simulation cell. Such reciprocal space sums can be evaluated with high a degree of efficiency (N log N) using particle-mesh methods(PME)[**?**]. An extension of PME is the smooth PME. With respect to PME, this method uses a fixed cuttoff in the direct sum and uses the B-spline interpolation of the reciprocal space structures onto a rectangular grid, permitting the use of fast Fourier transforms to efficiently calculate the reciprocal sum. In this thesis we use SPME method to evaluate the electrostatic energies[**?**].

1.3.6 *Statistical Ensembles*

Molecular dynamics can be performed in different statistical ensembles. The traditionally used ensemble to perform MD is the micro-canonical ensemble (NVE), where the number of particles (N), the volume (V), and the total-energy (E) of the system are fixed to a constant value.
The simple extension of NVE ensemble is the canonical one (NVT), where the number of particles, the volume and the temperature are fixed to a constant value. The temperature T, in contrast to the number of particles N and volume V, is an inten-

sive parameter. The temperature T is related to the time average of the kinetic energy given as:

$$\mathbf{T} = \frac{2}{3}\frac{E_{kin}}{Nk_B} = \frac{1}{3Nk_B}\sum_{i=1}^{N}\frac{{p_i}^2}{m} \tag{1.23}$$

where, E_{kin} is the kinetic energy, k_B is the Boltzmann constant. The simplest way to control the temperature, is to rescale the velocities at each step by the factor $\lambda = \sqrt{\frac{T_{req}}{T_{curr}}}$, where T_{curr} is the current temperature calculated from the kinetic energy and T_{req} is the desired temperature (for instance 300 K). However, an alternative way to maintain is to couple the system to an external heat bath that is fixed at the desired temperature. The bath acts as a source of thermal energy, supplying or removing heat from the system as appropriate. This thermostat is named as the "Berendsen" thermostat. It is extremely efficient for relaxing a system to the target temperature, but once the system has reached equilibrium, it might be more important to probe a correct canonical ensemble.

Extended system methods, was originally introduced for performing constant MD simulation by Nosè in 1984, and subsequently developed by Hoover in 1985. The idea of the method was to reduce the effect of an external system, acting as a heat reservoir, to an additional degree of freedom s. This reservoir has a potential energy $(f+1)k_B T \ln s$, where f is the number of degrees of freedom in the physical system and T is the desired temperature. The kinetic energy of the reservoir is given as $(\frac{Q}{T})(\frac{ds}{dt}^2)$, where Q is considered as the fictitious mass of the extra degree of freedom. The magnitude of Q determines the coupling between the reservoir and the real system and so influences the temperature fluctuation. If Q is large then the energy flow is slow; in the limit of infinite Q, conventional molecular dynamics is regained. However, if Q is small then the energy oscillates, resulting in equilibrium problems. It has been suggested that Q should be proportional to $fk_B T$.

Another ensemble we discuss here it the NPT ensemble, an extension of NVT ensemble, where together with temperature the pressure of the system is maintained to a constant vaue. As most experimental measurements are usually made under conditions, which include a fixed pressure P, temperature T, and number of atoms N (constant-NPT ensemble), and so simulations in the isothermal-isobaric ensemble are the most directly relevant to experimental data. A simulation in NPT ensemble maintains the constant pressure by changing the volume of the simulation cell.

The amount of volume fluctuation is related to the isothermal compressibility, κ

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{1.24}$$

An alternative to maintain constant pressure is to couple the system to a "pressure" bath, analogous to the temperature bath. The rate of change of pressure is given by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p}(P_{bath} - P(t)) \tag{1.25}$$

where τ_p is the coupling constant, P_{bath} is the pressure of the 'bath', and P(t) is the actual pressure at time t. The volume of the simulation box is scaled by a factor λ, which is equivalent to scaling the atomic coordinates by a factor $\lambda^{1/3}$. Thus:

$$\lambda = 1 - \kappa\frac{\delta t}{\tau_p}(P - P_{bath}) \tag{1.26}$$

and the new position are given by:

$$r_i{}^{new} = \lambda^{1/3} r_i \tag{1.27}$$

In the extended pressure-coupling systems, an extra degree of freedom, corresponding to the volume of the box, is added to the system. The kinetic energy associated with this degree of freedom (which can be considered to be equivalent to piston acting on the system), is $(\frac{1}{2}Q)(\frac{dV}{dt}^2)$, where Q is the 'mass' of the piston. The piston also has a potential energy PV, where P is the desired pressure and V is the volume of the system. The volume varies in the simulation with the average volume being determined by the balance between the internal pressure of the system and the desired external pressure. In this thesis, we have performed MD simulation in both NVT and NPT ensembles.

1.3.7 *Solvation effect*

In most cases, we are interested in the properties of molecules in solution, usually in aqueous solution. therefore, it is essential to calculate not only the interaction between the atoms of biomolecules, but also the solvent effect on the interaction. there are two different approches to include the solvent effects in molecular dynamics. One is explicit solvent model, the other one is the continuum solvent model, also known as the implicit solvent model.

In explicit solvent models we need to calculate the interaction between the particles and every solvent molecule individully, therefore it is a very "computationally expensive " job, and the

cost increase with the number of interaction sites in the water model. Algorithms such as Particle Mesh Ewald (PME) method, particle-particle/particle-mesh(P3M) method are developed to accelrate the simulations.

Even with advanced algorithms, one obvious drawback of the explicit solvent model is the large system size due to numerous solvent molecules. An alternative approch is using the continuum solvent models, such as the Generalized Born(GB) model. This model represents the solvent implicity as a continuum with dielectric properties of water, and also includes the charge screening effects of salt. this lowers the calculation expense in two aspects: first, it removes the calculation of the interactions and motions involving solvent molecules;second, the absence of solvent friction accelerates the dynamics of the solute. However, the computing cost of GB model scales up with the size of the system. for large systems, the computing cost may be greater than using explicit solvent model.

Both explicit solvent and implicit solvent models have certain strengths and weakness. Calculations using explicit solvent generally yield more accurate results. In some cases, such as simulations involving water bridges, explicit water molecules are essential for the calculation. However, systems using explicit solvent have many more atoms (in most case mostly water molecules), demanding additional computing resources. Due to the fraction force from water molecules, the dynamics of the solute is also slower in the explicit solvent. this may be useful for understanding the real time scale of a biological event. However, for some studies where rate is not a factor, but the results of the motion are important, the implicit solvent model will be more efficient. Another important advantage of implict solvent model is that it calculates solvent free energy, instead of energy. In the explicit solvent model, water atoms are considered to be part of the system, and solvent-solute interactions are included in the non-bonded calculations. There is no independent solvation term. In implicit solvent the simulated system only includes solute. the molecular mechanic interactions are calculated first in vacuum. then an additional term, the solvation free energy is calculated and included in the energy of the system. the implicit solvent mimics the average effects of the water molecules, therefore it genrates solvation free energies. it is essentail for methods such as Molecular Mechanics-Poisson Boltzmann (Generalized Born) Surface Aea (MM-PB(GB)SA) approch, which needs an explicit solvation free energy term in calculation. just a note, MM-PB(GB)SA is a post-processing analysis, which means it is still recommended to genrate the structures using explicit solvent first, then using implicit solvent model to analyze the solvation free energy on existing structures.

Molecular Dynamics (MD) simulations allow investigating processes occurring on timescales of ~100ns. However, most interesting and relevant biological process happen on time scales that are orders of magnitude larger, and are therefore termed as rare events. For example, protein folding (μs-few seconds), protein-protein interactions, transport of molecules across membrane channels (order ~ μs) and many others. Over the years, we have observed an astounding increase in computer power (Blue gene, DESRES), which promise to increase utility of MD simulations to investigate more and more complex systems on μs timescale. However, these supercomputing machines are not available to all the research groups. Therefore another approach to overcome the timescale problem is to renounce the all atom approach and to use coarse grained models. This would retain the essential characteristics, however you require a detailed knowledge of system, that is often not available.
For systems, where its important to maintain the atomistic description, one can exploit methodology aimed at accelerating rare events to timescales reachable in MD simulations. Notable success has been achieved is using the accelerating methodology in diverse fields of interest. From their scope and range of applicability, they are classified in four categories [**?**]:

1. Methods aimed at improving sampling, in a subspace of few predefined collective variables (CVs), that allow reconstructing the probability distributions as a function of chosen CVs. Examples of these methods include thermodynamic integration [**? ?**], free energy perturbation [**?**], umbrella sampling [**?**], conformational flooding [**?**], weighted histogram [**? ? ?**], steered MD [**?**], Jarzynski's identity based methods [**?**] and adaptive force bias [**?**]. The power of these methods in highly dependent on judicious choice of CVs, and computational performance degrades as a function of the number of variables.

2. Methods aimed at exploring the transition mechanism. Examples in these catogeries are transition path sampling[**? ?**], finite temperature string method [**? ?**], transition interface sampling [**?**] and forward flux methods [**?**]. These methods do not require in most cases, an explicit definition of a reaction co-ordinate, but require a priori knowledge of initial and final states of process under investigation.

3. Methods for exploring the potential energy surfaces and localizing the saddle points that corresponds to a transition state. Examples in these catogeries are dimer method [**?**], hyperdynamics [**?**], multiple time scale accelerated MD [**?**]

and event based relaxation [**?**]. The power of these methods is limited to low dimensionality, and reliability degrades with the complexity of system.

4. Methods in which the phase space is explored simultaneously at different values of temperatures, are parallel tempering [**?**] and replica exchange [**?**], or as a function of the potential energy, such as multicanonical MD [**?**] and Wang-Landau [**?**].

1.4.1 *Metadynamics*

Metadynamics method encompasses several features of techniques mentioned earlier and provides a unified frame work for computing free energies and accelerating rare events. It is a powerful algorithm, based on dimensional reduction that is used for accelerating rare events in system described by complex Hamiltonians, at a classical or quantum level [**?**]. Before to use the algorithm, the requirements are to identify a set of CVs, which are assumed to describe well the process of our interest. The power of the algothim lies in treating the CVs simultaneously and in its flexibility: the method can be proficientely used both for reconstructing free energy and for accelerating rare events.

1.4.2 *The algorithm*

Consider a system described by a set of co-ordinates x and a potential V(x) evolving under the action of molecular dynamics, whose equilibrium distribution is canonical at a temperature $T(\frac{1}{\beta})$. Since our interest is to describe the system using a set of CVs, $S_\alpha(x)$, α=1,d where d is a small number assuming they provide a good coarse-grained description [**?**]. The equilibrium behavior of these CVs in defined by probability distribution:

$$P(s) = \frac{\exp\left(-(\frac{1}{T})F(s)\right)}{\int ds \exp\left(-(\frac{1}{T})F(s)\right)} \tag{1.28}$$

where s denotes the d dimensional vector (s_1,....,s_d), with the free energy given by:

$$F(s) = -T \ln\left(\int dx \exp\left(-\frac{1}{T}V(x)\right)\delta(s - S(x))\right) \tag{1.29}$$

in equation 1.29, capital S is used for representing the function of the coordinates S(x), while lower case s is used for denoting the value of the CVs.
Consider now a trajectory x(t) of the system at temperature T.

On computing the trajectory for a very long time, probability P(s) can be obtained by taking the histogram of CV s along this trajectory. At time t, P(s) ~ $\frac{1}{t}\int_0^t dt'\delta$ (S(x(t') - s). If the system displays metastability, the motion of s will be bound in some local minimum of F(s) (i.e., in local maximum of P(s)), and will escape from this minimum with a very low probability on the timescale determined by the potential V(x) alone. In Metadynamics, the metastability is eliminated by modifying the underlying potential V(x), by adding a history dependent term consisting of Gaussians centered along the trajectory in s space during the evolution of the system [**? ?**]. Each time a new gaussian is added at time τ_G, the biasing potential at time t is given as:

$$V_G(S(x),t) = \omega \sum_{\substack{t'=\tau_G,\tau_{2G}... \\ t'<t}} \exp(-\frac{(S(x)-s(t'))^2}{2\delta s^2}) \qquad (1.30)$$

In equation 1.30 capital S is used for denoting the function of the co-ordinates S(x), while lower case s is used for denoting the value of the CVs, s(t)=S(x(t)) is the value taken by the CV at time t, ω is gaussian width, δs is the gaussian width and frequency τ_G at which gaussians are added. The basic assumption of metadynamics is that V_G(s,t) defined in equation 1.30 after a sufficiently long time provides an estimate of the underlying free energy:

$$\lim_{t\to\infty} V_G(s,t) \sim -F(s). \qquad (1.31)$$

Here, in equation 1.31, a equilibrium quantity free energy (F(s)) is estimated by a non-equilibrium dynamics. Equation(1.31) can be qualitatively understood as slow 'deposition' (i.e. $\omega \to 0$), and in this limit V_G(s,t) varies very slowly, with the probability to observe s is the approximately proportional to exp[$-\frac{1}{T}F(s) + V_G$ (s,t))]. If the function F(s)+V_G(s,t) has some local minimum, s will be preferentially localized in the of this minimum and Gaussians will be deposited, till the minimum is filled. If we consider the situation where F(s) ~ $-V_G$(s,t) in a region Ω(s). In this case the probability distribution would be flat and the locations of the gaussians will not be affected by the bias. Hence, if $\omega \to 0$, only corrugations in the free energy that are not flattened by the dynamics will be the order of the size of the newly added gaussians.

In most cases, it is not sufficient to describe the process of interest using a single CV, therefore in cases where we use more than one CV at the same time, the metadynamics potential is given by

$$V_G(S(x),t) = \omega \sum_{\substack{t'=\tau_G,\tau_{2G},... \\ t'<t}} \exp\left(-\sum_{\alpha=1}^{d} \frac{(S_\alpha(x)-s_\alpha(t'))^2}{2\delta s_\alpha^2}\right)$$

(1.32)

and it is necessary to choose metavariable δs_α for each CV. The time required to escape from a local minimum is determined by number of Gaussians needed to fill the basin. This number is proportional to $(\delta s_\alpha)^d$ where d is the number of CV chosen to describe the system. The efficiency of the method scales exponentially with number of CV. The added gaussians can reproduce features of the FES on a scale larger than δs. The quality and accuracy of the free energy construction is highly influenced by the parameters of metadynamics. The parameters need to be chosen wisely, striking a balance between accuracy and efficiency.

1.4.3 *How to Choose CVs*

Reliability of Metadynamics is highly dependent on the choice of CVs, and this in turn depends on the process of investigation. Therefore if not essential it is important to have good knowledge of the process under investigation, which aids in chosing the right CVs. Important features that CVs must satisfy are:

1. They should be slow variables able to describe well the process of interest.
2. They should be able to distinguish between the initial, intermediate and final states.

Practical Example of choosing CVs

For example, lets consider the diffusion of antibiotics through OmpF channel. The OmpF channel has an axis of symmetry along Z-axis, with a hourglass shape. Here we want to follow the exit of antibiotic from the channel. In this case, one appropriate CV would be distance: defined as difference between the center of mass (com) of the antibiotic (a_1) and the center of mass of OmpF (p_1) along z-axis. The CV distance, s_{dist} is then defined as:

$$s_{dist} = \frac{\sum_{i \in a_1} m_i z_i}{\sum_{i \in a_1} m_i} - \frac{\sum_{i \in p_1} m_i z_i}{\sum_{i \in p_1} m_i} \tag{1.33}$$

We make a assumption in equation 1.33, com of OmpF (p_1) is constant respect to com of the antibiotic(a_1). Therefore using this CV s_{dist} we caputure the location of antibiotic in the OmpF channel.

1.4.4 *Estimation of Error*

Error estimation in Free energy calculations is very much dependent on the meta variables ($\omega, \delta s, \tau_G$). In [**?**] it has been shown

that the error on reconstruction profile is determined by the ratio of (ω/τ_G) and not by ω and τ_G alone. For instance, adding Gaussian of height ω=1.0 Kcal mol^{-1} every τ_G = 4ps, is equivalent to adding Gaussian of height ω=0.5 Kcal mol^{-1} every τ_G = 2 ps, with a requirement that τ_G remains much shorter than the time required to fill the free energy basin. The error ϵ in metadynamics is a measure of expected deviation of V_G(s,t) from -F(s):

$$\varepsilon^2(s) = \left\langle (V_G(s) + F(s))^2 \right\rangle_M = \tag{1.34}$$

$$= \left\langle (V_G(s) - \langle V_G(s) \rangle_M)^2 \right\rangle_M \tag{1.35}$$

Remarkably, the error does not depend on $F(s)$. An alternative expression for for the error, dependent on parameters: w/τ_G, T, D and S, is given as:

$$\bar{\varepsilon}^2_{\text{approx}} = C_d \frac{S^2 w T}{D\tau_G} \frac{\delta s}{S} \tag{1.36}$$

where C_d is a constant that depends only on the dimensionality. The two expressions for the error share the same functional dependence on w/τ_G, T, D and S. The ratio between the two expressions is approximately a constant as a function of δs only for $d = 1$ and $d = 2$, while significant deviations are observed in higher dimensions.

The dependence of the error on the simulation parameters becomes more transparent if $\bar{\varepsilon}$ is expressed as an explicit function of the total simulation time. Consider in fact a free energy profile $F(s)$ that has to be filled with Gaussians up to a given level F_{max} , for example the free energy of the lowest saddle point in $F(s)$. The total computational time needed to fill this profile can be estimated as the ratio between the volume that has to be filled and the volume of one Gaussian times τ_G :

$$t_{sim} \approx \tau_G \frac{F_{max}}{w} \left(\frac{S}{\delta s}\right)^d \tag{1.37}$$

Substituting in Eq. 3.36 yields

$$\bar{\varepsilon}^2 \approx \frac{\tau_S}{t_{sim}} F_{max} T f_d(\frac{\delta s}{S}) \tag{1.38}$$

where $\tau_S \doteq \frac{S^2}{D}$ is the average time required for the CVs to diffuse on a distance S and

$$f_d(\frac{\delta s}{S}) = (2\pi)^{d/2} \sum_k \frac{1}{k^2\pi^2} \exp\left(-\frac{k^2\pi^2}{2}\left(\frac{\delta s}{S}\right)^2\right)$$

is a function of $\frac{\delta s}{S}$ and of the dimensionality alone. Eq. 1.38 states that the error of a metadynamics reconstruction is inversely proportional to the square root of the total simulation time, measured in units of the diffusion time. The error will be large for slowly diffusing systems, in which the walker takes a long time to explore the CVs space.

1.5.1 *MM-PBSA*

one of the major goals of the computational chemistry is to develop methods to accurately predict the binding eergy of a ligand to a protein. This is of central interest in medicinal chemistry, because the action of most drugs (inhibition, activation etc.) is caused by the binding of the drug to its target receptor. However, many biochemical problems can be treated in a similar way For example, the reactivity of an enzyme can be estimated by comparing the free energy of the reactant and tranisition states of the active site in protein. Therefore, many methods have been developed with this aim [**?** **?**]. MM-PB(GB)SA method is another free energy calculation method(107). Unlike other free energy calculation methods like free energy perturbation (FEP) or Thermodynamic Integration(TI) method, this is the post-processing method, which means it calculates the free energy based on existing structure snapshots. MM-PBSA restrict the molecular simulations to the states before and after the binding process.It is an attractive approch because it does not contain any parameters that var for different ligands-receptor systems and it involves a set of physically well defined term: The binding affinity is estimated from the free energies of the receptor, the ligand and the complex. In MM-PB(GB)SA calculation, the free energy of the system is divided into three parts: molecular mechanics energy(MM), and non polar solvation free energy (SA). The MM energy usually consists of electrostatic energy and vander walls energy of the system in vacuum. In some applications the bond energies, angle energies, and dihedral angle energies are also included in MM energy. the entropy of the system can be included by using normaol mode analysis or other methods.

In this method to estimate the free eergy of the complex system, one carries out a molecular dynamics simulation, typically in a periodic box with water and counter ions, and correct representation of long-range electrostatic effects, saving a set of representative structures. then one post process these structures, remove any solvent and counter ions molecules, and calculates the free energy, G, accordign to the following equation:

$$\triangle G_{bind} = \triangle E_{MM} + \triangle G_{PBSA} - TS_{MM} \tag{1.39}$$

or

$$\triangle G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \tag{1.40}$$

where $\triangle G_{bind}$ is the calculated average free energy, and $\triangle E_{MM}$ is the average molecular mechanical energy, $G_{complex}$, $G_{receptor}$,

and G_{ligand} are the free energies of the complex, the protein, and the ligand, respectively.

For each system, the free energy can be estimated in terms of molecular mechanic potential energy E_{MM}, the solvation free energy G_{solv}, and the entropic contribution (TS) as fallows:

$$G_{bind} = E_{MM} - TS + G_{solv} \tag{1.41}$$

$$E_{MM} = E_{internal} + E_{elec} + E_{vdW} \tag{1.42}$$

$$E_{internal} = \triangle E_{bond} + \triangle E_{angle} + \triangle E_{tors} \tag{1.43}$$

$$G_{solv} = G_{elec,solv} + G_{nonpolar,solv} \tag{1.44}$$

where these correspond to the bond, angle, torsion, van der waals, and electrostatic terms in molecular mechanical force field, evaluated with no non-bonded cutoff. All the terms in Eq. 1.42 are averages of energies obtained fron a number of snapshots taken from the MD simulations [**?**]. In order to reduce the the time consumption and to obtain the stable energies, the same geometry is normally used for all three reactants (complex, receptor,ligand), i.e only the complex is explictily simulated by MD.There by, $E_{internal}$ cancels in the clculation of $\triangle G_{bind}$. The MM-PBSA methos has been successfully applied to many systems [**? ?**], [**? ? ?**]. G_{pbsa} is the solvation free energy calculated with the numerical solution of the Poisson-Boltzmann equation and an estimated of the non-polar free energy with the simple surface area term. The nonpolar solvation contribution is estimated as

$$G^{S}_{nonpolar,solv} = \gamma A + b \tag{1.45}$$

Where A is solvent-accessible surface area and the solvent parameters, γ and b, are 0.00542 kcal/mol $Å^2$ and 0.92 kcal/mol, respectively.the probe radius of the solvent was set to 1.4 . the atomic radii of of the solute were taken from PARSE parameter set [**? ?**]. $-TS_{MM}$ is the solue entropy, which can be estimated by quasi harmonic analysis of the trajectory or by using normal-mode analysis (108). the solvant entropy is included tin the solvation free energy term. The free energy due to ionic strength effects can be added with a continuum approch.

The ability to accuratly calculate $\triangle G$, the average free energy, we can calculate $\triangle G$ for a given macromolecular system in various different confermations or structures, adds a very important methodology to our computational arsenal. This has been possible before with free energy perturbation but only for small

systems and very limited conformational or topological changes. By using continuum model, it is implicity integrating out all the solvent coordinates and simplifying the problem. Also, by calculating the absolute free energy directly with eq 1.2 between the two "end points" instead of calculating the relative free energy along a mapping coordinate, we are avoiding computations of eq 1.1 will have intrinsically much larger errors than free energy perturbation/thermodynamic integration calculations. What is surprising is that, despite these larger uncertainties, we can often calculate $\triangle$G in respectable agreement define with experiment.(107)

III

Effect of Resistance Mutation on BVDV RdRp

The Structural and Dynamical Aspect

Point mutation I261M affects the dynamics of BVDV and its interaction with Benzimidazole antiviral 227G.

Abstract

Bovine viral diarrhea virus (BVDV) is a *Pestivirus* of the *Flaviviridae* family and represents a major viral pathogen in cattle and other ruminants. Infection with BVDV can result in a wide assortment of disease manifestations including resorption, mummification, or abortion of the dead fetus. Extensive study of BVDV is required not only because it poses heavy agronomic losses but also it considered to be a valuable surrogate model for the study of the hepatitis C virus (HCV). Recently the point mutation I261M on the thumb domain was shown to confer resistance to BDVD against 227G and other benzimidazole compounds.

Here, we investigated the role of this mutation by using a non conventional approach, which is not based on binding free energy calculations on structures of the mutated complex which are taken a priori similar to those of the wild one. Namely, we firstly performed MD simulations on the wild and mutated BVDV RdRp proteins in aqueous solution. Then, we selected representative equilibrium conformations by performing a cluster analysis, and ran docking calculations of 277G on representative of the 5 most populated clusters of each protein. Finally, high-score poses were subjected to MD simulations to assess structural and dynamical differences between wild and mutated 227G-protein adducts.

Interestingly, the mutation affects the structure and the dynamics of the protein, particularly in the region of binding of the ligand, and this results in different binding sites of 227G onto the two proteins. Moreover, while 227G closes the entrance for the RNA primer in the case of the wild protein, in presence of the mutation a gate and a channel leading to the catalytic site are still present. These results could furnish a possible molecular explanation of the resistance mechanism by mutation I261M.

In a previous paper, we reported mutagenesis experiments aimed at confirmCozing that BVDV RdRp is the target of our lead benzimidazole compound 227G. Here, we perform a multidisciplinary computational study to furnish a molecular-level explanation of effect of the resistant mutation. Our results are fully compatible with the mechanism of action furnished in the previous paper. Namely, we see that the mutation I261M alter the binding mode of the ligand, leaving the channels for the entrance of the template RNA, the entrance of NTP and the exit of double strand RNA all open. The dynamics of the protein is also altered to a less extent as compared to the wild-type.

Introduction

"Bottom-up approch : To develop antivirals from the information retrieve from the resistance mechanism "

Bovine viral diarrhea virus (BVDV), is a major viral pathogen in cattle and other ruminants [1][Erica Weiskircher et al , 2009, virology journal, Brock KV: The persistence of bovine viral diarrhea virus**.** *Biologicals* 2003, **31:**133-135]. It along with classical swine fever virus and border disease virus of sheep, belongs to the genus Pestivirus in the family *Flaviviridae.*

Infection of cattle with BVDV can result in a wide assortment of disease manifestations which include resorption, mummification, or abortion of the dead fetus while those fetuses who survive early infection may be malformed or blind, may have skeletal defects, respiratory problems, underdeveloped brain or weak immune system.[2]

which resulted in a high mortality rate in cattle throughout the world. Diseases related to BVDV cause economic losses to cattle producers throughout the world due to decreased performance, loss of milk production, reproductive wastage, and increased risk of morbidity and mortality.[3].It is the most costly viral disease in US cattle herds, with losses estimated at $10–$88 per head—or $2 billion dollars per year.[4]

As the disease is so widespread vaccination is a vital part of the bio security of the herd. Vaccines though highly efficacious, are mainly available for BVDV1A and BVDV2a strains, while different variations of the strain exist and would have serious consequences if introduced into a population of cattle never having been exposed to that particular variant, thus it remains an agronomical burden.[5,6,7]. Absence of effective therapy broadly targeting all genotypes has led to development of new strategies to combat BVDV infections. Potent Antiviral compounds specifically targeting the replication in RNA dependent RNA polymerase (RdRp), are now being used as alternative therapeutic option[8-22]. A number of selective anti-BVDV compounds have been reported.

Extensive study of BVDV is required not only because it posses heavy agronomical losses but also it considered to be a valuable surrogate model for the study of hepatitis C virus (HCV), belongs to the Flaviviridae family[23-28]. It is the frequent cause of acute and chronic hepatitis, liver cirrhosis and hepatocellular carcinoma worldwide, and creates a significant burden to healthcare systems due to mortality, morbidity and treatment costs[29-32] .The current standard therapy consists in administration of pegylated alpha interferon along with nucleoside analogue inhibitor (NI) ribavirin[33-36] , which is effective in only about half of patients who suffer from chronic HCV infection, and often produce toxicity and significant side effects[37-38] . Given the high prevalence of HCV, there is an urgent need to develop more effective and well tolerated therapies for chronic hepatitis C. [39-40]. Historically, the search for anti-HCV antiviral agents has been challenging due to the inability to infect cultured cells with HCV, and the lack of readily available animal models for HCV infection . A major breakthrough for evaluating candidate antiviral molecules in cell cultures is represented by the HCV replicons, i.e. engineered subgenomic HCV RNAs capable of autonomous replication once introduced in cultivated cells [41-42]. However, besides being expensive, HCV replicons do not produce infectious virions: thus, the replicon system can be used only for the identification of those inhibitors affecting genome replication and expression. In some aspects of viral replication, BVDV is more advantageous than the currently used HCV replicon systems. The latter do not undergo a complete replication cycle; hence, early stages (attachment, entry, uncoating) or late stages (virion assembly and release) of the viral replication cycle cannot be studied in the HCV replicon system[28,43-45]. Therefore, insight into the mechanism of antiviral activity of anti-pestivirus compounds may provide valuable information for the design of novel antiviral strategies against HCV[46].

Recently reported by us, the polymerase inhibitor 227G (2-phenyl benzimidazole derivative), is one of the most potent inhibitor against BVDV RdRp, strongly inhibiting BVDV RdRp activity in a dose-dependent man-

ner, with an IC_{50} value of 0.002 µM; HCV1b RdRp was also inhibited, even though with an higher IC_{50} value of 0.4 µM. Thus, viral polymerases were confirmed to be the molecular target of 227G, which appears to be in our knowledge the first one belonging to this class which is active against both viruses.

Several drug resistant point mutations of BVDV RdRp, such as F224S [8,10,21,12,15], F224Y have been reported in finger domain. In our previous study, we have mentioned compound 42 [47-48] and ASCC170 (Benzimidazole) lyes in finger domain, having resistant mutation I261M. Interestingly, we found that 227G, showed cross-resistance with compound42 and ASCC170. This mutation was found to lye in the fingertip region of BVDV RdRp. The fingertip region contains polymerase motifs I and II, which are involved in RNA template and NTP binding and hence is crucial for the polymerisation[49,50]. To the best of our knowledge, 227G is the most potent inhibitor against BVDV reported till date. Hence it becomes more and more intriguing to understand the mechanism by which the mutation I261M renders the protein resistant against 227G and thereby provide the basic information about general feature involved in drug resistance as well as for the direct application involving the design of the inhibitor that is effective against that mutant.

In this light, here we present extensive atomistic molecular dynamics (MD) simulations carried out to identify structural and dynamical features of the interaction between the mutated RdRp and 227G. We found that the structure and dynamics of the mutated protein feature some differences as compared to those of the wild, which reflect on the binding of 227G in a crucial way. On the basis of our results we provide a likely explanation of the molecular mechanism of resistance .

The aim of this work was to create a link between the resistant mutation and the binding property of 227G. The medicinal chemist can use this approach to predict the effect of NNIs relevant mutations in the presence of novel inhibitors.

Material and Methods

Biological

Cells and viruses:
Madin-Darby Bovine Kidney (MDBK) cells were used for growth and selection of mutants resistant of Bovine Viral Diarrhea Virus (BVDV). Either cells (CCL-34) and viruses (VR-534 strain NADL) were purchased from American Type Colture Collection (ATCC). The absence of mycoplasma contamination was checked periodically by the Hoechst staining method.

Selection of Drug-Resistant Mutants :
Drug-resistant variants to compound 227G were selected by serial passages of parental BVDV wild type in the presence of stepwise doubling drug concentrations. After 3 days of incubation, MDBK cells were seeded in 24-well plates and incubated overnight at 37°C in a humidified CO_2 (5%) atmosphere. Cell monolayers were infected, in the first passage, with a m.o.i of 0.01 and after two hours treated with a drug concentration equal to the EC_{50}. Multiwell were daily observed until they exhibited a 75-80% of cytopathic effect. The procedure was repeated (4 passages) and the amount of virus obtained after each passage was sufficient to determine infection of the next cell culture which, after infection and washing, was incubated with a double amount of the selecting drug. Drug-resistant virus population was selected up to a drug concentration 16-fold greater than the EC50.

Molecular Analysis of Resistant Viruses:
Resistant virus preparation was subjected to RNA extraction, RT-PCR and genome sequencing to identify the mutation patterns responsible for resistance. Viral RNAs from wt and drug-resistant mutants were obtained using the QIAamp viral RNA minikit (QIAGEN), starting from 140 µL of cell-free viral suspensions containing about 10^6 PFU/mL, in order to determine the nucleotide sequence of NS3 and NS5B genes of nonstructural region of BVDV genome. Reverse transcriptions were carried out using the Superscript II enzyme (INVITROGEN) and cDNAs were amplified by PCR using Platinum Pfx polymerase (INVITROGEN) following the manufacturer's protocol.

PCR fragments were purified using the QIAquick PCR Purification kit (QIAGEN) and analyzed using the cycle-sequencing method (CIBIACI service of University of Firenze). Both DNA strands were sequenced with specific primers. The comparative analysis of the chromatograms allowed us to deduce the mutation pattern responsible for resistance to 227G ($BVDV^{227G}$).

Primers and parameters used in the RT-PCR reactions.
The primers used in the reverse transcription reaction were RT3 5'-CCCCACAAACCATATCTGATTATTTCTTCTTTA-3' and RT5b 5'-GTAGATAATCTTGACTACTGTTTAGCTCTTGAG-3', both of 33nt, that respectively bind 360bp downstream the NS3 gene and 90bp downstream the NS5b gene. cDNAs were amplified by PCR using Platinum *Pfx* DNA Polymerase (INVITROGEN), following the manufacturer's protocol. The region containing *NS3* gene and was amplified by PCR reactions carried out with the primers A (5'-TAAAAATGCTCATGGTAGGCAACCT-3') and B (5'-TTATCATTGGGACATGCCTCTTTGA-3'), resulting in a PCR fragment of 2205bp. PCR consisted of an initial denaturation step of 3 min; 34 cycles of denaturation at 94 °C for 30s, annealing at 56°C for 30s and extension at 68°C for 2.5 min; final extension at 68 °C for 5 min. The region containing *NS5B* gene was amplified by two separated PCR reactions, carried out with the primers C (5'-ATTATAAAGGAGGTAGGCTCAAGGA-3') and D (5'-CCATCTGCTGTTATAACTGGTACTT-3') and with the primers E (5'-ACCCCCTTGTTCAACATCTTTGATA-3') and F (5'-GTGGACGGTCCCAACTATATTTATA-3'), respectively, resulting into two PCR fragments of 1223bp and 1792bp. The conditions are similar to first one the described above, except that the annealing temperature was 52.5 °C and the extension time was 2 min.

Computational

Structures of Receptors

The published X-ray crystal structures of BVDV RdRp [49] (PDB code 1S48; resolution 3.0 Å) was used as starting structures for molecular simulations. BVDV RdRp consist of 588 amino acid residues respectively (Figure1). Simulations of BVDV RdRp receptors in their un-complexed form (vide infra) were performed in order to assess the presence of major conformational changes in the proteins and obtain starting structures for docking. We performed another MD simulation with the Met modeled at residue 261, in APOm using VMD [54]

MD Simulation protocol

Parametrization.

The AMBER99, TIP3P, and AMBER-modified Aaqvist [55] force fields [56]were used to model proteins, water and ions respectively. Concerning 227G, since no experimental structure is available, the molecule was first drawn with ACD Chem. Sketch 11.0. The resulting molecular geometry was then optimized at the HF/6-31G(d)

level up to a convergence in energy of 10^{-5} au using the Gaussian03 package[57,58,59] . The *gaff* force field [60] was used except for charges, calculated following the standard AMBER protocol: from the electrostatic

potential map generated by the molecule (calculated with Gaussian03), the RESP [61] charges were derived using the *antechamber* module of AMBER.

Dynamics.

State-of-the-art all-atom MD simulations in the presence of explicit water and counter-ions were carried out with the ORAC package [62][REF] for both receptors, in their apo forms and in complex with 227G. Details of the simulation are reported in Table 1. For all of the system the following procedure was used: First, geometry optimizations were carried out with a two-step protocol: (i) 10000 cycles (2000 steepest descent plus 8000 conjugate gradients) with harmonic restraint of k=10 kcal/(mol Å^2) on each heavy atom of the solute; (ii) 20000 conjugate gradients cycles with no restraints. Next, heating up to 300 K was achieved by linearly increasing the temperature in 100 ps of *NVT* MD, while imposing restraints of 1 kcal/(mol Å^2) on the solute. Restraints were then released for 100 ps, and as a last step preceding the productive dynamics, 1 ns of NPT MD was carried out in order to relax the simulation box. Finally, 20 and 10 ns long simulations were performed respectively for each apo-enzyme and adduct in explicit water solution under the NPT ensemble. Pressure and temperature were regulated at 1 atm and 310K using the isotropic Andersen–Parrinello–Rahman barostat [63-65] and the Nosé-Hoover thermostat [66-67] respectively.

Electrostatic interactions were evaluated using the soft Particle Mesh Ewald schemes [68-69] with 1 Å grid spacing and a cut-off of 10 Å, the same used for Lennard-Jones interactions. A MTS-Respa algorithm [70-71] with five shells (time steps of 0.5-1-2-4-12 fs) was used to integrate equations of motion, in conjunction with the SHAKE algorithm [72] applied to bonds involving hydrogens.

Clustering:

Clustering of protein structures performed for each system over the last 20 ns of MD simulation (2000 frames). The clustering method developed by [73] was used together with the *g_cluster* module of the GROMACS 4.0.7 package[74-75]. A cutoff of 0.2 nm was used, and only the “scaffold” of the protein (not including the most flexible regions of loops L2 to L4, N-terminal domain – including loop L1 – and fingertip – including motifs I and II) was selected for structural alignment, with the purpose of highlighting the structural differences in the region of interest, where 227G is believed to bind [76]. In each system two main clusters were found grouping together more than 75% of all conformations, while 95% of conformations are contained in the first four clusters. Interesting, a really dominant cluster has been found only in the presence of mutation. The corresponding representative structures of Clus1 and Clus2 for each system are shown in Figure 2A1.

Structures of Complexes

Docking:

Structures of the complexes 227G-BVDV was generated through molecular docking using the Autodock 4 package [77-79]. As a result top four conformations of proteins were picked for further molecular docking study in both cases. The cluster analysis was performed on the backbone atoms of the molecules, with the 0.15 RMSD cut-off (nm). We ranked the clusters (Table 1) according to their relative population.

In the absence of literature data indicating the presence of allosteric sites in BVDV RdRp, the C_α atom of Ile261 was chosen as the grid center for all docking runs. The choice is guided by mutagenesis experiments performed on BVDV, showing that mutation Ile261Met produces resistance to 227G (see Ref. Coupled paper; Ile261Met mutation is resistent mutant for phenyl-benzimidazole[80] and arylazoenamine [81]. For the complexes grid maps were made of 30x40x40 points in x, y, z directions, distant 0.375 Å from each other.

The Lamarckian genetic algorithm was used [79] with the following parameters: The global optimization started with a population of 150 randomly positioned individuals, a maximum of 2.5 x 10^8 energy evaluations, and a maximum of 27,0000 generations, as mentioned the protocol for blind docking[82]; mutation rate of 0.02; crossover rate of 0.80; elitism value of 1. For local search, pseudo-Solis and Wets algorithms were applied by the default parameter [83,79]. A total of 200 runs was performed, while all the remaining run parameters were maintained at their default settings. A cluster analysis was carried out using 2 Å as the root-mean-square deviation tolerance.

Finally, the conformer for 227G-BVDV complex was selected on the basis of 1) cluster population 2) binding energy [84-86], and 3) The vicinity to the mutation in the case of BVDV-227G complex. These complexes were used as the starting structures for MD simulations.

Analysis of structures and dynamics.

For each analysis we took 10ns of production run. An inventory of the structural details and interactions was obtained by performing an extensive qualitative and quantitative analysis of the MD simulations in terms of hydrogen bonds (HB), hydrophobic contacts (HpH) and water-mediated interactions. The HB's between 227G and RdRp are counted using VMD scripts[54] with the following threshold parameters: a distance of at most 3.2 Å and donor-hydrogen acceptor angle larger than 150°. HpHs are counted when non-polar atoms are separated by at most 3 Å. The durable HB is defined as the one with a lifetime equal to or higher than 20% of MD simulation time, and transient HB are the one with a lifetime shorter than 20% of simulation time [87]. For SB (Salt-Bridge), the cutoff distance is 3 Å. The criterion to determine a π-π interaction was that the SICD (Short inter atomic carbon-carbon distance) should be smaller than 4.8 Å [88].The RMSD residues wise was plotted taking reference average snapshot from the equilibrium dynamics to characterize the flexibility of the protein, in particular in the template entrance region, the behaviour of the minimum distances between key residues in flexible loops in this region, and the area of the entrance region along the whole equilibrium dynamics.. The energetic analysis was performed in the usual manner: the interaction energy was decomposed as the sum of individual components, each with a physical meaning and evaluated via the force field terms. In particular, we examined variations in the 227G-BVDV RdRp electrostatic and van der Waals interactions.[89-92]

Area Analysis:

In order to calculate this area, we considered the triangle defined by the minimum distances among the loops. At every MD step minimum distances among L1, L2, and L4 were calculated. Since the atom of L1 at the smallest distance from L2 is generally different from that at the smallest distance from L4, in order to simplify the calculation we calculate the area of the triangle whose vertex are defined by the centers of mass of minimum-distance atoms in every loop. We are aware that the area calculated in this manner is an approximation and its absolute value has no quantitative meaning. Nevertheless, a comparison of the mouth areas provides a qualitative indication of the space available to the template RNA strand to enter the polymerization site.

Results and Discussion

Biological

Molecular characterization of drug-resistant mutants.

In an effort to determine the molecular target underlying the antiviral activity, mutants resistant to compound 227G were selected ($BVDV^{227G}$) by serial passages of WT-type (WT) BVDV in the presence of stepwise doubled drug concentrations. Selection started from cell cultures infected with a m.o.i. of 0.01 and, for the first two passages, treated with a drug concentration equal to EC_{50}. Thereafter, the amount of virus obtained after each passage was sufficient to carry on the infection of the next cell cultures, which, after infection and washing of the unabsorbed inoculum, were incubated with stepwise doubling amounts of the selecting drug. At final passage, 227G resistant mutants were grown up to a drug concentration 16-fold higher than the EC_{50}. A single-cycle growth experiment was finally performed to determine the replication ability of resistant strains in the presence and absence of corresponding inhibitor. Cells were infected with either wt BVDV or resistant strain, in the presence or absence of each compound (at finally concentration of $16xEC_{50}$), and progeny virus was harvested and titrated on MDBK cells without compound. The yield of wt virus was inhibited at least 1,000-fold in the presence of compound, while no apparent reduction was observed with the resistant strains.

In order to identify the mutation pattern responsible for drug resistance, the selected virus population was subjected to RNA extraction, RT-PCR and sequencing of region of the BVDV genome encoding for nonstructural proteins NS3-NS5B, the principle enzymes involved in replicative complex: NS5B, the RNA-dependent RNA polymerase (RdRp); NS3, a multifunctional protein with helicase, NTPase and protease activity. Comparative genomic analysis of parental $BVDV^{wt}$ and $BVDV^{227G}$ showed, in the latter, a mutation located in the gene encoding for NS5B protein, while no mutations were detected in the gene encoding for NS3 protein. The mutation consists in an A-to-G substitution at nucleotide (NT) 783 that changed the amino acid coding sequence from an isoleucine to a methionine at amino acid 261 of NS5B, suggesting an important role of this residue in the binding of enzyme with the benzimidazole structure of inhibitor.

The mutation occurring in $BVDV^{227G}$ is localized in the fingers domain of the RdRp structure, suggesting that this domain could be involved in the binding of benzimidazole compounds. This is in good agreement with available literature knowledge [10,12], while is worth noting that, in HCV, the mutations conferring resistance in particular to NNI and to benzimidazoles [93] , are instead located in the thumb domain. Relevant concentration at which $227G^{r}$ virus was isolated was found to be 16^{X}. High level of resistance of mutation I261M in BVDV RdRp against of 227G made it necessary to understand as to how the mutation renders the protein resistant. Thus, an in depth study, at microscopic level was performed as mentioned in the following subsections.

Computational

In order to understand the molecular determinants behind occurrence of resistance in the mutant BVDV RdRp, we used here a multidisciplinary approach including docking calculations and molecular dynamics (MD) simulations. Four systems were investigated, namely the wild type and mutated apo-proteins (APO and APOm respectively) and the corresponding complexes with 227G (COM and COMm respectively).

As reported in 1st paper, while the docking pose of 227G on HCV is found to be consistent with available experimental data, such information is missing for BVDV. Therefore, building a reliable initial model of the complex requires more care in this case. It is well known that docking algorithms are extremely sensitive to small adjustments occurring within a binding site, and that flexibility, in particular of the target, is often crucial to obtain reliable results. For this reason, we use the following protocol to find significant structures of the complexes between 227G and the proteins (Figure 1). At first, we ran 30 ns long MD simulations of the two apo-proteins (wild and mutated, respectively APO and APOm hereafter). Then, we performed a cluster analysis to identify the different conformations assumed by the proteins and use them for multi-conformer docking. Only the palm region of the protein was selected for structural alignment, with the purpose of highlighting the structural differences in the region of interest, where 227G is believed to bind . In each system two main clusters were found grouping together more than 75% of all conformations, while 95% of conformations are contained in the first four clusters (Table 1). Interestingly, a really dominant cluster has been found only in the presence of mutation, while two almost equally populated ones are present in APO.

In Table2 are reported the values of the RMSD of the whole protein along with those of the most important regions with respect to the wild type X-ray structure in [49]. It can be seen that the highest deviations are systematically found in APOm, while the RMSD values of the APO protein are quite standard. The difference in RMSD between APO and APOm values can be as large as 2 A for the entire protein, and the largest values are found again in the N-terminal domain. Indeed, the presence of the point mutation I261M affects quite significantly the structure (and the dynamics, vide infra) of the protein: the RMSD has a value of 4.3 A between the closest-to-average structures along the dynamics, and even larger values are found in the N-terminal domain and between the conformations assumed by the linker (Table 3). This can be seen more clearly in Figure 2, where we report the superimposition among the closest-to-average structures assumed by the proteins in the two systems, as well as the conformations of the binding pocket region (enclosed by loops L1, L2, L3 and L4 - linker - and by motifs I and II) in the first four clusters of each system. In particular, it can be seen that the two N-terminal domains assume very different conformations. Furthermore, with respect to the X-ray structure, the linker loop is significantly displaced in APOm in all but the second cluster, while in APO it keeps closer. Interestingly, these structural distortions occurr with minor changes in the secondary structure of the apo protein (Figure 3, vide infra). In the following, we will analyze the effects of such distortion on the interaction with 227G.

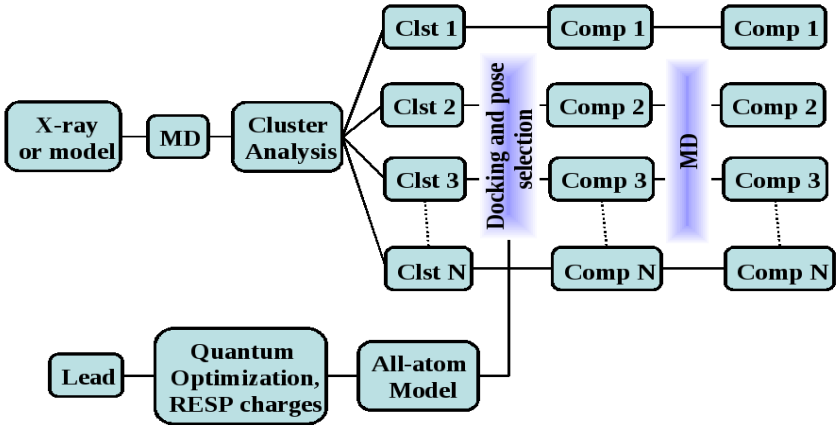


Figure1. Scheme used to find structures of the complexes between 227G and BVDV RdRps.

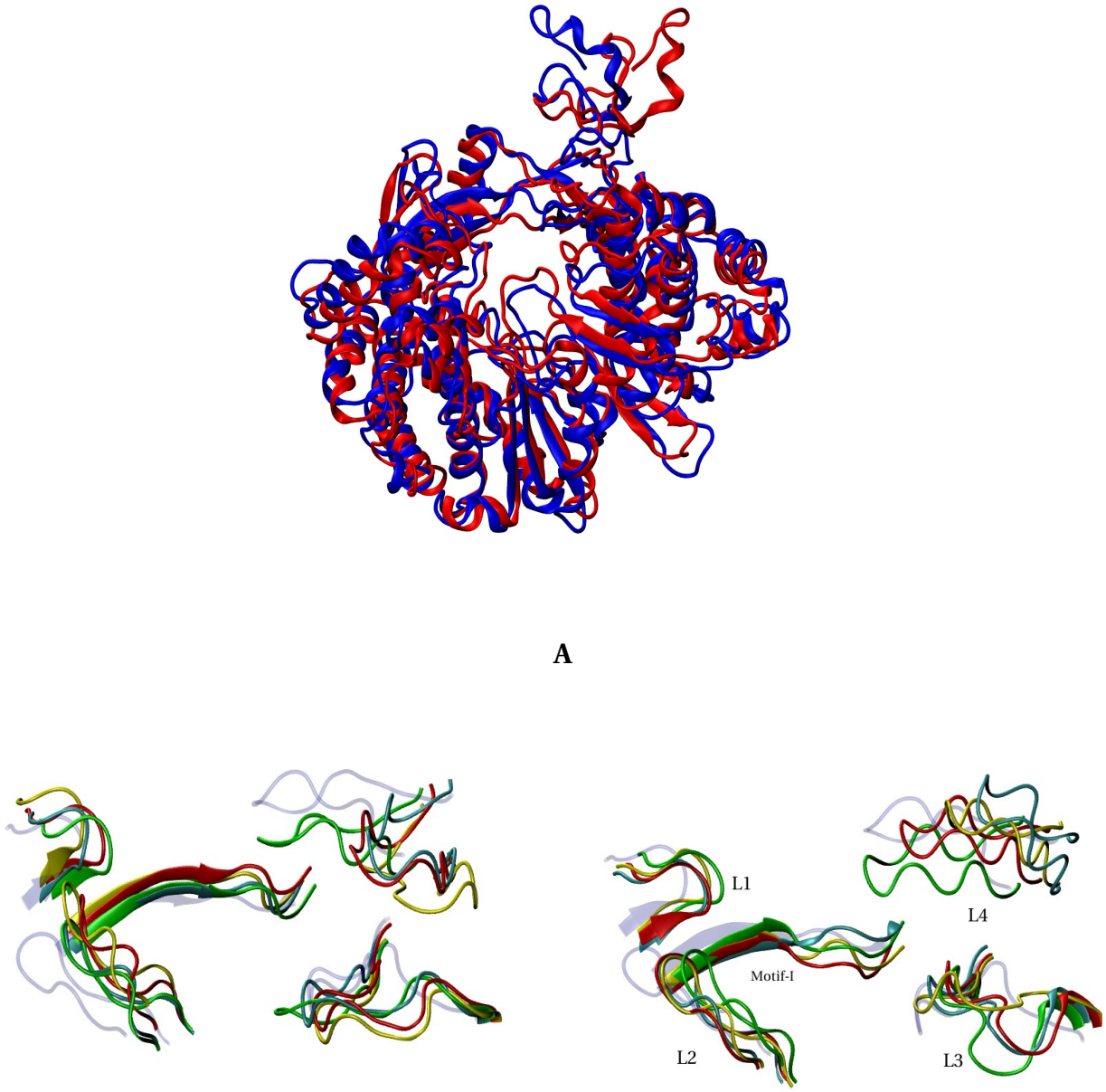


Figure2. A) Superimposition of Apo(Red) and APOm(blue) close to average structures; B and C) Schematic representation of the binding sites of APO (B) and APO_m (C). The crystal structure is colored in blue (transparent), clus1 in red, clus2 in green, clus3 in yellow and clus4 in cyan.

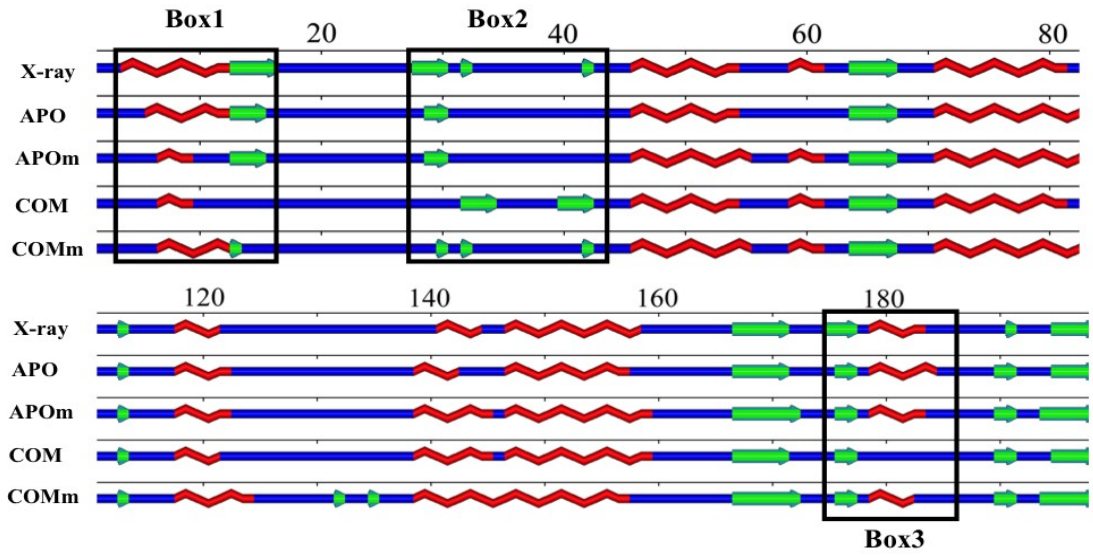


Figure3. Secondary Structure (SS) diagram of the BVDV RdRp in the five systems X-ray, APO, APOm, COM and COMm. Beta strands and alpha-helices are represented by green and red colors respectively. The protein conformations selected for SS calculation are those with the lowest RMSD from the average structure calculated over the equilibrium dynamics. Only part of the sequence is shown, namely that containing regions with major changes along the dynamics or biologically relevant domains, highlighted by black boxes.

Cluster	**APO**	**APOm**	**COM**	**COMm**
Clus1	902 (45.1%)	1473 (73.6%)	906 (45.3%)	1307 (65.3%)
Clus2	795 (39.7%)	445 (22.2%)	669 (33.4%)	386 (19.3%)
Clus3	144 (7.2%)	51 (2.5%)	325 (16.6%)	192 (9.6%)
Clus4	68 (3.4%)	24 (1.2%)	79 (3.9%)	113 (5.6%)
Clus5	63 (3.1%)	8 (0.4%)	20 (1.0%)	3 (0.1%)
Clus6	18 (0.9%)	-	1 (0.0%)	-
Clus7	11 (0.5%)	-	1 (0.0%)	-

Table 1. Clustering of protein trajectories. Clustering of protein structures performed for each system over the last 20 ns of MD simulation (2000 frames). In each system two main clusters were found grouping together more than 75% of all conformations, while 95% of conformations are contained in the first four clusters. Interesting, a really dominant cluster has been found only in the presence of mutation. The corresponding representative structures of Clus1 and Clus2 for each system are shown in Figure A1

	Clus1		**Clus2**		**Clus3**		**Clus4**	
	APO	APOm	APO	APOm	APO	APOm	APO	APOm
Protein	2.4	4.4	2.6	4.0	2.9	4.9	2.6	3.7
Binding site	2.3	3.4	2.6	2.5	3.0	4.5	3.4	3.5
N-terminal	3.4	5.4	3.2	5.5	3.4	5.9	3.6	4.8
Linker	1.5	2.1	1.6	1.8	1.7	1.9	1.7	1.8

Table 2. RMSD (Å) of cluster representatives with respect to the X-ray structure of wild BVDV RdRp [49]. In the table are reported the values of the RMSD of the whole protein along with those of the most important regions. It can be seen that in case of the wild protein (APO) clus1 and clus2 are almost equidistant from the X-ray structure. The values for APOm are usually larger than those for APO, which could have been expected.

Average Structure			
Protein	Binding-site	N-terminal	Linker
4.3	5.4	11.5	8.4
Cluster (protein)			
Clus1	Clus2	Clus3	Clus4
3.8	3.6	4.9	3.9

Table3. RMSD (A) between the closest-to-average and cluster representative structures of APO and APOm.

Docking of 227G on APO and APOm

Extensive docking calculations (8 runs in total) were performed using as target the representative structures of the first four clusters extracted from APO and APOm dynamics (see Table 1). In Table 4 and Figure 4 are reported the results and the conformations of the 16 highest-score poses. The criterion we used to select these poses is the standard one [86-88](see Methods for a deeper description and figure 2 for scheme), i.e. we choose, for each docking run, the representative pose representing the largest cluster of ligand conformations (to not be confused with the cluster of proteins conformations described above) among those with the best energetic score. This resulted, for each protein and each protein cluster, in two different orientations of the ligand within the binding pocket.

Due to limited computational time, we initially selected only one representative complex for each protein and for each orientation of the ligand (four complexes in total) for the subsequent MD simulations. This somewhat limits the purpose of our approach, because it can not be excluded *a priori* that a different starting conformation could lead to different dynamics on the time scale explored here[97] . Nevertheless, it turns out *a posteriori* that the choices we made on the starting structures do indeed reinforce the outcome of our study (vide infra). Indeed, looking at Figure 4 it can be seen that for both the wild and mutated systems the overall features of the binding region in complexes 1 are similar to those in complexes 3 and 4, while are significantly different from those of complexes 2. In particular, the linker loop which is in closer contact with 227G only in complexes 2. Because of this, and since clus1 and clus2 are the most populated in the dynamics of the apo proteins (Table1), we restricted the choice to these two clusters only. On the basis of the same criteria discussed above for the selection of the highest-score poses, in the case of APO we selected clusters number 2 and 1 when ligand assumes respectively orientations 1 and 2, while in the case of APOm cluster number 2 was chosen for both orientations.

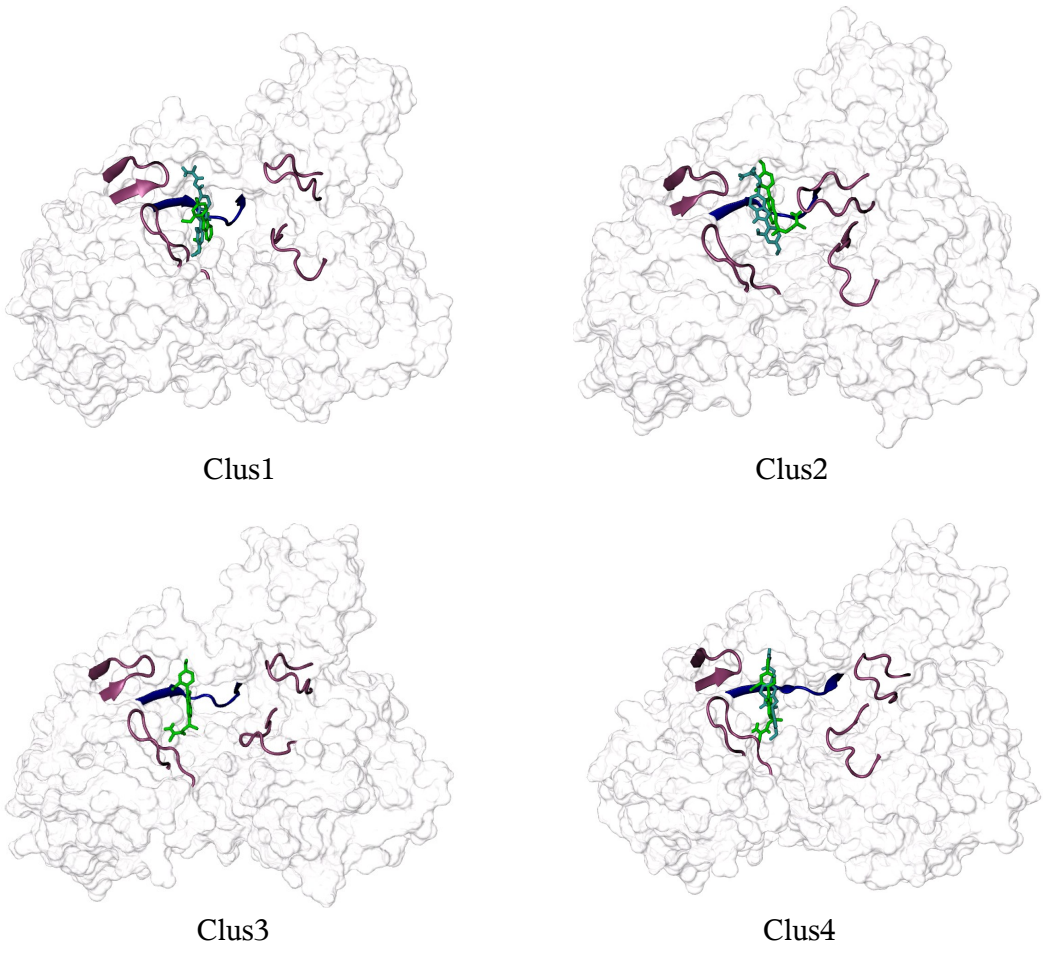
Interestingly, it turned out that, during the subsequent equilibrium MD simulations only orientation 1 is stable in the case of APO, while the ligand dissociates from the protein after ~3.5 ns when assuming orientation 2 (data not shown). At opposite, only the second orientation is stable in APOm. Indeed when starting from orientation 1 the ligand flips by ~180% after ~3 ns assuming the orientation 2 (Figure 5). In view of these results, we decided to perform two additional MD simulations using the structures of the protein in clusters number 1 with 227G assuming orientations 1 and 2 in APO and APOm respectively. In this way, for each system we can investigate the effect of two significantly different starting conformations on the dynamics of the complexes. As we will see later, the main results of our simulations are very robust against this parameter.

Hereafter we will indicate as COM and COMm the adducts between APO and APOm representatives of clus2 and 227G with orientation 1 and 2 respectively. The same adducts with representative proteins' structures of clus1 will be referred as COM' and COMm'. We performed four 30 ns long all-atom MD simulations of the aforementioned complexes. For the sake of clarity, we discuss in the following the analysis of the structural and dynamical features of APO, APOm, COM and COMm, leaving the comparison with COM' and COMm' to the end.

		Orientation1			Orientation2		
	Cluster number	Energy (kcal/mol)	# Confs	%	Energy (kcal/mol)	# Confs	%
APO	Clus1	**-7.3**	**60**	**30**	**-7.9**	**88**	**44**
	Clus2	**-7.9**	**174**	**87**	-8.2	5	2.5
	Clus3	-7.3	97	48	0	0	0
	Clus4	-8.5	52	26	-8.3	36	18
APOm	Clus1	-8.2	76	38	**-7.9**	**97**	**49**
	Clus2	**-8.8**	**122**	**61**	**-8.7**	**110**	**55**
	Clus3	-7.2	122	61	-7.0	36	18
	Clus4	-7.4	39	20	-7.4	106	53

Table 4. Results from docking of 227G on representative structures of the first four clusters found in APO and APOm systems (see Table 1). For each representative, 200 docking poses were evaluated, and for each protein two opposite orientations of the ligand were found in the same binding pocket. The columns heading "#Confs" and "%" represent respectively the number and the percentage of poses which belong to the same cluster of binding positions. Orientations 1 and 2 refer respectively to binding positions in which the benzyl group and the sulphur atom of the ligand are pointing towards the template entrance gate of the protein (see Figure 6). For details on the docking scheme and the criteria used see Materials and Methods section.

APO



APOm

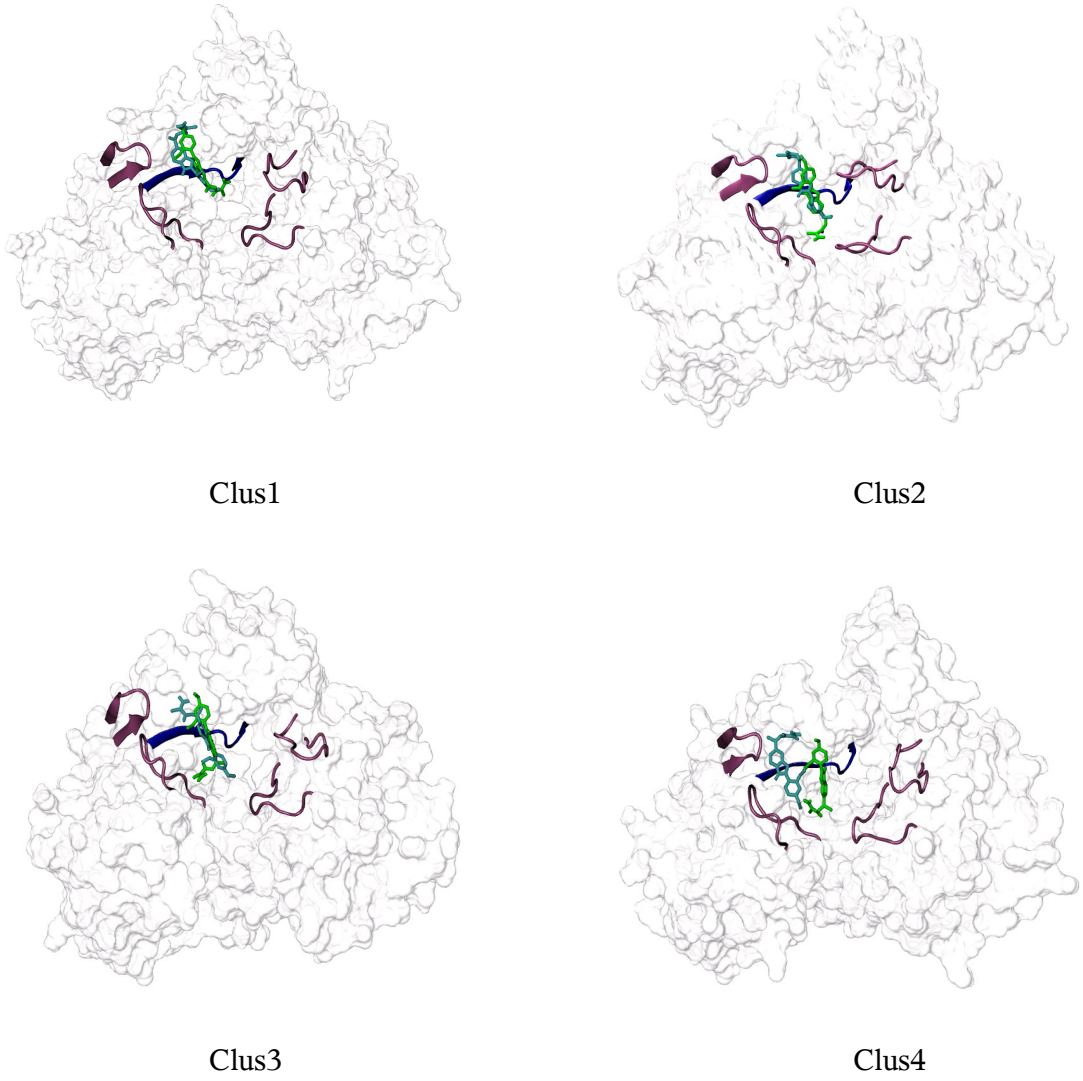


Figure4. Orientations 1 and 2 of 227G obtained from docking, refer respectively to binding positions in different clusters of protein in which the benzyl group and the sulphur atom of the 227G are pointing towards the template entrance gate of the protein in all most at the same position but opposite in orientation. The loops L1 to L4 and motif-1 render in cartoon while rest of the protein in surf view.

Figure 5. The movement of 227G from the initial pose with orientation 1 in APOm (between L1, L4 and motif I, red stick representation) to the final and stable one (yellow) enclosed by L2 and motif I. Blue and green conformations are intermediate ones along the flipping process.

Due to limited computational time, we initially selected only one representative complex for each protein and for each orientation of the ligand (four complexes in total) for the subsequent MD simulations. This somewhat limits the purpose of our approach, because it can not be excluded *a priori* that a different starting conformation could lead to different dynamics on the time scale explored here [94-96,103]. Nevertheless, it turns out *a posteriori* that the choices we made on the starting structures do indeed reinforce the outcome of our study (vide infra). Indeed, looking at Figure6 it can be seen that for both the wild and mutated systems the overall features of complexes 1 are quite similar to those of complexes 3 and 4 and different from those of complexes 2(Figure 1). In particular, the linker loop is in closer contact with 227G only in complexes 2. Because of this, and since clus1 and clus2 are the most populated in the dynamics of the apo proteins (Table1), we restricted the choice to these two clusters only. Based on the same criteria discussed above, in the case of APO we selected clusters number 2 and 1 when ligand assumes respectively orientations 1 and 2, while in the case of APOm, cluster number 2 was chosen for both orientations. Interestingly, it turned out that in the case of APO, only orientation 1 is stable, while the ligand dissociates from the protein after ~3.5 ns when docked with orientation 2 (Figure 1). At opposite, only the second orientation is stable in APOm. Indeed when starting from orientation 1 the ligand flips by ~180% after ~3 ns assuming the orientation 2. In view of these results, we decided to perform two additional MD simulations using the structures of the protein in clusters number 1 with 227G assuming orientations 1 and 2 in APO and APOm respectively. In this way, for each system we investigate the effect on the dynamics of two significantly different complex starting conformations. As we will see below, the main results of our simulation are very robust against this parameter. Hereafter we will indicate as COM and COMm the adducts between APO and APOm representative structures of clus2, and 227G assuming respectively orientations 1 and 2. The same adducts with representative proteins' structures of clus1 will be referred as COM' and COMm'. We performed four 30 ns long all-atom MD simulations of the above complexes.

For the sake of clarity, we discuss in the following the analysis of the structural and dynamical features of APO, APOm, COM and COMm, leaving the comparison with COM' and COMm' to the end.

Analysis of the dynamics trajectory:

All the simulations were converged to stable values of the RMSD and energy after ~10 ns (Figure 6). Kinetic and potential energies of the system also converged after few ps of MD simulation (data not shown). Therefore, all the results we present below refer to the last 20 ns of equilibrium dynamics.

Secondary structure changes.
As reported in the previous section, secondary structural analysis were performed using polyview [105] on the four average structures extracted from the dynamics. Some notable structural changes were found and detailed analysis was performed. The major differences with respect to the X-ray structure were seen in linker (which is believed to recruit other proteins [50]), hood (involved in dimerization [49,50]), and tip of the finger-tip region (which has been shown to regulate the coordinated movements of the fingers and thumb domains during the polymerase reaction cycle of BVDV and other RdRps, while rest of the region including the catalytic site remain conserved (Figure 3). These three regions are enclosed by black boxes in Figure 3:

Box1 corresponds to the hood of the N-terminal. The beta strand/alpha-helix motif present in the X-ray structure has almost disappeared in COM, whereas it was found to be shortened in COMm. A visual inspection of the trajectory showed that this can be due to the tight binding of 227G to the linker in COM. This interaction renders the linker rigid, inducing a strain in the rest of the hood region (residues Val92 to Ser122), which at the end cause the rupture of an intra HB pattern of hood helix (picture of this? What about beta-strand?). In APO instead the secondary structure is almost fully retained, while a partial loss of helical arrangement is seen in APOm.

Box2 encloses the linker region, which is believed to recruit other proteins in the polymerization process [49]. As said above, when compared to the X-ray, COM shows the largest structural changes due to the strong interaction with the ligand which cause the formation of a new beta strand.

Box3 delimits the region of the finger-tip in contact with the thumb domain. Here a major change is only observed in COM, where a complete loss of the helical structure occurs, which could influence the interaction with incoming nucleotides during polymerization.

In conclusion, the largest changes are generally found in COM, which points to a different and crucial effect of the ligand on the structure of the wild protein. In particular, as we will see later the outcome of structural motifs in the linker is consistent with a strong reduction in the conformational flexibility of this region as compared to the other systems.

Interestingly, only minor changes are seen in APO with respect to APOm, which is consistent with the retain of functionality by the mutated enzyme.

Overall pertubative changes in the four models

The RMSD (Root Mean Square Deviation) of protein's backbone atoms with respect to starting structures of the APO, APOm, COM and COMm systems are reported in Figure 6A. As described in material and methods section, structural alignments have been performed onto the palm sub-domain. To take into account completely stabilized systems all the analysis has been carried out in the last 20 ns of the simulations. It can be clearly seen that RMSD of COM and COMm were stable after 4ns, showing similar dynamical evolution. In case of APO the system seems to show a periodic behavior after 8 ns of simulation and oscillates around an average value of 3.5Å. A similar behavior is shown by APOm system (average RMSD at 3.8Å) although the amplitude and frequency of the oscillations are respectively larger and lower as compared to APO.
While this picture is helpful in conveying the dynamics of the overall structures, we were interested in understanding which parts of the protein were more flexible and affected by the I261M mutation by the interaction with 227G. Henceforth, we monitored the RMSD of the backbone atoms in function of the residue number of protein and ligand (Figure 6B). It is obvious from the plot that the systems undergo different conformational changes in the four simulations.

The maximum deviation is observed always at the N-terminal domain (residue 92 to 142). The N-terminus domain along with a long insert in the finger domain (260 to 288), form a fingertip region. The fact that flexibility of the finger-tip region is an intrinsic property of the polymerase, which is necessary for the binding and translation of template during elongation [49], justifies the deviation of N-terminal domain in the APO protein. Nevertheless this region exhibits even higher deviation in the APOm protein, again pointing

Using polyview [104], secondary structural analysis were performed for the four average structures over the whole simulation period, and their respective changes are shown in figure2. We compared the changes in alpha helices and beta sheets of our systems with respect to the x-ray structure, targeting mainly putative NNI binding and polymerase catalytic site. Some notable structural changes were found in our systems and detailed analysis were performed. We summarize the characteristics of the secondary structure in both the modeled and x-ray crystal structure.

In the four models, we observed the major notable difference in linker (recruiting other protein[50]), hood (in dimerization[49,50]), and tip of the finger-tip region (concerted movement of thumb and finger domain(Di macro, 2005; JBC) when compared with the x-ray, while rest of the area including catalytic site remain conserved. In figure2 the box2 correspond to linker or the L4 loop. The L4 loop, in the COM exhibited major difference with respect to other three system, where in formation of newly beta sheet was observed, which lacked in x-ray, APO and APOm system respectively, while it is comparatively short in COMm, originally formed with residue Glu123 to Asn132. In fact, in APO, APOm, and COMm system this beta sheet is not consistent during the time evolution of MD simulation but it is very much prominent in case of COM. Interestingly, the length of the sheet in COM, is the highlighted feature, which is absent in rest models including x-ray.

Another noteworthy secondary structural change observed in unique hood region (figure2, box-1). Here, the most substantial change is the loss of alpha helix and beta sheet in COM with respect to others models. In x-ray beta sheet consist of ten residues started from Glu95 to Arg105, whereas in other model APO, APOm and in COMm it is shortened. A visual inspection of the trajectory showed that due to the tight binding of 227G with the linker region, which makes linker rigid, inducing a strain in the rest part of the hood region residue Val92 to Ser122, causing a local structural change leading to breaking of intra HB with in the helix and which eventually results into the reduction of alpha helix. With respect to x-ray there is a consistency in APO while reduction of alpha helices observed in APOm, while in COMm it persist with shortened beta sheet.

The third major change was observed in the tip of the fingertip region. A complete loss of alpha helices occurred in COM with respect to x-ray originally formed by residues Val269 to Ala275. On the contrary all the rest models retain their secondary structural property.

This initial information already suggested that in the complex system the loops near the putative binding site as well as the fingertip region might be more constraint and less flexible than in other three system viz, APO APOm and COMm.

Overall pertubative changes in the four models

The Carbon α RMSD (Root Mean Square Deviation) from the formalized starting structure of the APO, APOm, COM and COMm systems are reported in Figure 5. As described in material and methods section, the structural alignments have been performed onto the most rigid palm sub-domain. To take into account completely stabilized systems all the analysis has been carried out in the last 20ns of the simulations. It can be clearly seen that RMSD of COM and COMm were found to stabilize after 4ns at average of 2.5 Å, showing similar dynamical evolution. In case of APO the system seems to show a periodic behavior after 8ns of simulation and it oscillates around an average value of 3.5Å. A similar behavior is shown by APOm system wherein although system evolves periodically but with a larger amplitude and lower frequency as compared to the APO system and averages around 3.8 Å.

Figure6 is helpful in conveying the dynamics of the overall structures; however we were interested in

understanding which parts of the protein were more flexible and affected by the I261M mutation and also in particular residues that are involved in interaction with 227G. Henceforth, we monitored the RMSD of the backbone atoms in function of the residue number of the entire system, for each trajectory, (shown in Figure 16 supplementary). The difference in the dynamical behavior of the four systems is more highlighted in Figure 16. It is obvious from the plot that the systems undergo different conformational changes in the four simulations.

The maximum deviation is observed in the N-terminal domain (residue 92 to 142). The N-terminus domain along with a long insert in the finger domain (260 to 288), form a fingertip region. The fact that flexibility of the finger-tip region is an intrinsic property of the polymerase, which is necessary for the binding and translation of template during elongation [49], justifies the deviation of N-terminal domain in the APO protein. Nevertheless this region exhibits even higher deviation in the APOm protein.

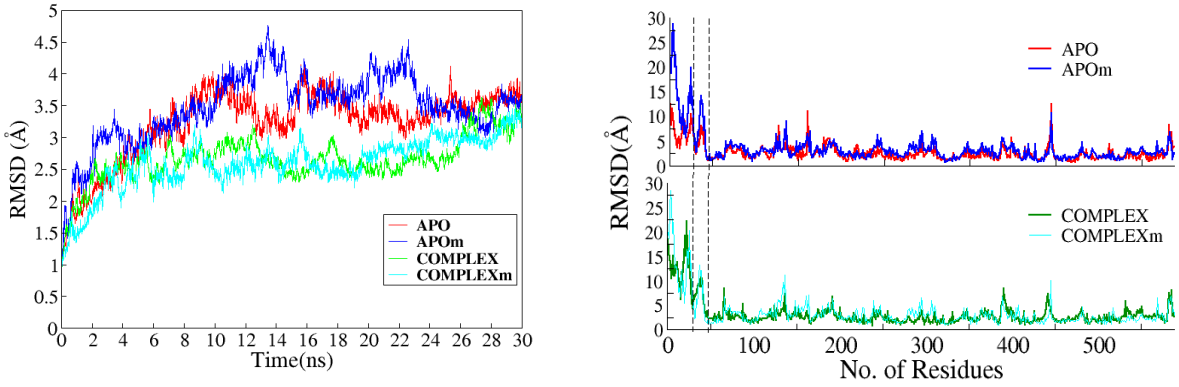


FIGURE 6 A)-Root-mean-square deviation (RMSD) of Four system (APO, APOm, COM and COMm) of backbone atoms as a function of time. B)-RMSD value per residue averaged over last 20ns with respect to starting structure.

Interestingly, the loop L4 (linker), a part of the N-terminal domain, along with loops L3 (thumb domain), L2 and L1 (finger domain) constitutes 227G binding cavity (1st paper). These loops along with the fingertip region make the template entrance channel. Figure16 of supplementary, shows striking dissimilarity in the RMSD of the loop L4, wherein the APOm has a maximum RMSD of 16Å and APO has 11 Å. By visual inspection of the trajectory it was found that there was a remarkable deviation of the loop L4 from the putative binding cavity with respect to the crystal structure as compared to the APO system. Similarly, in case of the COM and the COMm system, the RMSD of loop L4 was found to be 12 Å and 15 Å respectively. During the simulation, in case of the complex system, due to the presence of the 227G, the loop L4 moves closer and interact with it, this leads to the high deviation of the L4 region with respect to crystal structure. While on the contrary, in the COMm, there was no any interaction of 227G with L4 loop so, it is free to move and hence a high RMSD value is perceived. The dynamical behaviour of loop L1 and L2 was found to be similar in APOm and APO protein in comparison to the complex protein which showed the least deviation. Loop L3 region exhibited almost the same behaviour in the mutated and wild protein except for high peak residue Arg536 which is Hbonded to residue His578 in APO, while this type of interaction was never present in the APOm protein dynamics: overly the complex dynamics was found to be most stable with least RMSD.

The above results assisted us to throw some light on the probable mechanism of resistance of I261M mutated protein against 227G, i.e., by enhancing the mobility of the loops, in particular Loop L4. In the APO protein, the L4, if flexible, while in the COM system, 227G interacts with the loop L4 thereby is restricting its movement, as can be very well interpreted while comparing the RMSD of the three simulations. In the mutated protein the loop L4 shows an enhanced flexibility leading to a conformational change. Thus, apparently, the interactions provided by the residues of L4, which is important for the stability of the 227G is lost and this leads to the inhibitors resistance.

However, we were also interested in seeing which part of the protein were more flexible and affected by the resistant mutation, so we monitored the root mean square fluctuations of the Cα atoms (i.e, somewhat analogous to crystallographic B-factors) for each residue during MD simulations. figure7 compares the calculated B-factors with the experimental ones from the x-ray structure. our data indicate that the larger fluctuations are concentrated in two regions corresponding to residues 123 to 133 called linker and 436 to 446 called loop3 of the binding site, which are flexible loops on the surface of the protein as reported in uncomplexed core domain structures. The second loop 436 to 446, exhibits almost the same behavior in all three simulations except APO. The reduced fluctuation in case of APOm is due to the interaction with L4 loop, which deviate from it initial position and comes closer to L3 loop. In case of COMP the loop L3 has less flexibilility than COMm (peak2), as the flexibility reduces because also this loop is in contact with the 227G making hydrogen bond having occupancy more than 46 %, whereas this interaction is never present in COMPm. In any event, this region is far away from the active site, so we focused out attention on the loop containing residues only on linker, because this region has been found to adopt several different conformations in previous structure, and its mobility appears to be correlated with RdRP catalytic activity. Moreover, our secondary structural analyses showed that the length of the loop is not constant in two complexes, being absent in COM system in comparison to mutant.

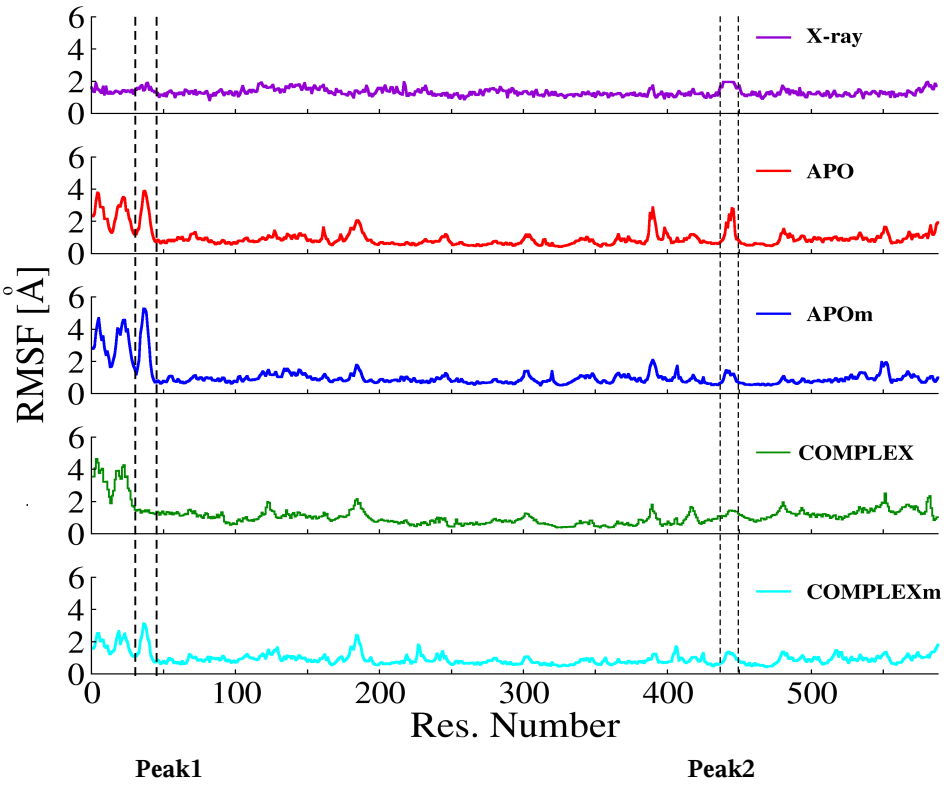


Figure7: Root mean square fluctuation (RMSF) of BVDV RdRp, extracted from B-factors (X-ray) or calculated from MD simulations (APO, APOm, COM and COMm). The region defined by residues 1 to 42 (numbered 92 to 134 in the X-ray structure) shows high flexibility in all MD simulations . In particular Peak1, which corresponds to linker L4 loop (residues 32 to 42), fluctuates in all simulations but COM, where it is in tight contact with 227G, which limits its conformational flexibility. The stability of L4 in the wild complex is also due to intra protein interaction such as a strong HB with D126, having occupancy higher than 95% (see Results section) (see 1st paper). This interaction is not observed in the case of COMm. Peak 2 corresponds to the L3 loop, which is particularly flexible in APO. The reduced fluctuation in case of APOm is due to the interaction with the linker, which deviate from it initial position and comes closer to L2 (see Fig. XX).

It is obvious from the plot of RMSF figure 7 that the loop did not under go the same conformational changes in the two simulations. The I261M complex shows larger fluctuations for the residues 32 to 42 than in APO and COM model indicating that the mutation affected the dynamics of this region.

This results may explain a portion of the mechanism of inhibition by 227G, i.e by decreasing the mobility of the allosteric loop. Of course, 227G bound to the COM physically blocks substrate access to the active site. It also suggests that the mechanism by which the mutation leads to inhibitor resistance is by restoring the loop to wild type flexibility , apparently necessary for catalytic activity. Furthermore, the COMm shows a slightly greater structural fluctuation of residues 32 to 42 compared to the COM.

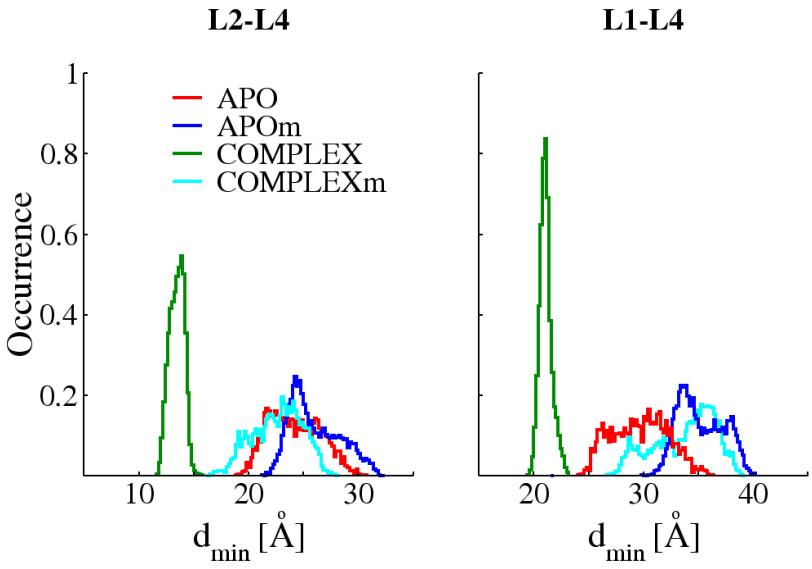
4)-Dynamic behavior of the binding site Site :

As mentioned in Ref. Paper 1, the active site of 227G lye between 4 loops, L1 (Pro388-Ile398) and L2 (Ala221-Asn229) originate from the finger domain, loop L3 (Leu530-Gly537) belongs to the thumb domain, while loop L4 (Leu125-Arg132) is the part of the N-terminal domain. Further, the cavity is encircled by the motif-I(260 to 266) of BVDV RdRp, which is located in the finger-tip region. Resistant mutation I261M lye in the motif I and hence this region needs to be carefully explored. A comparative study, closely monitoring the dynamics, of the four systems, viz APO, APOm, COM, COMm specifically focusing on the binding site region, were performed. This was done in order to have an in depth understanding of the influence of mutation on the local conformational changes suggesting the mechanism by which the mutation renders protein resistant against inhibitor. The results are discussed in the following subsections.

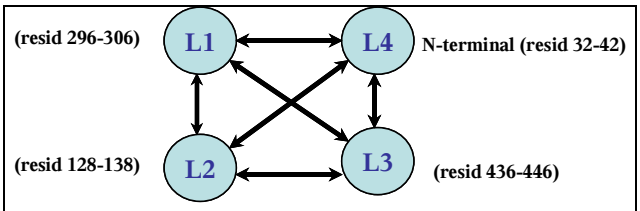
4.1)-Mobility of the loops encircling the cavity:
In order to gain deeper and more quantitative insights into the local conformational changes, we calculated the minimum distances between the four loops that define the putative binding cavity in BVDV RdRp. The distances calculated were a minimum over the all the snapshot that were sampled by the loops L1-L2, L2-L3, L2-L4, L1-L4, L3-L4 and L1-L3(figure8,B) for last 20ns of simulation of APO, APOm, COM and COMm models respectively. Figure8 A clearly shows that, in the COM system the loop pair sampled the least distances when compared to the APO, APOm and COMm system. This indicates that 227G restrains the movements of the loops by making strong interacting with the residues of the loop which was in accordance with the RMS fluctuation shown by complex simulation (figure7).

A more clear picture highlighting the differences between the four systems can be observed in figure 8(B) wherein an average of the minimum distances obtained in the last 20 ns are listed. In case of loop L1-L4, the minimum average value is sampled by complex (14.8 A) while the maximum average value is sampled by APOm(35.2A). The COMm indicated an average of (33.8A) and was trailing behind with an average value of 29A. This indicated a greater opening of loop L4 in mutated system as compare to APO and COM. Similar behavior is also observed in loop L2-L4, wherein the least average value of distance observed in COM (18.5A), while 30.5, 31.2 and 30.7 values were observed in APO, APOm and COMm system respectively. This indicated that in the mutated system loop L4 opens farther, sampling much larger distance values, and the L1-L4 and L2-L4 distances fluctuate more erratically than they do in wild proteins.

A visual depiction of the differences in the fluctuation of loops in all four systems, see Figure14, The superimposition of the averaged structures confirms a more open structure in APOm and COMm system in comparison to APO and complex systems. It is readily apparent that the simulation on the APOm and COMm system are characterized by much larger L1-L4 and L2-L4 distances values than the APO and least in complex system.



A



Loops	APO	APOm	COMPLEX	COMPLEXm
L1-L2	12.4(±1.4)	12.4(±0.9)	9.6(±1.0)	11.8(±1.0)
L1-L3	24.2(±2.3)	26.1(±2.3)	12.0(±1.2)	22.7(±2.3)
L1-L4	29.7(±2.5)	35.2(±2.0)	14.8(±1.5)	33.8(±2.8)
L2-L3	21.9(±0.9)	21.3(±1.0)	11.7(±1.4)	19.2(±1.3)
L2-L4	30.5(±0.9)	31.2(±0.8)	18.5(±1.8)	30.7(±1.2)
L3-L4	15.6(±0.3)	15.4(±0.4)	10.5(±1.5)	17.9(±0.3)

Simulation length : 30ns

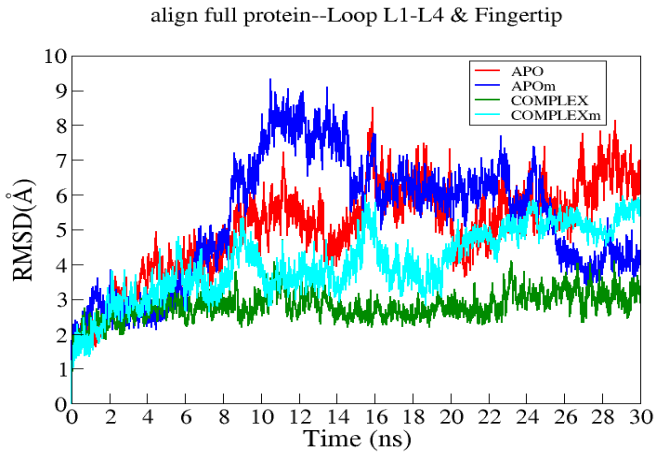
B

Figure8: Minimum distance between the loop L2-L4 and L1-L4 along the equilibrium trajectory: A)-The graph (left) indicates the minimum distance values that were sampled by loop L2-L4 throughout the trajectory, while the graph on the right the values sampled by loops L1-L4 for all the four systems. All the distribution are resemblance but that of COM, which features a sharp peak at lower values than in the other systems. This feature persists when calculating the minimum distance for the other couples of loops (see Table SXX(NOT Clear)).
Table8(B) Average distances (and standard deviations) between the loops L_4, L_2, L_3 and L_1. It can be seen that for all distances the lower value occurs for COM.

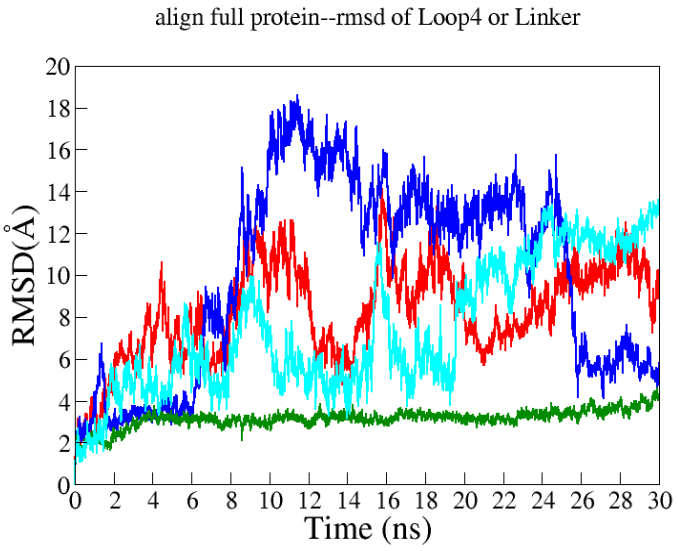
4.2)-Linker Region :

1)-RMSD timewise only linker region:

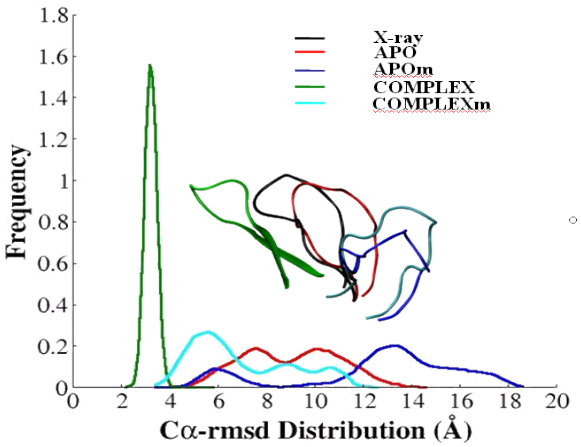
It was evident from the analysis of the RMSD-residue wise graph and the average distances between the loops, that, in particular, loop L4 showed conformational differences between the APO, APOm, COM and COMm trajectories. This region therefore needed to be carefully explored in an effort to understand the influence of its local motion on the drug binding affinity.



A



B

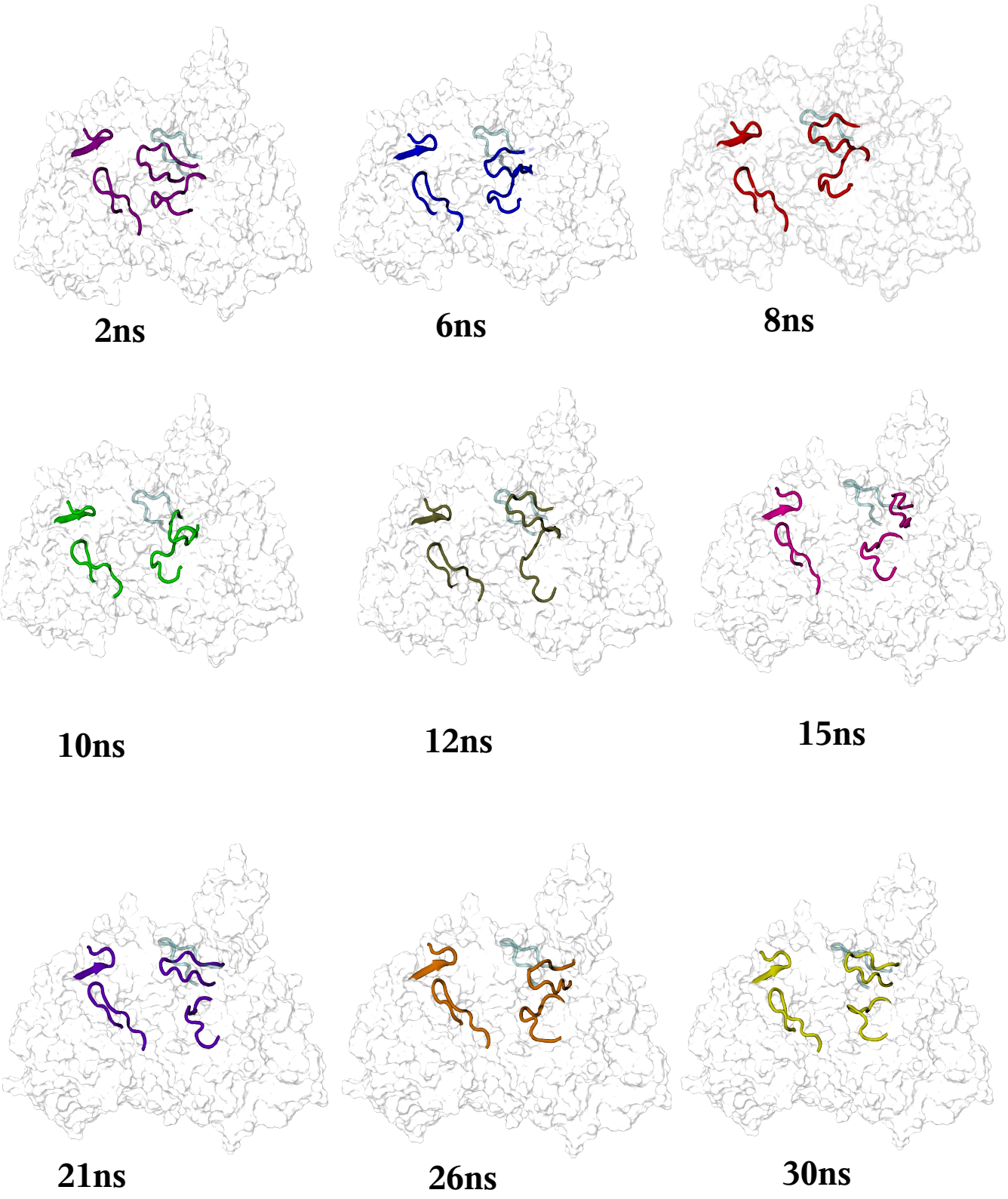


C

Figure9: RMSD

Time evolution of protein (APO and APOm) or complex (COM and COMm) Cα-RMSDs, calculated with respect to starting structures. All systems were found to reach quite stable values within ~10 ns, although ~1 A oscillations are visible even after the equilibration time. These are mostly due to oscillations in the region around drug binding site, including loops L1 to L4 and hood of the N-terminal, as can be seen comparing frames b) and c), showing respectively the RMSD of the "scaffold" (all but "flexible") and "flexible" . In particular, the linker region (d) (loop L4) features the largest flexibility. The inset shows the distribution of RMSD values calculated over the last 20 ns of MD.

A



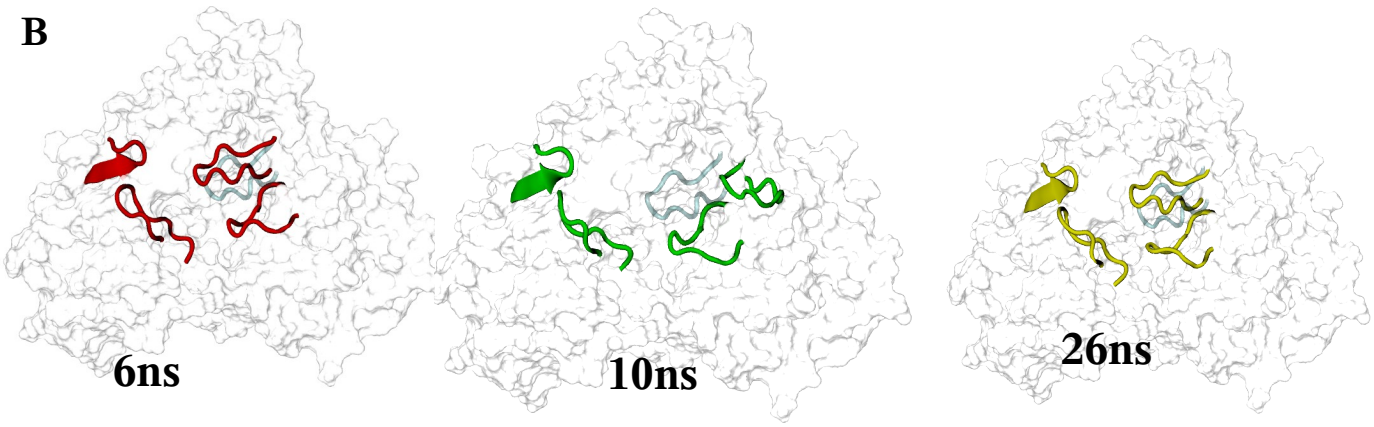


Figure10 Superimposition of the x-ray with APO (A) and APOm(B) taking snapshots at time corresponding to peaks and dips in the rmsd of linker (as denoted by arrows and dots in figure fig6c. The movement of loops which corresponds to binding pocket during the MD simulation. A)- periodic movement of loops focusing linker in APO while a transient movement observed in (B)APOm .

The figures9 and 10 clearly show the difference in the behavior of the linker region, loop L4, in APO, APOm, COM and COMm over the entire length of simulation. In case of the complex the conformation of the L4 remains stable during the course of the trajectory, showing an average RMSD below 1.7 Å while on the contrary, the rest three system viz APO, APOm and COMm showed varying level of periodicity(figure9 (B).
In case of APO, the system was found to oscillate around an average value of 8.3A. At first, around 5ns the system deviates from the initial structure to an RMSD of 10.5A, then the RMSD decreases to 5A at 8ns, preceded by another hike in RMSD to 12 A between 9 to 12 ns with an eventual dip to 5A at 13 ns. The system continues to evolve with respect to time periodically as the RMSD increases to 14A at 16ns, decreasing to 5A and eventually increasing again to 12A at 28ns. Since the loop L4, lye in the N-terminal domain hence its large mobility is consistent with the necessity of the N-terminal domain to be more flexible Unlike, in APO system, although the RMSD in case of APOm showed periodic behavior but there is a large difference in amplitude and frequency of the oscillations. As already mentioned roughly four oscillations are observed in the APO system while APOm shows just one oscillation within 30ns of simulation. In case of APOm, the system shows a huge deviation from the starting structure at 10 ns with an RMSD of 18A , which decreases to 14A between 16 to 24ns and then finally to 4A at 26ns. The RMSD shows different dynamical behavior of APO and APOm systems. While in APO system the Linker exhibits large mobility, which relates to its function of interacting with the template during polymerization.

In order to explore in depth the local conformational changes taking place in the APO and APOm systems, snapshots corresponding to each peak and dips were extracted from APO(arrow) and APOm(yellow dots) RMSD curve from figure 9(B).We obtained five dips 2, 8, 12, 21, and 30ns and four peaks at 4, 10, 15, and 26ns in case of APO, indicating a periodic nature of linker. While only two dip at 6 and 26ns and one long and extended peak obtained in APOm, clearly showing the transient nature of movement of linker. (Figure 10)

In the Inset of figure 9(C) the respective linker representative conformations as compared to that found in the X-ray structure of wild BVDV. The linker belongs to N-terminal domain; as reported previously [10,12,49] it is involved in recruiting other polymerases and in the translocation of the template during polymerization. The first thing of notice is the different behavior of APO and APOm systems: although a periodic behavior is seen in both of them, there is a large difference in amplitude and frequency of oscillations. Namely, in APO roughly four oscillations occurs between RMSD values of 6 Å and 12 Å, and the distribution is quite narrow. In APOm the periodicity is larger, with roughly one oscillation between 4.5 Å and 18 Å, as confirmed from the distribution having the largest peak at ~13 Å. This results are consistent with those in Figure8 in RMSF, indicating a larger flexibility of the linker region in APOm with respect to APO.

The linker shows large fluctuations also in COMm, where a rough bimodal distribution with centers at 5 and 9-10 Å can be envised (see inset). Moreover, in the last part of the trajectory the displacement from the initial position increases up to ~14 Å, confirming that the ligand has a small effect on this dynamics of the mutant protein. On the contrary, the distribution for COM is centered at 4 Å with a range between 2 to 4 Å, indicating that the binding of 227G to the APO protein strongly stabilizes the linker in a definite conformation (that closes the entrance of the template channel, cfc. Figure 3.

4.3)-Role of mutation and its effect:
Image loops bundle one and 4 loops together.
Previous study has reported that motif I, lying in the fingertip region, is involved in NTP binding during the polymerisation. I261 strategically located in motif I and is therefore believed to play role in interacting with the incoming nucleotide. In addition, as reported in our previous studies (REF), I261 is involved in stabilising and properly orienting 227G by making strong HpH interaction (coupled paper). I261 when mutated to M261 confers resistance to 227G. Mutation I261M, apart from probably changing the general binding mode of 227G by causing a local spatial bump or property mismatch, because of its chiral side chain, isolecine being more hydrophobic than methionine, with the original binding mode of the ligand was also observed to cause huge conformational changes in the loops as discussed earlier. Interestingly, I261, part of motif-I and the binding site, is located relatively close to motif-II. Thus the combination of the closely located functional groups could have possible structural and/or mechanistic consequences.

The N-Terminal domain along with a long insert in the finger domain constitutes the Fingertip region. Fingertip is important not only for the translocation of template during the polymerisation but also as it contains two highly conserved motif I(NTP binding) and motif II(Template Binding)[Choi 2004, 2006]. Previous studies by choi and nytes, have reported resistant mutation for different classes of NNI in (E291G) and around (F224S/Y) in Motif II. The resistant Mutation I261M against 227G is found to lie in the Motif I of the fingertip region. It was previously hypothesized that the linker region of fingertip region involved in dimerazition and provide the site for binding of other proteins required for polymerization.

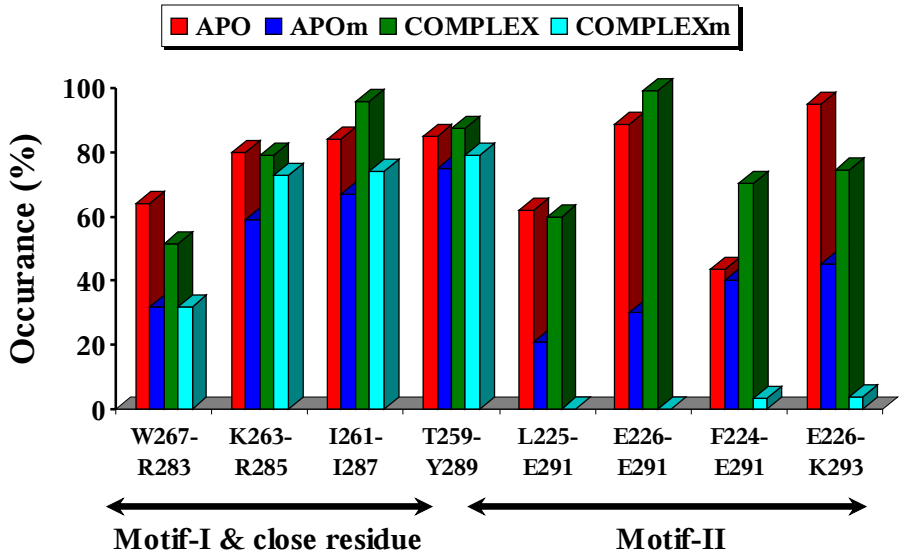


Figure13 Intra-protein H-bonds network formed between residues of the fingertip and surroundings. A reduced pattern of interactions is clearly visible due to the mutation, present both in the apo and holo proteins and particularly pronounced in motif II. In addition, binding of 227G does not influence significantly the interaction network in the wild protein, while a dramatic effect is observed again in motif II. This reflects the different binding mode of 227G in the resistant mutant , and provides further information about the possible mechanism of resistance.

The N-Terminal domain along with a long insert in the finger domain constitutes the fingertip region. Fingertip is important not only for the translocation of the template but also as it contains two highly conserved motifs motif I, which , and motif II, [10,12,49]. Previous studies by Choi and Nytes have reported resistant mutations for different classes of NNIs to lie in (E291G) and close to (F224S/Y) Motif II. The resistant mutation I261M against 227G is found to lie in the Motif I of the fingertip region. It was previously hypothesized that the linker of fingertip region is involved in protein dimerization and also provides a binding spot to other proteins during the polymerization process [49,50]. All these mentioned facts highlights the importance of fingertip and also a need to study in detail the effect of mutation in this region.

All above mentioned facts indicates, fingertip with its myriad of functions needs to be exploited in detail.. the effect of mutation which may also help us to throw some light on the underneath lying mechanism of resistance.

The differential behavior of the loop and the ligand in the two systems needs to be explained in the context of the mutation of residue I261M. An in depth analysis of the intra hydrogen bond network formed within the Fingertip residues and those surrounding them was performed for all the four systems APO(red), APOm (blue), COM(green) and COMm (cyan) respectively(Figure13). It displays a reduction of the intra hydrogen bonding in the mutated protein as compared to the wild protein in the Motif I region. While the COM showed a further stability by increased probability of Hbond (I261-I287), and other residues W267-R283, K263-R285 and T259-Y289. The COMm system showed reduce probability of Hbond formation when compared to the APO and COM. In case of Motif-II the reduction of hyderogen bond is more visible, as in COMm it all most lost the interaction with L225-Y289, L225-E291, E226-E291, F224-E291 and E226-K293. the change in hydrogen bonding competence was noticeable in mutated systems only while in APO and COM system trajectories all of these interactions were conserved.

The difference in the Hbonding propensity between four systems was more pronounced in the residues of Motif II (Figure13). There was a complete disruption of the Hbond in the COMm system, while there was a reduction in the Mutated system when compared to the APO and COM system. This indicated an underlying local structural change occurring in the fingertip region in the presence of mutation.

Role of Ile261:

The mutation Isoleucine to Methionine at position 261 was found to confer resistant to 227G. I261M is also reported mutation of other compounds [47,48](Gabriele et al; 2010, tonelli et al 2010), so this cause a very significant changes and could be a hot-spot. It is well known fact that the probability of occurrence of Ile as a part of the binding cavity is greater than 50% (REF NCBI). The effect of the point mutation depends on both the nature and position of mutation. Generally, non-conservative mutation on buried residue can have profound effect. Ile261 located in the motif I of the finger tip region, non-conservatiove [49] and buried into the putative binding pocket, found to play a very important role in the stability of 227G. Therefore, we focused some attention on this residue. It belongs to motif I, whose function is to bind with incoming NTP in the process of polymerization. This residue might be involved in the interaction with template during polymerization as its location is close to the template entrance site.

The mutation from I261 to M261 is a change of small(S) amino acid to large (L) amino acid, as methionine has longer side chain. It is well know fact that because of its longer side chain Met has more flexibility than isoleucine.(NCBI). The S-L change mutation might probably change the general binding mode of inhibitors because these types of mutations often cause a local spatial bump or property mismatch with the original binding mode of the ligand.[105](Cunningha, B. C.; Science1989 and Ge-Fei Hao et al. J. Phys Chem B 2010). This substitution is believed to cause a reduction of high affinity binding of 227G into the binding pocket of BVDV RdRp. In complex study 227G has maximum HpH contacts with I261, this residue is one of the most key determinant in the stability of 227G.

An important difference in the conformation and dynamics of Ile261 has been observed in the trajectory analysis of the mutant. The M261 has longer side chain which doesn't let 227G bind into the cavity. As we observed after 4ns of simulation 227G moved from its original position and stabilize itself in another binding mode.

Electrostatic potential map of binding pocket:

To investigate electrostatic effect of the I261M mutation, the electrostatic potential grids were computed by numerical solution of the linearized Poisson-Boltzmann equations (APBS) and mapped onto the surface of the binding pockets of APO and AOm[Fig. 11].

As shown, I261M mutation was found to perturb the electrostatic potential inside the binding pocket. The electrostatic potentials of the bound 227G, are fairly complementary with the binding pocket in COM and COMm

A)-COM Front view Back view

B)-COMm

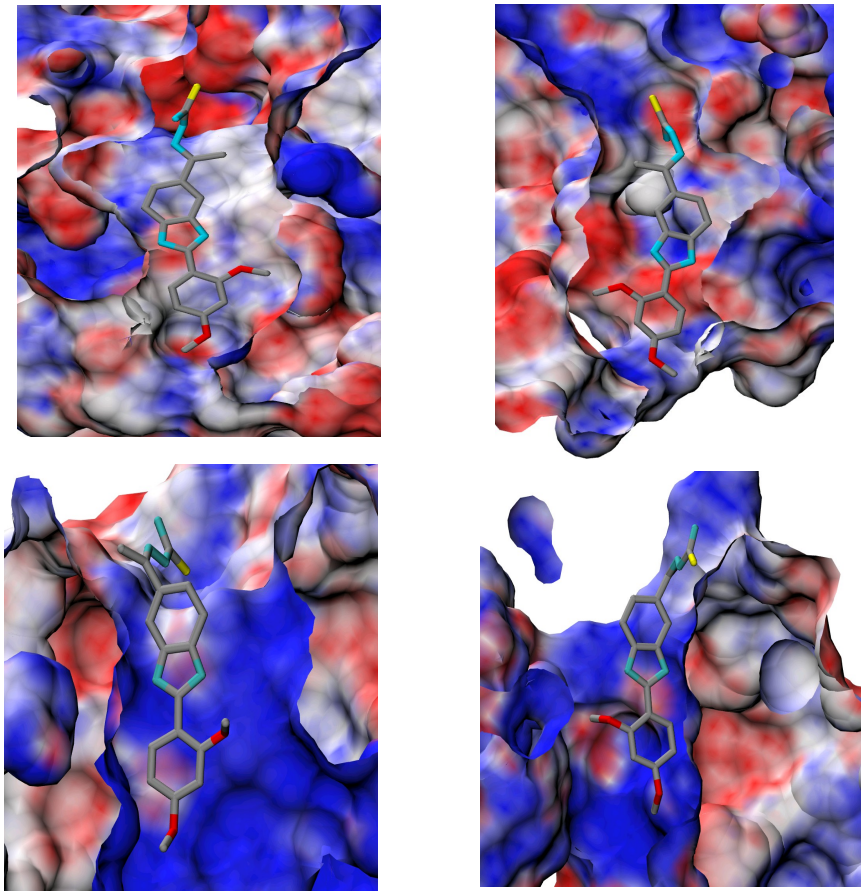


Figure11: Electrostatic potential maps of the 227G binding pockets of RdRp: (A) COM(I261); (B) COMm (M261) surface (positive in blue, negative in red, and neutral in white). 227G is shown in stick and colored according to the atomic-coloring scheme (C in silver , O in red, and N in cyan and Sulfur in yellow).

Inhibitor binding mode:

The interaction between RdRp and 227G in the two complexes as well as the influence of the inhibitor on the structural and dynamical properties of the binding site region have been clarified by analyzing the trajectory data obtained from the MD simulations.

It is evident from the RMSD of all non hydrogenous atoms of 227G that the inhibitor behaves quite differently in the COM versus how it behaves in the resistant mutant, and the binding mode is more stable than in the mutant model. In fact, in the I261M simulation the inhibitor didn't remains in the same conformation during the entire simulation. Initially the binding pose of COM and COMm is similar but after ~3.5ns of simulation it moves from its original position and relocate itself into a new position of the same binding site[figure movement]. In fact, in the COM the binding site remains stable with RMSD 2.5 A through out the entire simulation whereas in COMm its position did go under substantial changes, upto a maximum RMSD of 9 A at 10ns , stabilizes in the latter part of the trajectory with RMSD 3.8A. Other analysis of the trajectory show that the inhibitor does not move in a simple way in COMm as in the mutated structure the 227G bind on the same place, it is stable in COM interacting with linker and surrounding residues mainly with motif I and loop residues but due to mutation there is a conformational change and by which the linker is moved away from the binding pocket, because of this the 227G didn't find its interaction with linker. In the absence of proper interaction it moved from the initial binding mode and stabilizes itself in another area of the same binding site.

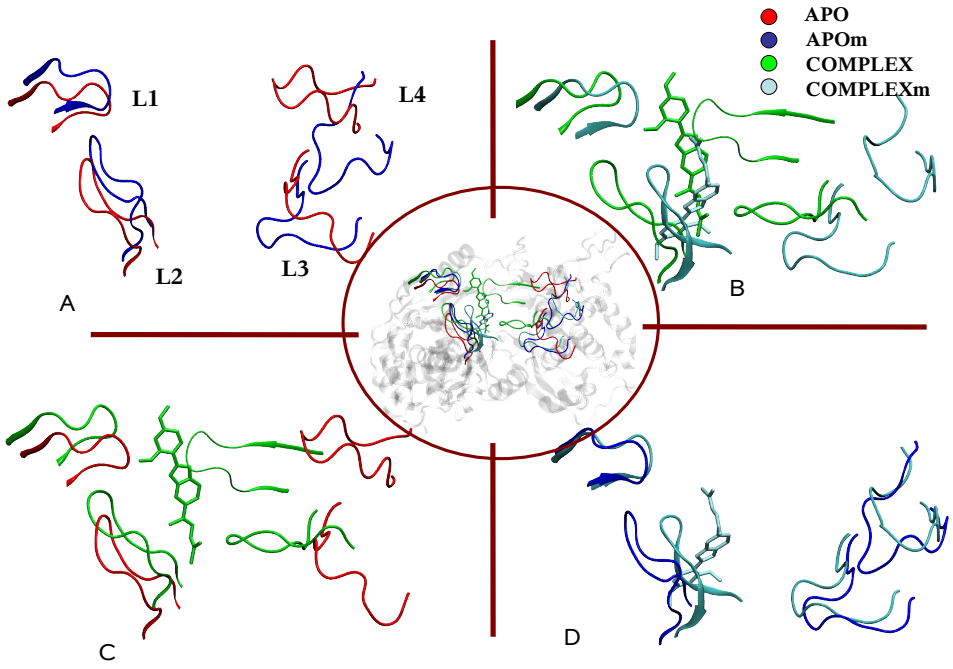


Figure14 a) Superimposition of the four loops L1, L2, L3 and L4 from the four systems APO, APOm, COM and COMm: red, blue, green and cyan color indicate APO, APOm, COM and COMm system respectively. Shown structures are the closest ones to the average of any simulation, extracted from the last 20ns of MD. In the lower pictures the same superimposition is zoomed for couples of structures, in order to highlight the conformational changes due to the I261M mutation, and its influence on the binding properties of 227G.
a) APO vs APOm. A considerable movement of L3 and linker (L4) loops is seen in APOm as compared to APO, indicating an opening of the binding cavity. This opening in the mutated system has been observed in two independent simulations started from the two highest-score poses (Clus1 and Clus2 in Tab. 3). This is true also for APO, where the closure of the template entrance gate has been verified by two independent simulations started from clus1 and clus2 ;
b) COM vs COMm. 227G binds to different places in the two systems (see also Fig. XX where the ligand in COMm travels around). As a result, the ligand is able to draw all the loops towards itself in COM, closing in this way the entrance gate; this is more evident in figure (d), where the comparison between APO and COM is reported. Movements of loops L3 and L4 toward the ligand are clearly recognizable. This is not true in COMm: here the loops L4, L3 and L2 are not significantly affected by binding of the ligand, as more evident in figure (c).

Wild-type complex

227G accommodate itself in a cavity encircled by four loops (L1, L2, L3 and L4). Loop L1 (Pro388-Ile398) and L2 (Ala221-Asn229) originate from the finger domain, loop L3 (Leu530-Gly537) belongs to the thumb domain, while loop L4 (Leu125-Arg132) is the part of the N-terminal domain. Further, the indole moiety of 227G stacks against the motif-I (Ile261-Lys266), which is located on the flexible finger-tip region Figure 15. The residues lining the compound include mostly hydrophobic (Ala221, Ala222, Phe224, Ile261, Pro262, Ile287, Ala392) and polar (Thr160, Thr162 Asn217, Asn264 and Ser533) amino acids, although also three basic (Arg127, Arg130, Arg132) and two acidic (Asp126, Glu128) are present. This is consistent with the chemical properties of the compound, whose main body is made of aromatic rings and tail is polar in nature. Consistently, the major stabilizing factor of 227G is provided by the strong hydrophobic interaction with the receptor (-47.6Kcal/mol), with main contributions from residues Arg127, Glu128, Ile261 and Pro262. This feature is clearly identifiable in where we collect the interaction energies and the (dynamic) interaction pattern related to HpH contacts, respectively(See paper1).

In addition to hydrophobic interactions, a network of inter and intra-molecular HBs stabilizes 227G in the pocket (Suppl Fig.1). In particular, a strong HB is established between N2 of 227G's indole moiety and the main chain of Asp126 (Average Dynamic length (ADL) = 2.3 Å). This interaction is present throughout all the simulation, has occupancy larger than 80%, and involves alternatively atoms OE1@Asp126 and OE2@ASP126 of the L4 loop. In addition to intermolecular bonds, a persistent intra-molecular HB exists during the entire simulation between side chain of Asp126 and HD@ASN264 of Motif-I. This interaction plays a significant role in stabilizing the former mentioned HB, which effectively locks the protein-ligand complex thereby securing the ligand stable inside the binding site.

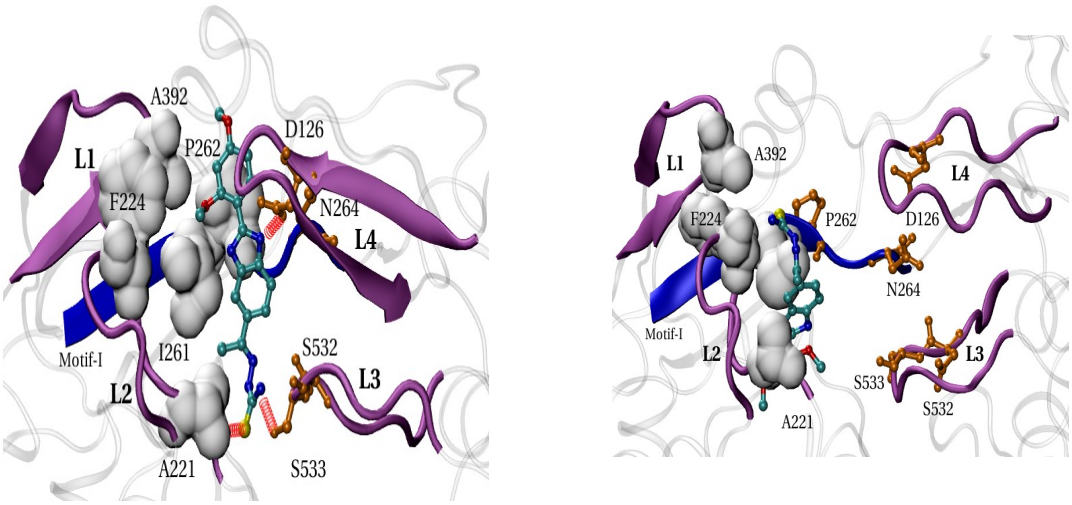


Figure15: Interaction map of COM-227G and COMm-227G with key residues of four loops represented L1, L2, L3 and L4 in magenta color and motif-I is blue in color. 1A)- The key residues which interact in COM-227G are depicted here. Three HB with D126, A221 and S533 are displayed as red color spirals. HpH contribution to the stability of 227G is provided by residues A221, F224, I261 P262 and A392 are in surface view. 1B)- represents the rearrangement of 227G in the presence of mutation. The HB made in wild with loop L4 and L3 residues are lost in COMm. 227G in mutant is totally different as in wild type it is close to loop L4 and L3 while in mutant it's close to loop L1 and L2 and in opposite orientation. This orientation is mainly stabilized by HpH contacts with the same residues as in wild except residues P262.

Another source of stabilization of the complex comes from water-mediated interactions between Glu128 and N1@227G: although the mediating water molecule changes during the simulation (see 1st paper), the interact persists for the entire simulation time. This water mediated HB is stabilized by another set of intra-molecular HBs between residues Asp126 and Glu128 of the protein. The interaction between these two residues, weak at the onset of the simulation, is strengthened once the water-mediated interaction comes into picture. Another HB is formed between H18@227G and OG@Ser533 with ADL=2.5 Å and occupancy of 46%, which creates a linkage with loop L3 (Table3 & Figure10-suppli in 1st paper). We also found a transient HB formed between Ala221 and S1@227G (~occupancy 12%), further stabilizes 227G in the binding pocket.

Interestingly, such intra-molecular bonds (either direct or water-mediated) were not observed in the simulation of the apo of the protein; thus, they are due to the presence of 227G and build a strong network that stabilizes ligand in the cavity.

4.4.2)-Mutant complex

In COMm, the drug stabilizes itself in totally different place from its original docking pose. The initial adduct in mutant was choosen in the same orientation as in case of COM. in this model 227G occupies a cleft that is formed with the same residues as in case of COM system. Here this pose was stable upto ~3.5ns making a HB with O@Pro262 with ADL 2.8 A and angle 154. apart from this HBond it also makes some HpH contacts with Met261 and Pro262. at this ~3.5ns of simulation the system was stable but the RMSD was not consistent as it was deviating with a variation from 2.6 to 3.8 A. after ~3.5ns of simulation 227G looses its interaction with Pro262 and it was rotated into the binding pocket and relocates itself in a new binding pose, which is close to loop L2 . the inhibitor binding mode is therefore preserved along the rest of the dynamics simulation. in this orientation drug was totally inverted in shape. The head of the 227G is completely buried while tail portion is solvent exposed. At this orientation drug didn't have any prominent HB but it has some strong HpH contacts with Gly132, Tyr198, Ile196, Tyr583 and Met170. Furthermore, it is important to note that Met261 drug is in close contact with 227G. but the new binding pose in COMm , have different set of binding residues than COM.

Hypothetical mechanism for Drug resistance:

All of these results and observation led us to hypothesize a mechanism for the inhibition (see coupled paper) and drug resistance of 227G. the docking pose in COM was stable and conserved through out the entire simulation. The most strong contribution in the stability of 227G is provided by the linker region, residue Asp126 and Glu128 along with water mediated interaction with Asn264. These interaction from linker makes hydrogen bonds with 227G, secures it stable into the binding site. The other interacting component which almost encircling the binding cavity is motif I, which contain resistant mutation I261. this residue contribute maximum HpH contacts with 227G along with Ala392, Pro262 and Ala221. These strong HB, HpH interaction locks 227G into the binding pocket. As we reported earlier the mechanism of inhibition of 227G mostly depend it location as it binds into the passage of template entrance site, and it blocks completely this passage. Some other reasons are the binding of 227g with motif-I, which is well known location to bind with incoming NTPs. Apart from these two major concern it could be very much probability the binding of 227G hinder the dimerazition, as it, itself bind on the surface and secondary it attracts the linker region towards cavity. Previously it was hypothesized that the linker involved in the recruiting other protein, essential for the polymerization [50](choi et al 2006). Thus the displacement of linker could hinder the dimerazition.

In place of Ile261 when Met came into the picture, scenario effected heavily, in case of APOm and COMm, we observed there is a large displacement of linker region from its original position. In the absence of linker at binding site, 227G was not able to stabilize itself at docking pose (initial orientation), thought it has HB with

Pro171 of Motif-I apart from some HpH contacts. After ~3.5ns it moved from the initial position. Another reason of the movement of the 227G from its original, could be the large side chain of Met at 261 position. The large side chain of Met261 in COMm could be a reason of dissociation of 227G from its original position as mostly its pointing towards binding site. This movement of 227G led it to bind into the different reason of same binding site.

When the Ile261 to Met261 residues are mutated in RdRp-227G complexes introduce a functional confirmation in the loops, fingertip and N-terminal region. Although there is not as such remarkable changes observed in the catalytic site. Above mentioned seems to be an acceptable argument involved in drug resistance mechanism.

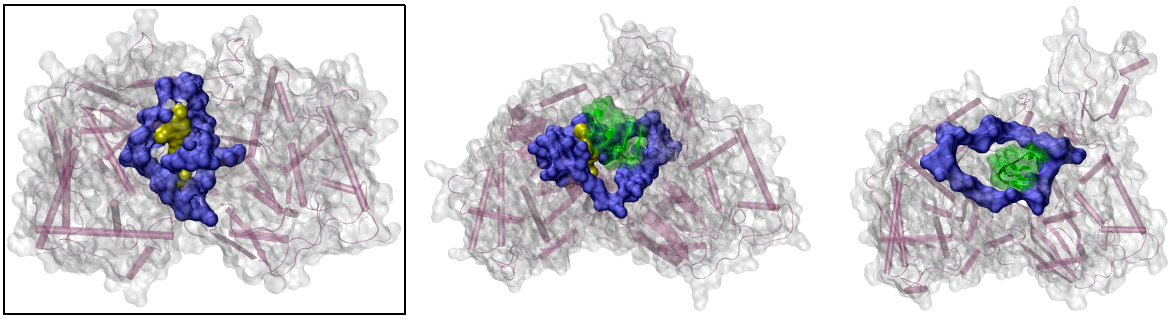


Figure17: Top view of COM (a) and COMm (b) representative conformations compared with the X-ray structure of BVDV RdRp (c) . For the sake of clarity the views are slightly rotated from the the axis entering the RNA template gate, and different in each picture. Structures in (a) and (b) are extracted from the last 20 ns of MD simulation (i.e. the equilibrium dynamics) and have the lowest RMSD with respect to the average. The protein is represented as transparent magenta cartoon and grey molecular surface, with residues lining the entrance of the RNA template highlighted in opaque blue. 227G in shown in yellow surface.
In order to better explain our proposal on a possible mechanism of resistance, we create a 5-mer ssRNA template model (shown in green) and manually docked it onto the entrance gate. It can be seen that a complete closure of entrance channel occurs in COM, where loop L4 “embraces” 227G; therefore the RNA template would be sterically hindered to enter into the polymerase. On contrary, the entrance remains partially open in COMm, allowing for a easy fit of the template RNA's head. This situation well compares with that shown in (c) for the wild protein, where the RNA template model well fits within the entrance gate. In particular, comparing figures (b) and (c) one can see that 227G sits on the left-side subpocket in the presence of I261M mutation, leaving the larger right-side pocket open.
These results are consistent with Fig. 4. Moreover, while this picture refers to the most representative configuration from the MD, in Fig. 20 we use a different approach to show that this results are valid in general for the most representative conformations of the complexes along dynamics.

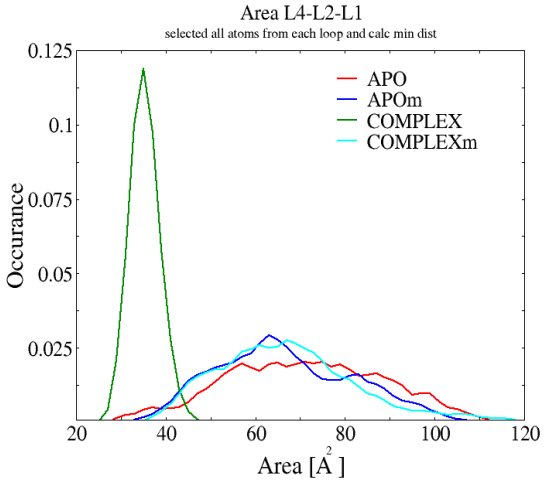


Figure18. Area of the region enclosed by loops L1, L2 and L4 lining the entrance gate:

Frequency plot of values of the area enclosed by loops L1, L2 and L4 lining the entrance gate (see Methods for further details). We observed that there is a complete closure of the template entrance channel due to presence of 227G only in COM, where the area has a value of 36±9 A; this fact is in agreement with Fig 3, where is shown that 227G completely blocks the entrance of the template channel.
The distribution is significantly spread in COMm as is evident in Fig 8. wherein despite of binding of 227G, the template channel is partially open. Indeed this latter is very similar to the distribution found in APOm, pointing to the fact that the ligand is not able to close the entrance for the template in this case (ADD avg + std in COMm, APOm and APO respectively).

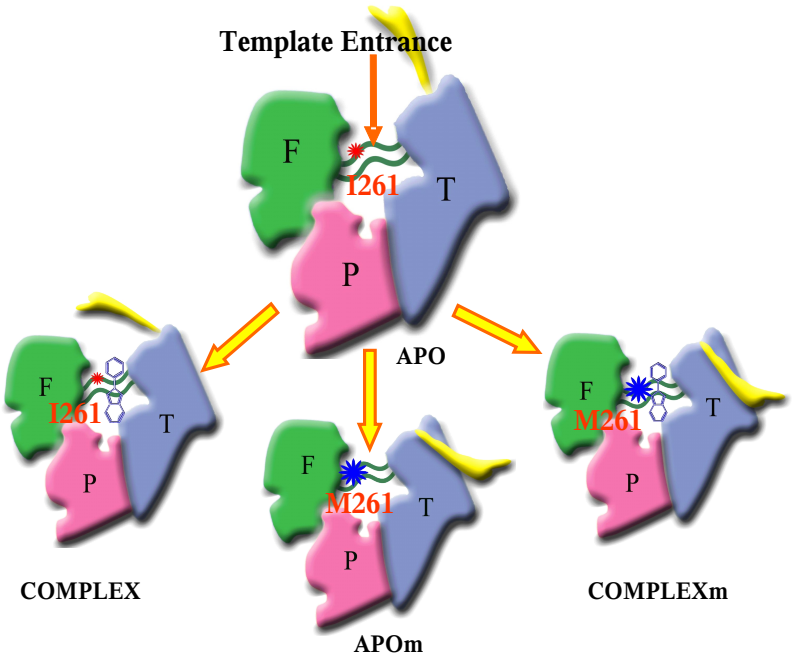


Figure19: Model illustrating the putative mechanism of BVDV RdRp resistance against 227G. A cartoon representation of BVDV RdRp is shown with the thumb (blue), finger (green) and palm (pink) domain. Residue I261 is represented by red star, Linker region (L4) represented by yellow (the entire N-terminal domain is not shown for the clarity) while the mutated residue M261 is represented by blue star.
While comparing the four structures, major differences can be pointed out in the positioning of the linker region. In case of COM, binding of 227G in the putative binding pocket brings the linker closer when compared to the crystal structure. On contrary, during the APOm simulation the linker appears to move away from its original position. Due to the movement of the linker 227G loses some of its major interactions with the linker and this leads to the unstability of the 227G in the APOm system.

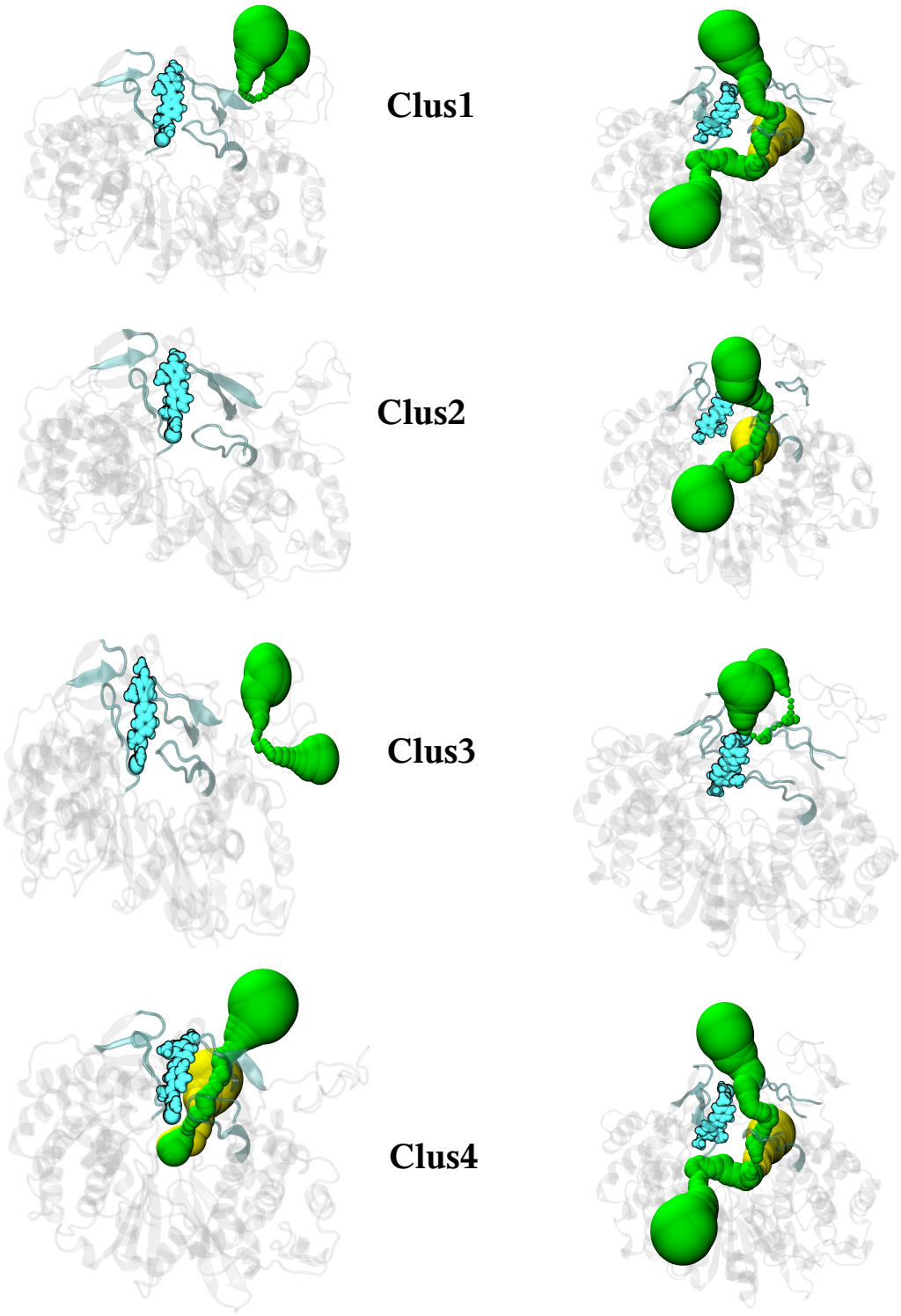


Figure20: chunnel analysis- We find only the three known tunnels with chunnel analysis. Left pannel correspond to clusters obtained in COM while right panel correspond to clusters of COMm. In both cases 227G is in Cyan in color and rendered in VdW. The green and yellow row of beads represent the probable tunnels exist normally in RdRp. Beads (green) coming from top represent template entrance channel, yellow beads correspond to NTP entrance tunnel and another green beads towards z axis representing the exit tunnel. The whole protein rendered in cartoon (ghost) while loops L1, L2, L3 and L4 in cyan (transparent).

Conclusion:

We have reported results herein of four systems more than ~120ns molecular dynamics (MD) simulations of the APO, APOm, COM and COMm. Detailed analysis of structural perturbation in the APO-APOm and COM-COMm systems indicate that the systems display a distinct dynamical behavior. Major and significant structural differences were found at the loops involved making putative binding site, are fingertip region and the hood of Nterminal. These changes are notable as they are close to template entrance site and catalytic site respectively. I261 is the cross resistant, as apart from 227G, it is also a resistant mutation of other compounds, previously reported by our group. The location of this resistance mutant is very strategic, as it is in the passage of template entrance channel and lies in motif I of fingertip. Considering these observation, it is reasonable to investigate the effect of this mutation on inhibitor binding, as well as to understand the drug resistance more in depth.

Infact the binding of 227G in COM caused several conformational changes especially in the linker region. which achieved a rigid configuration after binding of the 227G in contrast to its flexible behavior. Simultaneously other loops apart from linker comes closer towards binding site making interaction with 227G and inducing a contraction into the entrance of the template site, led a complete closer of this tunnel.

Conversely, the MD simulation of the resistant mutation in APOm and COMm displayed a different conformational transition in the linker and a high degree of mobility in this region. Comparatively in APO, we observed a periodic behavior of the linker with high oscillation and less displacement while in APOm and COMm a transit movement observed with low oscillation and large displacement.

Here, we reported that I261M is the resistant mutation of 227G. it lost its complete potency in case of M261 as regained itself a basic setup to perform polymerization. From experimental it was validated that polymerization takes place in M261 strain. we observed there is enough space for the entrance of template in the COMm model though 227G was binding into the binding pocket. While in COM, 227G blocks completely the passage of entrance of template . Figure20 explains completely the enormous reduction of area in COM while other three models behave similarly with slight changes. To validate our area analysis we performed chunnel analysis [106] on COM and COMm, to identify the functionally important tunnels reported experimentaly such as template entrance channel, NTP entrance channel and the exit channel from the C-terminal.(REF). Interestingly, we got a very nice match with our experimental results. Furthermore, the chunnel analysis in all four obtained clusters of COM and COMm. In COM, we obtained a complete blockage of all tunnels required for biological functioning except clus4 as indicated in figure[chunnel wali], While in COMm all above mentioned tunnels are completely open except clus3 which is also partially closed. So the availability of these tunnels in COMm, provides a possibility for RdRp to retain its polymerization by allowing the possible entry of template into the tunnels, even in the presence of 227G. From above explanation it seems that the 227G is completely ineffective in COMm which is a good agreement with the experimental results .

Mutation I261M as indicated by the mutagenic studies reported in this paper lye in the motif-I (finger-tip region). Our computational results confirmed the mutagenesis experiments that have pinpointed mutations in the finger domain of the BVDV RdRp as responsible for an increased resistance against NNI. Thus, the indirect evidences of the location of the binding site in the fingers domain of BVDV RdRp extracted from the genotypic characterization of our resistant mutants received a strong microscopically well-funded support from the computer simulations. Analysing our MD trajectories, we found that the linker contributes to stabilize 227G in the putative binding cavity. In the presence of the 227G, our MD results indicated that these loops of the binding cavity are even closer than other system, thereby resulting into a further closing of the cavity. As mentioned in the result, in order to quantify this hypothesis we did area analysis of the mouth of the cavity in all three cases (COM, COMm and X-ray), we observed there is wide reduction in area of COM. As a consequence, in the mutant we observed, there is an opening of putative binding site as loops, mainly linker moves from the binding site and doesn't let 227G to stabilize into the binding pocket, causing drug resistance. Another probable possibility of drug resistance could be the longer side chain of M261, which cause hindrance to 227G, as in COM it gain maximum HpH interaction with I261.

References:

1. Erica Weiskircher et al , 2009, virology journal, Brock KV: The persistence of bovine viral diarrhea virus. *Biologicals* 2003, 31:133-135
1. E.J Richey.(2009).http://www.cattlenetwork.com/Cattle-Health--Treatment--Prevention---Control-Of-BVD/2009-06-01/Article.aspx?oid=496856&fid=VN-ANIMAL_HEALTH-BVD-ARTICLES
2. Daniel L. Grooms1, et al *Bov Pract* 43:106-116, 2009, peer reviewed
3. Bartlett, B, Grooms, D. BVD-PI Eradication: Unintended Consequences. Michigan Dairy Review. 2008;13(3) July 2008.)
4. Fraser, Clarence, Ed. The Merck Veterinary Manual. 7th Ed. 1991. Pp. 166-168.;
5. Radostits, Otto and Clive Gay, et al. Veterinary Medicine. Saunders/Elsevier. 10th Ed. 2007. Pp 1248-1275.;
6. Ridpath, Julia and Robert Fulton. “Knowledge gaps impacting the development of Bovine Viral Diarrhea Virus control programs in the United States.” JAVMA, Vol 235, No. 10. November 15, 2009. Pp. 1171-1179.)
7. Baginski, S.G., Pevear, D.C., Seipel, M., Sun, S.C., Benetatos, C.A., Chunduru, S.K., Rice,C.M., Collett, M.S., 2000. Mechanism of action of a pestivirus antiviral compound. Proc. Natl. Acad. Sci. U.S.A. 97, 7981–7986.
8. Yanagida, K., Baba, C., Baba, M., 2004. Inhibition of bovine viral diarrhea virus (BVDV) by mizoribine: synergistic effect of combination with interferon- . Antiviral Res. 64, 195–201.
9. Paeshuyse, J., Leyssen, P., Mabery, E., Boddeker, N., Vrancken, R., Froeyen, M., Ansari,I.H., Dutartre, H., Rozenski, J., Gil, L.H., Letellier, C., Lanford, R., Canard, B.,Koenen, F., Kerkhofs, P., Donis, R.O., Herdewijn, P., Watson, J., De Clercq, E.,Puerstinger, G., Neyts, J., 2006. A novel, highly selective inhibitor of pestivirus replication that targets the viral RNA-dependent RNA polymerase. J. Virol. 80,149–160.
10. Tabarrini, O., Manfroni, G., Fravolini, A., Cecchetti, V., Sabatini, S., De Clercq, E., Rozenski, J., Canard, B., Dutartre, H., Paeshuyse, J., Neyts, J., 2006. Synthesis and anti-BVDV activity of acridones as new potential antiviral agents. J. Med. Chem. 49, 2621–2627.
11. Paeshuyse, J., Chezal, J.M., Froeyen, M., Leyssen, P., Dutartre, H., Vrancken, R., Canard, B., Letellier, C., Li, T., Mittendorfer, H., Koenen, F., Kerkhofs, P., De Clercq, E., Herdewijn, P., Puerstinger, G., Gueiffier, A., Chavignon, O., Teulade, J.C., Neyts, J., 2007. The imidazopyrrolopyridine analogue AG110 is a novel, highly selective inhibitor of pestiviruses that targets the viral RNA-dependent RNA polymerase at a hot spot for inhibition of viral replication. J. Virol. 81, 11046–11053.
12. Okamoto, M., Sakai, M., Goto, Y., Salim, M.T.A., Baba, C., Goto, K., Watashi, K., Shimotohno, K., Baba, M., 2009. Anti-bovine viral diarrhea virus and hepatitis C virus activity of the cyclooxygenase inhibitor SC-560. Antiviral Chem. Chemother. 20, 47–54.
13. Chang, J., Wang, L., Ma, D., Qu, X., Guo, H., Xu, X., Mason, P.M., Bourne, N., Moriarty, R., Gu, B., Guo, J.T., Block, T.M., 2009. Novel imino sugar derivatives demonstrate potent antiviral activity against flaviviruses. Antimicrob. Agents Chemother. 53,1501–1508.
14. Paeshuyse, J., Letellier, C., Froeyen, M., Dutartre, H., Vrancken, R., Canard, B., De Clercq, E., Gueiffier, A., Teulade, J.C., Herdewijn, P., Puerstinger, G., Koenen, F.,Kerkhofs, P., Baraldi, P.G., Neyts, J., 2009. A pyrazolotriazolopyrimidinamine inhibitor of bovine viral diarrhea virus replication that targets the viral RNA dependent RNA polymerase. Antiviral Res. 82, 141–147.
15. Sako, K., Aoyama, H., Sato, S., Hashimoto, Y., Baba, M., 2008. -Carboline derivatives with anti-bovine viral diarrhea virus (BVDV) activity. Bioorg. Med. Chem. 16, 3780–3790.
16. Aoyama, H., Sako, K., Sato, S., Nakamura, M., Miyachi, H., Goto, Y., Olamoto, M., Baba, M., Hashimoto, Y., 2009. Polymethylated -carbolines with potent anti-bovine viral diarrhea virus (BVDV) activity. Heterocycles 77, 779–785.
17. Luscombe, C.A., Huang, Z., Murray, M.G., Miller, M., Wilkinson, J., Ewart, G.D., 2010. A novel Hepatitis C virus p7 ion channel inhibitor BIT225, inhibits bovine viral diarrhea virus in vitro and shows synergism with recombinant interferon- -2b and nucleoside analogues. Antiviral Res. 86, 144–153.
18. Mohammed TA Salim, Mika Okamoto, Shinnosuke Hosoda, Hiroshi Aoyama, Yuichi Hashimoto, Masanori Baba; Anti-bovine viral diarrhoea virus activity of novel diphenylmethane derivatives;Antiviral Chemistry & Chemotherapy, 2010;20:193-200
19. Mohammed T.A. Salim, Yukinori Goto, Takayuki Hamasaki, Mika Okamoto, Hiroshi Aoyama, Yuichi Hashimoto, Simone Musiu, Jan Paeshuyse , Johan Neyts, Matheus Froeyen, Piet Herdewijn, Masanori

Baba; Highly potent and selective inhibition of bovine viral diarrhea virus replication by -γ carboline derivatives; Antiviral Research 88 (2010) 263–268.
20. Sun, J.H., Lemm, J.A., O'Boyle, D.R.2., Racela, J., Colonno, R., Gao, M., 2003. Specific inhibition of bovine viral diarrhea virus replicase. J. Virol. 77, 6753–6760.
21. 24. King, R. W., H. T. Scarnati, E. S. Priestley, I. De Lucca, A. Bansal, and J. K. Williams. 2002. Selection of a thiazole urea-resistant variant of bovine viral diarrhoea virus that maps to the RNA-dependent RNA polymerase. Antivir. Chem. Chemother. 13:315–323.
22. Ni. Zhang, Zhengwen.Liu, Qunying.Han , Jinghong.Chen, Yi.Lv ;Xanthohumol enhances antiviral effect of interferon α-2b against bovine viral diarrhea virus, a surrogate of hepatitis C virus;Phytomedicine17(2010)310–316
23. Mohammed TA Salim, Mika Okamoto, Shinnosuke Hosoda, Hiroshi Aoyama, Yuichi Hashimoto, Masanori Baba; Anti-bovine viral diarrhoea virus activity of novel diphenylmethane derivatives;Antiviral Chemistry & Chemotherapy, 2010;20:193-200
24. Finkielsztein, L.M.;Moltrasio, G.Y.;Caputto, M.E.;Castro, E.F;Cavallaro, L.V;Moglioni, A.G., What is Known About the Antiviral Agents Active Against Bovine Viral Diarrhea Virus (BVDV)? , Current Medicinal Chemistry, Volume 17,Number 26, September 2010 , pp. 2933-2955(23)
25. Steinmann J , Becker B 1, Bischoff B 1, Paulmann D 1, Friesland M , Pietschmann T, Steinmann J, Steinmann E ; Virucidal activity of 2 alcohol-based formulations proposed as hand rubs by the World Health Organization; AMERICAN JOURNAL OF INFECTION CONTROL, Volume: 38,Issue: 1,Pages: 66-68,Published: FEB 2010
26. Ni Zhang, Zhengwen Liu , Qunying Han , Jinghong Chen, Sai Lou, Jianming Qiu, Guoyu Zhang; Inhibition of bovine viral diarrhea virus in vitro by xanthohumol: Comparisons with ribavirin and interferon- and implications for the development of anti-hepatitis C virus agents; European Journal of Pharmaceutical Sciences 38 (2009) 332–340
27. Buckwold, V.E., Beer, B.E., Donis, R.O., 2003. Bovine viral diarrhea virus as a surrogate model of hepatitis C virus for the evaluation of antiviral agents. Antiviral Res.60, 1–15.
28. Choo, Q. L., G. Kuo, et al. (1989). "Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome." Science 244(4902): 359-362.
29. Saito I Fau - Miyamura, T., A. Miyamura T Fau - Ohbayashi, et al. (1990). "Hepatitis C virus infection is associated with the development of hepatocellular carcinoma." PNAS 87(17): 6547–6549.(0027-8424)
30. Cohen, J. (1999). "The Scientific Challenge of Hepatitis C." Science 285(5424): 26-30.
31. De Francesco, R. and A. Carfí (2007). "Advances in the development of new therapeutic agents targeting the NS3-4A serine protease or the NS5B RNA-dependent RNA polymerase of the hepatitis C virus." Advanced Drug Delivery Reviews 59(12): 1242-1262.
32. Castet, V., C. Fournier, et al. (2002). "Alpha Interferon Inhibits Hepatitis C Virus Replication in Primary Human Hepatocytes Infected In Vitro." J. Virol. 76(16): 8189-8199
33. Hoofnagle, J. H. and L. B. Seeff (2006). "Peginterferon and Ribavirin for Chronic Hepatitis C." New England Journal of Medicine 355(23): 2444-2451
34. Sharma, S. D. "Hepatitis C virus: molecular biology & current therapeutic options." (0971-5916)
35. Magdalena Sarasin-Filipowicz, Markus H Heim.Interferon-induced gene expression in chronic hepatitis C. Future Virology 2010 5:1, 25-31
36. [Lauer, G. M., and B. D. Walker. 2001.N. Engl. J. Med. 345:41–52,
37. Cao F. et. al, J. Viral. Hepat, 2010].
38. [*J. Med. Chem.*, **2005**, *48* (1), pp 1–20; Licia Tomei et al; Antiviral Chemistry & Chemotherapy 16:225–245, 2005,
39. Bernd et al, Best practice and Research clinical gastroenterology,2008, and annals of Hepatology; 2009, review].
40. Bartenschlager, R.; Lohmann, V. Novel cell culture systems for thehepatitis C virus. AntiViral Res. 2001, 52, 1-17. Bartenschlager, R. Innovation: Hepatitis C virus Replicons: Potential role for drug Development. Nat. ReV. Drug DiscoV. 2002, 1, 911-916.
41. De Francesco R & Megliaccio G. Nature, 2005; 436 (18): 953-960]
42. Lindenbach, B. D. et all; Complete Replication of Hepatitis C Virus in Cell Culture. Science 2005, 309, 623-626.

43. Wakita, T. et al, Production of Infectious Hepatitis C Virus in Tissue Culture from a Cloned Viral Genome. Nat. Med. 2005, 11, 791-796.
44. Zhong, J. et al, Robust Hepatitis C Virus Infection in Vitro. Proc. Natl. Acad. Sci. U.S.A. 2005, 102, 9294-9299.
45. Buckwold, V. E., B. E. Beer, and R. O. Donis. 2004. Antivir. Res. 60:1–15.
46. Gabriele Giliberti, Cristina Ibba, Esther Marongiu, Roberta Loddo, Michele Tonelli, Vito Boido,Erik Laurini, Paola Posocco, Maurizio Fermeglia, Sabrina Pricl,Synergistic experimental/computational studies on arylazoenamine derivatives that target the bovine viral diarrhea virus RNA-dependent RNA polymerase, Bioorganic & Medicinal Chemistry xxx (2010) xxx–xxx
47. Michele Tonelli , Vito Boido, Paolo La Colla, Roberta Loddo, Paola Posocco, Maria Silvia Paneni ,Maurizio Fermeglia, Sabrina Pricl,Pharmacophore modeling, resistant mutant isolation, docking,and MM-PBSA analysis: Combined experimental/computer-assisted approaches to identify new inhibitors of the bovine viral diarrhea virus (BVDV); Bioorganic & Medicinal Chemistry xxx (2010) xxx–xxx
48. Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevear, D.C., Rossmann, M.G., 2004. The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in de novo initiation. Proc. Natl. Acad. Sci. U.S.A. 101, 4425–4430.
49. Choi, K. H., A. Gallei, P. Becher, and M. G. Rossmann. 2006. The structure of bovine viral diarrhea virus RNA-dependent RNA polymerase and its amino-terminal domain. Structure 14:1107–1113.
50. Daniele Di Marino, Francesco Oteri, Blasco Morozzo della Rocca, Giovanni Chillemi, Mattia Falconi, ADP/ATP mitochondrial carrier MD simulations to shed light on the structural–dynamical events that, after an additional mutation, restore the function in a pathological single mutant; Journal of Structural Biology 172 (2010) 225–232
51. Maria L. Barreca, Keun Woo Lee,y Alba Chimirri, and James M. Briggsy; Molecular Dynamics Studies of the Wild-Type and Double Mutant HIV-1 Integrase Complexed with the 5CITEP Inhibitor: Mechanism for Inhibition and Drug Resistance, Biophysical Journal Volume 84 March 2003 1450–1463
52. Alcaro S, Artese A, Ceccherini-Silberstein F, Ortuso F, Perno CF, Sing T, Svicher V.; Molecular dynamics and free energy studies on the wild-type and mutated HIV-1 protease complexed with four approved drugs: mechanism of binding and drug resistance. J Chem Inf Model. 2009 Jul;49(7):1751-61.
53. VMD reference Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", J. Molec. Graphics, 1996, vol. 14, pp. 33-38)
54. Aaqvist J. Ion-water interaction potentials derived from free energy perturbation simulations. J. Phys. Chem. (1990) 94:8021–8024; J. Phys. Chem. B, 2008, 112 (30), pp 9020–9041
55. [Cornell 1995] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW and Kollman PA (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc 117: 5179-5197.
56. [Frisch 2003] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, GAUSSIAN. Gaussian, Inc. (2003) Pittsburgh PA.
57. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JJA, Vreven T, Kudin KN, Burant JC, et al, Revision C.02 ed. Gaussian, Inc. (2004)
58. Wallingford CT.; Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, GAUSSIAN. (2003) Pittsburgh PA: Gaussian, Inc.
59. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J. Comp. Chem. (2004) 25:1157–1174.
60. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. (1993) 97:10269–10280
61. [Procacci 1998] Procacci P, Darden T, Paci E, Marchi M (1998) Orac: a molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions. J Comput Chem 18: 1848-1862.
62. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J. Chem. Phys. (1980) 72:2384–2393;
63. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. (1981) 52:7182–7190
64. [Parrinello 1980] Parrinello M and Rahman A (1980). Crystal structure and pair potentials: a molecular-dynamics study. Phys Rev Lett 45: 1196-.

65. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. Mol. Phys. (1984) 52:255–268;
66. Hoover WG. Canonical dynamics: equilibrium phase-space distributions. Phys. Rev. A (1985) 31:1695–1697
67. [Tom 1993] Tom D, Darrin Y and Lee P (1993) Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. J Chem Phys 98: 10089-10092.
68. [Essman 1995] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103: 8577–8593.
69. [Bayly 1993] Bayly CI, Cieplak P, Cornell W, and Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97: 10269-10280.
70. [Tuckerman 1992] Tuckerman ME, Berne B, and Martyna G (1992) Reversible multiple time scale molecular dynamics. J Chem Phys 97: 1990–2001.
71. [Ryckaert 1997] Ryckaert JP, Ciccotti G, and Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23: 327–341
72. Daura et al. (Angew. Chem. Int. Ed. 1999, 38, pp 236-240)
73. H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. Comp. Phys. Comm., 91:43–56, 1995.
74. E. Lindahl, B. Hess, and D. van der Spoel. Gromacs 3.0: A package for molecular simulation and trajectory analysis. J. Mol. Mod., 7:306–317, 2001.
75. Carlson HA, Masukawa KM and McCammon JA (1999) Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem A 103:10213–10219.
76. [Morris 1998] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK and Olson AJ (1998) Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. J Comp Chem 19: 1639-1662.
77. R. Huey, G.M. Morris, A.J. Olson and D.S. Goodsell, J. Comput. Chem. 28 (2007), p. 1145;
78. Morris GM, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ; J Comput Chem. 2009 Dec;30(16):2785-9).
79. (Tonelli, M.; Bioorg. Med. Chem. 2010, 18, 2304)
80. (Gabriele Giliberti, et al; Bioorg. Med. Chem. 2010)
81. (Hetenyi et al; 2002; protein science)
82. F.J. Solis and R.J.-B. Wets. (1981) "Minimization by random search techniques", Mathematical Operations Research, 6, 19-30,
83. Bjelic S. et al; 2004; Biochemistry,
84. Osterberg F. et al; FEBS letters ; 2005,
85. Ragno, R. J. et al.; Med. Chem. 2005, 48, 200–212.
86. (Kumar et al; J Phys. Chem; 2010 REF for this)
87. (Singh & Thornton, 1990, JMB)
88. [Cheatham 1999] Cheatham III TE, Cieplak P and Kollman PA (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J Biomol Struct Dyn 16: 845-862.
89. T. E., P. Cieplak, and P. A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. J. Biomol. Struct. Dyn. 16:845–862.)
90. [MacKerrel 1998] MacKerell AD, Jr, Bashford D, Bellot M, Dunbrack RL, Jr, Evanseck JD, Field MJ et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102: 3586–3616.
91. [Xiang 2002] Xiang Z, Soto CS, Honig B (2002) Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction.Proc Natl Acad Sci U S A. 99(11): 7432-7.
92. [Kukolj 2005] Kukolj G, McGibbon GA, McKercher GJ, Goulet S, Poupart MA et al. (2005) Binding site characterization and resistance to a class of non-nucleoside inhibitors of the hepatitis C virus NS5b polymerase. J Biol Chem 280: 39260-7.

93. Amaro, R.E., D.D.L. Minh, L.S. Cheng, W.M. Lindstrom, Jr., A.J. Olson, J.-H. Lin, W.W. Li, J.A. McCammon. Remarkable Loop Flexibility in Avian Influenza N1 and its Implications for Antiviral Drug Design. *J. Amer. Chem. Soc.* 129, 7764-7765 (2007).

94. Amaro, R.E., R. Baron, J.A. McCammon. An Improved Relaxed Complex Scheme for Receptor Flexibility in Computer-Aided Drug Design. *J. Comput.-Aided Mol. Des.* 22, 693-705 (2008).

95. Carlson A. H. and J. ANDREW MCCAMMON; Accommodating Protein Flexibility in Computational Drug Design; December 14, 1999

96. Rommie E. Amaro, and Wilfred W. Li; Emerging Methods for Ensemble-Based Virtual Screening; Current Topics in Medicinal Chemistry, 2010, 10, 3-13

97. Pietro Cozzini et al; Target Flexibility: An Emerging Consideration in Drug Discovery and Design; J. Med. Chem; Volume 51, Number 20 October 23, 2008;

98. Jacob D. Durrant, and J. Andrew McCammon; Potential Drug-Like Inhibitors of Group 1 Influenza Neuraminidase Identified through Computer-Aided Drug Design; Comput Biol Chem. 2010 April ; 34(2): 97–105

99. Jung-Hsin Lin, Alexander L. Perryman, Julie R. Schames, and J. Andrew McCammon; Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme; J. AM. CHEM. SOC. 2002, 124, 5632-5633

100. Kristin L. Meagher and Heather A. Carlson; Incorporating Protein Flexibility in Structure-Based Drug Discovery: Using HIV-1 Protease as a Test Case ; J. AM. CHEM. SOC. 2004, 126, 13276-13281

101. J. Andrew McCammon; Target flexibility in molecular recognition; Biochimica et Biophysica Acta 1754 (2005) 221 – 224

102. B-Rao et al. Drug Discovery Today, 14, 394 – 2009
103. (porollo, A. et al 2010; Nucleic Acid Research)
104. (Cunningha, B. C.; Science1989 and Ge-Fei Hao et al. J. Phys Chem B 2010)
105. porollo, A. et al 2010; Nucleic Acid Research
106. (Coleman et al 2009; biophysical)

III

Effect of Resistance Mutation on BVDV RdRp

An Energetic Aspect

Inhibition of viral RNA polymerase investigated by computer simulations

Shailendra Asthana ,[†] Saumya Shukla ,[†] Matteo Ceccarelli,[‡] Paolo La Colla,[†] Paolo Ruggerone,[*,‡] and Attilio V. Vargiu[‡]

Department of Biomedical Sciences and Technologies, University of Cagliari, S.P. Monserrato-Sestu Km 0.700, I-09042 Monserrato (CA), Italy , and Department of Physics, University of Cagliari, S.P. Monserrato-Sestu Km 0.700, I-09042 Monserrato (CA), Italy

E-mail: paolo.ruggerone@dsf.unica.it

Abstract

The impressive progress in computational techniques applied to biologically and pharmacologically inspired problems is facing the challenging task of extending even more the range and the complexity of the questions that should be addressed. A single technique may not be well suited to shade light on a specific problem and the combination of different computational approaches will provide the appropriate answer. The example of such combined strategy is discussed in the present paper in relation to the inhibition of replication in Bovine Viral Diarrheal Virus (BVDV), a member of the Flaviviridae family that contains also the Hepatitis C Virus (HCV). In a previous work, a promising inhibitor (227G) targeting the virally encoded RNA-dependent RNA polymerase (RdRp) of both BVDV and HCV was identified by combining biological assays and computer simulations, and the effects of a resistant mutation (I261M) was identified. In the present work we applied several biophysical computational approaches,

[*]To whom correspondence should be addressed
[†]University of Cagliari
[‡]University of Cagliari

ranging from MM/PBSA to metadynamics to provide a more quantitative understanding of the inhibition action and of the resistance mechanism developed by the virus after the I261M mutation. We calculated the free energy of binding of 227G and characterized its dissociation from the wild-type and mutated RdRps of BVDV in order to provide an estimate of binding residence times, which are related to activity. The evaluated free energies of binding agree well with the experimentally derived value, confirming the validity of the approach. The biophysical picture of the inhibition action and the quantitatively results obtained in this study may be of support in defining an efficient strategy to design new, more potent and selective antivirals.

Introduction

Understanding ligand-protein interaction processes is crucial for modern drug discovery process, which is often fraught with critical junctures at which decisions are difficult and a molecular knowledge is crucial to gain the upper hand thereon. Computational approaches can play an important role in this challenging task. This role can thrive even more if different computational approaches from bioinformatics to state-of-the-art molecular simulations at the desired level of accuracy are combined properly. Models (if properly validated) may provide insights in the details of the processes, allowing to access features often difficult to be observed experimentally. In the absence of experimental data, structural models can be built and represent the starting point of simulations that have the double functions of gaining insights in the details of the processes and of validating the models. Computational molecular biology approaches are keys to face this challenge, providing a way to effectively bridge the gap between the time and size scales reachable by theoretical methods and those biologically relevant.

An application of this combined strategy based on different computational approaches, of which the appropriate interplay is a challenging and delicate issue itself, is represented by the study of the action of antiviral compounds. In particular, our attention focused on the inhibition action of 227G (see Figure 1a, where also the numbering of the atoms is reported), which has been demonstrated to be promisingly active against both Hepatits C Virus (HCV) and Bovine Vi-

ral Diarrheal Virus (BVDV) belonging to the Flaviviridae family. HCV and BVDV have created havoc by effecting Human race directly or indirectly. HCV (officially recognized as the *silent epidemic*) infected approximately 180 million people worldwide, and it is the leading reason for liver transplantation in the United States.[1,2] It creates a significant burden to health-care systems attributable to mortality (250,000 deaths can be attributed to HCV each year in the world), morbidity and treatment costs.[1,3,4] Unfortunately, more than 20 years after discovery of HCV[5] therapeutic options remain limited. No vaccine or effective therapy broadly targeting all genotypes of HCV[6] is disposable at present. Antiviral treatment is unlikely to induce significant population-wide reductions, and mortality is expected to continue to increase, at least in the next 10-20 years. BVDV causes a range of clinical manifestations in cattles,[7] thereby continuing to be a financial burden to the farming industry. In fact, diseases related to BVDV cause decreased performance, loss of milk production, reproductive wastage, and increased risk of morbidity and mortality. It is the most costly viral disease in US cattle herds, with losses estimated at $2 billion dollars per year. Analogously to HCV, an appropriate treatment of BVDV infections is still far from having been developed.[8] Thus there is a call for highly effective and selective inhibitors of HCV and BVDV replication to develop improved therapeutic options and to keep the upper hand on viral infections.

The major efforts to develop Flaviviridae-directed antiviral agents have focused on the inhibition of essential virally-encoded enzymes such as viral polymerases, following the successful paradigm established for HIV, for which inhibitors of the reverse transcriptase have reached clinical use.[9,10] The Flaviviridae NS5B RNA-dependent RNA polymerase (RdRp) is considered one of the most interesting targets, since polymerase activity is essential for viral replication, and mammalian host cells are devoid of such RdRp, thus offering the opportunity to identify very selective inhibitors of the viral enzyme.[11] The structure of BVDV RdRp is shown in Figure 1b, with the important regions highlighted in different colors.

In the search of antivirals, the benzimidazole class of compounds has been extensively pursued as potent inhibitors[12–15] Among them we identified a promising candidate, 227G, which appears powerful in inhibiting BVDV RdRp.[16,17] In the present paper we deepen the analysis of 227Gs'

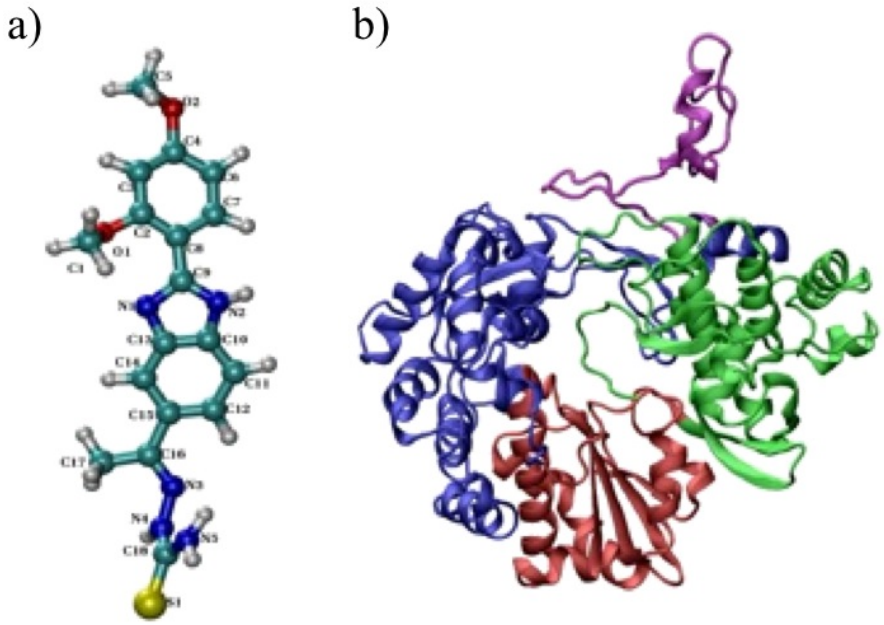


Figure 1: (a) Ball and stick representation of 227G with oxygen atoms in red, nitrogen atoms in blue, carbon atoms in cyan, sulfur atom in yellow, and hydrogen atoms in white. (b) Cartoon representation of BVDV RdRp. Different domains of RdRp are differently colored: thumb in green, finger in blue, palm in red, and the unique N-terminal in magenta.

action on wild type and resistant mutated BVDV RdRps via a thorough application of different computational approaches, ranging from standard molecular dynamics (MD) simulations, metadynamics, and molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) calculations. The I261M mutated RdRp has been proven to mantain its functionality also in the presence of 227G, making the inhibitor useless. Our study on the wild type and mutated RdRps identified the key residues involved in the action mechanism and in the resistance development by mapping the interactions during the unbinding process of 227G from the RdRps. Our hypotheses are conforted by the good agreement of computational results with the experimentally available data. In conclusion, this study allowed not only to identify additional determinants of 227G action, but also to validate a computational protocol that might be very powerful in supporting a more efficient and structure-based drug design.

Results and discussion

The computational approaches we used in the present work aim at providing different and complementary answers to the problem of 227G action. While well suited to simulate a rare event such the dissociation process of 227G from RdRp and to furnish the values of the associated activation barriers, metadynamics is not always a convenient tool to evaluate the binding free energy. This is due to the fact that the calculation of this latter quantity requires the evaluation of the association process, for which the reaction coordinates chosen to investigate the unbinding might be inappropriate and unable to provide a good converged free energy. Thus, we estimated the binding free energy by performing MM/PBSA calculations.

Unbinding of the compound: use of metadynamics

To quantitatively assess the different behaviour of 227G bound to the wild type and I261-mutated BVDV RdRps (hereafter COMPLEX and COMPLEXm, respectively) we performed metadynamics simulations of the unbinding process of the compound from the two proteins (see *Materials*

and Methods for more details on metadynamics). By accelerating the dynamics of the collective coordinates that describe the unbinding process, i.e., the distance between the centers of mass of 227 and of the binding pocket (d_{com}) and the number of hydrophobic contacts (n_{hph}), it is possible to catch the essential features of a process otherwise unaccessible by standard MD simulations. The different paths followed by 227G by leaving COMPLEX and COMPLEXm are visualized in Figure 2a and b. Remarkable is the behaviour of the loops and its correlation with the unbinding of 227G, which is correlated to the different initial *packing* of the compound in the two complexes. Indeed, only in COM the ligand is *covered* by the loops, in particular the linker L4. Therefore, it is consistent that the dissociation of 227G is accompanied to a significant displacement of L4 (see upper and lower panels of Figure 2a). At opposite, the path of 227G to escape from the mutated RdRp goes aside of the linker, and is not accompanied by any significant movement of the linker but only by fluctuations.

Beside the qualitative picture of the unbinding process, from the metadynamics runs the free energy underlying the process can be extracted in function of the accelerated coordinates. For COMPLEX and COMPLEXm the free energy surfaces (FESs) extracted from the simulations are shown in Figure 3 and Figure 4. The FES associated with the undocking from COMPLEX (Figure 3) presents two minima: the deepest one (1 in Figure 3) corresponds to the initial orientation identified in a previous work of us.[17] In this initial position 227G points its benzyl group toward the template entrance gate of the polymerase and is stabilized in the cavity by a strong hydrogen bond (HB) formed by the atom N2 of azole moiety with the residue Asp126, which lyes in the linker. During the standard MD simulations this HB is present for more than 95% of the simulation. A second HB appears between the atom N5 of 227G and the side chain oxygen of Ser533 and is observed for more than 50% of the time. Additionally, several hydrophobic (HpH) contacts contribute to the stabilization of 227G in this minimum. In particular, HpH contacts involve Pro262, Ile261, Arg127, Glu128, Arg132, Gly220, Ala221, and Tyr289. Leaving this stable position in the binding pocket requires 227G to overcome an energy barrier of $\sim$12kcal/mol. The departure from Minimum 1 starts with the disruption of the Hbond connecting 227G to Asp126. As a conse-

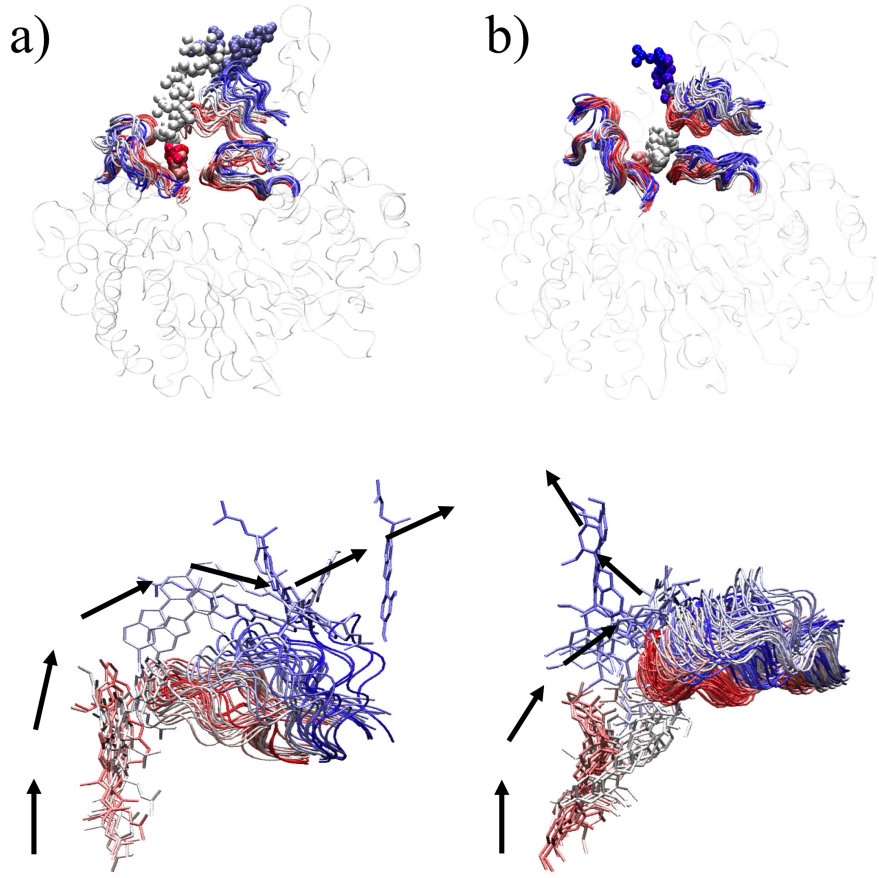


Figure 2: Evolution of the center of mass of 227G in (a) COMPLEX and (b) COMPLEXm during the metadynamics simulations of the undocking process. In the upper panels the whole system (in ghost) is represented, while in the lower panels only 227G and the relevant loops are shown. Loops and 227G are represented in a time-step-coloring fashion, which colors differently the configurations at different steps: initial states in red, half-trajectory states in white, and final states in blue. The arrows in the lower panels highlight the exit paths of 227G from COMPLEX and COMPLEXm. Pictures were made with VMD.[18]

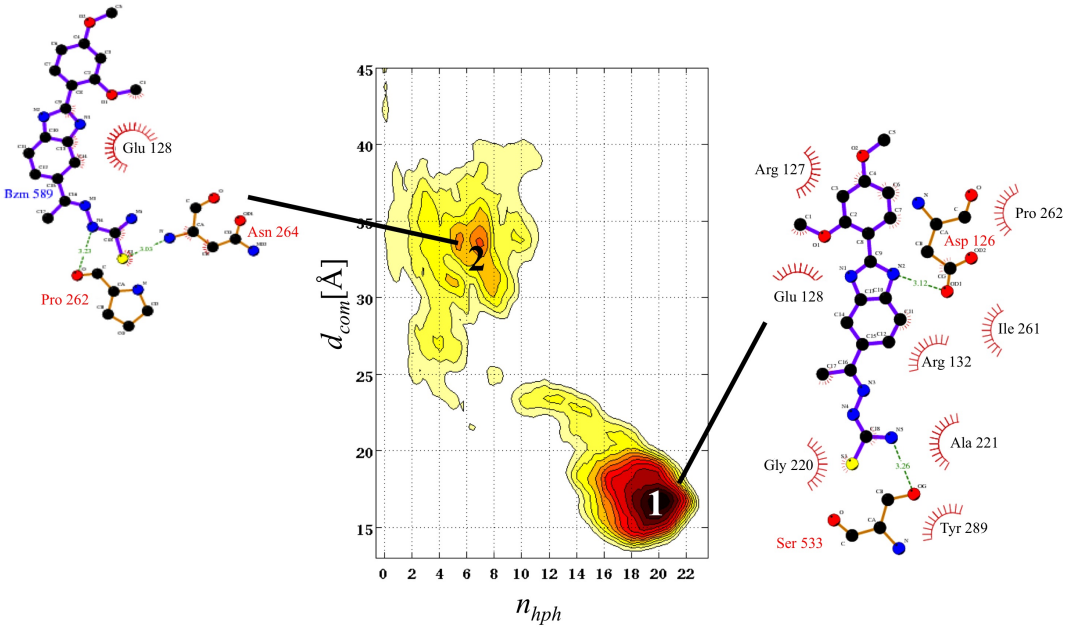


Figure 3: Free energy surfaces as a function of d_{com} and n_{hph} for the unbinding of 227G from COMPLEX. Isosurfaces are drawn one per 1 kcal/mol. Insets show Ligplot of snapshots of 227G in the minima extracted from the metadynamics runs.

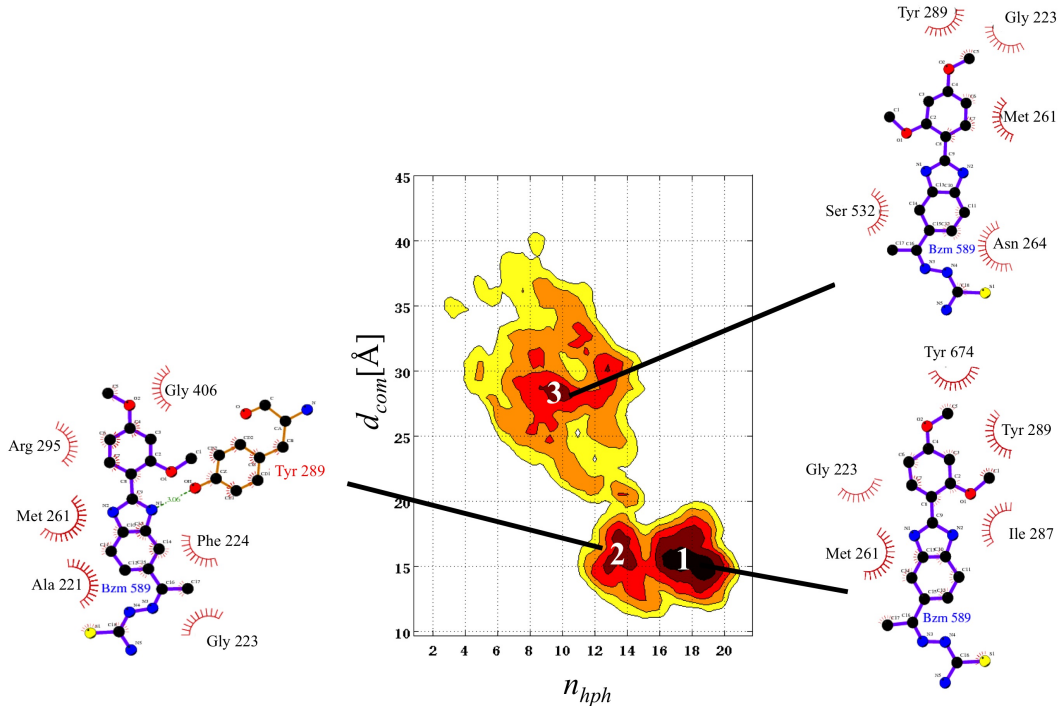


Figure 4: Same as Figure 3 for the unbinding of 227G from COMPLEXm.

quence, 227G moves slightly deeper into the cavity with a small rotation of the azole moiety around the compounds' major axis. The displacement brings the N2 atom of 227G closer to the residue Pro262 of Motif I and a HB is established between them. The next step involves the breaking of the HB between N5:227G and Ser533:O, this latter a residue of loop L3. As a consequence the tail of 227G acquires mobility and the compound rearranges itself coming closer to Loop L2 with the formation of a new HB, Asn217:ND2 @S1:227G. Successively, the unbinding proceeds and the two HBs bonds, Asn217:ND2 @S1:227G and Pro262:O@N2:227G, are disrupted. This causes a slight jump of the 227G toward the mouth of the cavity, which is flanked mainly by Arg127 of loop L4, Ala392 of Loop L1 and Phe224 of Loop L2. Due to the movement of 227G an opening of the entrance starts as Arg127 of Loop L4 moves away from Loop L1.

227G moves further upwards and lies mainly between Loops L4, L2 and L1. At this stage the tail of 227G is held by two HBs with residues Pro262 of Motif I and Arg132 of Loop L4. A second, very shallow minimum (2 in Figure 3) is reached by 227G along the unbinding route at

$d_{com} \sim 34$ Å and n_{hph} between 5 and 7. In Minimum 2 the major stabilizing contributions in form of HpH contacts are essentially lost with the exception of that involving Glu128. Additionally, two weak Hbonds with residues Pro262 and Asn264 belonging to Motif I are present. The role of Minimum 2 in the unbinding process needs further analysis. Indeed, Minimum 2 might act as an intermediate along the escape route, from which 227G has an equal probability to definitely dissociate or to re-associate, being the free energy barrier for both processes ~ 5 kcal/mol.

As shown in Figure 4, the topology of the FES for COMPLEXm is different from that of COMPLEX, as three pronounced minima are clearly recognizable. The first one (Minimum 1 in Figure 4) is the stable position in which 227G were found after a docking study and standard MD simulations. Note that here 227G has its sulfur atom directed toward the template entrance gate, assuming an orientation opposite to that characterizing the deepest minimum found in the wild type polymerase. In Minimum 1 227G is stabilized by strong HpH contacts with Met261, Ile287, and Try289. Additional HpH contacts are provided by Gly223 and Tyr674. 227G positions itself in between Loops L1 and L2. A small barrier of ~ 2 kcal/mol separates Minimum 1 to a second basin, Minimum 2, where 227G interacts essentially via HpH contacts with residues Ala221, Gly223, Phe224, Met261, Arg295, and Gly406 belonging to Loop L2 and Motif I. A short-life HB connects residue Tyr289 and the azole moiety of 227G. This HB is soon disrupted and the tail of 227G starts interacting with Glu128 of loop L4. From Minimum 2, 227G reaches another intermediate state, Minimum 3, with a free energy cost of ~ 5 kcal/mol. In this minimum the major contribution to the stabilization of 227G comes from Met261, Asn264, Tyr289, and Ser532, which are residues of Loop L3 and Motif I, in form of HpH contacts. From Minimum 3 a barrier of ~ 4 kcal/mol separates 227G from the aqueous environment. Before reaching the solvent 227G keeps a weak linkage to the protein by interacting with the residues of the hood of the N-terminal (residues Val92 to Ser122).

Comparing the two FESs, it immediately appears that the deepest minimum in COMPLEX is more pronounced than in COMPLEXm. Additionally, the free energy barriers are also larger in COMPLEX than in COMPLEXm, assessing an enhanced stability of 227G in the wild type RdRp

than in the mutated polymerase. In particular, the largest barrier along the escape route is ~ 7 kcal/mol smaller than in COMPLEX.

To gain a more quantitative assessment of the stabilization in COMPLEX and COMPLEXm we evaluated the interaction energies of the compound with the relevant residues within the framework of the Amber force-field description. The results are reported in Figure 5. Note the large reduction in the interaction with residue Asp126 (which has the most stabilizing interaction with 227G in COM) when the mutation occurs. This is consistent with the different position of the linker in the two complexes, as Asp126 belongs to this loop, which is approaching very closely 227G in COM. This structural arrangement of L4 has the double effect of stabilizing 227G and of reducing the space available to the template RNA strand to enter the polymerization site. In COMm the loop L4 undergoes more pronounced fluctuations and on average it is not able to approach and to interact with 227G. Such a lack of interaction would probably leave enough space for the entrance of the RNA template in the polymerase an consequently for the functioning of the system.

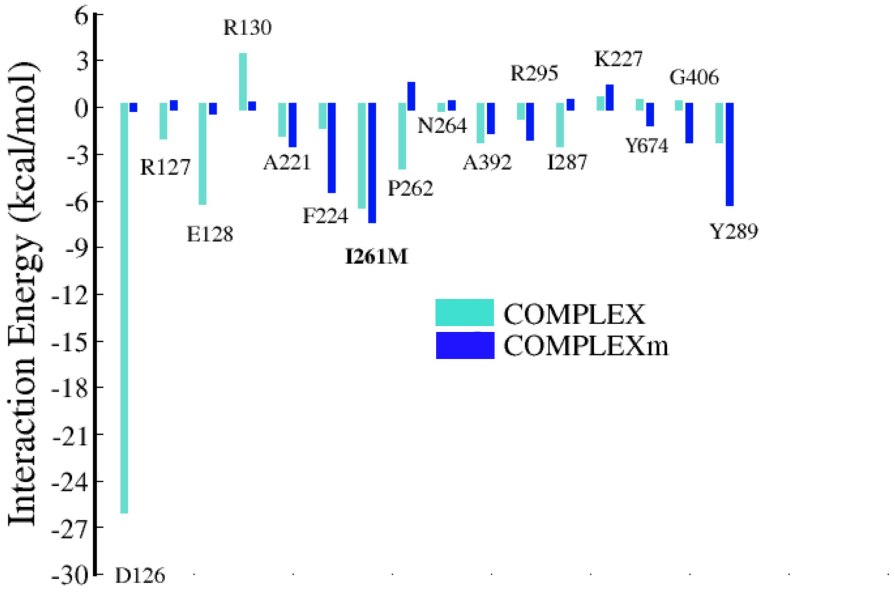


Figure 5: Ligand-residue interaction energy (in kcal/mol) for COMPLEX (cyan) and COMPLEXm (blue).

Binding free energies

The energy barriers required to leave the binding sites in COMPLEX and COMPLEXm offer a first quantitative (although approximated) indication of the different activity of 227G in the wild type and mutated RdRp. However, a validation of the results obtained in the present work needs a more close comparison with the available data coming from biological assays. To achieve this goal we evaluated the so-called half maximal inhibitory concentration (IC_{50}), a measure of the effectiveness of a compound in inhibiting biological or biochemical function. IC_{50} can be related to the binding free energy ΔG_{bind} by the following equation:[19,20]

$$\Delta G_{bind} = RT\,ln\mathrm{IC}_{50}, \tag{1}$$

where R is the universal gas constant and T the temperature. To calculate ΔG_{bind} we used the

Table 1: Calculated values of the contributions to ΔG_{bind} in kcal/mol. The experimental value reported for comparison is extracted from measured IC_{50} [16] by using Equation Eq. (1) at T = 298 K. The third column contains for each term the difference between the values of COMPLEX and COMPLEXm.

	COMPLEX	COMPLEXm	Δ(COMPLEX-COMPLEXm)
$\Delta E_{electrostatic}$	-36.0 (± 1.9)	-42.9 (± 2.7)	6.9
$\Delta E_{vanderWaals}$	-52.2 (± 2.9)	-39.9 (± 1.8)	-12.3
ΔG_{PB}	-24.1 (± 2.7)	-13.7 (± 3.2)	- 10.4
ΔG_{NP}	-5.8 (± 1.2)	-5.6 (± 0.9)	-0.2
PB_{tot}	-33.9 (± 3.1)	-29.8 (± 3.4)	- 4.1
$\Delta G_{electrostatic+PB}$	-11.9 (± 3.9)	-29.2 (± 2.6)	17.3
$\Delta G_{vanderWaals+NP}$	-57.9 (± 4.7)	-45.6 (± 3.8)	-12.3
TS_{solute}	-18.9 (± 2.3)	-17.6 (± 1.8)	-1.3
ΔH	-69.8	-74.8	5.0
$\Delta G_{binding}$	-14.0	-12.2	-1.8
Experiment			
$\Delta G_{binding}$	-12.0	not available	

MM/PBSA method[21–23] that allows ΔG_{bind} to be dissected in the single contributions within the framework of the Amber force-field. The results for COMPLEX and COMPLEXm are listed in Table 1. The binding free energy for COMPLEX was found to be -14.0 kcal/mol. The binding free

energy obtained from experimental IC_{50} value by inverting Equation Eq. (1) was -12.0 kcal/mol, which was in good agreement with the result extracted from the MM/PBSA calculation and with the largest energy barrier extracted from the metadynamics simulations. Following the same procedure, we calculated the binding free energy of 227G in COMPLEXm which came out to be -12.2 kcal/mol.

Hydration of the compound along dissociation path

An important role in the drug-target interaction, and more in general in the ligand-protein interaction, is played by the solvent.[23–27] A way to characterize this contribution is to evaluate the first shell of water molecules surrounding the 227G in COMPLEX and COMPLEXm during the simulations. The calculated probability distributions of the water population in this first shell are

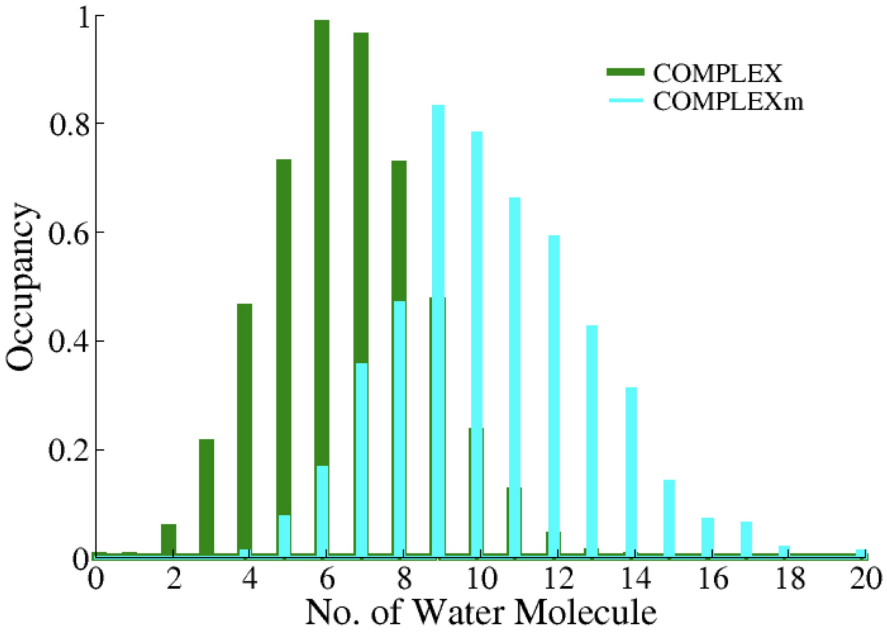


Figure 6: Histogram of number of water molecules around the drug during the simulation in COMPLEX and COMPLEXm system represented by green and cyan, respectively.

reported in Figure 6. In green are represented the data related to COMPLEX while in cyan those

obtained for COMPLEXm. Remarkable differences are present. With respect to COMPLEX the distribution for COMPLEXm is shifted toward larger numbers of water molecules around 227G. In fact, the average number of waters ringing the inhibitor is ~ 6 in the case of COMPLEX and ~ 10 in COMPLEXm. Thus, the compound is more exposed to the solvent in the mutated RdRp than in the wild type, indicating a cavity more open to the aqueous environment in the former than in the latter system. The interaction of the inhibitor with Loop L4 present in COMPLEX but not in COMPLEXm is probably responsible for the hindered access of water to the cavity. A further confirm of the different exposure to the solvent in COMPLEX and COMPLEXm comes from the calculation of the solvent accessible surface area (SASA) for the two adducts by Platinum.[28] The binding pocket in COMPLEX results more hydrophobic than in COMPLEXm.

Conclusions

To develop more efficient antiviral agents a more detailed microscopic knowledge of the features determining mechanism of action and stability of compounds is necessary. A structure-based drug design has to include structural features at large, intending with this dynamical effects and interaction with the solvent. In the present work, we applied to the problem of an inhibitor, 227G, acting on the wild type and mutated BVDV RdRp, a computational strategy that accounts for all these effects in an all-atoms description. By studying the undocking mechanism we were able to accomplish an inventory of the interactions between 227G and the various residues of RdRp. At the same level of accuracy, the hydration of the system along the dissociation route was evaluated pointing out the different degree of hydration in the wild type and mutated RdRps. Finally, we estimated the protein ligand binding free energy, which came out in good accordance with the binding free energy extracted from the available experimentally determined IC_50. Such agreement represents a first validation of a computational protocol that can be used to provide hints and suggestions for the rational drug design mentioned above.

Materials and Methods

Standard MD and Metadynamics

Initial relaxation of the systems of COMPLEX and COMPLEXm was performed for 30ns at constant volume and temperature using the MD software package ORAC.[29] The process of undocking, a rare event, which can not be reached by standard MD simulations with an all-atom representation. The metadynamics algorithm[30] is based on a history dependent biasing potential added in a subspace defined by a chosen set of reaction coordinates $s_\alpha(x)$, aimed at reconstructing the multidimensional free energy of a given process. At time t the biasing potential V_G is given as the sum of repulsive Gaussian functions added with a frequency $1/\tau_G$:

$$V_G[s(x),t] = \sum_{t'=\tau_G,2\tau_G,3\tau_G,\ldots;t'<t} We^{-\frac{[s(x)-s(t')]^2}{2(\delta s)^2}}, \tag{2}$$

where W is the Gaussian height, δs the Gaussian width. Due to this potential, the system is discouraged from revisiting the configurations already sampled. Metadynamics not only allows the acceleration of rare events, but also the reconstruction of the free energy $F_G(s,t) = -V_G$, which is an approximation of $F(s)$ in the region $\Sigma(s)$ explored by $s(x_G(t))$ up to time t.[31,32] The accuracy of free energy reconstruction is dependent upon the Gaussian parameters W, δs, and τ_G. Details of the metadynamics algorithm have been previously described.[30,31,33]

The choice of reaction coordinates (RCs) is pivotal to obtain the best approximation of the free energy. The RCs used here to describe COMPLEX and COMPLEXm dissociation are:

1. The distance d_{com} between the center of mass of the ligand and the center of the mass of the protein. A similar choice of RCs has been applied e.g. in Ref.[34]

2. The number of hydrophobic contacts n_{hph} between non polar carbons on the ligand and on the residues that it covers XXARE YOU SURE?XX in the starting structure, modeled as a

coordination number:

$$n_{hph} = \sum_{i,j} \frac{1-(r_{ij}/r_0)^a}{1-(r_{ij}/r_0)^b}, \tag{3}$$

where r_{ij} is the distance between the atoms of interest, a and b have values of 6 and 12, respectively, and $r_0 = 6$ accounts for the typical carbon-carbon distance (4/4.5 Å) and the thermal motionsÕ amplitude (1.5/2 Å). A similar RC has been used e.g. in Refs.[27,35,36]

The Gaussian parameters were W = 0.25 kj/mol, $\delta s_{d_{com}}$ =0.5 Å, and $\delta s_{n_{hph}}$ =6 in both the case of COMPLEX and COMPLEXm. The time interval between two successive Gaussian depositions was set to 0.5 ps in all the simulations.

Free energies surfaces were calculated as a function of (d_{com}, n_{hph}). On COMPLEX and COMPLEXm we performed 22ns-long and 12ns-long metadynamics runs, respectively. The activation free energies associated to the detachments were extracted by stopping the summation over Gaussians just after the complete detachment of the drug.[27,34]

MM/PBSA Calculation

Since the minimum energy pose obtained from metadynamics corresponded to the initial docking position, MM/PBSA calculations were performed from the snapshots taken from the 30ns-long standard MD simulations of the initial docking pose in both COMPLEX and COMPLEXm. According to the MM/PBSA method binding free energy (ΔG_{bind}) of each system could be conceptually summarized as follows

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{lig}). \tag{4}$$

Here G_{com}, G_{rec}, and G_{lig} the free energy for the RdRp-inhibitor complex, receptor (RdRp), and ligand (inhibitor), respectively. Each term is calculated by averaging for each system the following free energy energy $\Delta G = \Delta E_{MM} + \Delta G_{solv} - T\Delta S$, where ΔE_{MM} refers to the molecular mechanics energy, ΔG_{solv} to the solvation free energy, and $T\Delta S$ to the vibrational entropy and are calculated as follows:

$$\Delta E_{MM} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{torsion} + \Delta E_{vdw} + \Delta E_{elect}, \tag{5}$$

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{SA}, \tag{6}$$

$$\Delta G_{SA} = \gamma SA + b. \tag{7}$$

E_{MM} includes the average molecular mechanics energy contributed by bonded (E_{bond}, E_{angle}, and $E_{torsion}$) and nonbonded (E_{vdw} and E_{elect}) terms of the force field. Here E_{MM} was further decomposed in the individual nonbonded contributions from residues belonging to the binding pocket.[37]

ΔG_{solv} is the solvation free energy, which is due to polar (ΔG_{PB}, evaluated using the Poisson-Boltzmann equation) and nonpolar (ΔG_{SA}, proportional to the surface area[38]) terms. The electrostatic solvation free energy was calculated using DELPHI[39,40] software, with low dielectric medium for solute (ε=1) and high dielectric medium for solvent (ε=80). Atomic radii were taken from PARSE[38] with an additional value of 1.90 Å for phosphorus,[41] and in order to be consistent with molecular mechanics energy calculation, the partial charges on solute were taken from the Amber 99 (RNA) force field. An 80% boxfill lattice with grid spacing 0.5 grid/Å was applied, and 10,000 linear iteration steps were required to obtain energy convergence. The nonpolar contribution to solvation free energy was determined from Equation Eq. (7), where the surface area was calculated using MOLSURF,[42] and γ and b are 0.00542 kcal/mol Å^2 and 0.92 kcal/mol, respectively, for use with PARSE atomic radii.[38] The solvent probe radius was set to 1.4 Å. Residues within 25 Å from the mass center of ligand were used for PBSA calculation. This included the ligand and water molecules located within 5 Å from the drug.[37]

The solute entropy contribution ($-T\Delta S$) was estimated by normal-mode analysis[21] using the NMODE module in AMBER 9.0. Each of the complex, receptor, and ligand systems were first minimized before switching over for normal-mode analysis, using a distance dependent dielectric constant of $\varepsilon = 4r$ with r interatomic distance, in the absence of solvent to mimic solvent screening[43] until the convergence criteria of 0.0001 kcal/molÅ was achieved. Normal mode calculation was extremely time-consuming and computationally expensive; therefore, only residues within 10 Å from the inhibitor center of mass (including water molecules and ions) were used here.[37] As discussed in Refs. 44 and 45 the differences of the calculated entropy value is quite small for dif-

ferent conformations, and normal-mode analysis calculation usually gave an inaccurate estimation for solute entropies, thus giving merely qualitative estimates of the solute entropy.[46] Therefore, our calculation was only based on the average entropy value obtained from 100 snapshots taken from the final 20 ns of MD trajectories with a time interval of 0.2 ns.

Acknowledgement

We thank Prof. N. Marongiu and Dr. G. Giliberti (University of Cagliari) for fruitful discussions. Calculations were performed thanks to the availability of CPU time at Cybersar (c/o Physics Department, University of Cagliari).

Supporting Information Available

References

(1) Cohen, J. *Science* **1999**, *285*, 26–30.

(2) Ghany, M. G.; Strader, D. B.; Thomas, D. L.; Seeff, L. B. *Hepatology* **2009**, *49*, 1335Ð1374.

(3) Choo, Q. L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M. *Science* **1989**, *244*, 359–362.

(4) Saito, I.; Miyamura, T.; Ohbayashi, A.; Harada, H.; Katayama, T.; Kikuchi, S.; Watanabe, Y.; Koi, S.; Onji, M.; Ohta, Y. *Proc Natl Acad Sci USA* **1990**, *87*, 6547Ð 6549.

(5) Shi, S. T.; Herlihy, K. J.; Graham, J. P.; Fuhrman, S. A.; Doan, C.; Parge, H.; Hickey, M.; Gao, J.; Yu, X.; Chau, F.; Gonzalez, J.; Li, H.; adn Amy K Patick, C. L.; Duggal, R. *Antimicrob Agents Chemother* **2008**, *52*, 675–683.

(6) Sarasin-Filipowicz, M.; Heim, M. H. *Future Virology* **2010**, *5*, 25–31.

(7) Meyers, G.; Thiel, H. J. *Adv Virus Res* **1996**, *47*, 53–118.

(8) Finkielsztein, L. M.; Moltrasio, G. Y.; Caputto, M. E.; Castro, E. F.; Cavallaro, L. V.; Moglioni, A. G. *Curr Med Chem* **2010**, *17*, 2933–2955.

(9) Clercq, E. D. *Curr Opin Microbiol* **2005**, *8*, 552–560.

(10) Das, K.; Bauman, J. D.; Clark, A. D.; Frenkel, Y. V.; Lewi, P. J.; Shatkin, A. J.; Hughes, S. H.; Arnold, E. *Proc Natl Acad Sci USA* **2008**, *105*, 1466–1471.

(11) Francesco, R. D.; Migliaccio, G. *Nature* **2005**, *436*, 953–960.

(12) Ishida, T.; Suzuki, T.; Hirashima, S.; Mizutani, K.; Yoshida, A.; Ando, I.; Ikeda, S.; Adachi, T.; Hashimoto, H. *Bioorg Med Chem Lett* **2006**, *16*, 1859–1863.

(13) Hirashima, S.; Suzuki, T.; Ishida, T.; Noji, S.; Yata, S.; Ando, I.; Komatsu, M.; Ikeda, S.; Hashimoto, H. *J Med Chem* **2006**, *49*, 4721–36.

(14) Hwu, J. R.; Singha, R.; Hong, S. C.; Chang, Y. H.; Das, A. R.; Vliegen, I.; Clercq, E. D.; Neyts, J. *Antiviral Res* **2008**, *77*, 157–62.

(15) Ikegashira, K.; Oka, T.; Hirashima, S.; Noji, S.; Yamanaka, H.; Hara, Y.; Adachi, T.; Tsuruha, J.-I.; Doi, S.; Hase, Y.; Noguchi, T.; Ando, I.; Ogura, N.; Ikeda, S.; Hashimoto, H. *J Med Chem* **2006**, *49*, 6950–3.

(16) Asthana, S.; Shukla, S.; Vargiu, A. V.; Giliberti, G.; Marongiu, N.; Ruggerone, P.; Colla, P. L. *to be submitted*.

(17) Asthana, S.; Shukla, S.; Vargiu, A. V.; Giliberti, G.; Marongiu, N.; Ruggerone, P.; Colla, P. L. *to be submitted*.

(18) Humphrey, W.; Dalke, A.; Schulten, K. *J Molec Graphics* **1996**, *14*, 33–38.

(19) Carta, A.; Loriga, M.; Paglietti, G.; Ferrone, M.; Fermeglia, M.; Pricl, S.; Sanna, T.; Ibba, C.; Colla, P. L.; Loddo, R. *Bioorganic & Medicinal Chemistry* **2007**, *15*, 1914–1927.

(20) Mazzei, M.; Nieddu, E.; Miele, M.; Balbi, A.; Ferrone, M.; Fermeglia, M.; Mazzei, M. T.; Pricl, S.; Colla, P. L.; Marongiu, F.; Ibba, C.; Loddo, R. *Bioorganic & Medicinal Chemistry* **2008**, *16*, 2591–2605.

(21) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc Chem Res* **2000**, *33*, 889–897.

(22) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. *J Med Chem* **2006**, *49*, 6596–606.

(23) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Biophysical Journal* **2008**, *94*, 550–61.

(24) Zhang, L.; Wang, L.; Kao, Y.-T.; Qiu, W.; Yang, Y.; Okobiah, O.; Zhong, D. *Proc Natl Acad Sci USA* **2007**, *104*, 18461–6.

(25) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. *Proc Natl Acad Sci USA* **2007**, *104*, 808–13.

(26) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. *J Am Chem Soc* **2008**, *130*, 2817–31.

(27) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. *Nucleic Acids Research* **2008**, *36*, 5910–21.

(28) Pyrkov, T. V.; Chugunov, A. O.; Krylov, N. A.; Nolde, D. E.; Efremov, R. G. *Bioinformatics* **2009**, *25*, 1201–2.

(29) Procacci, P.; Darden, T. A.; Paci, E.; Marchi, M. *J Comput Chem* **1997**, *18*, 1848–1862.

(30) Laio, A.; Parrinello, M. *Proc Natl Acad Sci USA* **2002**, *99*, 12562–6.

(31) Laio, A.; Gervasio, F. L. *Rep Prog Phys* **2008**, *71*, 126601.

(32) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(33) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *The journal of physical chemistry B* **2005**, *109*, 6714–21.

(34) Gervasio, F. L.; Laio, A.; Parrinello, M. *J Am Chem Soc* **2005**, *127*, 2600–2607.

(35) Fiorin, G.; Pastore, A.; Carloni, P.; Parrinello, M. *Biophysical Journal* **2006**, *91*, 2768–77.

(36) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J Am Chem Soc* **2006**, *128*, 13435–13441.

(37) Kongsted, J.; Ryde, U. *J Comput Aided Mol Des* **2009**, *23*, 63–71.

(38) Sitkoff, D.; Sharp, K. A.; Honig, B. *J Phys Chem* **1994**, *98*, 1978Ð1988.

(39) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J Comput Chem* **2002**, *23*, 128Ð137.

(40) Honig, B.; Sharp, K. A.; Yang, A. S. *J Phys Chem* **1993**, *97*, 1101Ð1109.

(41) Rashin, A. A. *Biopolymers* **1984**, *23*, 1605Ð1620.

(42) Norinder, U.; Österberg, T.; Artursson, P. *Pharm Res* **1997**, *14*, 1786–1791.

(43) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. *Proc Natl Acad Sci USA* **1999**, *96*, 14330Ð14335.

(44) Reyes, C. M.; Kollman, P. A. *J Mol Biol* **2000**, *297*, 1145Ð1158.

(45) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J Am Chem Soc* **2001**, *123*, 5221Ð5230.

(46) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J Biomol Struct Dyn* **1998**, *16*, 265Ð280.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

III)-
DABO snuggles its way to HIV-RT Inhibition

Combining Docking, Molecular Dynamics to Predict Binding Modes and Affinities for Non-nucleoside Inhibitors to HIV-1 Reverse Transcriptase

Abstract

The reverse transcriptase of human immunodeficiency virus (HIV) catalyzes a series of reactions to convert the single-stranded RNA genome of HIV into double-stranded DNA for host-cell integration, This task requires the reverse transcriptase from the viral genome. Docking, scoring, molecular dynamics (MD), methods are used here to predict binding modes and affinities for a set of 4 non-nucleoside inhibitors to HIV-1 reverse transcriptase. The location of each drug outside the binding pocket was determined by an automated docking program, and steering into the binding pocket followed a route that is likely to represent the actual entrance pathway. The comparison of the dynamics of DABO with crystal structure of structurally similar compounds has provide information on possible molecular mechanisms of ligand binding, specificity and regulation of RT. The present calculations provide a validation of the combination of docking, MD as a powerful tool in structure-based drug design, and the methodology is easily scalable for attaining a higher throughput of compounds. This coherent picture strongly suggests that attempts to understand through the structure-based drug design may be considerably more successful if dynamic structural aspects of the type studied here are considered,particularly in those region which is more flexible but functionally important.

Epidemiology

Projected as the third leading cause of death, by the year 2030, Human immunodeficiency virus reverse transcriptase(RT) inhibition has become a major cause of concern worldwide .[1] One in 155 adults and adolescents are infected with human immunodeficiency virus (HIV) globally, with an astonishing rate of one in 21 in Sub-Saharan Africa.[2] In the USA alone, more than half a million people were living with HIV in 2007, and 44,000 new cases were reported in the same year.[3] major focus for the development of a range of pharmaceutical compounds that interfere with the various stages of HIV-1 retro viral life cycle. The gravity of this infectious disease has led to the development and approval of a range of pharmaceutical compound that interfere with various stages of the HIV-1 retro viral life cycle.

Drug combination therapy, commonly known as highly-active antiretroviral therapy (HAART), has for some 15yrs have been considered as the standard treatment for patients with HIV infection, whether antiretroviral drug-naive or drug-experianced.[4] .Thus far, the anti-HIV drug armamentarrium have grown to 25 licensed drugs [5,6,7], each targeting one of the three key events in the HIV Replication cycle; viral entry, involving the HIV viral membrane fusion protein gp41; the multistep conversion of single stranded viral genome RNA genome into double stranded DNA, catalyzed by reverse transcriptaseRT; and the cleavage of viral precursor proteins by the vital encoded protease.

Currently the most attractive target for development of antiretroviral compounds is HIV RT, with nearly 11 of the approved anti-HIV targeting (Pauwels R. et al 2004; current opinion in pharmacology)[8] it, and encompass two classes of agents that inhibit RT by two different mechanism; nucleoside RT inhibitors (NRTIs), which mimic the endogenous substrates and bind competitively at a catalytic site, and non-nucleoside RT inhibitors (NNRTIs), which are a diverse group of compounds that bind to a allosteric cavity which is created upon formation of the RT-inhibitor complex, thereby inducing a conformational change that inhibits the function of RT.[9,10]

Over the course of 25 years that followed after this seminal discovery, seven nucleosides have been approved by the United States Food and Drug Administration(FDA) for the treatment of HIV infection starting with the approval of AZT in 1987 and followed by didanosine (ddI), zalcitabine (ddC), stavudine (d4T), lamivudine (3TC), abacavir (ABC), tenofovir disoproxil fumarate [TDF; prodrug for the oral delivery of the nucleotide analog tenofovir (TFV)] and, most recently in 2003, emtricitabine (FTC). Despite of the approval of NRTIs as effective therapies for HIV-1/AIDS, side-effects of these antiretrovirals were found to be real and certainly not to be discounted. Despite of the fact that NRTIs generally (but not always) act with greater specificity for the HIV-1 RT, compared to mammalian DNA polymerases [11], there is a separate enzyme (polymerase-gamma) inside the cell that replicates mitochondrial DNA which can be affected by NRTIs. NRTIs can deplete or impair the function of this enzyme under certain circumstances. Another drawback of NRTIs is that they require three phosphorylation steps, catalyzed by cellular kinases, to be converted into active triphosphate metabolites. The active metabolites then act as competitive inhibitors or alternative substrates with respect to the normal substrates (either dATP, dGTP, dCTP or dTTP) and lead to the termination of chain elongation . Thus, their activation and efficacy depend on the metabolic state of the infected host cells. Given the properties of existing drugs, for new NRTIs , it is become increasingly difficult to comply with the demands for higher activity (potency), lower toxicity (side effects) and the more favorable resistance profile required for approval as antiviral drugs. These drawbacks of NRTI are not present in NNRTIs and hence the latter are being used as powerful weapons to combat HIV.

NNRTI form a group of chemically diverse compounds that specifically inhibit HIV-1 RT by targeting a non-substrate binding site on the enzyme, termed NNRTI-BP. Structural, computational and biochemical studies have demonstrated that NNRTI binding to HIV-1 RT induces short-range and long range conformational changes, in enzyme structure. Whereas the ddN analogues (i.e., AZT, ddI, etc.), following their intracellular phosphorylation to the triphosphate form, interact with the substrate binding site of the HIV reverse transcriptase (RT), the NNRTIs block the HIV-1 RT reaction through interaction with an allosterically located, non-substrate binding site [12, 13, 14 15]. This NNRTI-binding pocket is located at a close distance from the substrate binding site[16]. It is not only spatially but also functionally [16,17] associated with the substrate binding site. The cooperative interaction between these two sites [18] provide a rationale to increase the effectiveness of nucleoside RT inhibitors and non-nucleoside RT inhibitors by using them in combination therapy.

NNRTI can be further classified into 'FIRST generation' and 'SECOND generation' compounds. the First generation compound included nevirapine[17,18], efavirenz[19,20] and delaviridine[21]. Despite of them being considered as cornerstones for HIV therapy because of their full potential as a component of HAART, they succumbed to development of drug resistance [8]. In fact, the major challenge faced for the success of first generation compounds are the high-mutability and quasi species characteristics of HIV, which could be attributed to the RT's lack of proof-reading of nucleotide sequences during DNA synthesis and HIV's capacity for genetic recombination [22,23]. All of these properties allow HIV to evade immune surveillance and to select and evolve drug escape viral/RT mutants that can replicate in the presence of drug(s) (i.e. they are drug resistant). The 'second generation ' refers to compounds with improved activity spectrum against the most common NNRTI-resistant HIV variants which were developed based on approach that involved the the concurrent study of several different parameters, as opposed to focusing on potency against wild-type virus, approach including: searching for new and/or further evolved chemical scaffolds that allow more flexible and multiple drug-binding modes; better 'quality' binding through aiming for multiple drug interactions, in particular, with more conservative and/or critical residues; and multiple interactions with the main chain of the amino acids that line-up the NNRTI binding site. The underlying assumption was that these properties would allow for functional binding to the NNRTI pocket, even in the presence of mutated amino acids. These 'Second' generation included emivirine (formerly known as MKC-442) [24], HBY097 [25], Calanolide A[26], the phenylethylthiazolylthiourea (PETT) derivative MIV-150 [27] and the quinazolinone analogues DPC083 and DPC961 [28].To date, four NNRTI have been approved for clinical use by FDA: nevirapin, delaviridin, efavirenz, and most advanced etravirine.

DABO(dihydro-alkoxy-benzyl-oxypyrimidines) derivatives are new promising NNRTI drug candidates which are present in the earlier stage of preclinical trials[29] and is the focus of this work. NNRTI such as DABO and its derivatives should be envisaged as potential microbicide to prevent mucosal HIV transmission: under the experimental conditions where the NRTIs only delayed viral breakthrough, the NNRTI such as MKC-442, DABOs(MC1220), were able to suppress HIV-1 replication for an entire period of 40 days[30].

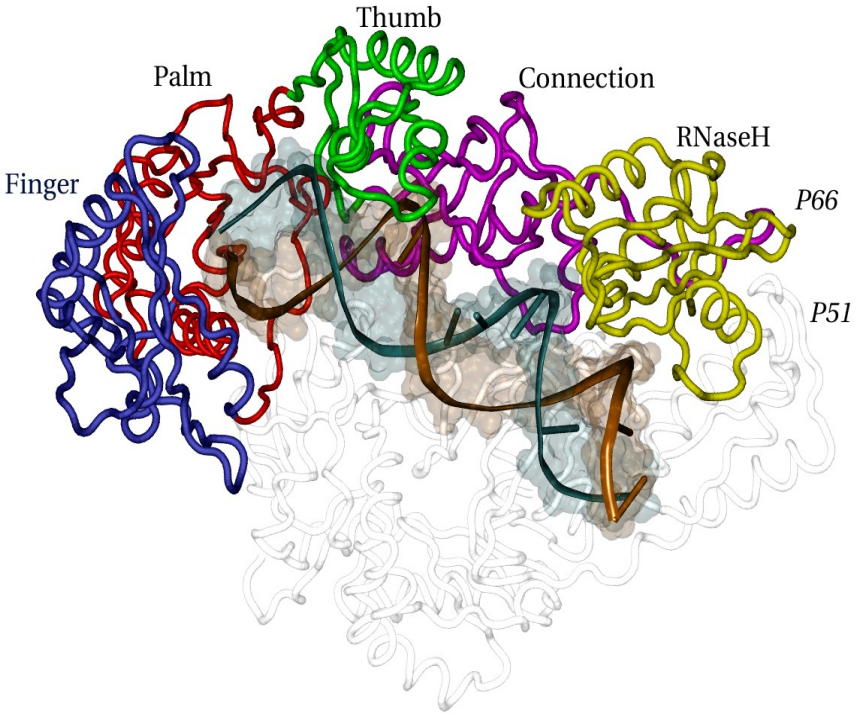


Figure1: Cartoon representation of HIV-RT. Two subunits P66(colored according to the subdomains) and P51(white) are depicted here. Finger is represented in blue, Palm in red, Thumb in green, the connection in magenta and RNaseH in yellow. The template is also represented here with the two strands being colored in orange and cyan.

Several studies have revealed a common mode of binding for chemically diverse NNRTIs with their target site at HIV-1RT. It has been suggested that the NNRTIs inhibit HIV1-RT by locking the active catalytic site in an inactive conformation, reminiscent of the conformation observed in the inactive P51 subunit.Our understanding of the molecular basis of RT function asymptotically increased after x-ray elucidation of RT structures in multiple forms, Apo (without any ligand), Holo (complexed with NNRTI[31,32]) and other complexes (nucleic acid template primers,[33,34]) , have provided an in depth understanding, and explores variability of RT at atomic level. RT is a heterodimer consisting of two two subunits of 66kDa(p66) and 51 kDa(p51).[35-40] The p66 and p51 subunits consist of four subdomains called thumb , palm fingers and connection but the organization of these domains are different in p66 and p51 subunit[Fig1]. The p66 subunit also contains the RNase H subdomain. The p66 finger,palm and thumb subdomain domains forms a large cleft that binds the template-primer DNA.[41] NNRTI bind to a common hydrophobic site, the non-nucleoside inhibitor binding pocket (NNIBP), located in the p66 palm subdomain approximately 10A away from the polymerase active site [42-46]. As a result of NNRTI binding, certain RT domains actively participate in flexibility and mobility which, in turn, leads to a dramatic reduction in catalytic enzyme efficiency.[47-49] Knowledge of important structural features (periodic movement of thumb and finger domain, primer grip, catalytic triad,) and "**shrink-wrap**" [50] nature of NNRTI binding cavity have made computational structure based drug design as a promising approach to identify and optimize structure.

In this work, extensive docking, scoring and molecular dynamics simulations have been combined to predict the binding modes as well as to identify key binding interactions of four compounds belonging to the DABO series of NNRTI. Allosteric site residues which were involved in the binding with the compounds were characterize with an aim to further optimize structure as well as to provide platform for structure based drug designing efforts. The protein residues that are responsible for the difference in binding affinities between the inhibitors are also identified and these results are compared to experimental data on common mutants of RT. In a nutshell, comprehensive detail on the binding pattern of the compounds belonging to the DABO compounds, conformational changes taking place in protein on inhibitor binding as well relative binding free energies have been elaborated in this work.

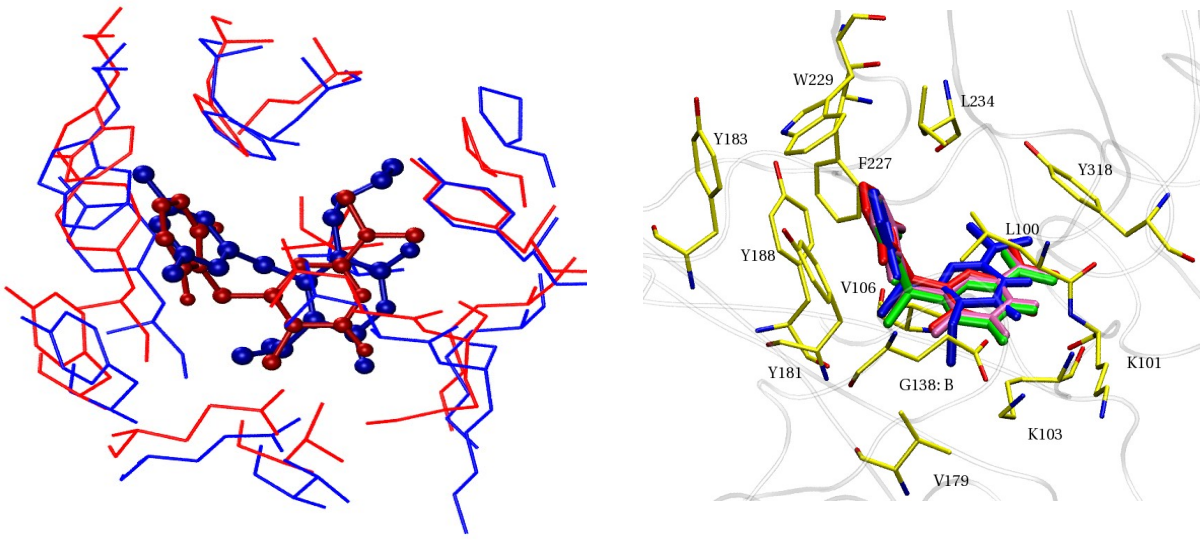
Results and discussion:

Recently many compounds belongs to DABO derivative were screened in cell-based assays and enzymatic assays against HIV-1 RT. Some of the analogs of our lead compounds MC1220[30], came out with very high potency viz MC1332, MC1195 and MC1346. We performed i*n-Silico* studies to explore the interaction of the DABO compounds with HIV-RT at microscopic level. Extensive Docking (Look M&M) was performed to identify biologically relevant binding pose. The poses were then refined using Molecular Dynamics protocol. All the results obtained are discussed in detail in the following subsection.

Here, it is worth mentioning that MC1220 is considered as the lead compound due to its 'knock out property' [ref..2010] which is lacking in other DABO analogs. Herein,1) we have tried to come up important residues playing key role in stabilizing the DABOs in the binding pocket, 2) we have tried to study the effect on protein dynamics on binding of the DABOs in the cavity, trying to throw some light on the mechanism of inhibition,3) we have also tried to compare the binding modes of DABO with those of NNI belonging to other classes, but a study investigating the knock out property of MC1220 is beyond the scope of this chapter and has been kept for future studies.

1)**Docking**

Docking has been done on the way as mentioned in the material and method section. HEPT[5,6](1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine) compound bound crystal structure 1RT1[51] was used for all the docking runs. The allosteric NNRTI binding pocket (NNIBP) is located approximately 10 Å away from the p66 polymerase active site and is not present in either the APO or the substrate-bound crystal structures. In these cases, the NNIBP is occluded, predominantly by the aromatic side chains of Tyr181 and Tyr188, which must undergo large torsional rotations in order to swing out of the pocket to accommodate NNRTIs [52]. Thus our reason of talking 1RT1 was to start with a protein structure with an existing allosteric binding site. For the docking runs the compound (HEPT) was removed from the binding cavity and then a relaxation of the protein was done by performing 5ns MD simulation. The relaxed structure obtained from Molecular dynamics was used for the docking.



A **B**

Figure2 : **A)** Docked conformation of DABO MC1220 superimposed on the Crystallographic structure of HEPT compound(1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine).(1RT1[51]) MC1220(depicted as red ball and stick) binds into an orientation similar to that of HEPT(represented as blue ball and stick) compounds. **B)** Docked structures of the DABO compound into the NNRTI binding cavity. The orientation acquired by the four DABO compounds were nearly identical. MC1220 is depicted in mauve, MC1332 in blue, MC1195 in red and MC1346 in green color.

The orientation as well as some of the key interactions obtained after docking the four DABO were found to be closely resembling that of the HEPT bound HIV-RT.[Fig.2A] A detailed description on the similarities and dissimilarities between the binding pattern of the DABO with the HEPT and other NNI's belonging to different classes are discussed later in this chapter.

As expected all the four DABO's docked into the binding site in nearly identical orientation. A detailed pictorial representation of the chosen binding mode and the key residues involved in docking for all four DABO compounds are present in Fig.1B. The RMSD of the MC1332, MC1195 and MC1345 with respect to the lead compound(MC1220) came out to be 0.6, 0.8.0.7 respectively, indicating similar binding modes for all the DABOs.

2)MD Simulation

2.1)Interactions

The starting structures obtained from docking where then used for further MD simulation runs for 20ns. DABO compounds accommodated themselves in the hydrophobic cavity present in the palm domain, sandwiched between two β sheets structures. These β-sheets house the primer grip (residues Met230 and Gly231) and catalytic aspartate triad (residues Asp110, Asp185, and Asp186). Owing to nearly identical binding modes of all the four DABO compounds under study, the Ring1 (pyrimidine ring) is surrounded by residues Val179, Lys103, Lys101 of p66 subunit and Glu138 of P51 subunit, likewise the Ring2 is surrounded by Tyr181, Tyr188, Trp229, Phe227, Leu234 and Pro95, while the di-methyl amine interacts with Val106, Pro236 and Tyr318 in nearly all the four DABO compounds.

RT residues are Important for Binding of DABO Compounds:

To improve rationally the efficacy of non-nucleoside compounds it is important to have a precise and detailed understanding of the important protein-inhibitor interaction. In an effort to dissect these interaction from the MD simulation, total interaction energy between DABO compounds and the binding pocket was calculated as described in material and method. Fig 4b) summarizes the favorable hydrophobic as well electrostatics interactions between MC1220, MC1346, MC1332 and MC1195 and the binding pocket respectively which were analyzed to identify residues that are important for binding.

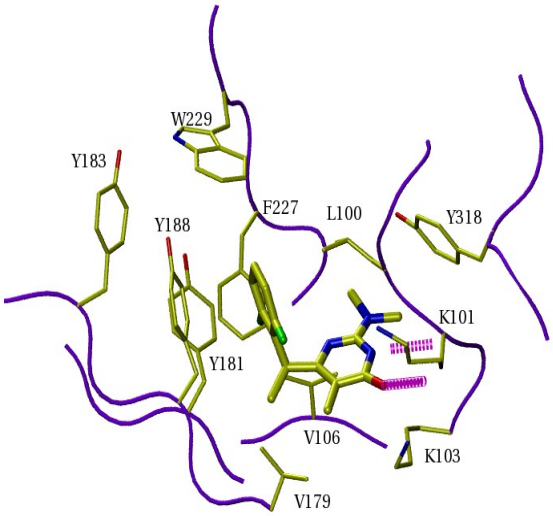
As predicted, major contribution to the stability of compounds in the pocket are provided by the HpH interactions. Leu101 makes the largest contribution to the ligand-residue VdW energies with value greater than -5kcal/mol in all the four DABO compounds. Similar consistent contribution are provided by residue Tyr188 and Tyr181 to the stability to all the four DABO compounds in the binding cavity. Tyr188 and Tyr181 makes π-π stacking against the wing2 throughout the simulation thereby locking the inhibitor into the binding cavity. Other residues which make favorable VdW interactions in MC1220, MC1332 and MC1195 include Lys103, Val106, Trp229, Leu234, Trp328 and Pro236. This shows that binding is clearly favired by VdW interactions with these residues.

Despite of the cavity being predominantly HpH in nature the NNI pocket provides ample possibilities for polar interactions as well. Here again, remarkable contribution is provided by Leu101 with electrostatic energy -16kcal/mol, -9kcal/mol, -18kcal/mol and -19kcal/mol in MC1220, MC1332, MC1346 and MC1195 compounds respectively. Additionally in all the four DABO compounds , two strong HB are formed between the O1(DABO) with the HN of Lys101 and N1(DABO) with O of Lys101(2.3, 2.0 and 1.9 A respectively). these HB persist throughout the 20 ns simulation with occupancy greater than 90% in all the four compounds. Other residue which contribute favorably in all the four DABO compounds is Lys103 with energy -8 kcal/mol, -10kcal/mol, -11kcal/mol and -3.5kcal/mol in MC1220, MC1332, MC1346 and MC1195 respectively. The residues which make largest VdW and electrostatic contribution to the binding of the compounds are depicted in Fig 3.

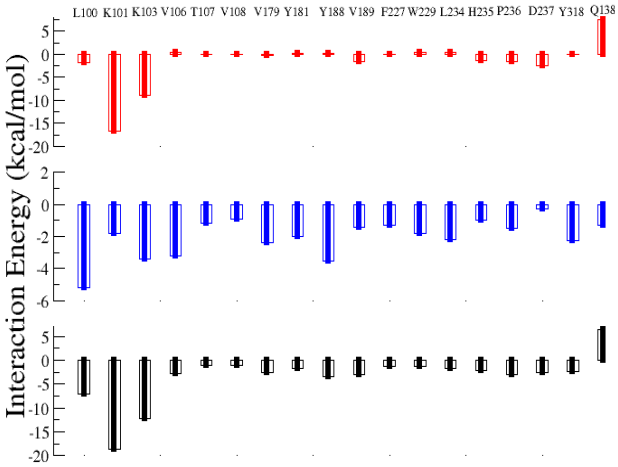
These results agree very well with the experiments carried out on the most prevalent resistant mutations for NNRTI . It has been increasingly clear that different RT mutation L100I + K103N/D, K103N+Y181C are required for engendering high level resistance to NNI . Such double mutant particularly K103N and Y181C mutant have been shown to raise in vivo in patients under NNRTU treatment. K103N mutation was most often the first, and

also the most common NNRTI resistant mutation observed in patient failing on NNRTI containing regimens; also frequently noticed the double K103N + L100I mutation. The K101E/C and L100I mutations were almost exclusively found as double mutants in combination with K103N.(123, 124). The data on prominent resistant mutation relate well with the residues contributing the most in terms of electrostatic and Van Der Waals energies identified by our simulations. Lys103, Lys101 and Leu100 were here found to have either highly favorable electrostatic or VdW interaction with the inhibitors, indicating their importance for the tight binding and inhibition. For instance, mutation K101E/C may result into the loss of HB which results into loss of binding affinity.

MC1220

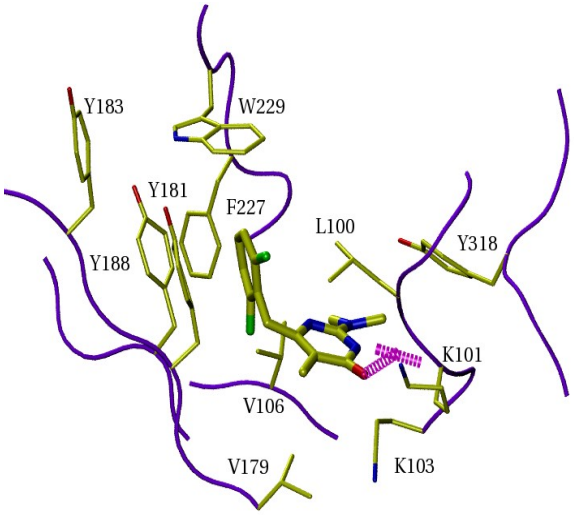


A)

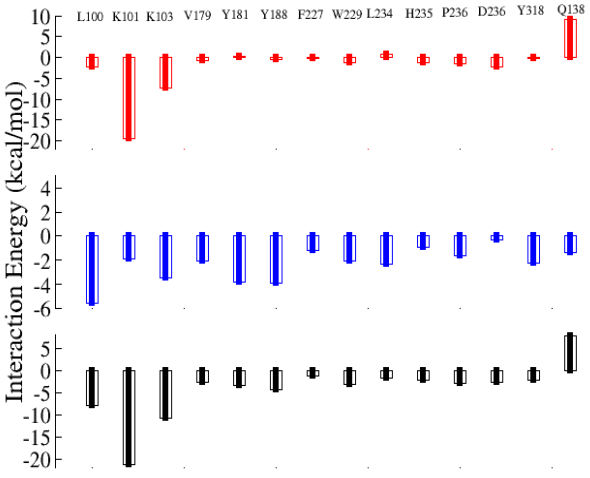


B)

1195



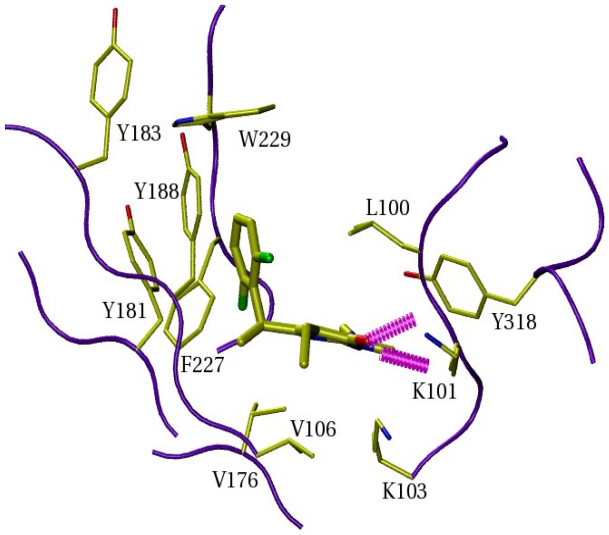
A)



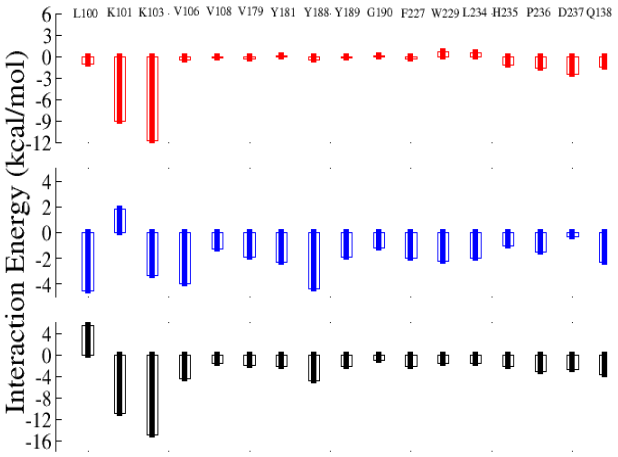
B)

Figure3: A) Pictorial representation of ligand-residue interaction spectrum of key residues of NNRTI binding site with DABO compounds. The residues are represented in stick while the HB are depicted in magenta color spirals. B) Ligand-residue interaction energies in kcal/mol for the residues that contribute most to the ligand-surroundings. The red color bar denote the electrostatic contribution, Blue correspond to the Van Der Waals energies and Black denotes the total interaction energies.

1332

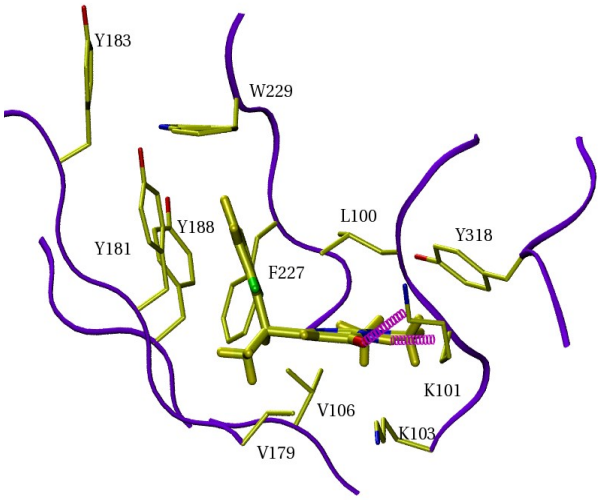


A)

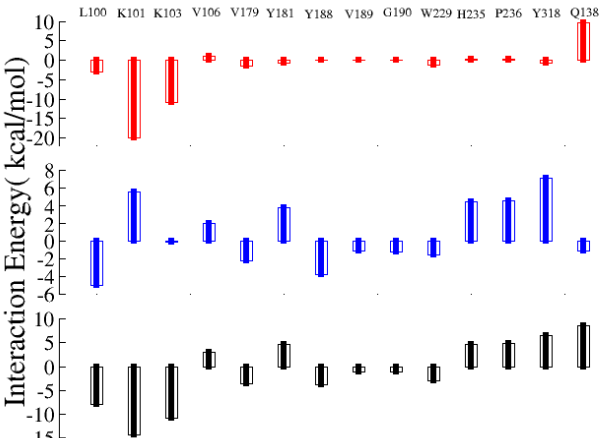


B)

1345



A)



B)

Figure4: A) Pictorial representation of ligand-residue interaction spectrum of key residues of NNRTI binding site with DABO compounds. The residues are represented in stick while the HB are depicted in magenta color spirals. B) Ligand-residue interaction energies in kcal/mol for the residues that contribute most to the ligand-surroundings. The red color bar denote the electrostatic contribution, Blue correspond to the Van Der Waals energies and Black denotes the total interaction energies.

1.2)Comparison with crystal.

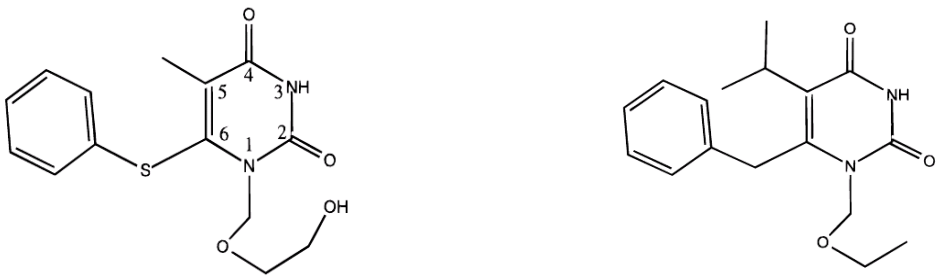
we did a comparative analysis of our DABO compounds with previously reported NNRTI bound crystal structures of structurally related HEPT[53,54] and derivatives of 6-Benzyl-1-(ethoxymethyl)-5-isopropyluracil, MKC-442[55,57], in order to characterize similarities and differences in binding of these related classes of NNIs . This was done with an aim to pick up ideas for further improvement of the potency of DABO compounds. For the comparison of the three related classes of NNIs we used crystal structure 1C1B for HEPT and 1RT1 for MKC-442. As for DABO, an average structure from the 20 ns simulations of MC1220 was used.

The structures of NNRTI, discussed here,revealed that the loops comprising of residues Pro225 and Pro236, flex to optimize contacts with the different compounds when compared with Apo protein While, with HEPT the loops occupy positions similar to those of the unliganded enzyme.[56] but in the presence of MKC-442[55], the loops move, contracting the binding pocket and tracking the shorter ethoxymethyl tail. In accordance with the behavior exhibited by HEPT, we observed that DABO compounds exhibited slight change in the orientation of loops harboring residue Pro225 and Pro236. In these compounds more substantial contractions around the inhibitors occur.

In the case of these more compact inhibitors the conformation is stabilized by a hydrogen bond from the carbonyl oxygen of residue Pro236 to the main chain amide NH of residue Lys103 with ADL(2.3A). The same intra HB between Lys103 and Pro236 with an average ADL of 2.5 A was observed in MC1220 and its analogs.

Earlier literature study indicated that the interaction with Tyr181 and the conformational change associated with it was an integral part of NNI binding. On the contrary, in case of HEPT and MKC-442 the interaction with Tyr 181 was not found to be the key interaction. Likewise, with the DABO series of NNIs the trigger for the conformational switch for Lys101, Leu100 and Tyr181 is the bulk of the 5-substituent(X position as shown in table 1): a methyl group is not large enough to force Tyr181 to move indicating that this interaction does not always exist in the new generation NNIs.

The hydrophobic nature of the NNI pocket provides relatively few possibilities for polar interactions and hydrogen bonding between the NNIs and the protein. However, the geometry of interaction for HEPT analogues appears to be constrained by a strong hydrogen bond from the 3-NH of the pyrimidine ring to the carbonyl oxygen of Lys101 (3.1 Å for HEPT and 2.8 Å for the tight binders MKC-442 while for MC1220 it is strongest as 2.6A. Since there is a slight difference in the orientation of the pyrimidine ring between MKC-442 ,maintenance of this H-bond necessitates a small movement of the polypeptide chain in the vicinity of residue Lys101. In addition, for the MKC-442 complexes, there is a water molecule in the NNI-binding pocket situated near a channel to the bulk solvent, which forms a triad of hydrogen bonds between the 4-carbonyl oxygen of the inhibitor, the main chain nitrogen of Lys101, and a carboxyl oxygen of Glu138 in the p51 chain. In the RT-HEPT complex, Tyr181 and Glu138 of the p51 chain, are in conformations which could still allow a water molecule to bind in this space although in this case hydrogen bonds would be formed with the hydroxyl oxygen of Tyr181 and the side chain of Glu138 . There are many other intra HB formed to make the cavity intact and positioned the key residues into a proper position as they can make proper interaction. These mainly intra HB formed between Leu227-Phe234, Thr107-Val189, Lys103-Pro236 and His235-Tyr318 residues. In comparison with DABO analogs we observed the same stability and architecture of amino acids as described above. As mentioned earlier the binding cavity is completely hydrophobic, except the region which is suppose to be known for providing the evidence for "***entry mode***" of NNRTIs [58]are near in the region of 225-226 and 105-Val106, and the region close to Lys101 and Lys103 along with Glu138 of p51 units, which was reported to explained as "***expulsion site***" of the NNRTIs from the binding pocket. We also observed water molecule on the surface of above mentioned region and also involved in providing stability to NNRTIs as making direct HB with 4-carbonyl oxygen of inhibitor and water mediated interaction with K103.

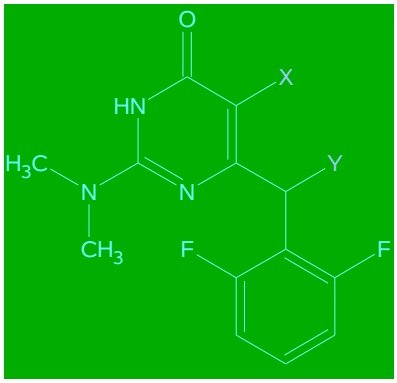


A **B**

Figure5: The chemical structure of the two Rt Inhibitors studied, A)-HEPT, B)-MKC-442. Images taken from Hopkins A.L. Et al 1996; J.Me. Chem.

The 5-6 Position Trigger:

Structure-activity relationship (SAR) studies show that a 5-position substituent on the pyrimidine ring is essential for DABO analogues to bind to RT. As previously reported in HEPT compounds the same trend we observed in our DABO compounds. In HEPT compounds, replacement of the 5-methyl (position X in table1) group of HEPT with a hydrogen atom leads to a complete loss of inhibitory properties,[59,60] while its replacement by a larger moiety, such as an ethyl or an isopropyl group (e.g. in MKC-442), produces much more potent inhibitors.[61-63] Comparison of the HEPT analogue complexes with the structure of MC1220 the drug-free RT(1RT1) shows that it is the 5-group that perturbs the conformation of Tyr181. The 5-methyl group of MC1220 produces only a slight perturbation, whereas the larger substitutes of MC1332 would clash properly with Tyr181 and Tyr188 in the unliganded RT conformations. Thus, the 5-methyl or ethyl group appears to force Tyr181, Tyr188 into the more open orientation, where it is then able to form strong interactions with the 6-phenyl ring of the inhibitor and other aromatic residues. The other residues which are affected, and change its position in presence of inhibitor was L100, although it still manages to shows very strong HpH interaction with position X in case of complex. Further evidence for a “trigger” role for these groups came into picture while speculating the position 6 (position Y in table 1). The pyridine C6 substituents is so important as it can make difference in MKC-442, GCA-186 and MC1220. the C6 substituents are positioned in the upper subpocket of the NNRTI binding site, which is formed between Leu100, Tyr181, Tyr188, Phe227 and Trp229. This region has methyl group in other analogs of MC1220, except MC1195 where it substituted with hydrogen. we observed a 10 fold loss of activity in MC1195 in comparison with other analogs of DABO's. While comparing with unliganded structure, we observed the Y position, hampered the position of W229, which relocate itself into new position, although having HpH interaction with inhibitor, contributing in the stability of DABO's into the binding pocket. Combined with a slight alteration in the orientation of the pyrimidine ring, the net effect is to place the methyl or ethyl groups in exactly equivalent places in the binding pocket, in each case preserving the trigger action. We have previously noted that all potent NNIs we have studied possess groups which occupy the same volume of space as the 5-substituent:25 the cyclopropyl group of nevirapine, the ethyl group of 1051U91, the 5(S)-methyl group of 9-Cl-TIBO, and the amide group of R-APA and 5-isopropyl groups of MKC-442 perform the same steric role as the 5- methyl or ethyl groups of DABO's.



A

Table:1 Experimental results DABOs

Compd	X	Y	[a]EC_{50}	[b,c]Long term activity
DABO 1332	CH_2CH_3	CH_3	0.01	no (>100 µM)
DABO 1195	CH_3	H	0.09	no (>100 µM)
DABO 1346	H	CH_3	0.00	yes (100 µM)
DABO 1220	CH_3	CH_3	0.004	yes (11 µM)

[a]Compound concentration (µM) required to achieve 50% protection of MT-4 cells from HIV-1-induced cytopathogenicity as determined by the MTT method.
[b]Compound capability to irreversibly inhibit virus multiplication in MT- 4 cells infected with a high multiplicity of infection (5 $CCID_{50}$/cell) , up to day 40 post infection.
[c]Compound present only during infection.

A)-The basic backbome of DABO compounds.

Protein Flexibility

2.2)**RMSF**

A quantitative analysis of the flexibility of the different sub-domains of RT is performed through the root-mean-square fluctuation (RMSF) of Cα atoms with respect to their time-averaged positions(Fig-4). Plotting the RMSF as a function of residue number reveals greatest movements from the fingers and thumb sub-domains of the p66 subunit. A smaller degree of flexibility is seen at the RNAse helicase (RNH) domain, while the p51 subunit generally shows much lower values than p66, consistent with its structural role in the heterodimer. Comparing the RMSFs of the different simulations reveals very similar profiles. However, a pronounced local effect of NNRTI binding is observed at the tip of the p66 thumb subdomain (indicated by the dotted lines. The APO simulations show a marked increase in thumb flexibility compared to the DABO simulations . It is clear that the outermost region of the p66 thumb domain is more mobile in the absence of DABO, specifically making stronger moves toward the fingers. This indicates that presence of D

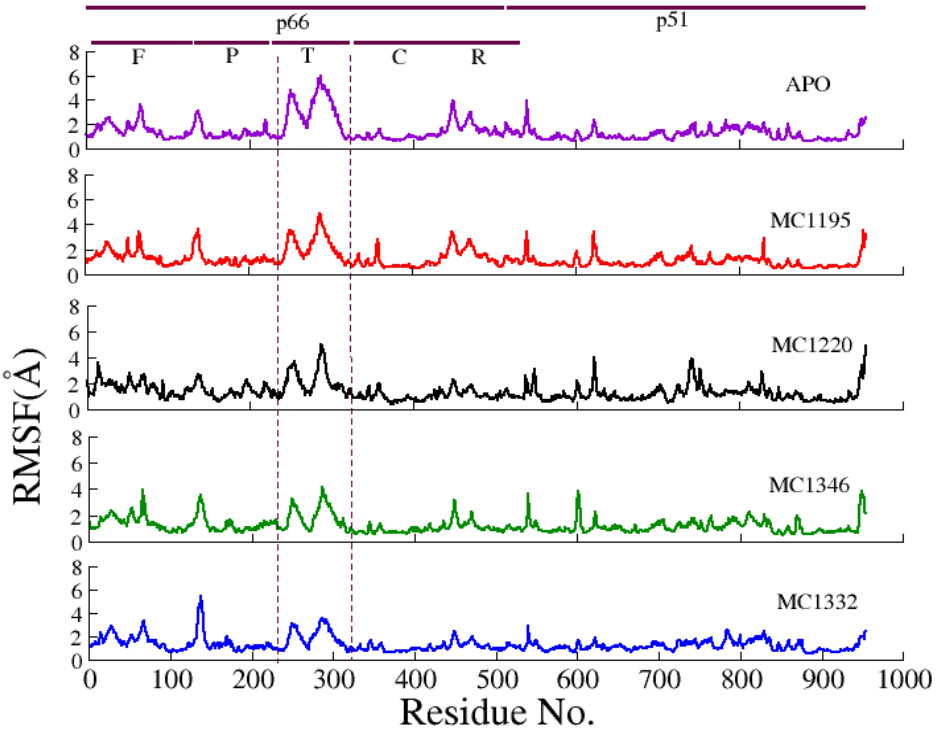


Fig.5: Conformational Flexibilty. RMSF of all C^{α} atoms from their time-averaged positions. Each subunit is indicated by black lines, along with the subdomains of the p66 subunit (F, fingers; P, palm; T, thumb; C, connection; R, RNH).

2.3) Mechanism of Inhibition.

Insightful comparative observations of crystallographic snapshots of "RT" have suggested three possible mechanism of inhibition of HIV-1 RT by NNRTIs:1) restriction of P66 thumb domain flexibility '**molecular arthritis**' 2) distortion of catalytically essential residues at the polymerase active site 3) displacement of primer grip. In the light of existing experimental findings we tried to detect the DABO induced changes in the native dynamics of the RT and thus clarify the inhibition mechanism of DABO compounds with special focus on MC1220.

The thumb, palm and the finger domain of the P66 subunit act as a clamp for the positioning of the double stranded template primer in position. In the presence of a nucleic acid substrate in the cleft ,widening of the fingers–thumb distance occurs which results into a more open cleft, analogous to the "**Venus flytrap**" models of substrate binding/recognition[64], which , on contrary, in the apo form the cleft remain occluded. In order to have a comparative understanding of the effect of the binding of MC1220 on the dimension of the nucleic acid clamp we sampled the minimum distances between the residues of the thumb domain and finger domain of all the four DABO simulations. We also sampled the distances between the thumb and finger domain of crystal structures 1DLO(without substrate) and 1RTD (with substrate). Fig.5 shows the distribution of the distances sampled by the MC1220, MC1195, MC1332 and MC1346 were centered around 47A, 44 A, 48 A and 46 A respectively, while on the contrary the minimum distances obtained in case of 1DLO and 1RTD came out to be 30 and 35 A respectively. This comparative analysis clearly indicated that while in the absence of the substrate the structure is in closed form, in the presence of the substrate it opens up to accommodate the nucleic acid. A further widening of the finger-thumb distance is exhibited in the DABO bound structures throwing it into an 'hyper-open' form of the empty cleft. More recently, fluorescence experiments[65] have suggested that NNRTI-induced loosening of the fingers–thumb grip , or in other words widening of the finger-thumb domain gap, as is observed in the binding of the DABO compounds, may cause RT to slide away from the 3′ primer end and thus prevent polymerization.30MAccOMON. A visual depiction of the movement of the thumb and finger domains of the DABO bound structure with respect to the 1DLO and 1RTD is shown in (fig. 5).

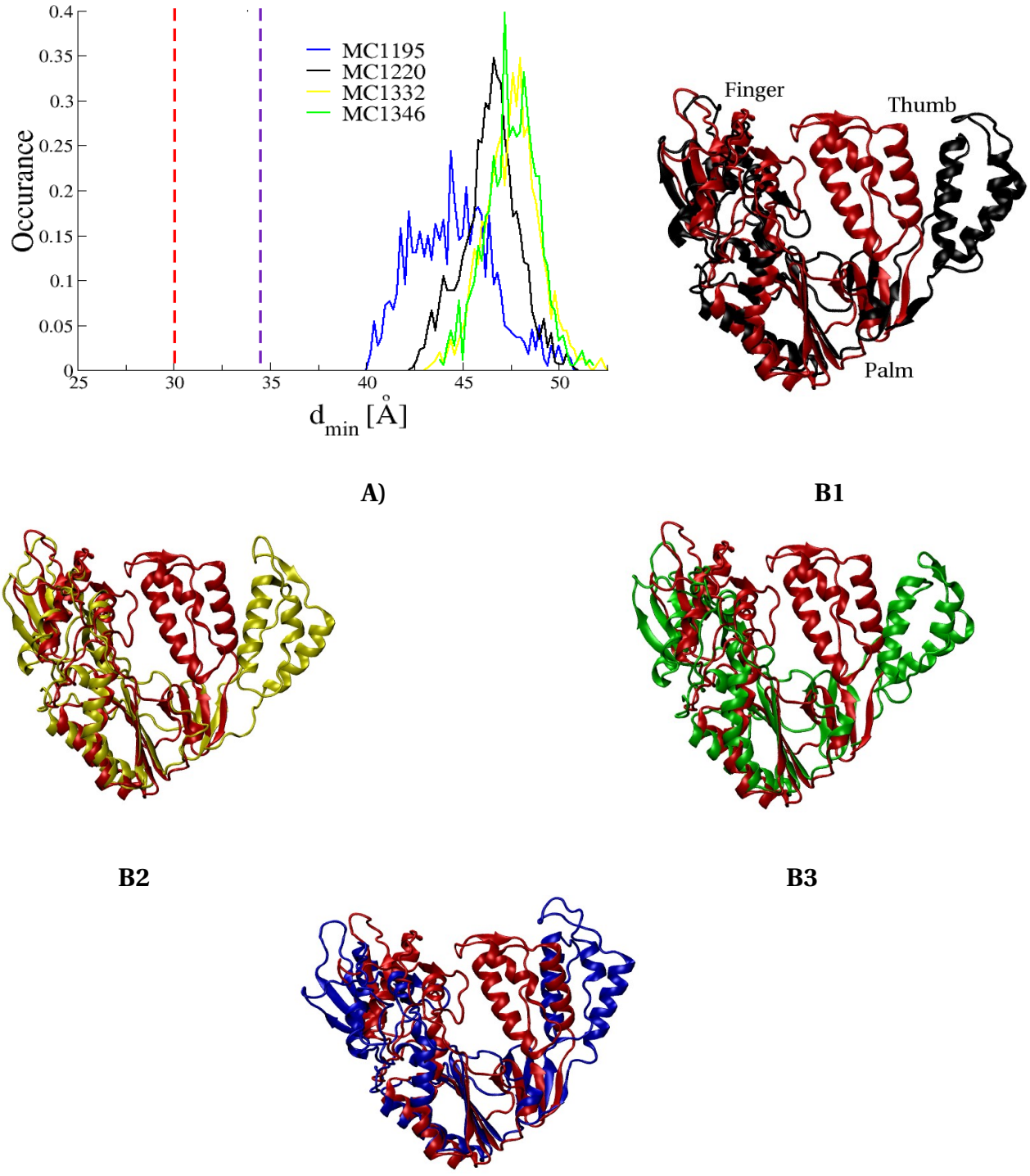


Fig.5: Minimum distance between the thumb and finger subdomains along the trajectory, A)Normalized histogram calculated for each system. The red and purple dotted line indicate the distance between the thumb and finger domain in 1DLO(without substrate) and 1RTD (with substrate) respectively. B) Cartoon representation of the polymerase region, illustrating the distance calculated. Here 1DLO, represented in red, is superimposed with MC1220 bound structure, in black, 1332-bound structure, in yellow, 1195-bound structure, in blue, and 1346 bound structure in green color.

Additionally, previous biochemical evidence have indicated that NNRTIs do not affect the ability of RT to form a ternary complex with nucleic acid and nucleoside triphosphate, but specifically inhibit the chemical step of nucleotide incorporation into the primer strand.[66,67]. This prompted us to deeply explore the possibility of probable distortions of residues near to the polymerase active site which may effect the polymerization process. MC1220 accommodates itself in such a way that it is flanked on both side by two beta sheets. these beta sheets harbor two important conserved structural features of HIV rt which include the Primer grip region(Met230 and Gly231), guide the correct positioning of the 3'end of the primer and catalytic triad(YMDD motif) which is involved in addition of nucleotide to the growing primer chain. Here also, we did a comparative analysis of minimum distances obtained between the two betasheet in all the four DABO simulations as well as crystal structure 1DLO and 1RTD. Histogram plot(Fig .6A) clearly indicated a widening of distances between the betasheets in the DABO simulations which centered around 12.8 A, 12.2 A, 12.9 A, and 14.3 in MC1220, MC1346, MC1195 and MC1332 while in 1DLO around 8.2 A and in 1RTd around 10.9 A. IT has been proposed that the shift of the primer grip, which is very well evident in the DABO bound structure, may causes incorrect placement of the 3′ primer end relative to the catalytic site and, consequently, a catalytic ally incompetent constellation of these key protein and nucleic acid atoms.[5][9]

Hence in a nutshell, widening of distance between the thumb-finger domain leading to a 'hyper open' of HIV-RT along with the distortions produced in the residues at the active site may result into inhibition of DABO bound HIV-RT structures.

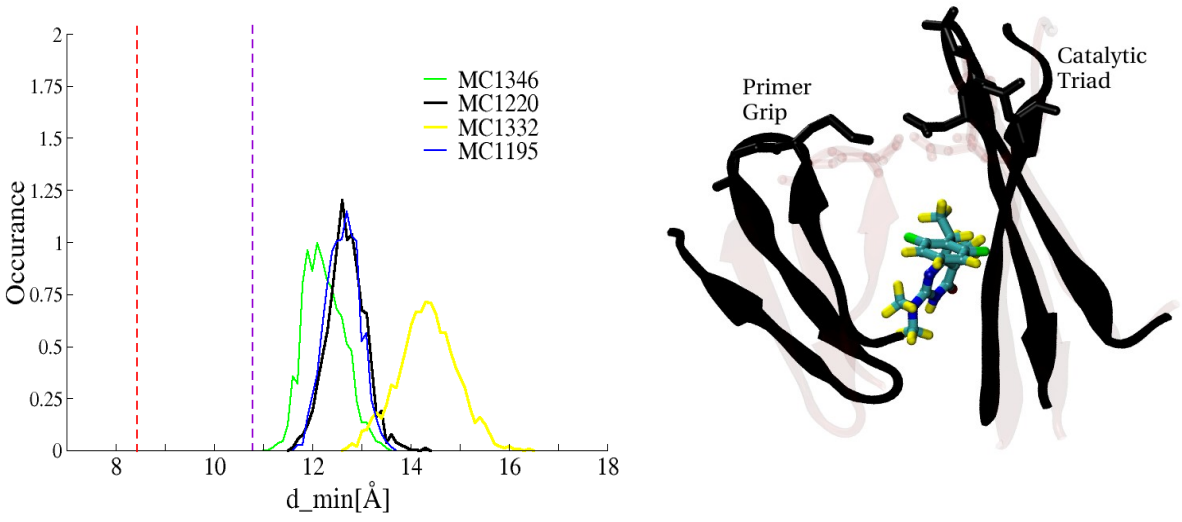


Fig.7: Minimum distances between two beta sheets of the NNRTI binding site which harbors two important conserved structural features of HIVRT, the Primer grip and catalytic triad(YMDD motif). A) Normalized histogram of the distances are repersented. The red and purple dotted line indicate the minimum distances between the beta sheets in 1DLO(without substrate) and 1RTD (with substrate) respectively. B) Illustration of the distances calculated. Superimposition of beta sheets of 1DLO(transparent pink) with the MC1220 bound structure(black)

Conclusion:

Many NNIs have been synthesized in the quest for new anti-AIDS therapies. DABO(dihydro-alkoxy-benzyl-oxypyrimidines) derivatives which are new promising NNRTI drug candidates, present in the earlier stage of preclinical trials(49), and are envisaged as potential microbiobcide , are the focus of this work.

Herein, we have performed insilico study, including docking and dynamics to understand the key interactions involved in Ligand-protein recognition process. We have performed all-atom MD simulations of the HIV-1 RT enzyme, in both the presence and absence of MC1220 and its analogs. The protein-drug inhibitor interaction were analyzed in detail. We were able to identify key residues crucial for binding of DABO compounds into the binding cavity. The result of this analysis also agree very well on the experiments on mutant of RT, which show that identifying the principal interaction for a class of inhibitor can give an important direction for further optimization of affinity and for the design of inhibitors that are less susceptible to the mutations.

The principal aim of this study was to identify how the binding of MC1220 at an allosteric site affects the conformational dynamics of RT such that it inhibits DNA polymerization. Focusing on the catalytic p66 subunit, we did a computed minimum distance between thumb and finger domain. The distribution of the distances sampled between the thumb and the finger domain revealed that while in the absence of the substrate(1DLO) the structure is in closed form, in the presence of the substrate it opens up to accommodate the nucleic acid(1RTD). A further widening of the finger-thumb distance is exhibited in the DABO bound structures throwing it into a 'hyper-open' form of the empty cleft. This results into a mechanical obstruction in the movement of thumb and finger domains required for proper positioning of the incoming DNA into the NNRTI- bound HIV-RT. We finally sought to link the mechanical obstruction to local conformational changes taking place in the catalytic site leading to impairment of polymerisation. We found that the hinge motion influences the geometry and positioning of two key regions of the active site, which lie on β-sheets on opposite sides of the NNIBP – the primer grip and the catalytic triad. Minimum distance analysis between the two beta sheets indicated a widening of distances between the betasheets in the DABO simulations which centered around. This widening in DABO bound structure, may causes incorrect placement of the 3′ primer end relative to the catalytic site and, consequently, a catalytically incompetent constellation of these key protein and nucleic acid atoms. Thus, by our simulation we were able to hypothesize that widening of distance between the thumb-finger domain leads to a 'hyper open' of HIV-RT along with the distortions produced in the residues at the active site may result into inhibition of DABO bound HIV-RT structures.

Work in progress :

In this study, four independent trajectories were generated for each system. The conventional all-atom MD simulations of proteins are still restricted to the nanosecond timescale and are liable to becoming trapped in energy minima. Therefore, techniques have been developed to increase transitions across energy barriers and thus enhance conformational sampling, while retaining the granularity of all-atom structures.
To explore the residence time and the escape mechanism of MC1220 and its analogs “rare events” a well validated Metadynamics, is in progress for future computational studies of RT. This would allow us to efficiently study the effect of a range of NNRTIs on the opening/closing motion of the p66 polymerase region, in isolation from other degrees of freedom. In particular, it could be used to predict the potency of NNRTIs and design novel compounds which optimally block the hinge region. We also noted the potential of using metadynamics from previews work and also from some of our preliminary results, for the discovery of new RT hinge regions which may also be susceptible to occlusion. Work is currently underway to use MD simulation to predict the mechanical impact of compounds which bind in this region, with the aim of debilitating motions necessary for the RNase activity of RT. another aspect is in progress using metaD is to understand the main key residue involving in stability of drug, while pulling the drug from the binding site “escape mechanism” and quantify it in form of energy barrier. Finally, MD simulations such as those described here can be combined with virtual docking methods to suggest leads for new drugs .

Reference:

1. Mathers, C. D. & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 3, e442.
2. World Health Organization. World Health Statistics 2009: Cause of spe cific mortality and morbidity. Available at: http://www.who.int/whosis/ whostat/EN_WHS09_Table2.pdf. Accessed on Dec 6, 2009.
3. Wheeler W, Mahle K, Bodnar U, et al. Antiretroviral drug resistance mutations and subtypes in drug-naïve persons newly diagnosed with HIV-1 infection, United States, Mar 2003–Oct 2006. 14th Conference on Retroviruses and Opportunistic Infections (CROI 2007). 2007 Feb 25–28; Los Angeles, CA.
4. Lee LM, Karon JM, Selik R, Neal JJ, Fleming PL: Survival after AIDS diagnosis in adolescents and adults during the treatment era, United States, 1994-1997. JAMA 2001, 285:1308-1315.
5. Erik De Clercq, Antiretroviral drugs;current opinion on pharamacoloy, 2010.
6. De Clercq E: Anti-HIV drugs:25 compounds approved within 25 years after the discovery of HIV.Int Jantimicrob Agents 2009,33;307-320.
7. De Clercq E:The histroy of antiretrovirals;key discoveries over the past 25 years.Rev med Virol 2009, 19:287-299.
8. Pauwels, R. (2004). New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections. Curr. Opin. Pharmacol. 4, 437–446 .
9. De Clercq, E. (2002). New developments in anti HIV chemotherapy. Biochim. Biophys. Acta, 1587,258–275.
10. De Clercq, E. (2004). Non-nucleoside reverse transcriptase inhibitors (NNRTIs): past, present, and future. Chem. Biodivers. 1, 44–64.
11. Martin, J., Hitchcock, M.J., De Clercq, E., Prusoff, W., 2010. A brief history of the first generation nucleoside HIV reverse transcriptase inhibitors. Antiviral Res. 85, 34–38.
12. E. De Clercq, Med. Res. Rev. 1996, 16, 125.
13. E. De Clercq, Rev. Med. Virol. 1996, 6, 97.
14. E. De Clercq, Med. Res. Rev. 1993, 13, 229.
15. E. De Clercq, Antiviral Res. 1998, 38, 153.
16. Tantillo, J. Ding, A. Jacobo-Molina, R. G. Nanni, P. L. Boyer, S. H. Hughes, R. Pauwels, K. Andries, P. A. J.Janssen, E. Arnold, J. Mol. Biol. 1994, 243, 369.
17. Smerdon S. J. et. al;Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. PNAS; 1994 .
18. Ren J, et. al;High resolution structures of HIV-1 RT from four RT-inhibitor complexes. Nature Structure Biology; 1995 .
19. Ren, J.; Milton, J.; Weaver, K. L.; Short, S. A.; Stuart, D. I.; et al. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. Struct. Fold Des. 2000, 8, 1089-1094.
20. Lindberg, J.; Sigurdsson, S.; Lowgren, S.; Andersson, H. O.; Sahlberg, C.; et al. Structural basis for the inhibitory efficacy of efavirenz (DMP-266), MSC194 and PNU142721 towards the HIV-1 RT K103N mutant. Eur. J. Biochem. 2002, 269, 1670- 1677.
21. Esnouf, R. M.; Ren, J.; Hopkins, A. L.; Ross, C. K.; Jones, E. Y.;et al. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this non-nucleoside inhibitor. Proc. Natl. Acad. Sci. U.S.A. 1997, 94, 3984-3989.
22. Coffin JM: HIV viral dynamics. AIDS 1996, 10 (Suppl 3):S75-S84.
23. Domingo E, Mas A, Yuste E, Pariente N, Sierra S, Gutierrez-Riva M, Menendez-Arias L: Virus population dynamics, fitness variations and the control of viral disease: an update. Prog Drug Res 2001, 57:77-115.
24. Baba M, Shigeta S, Yuasa S, Takashima H, Sekiya K, Ubasawa M, Tanaka H, Miyasaka T, Walker RT, De Clercq E: Preclinical evaluation of MKC-442, a highly potent and specific inhibitor of human immunodeficiency virus type 1 in vitro. Antimicrob Agents Chemother 1994, 38:688-692.
25. Kleim J-P, Bender R, Kirsch R, Weichsner C, Paessens A, Rosner M, Rubsamen-Waigmann H, Kaiser R,

Wichers M,Schneweis KE et al.: Preclinical evaluation of HBY 097, a new non-nucleoside reverse transcriptase inhibitor of human immunodeficiency virus type 1 replication. Antimicrob Agents Chemother 1995, 39:2253-2257.

26. Kashman Y, Gustafson KR, Fuller RW, Cardellina JH II, McMahon JB, Currens MJ, Buckheit RW Jr, Hughes SH, Cragg GM, Boyd MR: The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical rainforest tree, Calophyllum lanigerum. J Med Chem 1992, 35:2735-2743.
27. Sahlberg C, Noreen R, Engelhardt P, Hogberg M, Nangasmetsa J, Vrang L, Zhang H: Synthesis and anti-HIV activities of urea-PETT analogs belonging to a new class of potent non-nucleoside HIV-1 reverse transcriptase inhibitors. Bioorg Med Chem Lett 1998, 8:1511-1516.
28. Corbett JW, Ko SS, Rodgers JD, Gearhart LA, Magnus NA, Bacheler LT, Diamond S, Jeffrey S, Klabe RM, Cordova BC et al.: Inhibition of clinically relevant mutant variants of HIV-1 by quinazoline non-nucleoside reverse transcriptase inhibitors. J Med Chem 2000, 43:2019-2030.
29. Tramontano E, Marongiu ME, de Montis A et al. Characterization of the anti-HIV-1 activity of 3,4-dihydro-2-alkoxy-6-benzyl-4-oxopyrimidines (DABOs), new non-nucleoside reverse transcriptase inhibitors. New Microbiol 1994; 17: 269–79.
30. Pani A, Musiu C, Loi AG et al. DABOs as candidates to prevent mucosal HIV transmission. Antivir Chem Chemother 2001.
31. Hsiou, Y., Das, K., Ding, J., Clark, A. D., Jr., Kleim, J. P., Rosner, M. et al. (1998). Structures of Tyr188 Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097: inhibitor flexibility is a useful design feature for reducing drug resistance. J. Mol. Biol. 284, 313–323.
32. Ren, J., Nichols, C., Bird, L. E., Fujiwara, T., Sugimoto, H., Stuart, D. I. & Stammers, D. K. (2000). Binding of the second generation non-nucleoside inhibitor S-1153 to HIV-1 reverse transcriptase involves extensive main chain hydrogen bonding. J. Biol. Chem. 275,14316–14320.
33. Ding, J., Das, K., Hsiou, Y., Sarafianos, S. G., Clark,A. D., Jr., Jacobo-Molina, A. et al. (1998). 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. 284, 1095–1111.
34. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science, 282, 1669–1675.
35. Tantillo, C.; Ding, J.; Jacobo-Molina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A. J.; Arnold, E. J. Mol. Biol. 1994, 243, 369-387.
36. Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Science 1992, 256, 1783-1790.
37. Ding, J.; et al. Structure 1995, 3, 365-379.
38. Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A.; Hughes, S. H.; Arnold, E. Nat. Struct. Biol. 1995, 2, 407-415.
39. Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Jr., A. D. C.; Lu, X.; Tamtilo, C.; Williams, R. L.; Kamer, G.; Ferris, A. L.; Clark, P.; Hizi, A.; Hughes, S. H.; Arnold, E. Proc. Natl. Acad. Sci. U.S.A. 1993, 90, 6320-6324.
40. Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Science 1998, 282, 1669-1675.
41. Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. Science 1998, 282, 1669-1675.
42. De Clercq, E. Clin. Microbiol. ReV. 1995, 8, 200-239.
43. Larder, B. A. In ReVerse transcriptase; Skalka, A. M., G. S., Eds.; Cold spring Harbor Laboratory Press: Plainview, NY, 1993; pp 205-222.
44. Larder, B. A. In ReVerse Transcriptase; Skalka, A. M., G. S., Eds.; Cold spring Harbor Laboratory Press: Plainview, NY, 1993; pp 163-191.
45. Ding, J.; Das, K.; Hsiou, Y.; Zhang, W.; Arnold, E.; Yadav, P. N. S.; Hughes, S. H. Struct.-Based Drug Des. 1997, 41-82.
46. Pedersen, O. S.; Pedersen, E. B. AntiViral Chem. Chemother. 1999, 10, 285-314.
47. Huang H, Chopra R, Verdine GL, Harrison SC: Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 1998, 282:1669-1675.
48. Temiz NA, Bahar I: Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. Proteins 2002, 49:61-70.

49. 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.
50. Kristina A paris et al, 2009, *52* (20), pp 6413–6420, Conformational Landscape of the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Non-Nucleoside Inhibitor Binding Pocket: Lessons for Inhibitor Design from a Cluster Analysis of Many Crystal Structures.
51. Hopkins AL et al, 1996, Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors.J Med Chem. 1996 Apr 12;39(8):1589-600.
52. McCammon et al, 2009. Elucidating the Inhibition Mechanism of HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors through Multicopy Molecular Dynamics Simulations. J. Mol. Biol. (2009) 388, 644-658.
53. M. Baba, H. Tanaka, E. De Clercq, R. Pauwels, J. Balzarini, D. Schols, H. Nakashima, C.-F. Perno, R. T. Walker, T. Miyasaka, Biochem. Biophys. Res. Commun. 1989, 165, 1375.
54. T. Miyasaka, H. Tanaka, M. Baba, H. Hayakawa, R. T. Walker, J. Balzarini, E. De Clercq, J. Med. Chem.1989, 32, 2507.
55. M. Baba, S. Shigeta, S. Yuasa, H. Takashima, K. Sekiya, M. Ubasawa, H. Tanaka, T. Miyasaka, R. T. Walker, E. De Clercq, Antimicrob. Agents Chemother. 1994, 38, 688.
56. Esnouf, R.; Ren, J. S.; Ross, C.; Jones, Y.; Stammers, D.; Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside Inhibitors. Nature Struct. Biol. 1995, 2, 303- 308.
57. Andrew L. Hopkins et al, 1999. Design of MKC-442 (Emivirine) Analogues with Improved Activity Against Drug-Resistant HIV Mutants.J. Med. Chem. 1999, 42, 4500-4505.
58. Ren J., Structure of HIV-2 reverse transcriptase at 2.35-Å resolution and the mechanism of resistance to non-nucleoside inhibitors.PNAS. 2002.
59. Miyasaka, T.; Tanaka, H.; Baba, M.; Hayakawa, H.; Walker, R.T.; Balzarini, J.; De Clercq, E. A novel lead for specific anti-HIV-1 agents: 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine. J. Med. Chem. 1989, 32, 2507-2509.
60. Tanaka, H.; Baba, M.; Ubasawa, M.; Takashima, H.; Sekiya, K.; Nitta, I.; Shigeta, S.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Synthesis and anti-HIV-1 activity of 2-, 3-, and 4-substituted analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT). J. Med. Chem. 1991, 34, 1394-1399.
61. Tanaka, H.; Baba, M.; Saito, S.; Miyasaka, T.; Takashima, H.; Sekiya, K.; Ubasawa, M.; Nitta, I.; Walker, R. T.; Nakashima, H.; De Clercq, E. Specific anti-HIV-1 "acyclonucleosides" which cannot be phosphorylated: synthesis of some deoxy analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine. J. Med. Chem. 1991, 34, 1508-1511.
62. Tanaka, H.; Takashima, H.; Ubasawa, M.; Sekiya, K.; Nitta, I.; Baba, M.; Shigeta, S.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Structure-activity relationships of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio)thymine analogues: effects of substituents at the C-6 phenyl ring and at the C-5 position on anti-HIV-1 activity. J. Med. Chem. 1992, 35, 337-345.
63. Tanaka, H.; Takashima, H.; Ubasawa, H.; Sekiya, K.; Nitta, I.; Baba, M.; Shigeta, S.; Walker, R. T.; De Clercq, E.; Miyasaka, T. Synthesis and antiviral activity of deoxy analogues of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)-thymine (HEPT) as potent and selective anti-HIV-1 agents. J. Med. Chem. 1992, 35, 4713- 4719.
64. Mao, B., Pear, M. R., McCammon, J. A. & Quiocho, F. A. (1982). Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. J. Biol. Chem. 257, 1131–1133.
65. Liu, S. X., Abbondanzieri, E. A., Rausch, J. W., Le Grice, S. F. J. & Zhuang, X. W. (2008). Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. Science, 322, 1092–1097.
66. Spence, R. A., Kati, W. M., Anderson, K. S. & Johnson, K. A. (1995). Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. Science, 267, 988–993.
67. Rittinger, K., Divita, G. & Goody, R. S. (1995). Human immunodeficiency virus reverse transcriptase substrate-induced conformational changes and the mechanism of inhibition by nonnucleoside inhibitors. Proc. Natl Acad. Sci. USA, 92, 8046–8049.
68. Hashimoto, H., Mizutani, K. & Yoshida, A. in WO 00147883 (Japan Tobacco Inc., Published International Patent Application, 2001.

IV)-
Conclusion

The main goal of our investigation has been to understand the molecular basis of protein ligand interaction targeting viral polymerases, such as HCV, BVDV RdRp and HIV-RT. In order to combat the viral resistance to antiviral NNIs, we need to reinvent new ways to develop antivirals. It is a common opinion that a structure-based bottom-up approach can represent a fast and cheap route to new antivirals: starting from the knowledge of resistance mechanisms it would be possible to design rationally new molecule able to overcome resistance. Information gathered at the molecular level by MD simulations can contribute to follow this new route, as described in this thesis.

Herein, we have employed an arsenal of computational techniques to explore the changes in protein dynamics on ligand binding, to understand the molecular basis of drug interaction, to shed light on probable mechanism of inhibition and resistance, to understand the effect of mutation on the dynamics of protein, to identify principal interactions in a protein ligand system; all this with an aim to provide clues for further optimization of affinity and for the design of inhibitors that are less susceptible to the mutations.

In our previous work we identified a novel binding site for the binding of 227G, belonging to benzimidazole class of compound, in the finger domain of BVDV RdRp. However, resistant mutant I261M was found to confer resistance to 227G. This work was aimed at understanding as to how the mutation I261M renders the protein resistant against 227G.

To obtain our objective computationally extensive protocol which include Docking, Standard Molecular dynamics, clustering and many analysis tool from our armamentarium have been employed to understand the structural and dynamical effects of the I261M mutation on BVDV RdRp.

We were able to pin point difference in the dynamics of the APO and APOm protein. We observed that major conformational changes occur in the loop L4 which constitute a part of the putative binding pocket. RMSD analysis showed that APOm system evolves periodically but with a larger amplitude and lower frequency as compared to the APO. Docking on the clusters obtained from trajectory of APOm indicated a different positioning of 227G in the cavity as compared to the COM system. Comparative study of the RMSD of loop L4 in COM and COMm also indicated different dynamics. It was observed that 227G in the COM system interact with L4 and thereby restrict its movement while due to the reorientation of the 227G in the COMm, it no longer interacts with the L4 and hence it retains its periodicity as in case of APO system.

Area analysis of the cavity in all four system indicated a complete closure of the entrance to the template channel in case of COM as compared to the other three systems. Chunnel analysis, which were performed on the clusters obtained from the COMm and COM systems indicated that in COM due to the presence of 227G all the three tunnels viz: 1)template entrance tunnel,2) NTP entrance tunnel ,3) Adduct exit tunnel were absent while in COMm all above mentioned tunnels are completely open. The availability of these tunnels in COMm, provides a possibility for RdRp to retain its polymerization by allowing the possible entry of template into the tunnels, even in the presence of 227G. From above explanation it seems that the 227G is completely ineffective in COMm which is a good agreement with the experimental results .

In a nutshell we hypothesized that on introduction of mutation I261M in the APO there is a change in the dynamics of the APOm specially with respect to the loop L4. Our docking studies showed that although 227G binds to the APOm but it positions itself in a way that does'nt hamper the entrance to the template into the template entrance tunnel and thereby retaining the polymerization capacity which is lost in the COM wherein the entrance to the template tunnel is completely blocked.

Apart from structural and dynamical aspects of analyzing the effect of resistant mutant , we also performed energetic analysis, and escape of drug in wild type and mutant protein. In the present work, we applied to the problem of an inhibitors, acting on the wild type and mutated BVDV RdRp, a computational strategy that accounts for all these effects in an all-atoms description. By studying the undocking mechanism we were able to accomplish an inventory of the interactions between 227G and the various residues of RdRp. At the same level of accuracy, the hydration of the system along the dissociation route was evaluated pointing out the different degree of hydration in the wild type and mutated RdRps. At the last the binding free energy gives a good relation with experimental data.

At present one of the most influential dilemma for the pharmaceutical industry is, whether to develop new drugs or promote those presently exist. In this aspect the in depth information mainly of protein ligand recognition, and the effect of resistant mutation could be a useful step to develop multi-resistant drugs.

Another important point, to account is the flexibility of the protein, mainly the functionally significant region. As it is a well known fact that the flexible loops govern the functional properties of the protein, and in our case we noticed the flexibility of the loop take part in the stability of drug, and most effected in the presence of resistant mutation.

In Chapter four, we employed extensive Docking and MD simulations in order to identify probable binding poses as well as pin point key residues crucial for the binding of DABO compounds into the binding cavity. The result of this analysis also agree very well on the experiments on mutant of RT, which show that identifying the principal interaction for a class of inhibitor can give an important direction for further optimization of affinity and for the design of inhibitors that are less susceptible to the mutations.

We also did a comparative analysis of the binding pattern of DABO with respect to structurally related HEPT(1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine) class of compounds, whose x-ray (1RT1) is already known. This was done to characterize similarities and dissimilarity in the binding of the two structurally related classes of compounds, with an aim to identify key determinants of binding and provide clues for further improvement of potency of DABO compounds.

Furthermore, we were also successful in throwing light on the probable mechanism of inhibition of the DABO compounds. Through our simulations we were able to observe that there is a widening of the thumb-finger distance in DABO bound structures throwing it in semi 'hyper-open' form. Fluorescence experiments, have suggested that NNRTI-induced loosening of the fingers–thumb grip, or in other words widening of the finger-thumb domain gap, as is observed in the binding of the DABO compounds, may cause RT to slide away from the 3′ primer end and thus prevent polymerization.

Thus binding of NNRTI results into a mechanical obstruction in the movement of thumb and finger domains required for proper positioning of the incoming DNA into the NNRTI bound HIV-RT. We also sought to link the mechanical obstruction to local conformational changes taking place in the catalytic site leading to impairment of polymerization. Through our simulations we observe that binding of NNRTI into the binding pocket resulted into a local conformational change within the binding pocket, by widening the distance between two beta sheets , necessary for correct positioning of 3' end of the template. Thus, by our simulation we were able to hypothesize that widening of distance between the thumb-finger domain leads to a 'hyper open' of HIV-RT which along with the distortions produced in the residues at the active site may result into inhibition of DABO bound HIV-RT structures.

To sum up, in the current study, we have presented a complete analysis of the effect of resistant mutant at dynamical and energetic level, as well as tried to understand, how the polymerization regained by RdRp in mutant protein though drug is present into the cavity.

It is noteworthy that, due to the resistance mechanism, designing a potent antiviral is a very challenging task . In future, we have plan to extend the investigation to understand the resistance mechanism as it is one of the tough task in current scenario of drug development pipe line.