(Research Article)

INTERNATIONAL JOURNAL
OF
PHARMACEUTICAL SCIENCES
AND
RESEARCH

HOMOLOGY MODELING AND DOCKING STUDIES ON HIV REVERSE TRANSRIPTASE INHIBITORS

T.B. Kakade

College of Pharmacy, Medha, Satara, Maharashtra, India

Correspondence to Author:

T.B. Kakade

College of Pharmacy Medha, Satara, Maharashtra, India
E-mail: dipali.1386@gmail.com

ABSTRACT: The usefulness of computational methods in drug design process is no longer a point of debate. The contributions from this evolving field have become apparent to both pharmaceutical scientists and academic biomedical researchers. As all the computational drug design are based on statistical and mathematical calculations there is strong evidences present for each and every step, thus research by chance is no longer exist. From statistical analysis of large data set, up to quantitative prediction, one can get simply by seating in front of graphical terminal and using this technique. Success stories from these endeavors are appearing in the literature at an impressive rate. Since the search for new antiviral drugs with reduced toxicity and lower side-effect is continues. Antiviral of class HIV reverse transcriptase are found to have wide range of activities but still this drugs are not very effective in curing HIV infected persons. So this drugs study is carried out and will show more biological and therapeutic activity by performing work of homology modeling and docking study of HIV reverse transcriptase as this work study has not been done earlier by following literature review This is an attempted with docking studies that would be helpful in guiding further development of more active and selective HIV reverse transcriptase inhibitors with more affinity towards receptor with less binding energy.

INTRODUCTION: Computational chemistry is a branch of chemistry that uses principles of computer science to assist in solving chemical problems. It uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of molecules and solids. Its necessity arises from the well-known fact that apart from relatively recent results concerning the hydrogen molecular ion (see references therein for more details), the quantum n-body problem cannot be solved analytically, much less in closed form. While its results normally complement the information obtained by chemical experiments, it can in some cases predict hitherto unobserved chemical phenomena. It is widely used in the design of new drugs and materials

The discovery of drug and drug molecule has always been the aim of medicinal chemistry. Occasionally new drugs are found by accident. More frequently they are developed as part of organized efforts to discover new ways to treat specific disease. The discovery of new drug has gone through an evolution over the years and new technologies are being added namely Chemo informatics and Computer aided drug design.

Although the phrase computer aided drug design may seems to imply that drug discovery lies in the hand of computational scientist who are able to manipulate molecule on their computer screen. But drug design process is actually a complex and interactive method [1].

Successful drug design is multi-step, multi-disciplinary procedure. Drug design is not merely a technology that generates drug for human on basis of biological advances. From a very broad perspective drug design may be divided into two phases.

A) Basic concept about the drug, receptor and drug-receptor interaction.

B) Concept of drug receptor interaction applied to human disease.

First step of this involves knowing what properties turn a molecule into a drug Computer aided molecular designing (CAMD) is a very important and valuable tool in developing first step in drug designing. Another part of first step knows properties of macromolecule which turns the macromolecule into a receptor. And the third and important part involves designing of specific drug like molecule to fit into a particular target [2].

There are various approaches for identification of lead compound.

1) Rational drug design.

2) Random high through put screening.

3) Focused library screening.

Out of these three approaches rational drug designing is widely used approach, this drug design approach again has two different approaches,

1) Structure based drug designing.

2) Analogue based drug designing.

1) **Structure based drug designing** [3]: Structure based drug design is possible only when sufficient information is known about the receptor or target, the attempt can be made to design directly a drug molecule to fit to the target. Molecular modeling is one of the important approaches of structure based drug design. Docking is a study of receptor-ligand binding and is structure based drug design.

2) **Analogue based Drug Designing:** The easiest way to create a new drug is to modify an existing one. These so-called drug analogs account for about half of all "new" drugs and are of tremendous importance both for medical and financial reasons. In addition, drug analogs can be modeled on naturally occurring bioactive substances that do not make good drugs on their own, thereby considerably increasing the therapeutic range of pharmaceutical medicine. This analogue based drug design is solely related with the biological activity of drug and development of new drug by taking into consideration the physicochemical properties of molecule thus, QSAR can be considered as an important approach of drug design.

Once the lead compound is successfully identified, it must be optimized. This optimization is achieved using Quantitative Structure Activity Relationship (QSAR) [4]. Once the drug is taken into the body in any form it goes through pharmacodynamic-pharmacokinetic- pharmacodynamic phase. First phase that is pharmacodynamic phase in which dissociation, dissolution etc are considered, the second phase is pharmacokinetic phase refers to absorption, distribution, metabolism and excretion of drug.

After the drug molecule is surmounted the barrier it has been distributed through the body it ultimately reaches to macromolecular environment of receptor where it exerts its biological effect. The binding of drug molecule to receptor takes place via variety of interaction i.e. ion-ion, dipole interaction, hydrogen bonding etc.

Thus, the first approach of rational drug design is solely bases on this interactions, structure based drug design i.e. quantitative structure activity relationship (QSAR).

Homology Modeling and Docking:

1) **Homology Modeling** [12]:

a. **Protein structure visualization, comparison, and classification:**
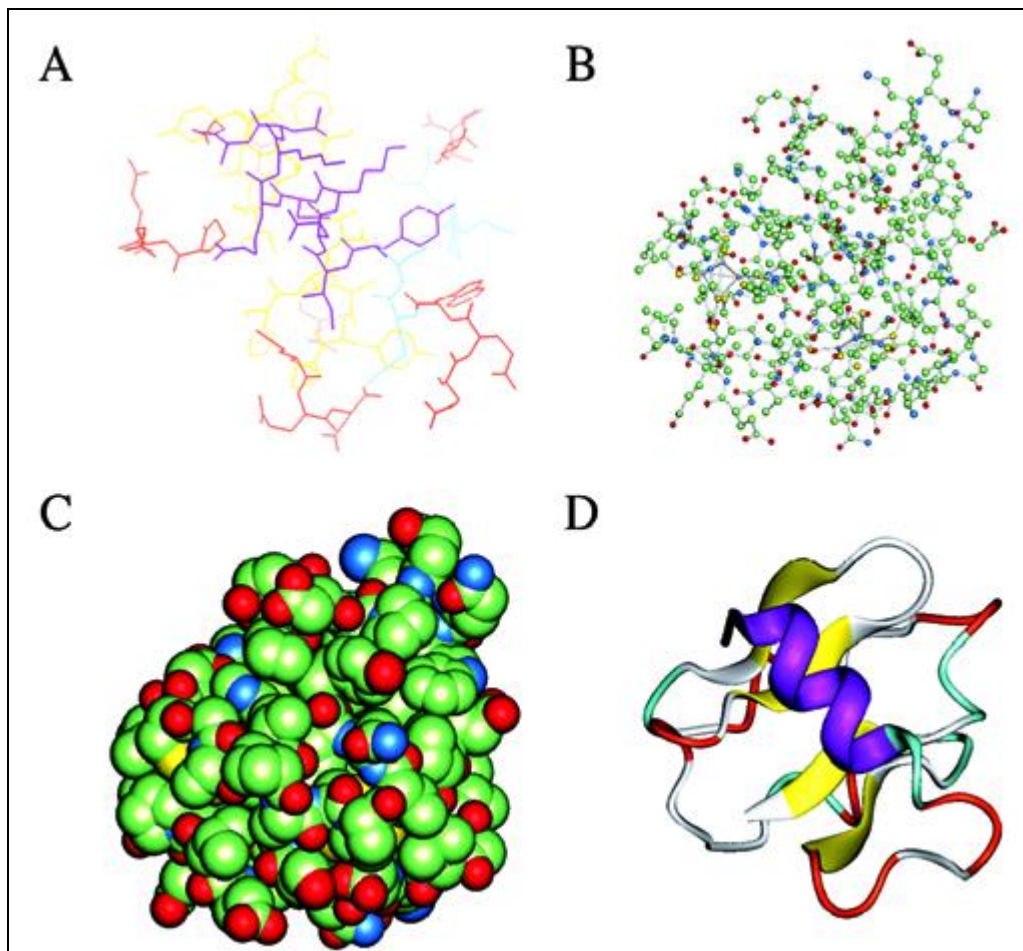
EXAMPLES OF MOLECULAR STRUCTURE VISUALISATION FORMS; (A) WIREFRAMES (B) BALLS AND STICKS (C) SPACE FILLING SPHERES (D) RIBBONS

FIGURE 1: PROTEIN STRUCTURE

Protein structure prediction: Protein structure prediction is one of the most important goals of bioinformatics and theoretical chemistry. Its aim is to predict 3 Dimensional structure of protein from amino acid sequence. It deals with prediction of proteins tertiary structure from primary structure. Protein structure prediction is of high importance in medicine.

Primary structure: The amino acid sequence of the peptide chains.

Secondary structure: Highly regular sub-structures (*alpha helix* and *strands of beta sheet*) which are locally defined, meaning that there can be many different secondary motifs present in one single protein molecule.

Tertiary structure: Three-dimensional structure of a single protein molecule; a spatial arrangement of the secondary structures. It also describes the completely folded and compacted polypeptide chain.

Quaternary structure: complex of several protein molecules or polypeptide chains, usually called protein subunits in this context, which function as part of the larger assembly or protein complex.

b. **Types of Protein modeling:**

i. **Experimental methods of Protein modeling** [5]**:**

1) **X-Ray Crystallography:** This is a method of determining the arrangement of atoms within a crystal, in which a beam of X-rays strikes a crystal and scatters into many different directions. From the angles and intensities of these scattered beams, a crystallographer can produce a three-dimensional picture of the density of electrons within the crystal. From this electron density, the mean positions of the atoms in the crystal can be determined, as well as their chemical bonds their disorder and sundry other information.

2) NMR Spectroscopy (Nuclear Magnetic Resonance): Due to poor results and unclear band obtained with this technique. NMR is not generally used for structure determination.

3) MS (Mass Spectroscopy): Due to poor resolution of this techniques and gap in template results in development of new techniques for homology modeling.

ii. **Computational Protein modeling:** There are mainly three methods in computational modeling,

1) **Protein threading or Fold recognition** [6]**:** In this method the amino acid sequence of an unknown structure is scanned against a database of solved structures. In each case, a scoring function is used to assess the compatibility of the sequence to the structure, thus yielding possible three-dimensional models.

2) *Ab-initio* **Protein modeling:** *Ab-initio* or *de-novo* protein modeling methods seek to build three-dimensional protein models "from scratch", i.e., based on physical principles rather than (directly) on previously solved structures. To predict protein structure *de novo* method for larger proteins will require better algorithms and larger computational resources like those afforded by either powerful supercomputer. Although the computational barriers are vast but the potential benefits of structural genomics (by predicted or experimental methods) make *ab initio* structure prediction an active research field.

3) **Homology modeling** [7]**:** Homology modeling is also known as comparative modeling, in this methods protein structure prediction for constructing an atomic-resolution model of a protein from its amino acid sequence is done. Almost all homology modeling techniques rely on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target.

There are several ways of developing homology model for certain protein, few of them are reported.
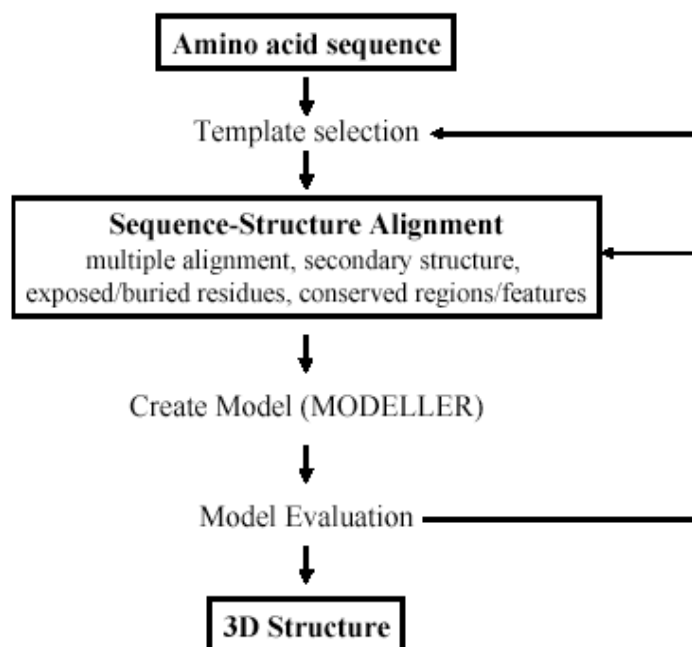
Comparative Modelling

FIGURE 2: COMPARATIVE MODELING

iii. **CASP:** CASP (Critical Assessment of Structure Prediction) blind prediction experiments aim to assess the prediction capabilities in the field. A limitation of CASP is that predictions are prepared and filed by humans using programs, and thus, what is being evaluated is the performance of the predicting groups rather than the performance of the programs themselves.

iv. **CAFASP:** To address this limitation, the Critical Assessment of Fully Automated Structure Prediction (CAFASP) experiment was initiated in 1998. In CAFASP, the participants are programs or Internet servers, and what is evaluated are their automatic results without allowing any human intervention.

v. *An in silico* **technique** [8]: newly developed *In silico* protein designing method has recently been applied to solve problems in protein stabilization. Primary advantage of *In silico* designing is the vast number of sequences that can be rapidly screen in the search for an optimal design. This feature allows for large changes in protein properties to be discovered.

There are total seven steps of homology modeling [9]:

1. Template recognition and initial alignment

2. Alignment correction

3. Backbone generation

4. Loop modeling

5. Side-chain modeling

6. Model optimization

7. Model validation

1) **Template recognition and initial alignment:** In the safe homology modeling zone the percentage identity between the sequence of interest and a possible template is high enough to be detected with simple sequence alignment programs such as BLAST or FASTA.

2) **Alignment correlation:** Aligning the sequences of the target protein and the template protein. The aim of this step is to match each residue in the target sequence to its corresponding residue in the template structure, allowing for insertions and deletions.

3) **Backbone generation:** When the alignment is ready, the actual model building can start. Creating the backbone is trivial for most of the model: One simply copies the coordinates of those template residues that show up in the alignment with the model sequence. If two or more aligned residues differ, only the backbone coordinates (N, Cα, C and O) can be copied.

4) **Loop modeling:** In the majority of cases, the alignment between model and template sequence contains gaps. Either gaps in the model sequence or in the template sequence (insertions). In the first case, one simply omits residues from the template, creating a hole in the model

5) **Side-chain modeling:** Building the side chains/loop. Coordinates of identical residues can be imported directly from the template to the target, but the side-chain conformations of the nonidentical ones cannot, and thus they need to be predicted. The prediction of the side-chain coordinates is usually based on empirical data collected from proteins of known structure.

6) **Model optimization:** The problem just mentioned above leads to a classical chicken-and-egg situation. To predict the side-chain rotamers with high accuracy; we need the correct backbone this in turn depends on the rotamers and their packing. The common approach to such a problem is an iterative one: predict the rotamers, then the resulting shifts in the backbone, then the rotamers for the new backbone, and so on, until the procedure converges.

7) **Model validation:** Every homology model contains an error, the percentage sequence identity between template and target. If it is greater than 90%, the accuracy of the model can be compared to crystallographic ally determined structures, except for a few individual side chains.
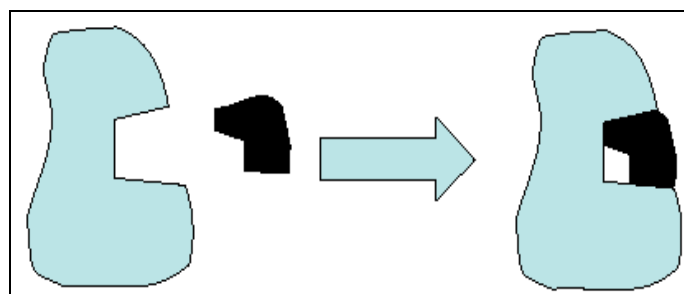
Docking [10]**:**

FIGURE 3: DOCKING PROCESS

Docking is another approach of rational drug design i.e. structure based drug design in which protein structures is known and drug molecules are designed and are checked for binding affinity towards receptor and determined whether it can be used as drug or not. Docking is a program tool used to predict how small molecule or drug can fit into the receptor site of known 3D structure of protein.

The dock program calculates the affinity grids for a ligand-receptor interaction and helps the scientist to design drug with better affinities. The procedure of docking is accurate with reasonable demand of computational procedures. The first step is to find global minimum energy (GME) level in the interaction energy between the receptor and ligand along with exploring all the possible degree of freedom for complex.
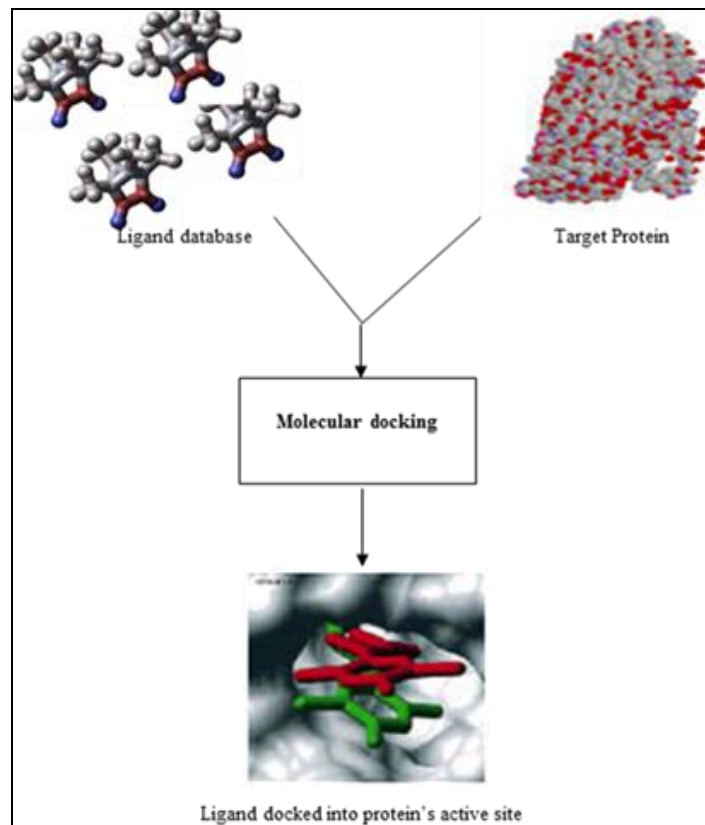
FIGURE 4: LIGAND DOCKING

1. **Scoring function [11]:** Scoring functions are fast approximate mathematical methods used to predict the strength of the non-covalent interaction between two molecules after they have been docked. Most commonly one of the molecules is a small organic compound such as a drug and the second is the drug's biological target such as a protein receptor. Scoring functions have also been developed to predict the strength of other types of intermolecular interactions, for example between two proteins or between protein and DNA. Scoring functions play an essential role in structure based virtual screening.

They are required to guide the docking of candidate compounds to structures of receptor binding sites, to select probable binding modes, and to discriminate binders from non-binders, although many scoring functions have successfully been used to identify novel. Ligands for a wide variety of targets, much work remains to be done to avoiding correct prediction of binding modes and high numbers of false positives. This review gives an overview of the current state of the field and outlines key issues for the further development of scoring functions:

Types of scoring function:

The three different types of scoring functions are there:

A. Force field based scoring function.

B. Empirical scoring function.

C. Knowledge based scoring function.

Types of Docking:

1) **Semi flexible docking:** When the large protein is rigid and small drug molecule is flexible this is called semi flexible docking. Example is Protein- ligand docking.

Protein - ligand docking: Protein ligand docking is a molecular modeling technique. The goal of protein ligand docking is to predict the position and orientation of ligand when it is bound to a protein receptor.
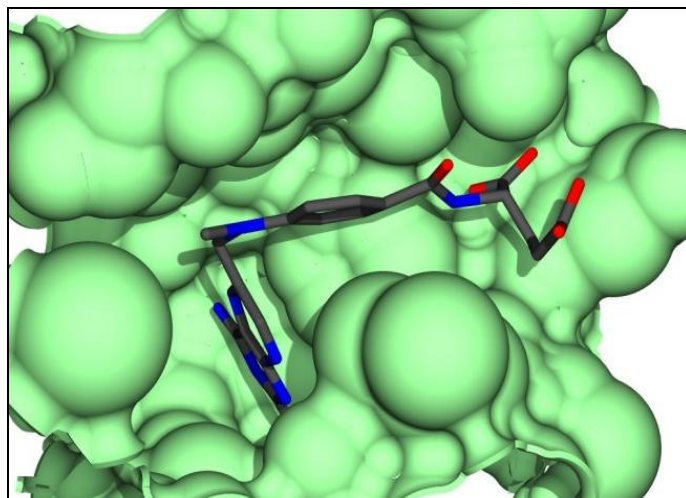

FIGURE 5: PROTEIN LIGAND DOCKING

2) **Rigid docking:** When both the protein molecules are rigid in nature then this docking is said to be rigid docking. Example is Protein-protein docking is more rigid**.**

Protein-protein docking: A protein-protein interaction occurs between two proteins that are similar in size. The interface between two molecules tends to be flatter and smoother that those in protein ligand interaction. Protein-protein interactions

3) **Flexible docking:** When both the molecules are flexible in nature then this docking is turned as flexible docking. Example is Protein-protein docking.

Protein-protein docking: A protein-protein interaction occurs between two proteins. This docking can be of either the rigid type or flexible type.
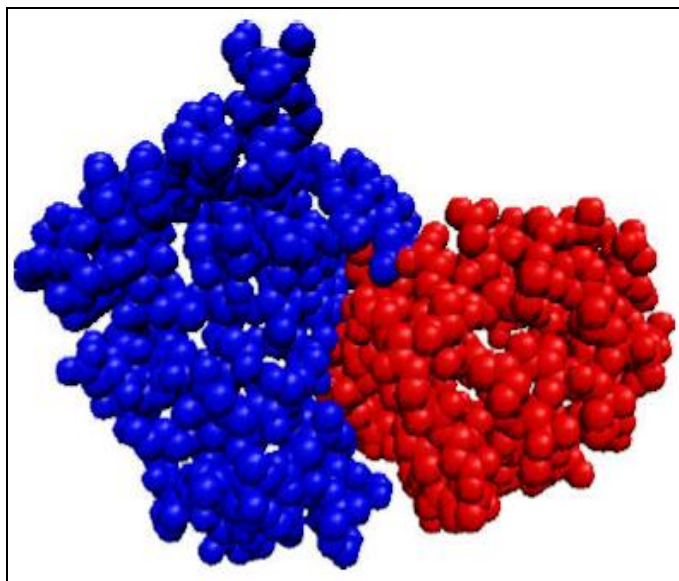

FIGURE 6: PROTEIN-PROTEIN DOCKING

AIDS and HIV [12]**:**

Acquired Immune Deficiency Syndrome or **Acquired Immunodeficiency Syndrome** (**AIDS**) is a disease of the human-immune system caused by the human immunodeficiency virus (HIV). The illness interferes with the immune system making people with AIDS much more likely to get infections, including opportunistic infections and tumors that do not affect people with working immune systems. This susceptibility gets worse as the disease continues.

HIV is transmitted in many ways, such as anal, vaginal or oral sex, blood transfusion, contaminated hypodermic needles, exchange between mother and baby during pregnancy, childbirth, and breastfeeding. It can be transmitted by any contact of a mucous membrane or the bloodstream with a bodily fluid that has the virus in it, such as the blood, semen, vaginal fluid, preseminal fluid, or breast milk from an infected person. The virus and disease are often referred to together as HIV/AIDS.

The disease is a major health problem in many parts of the world, and is considered a pandemic, a disease outbreak that is not only present over a large area but is actively spreading. In 2009, the World Health Organization (WHO) estimated that there are 33.4 million people worldwide living with HIV/AIDS, with 2.7 million new HIV infections per year and 2.0 million annual deaths due to AIDS. In 2007 UNAIDS estimated: 33.2 million people worldwide were HIV positive; AIDS killed 2.1 million people in the course of that year, including 330,000 children, and 76% of those deaths occurred in sub-Saharan Africa. According to UNAIDS 2009 report, worldwide some 60 million people have been infected since the start of the pandemic, with some 25 million deaths, and 14 million orphaned children in southern Africa alone.

HIV-1 Reverse Transcriptase: The enzyme reverse transcriptase (RT) is used by retroviruses to transcribe their single-stranded RNA genome into single-stranded DNA and to subsequently construct a complementary strand of DNA, providing a DNA double helix capable of integration into host cell chromosomes. Functional HIV1-RT is a heterodimer containing subunits of 66 kDa (p66) and 51 kDa (p51) P66 contains two domains, the N-terminal polymerase domain (440 residues) and the C-terminal RNase H domain (120 residues). P51 is processed by proteolytic cleavage of p66 and corresponds to the polymerase domain of the p66 subunit.

Portions of both p51 and the polymerase domain of p66 can be described as a "right hand" that contains three subdomains: fingers, palm, and thumb. The connection sub domain connects the hand of the polymerase domain and the RNase H domain in p66, which provides the ribonuclease activity of HIV-RT. Although p51 contains a connection sub domain, it lacks an RNase domain. The connection subdomains and the palm subdomains contain three-stranded b-sheets with a-helices on one side. The thumb subdomains comprise three a-helices. Interestingly, although the two subunits are identical in their primary amino acid sequence (except for length), they are structurally very different. This can be clearly seen by observing the subunits separately.

Three catalytic residues are exposed in the nucleic acid binding cleft of p66, but are buried in p51, which lacks a discernable cleft. Another striking difference between the two subunits is the orientation of the connection subdomain; in p51 it is tucked into a central position and contacts all of the other subdomains, but in p66 it contacts only RNase H and the thumb.

A question arises as to why HIV has evolved a heterodimer in which the smaller subunit (p51) is a cleavage product of the larger.

Antivirals [13]: Most of the antiviral drugs generally fall into the following groups:

- Nucleoside analogues that inhibit the viral reverse transcriptase enzyme, preventing replication e.g.: lamivudine, zidovudine.

- Non-nucleoside analogues that have the same effect e.g.: Efavirenz.

- Inhibitors of proteases that prevent viral protein processing e.g.: Saquinavir, Indinavir.

- Inhibitors of viral DNA polymerase that prevent replication e.g.: Acyclovir.

- Inhibitors of viral capsule disassembly e.g.: Amantidine.

- Inhibitors of neuraminidase that prevent viral escape from infected cells e.g.: Oseltamivir

- Immunomodulators that enhance host defences e.g.: Interferons.

Software Profile:

Tools and databases used in Homology modeling:

1. **NCBI** [14]: The National Center for Biotechnology Information (NCBI) is part of United States National Library of Medicine (NLM), a branch of National Institute of Health. The NCBI is located in Bethesda, Maryland and was established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conduct research in computational biology, develops software tools for analyzing genome data. NCBI houses genome sequencing data in Genbank and an index of biomedical research articles in Pubmed central, Pubmed and other information relevant to Biotechnology and Pharmacy.

2. **BLAST** [15]: Basic Local Alignment Search Tool (BLAST) is an algorithm for comparing primary biological sequence information, such as amino acid sequences of different proteins or the nucleotide of DNA sequence. A BLAST search enables a researcher to compare a query sequence with library or database of sequence and identify library sequences that resemble the query sequence above a certain threshold.

3. **Swiss-Prot** [16]: Swiss-Prot is a manually created biological database of protein sequence. Swiss-Prot was created in 1986 by Amos Bairoch during his PhD and developed by Swiss Institute of Bioinformatics and European Bioinformatics Institute. Swiss-Prot strives to provide reliable protein sequences associated with a high level of annotation, a minimum level of redundancy and high level of integration with other databases.

4. **PDB** [17]: Protein Data Bank was founded in Brookhaven National Library in year 1971 by Dr. Edgar Meyer and Dr.Walter Hamilton. Then in 1999 it was transferred to members of Research Collaboratory for Structural Bioinformatics (RCSB).Mission of the PDB is to maintain a single Protein Data bank Achieve of macromolecular structural data that is freely and publicly available to the global community. The PDB is a key resource in structural biology and is critical to more recent work in structural genomics.

Online server used in Homology modeling:

1. **Cph 2.0 Model server** [18]: Cph 2.0 model server is an improved method for fold recognition/homology modeling, in which a large sequence database is iteratively searched to construct a sequence profile until a template can be found in a database of proteins with known structure. Query and template sequences are subsequently aligned using a score based on profile-profile comparisons.

2. **Phyre model server** [19]: It is a homology modeling server. The models produced by Phyre are based on finding a sequence alignment to a known structure, copying the coordinates and relabelling the residues according to your sequence. It can detect remotely homologous structures that can't be found by conventional methods. This is because of use profiles generated by PSI-Blast for both our sequence and the sequences of the known structures. Phyre performs a profile-profile matching algorithm together with predicted secondary structure matching.

3. **D Jigsaw Model server** [20]: This protein server builds three-dimensional models for proteins based on homologues of known structure. Templates are identified using HMM (Soding, 2005). The returned alignments are used to build the models. All models are preselected using populus energy. Gaps and missing residues are closed and filled using populus repair. Finally all models are recombined using the basic populus approach.
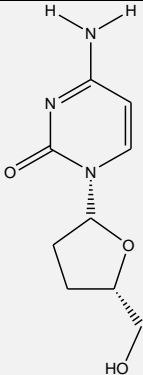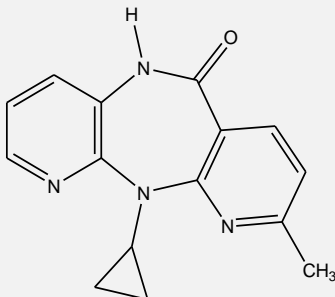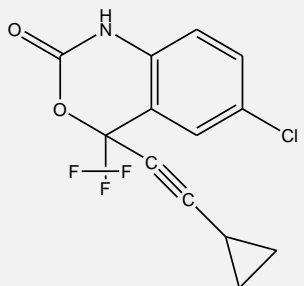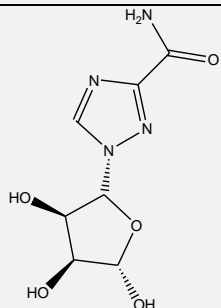
Software used in Homology modeling:
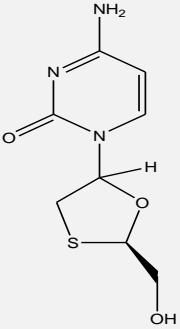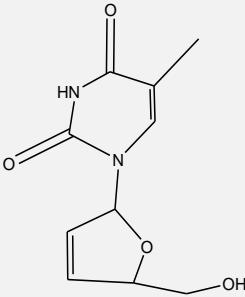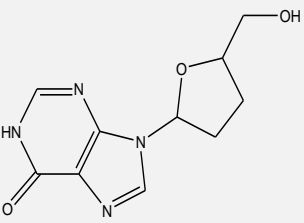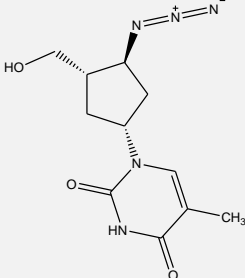
1. **Procheck** [21]: Procheck is used to assess both the overall stereochemical quality of a given protein structure, as compared with well-refined structures at the same resolution. And also to give an indication of its local, residue-by-residue reliability. To assess a structure, the program makes use of a number of parameters that have been found to be good indicators of stereochemical quality. These parameters, which are for the most part not, included in standard refinement procedures. The checks also make use of "ideal" bond lengths and bond angles, as derived from a recent and comprehensive analysis of small molecule structures in the Cambridge Structural Database.

Software used for docking study:

1. **Molegro virtual Docker (MVD)** [22]: Molegro Virtual Docker (MVD) is an integrated environment for studying and predicting how ligands interact with macromolecules. The identification of ligand binding modes is done by iteratively evaluating a number of candidate solutions (ligand conformations) and estimating the energy of their interactions with the macromolecule. The highest scoring solutions are returned for further analysis. MVD requires a three-dimensional structure of both protein and ligand (usually derived from X-ray/NMR experiments or homology modeling). MVD performs flexible ligand docking, so the optimal geometry of the ligand will be determined during the docking. High docking accuracy; the docking engine has been proven to correctly identify binding modes with high accuracy. It has been shown to outperform other docking programs with regard to identification of correct binding modes. Easy to use build in wizards enable the user to easily setup and perform docking runs.
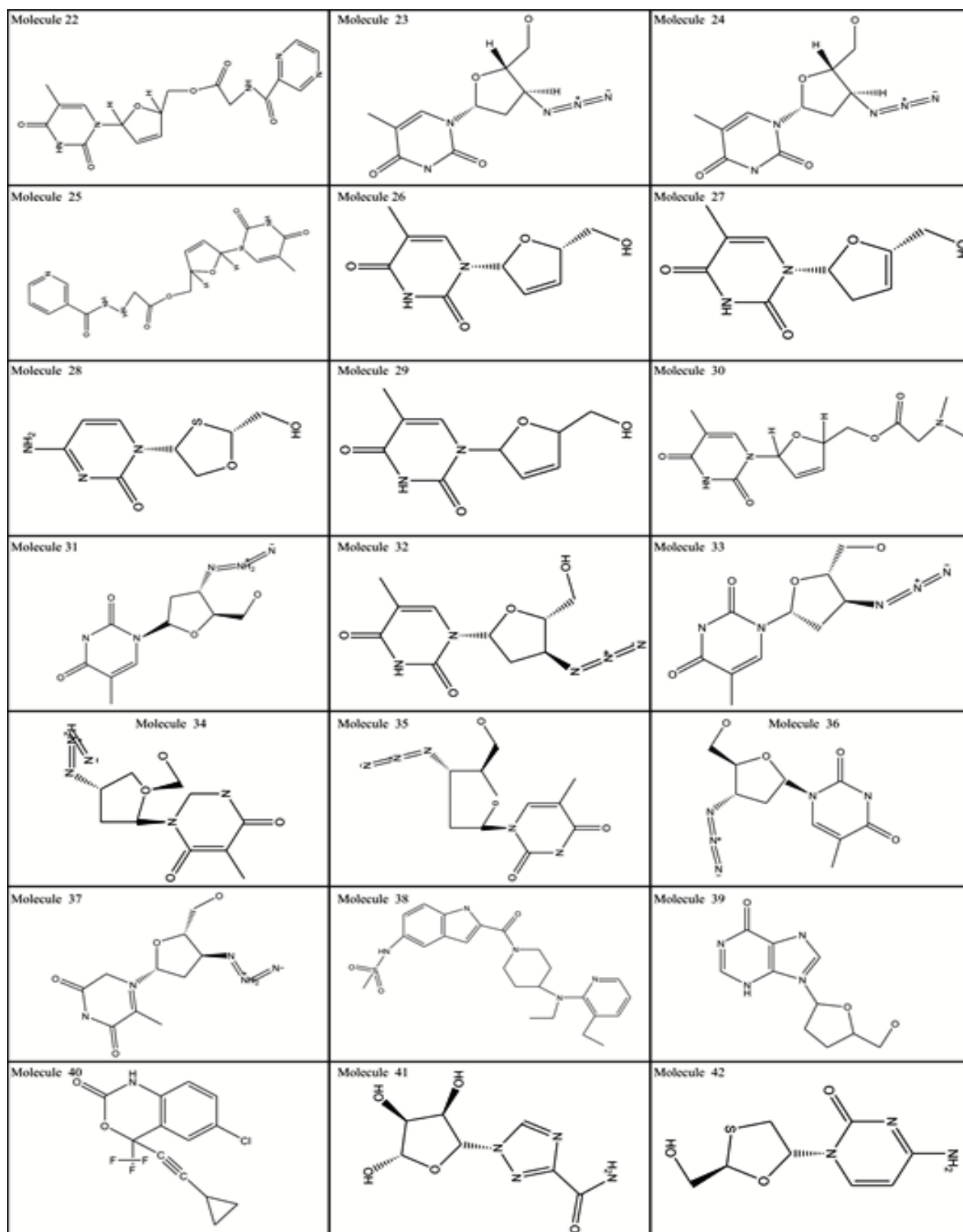
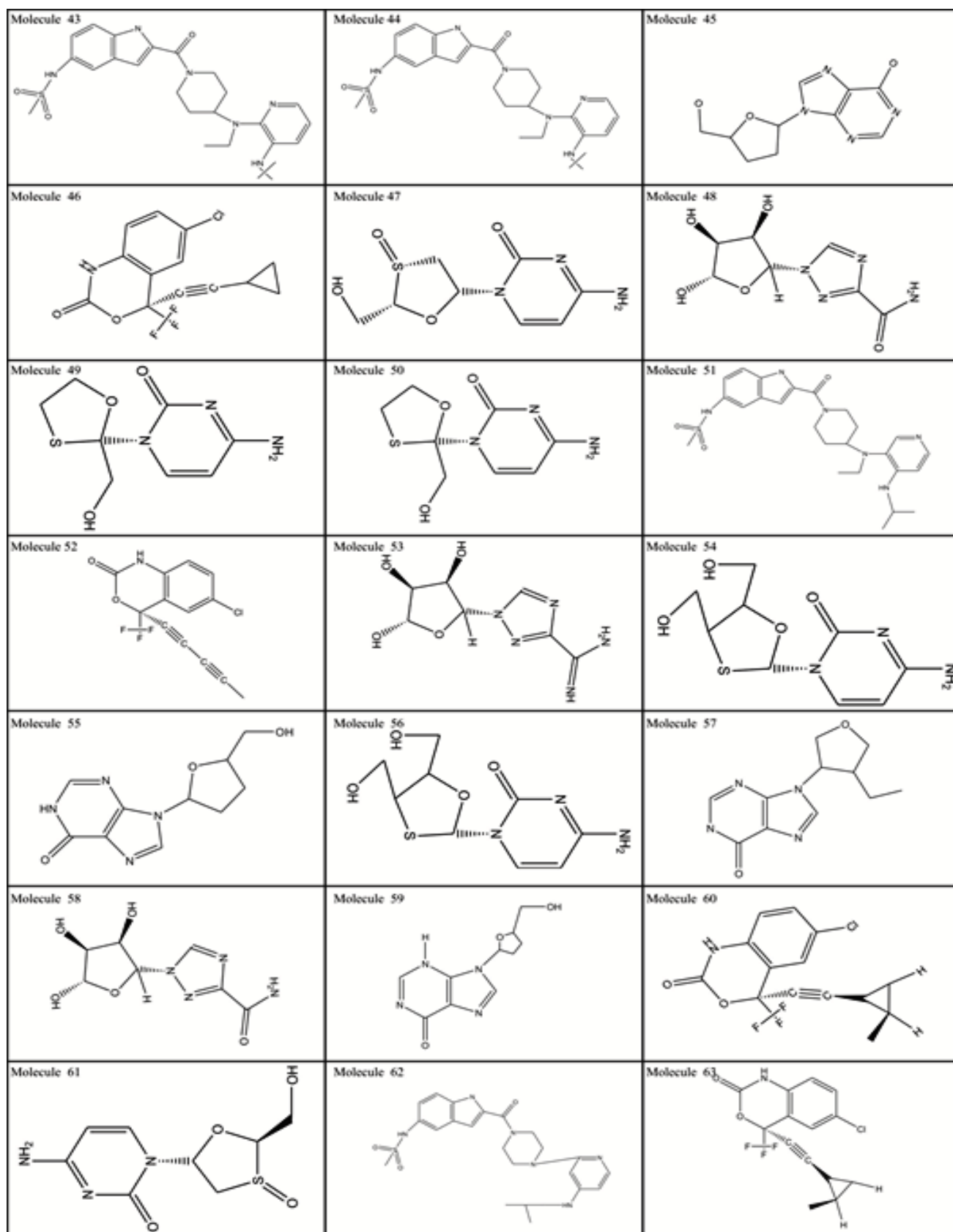**EXPERIMENTAL AND RESULTS:** Presently available drugs in market are shown in following **Table 1**:

Drug name	structure	Drug name	Structure
Zalcitabine		Nevirapine	
Efavirenz		Ribavirin	

Lamivudine		Stavudine	
Didanosine		Zidovudine	

Proposed structures of drugs shown in following **Table 2**:

Molecule 43	Molecule 44	Molecule 45
Molecule 46	Molecule 47	Molecule 48
Molecule 49	Molecule 50	Molecule 51
Molecule 52	Molecule 53	Molecule 54
Molecule 55	Molecule 56	Molecule 57
Molecule 58	Molecule 59	Molecule 60
Molecule 61	Molecule 62	Molecule 63

Molecule 85, Molecule 86, Molecule 87, Molecule 88, Molecule 89, Molecule 90, Molecule 91, Molecule 92, Molecule 93, Molecule 94, Molecule 95, Molecule 96, Molecule 97, Molecule 98, Molecule 99, Molecule 100, Molecule 101, Molecule 102, Molecule 103, Molecule 104, Molecule 105
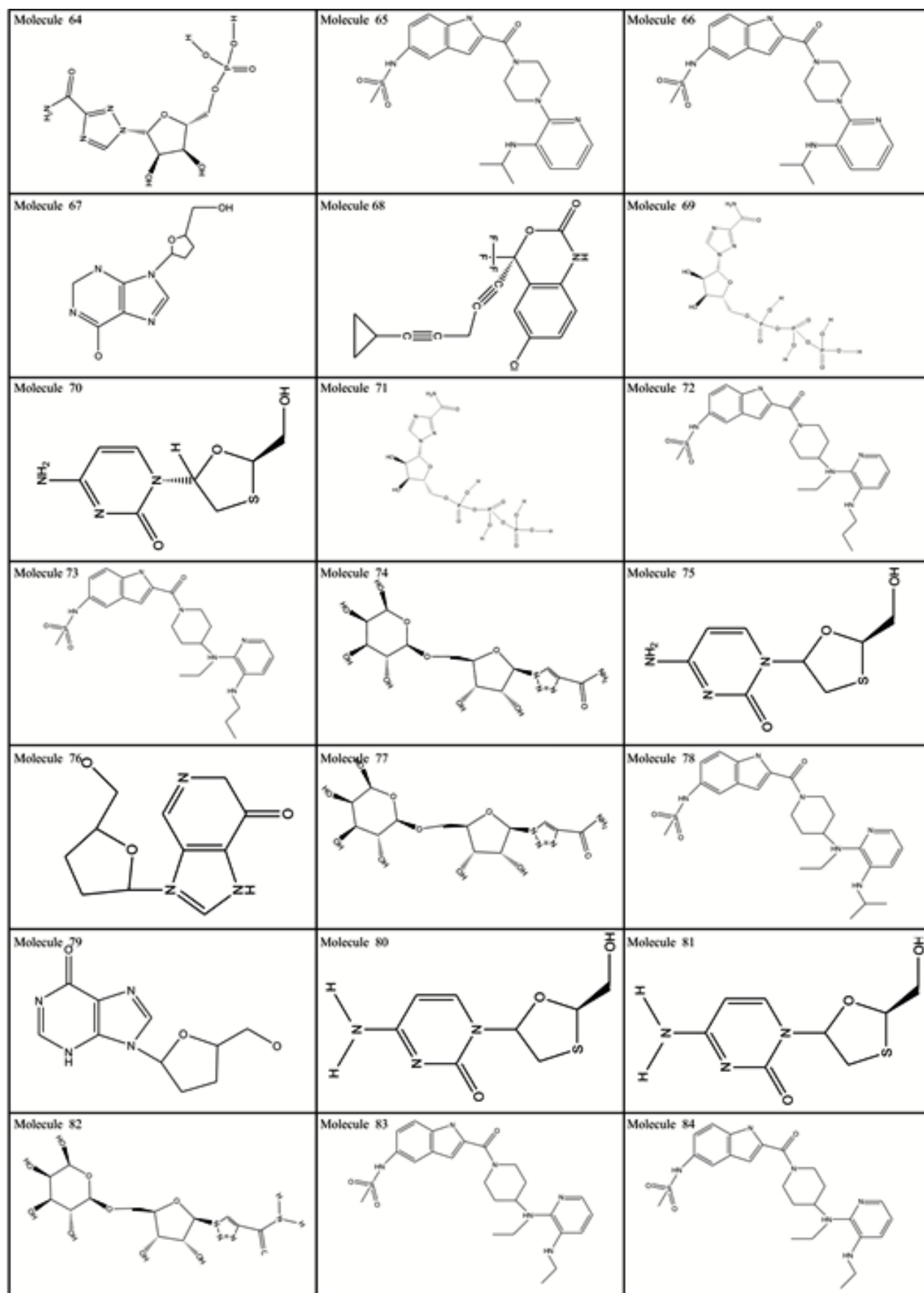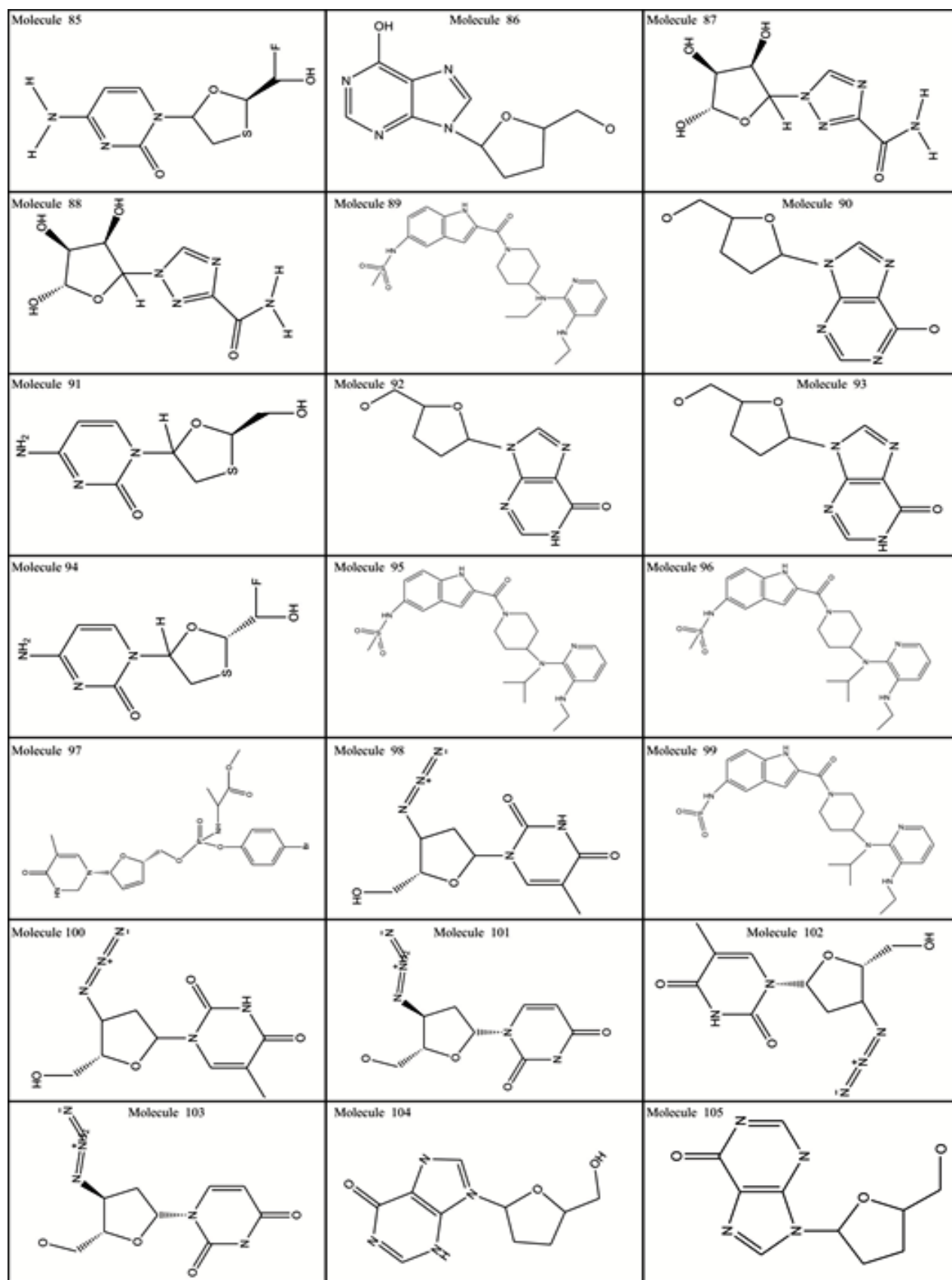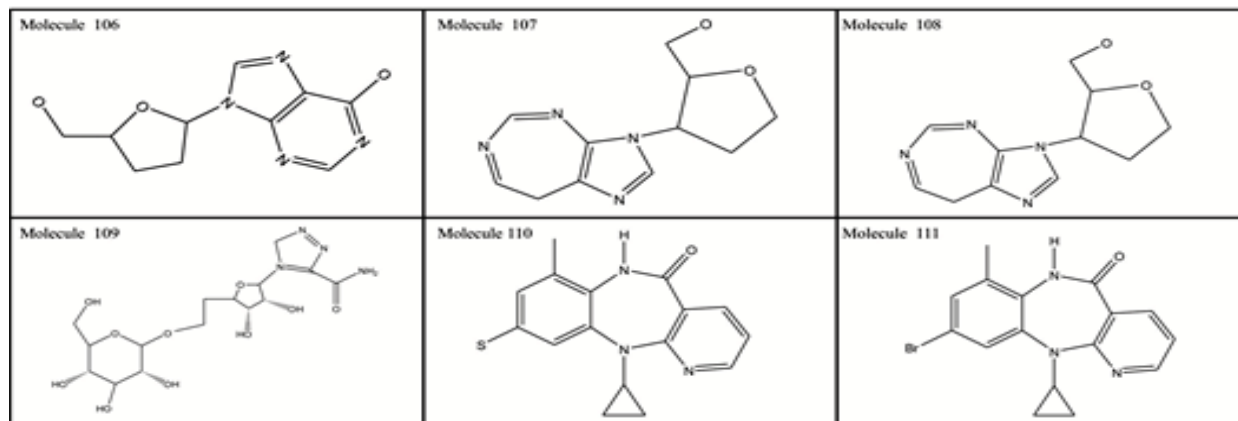
The proposed Structures of molecules where drawn in Argus lab and saved in MDL mole format. The following **figure 7** is a screenshot of Argus lab.
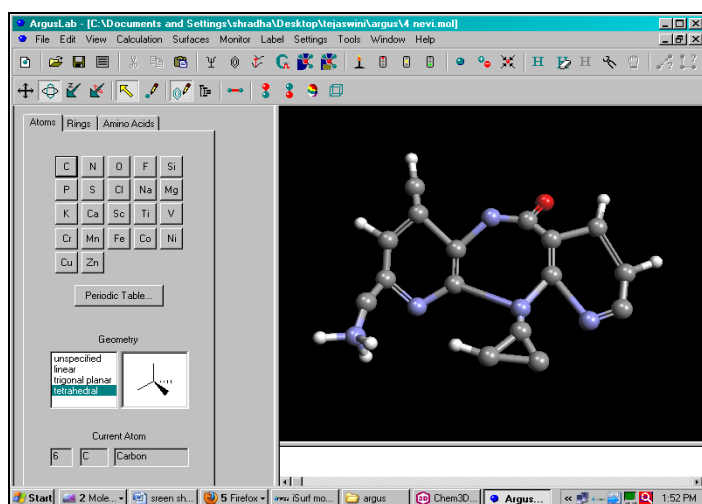

FIGURE 7: STRUCTURE DRAWING IN ARGUS LAB SOFTWARE.

Energy minimization is done in Chemoffice 3D Ultra for getting structure with minimum energy conformation.
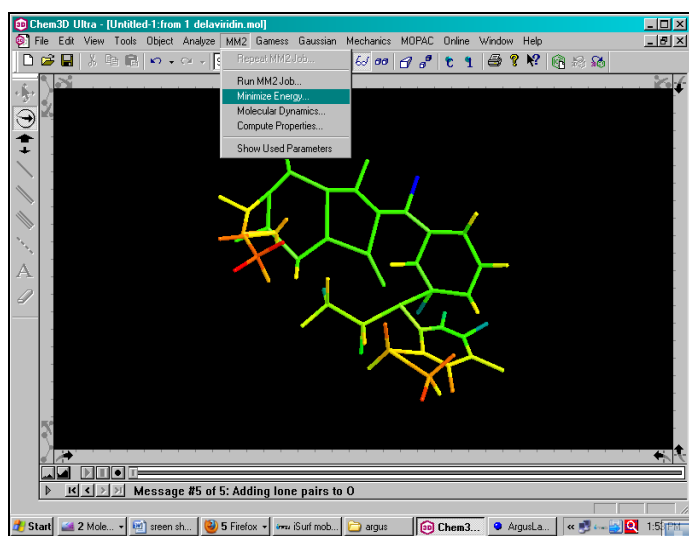

FIGURE 8: ENERGY MINIMIZING IN CHEMOFFICE 3D ULTRA

Homology modeling and its observations:

1. Material and Method:

- The sequence of HIV reverse transcriptase enzyme was obtained from Swiss-Prot database. Sequence of HIV reverse transcriptase in FASTA format.

- The next step in homology modeling was search for getting 3D structure of protein, for that online servers were opened and the sequence was given for 3D structure search. The results were obtained from all the three server used and in form of 3D structure. Screen shots for all the three servers are given in following **figures 10, 11, 12**.

Reverse Transcriptase [Human Immunodeficiency Virus 1]:

PISPIDTVPVTLKPGMDGPKVKQWPLTEEKIKALTEI CKEMEKEGKISKIGPENPYNTPVFAIKKKKDSTK WRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKS VTVLDVGDAYFSVPLDESFRKYTAFTIPSMNNETP GIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRIKNPE MVIYQYMDDLYVGSDLEIGQHRTKIEELRAHLL RWGFTTPDKKHQKEPPFLWMGYELHPDRWTVQPIE LPEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLC KLLRGAKALTEVVPLTEEAELELAENREILKTPVHG VYYDPSKDLVVEVQKQGQDQWTYQIYQEPFKNLK TGKYARKRSAHTNDVRQLAEVVQKVATESIVIWGKI PKFRLPIQRETWETWWMEYWQATWIPEWEFVNTP PLVKLWYQLEKDPIVGAETFYVDGAASRETKLGKA GYVTNRGRQKVVSLTETTNQKTELHAIHLALQDSG SEVNIVTDSQYALGIIQAQPDRSESEVVNQIIEELIKKE KVYLSWVPAHKGIGGNEQVDKLVSSGIRKVL FLDGIDKAQEEHERYHSN

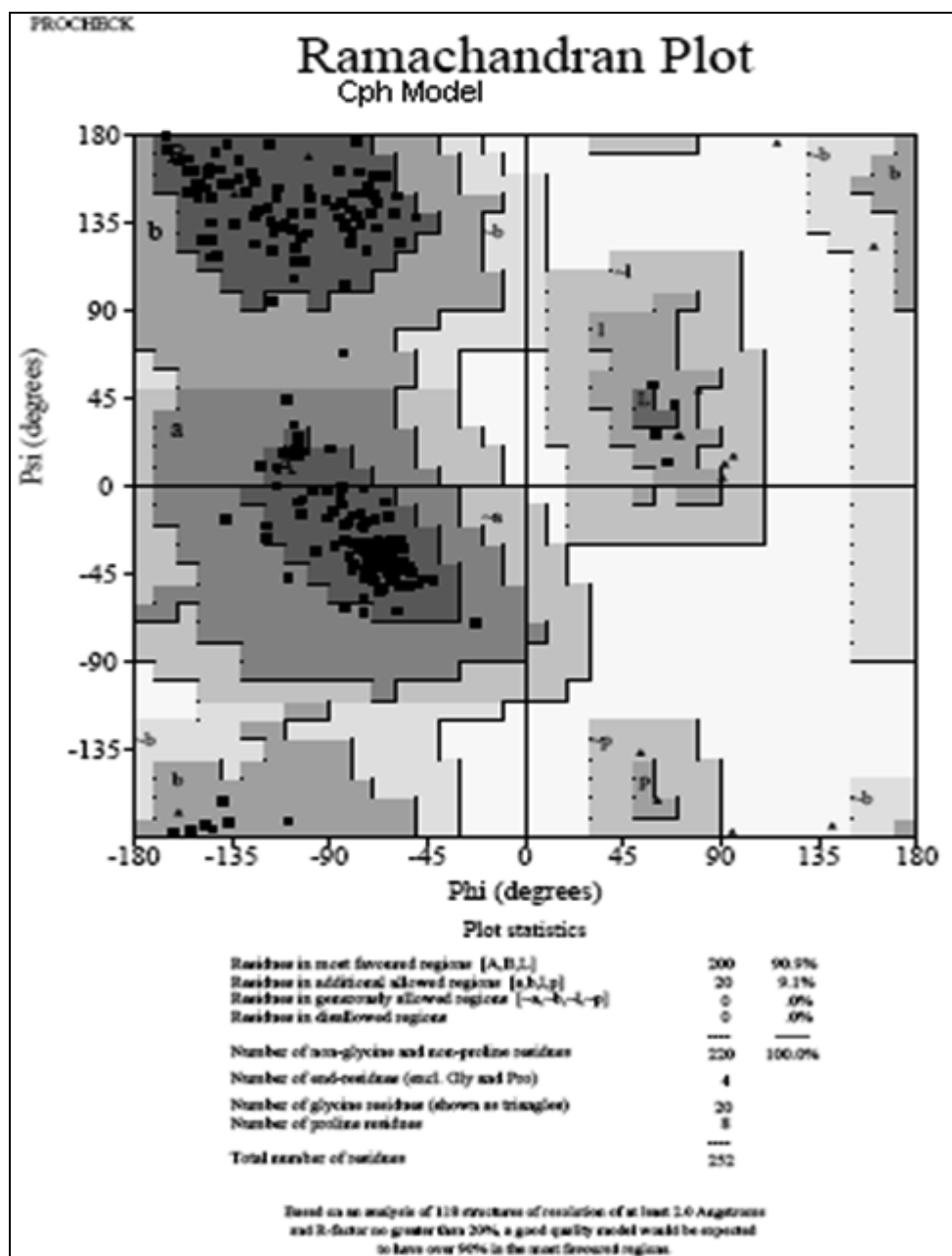**FIGURE 9: SEQUENCE OF HIV REVERSE TRANS-CRIPTASE ENZYME WITH SWISS-PROT**

Online server used in Homology modeling: Cph   Model Server:
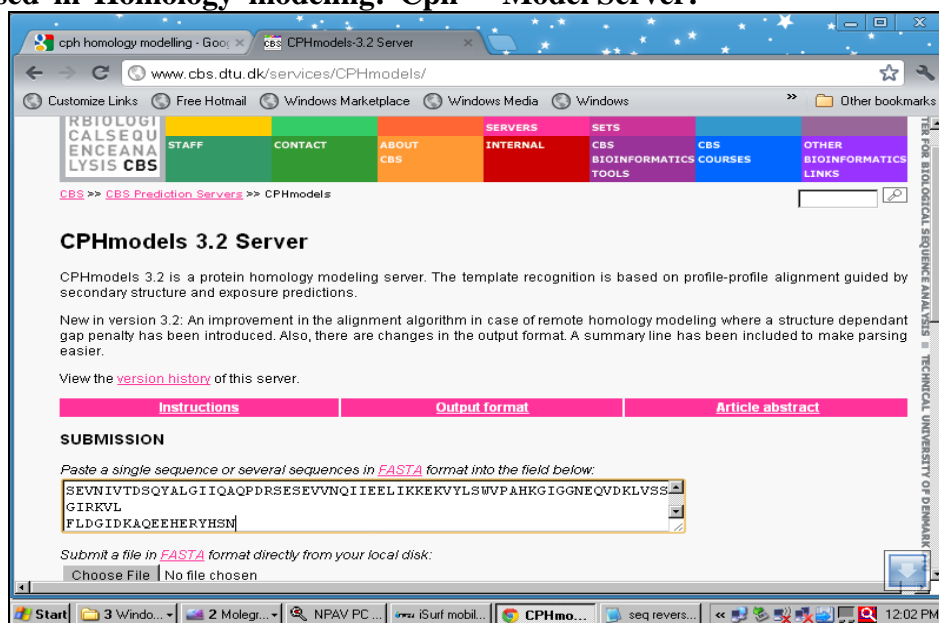
FIGURE 10: CPH MODEL SERVER

Phyre Model Server:
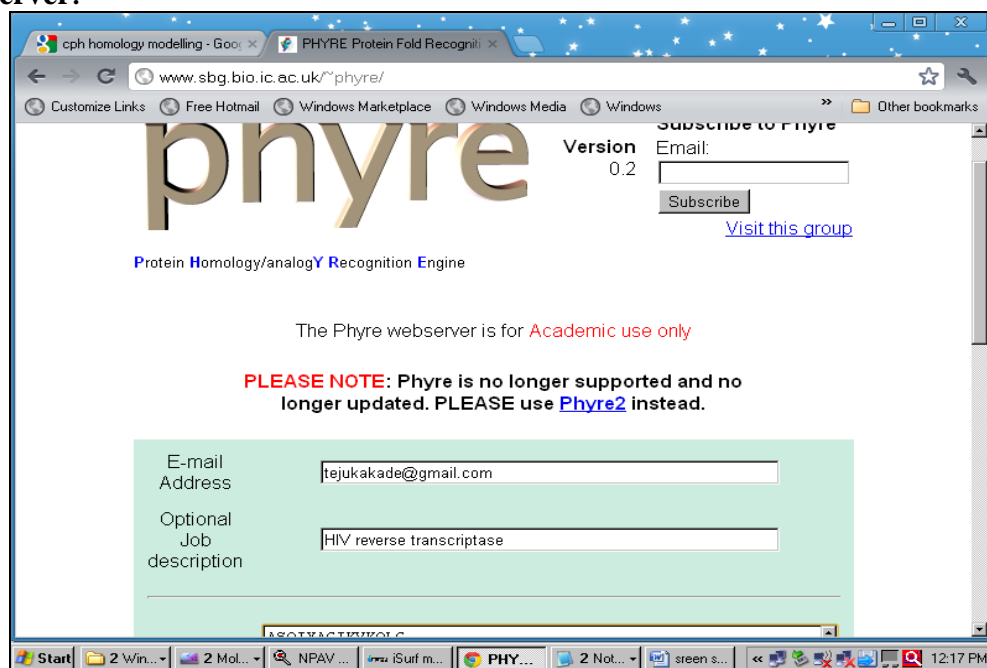
FIGURE 11: PHYRE MODEL SERVER.

----------<<< P R O C H E C K   S U M M A R Y >>>----------
CPH MODEL SERVER 2.0 252 RESIDUES
RAMACHANDRAN PLOT: 90.9% CORE   9.1% ALLOW   .0% GENER   .0% DISALL
GLY & CYS RAMACH: 2 LABELED RESIDUES (OUT OF 28)
CHI1-CHI2 PLOTS: 11 LABELED RESIDUES (OUT OF 150)
MAIN-CHAIN PARAMS: 6 BETTER   0 INSIDE   0 WORSE
SIDE-CHAIN PARAMS: 5 BETTER   0 INSIDE   0 WORSE
RESIDUE PROPERTIES: MAX.DEVIATION: 12.2   BAD CONTACTS: 4
BOND LEN/ANGLE: 9.6   MORRIS ET AL CLASS: 1 2 2 3 CIS-PEPTIDES   G-
FACTORS   DIHEDRALS: -.50 COVALENT: -.53   OVERALL: -.49
M/C BOND LENGTHS: 81.3% WITHIN LIMITS 18.8% HIGHLIGHTED
M/C BOND ANGLES: 68.4% WITHIN LIMITS 31.6% HIGHLIGHTED   4 OFF GRAPH
PLANAR GROUPS: 50.9% WITHIN LIMITS 49.1% HIGHLIGHTED   16 OFF GRAPH

FIGURE 6: PROCHE SUMMARY

Ramchandran Plot:

FIGURE 12: THE RAMCHANDRAN PLOT

Finally the 3D structure of HIV reverse transcriptase enzyme was obtained by An *In-silico* homology modeling technique and is shown in Figure and secondary structure of HIV reverse transcriptase enzyme is shown in **Figure 13**.
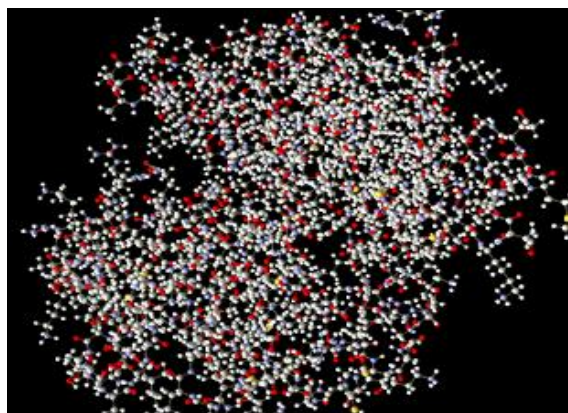

FIGURE 13: 3D STRUCTURE OF HIV REVERSE TRANSCRIPTASE ENZYME BY A HOMOLOGY MODELING TECHNIQUE

FIGURE 14: SECONDARY STRUCTURE OF AMPA RECEPTOR

2. **Docking:**

i. **Material and method:**

- Docking studies were performed using Molegro Virtual Docker MVD 2007 software.

- Firstly the protein molecule was imported from source in MVD2007; structure of protein molecule is shown in Figure 15.

- Then the surfaces was created of protein molecule by selecting create surface option for docking.

- Then next step was detection of cavities for binding, there should be at least 5 cavities and maximum 15 cavities. The Figure 10 shows protein molecule with cavities for docking.
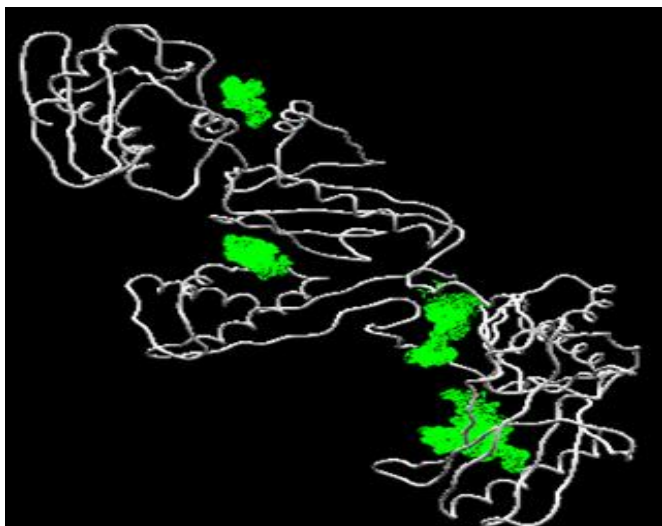

FIGURE 15: PROTEIN WITH CAVITIES FOR DOCKING

- Next step was import of SDF file of drug molecule on screen.

- Docking wizard was then selected and clicked on select molecule, the radius was increased up to 25 and the protocol was run.

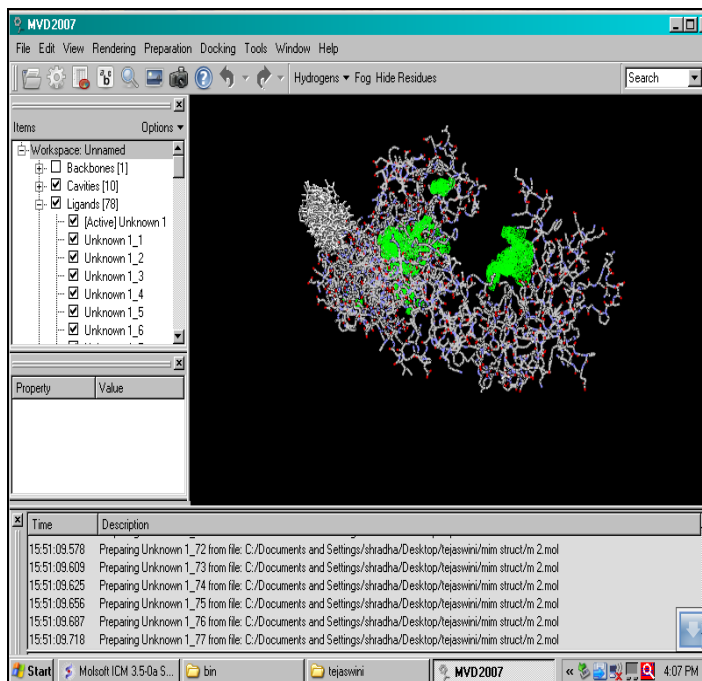- The running screen shot of Molegro virtual docker is shown in Figure 16 and 17.

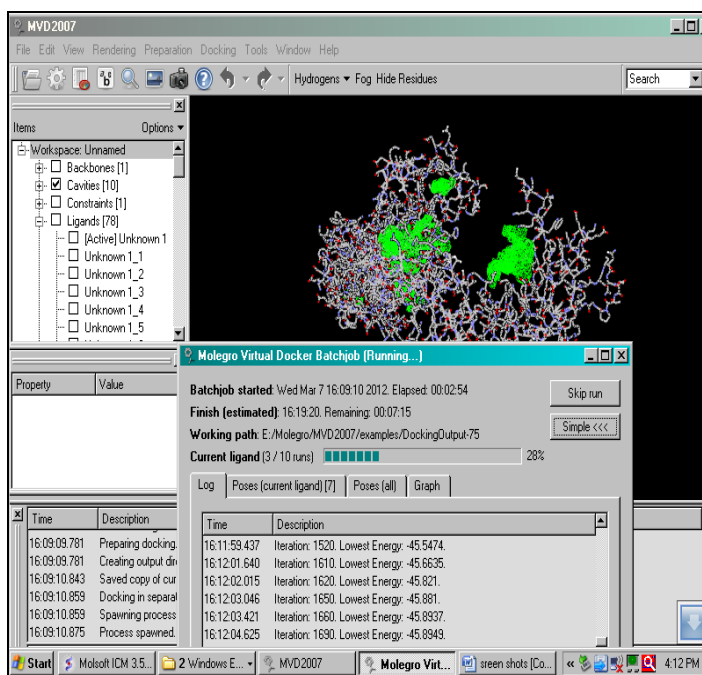**FIGURE 16: PROTEIN STRUCTURE IN MOLEGRO VIRTUAL DOCKER**

FIGURE 17: THE RUNNING SCREEN SHOT OF MOLEGRO VIRTUAL DOCKER

This much work has been performed on workstation of Molegro Virtual Docker (MVD). All most all molecules of test set were found with good binding or docking score and few molecules found with great binding score. Tables having molecule ID and its binding score is given in result. Also, presently available drugs with its binding score are given in result.

RESULT AND DISCUSSION: The docking study was carried out by taking the reverse transcriptase enzyme, this reverse transcriptase receptor contains A, B. It contains Gly231, Asp 186, Met 184, Cys 421 and Trp229. The drug molecule at binding site with amino acid residues of Reverse Transcriptase is given shown figure18 below
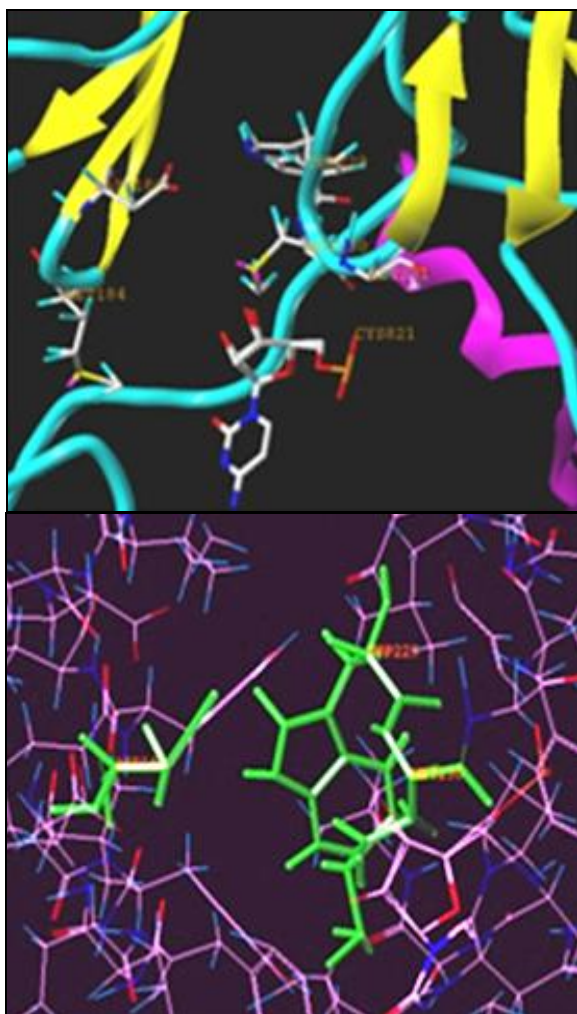

FIGURE 18: SHOWING AMINO ACID RESIDUES AT WHICH DRUG BINDS IN ACTIVE SITE

The docking study was carried out by taking the reverse transcriptase enzyme which was crystallized by homology modelling method and was then confirmed by plotting Ramachandran plot. Didanosine, Efavarine, Lamivudine, Nevirapine,

Ribaverine, Zalcitabine, Stavudine, Zidovudine are the reported drug. This reverse transcriptase receptor contains A, B. All the above mentioned drugs bound in the A and B chain. The active site of the drug is determined for the purpose of docking the set of ligands. It contains Gly231, Asp 186, Met 184, Cys 421 and Trp 229. The considered drugs were docked against this particular receptor.

The existing inhibitors (drugs available in market for treatment of AIDS) having activity on same receptor had values higher that -25Kcal/Mol. All such drugs with docking score are given in **table 3**.

TABLE 3: DRUGS PRESENT IN MARKET WITH ITS BINDING / DOCKING SCORE

Drugs	Binding energy / Docking score (Kcal/mol)
Didanosine	-53.2699
Efavirenz	-58.6239
Lamivudine	-46.4597
Nevirapine	-53.177
Ribavirin	-39.711
Zalcitabine	-52.9785
Stavudine	-25.2098
Zidovudine	-40.7162

All the molecules of test set were docked with HIV reverse transcriptase. Nearly all of them were found active but varying in magnitude of energy. The docking values are listed in Table 4 .The docking values more than -25Kcal/mol will give potentially active compound. According to this; all the molecules of test set can give reasonable activity still the molecules with best docking score are listed in **table 4.**

TABLE 4: MOLECULES FROM TEST SET WITH DOCKING SCORE

Molecule No.	Binding energy / Docking score (Kcal/mol)
Molecule 05	-112.651
Molecule 11	-107.982
Molecule 22	-94.4041
Molecule 27	-137.299
Molecule 48	-100.336
Molecule 55	-106.724
Molecule 70	-134.144
Molecule 77	-109.634

In the MVD docking result, the molecule 05, 11, 22, 27, 48, 55, 70, 77 (structure given in table no 4) gave the most negative score of -112.65 Kcal/mol, -107.98 Kcal/mol, -94.40 Kcal/mol, -137.29 Kcal/mol, -100.33 Kcal/mol, -106.72 Kcal/mol, -134.14 Kcal/mol, -109.63 Kcal/mol respectively.

Out of all, Molecule 27 and Molecule 70 found to have most negative score and which is having the highest interaction with receptor, while in Molegro result also molecule 27 and 70 are coming in the best ranking list.

CONCLUSION: HAART (Highly Active Anti Retro Viral Therapy) is one of the most promising treatment method for AIDS where different combination of Reverse Transcriptase inhibitor (RTI) or Protease inhibitors are prescribed to the patient to prevent the virus in different stages of its life cycle. This kind of computational approach will help us to extract the important descriptors from the existing drug series to add up a new molecule in the RTI drug list and to combat with fast drug resistance property of the virus. The docking study will help us to check out the binding capability of the drug. Just the addition of a single drug in the series often results in offshoot of the drug combinations for HAART therapy.

FUTURE SCOPE:

- QSAR studies of the same test set to get more refined molecules.

- Building of few more sets with bioisosteric changes.

- Laboratory synthesis of molecules of test set (27 and 70) and biological evaluation of same.

REFERENCES:

1. Perun, T.J., Propst, C.L., Computer aided drug design, Methods and applications, Infroma Health care, (1989), 1-5.
2. Nogrady, T., Weaver, D.F., Medicinal chemistry; A molecular biochemical approach, 3rd Edi, Oxford University press, US. (2005), 5-7.
3. Fischer, J., Ganellin, R., Analogue based drug discovery, Wiley- VCH, xxi. (2006),
4. Nogrady, T., Weaver, D.F., Medicinal chemistry; A molecular biochemical approach, 3rd Edi, Oxford University press, US. (2005), 8-10.
5. Gasteiger, J., Engel, J., Chemoinformatics, Wiley- VCH GMBH and KGaA, 518.
6. Xiong, J., Essential Bioinformatics, Cambridge university press, New-York, 223.
7. Siew, N., Fischer, D., Convergent evolution of protein structure prediction and computer chess tournament: CASP, Kasparov and CAFASP, IBM system journal, Vol-40, (2001), 315.
8. Dahiyat, B. I., In silico design for protein stabilization, current opinion in Biochemistry, (1999), 387-390.
9. Philip, E., Bourne. Weissig, H., Structural Bioinformatics, Wiley-Liss, Inc., (2003), 507-519.
10. Taylor, Jewsbury, Essex, A review of protein-small molecule docking methods.
11. Krovat, E. M., Steindl, T., Langer, T., Recent Advances in Docking and Scoring, Current Computer-Aided Drug Design, Vol-1, (2005), *93*-102.
12. Dale, M. M., Rang, H. P., Ritter, J. M., Moore, P. K., Pharmacology, 5th Edi, Elsevier India Pvt. Ltd., New-Delhi, 463-467.
13. www.wekipedia.org
14. NCBI: http://www.ncbi.nlm.nih.gov/
15. BLAST: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
16. Swiss-Prot: http://www.expasy.org/sprot/
17. PDB: http://www.rcsb.org/pdb/home/home.do
18. Cph protein server: http://www.cbs.dtu.dk/services/CPH models/
19. Phyre protein server: http://www.sbg.bio.ic.ac.uk/phyre/ html/help.html
20. 3d Jigsaw protein server: http://bmm.cancerresearchuk.org/ ~3djigsaw/
21. Laskowski, R. A., Macarthur, M. W., Smith, D. K., Jones, D.T., Molegro virtual docker MVD-2007, user manual, Molegro Aps, Denmark, 10.
22. John H. Block, John M.Beale, Wilson and Gisvolds's textbook of Organic medicinal and Pharmaceutical chemistry, 11th edition.