CHAPTER 1

INTRODUCTION

1.1 Research rationale

1.1.1 Molecular Modeling: The most potentially method for drug design

There are two basic approaches in developing of antiviral agents for therapy of AIDS (Acquired Immune Deficiency Syndrome): the empirical and the rational approach. Most of drugs in clinical use have been provided by the empirical approach. The strategy of the approach is to identify the lead compound and subsequent modify repeated structure of the lead compound, evaluate activity to produce and achieve the optimal improvement. Random screening or structural modifying of a critical metabolite is the way to obtain the lead compound. AZT, actives against HIV-1 (Human Immunodeficiency Virus type 1), the virus responsible for AIDS, is an example of the lead compound. The second approach for drug development, which is more challenge, is the rational approach. This approach requires a sophisticated knowledge of the target structures as well as their biochemical pathways. The interaction between target receptor and a compound is determined by identification of their structural, spatial and electronic requirements. Rational drug design is a paradigm of this approach. This application is the most potentially method to design the specific drug based on the knowledge of how the atoms are arranged by using molecular modeling and interactive computer graphics. From theoretical point of view, understanding the details of the activity of proteins requires an investigation of the dynamics of the structural fluctuations and their relation to reactivity and conformational change. Studies of the dynamics are utilized for determining thermodynamic properties as well as for providing information concerning their motions. Stability of proteins and their interactions with drugs and ligands are of special interest. Many theoretical methods are used as tools to investigate these properties. **(1, 2)**

Molecular modeling was considering as a way to mimic the behavior of the molecule and molecular systems. It is associated with the computer modeling as

a mechanical tool. Molecular modeling including both quantum mechanics and molecular mechanics and other computer-based methods is the most powerful method to understand and predict the behavior of molecular systems particularly in the drug design development. **(3)**

1.1.2 HIV-1 Integrase: The third target for AIDS therapy

Reverse transcriptase and protease are served to be the first two pharmacological targets for AIDS therapy as their important roles in viral replication cycle. **(4)** The first group of drug to be used, the inhibitor that inhibits reverse transcriptase action, becomes the drugs approved for treatment of AIDS. Consequently, protease inhibitors were accepted to be the second drugs group. Due to the rapid mutation of the two enzymes among these two drugs, the combination therapies in which two or more drugs have been used simultaneously were introduced, known well as drug cocktails. However, the resistance of these two targets along with the efficacy of the existing drugs continues to occur rapidly. Therefore, novel drugs focusing at other targets for maximizing the anti-viral therapy are robustly required. **(4)**

HIV-1 integrase is a third viral enzyme for AIDS therapy. It plays a crucial role in the viral replication by catalyzing the incorporation of the reverse transcribed viral DNA into the human chromosome. The complete structure (full-length) of the HIV-1 IN composes of three domains. There is not any cellular analogue for this enzyme in human was found. Moreover, the full-length structure is still experimentally unavailable. These made this enzyme an ideal and attractive target for curative involvement. **(5)**

1.1.3 Aims of this study

In this thesis, the full-length model, which is currently experimentally unavailable, was theoretically synthesized. Various molecular modeling method including the MD (molecular dynamics) simulation, homology modeling and the molecular docking were applied to explore the dynamical behavior of HIV-1 integrase, as well as the interaction with viral and host DNA. Further details on each technique will be described in the following chapters.

1.2 AIDS

1.2.1 Historical Outline

1977 – 1978 The viruses were found in the United States, Haiti and Africa.

1981 The disease was first recognized and become a worldwide severe epidemic.

1982 The Centers for Disease Control and Prevention (CDC) officially named the condition AID.

1983 – 1984 The viruses responsible for weakening the immune system was identified as HIV by Luc Montagnier of France and Robert Gallo of the United States

It was believed that the HIV-1 virus, the human AIDS virus, is originally from African chimpanzees and being "crossed over" into the human population through blood contact when the chimpanzees were hunted as food **(6)**.

1.2.2 World epidemic situation

From the UNAIDS/WHO report update in 2004, it was shown that the number of people living with HIV/AIDS and the number of AIDS deaths continue to increase every year (see Table 1.1). The global situation of the HIV/AIDS was summarized in Table 1.2. There were an estimated 40 million people living with HIV/AIDS (37 million adults and 2.5 million children under 15). The report stated that there were the steady increases in the number of people living with HIV/AIDS, as well as number of AIDS death. The numbers continue to increase in several regions. Asia and the Pacific, Eastern Europe and Central Asia face the expanding epidemics, with the number of people living with the disease growing year by year. **(7, 8)**

1.2.3 HIV/AIDS Symptoms

HIV infection occurs in three main stages; acute infection, chronic infection and AIDS **(9)**. Detailed information of the infection can be clearly found in Figure 1.1 which plots between the relative values of the CD4 T-cells, antibody, HIV cytotoxic lymphocyte (CTL) and the HIV viral load versus time post infection. The first stage is the earliest and shortest of the HIV infection. During the first three

months, the viral load increases with the cytotoxic T-lymphocyte while the CD4 T-cell count drop in the absence of antibody. The symptoms are flu-like illness for three to six weeks after infection. Fever and fatigue may occur and last for a week or two. Other symptoms, such as sore throat, diarrhea, nausea, vomiting, etc., may or may not occur. The second stage; chronic infection, begins three to six months after a person gets HIV and lasts about 10 years. The CD4 count is gradually decreased and the viral load is maintained as for the releasing of the antibody. There are not any symptoms in this stage. The immune system continues to slowly run down even there are no symptoms. For final stage, this reached the stage of AIDS, the CD4 T-cell counts drop from 450 – 1,200 cells per microliter to 200 lower. Since AIDS itself has no symptoms, the disease symptoms are due to the devastation of the body immune system and the kind of infections that a person may have; *i.e.*, being tired all the time, swollen lymph nodes in the neck or groin, fever lasting than 10 days, etc.

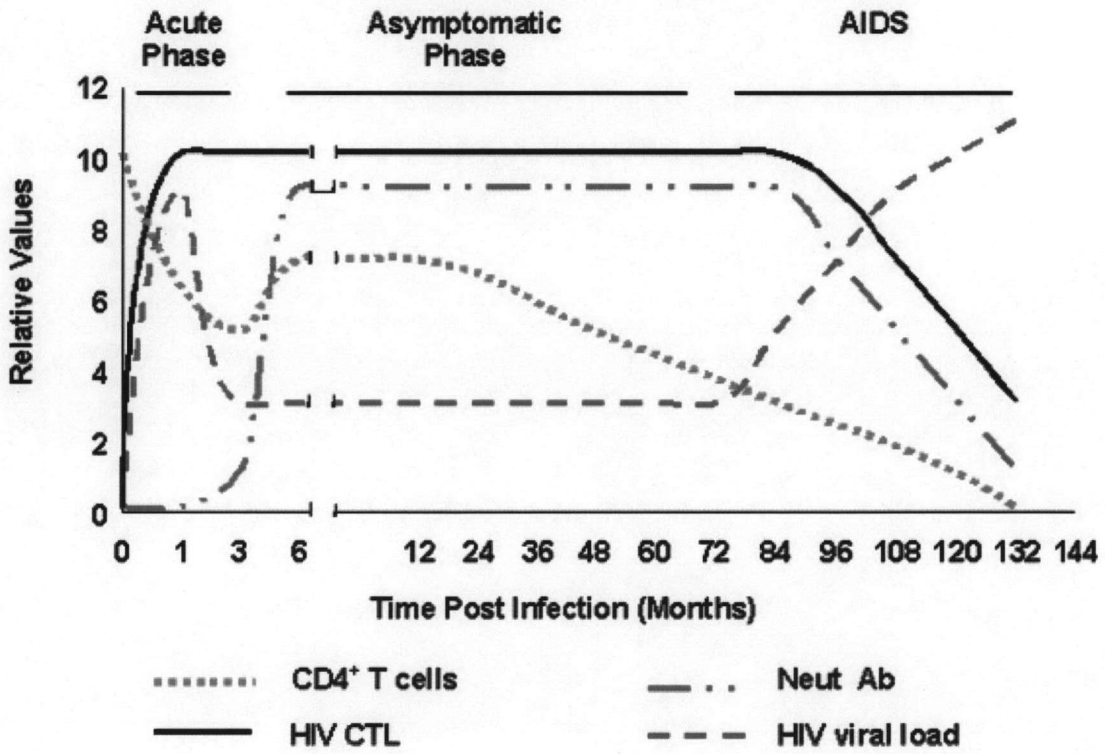


Figure 1.1 Stages of HIV/AIDS infection, acute, chronic (asymptomatic) and AIDS.

Table 1.1 Estimated number of people living with HIV/AIDS and number of deaths due to AIDS globally 1999 – 2003. **(7)**

Year	Estimated number (millions)	
	People living with HIV/AIDS	Deaths due to AIDS
1999	33	2.1
2000	35	2.3
2001	37	2.6
2002	39	2.8
2003	40	3.0

Table 1.2 Global summary of the HIV/AIDS (December 2003), the number is an average while the number in parenthesis is for range. The data is from UNAIDS 2004 report on the global AIDS epidemic. **(7)**

Number of people living with HIV/AIDS (millions)	Total	40 (34 – 46)
	Adults	37 (31 – 43)
	Children under 15 years	2.5 (2.1 – 2.9)
People newly infected with HIV in 2003 (millions)	Total	5.0 (4.2 – 5.8)
	Adults	4.2 (3.6 – 4.8)
	Children under 15 years	0.70 (0.59 – 0.81)
AIDS deaths in 2003 (millions)	Total	3.0 (2.5 – 3.5)
	Adults	2.5 (2.1 – 2.9)
	Children under 15 years	0.50 (0.42 – 0.58)

1.3 HIV

1.3.1 Historical Outline (10)

Two species of the virus have been recognized: HIV-1 and HIV-2. The differences are due to their genome and infectivity, HIV-1 is three time infective **(10)**. HIV-2 is mostly found to infect in Western Africa and India; HIV-1 is found to widespread in the rest of the world. The structure of the HIV-1 and its components including single strand viral genome RNA, three associated enzymes (RT, PR and IN), p24 capsid protein and p17 matrix protein, gp41 and gp120 virally membrane proteins, was shown in Figure 1.2.

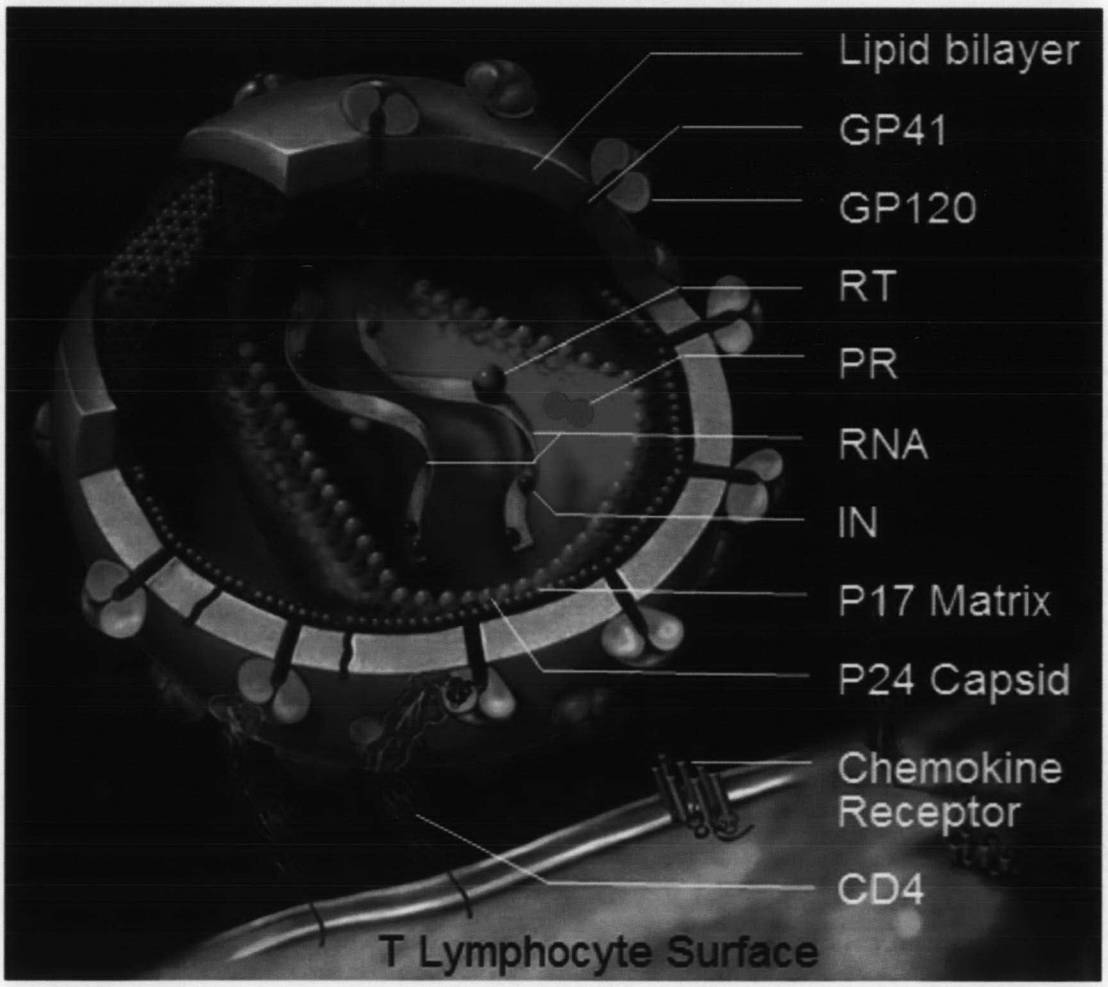


Figure 1.2 Structure of an HIV-1 virion and its component **(11)**

1.3.2 Biology of HIV

HIV belongs to the retrovirus family in which the genetic material is its RNA **(4, 12)**. The virus is a spherical in shape with a diameter of ~110 nanometers (see Figure 1.2). HIV-1 is considered to compose of ~9-kB RNA and 15 proteins (Figure 1.3).

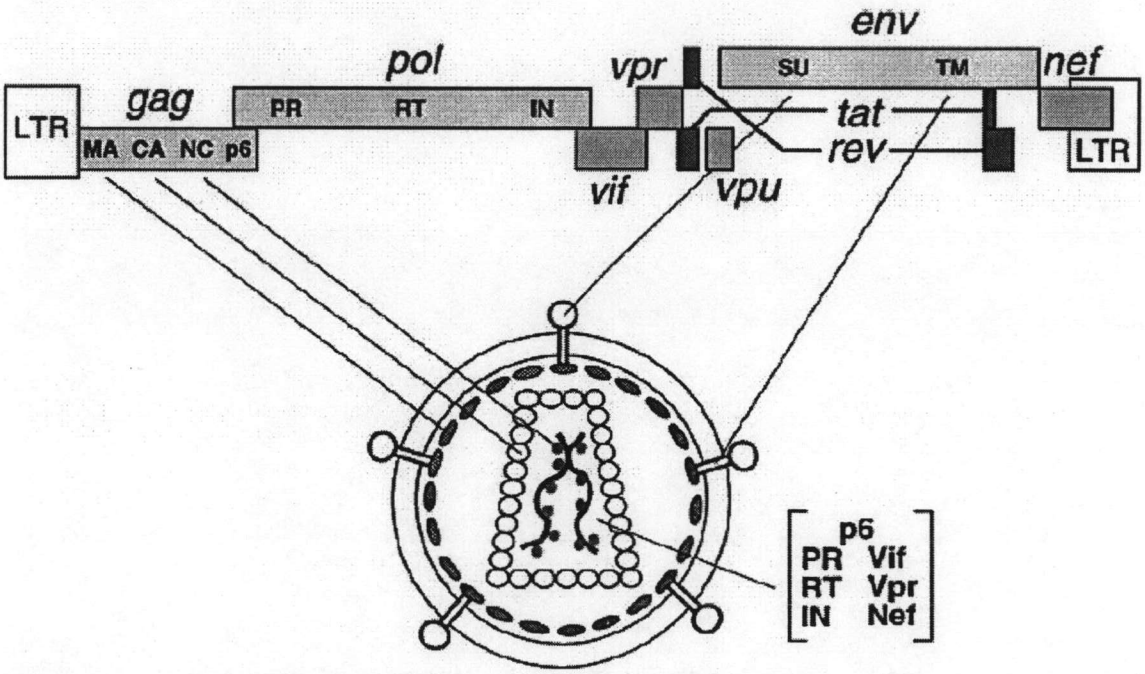


Figure 1.3 Organization of HIV-1 genome and virion **(12)**.

The proteins are from the major three genes, *gag* (Group specific Antigen gene), *pol* (Polymerase) and *env* (Envelope), which are common to all retroviruses. *Gag* provides four structural proteins; MA (matrix), CA (capsid), NC (nucleocapsid) and p6. *Pol* gene encodes the three important enzymes, RT (reverse transcriptase), PR (protease) and IN (integrase). *Env* consists of gp41 transmembrane portion, TM and gp120 surface molecule, SU. There are six additional proteins, three (*Vif*, *Vpr* and *Nef*) are found in viral particle, two other regulatory proteins, *Tat* and *Rev* as for transactivator of transcription and regulates late gene expression, respectively, and a *Vpu* in which assists in assembly of the virus. LTR represents long terminal repeat at 3' and 5' ends of the viral genome.

1.3.3 Replication Cycle

Replication of HIV-1 within the host cells begins by entering of the viral cell to the host cell through the particular receptor called CD4 T-lymphocyte or CD4 cell **(13)**. The replication cycle can be divided into 15 steps as depicted in Figure 1.4.

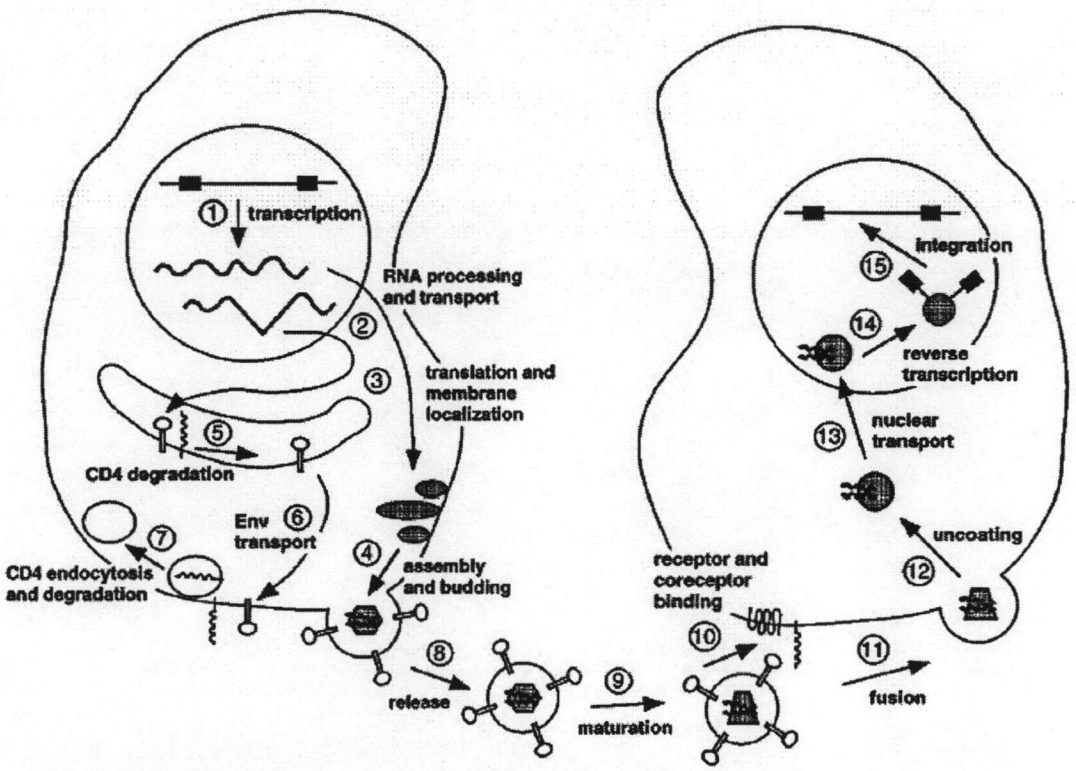


Figure 1.4 The HIV infections cycle **(13).**

- Step 1 Expression of viral transcripts at 5' long terminal repeat (LTR), enhances by *Tat*.
- Step 2 Transportation of a set of spliced and genomic RNA from nucleus to cytoplasm, regulates by *Rev*.
- Step 3 Translation of viral mRNA in the cytoplasm and localization of the *Gag* and *Gag-Pol* polyprotein to the cell membrane.
- Step 4 Assembly of the core particle from the *Gag*, *Gag-Pol* polyprotein (later posses to MA, CA, NC, p6, PR, RT and IN), *Vif*, *Vpr*, *Nef* and genomic RNA, an immature virion begins to bud from the cell surface.
- Step 5 Providing of SU and TM proteins for the outer membrane coat.
- Step 6 Transportation of the *Env* to the cell surface.

- Step 7 Promotion of endocytosis and degradation of CD4 by *Nef*.
- Step 8 Budding and releasing of the particle from cell surface coated with SU and TM.
- Step 9 Maturation of the virion includes proteolytic processing of *Gag* and *Gag-Pol* polyprotein, catalyzes by PR.
- Step 10 Infection to another cell by mature virion.
- Step 11 Fusion and entry of the core into the cell.
- Step 12 Uncoating of the virion core, exposing of the viral nucleoprotein complex, which are MA, RT, IN, *Vpr* and RNA.
- Step 13 Transportation of the nucleoprotein to the nucleus
- Step 14 Reverse transcription of the viral RNA into a partially duplex DNA, catalyzes by RT
- Step 15 Integration of the viral DNA into host chromosome, catalyzes by IN.

1.4 Treatment and FDA Approved Drugs

1.4.1 Historical Outline

When AIDS first surfaced, none of medicines was found to combat the original immune deficiency. During the past 10 years, drugs for fighting against the disease have been developed by many researchers in many areas. Following, a complex regimen of drugs that attack HIV at various stages in its infection cycle, known as antiretroviral drugs, is given to patients nowadays. The drugs include Reverse Transcriptase inhibitors (RTIs), Protease inhibitors (PIs) and Entry inhibitors in which being approved by the Food and Drug Administration (FDA) as shown in Table 1.3. The treatment for HIV, so called antiretroviral therapy, by using the existing drug comes with common but serious side effects include extreme nausea and diarrhea, liver damage and failure, and jaundice. There is not any evidences indicate that the HIV infection can terminate the infection.

1.4.2 Progress of inhibitors design for anti-retroviral drugs

Inhibitor design for anti-retroviral drugs in both experimental and theoretical areas attracts interest researchers worldwide in broad areas. For example,

organic chemists try to synthesize new substances, extract and isolate the bioactive compound from the natural product **(14-16)**. Biologists and biochemists attempt to test the activity of the ligand against the virus. Theoreticians introduce new tools to design new potent ligand by using many techniques such as molecular docking, pharmacophore searching, *etc.*, **(17-20)** and try to relate the structure of the ligands and their activity using QSAR, COMFA, COMSIA methods **(21, 22)**. Furthermore, several pharmacological enzymes both in free and complex forms with the existing drugs have been studied and compared using MD simulations in order to understand their structure, dynamical behavior and various properties in molecular levels **(23, 24)**. The interactions between enzyme and drugs were precisely calculated using different levels if quantum chemical calculations **(25, 26)**. The main goal of all attempts is to retrieve the powerful novel anti-retroviral drug to get over the AIDS disease. Detailed summary of the study on the retroviral inhibitors design using various methods was displayed in Figure 1.5.

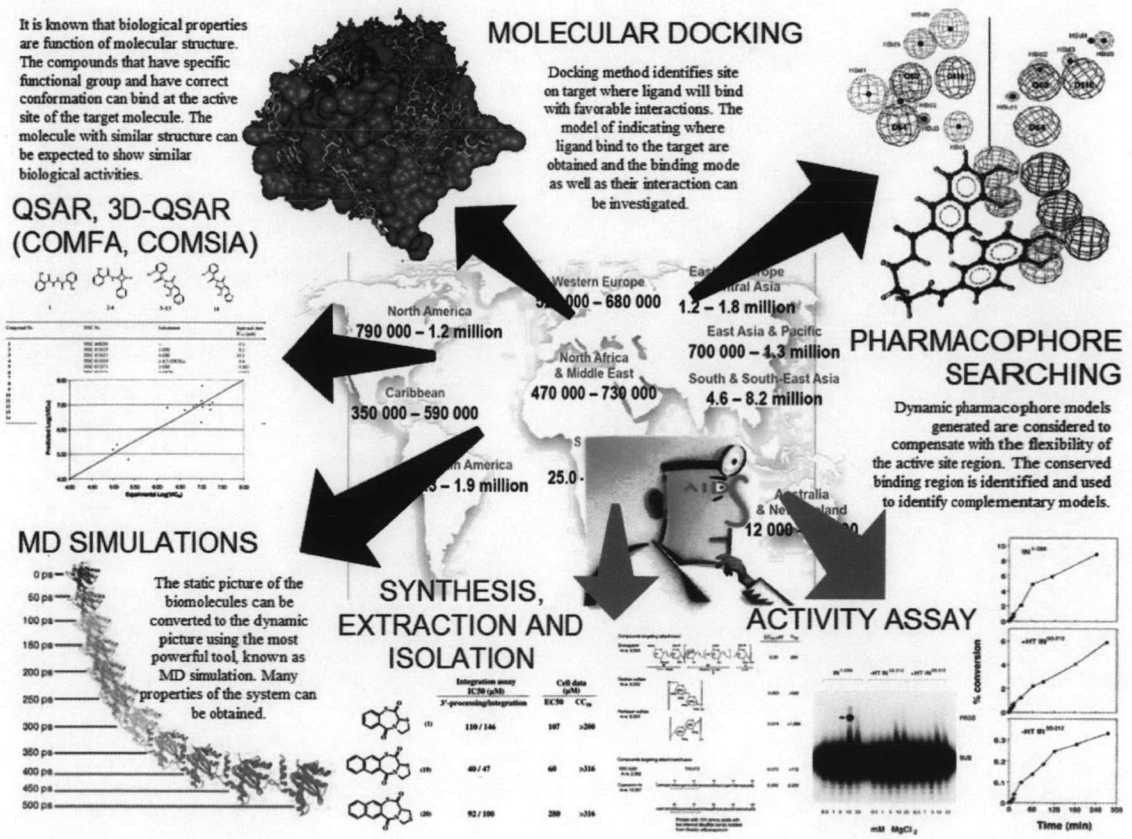


Figure 1.5 Study of retroviral inhibitors design using various experimental and theoretical techniques.

Table 1.3 FDA approved anti-HIV drug **(4, 27).**

FDA approved anti-HIV drug	Brand name	Generic name	Abbreviation
Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTI)	Combivir	Zidovudine + Lamivudine	AZT + 3 TC
	Emtriva	Emtricitabine	FTC
	Epivir	Lamivudine	3TC
	Epzicom	Abacavir + Lamivudine	ABC + 3TC
	Hivid	Zalcitabine	ddC
	Retrovir	Zidovudine	AZT or ZDV
	Trizivir	Abacavir + Zidovudine + Lamivudine	ABC + AZT + 3TC
	Truvada	Tenofovir + Emtricitabine	TDF + FTC
	VIDEX	Didanosine	ddI
	VIDEX EC	Didanosine: delayed-release capsules	ddI
	Viread	Tenofovir DF	TDF
	Zerit	Stavudine	d4T
	Zerit XR	Stavudine: delayed-release	d4T
	Ziagen	Abacavir	ABC

Table 1.3 **(cont.)**

FDA approved anti-HIV drug	Brand name	Generic name	Abbreviation
Non Nucleoside Reverse Transcriptase Inhibitors (nNRTI)	Recriptor	Delavirdine	DLV
	Systiva	Efavirenz	EFV
	Viramune	Nevirapine	NVP
Protease Inhibitors (PI)	Agenerase	Amprenavir	APV
	Crixivan	Indinavir	IDV
	Fortovase	Saquinavir (soft gel cap)	SQV (SGC)
	Invirase	Saquinavir (hard gel cap)	SQV (HGC)
	Kaletra	Lopinavir/Ritonavir	LPV/r
	Lexiva	Fosamprenavir	FPV
	Norvia	Ritonavir	RTV
	Reyataz	Atazanavir	ATV
	Viracept	Nelfinavir	NFV
Entry / Fusion Inhibitors	Fuzeon	Enfuvirtide	ENF

1.5 HIV-1 Integrase

1.5.1 Historical Outline

Recently, many researchers pay attention to studying HIV-1 IN enzyme as a novel and attractive target for AIDS therapy, since, the enzyme plays important role in the replication cycle of the virus by carrying out the integration of the viral DNA into the host chromosome. Furthermore, there is not any cellular homologue for this enzyme **(28)**.

1.5.2 Structure and Function

HIV-1 IN, a 288-amino acid full-length, consists of three domains; N-terminal, core and C-terminal domains **(29)**. The studies revealed that all three domains are required for the integration process but the core domain only response for the disintegration reaction. The experimental structures of each domain which are resolved by X-ray crystallography and NMR technique are available at the Protein Data Bank **(30)**. Figure 1.6 displays schematic representation of the full-length HIV-1 IN and its domain.

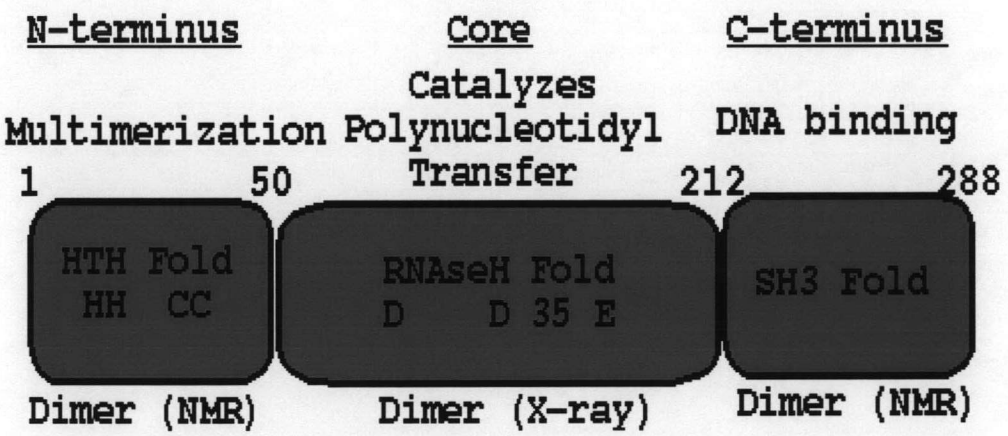


Figure 1.6 Schematic representation of the full-length HIV-1 IN **(28)**.

1.5.2.1 N-terminal domain

N-terminal domain, residues 1 – 50, contains an HHCC zinc binding motif **(30)**. The structure resolved using NMR technique shows that the structure forms dimer and composes of only α-helices, which is similar to helix-turn-

helix (HTH) DNA binding domain (Figure 1.7) **(31)**. Previous studies demonstrated that binding of zinc in this domain promote the tetramerization of the enzyme and increased enzymatic activity **(32, 33)**.

Figure 1.7 N-terminal domain of the HIV-1 IN **(31)**.

1.5.2.2 Catalytic core domain

The catalytic core domain, residue 51 – 212, contains a conserved DD-35E, which are two aspartate residues (Asp64 and Asp116) and a glutamate residue (Glu152) as an active site **(29, 34)**. The structure resolved from x-ray crystallographic study shows that the structure consists of a central stranded β-sheet surrounding with helices and form dimer as can be seen in Figure 1.8 **(35)**. This domain was responsible for polynucleotidyl transfer reaction and was found to bind Mg^{2+} **(35)**. (Figure 1.9 a) The mechanism of this reaction is still not clear. A possibly catalytic mechanism of the *E. Coli* polymerase I has been proposed **(36)** (Figure 1.9 b). The first Mg^{2+} stabilizes the pentacoordinate transition state and the leaving alkoxide group. The second Mg^{2+} activates a nucleophilic attack on the P atom of the terminal nucleotide carried out by hydroxide ion. The position of the second Mg^{2+} is still experimentally unresolved. This led to the thought that this ion may come together with the binding of the DNA **(37)**.

Figure 1.8 Core domain of the HIV-1 IN **(35)**.

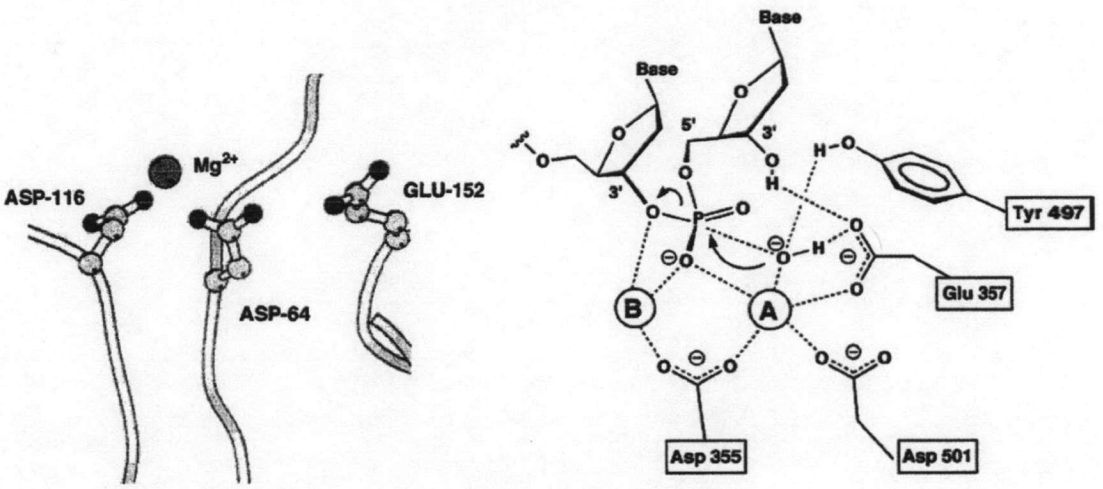


Figure 1.9 (a) Two conserved aspartate residue (Asp64, Asp116) and a glutamate residue (Glu152) coordinated with Mg^{2+}, and (b) the proposed two ion mechanism (27).

1.5.2.3 C-terminal domain

The C-terminal domain, residues 213 – 288, is responsible for non-specific binding to the target DNA **(29)**. The structures were resolved using NMR studies **(38)**. The studies show that the structure composes of strand barrels. The structure resembles the Src homology 3 (SH3) domains which is responsible for DNA binding **(29)**. Previous studies revealed that residues 220 – 270 are the minimal region

for DNA binding **(39)**. The residues 271 – 288 are still not available. Structure of the C-terminal domain was shown in Figure 1.10.

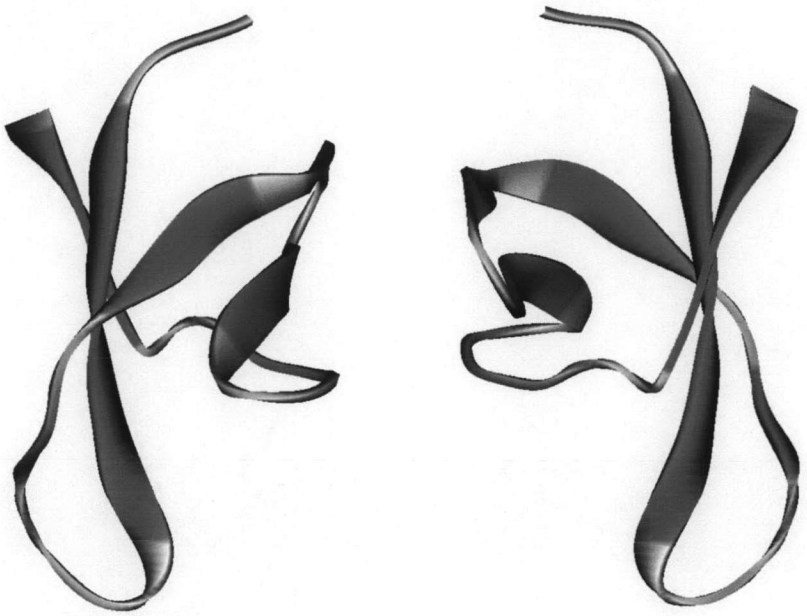


Figure 1.10 C-terminal domain of the HIV-1 IN **(38)**.

1.5.3 Integration process

The integration process occurs into two main steps, namely, 3' processing and strand transfer **(40)**. In the first step, two nucleotides from each 3' end of the viral DNA, synthesized by HIV-1 RT from viral RNA, were removed. The viral DNA was then left by the recessed CA OH' at the 3' end. In a second step, the previously processed 3'ends were transfer to the site of integration and joined to the 5'ends of the target DNA strands, in which being produced by 5 base pairs integrase-catalyzed staggered cut. Finally, the viral DNA was integrated into the host chromosome, so called integration. The intermediate recombinants viral-host DNA were still unjoined by a gap of 5 bp. These gaps were filled in and then ligated by the host enzyme. The HIV-1 IN also catalyzes the disintegration reaction (Figure 1.11).

1.5.4 HIV-1 IN Inhibitors

HIV-1 IN is considered as an attractive target for AIDS therapeutics. Numerous small molecules have been described as HIV-1 integrase inhibitors. The inhibitors were designed to act at the following possible sites **(41, 42)**.

1. attachment (*att*) sites
2. assembly of stable IN-DNA complex
3. 3' processing
4. 3' end joining (strand transfer)
5. 5' end joining (gap filling)
6. preintegration complex
7. N-terminal domain
8. C-terminal domain
9. catalytic domain and metal binding
10. oligomerization and IN-protein interaction

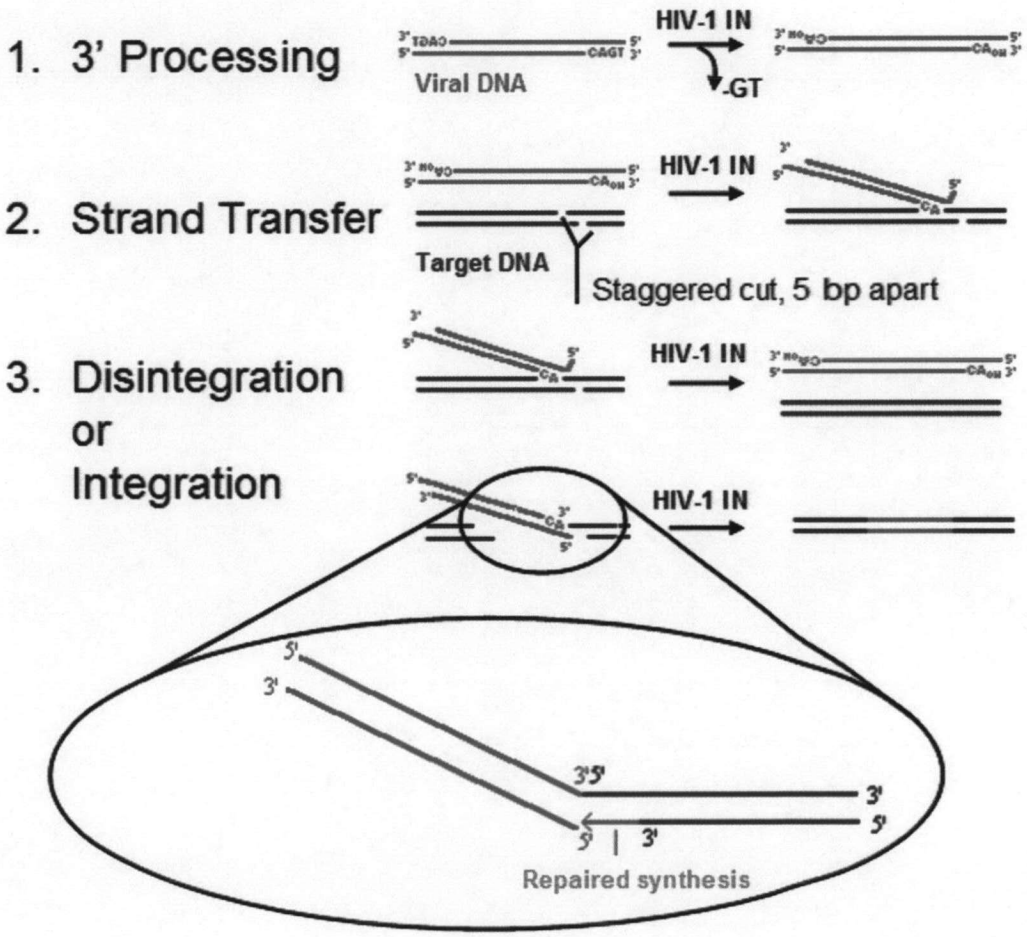


Figure 1.11 Integration process catalyzed by HIV-1 IN.

Numbers of HIV-1 IN inhibitors reported to date can be grouped into 5 categories.

1. DNA binders
2. Nucleotides
3. Hydroxylated aromatic
4. Sulfonates, Sulfones and Sulfa drugs
5. Peptides

The structures of HIV-1 IN inhibitors reported were summarized in Table 1.4.

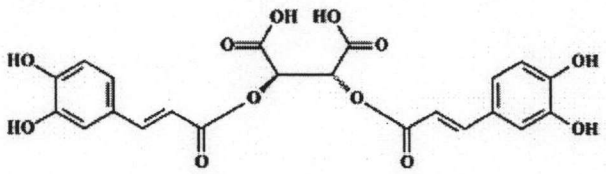
Table 1.4 Structure, group and sites of action of the HIV-1 IN inhibitors **(41)**.

Groups/Representative drugs	IC50 (μM)	Antiviral activity
1. DNA binders		
Doxorubicin, mitoxantrone	1-10	No
Chloroquine	10	No
Phenanthroline-cuprous complex	1	No
Triple helix-forming oligonucleotides	0.05	No
Minor groove binders	0.1	yes
2. Nucleotides and analog		
AZT nucleotides	100	yes (also RT)
Mononucleotides	40	yes (also RT)
Di- and tetranucleotides	3-300	no
Guonosine quartets (T30177)	0.05	yes (also gp120)
Coumermycin A1	10	no
RNA ligands	0.01	?
3. Hydroxylated aromatic		
Aurintricarboxyl acid	<1	yes (also gp120)
Cosalane derivatives	2	yes (also PR)
Dihydroxynaphthoquinone	2	no
Anthraquinones (purpurin, alizarin)	10	no
Flavones (quercetagetin)	0.1	no
CAPE and derivatives	3	some derivatives
Curcumin	40	moderate

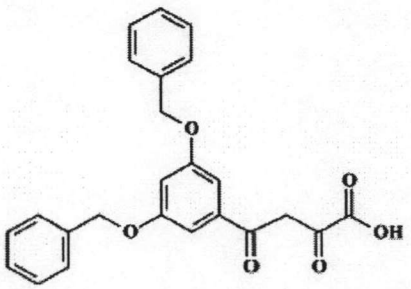
Table 1.4 (cont.)

Groups/Representative drugs	IC50 (μM)	Antiviral activity
3. Hydroxylated aromatic (cont.)		
Tyrphostins	0.2	one derivative
Biscatechols (β-conidendrol, hematoxylin)	0.5	no
Lignans and lignaloides	5	metabolites ?
Biscoumarins	1	yes
Dicaffeoylquinic/chicoric acids	0.2	yes
Arylamides	5	?
Hydrazides	0.2	?
Depsides (granulatine)	5	some derivatives
Tetracyclines	1-30	some
4. Sulfones and Sulfonates		
Diaryl sulfones	2	no
Aryl sulfonates	5	?
Suramin	0.2	yes
2-Mercaptobenzenesulfonamides	10	yes
5. Peptides		
Hexapeptide (HCKFWW)	3	no
Anti-IN antibodies	0.5	?

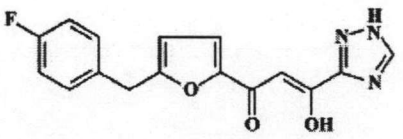
Recently, several HIV-1 IN diketo acid inhibitors have been studied. The chemical structures were shown in Figure 1.12.



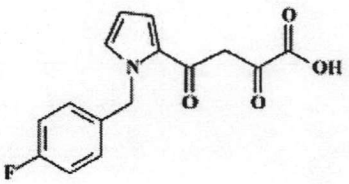
L-chicoric acid



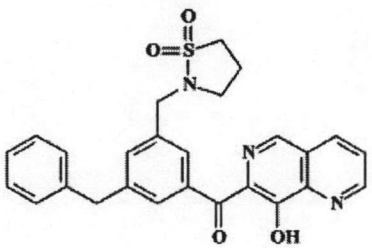
L,708,906



S-1360



L,731-988



8-Hydroxy-[1,6]naphthyridine

Figure 1.12 Chemical structures of some HIV-1 IN diketo acid inhibitors **(41)**.

1.6 Interaction with DNA

The process that the chromosomes are broken and rejoined to form a new genetic combination in which different from the original one is called genetic recombination. There are two kinds of the process, one is homologous recombination in which the process occurs between homologous DNA segments, and another is non-homologous recombination. The non-homologous recombination or site-specific recombination, the process of joining of non-homologous segments occurs at specific sites. It is responsible for the integration of bacteriophages into the host chromosome. There is an unusual kind of the genetic recombination known as transposition. The transposition is the process that short DNA segments (transposable element) named transposons changes their location within the host's chromosomal chromosome. This process is catalyzed by transposase. The retrotransposition occurs via RNA intermediates by which the RNA was copied to DNA by RT. The DNA is then incorporated into host chromosome by IN.

1.6.1 Structure of DNA

DNA (Deoxyribonucleic acid) is identified as a genetic material of the living organisms which are bacteria, protozoa, fungi, algae, plants and animals **(43)**. Geneticists, biochemists and biophysicists have been interested in molecular structure of the DNA for many decades. The DNA is macromolecule which composes of smaller building blocks known as nucleic acid. The structure of DNA is complex, *i.e.*, the structure has the repeating unit of nucleic acids called nucleotides, a strand of DNA is formed by linking together the nucleotides unit, a double helix is shaped by interaction of the two strands of DNA, and the double helix DNA folds and bends into the three-dimensional structure. DNA is associated with proteins within the living cells and plays essential roles in the genetic process. Unlike those organisms, the viruses used their own genetic material called RNA (ribonucleic acid).

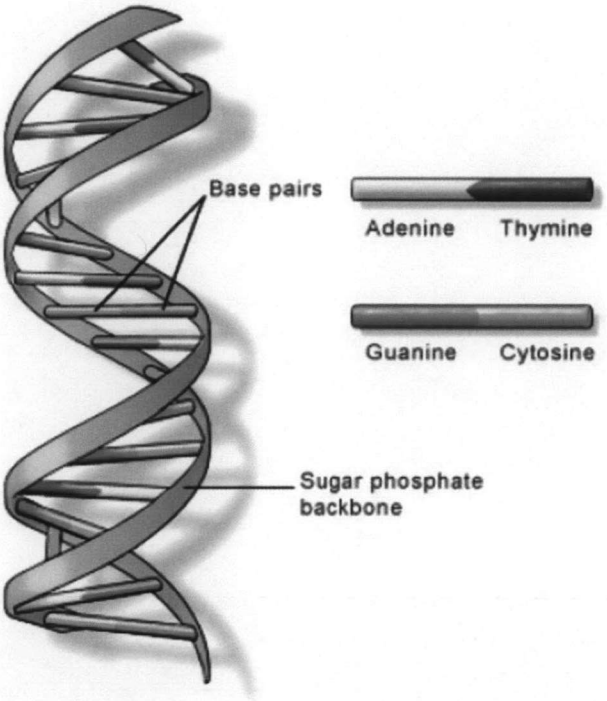


Figure 1.13 Structure of DNA **(44)**.

1.6.2 Nucleotides: A building block of DNA

A nucleotide composes of three main components, a phosphate group, a nitrogenous base and a pentose sugar **(45)**. There are five bases which are adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U). The bases are divided into two group based upon the structure in which the bases A and G are purine bases with a double-ring structure while the rest C, T and U are pyrimidine bases with a single-ring structure. There are two types of sugar which are ribose and deoxyribose. In the absence of the phosphate group, the pair containing base and sugar is called nucleoside. Note that the differences in chemical structure of the nucleotides are the key to distinguish the DNA and RNA. Figure 1.14 shows the components of nucleotides and the repeating unit of the DNA and RNA. The linkage between the nucleotide subunits to form the DNA molecules are phosphodiester bond in which forming by a phosphate group on one nucleotide and the sugar molecule on the adjacent nucleotide (Figure 1.14) The backbone of the DNA is, then, formed by the phosphates and sugar molecules with the negative charge on each phosphate.

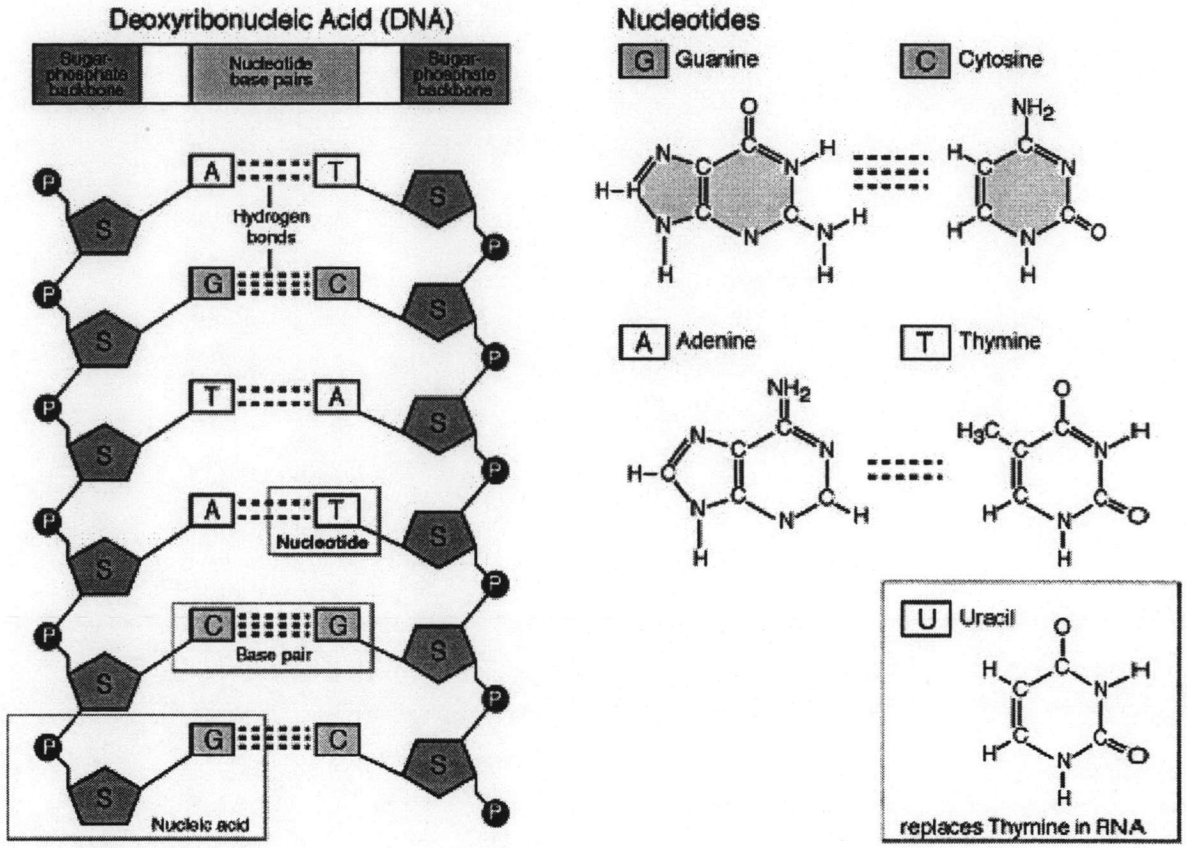


Figure 1.14 The component of DNA (left) and structure of five bases, A, T, C, G and U (right), the hydrogen bonds formed by base pairs were also shown, the image was downloaded from the site: http://www.accessexcellence.org/RC/VL/GG/nucleotide2.html **(46)**.

1.6.3 Double helical structure of DNA

In 1953, the structure of the DNA was determined by James Watson and Francis Crick (see Figure 1.15) **(45)**. It was realized that the hydrogen bonding between the bases, adenine to thymine and cytosine to guanine, are important in the helical structure of DNA. There are several key features for the molecular structure of the double helical DNA. (1) A right-handed double helix structure was formed by twisting together of the two strand of DNA. (2) The base pairs form hydrogen bonding according to the AT/GC rule, *i.e.*, adenine in one strand hydrogen bonds with thymine in another strand, likewise, guanine and cytosine. There are three hydrogen bonds forming by G and C while only two hydrogen bonds were found in A and T (Figure 1.16). The sequence of the base within the strand has important role in defining feature for the carrying information. (3) The orientation of the two DNA

strands is in antiparallel, 5' to 3', direction (Figure 1.16). (4) Each DNA strand has ~ 10.0 nucleotides per complete 360° turn of the helix. The key features of the double helical DNA was shown in Figure 1.16.

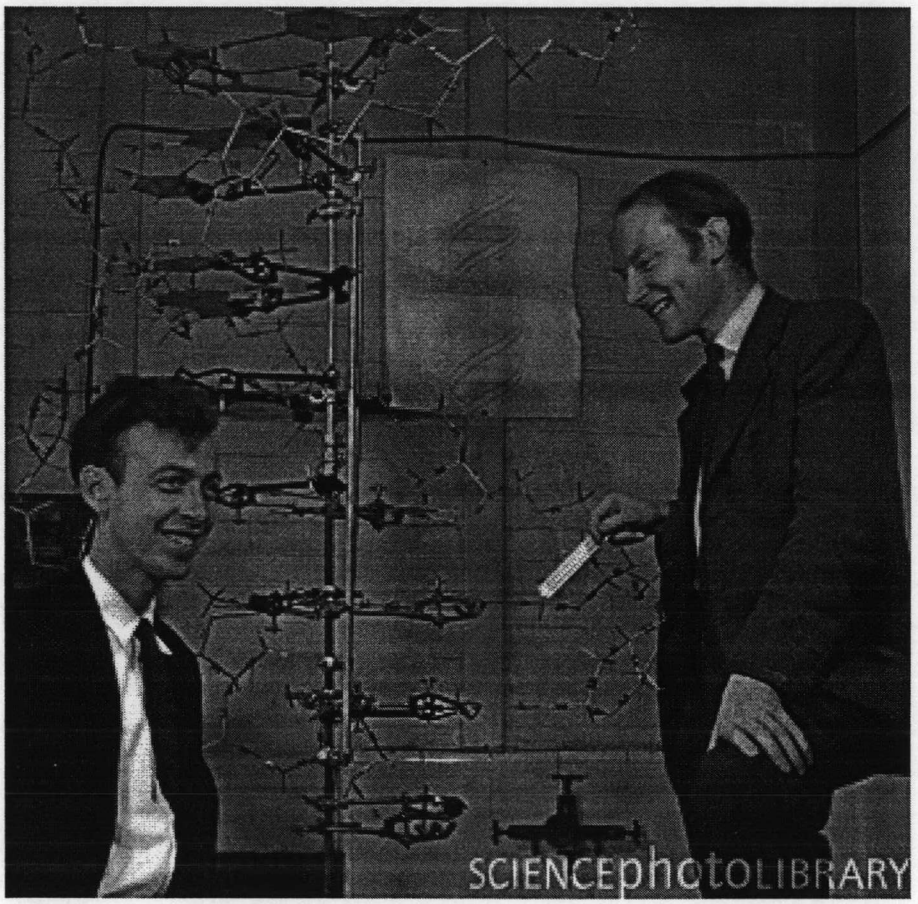


Figure 1.15 The double helical model of DNA by Watson and Crick. The picture was downloaded from Science photo library (by A. Barrington Brown).

There are three different types of secondary structures of the DNA, which are A DNA, B DNA and Z DNA. The predominantly form which is found in the living cells is B DNA. Unlike the A and B DNA which are right-handed helix, the Z DNA has left-hand orientation. Furthermore, the backbone of the Z DNA, differ from the other two forms, is slightly zigzag (Figure 1.17). The number of base pairs per 360° turn of A and Z DNA are 11.0 and 12.0, respectively, while the number for the B DNA is 10.0 as described above.

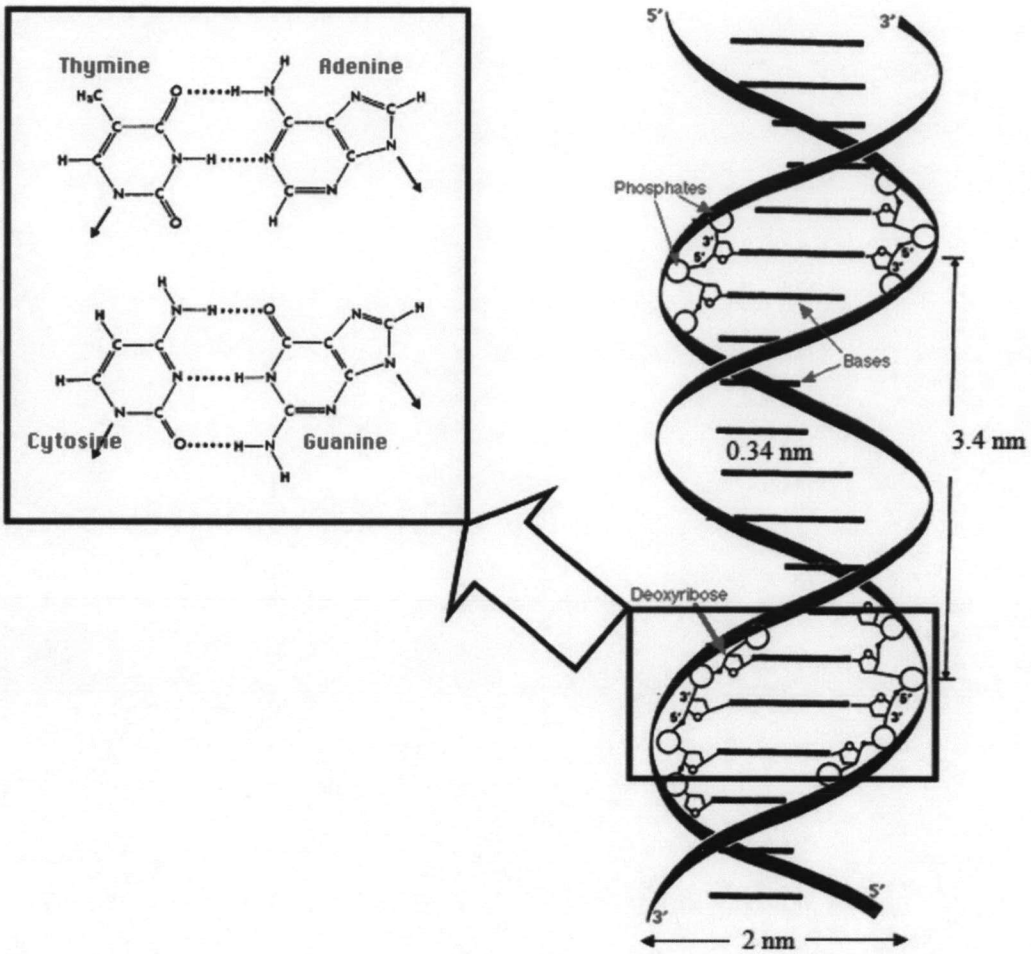


Figure 1.16 Structure of the double helix DNA and its key features.

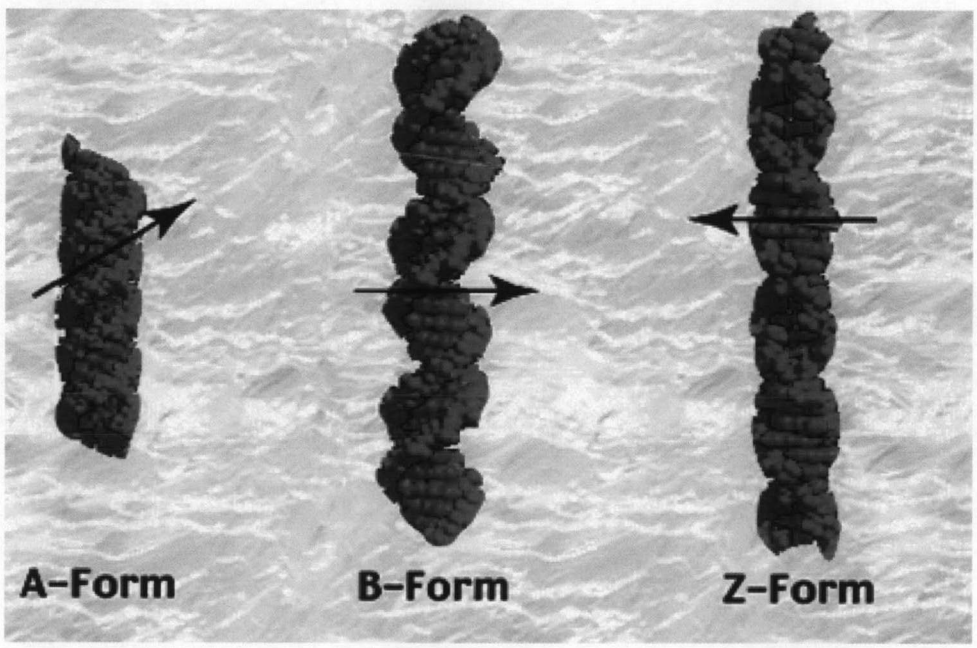


Figure 1.17 The structures of DNA in A, B and Z form.

1.6.4 HIV-1 IN-DNA interaction

As the function to incorporate the viral DNA into host chromosome, HIV-1 IN needs to interact with both viral and host DNA (Figure 1.18). Therefore, the viral and host DNA are substrates for integration catalyzed by HIV-1 IN. The viral DNA has several hundreded-base-pair encoding by three regions organized at each long terminal end (LTR). Highly conserved CA nucleotides locate at the 3' end of both strands serve as an attachment (*att*) sites for integration. Unlike the viral DNA, there is no obvious rule for the host DNA. Previous studies by activity assays state that viral DNA and other oligonucleotides can be distinguish by the IN **(47)**. However, the discrimination mechanism is still unclear. The metal cofactor encounters the HIV-IN-DNA interaction **(48)**.

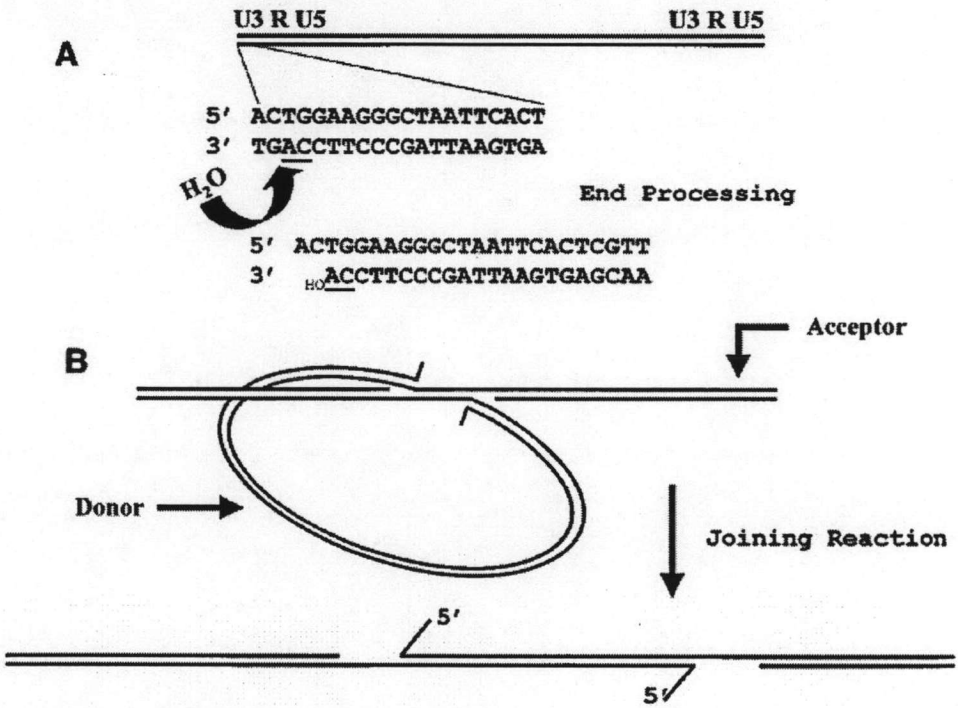


Figure 1.18 Integration mechanism of the HIV-1 IN **(49)**.

1.6.4.1 Recognition of the viral and host DNA

The viral DNA which is the substrate for DNA joining process catalyzed by HIV-1 IN must be double-stranded **(50, 51)**. However, the conserved end terminal dinucleotides CA of the cleaved viral DNA does not need to be base-paired **(51)**. An approximate length for the substrates for HIV-1 IN is at least 15 bp long. The

larger substrates are acceptable but seem to interact with HIV-1 IN in an undefined ways and may be cleaved to a lesser unit **(50)**. Many studies revealed that the conserve CA dinucleotides is critical for the HIV-1 IN **(47, 50, 51)**, *in vitro*. On the contrary, the distance of the cleavage site form the viral end is not critical for *in vitro* processing. From the cross-linking studies, the C-terminal domain stays close to the internal site of the CA of the viral DNA **(52)**. The C-terminal domain was suggested to bind viral DNA nonspecifically and the core domain plays a role in specific recognition of viral DNA ends and position of nucleophilic attack for host DNA **(53)**.

Unlike the recognition of the specific sequence of the viral DNA, the target for integration can be many host DNA sequences **(53)**. Target DNA recognition requires structure featured of the host DNA, architecture of the viral preintegration complex and some cellular factors. HIV-1 IN plays a major role in target recognition. The C-terminal portion binds DNA nonspecifically. The core region recognized the viral DNA end specifically and positions the host DNA for nucleophilic attack.

1.6.4.2 Retroviral integration mechanism

HIV-1 IN catalyzed the insertion of viral DNA into host chromosome, the reaction is known well as the integration process. The process occurs in two main steps, end processing and joining. In the first step, a dinucleotide which is adjacent to the highly conserved CA dinucleotide was removed from the 3' strand of the U3 and U5 LTR of the viral DNA. Water molecules as well as other nucleophiles are involved in the reaction. An exposed 3' hydroxyl group was then used as a nucleophile to attack the host DNA during the second step. The viral DNA was, thus, inserted to the host DNA **(49)**, see Figure 1.18.

1.6.4.3 Interaction with HIV-1 IN

The integrase can form stable complex with viral DNA *in vitro* **(48)**. The N-terminal zinc binding domain promotes enzyme multimerization **(31)**. Photo cross-linking studies reveal that there were close DNA contacts were found among all three domain. The N-terminal domain is close to both viral and host DNA. The catalytic core domain also participate in viral and host DNA binding while the C-terminal domain involves in viral DNA binding and recognize the viral end **(52)**. Experimental studies including X-ray crystallography, photo cross-linking suggesting

the model complex of the HIV-1 IN and DNA **(52, 54)**. Figure 1.19 shows model of the core and C-terminal bound to integration intermediate including the viral DNA end. Previous studies reveal that the core domain plays a role in recognition of the CA at the viral DNA end. The core and the C-terminal domain are capable of binding DNA **(52)**. The stable complex formed by HIV-1 IN and DNA required all three domains and the metal ions, Mg^{2+} or Mn^{2+}, are of important for stable complex formation.

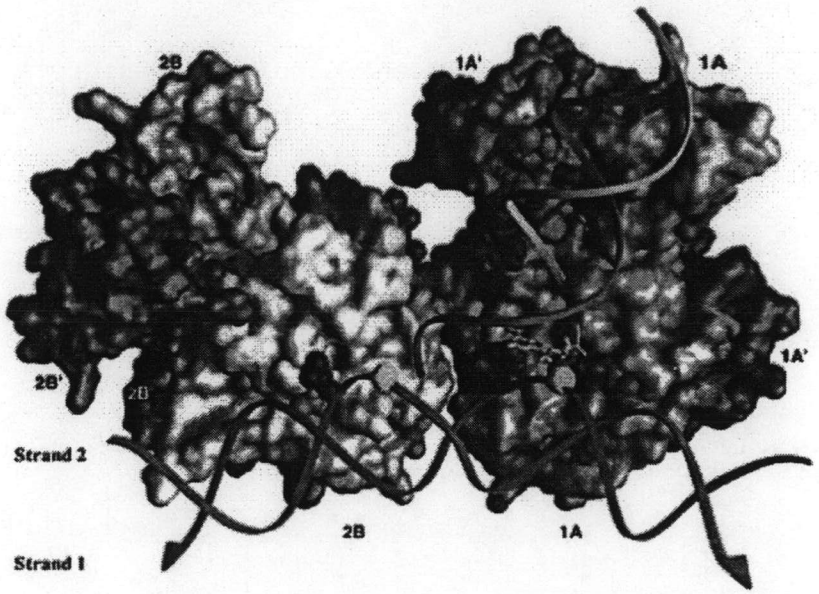


Figure 1.19 Core and C-terminal domain of HIV-1 IN complex with target and viral DNA from photo cross-linking study **(54)**.

1.7 Previous studies of the HIV-1 IN

1.7.1 Molecular dynamics simulation studies

There have been several studies on the HIV-1 IN using the MD simulation technique to provide useful information on the structural and the dynamical behaviors of this enzyme and insight into the enzyme-inhibitor complex behavior **(23, 24, 55-58)**. Some selected simulations from various groups were historical reviewed and summarized in Table 1.5. Almost all simulations take into account only the core domain in which evaluations were mainly focused to the binding mode of the active site for metal ion **(24, 58)**. The rests are the studies of the HIV-1 IN binding with inhibitor and DNA **(23, 55, 57)**. The effect of mutation was also evaluated **(23)**. The dynamical behaviors of the whole full-length HIV-1 IN, as

well as the two-domain fragment structures have been under studied by many research groups, Briggs group, McCammon group, Chimirri group **(59, 60)**. Such a study would be able to address many important questions such as how one domain affects the structural and dynamical behaviors of the others in the complete system. Why such question is important? Changes of structural and dynamical properties of the core due to the other domains means that the full-length HIV-1 IN is required in the study of drug/enzyme as well as enzyme/DNA interactions, and hence in the drug screening and drug design.

Table 1.5 Selective historical review of the previous MD studies on the HIV-1 IN.

Year	System of interest	Simulation time (ns)	Interested properties	Obtained results	Ref.
1999	Core domain: system with and without Mg^{2+}	1	- Structure determination of HIV-1 IN core domain. - Binding of metal ion in the catalytic core domain.	- Important conformational changes in the active region, particularly residues 141–148, which are part of the unsolved region in the crystal structures, were observed.	**24**
2000	Core domain (HIV-1 and ASV IN): system with two Mg^{2+}	2	- Similarities in the active site upon metal cofactor binding	- The binding of a second divalent ion does not decrease the flexibility in the region of residues 140–149.	**58**

Table 1.5 (cont.)

Year	System of interest	Simulation time (ns)	Interested properties	Obtained results	Ref.
2000	Core domain: system with two Mg^{2+} and two Mg^{2+}/HPO_4^{2-}	0.5	- Active site conformation - Binding site for second metal ion - Binding site for DNA	- The second metal ion is likely to be carried into the HIV-1 IN active site by a DNA strand. - A conversion of the HIV-1 IN into a complete and stable active form would only happen upon its binding of substrate. - A conformational change is observed in the HIV-1 IN active site only when the enzyme is in the presence of two divalent cations.	**55**

Table 1.5 (cont.)

Year	System of interest	Simulation time (ns)	Interested properties	Obtained results	Ref.
2001	Core domain: system with and without Mg^{2+}	6 (with ion)/ 4 (without ion)	- Fluorescence properties	- The binding of Mg21 drastically slowed down the rate of conformation changes. - The loop F139–S153 is remained flexible except for the residues 151–153 that are transitorily part of the α4 helix.	**56**
2001	Core domain: system with 5-CITEP inhibitor and one Mg^{2+}	2	- Active site binding mode	- The structural water molecule binding in the active site region should be taken into consideration by either utilizing it as a bridge group or displacing it for the future design of novel specific and tight-binding ligands.	**57**

Table 1.5 (cont.)

Year	System of interest	Simulation time (ns)	Interested properties	Obtained results	Ref.
2003	Core domain (wild type and double mutant forms): system with 5-CITEP inhibitor and one Mg^{2+}	1.75	- Ligand-IN complex formation and functional relationship - Role of loop close to the active site - Molecular mechanism for HIV-1 IN drug resistance - Receptor based drug design	- Substantial differences between the two forms during the MD simulations, particularly in the inhibitor binding mode and in the structure and dynamics of the surface loop located near the three catalytic residues in the active site were found. - The binding mode of inhibitor is quite different for both systems.	**23**

Note that the MD studies of the two-domain structure as well as the full-length from various research groups (Briggs group, McCammon group, Chimirri group, *etc.*) are still in progress and have not been published at the mean time.

1.7.2 Theoretical studies of the HIV-1 IN-DNA complex

There are several studies proposed the complex structure between the HIV-1 IN and DNA **(61-64)**, see Figure 1.20. The docking calculations were performed to find the possible binding region of the DNA. Some key residues were proposed to interact with DNA, *i.e.*, Lys156, Lys159, Lys186, Arg187 and Lys188, Lys211, Lys215 and Lys219 in the core domain provide a platform for attachment site. Residues Leu263, Lys264 in the C-terminal domain was observed to cross-link to the viral DNA. It was stated that a flexible elbow in the linking sequence allows the c-terminal domains to help tether the DNA during the integration process. The hydrophobic groups on the surface of an SH3-fold β-sheet in each protein intercalate into the minor groove of the bound DNA. The dinucleotides formed hydrogen bonds with Arg228, Lys244, Glu246, Arg262, Arg263, Lys264, Lys266, Ile267 and Arg269 on the c-terminal domain **(61)** (Figure 1.20 a). The docking calculation of the nucleotides to the core domain showed that most nucleotides dock near Lys111, Lys136, Glu138, Lys156, Lys159, Lys160, Lys185, Lys186 and Lys188, which are known to react with DNA, see Figure 4.8b **(62)**. There are two model dimer and tetramer full-length HIV-1 IN complexed with DNA **(63, 64)**. The dimer structure proposed by Luca *et al.* showed that the following residues interact with DNA, Lys156, Lys159, Lys160, Lys186 and Lys188 in the core region of chain B, Arg20 (N-terminal chain B) and Ser230, Arg231, Trp243, Lys244, Lys263, Lys264 (C-terminal chain A), Figure 1.20 d. There are some hydrophobic interaction between backbone of the DNA and some residues in the HIV-1 IN, *i.e.* Phe1, Leu2, Trp19, Ala21, Met22, Ala23, Phe26, Pro30, Met50 (N-terminal chain B), Met154, Ile191 (core chain B) and Trp243, Pro261 (C-terminal chain A). Podtelezhnikov et al. proposed the tetramer complexed with both viral and host DNA **(60)**, Figure 1.20 c. The results stated that, for binding of viral DNA, the terminus of viral DNA bind to catalytic domain near the site of integration and the more distal in the viral DNA are in contact with 247 – 270 of the c-terminal domain. Furthermore, for binding of host DNA, the core domain of the HIV-1 IN bind host DNA near the integration site. The C-terminal region is close to the internal segment between the integration sites of host DNA while residues 1 – 12 of the N-terminal domain bind outside the internal segment.

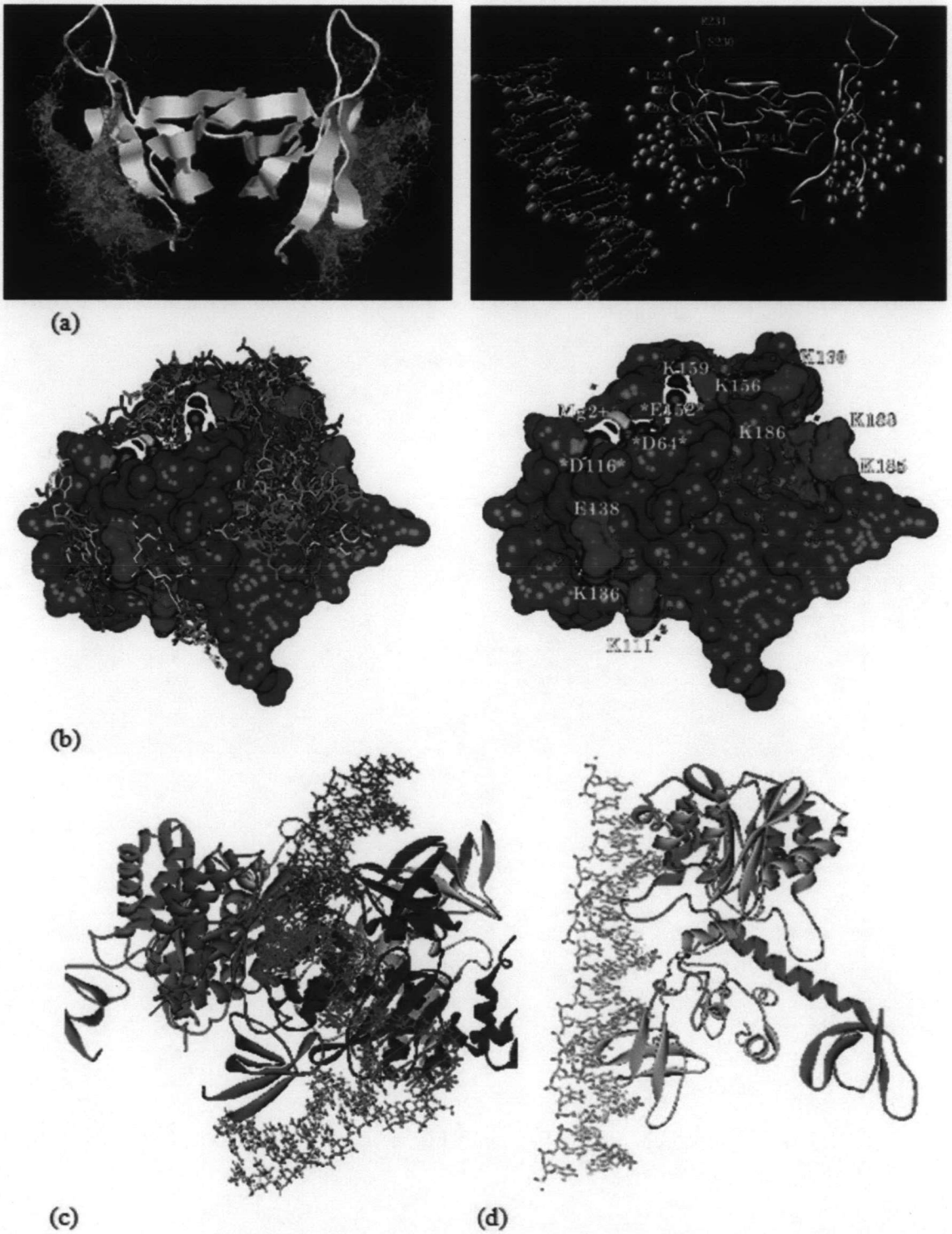


Figure 1.20 Proposed theoretical models for HIV-1 IN-DNA interaction **(61-64)**.

1.8 Objectives

Several methods, *i.e.*, molecular modeling, MD simulation and the molecular docking techniques were applied to study the HIV-1 IN with the following objectives

1.8.1 To propose the full-length HIV-1 IN structure which composing of the three domains.

1.8.2 To explore the structure and dynamical behavior of core only, two-domain fragment structure and the full-length structure.

1.8.3 To investigate the effect of the divalent metal ion in the active site on the catalytic residues in the core domain.

1.8.4 To investigate the effect of the two terminal-end domains on the catalytic core domain.

1.8.5 To propose the dimeric and tetrameric full-length HIV-1 IN complexed with DNA.