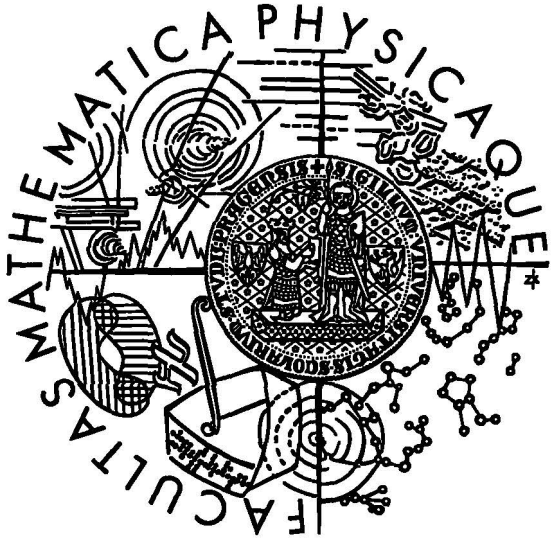
Charles University in Prague
Faculty of Mathematics and Physics

DIPLOMA THESIS



Sompob Saralamba

Molecular dynamics simulations of HIV reverse transcriptase in complex with potential inhibitors

Institute of Physics
Supervisor: RNDr. Ivan Barvík PhD.

Physics
Mathematical and Computational Modeling
in Physics and Engineering

I hereby declare that I worked out this diploma thesis by myself, using the cited references. I grant Charles University in Prague the right to display or copy any part or all of the thesis for use within the University and make available the thesis to other persons or organizations for educational or research purposes excluding any commercial purposes.

In Prague, 19. 8. 2006 Sompob Saralamba

Název práce: *Molekulárně-dynamické simulace komplexů HIV reverzní transkriptázy a jejích potenciálních inhibitorů*
Autor: *Sompob Saralamba*
Katedra (ústav): *Fyzikální Ústav MFF UK*
Vedoucí diplomové práce: *RNDr. Ivan Barvík, PhD*
e-mail vedoucího: *ibarvik@karlov.mff.cuni.cz*

Abstrakt: *Hlavním cílem této diplomové práce bylo studium dynamiky enzymu HIV reverzní transkriptázy v komplexu s dvojšroubovicí DNA:DNA a inhibitory (Adefovirem a Tenofovirem) pomocí prostředků MD simulací. Předmětem zájmu byly jemné diference ve způsobu vazby nukleotidů a inhibitorů do aktivního místa HIV-RT (v okamžiku před či po jejich navázání k enzymem syntetizovanému vláknu DNA). Struktury HIV-RT + DNA:DNA + nukleotidu/inhibitoru byly obklopeny vodní obálkou. Výsledné systémy sestávaly z přibližně 130.000 atomů. Extrémně časově náročné MD simulace byly provedeny pomocí velkých multiprocesorových systémů. Bylo zjištěno, že pouze Adefovir-DP (PMEApp) prodělává během MD simulace konformační změny (kvantifikované pomocí torzních úhlů). Tenofovir-DP (PMPApp) obsahující dodatečnou methylovou skupinu byl zcela stabilní. Optimální vazba Tenofoviru-DP (PMPApp) do aktivního místa enzymu by mohla být jedním z důvodů jeho větší terapeutické účinnosti. Residuum Arg72, které stabilizuje nukleotid/inhibitor v aktivním místě, se váže účinněji pomocí vodíkových vazeb k trifosfátovému motivu inhibitorů spíše než přirozených substrátů.*

Klíčová slova: *Molekulární dynamika, NAMD, HIV-1 RT*

Title: *Molecular dynamics simulations of HIV reverse transcriptase in complex with potential inhibitors*
Author: *Sompob Saralamba*
Department: *Institute of Physics MFF UK*
Supervisor: *RNDr. Ivan Barvík, PhD*
Supervisor's e-mail address: *ibarvik@karlov.mff.cuni.cz*

Abstract: *The main objective of this work was to study the dynamics of the HIV-1 RT enzyme in complex with DNA:DNA and inhibitors (Adefovir and Tenofovir) by using MD simulations. We were interested in subtle differences in binding of the nucleotides/inhibitors into the HIV-RT polymerase active site (before and after they were incorporated into the primer DNA strand). HIV-RT + DNA:DNA + nucleotide/inhibitor structures were solvated with water molecules. It led to simulate the systems consisting of 130,000 atoms. Extremely time-consuming MD simulations were produced using large multiprocessor systems. It was found, that only Adefovir-DP (PMEApp) underwent conformational transitions within a MD run (quantified using backbone torsion angles). Tenofovir-DP (PMPApp) carrying additional methyl group in the backbone was completely stable. Tenofovir-DP (PMPApp) optimal binding into the polymerase active site could be one from reasons of its higher therapeutic potency. Interestingly, Arg72 residue (sandwiching incoming nucleotide/inhibitor) was found to be more firmly anchored by hydrogen bonds to the triphosphate moiety of inhibitors rather than natural substrates.*

Keywords: *Molecular Dynamics, NAMD, HIV-1 RT*

Acknowledgements

I would like to express my deepest gratitude to my supervisor, RNDr. Ivan Barvík for his teaching and tremendous help. I would also like to thank Mr. Jaturong and Mrs.Nagmta Sriwongtrakul and the Czech Republic Government for providing me an opportunity to study in Czech Republic and the financial support during my study.

Contents

Chapter 1

Basic information about HIV-1

The human immunodeficiency virus type I (HIV-1) has been identified for many years as the primary cause of the acquired immunodeficiency syndrome (AIDS). Nowadays the number of people living with HIV exceeds 40 million [1]. The majority of HIV infected persons live in the developing countries of Asia and Africa. Many researchers around the world are studying AIDS from different points of view for finding the way to cure it.

To reach the goal of finding the way for curing AIDS, an understanding of the structure and replication cycle of HIV-1 is essential. This first chapter gives some basic information about it.

1.1 Classification and Structure of HIV-1

HIV is a retrovirus. It has ribonucleic acid (RNA) genome able to integrate (after transcription into DNA) into the host chromosome. HIV is categorised into the Family of *Retroviridae*, Genus *Lentivirus* [2].

There are two types of HIV: HIV-1 and HIV-2. Both types are transmitted by sexual contact, through blood, and from mother to child. However, it seems that HIV-2 is less easily transmitted, and the period between initial infection and illness is longer in the case of HIV-2 [3]. Worldwide, the main problematic virus is HIV-1. Generally, when people refer to HIV without specifying the type of virus they will be referring to HIV-1. The relatively uncommon HIV-2 type is concentrated in West Africa and is rarely found elsewhere [1, 3].

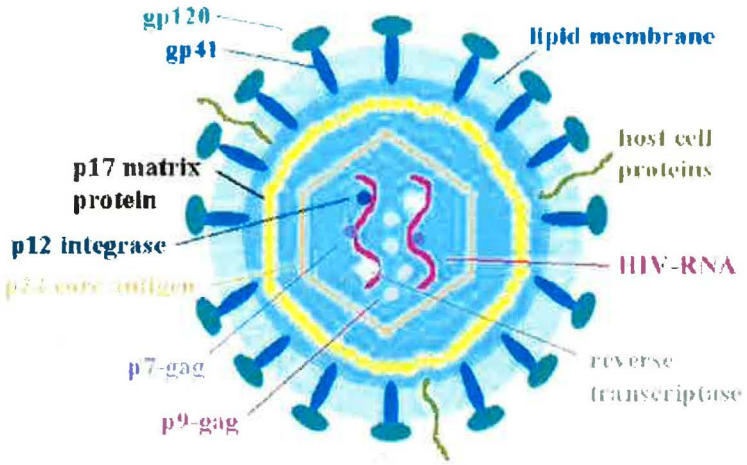


Figure 1.1: Schematic structure of the HIV-1 viral particle (*from HIV Medicine 2003*) [4]

The HIV-1 viral particle has a diameter of 100 nm. It consists of a cylindrically shaped nucleoid and surrounding proteins. They are both enveloped by a lipid membrane. The nucleoid is composed of two proteins known as the nucleoid core protein p24 and the matrix core protein p17. Inside the nucleoid are two copies of single-stranded RNA, nucleoprotein p7 and the associate enzyme called reverse transcriptase (RT). Outside the lipid membrane are attached two proteins: an external glycoprotein gp 120 and a transmembrane spanning protein gp41. These two external proteins help the virus to bind and fuse with another target cell. The viral particle contains all essential enzymes for replication: a reverse transcriptase (RT), an integrase p32 and protease p11.

1.2 The HIV-1 replication cycle

The main target cell which HIV-1 replicates in is an immune cell called lymphocyte, more specifically a CD4+ T cell or Helper T cell. It is an important part of the immune system. Its function is to detect cells in the body that are internally infected by viruses and bacteria. Decreased amount of T cells results in developing of AIDS.

After HIV-1 enters the human body, it comes in contact with CD4+ T cells. The

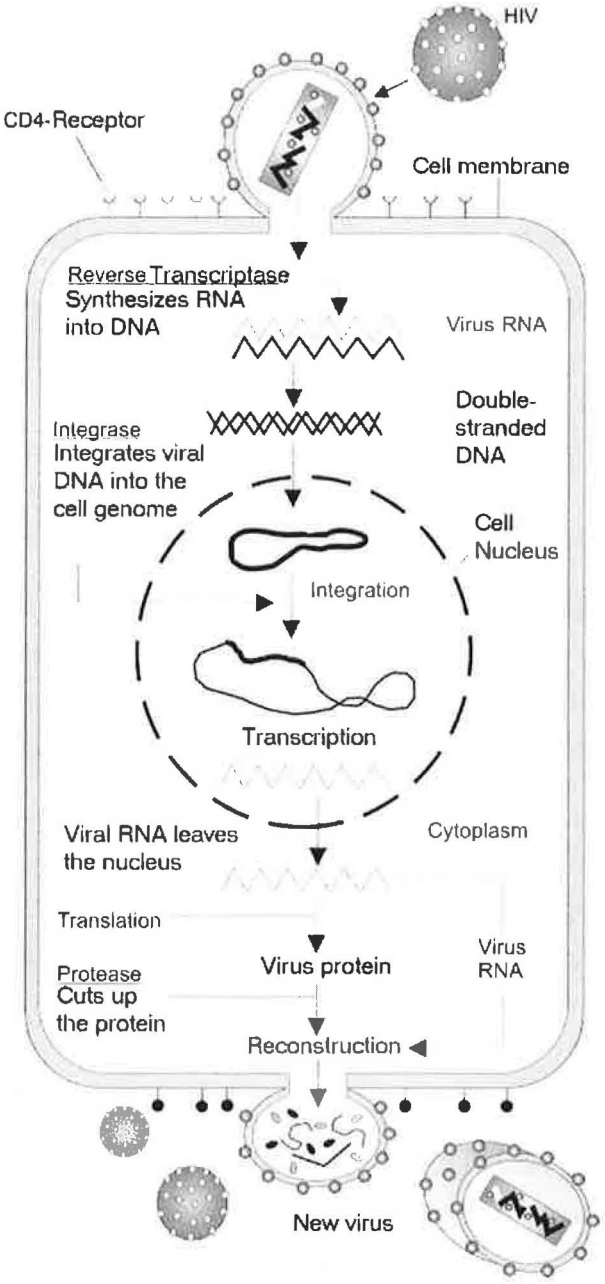


Figure 1.2: HIV-1 Replication cycle (*picture from www.wikipedia.org*)

virus will use the cellular machinery of the host cell to replicate itself. For this to occur, the virus has to complete many steps. It is theoretically thinkable to create a drug that will stop the virus at each step of its replication cycle.

1. **Binding and Fusion:** Replication cycle of HIV-1 starts when its envelope gp120

and gp41 proteins bind to the CD4 receptor and a chemokine receptor (either CCR5 or CXCR4). Three sugar-coated proteins (glycoproteins) which are parts of gp120 spread apart after gp120 attaches to CD4. This makes the gp41 to become bare and bind to the chemokine receptor. The viral envelope and the lipid membrane of the host cell are brought into direct contact and fuse into each other. The virus then releases RNA, its genetic material, into the cytoplasm of the host cell.

2. **Reverse Transcription:** Viral RNA is converted to DNA. It happens in the opposite direction comparing to the transcription process in human cells which converts DNA to RNA. The HIV-1 enzyme reverse transcriptase (RT) is used in this process. This enzyme transcribes single stranded viral RNA into double stranded DNA containing the instructions HIV-1 needs for reproducing itself. HIV-1 RT uses nucleotides from the host cell cytoplasm in this process.

3. **Integration:** Viral DNA from previous step (also called the preintegration complex) is transported across the nuclear membrane into the nucleus, where DNA or chromosomes of the host cell are stored. It is inserted into the chromosome by the HIV-1 integrase enzyme. The inserted viral DNA is referred to as provirus.

4. **Transcription:** When the host cell becomes active, the provirus uses an enzyme of the host cell called RNA polymerase to produce short strands of RNA called messenger RNA (mRNA) containing copies of the genetic information. These mRNAs are translated into long chains of HIV-1 subunits such as structural proteins, proteases, reverse transcriptases, integrases, glycoproteins, and regulatory proteins.

5. **Assembly:** The HIV-1 protease enzyme cleaves the long chains of proteins from previous step into smaller individual proteins. These proteins and HIV's RNA genetic material are assembled to becomes a new viral particle at the lipid membrane of the host cell.

6. **Budding:** The lipid membrane of the host cell is deformed to form the envelope of newly assembled virus. This viral envelope has gp120 glycoproteins and gp41 transmembrane spanning proteins. The newly synthesized virus then buds from the host cell and can now move on to infect other cells.

1.2.1 Examples of HIV-1 inhibitors

There are many kinds of drugs or inhibitors which are used to stop HIV-1 at each step of its life cycle. These drugs are usually used together to stop the HIV-1 replication efficiently.

- **Entry inhibitors (including Fusion inhibitors)** block the interaction between CD4 receptors and viral gp120 or gp41 proteins.
  - *Vicriviroc*
  - *Maraviroc*
  - *Enfuvirtide*
- **Reverse Transcriptase Inhibitors** block viral HIV-1 reverse transcriptase (RT) from using the nucleotides of the host cell.
  - *Tenofovir disoproxil fumarate (DF)*
  - *Zidovudine (AZT)*
  - *Emtricitabine (FTC)*
  - *Nevirapine (NVP)*
  - *Delavirdine (DLV)*
- **Protease Inhibitors** bind to the protease enzyme and block it from cleaving and separating subunits of the HIV-1 polyprotein.
  - *Tipranavir (TPV)*
  - *Fosamprenavir (FPV)*
  - *Saquinavir (SQV)*

Chapter 2

HIV-1 RT and its inhibitors

The reverse transcriptase (RT) is one of the main enzymes of HIV-1. It is used in the reverse transcription process. This process was discovered by Prof. Howard Martin Temin from the university of Wisconsin, USA. With this work he won the Nobel Prize in 1975 [5]. HIV-1 RT is responsible for transcribing the single-stranded viral RNA genome into double-stranded DNA. Because of this essential function in the HIV replication cycle, HIV-1 RT is a significant target for drugs used in the treatment of AIDS.

The structure of HIV-1 RT is described briefly in this chapter. Some details of the structure of inhibitors used in molecular dynamics simulations (tenofovir and adefovir) are also given.

2.1 Structure of HIV-1 RT

The HIV-1 RT structures have been studied and determined by using X-ray diffraction methods. Coordinates of these structures are available for download from the so-called Protein Data Bank (PDB www.rcsb.org/pdb).

HIV-RT is a multifunctional enzyme. It consists of a DNA polymerase site that is used to copy either an RNA or DNA template and a ribonuclease H (RNase H) domain able to degrade RNA. HIV-RT is a asymmetry heterodimer consisting of a 66 kDa (p66) and a 51 kDa (p51) subunit. The p51 subunit is composed of 450 amino acids (also called residues). The p66 subunit is composed of 560 residues. It contains two domains: the N-terminal polymerase domain (440 residues) and the C-terminal RNase H domain

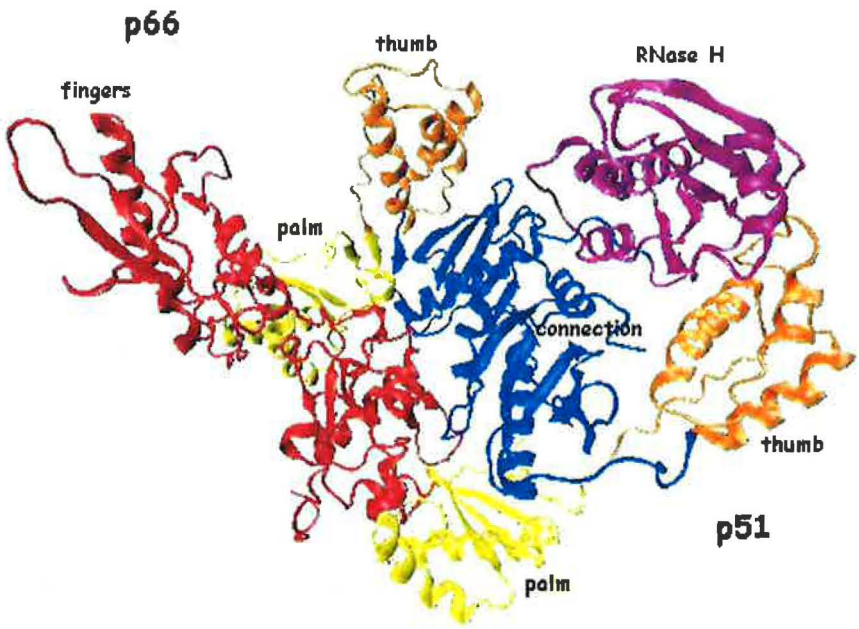


Figure 2.1: HIV-1 RT cartoon diagram. The corresponding PDB code is 1T05

(120 residues). Because of the similarity to a human right hand, the subdomains of p66 subunits are usually referred to as fingers, palm, thumb, and connection [6].

The fingers subdomain is composed of mixed β-strands and three α-helices, and the palm includes five β-strands that form hydrogen bonds with four β-strands positioned at the base of the thumb. A helical bundle forms the thumb subdomain of the enzyme, and the connection subdomain, which connects the polymerase and RNase H domains, is composed of a large β-sheet and two α-helices [7].

2.2 Active site of HIV-1 RT

There are two active sites in HIV-1 RT: polymerase site and RNase H site. Both active sites are in the p66 subunit. These subdomains of the p66 subunit pack together to form an open right-hand configuration, creating a large cleft in the polymerase site that exposes the three catalytic residues Asp110, Asp185, and Asp186. The The HIV-1 RT uses this active site to transcribe the single-stranded viral RNA into a double-strand of DNA.

The function of RNase H site is to cleave the 3′-O-P-bond of RNA in a DNA/RNA duplex to produce 3′ hydroxyl and 5′ phosphate terminated products. It is also used to destroy the RNA template after the synthesis of the first DNA strand by HIV-1 RT is completed.

2.3 RT and Nucleic acids

Several crystal structures of HIV-1 RT bound with double-stranded DNA and inhibitors have been determined by X-ray methods and can be found in the Protein Data Bank. In molecular dynamics simulations the HIV-1 RT structure in complex with a DNA:DNA duplex and tenofovir (PDB accession code 1T05) were used [8].

The 1T05 structure shows that the majority of the HIV-1 RT / DNA contacts is realized by residues of the p66 fingers, palm and thumb sub-domains. The palm and thumb are acting as a clamp that locates DNA relative to the active site residues. The 3′-OH hydroxyl group of the primer strand and tenofovir are positioned close to the polymerase active site. The tenofovir is non-covalently bound to the HIV-1 RT / DNA complex as an incoming substrate.

2.4 HIV-1 RT inhibitors

The HIV-1 RT inhibitors can be divided into three categories: 1) nucleoside reverse transcriptase inhibitors (NRTIs) 2) nucleotide reverse transcriptase inhibitors (NtRTIs) 3) Non-nucleoside reverse transcriptase inhibitors (NNRTIs) [9].

- **Nucleoside reverse transcriptase inhibitors** (NRTIs) act as chain terminators in the reaction catalyzed by HIV-1 RT. NRTIs lack the 3′ hydroxyl group. Therefore, after the inhibitors incorporate into the growing DNA chain, the newly made DNA strand is terminated. The well known drug in this class is *AZT* or *zidovudine*.

- **Nucleotide reverse transcriptase inhibitors** (NtRTIs) work by the same way as NRTIs but the first phosphorylation step (nucleoside to nucleotide) is not necessary [4]. Widely used drugs in this class are *tenofovir* (PMPA) and *adefovir* (PMEA).

Tenofovir Adefovir

Figure 2.2: Comparision of the chemical structure of tenofovir and adefovir

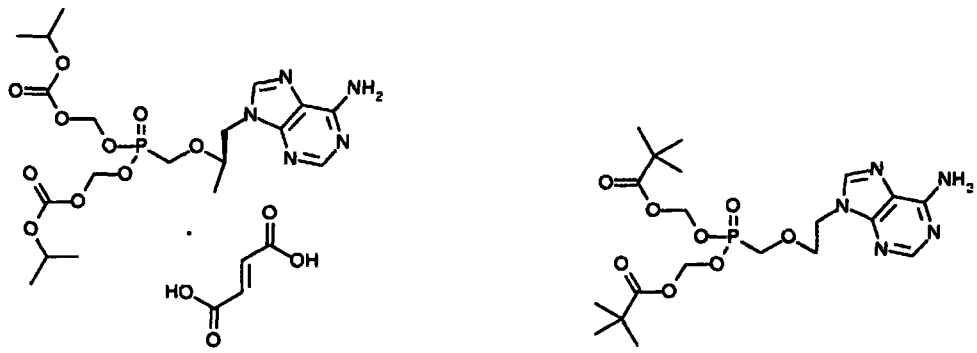
- **Non-nucleoside reverse transcriptase inhibitors** (NNRTIs) bind onto an allosteric site which is situated about 10 Å away from the HIV-1 RT polymerase active site [9]. After NNRTIs binding, the allosteric site becomes more restricted in its flexibility and mobility, it leads to a reduced enzymatic activity of HIV-1 RT [4].

In our molecular dynamics simulations, *tenofovir* and *adefovir* were used as incoming substrates. They were incorporated into the HIV-1 RT polymerase active site and subtle differences of their binding were compared. Despite of remarkable structural similarity, tenofovir is well known as a more potent inhibitor of HIV-1 RT.

2.4.1 Tenofovir

Tenofovir or PMPA is an anti-AIDS drug in the class NtRTIs. It was discovered by a group of Prof. Antonín Holý from the Academy of Sciences of the Czech Republic (IOCB) in Prague and Prof. Erik DeClercq from Rega Institute for Medical Research, Katholic University, Belgium[10].

Tenofovir was approved by U.S. Food and Drug Administration(FDA) in 2001 and is marketed under the name VIREAD™ by Gilead Sciences company. VIREAD™ is the brand name for the drug product, tenofovir disoproxil fumarate (tenofovir DF) tablets. The chemical name of tenofovir DF is
9-[(R)-2-[[bis[[isopropoxycarbonyl)oxy] methoxy] phosphinyl] methoxy] propyl] adenine fumarate (1:1) (IUPAC)
or
(R)-5-[[2-(6-amino-9H-purin-9-yl)-1-methy] methyl] -2,4,6,8-tetraoxa-5 -phosphanonanedioic acid, bis(1-methylethyl) ester, 5-oxide, (E)-2-butenedioate (1:1)



Tenofovir DF **Adefovir DP**

Figure 2.3: Comparision of the chemical structure of tenofovir disoproxil fumarate and adefovir dipivoxil

(CAS). It has molecular formula of $C_{23}H_{34}N_5O_{14}P \cdot C_4H_4O_4$. The structural formula of tenofovir DF is shown in Figure 2.3. Tenofovir DF is a fumaric acid salt of *bis*-isopropoxycarbonyloxymethyl ester derivative of tenofovir.

In oder for tenofovir DF to act as an inhibitor of HIV-1 RT it needs to go through two ester hydrolytic steps by the host esterases to form tenofovir and afterwards through two phosphorylations by cellular phosphokinase enzymes to form the active metabolite, the tenofovir diphosphate (PMPApp) [11]. The tenofovir diphosphate inhibits the viral reverse trancriptase activity by binding competition with the natural subtrate deoxyadenosine triphophate and by DNA chain termination after its incorporation into viral DNA [11].

2.4.2 Adefovir

Adefovir or PMEA was developed by Gilead Sciences company for the treatment of HIV. In November 1999, the U.S. Food and Drug Administration (FDA) refused to approve the adefovir as a drug for treatment of HIV due to concerns about the severity and frequency of kidney toxicity when dosed at 60 or 120 mg [12]. However, adefovir is used for treatment of Hepatitis B and it is effective with a much lower dose. FDA approved adefovir for use in the treatment of Hepatitis B on September 20, 2002. Adefovir is sold for this indication under the brand name *Hepsera* [12].

HEPSERA™ is the tradename for adefovir dipivoxil, a diester prodrug of adefovir. The chemical name of adefovir dipivoxil is *9-[2-[bis[(pivaloyloxy)methoxy]phosphinyl]-methoxy]ethyl]adenine*. It has a molecular formula of $C_{20}H_{32}N_5O_8P$. The structural formula of adefovir dipivoxil is shown in Figure 2.3.

The mechanism of action of adefovir is similar to the mechanism of tenofovir. Adefovir is an acyclic nucleotide analog of adenosine monophosphate. Adefovir is phosphorylated to the active metabolite, adefovir diphosphate (PMEApp), by cellular kinases. Adefovir diphosphate inhibits HIV-1 RT by competing with the natural substrate deoxyadenosine triphosphate and by causing DNA chain termination after its incorporation into viral DNA [12].

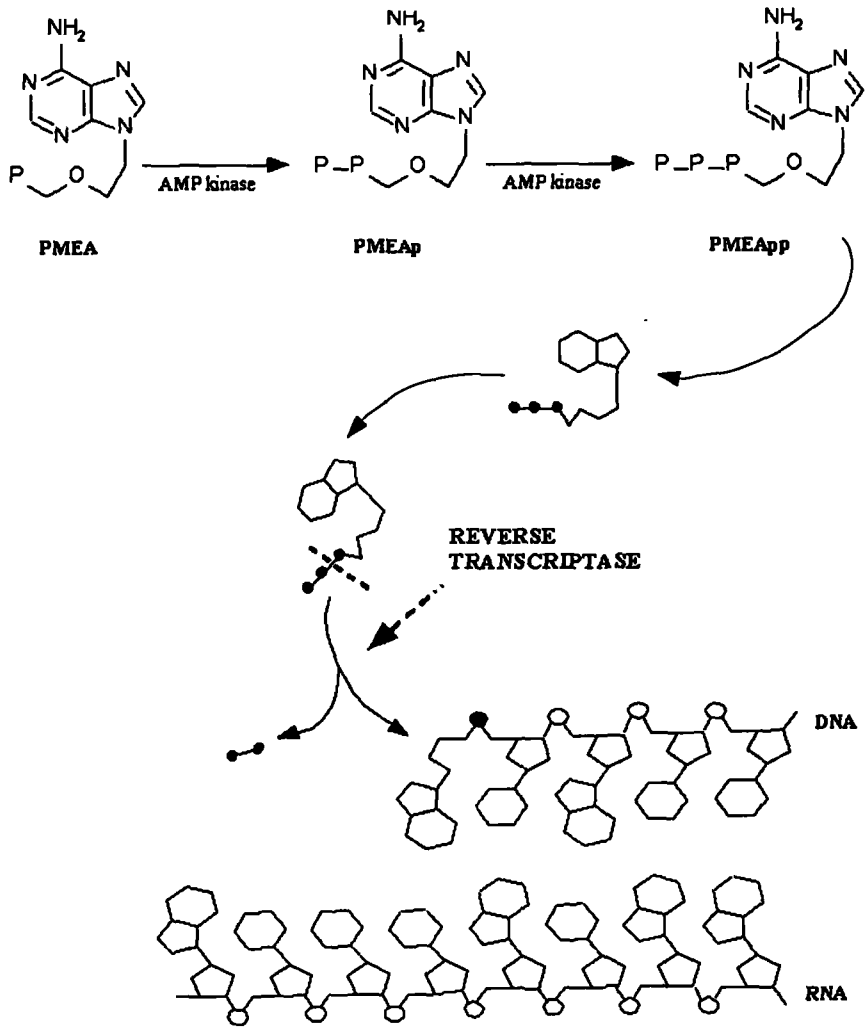


Figure 2.4: Mechanism of action of adefovir (PMEA). Similar mechanism of action applicable to tenofovir (PMPA)

Chapter 3

Molecular Dynamics Simulation

One of the most famous computational methods in studying biomolecules is Molecular Dynamics (MD). MD generates microscopic time series (trajectories) of the dynamic evolution of all atoms in a molecular system by integrating Newton's equations of motion. From these trajectories we are able to calculate the physical quantities of the molecular system. MD simulations are a suitable tool for studying both - kinetic and thermodynamic - properties of the simulated system. Because of this universality MD simulations are used to simulate proteins, peptides, nucleic acids(DNA and RNA), lipids, sugars, and other biomolecules.

This chapter explains briefly basics of MD simulation methods. The configuration keywords used for the NAMD software package are also described.

3.1 Molecular Mechanics and Force Field

The main idea of the molecular mechanics (MM) is that a molecular system can be viewed as a microscopic mechanical system. Molecular mechanics considers a molecule as a system of rigid balls connected via mechanical springs and the potential energy of the system is a sum of all interactions within and between the springs. We can write

this expression in mathematical terms as follows:

$$\begin{aligned} E_{total} &= E_{valence} + E_{non-bond} + E_{cross-term} \\ E_{valence} &= E_{stretch} + E_{angle} + E_{out-of-plane} \\ E_{non-bond} &= E_{vdW} + E_{elec} \\ E_{cross-term} &= E_{str-str} + E_{str-angle-str} + E_{angle-angle} + E_{torsion-str} + \ldots \end{aligned} \quad (3.1)$$

From above energy equations, the interactions between atoms can be simply divided: Bonded and Non bonded interactions.

3.1.1 Bonded interactions

Bonded interaction is the interaction between atoms which are connected together via covalent bonds. As molecular mechanics considers bonds between atoms as springs, the mathematics of spring deformations can be used to describe the ability of bonds to stretch, bend, and twist. The following equation is the generally used form for the bonded interactions:

$$\begin{aligned} E_{bond} = & \sum_{bonds} \frac{k_b}{2} (l - l_0)^2 + \sum_{angles} \frac{k_a}{2} (\theta - \theta_0)^2 \\ & + \sum_{torsions} k_d \left(1 + \cos(n\phi - \phi_0)\right) + \sum_{improper} k_i (\omega - \omega_0)^2 \end{aligned} \quad (3.2)$$

The first and second terms of this equation describe the bond stretching and bending respectively. These two potential energy terms are based on Hooke's law. The k_b parameter controls the stiffness of the bond spring. The k_a parameter controls the stiffness of the angle spring, while θ_0 defines its equilibrium angle. The third term is the potential which is required for the rotations about torsion or dihedral angles.

3.1.2 Non-Bonded interactions

Non-Bonded interactions are either intermolecular interactions (between two atoms from the different molecules) or intramolecular interactions (between atoms from the same

molecule but not connected together via the covalent bond). The non-bonded interactions consist of the electrostatic term, Van der Waals interaction, and hydrogen bonding as well as other possible interactions.

The electrostatic interaction between two charged atoms separated by at least 3 bonds is calculated from the Coulomb's law:

$$E_{elec} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{3.3}$$

The Van der Waals potential considers repulsion between atoms at small separations and weak attraction at larger distances. The common form of this potential for a pair of atom i and j is given by a Lennard-Jones potential function as:

$$E_{vdW} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\epsilon \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \tag{3.4}$$

3.1.3 Force Field

A force field refers to the potential function form and parameter sets used in the function. The force field used for biomolecules is usually empirical. There is no correct form for the force field. In principle, any set of potential energy functions can be established and parameterized on the basis of either experimental or ab initio data.

From the equation 3.1 it is common practice to represent the potential of a molecule by a set of potential energy functions, including bonded and non-bonded interactions. Additional terms, including out-of-plane interaction, hydrogen bonding, cross terms, etc., may be added.

In practice, there are many force fields available for using in the simulation. They are both in commercial version and free to download. The choosing of force field depends on the kind of the simulated system. The most widely used force fields are:

- **AMBER** (Assisted Model Building with Energy Refinement, amber.scripps.edu)
- **OPLS, OPLS-AA** (Optimized Potentials for Liquid Simulations, zarbi.chem.yale.edu)
- **CHARMM** (Chemistry at HARvard Macromolecular Mechanic,

www.pharmacy.umaryland.edu/faculty/amackere/force_fields.htm)

- **GROMOS** (GROningen Molecular Simulation, www.igc.ethz.ch/gromos)

These force fields are based on slightly different equations and it is not possible to substitute or mix parts of different force fields.

3.2 Molecular Dynamics (MD)

Molecular dynamics is a method for computing equilibrium and kinetic properties of the systems following the laws of classical (Newtonian) physics. In MD simulations the time evolution of a simulated system is described by the Newton equations of motion. The acceleration of the atom i of mass m_i, $\frac{d^2\mathbf{r}_i}{dt^2}$ depends on the force $\mathbf{F}_i$, acting on atom i:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \tag{3.5}$$

Newton equations 3.5 give a link between the molecular motion and potential energy function V. For the numerical solution of these equations, the integration algorithms (for example Verlet or Leap-Frog) are widely applied. The table 3.2 shows the equations of updating of the coordinates and velocities by using Verlet and Leap-Frog algorithms.

Taylor series expansion
$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2 + \cdots$
$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t + \frac{1}{2}\frac{d^3\mathbf{r}(t)}{dt^3}\Delta t^2 + \cdots$
Verlet algorithm
$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\Delta t^2$
$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{1}{2}\left(\frac{d^2\mathbf{r}(t)}{dt^2} + \frac{d^2\mathbf{r}(t+\Delta t)}{dt^2}\right)\Delta t$
Leap-frog algorithm
$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t+\frac{1}{2}\Delta t)\Delta t$
$\mathbf{v}(t+\frac{1}{2}\Delta t) = \mathbf{v}(t-\frac{1}{2}\Delta t) + \frac{d^2\mathbf{r}(t)}{dt^2}\Delta t$
$\mathbf{v}(t) = \frac{1}{2}\left[\mathbf{v}(t-\frac{1}{2}\Delta t) + \mathbf{v}(t+\frac{1}{2}\Delta t)\right]$

Table 3.1: Integration algorithms

In our works two widely known MD software packages were used: AMBER [13] and NAMD [14]. Both software packages can use the AMBER force field. However, the methods they use to control the physical quantities of the simulated system (temperature, pressure, volume) are different. In chapter 4 we explained briefly how to set up a MD simulation run for biomolecular structures and MD configuration parameters used in NAMD are discussed.

3.2.1 Molecular Dynamics at Constant Temperature

Rescaling of velocities

As the temperature of a simulated system is related to the average kinetic energy of the particles, the temperature can be controlled by scaling the velocities [15]. In this method, the velocities are multiplied by a factor λ and the associated temperature change can be computed as follows:

$$\Delta T = \frac{1}{3}\sum_{i=1}^{N}\frac{m_i(\lambda v_i)^2}{Nk_b} - \frac{1}{3}\sum_{i=1}^{N}\frac{m_i v_i^2}{Nk_b} \tag{3.6}$$

$$\Delta T = (\lambda^2 - 1)T(t) \tag{3.7}$$

$$\lambda = \sqrt{T_{req}/T(t)} \tag{3.8}$$

where $T(t)$ is the temperature at time t. At each time step the velocities are multiplied by the factor $\lambda = \sqrt{T_{req}/T_{curr}}$, where T_{curr} is the current temperature calculated from the kinetic energy and T_{req} is the desired temperature.

In the NAMD software package, the keyword `rescaleFreq` defines the frequency (in steps) of rescaling of velocities, and the keyword `rescaleTemp` sets the target temperature [16].

Weak coupling with a heat bath

In 1984, Berendsen and colleague [17] proposed an alternative way to maintain the temperature of the system. His method is to couple the system to an external heat bath which is fixed at the desired temperature and at each time step the velocities are scaled. The rate of change of the temperature is corresponding to the difference in the

temperature between the bath and the system:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{bath} - T(t)) \tag{3.9}$$

where τ is the coupling parameter. The diffence of temperature in each time step is:

$$\Delta T = \frac{\delta t}{\tau}(T_{bath} - T(t)) \tag{3.10}$$

The scaling factor is:

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right) \tag{3.11}$$

where δt is the time step, $T(t)$ is the current temperature, T_{bath} is the desired temperature.

NAMD provides this Berendsen thermostat as an extra option for temperature control. The keywords `tCouple` and `tCoupleTemp` switch the Berendsen coupling on and specify the bath temperature T_{bath} respectively.

Langevin dynamics (LD)

Langevin dynamics is another important method for temperature control. LD is different from two previous methods. LD does not directly modify the velocities, but introduces friction and random forces. The surrounding of a simulated molecule is modeled by including of only average interactions. These average interactions are assumed to have a friction term (with a friction coefficient or damping constant ξ) proportional to the atomic velocity and a random component (Γ_i). The random force is associated with a temperature and adds energy to the system, while the friction term removes energy [18, 19]. Consequently, the Langevin equation for an atom i can be written as follows:

$$m_i\frac{d\mathbf{r}_i^2}{dt} = \mathbf{F}_{intra} + \Gamma_i(t) - \xi m_i\frac{d\mathbf{r}_i}{dt} \tag{3.12}$$

The random force $\Gamma_i(t)$is derived from a Gaussian distribution with the properties

$$\langle\Gamma_i(t)\rangle = 0 \tag{3.13}$$

$$\langle\Gamma_i(t)\Gamma_i(t)\rangle = 2k_B\xi T_0\delta(t) \tag{3.14}$$

where k_B is the Boltzmann's constant and T_0 is the desired temperature.

In NAMD the keyword `langevin` turns on the LD option. In this case `langevinTemp` must be set to the desired temperature. The value of the damping constant ξ is given by `langevinDamping` expressed in ps^{-1}. There is also an option to turn off LD specifically for hydrogen atoms using the keyword `langevinHydrogen`. This may be used to improve stability of the integrator [16].

3.2.2 Molecular Dynamics at Constant Pressure

Berendsen pressure bath coupling

This pressure control method is analoguos to the weak coupling with heat bath method. The idea is to couple the system to a pressure bath and the pressure can be kept at a constant value by simply scaling the volume [17]. The rate of change of pressure is given by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{bath} - P(t)) \tag{3.15}$$

where τ_P is the coupling constant, P_{bath} is the pressure of the bath, and $P(t)$ is the pressure at time t. The volume of the simulation box is scaled by a factor λ which is

$$\lambda = 1 - \kappa\frac{\delta t}{\tau_P}(P - P_{bath}) \tag{3.16}$$

NAMD keywords for turning Berensen Pressure method on is `BerendsenPressure`. `BerendsenPressureTarget` and `BerendsenPressureRelaxationTime` specify the required pressure and the relaxation time of Berendsen's method respectively.

Extended system method

This constant pressure method was originally proposed by Andersen [20]. In this method, an extra degree of freedom, corresponding to the volume of the simulation box which adjusts itself to equalize the internal and applied pressures, is added to the system. The kinetic energy associated with this degree of freedom is $\frac{1}{2}W(dV/dt)^2$, where W is the mass of the piston and V is the volume of the system. The equations of motion for the

particles and the volume of the system are

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m} + \frac{1}{3}\frac{\dot{V}}{V}\mathbf{r}_i \tag{3.17}$$

$$\dot{\mathbf{p}}_i = \mathbf{f}_i - \frac{1}{3}\frac{\dot{V}}{V}\mathbf{p}_i \tag{3.18}$$

$$\ddot{V} = \frac{1}{W}[P(t) - P_{ext}] \tag{3.19}$$

where $P(t)$ is the pressure at time t, $\mathbf{f}_i$ is the force acting on the particle i.

Langevin Piston Algorithm

This algorithm was proposed by Scott E. Feller and colleague [21]. It was developed from the extened system method and Berendsen's method for molecular dynamics simulation at constant pressure. The equations of motion of the particles in the system are

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m} + \frac{1}{3}\frac{\dot{V}}{V}\mathbf{r}_i \tag{3.20}$$

$$\dot{\mathbf{p}}_i = \mathbf{f}_i - \frac{1}{3}\frac{\dot{V}}{V}\mathbf{p}_i \tag{3.21}$$

$$\ddot{V} = \frac{1}{W}[P(t) - P_{ext}] - \gamma\dot{V} + \Gamma(t) \tag{3.22}$$

where γ is the collision freequency and $\Gamma(t)$ is a random force taken from a Gaussian distribution with zero mean and vairance

$$\langle\Gamma(0)\Gamma(t)\rangle = \frac{2\gamma k_B T\delta(t)}{W} \tag{3.23}$$

where k_B is Boltzmann's constant.

Nosé-Hoover constant pressure method

Nosé-Hoover method is another method that is used to do the MD simulation at constant pressure. The volume is considered as a dynamical variable that changes during the simulation. The equations of motion proposed by Martyna *et al.* [22] for the positions

and the momenta are

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} + \frac{p_\epsilon}{W}\mathbf{r}_i \tag{3.24}$$

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \left(1 + \frac{d}{N_f}\right)\frac{p_\epsilon}{W}\mathbf{P}_i - \frac{p_\xi}{Q}\mathbf{P}_i \tag{3.25}$$

where N_f is dN. The equation of motion 3.24 and 3.25 are complemented with an equation of motion for the volume in d dimensions and N particles.

$$\dot{V} = \frac{dVp_\epsilon}{W} \tag{3.26}$$

$$\dot{p}_\epsilon = dV\left(P_{int} - P_{ext}\right) + \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i} - \frac{p_\xi}{Q}p_\epsilon \tag{3.27}$$

In these equations P_{ext} is the external pressure. P_{int} is the internal pressure, which can be calculated during the simulation

$$P_{int} = \frac{1}{dV}\left[\sum_{i=1}^{N}\left(\frac{\mathbf{p}_i^2}{m_i} + \mathbf{r}_i \cdot \mathbf{F}_i\right) - dV\frac{\partial E(V)}{\partial V}\right], \tag{3.28}$$

where E is the potential.

NAMD has a method, namely, Langevin piston Nosé-Hoover method for doing the constant pressure simulation. It is a combination of the Nosé-Hoover constant pressure method with the Langevin piston method [16]. The keyword `LangevinPiston` is for turning this method on. `LangevinPistonTarget` and `LangevinPistonTemp` specify the desired pressure and temperature respectively.

3.2.3 Molecular Dynamics Algorithm

A general MD simulation consists of steps that are given in Figure 3.1. Firstly, we initialize the simulated system by giving initial positions $\mathbf{r}_i(0)$, velocities $\mathbf{v}_i(0)$ for each particle and the parameters that detail the conditions of the MD run (the number of particles, initial temperature, time step, density). The next step is to compute the potential energy and the forces on all particles. After this we solve the equations of motion (3.5) by using the Verlet or Leap-Frog algorithm to update the positions and velocities. In the case that the methods of controlling the temperature and pressure are applied, these methods are used after solving these equations of motion. For example,

if the Berendsen's coupling methods of constant temperature and pressure are used, the positions, velocities are updated as follows:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\frac{\mathbf{F}(t)}{m}\Delta t^2 \tag{3.29}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{1}{2m}\left(\mathbf{F}(t+\Delta t) + \mathbf{F}(t)\right)\Delta t \tag{3.30}$$

scale: $\mathbf{r}, \mathbf{v}$

$$\mathbf{r}(t+\Delta t) \leftarrow \lambda_P^{1/3}\mathbf{v}(t+\Delta t) \tag{3.31}$$

$$\mathbf{v}(t+\Delta t) \leftarrow \lambda_T\mathbf{r}(t+\Delta t) \tag{3.32}$$

where

$$\lambda_P = 1 - \kappa\frac{\Delta t}{\tau_P}\left(P(t) - P_{bath}\right) \tag{3.33}$$

$$\lambda_T = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right)\right]^{1/2} \tag{3.34}$$

These steps and previous one are repeated until the calculation of a time evolution of the simulated system reaches the desired length of time. After the completion of the central loop, the average of measured quantities are computed and printed, and then the MD simulation stops.

In our works, MD simulations were carried out in isothermal-isobaric ensemble. The number of particles (N), the pressure (P) and the temperature (T) of the system were maintained at the constant values. This means the methods for controlling the temperature and pressure as described in 3.2.1 and 3.2.2 were used together to produce trajectories in NPT ensemble.

3.2.4 Stages for performing a common MD Simulation

The performing MD simulation can be divided in six stages including Initial input, Energy minimization, Heating, Equilibration, Production MD Simulations, and Data analysis.

- **Initial input** The stage of preparing of the structure of the simulated system (such as proteins or enzymes). The input data for MD simulations are coordinates of each atom and the molecular topology including force constants. The coordinates for simulated proteins can be taken from the Protein Data Bank (PDB) and the

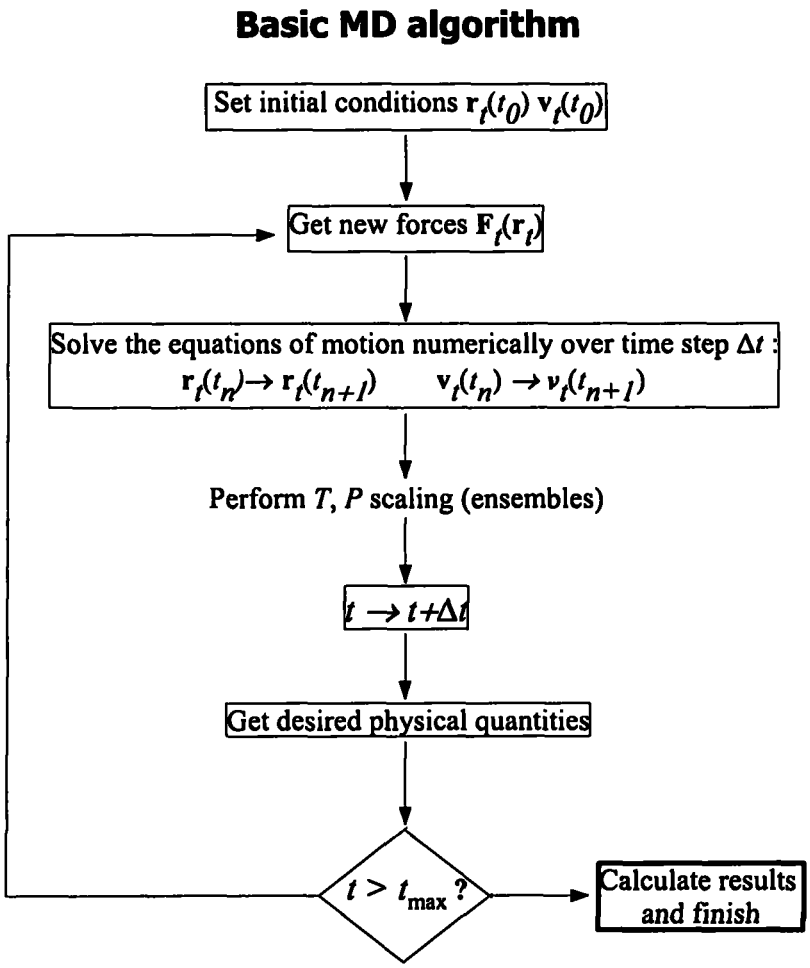


Figure 3.1: General MD algorithm

molecular topology can be generated using software packages such as AMBER or CHARMM.

- **Minimization** The purpose of the energy minimization stage is to reduce the potential energy of the simulated system. It is possible that random initial coordinates lead to steric conflicts between atoms. There are several widely used minimization methods such as *steepest descent* and the *conjugate gradient* method.
- **Heating** The temperature of the system is linearly increased from 0 K to a desired value. The temperature of the system involves the velocity of each atoms in the system. Consequently, the initial distribution of velocities can be drawn from the Maxwell-Boltzmann distribution.
- **Equilibration** The kinetic and potential energies of the system are equilibrated to become stable using a temperature control method. The Langevin dynamics method may be applied.
- **Production MD Simulations** is performed in a desired thermodynamic ensemble to produce MD trajectories sampling the structural characteristics and dynamics of the simulated system.
- **Data analysis** MD trajectory is used to analyze dynamical behavior and calculate physical quantities of the simulated system.

3.3 General Concepts

3.3.1 Periodic Boundary conditions

To reduce the size of a simulated system which has many particles and no boundary, the periodic boundaries are used. Periodic boundaries enable the number of particles in the simulated system remains constant. If there is a particle leaving the simulated system from one site, there is another image-particle entering the simulated system from the opposite side [19]. If the simulated system is a cubic box, we recreate the system and place its exact copies in all the directions to fill the entire 3D space. The original system is called unit cell, all other cells are called images. The atom in a unit cell may now interact with other atoms from the unit cell as well as the atoms from simulated system's images.

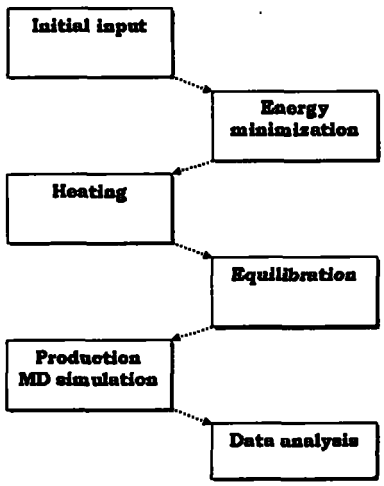


Figure 3.2: Stages for performing a common MD simulation

The NAMD configuration keywords for setting up the unit cell are `cellBasisVector` and `cellOrigin`. `cellBasisVector1(2,3)` are the components of the three basis vectors, which specify the geometry of the unit cell and its dimensions. `cellOrigin` is the position of the unit cell centre [16].

3.3.2 Ewald method

The Ewald summation method is the method for computing the electrostatic interation in a periodic system. In this technique, a charge particle i interacts with all the other particles in the simulation box (the unit cell) and with all of their images $\mathbf{n}$ in an infinite array of periodic cells [15, 23]. Consequently, for the neutral system($\sum_i q_i = 0$), the electrostatic energy is given by

$$E_{elec} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i), \tag{3.35}$$

where $\phi(r_i)$ is the electrostatic potential at the position of charge particle i:

$$\phi(r_i) = \sum_{j,\mathbf{n}}{}' \frac{q_i}{|\mathbf{r}_{ij} + \mathbf{n}L|} \tag{3.36}$$

where the prime on the summation indicates that the sum does not include the interaction $i = j$ for $\mathbf{n} = 0$ and L is the side of the cube unit cell.

It is known that the infinite series as in the equation 3.36 are poorly converging. Ewald method overcomes this problem by considering each charge to be surrounded by a neutralising charge distribution of equal magnitude but of opposite sign, namely, the screening charge clouds and the compensating charge clouds [15, 23] and rewrites the Coulomb energy equations (3.35 and 3.36) as the sum of three components. The first term($E_{Fourier}$) is the interaction of point charges with the compensating charge clouds. Because the distribution of compensating charge clouds is periodic, this term is calculated using Fourier transform in $\mathbf{k}$ space. The second term(E_{self}) is associated with the interaction of point charges with their own compensating charge clouds. The third term(E_{real}) is the interactions of point charges with the other point charges partially screened by the screening charge clouds.[15, 23] Therefore, the expression for E_{elec} can be written in the following form:

$$E_{elec} = E_{Fourier} - E_{self} + E_{real}, \tag{3.37}$$

where the Fourier part is

$$E_{Fourier} = \frac{1}{2V} \sum_{\mathbf{k} \neq 0} \frac{4\pi}{k^2} |\rho(\mathbf{k})|^2 \exp(-\frac{k^2}{4\alpha}) \text{ and } \rho(\mathbf{k}) = \sum_{i=1}^{N} q_i \exp(\mathrm{i}\,\mathbf{k} \cdot \mathbf{r}_i) \tag{3.38}$$

the self-interaction term is

$$E_{self} = (\frac{\alpha}{\pi})^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2, \tag{3.39}$$

and and the real-space term is

$$E_{real} = \frac{1}{2} \sum_{i \neq j}^{N} \frac{q_i q_j \operatorname{erfc}(\sqrt{\alpha} r_{ij})}{r_{ij}} \tag{3.40}$$

In these equations α determines the width of Gaussian distribution of screening and compensating charge clouds, V is the volume of the unit cell.

The NAMD keyword for using Ewal method is PME which stands for *Particle-Mesh Ewald*. PME computes the trigonometric functions in the equation 3.38 which require the evaluation of the function on the grid points within the interval L. The keyword

`PMETolerance` sets the accuracy in the computation of real-space term (equation 3.40). The keywords `PMEGridSizeX(Y,Z)` specify the number of grid points in three directions. The keyword `PMEInterpOrder` determines the quality of function interpolation between grid points. [16]

3.3.3 SHAKE

The bond constraints are used in the MD simulation for decreasing the time steps which are comsumed in the bond vibrations of light atoms (hydrogens). The most commonly used method for applying constraints in molecular dynamics is the SHAKE algorithm of Ryckaert *et al.* [15]. In the SHAKE algorithm, the atom positions are updated from

$$\mathbf{r}_i^{constrained}(t+\Delta t) = \mathbf{r}_i^{unconstrained}(t) - \frac{\Delta t^2}{m_i}\sum_{k=1}^{l}\lambda_k\frac{\partial\sigma_k(t)}{\partial\mathbf{r}_i} \tag{3.41}$$

where λ_k is the Lagrange multipier and σ_k is the function

$$\sigma(\mathbf{r}_i,\mathbf{r}_j) = r_{ij}^2 - d_{ij}^2 \tag{3.42}$$

where d_{ij} is the fixed distance.

SHAKE algorithm is used in NAMD to constrain bond lengths [16]. The keyword `rigidBonds` is for constraining all bonds involving light atoms. The `rigidTolerance` specifies the tolerance of bond length convergence and `rigidIterations` is the maximum number of SHAKE iterations.

3.3.4 TIP3P Water Model

In our works, the TIP3P water model was used in the solvation of the HIV-1 RT protein. In the TIP3P model, each water molecule is maintained in rigid geometry and the interaction between molecules is described by using Coulomb and Lennard-Jones potential functions [15, 24]. The TIP3P model considers a molecule of water as a rigid ball in the computing of Van der Waals interaction between two water molecules. The centre of each rigid ball are at the oxygen atoms. The hydrogen atoms are not used in the calculation of Van der Waals interactions.

The parameters of the TIP3P water model are shown in the table 3.2. The parameters

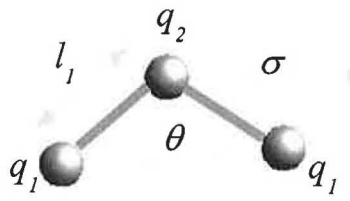


Figure 3.3: The TIP3P simple water model.

σ and ϵ are the Lennard-Jones parameters (see equation 3.4).

TIP3P					
σ Å	ϵ kJ/mol	l_1 Å	q_1(e)	q_2(e)	$\theta°$
3.15061	0.6364	0.9572	+0.4170	-0.8340	104.52

Table 3.2: TIP3P water model parameters.

The command `SolvateBox` with the parameter `WATBOX216` in the LEAP module from the AMBER software package was applied to build the box of TIP3P waters in our works.

Chapter 4

Simulated Systems: Building and Conformational Analysis

4.1 Simulated systems

The main objective of our works is to study the dynamics of the HIV-1 RT enzyme in complex with DNA:DNA and inhibitors (adefovir and tenofovir) by using MD simulations. We were interested in subtle differences in binding of the inhibitors into the HIV-RT polymerase active site before and after they were incorporated into the primer DNA strand.

To study inhibitors before their incorporation into the primer DNA strand, the adefovir diphosphate (PMEApp) and tenofovir diphosphate (PMPApp) were given into the polymerase active site. Of course, the reference MD simulations with natural substrates - adenosine triphosphate (ATP) and deoxyadenosine triphosphate (dATP) were produced as well. Moreover, the MD simulations of HIV-1 RT and DNA:DNA with DNA primer strand terminated by adefovir and tenofovir inhibitors were computed. Again, the reference MD simulations of HIV-1 RT complexed with the primer DNA strand terminated at the 3'-end by adenosine and deoxyadenosine were also carried out. Table 4.1 shows differences between the simulation systems. Each structure was solvated with water molecules. It led to simulate the systems consisting of 130,000 atoms. Extremely time-consuming MD simulations were produced using the large multiprocessor systems.

1	1T05 HIV-1 RT	+	DNA:DNA	+	PMEApp	+	Solvent
2	1T05 HIV-1 RT	+	DNA:DNA	+	PMPApp	+	Solvent
3	1T05 HIV-1 RT	+	DNA:DNA	+	3′-Terminal Deoxyadenosine	+	Solvent
4	1T05 HIV-1 RT	+	DNA:DNA	+	3′-Terminal Adenosine	+	Solvent
5	1T05 HIV-1 RT	+	DNA:DNA	+	dATP	+	Solvent
6	1T05 HIV-1 RT	+	DNA:DNA	+	ATP	+	Solvent
7	1T05 HIV-1 RT	+	DNA:DNA	+	Adefovir	+	Solvent
8	1T05 HIV-1 RT	+	DNA:DNA	+	Tenofovir	+	Solvent
9	1T05 HIV-1 RT	+		+		+	Solvent

Table 4.1: The difference of each simulation system. The systems no. 1, 2, 5 and 6 are the simulation systems of HIV RT complexed with the inhibitors and nucleotides before the incorporation with the DNA primer whilst the systems no. 3, 4, 7 and 8 are the simulations of HIV RT complexed with the inhitors and the terminal nucleosides after the incorporation with the DNA primer.

4.2 Building of simulated structures

There are many crystal structures of HIV-1 RT determined by X-ray methods and available for download from the Protein Data Bank(PDB),The 1T05 HIV-1 RT was chosen to serve as an initial structure in MD simulations. This HIV-1 RT structure is in complex with two DNA strands and there is a molecule of tenofovir diphosphate (PMPApp) bound in the polymerase active site.

The LEAP module from the AMBER software package [13] and the CHIMERA [25] program were used to modify coordinates of substrate atoms in 1T05 to produce all simulated systems listed above. The topology and initial coordinate files - input files for MD simulations - were generated using the LEAP module and the AMBER force field.

The 1T05 structure can be divided into three groups: a protein group composed of two chains of HIV-1 RT, a group of DNA strands and a group of remaining atoms and molecules: magnesium ions, water molecules and tenofovir. Initially, for the MD simulation of free HIV-1 RT in water (system no.9), the non-protein molecules were removed from 1T05. The hydrogen atoms were added by using the LEAP program from the AMBER software package. The molecules were then solvated with a large box of Monte Carlo TIP3P water moleucles. Each box side reaches at least 10 Å away from the nearest atoms of HIV-1 RT molecule.

For preparation of the simulation structures of solvated HIV-1 RT complexed with DNA:DNA and incoming residues (system no. 1-8 in the table 4.1), the water molecules and glycerol (GOL) were deleted from 1T05. The residues MRG[1] and DDG[2] in the primer chain in positions 817 and 822, respectively, were replaced by deoxyguanosine monophosphate (dGMP) to pair with the deoxycytidine monophosphate (dCMP) template residues. For replacing of MRG and DDG residues with dGMP we used the coordinates of all atoms from MRG and DDG, moreover, the oxygen atom was anchored at the $C_{3'}$ position of the sugar ring.

To produce initial structures (systems no. 1, 5 and 6) tenofovir-diphosphate(DP) at residue 823 was just replaced with the adefovir-DP, dATP and ATP moieties respectively. The adenine moiety of tenofovir-DP was used to create corresponding parts of adefovir-DP, ATP and dATP. Again, these systems were then solvated with TIP3P water molecules resulting in water box sides of wchich reached at least 10 Å away from the nearest atoms of the solute.

The initial structures of the systems no. 3, 4, 7 and 8 were prepared by the same way as the systems no. 1, 2, 5, and 6. The structure of tenofovir-DP was modified to produce terminal residues covalently linked to the 3′-primer terminus of DNA:DNA. These terminal residues were adefovir in the system no. 7 and deoxyadenosine or adenosine in the system no. 3 and 4 respectively.

4.3 Methods used in MD simulations

Our simulated systems consist of biomolecules and results of produced MD simulations should be relevant for comparison with real experiments which are made under conditions of constant temperature(T) and pressure(P). Therefore the isothermal-isobaric NPT ensemble was chosen. Performed MD simulations followed the stages explained in chapter 3. The NAMD software package was used for it.

Initially, all simulation structures were energy minimized. The smooth Particle-mesh Ewald (PME) method was employed for long-range electrostatic forces. The non-bonded cutoff for van de Waals interactions was set to 9 Å. The periodic boundaries were used to keep the number of particles constant. The SHAKE algorithm was applied to constrain bonds where the hydrogen atoms were involved. In the heating and production stages,

[1]MRG≡ $N2$ – (3 – *mercaptopropyl*) – 2′ – *deoxyguanosine* – 5′ – *monophosphate*
[2]DDG≡ 2′, 3′ – *dideoxy* – *guanosine* – 5′ – *monophosphate*

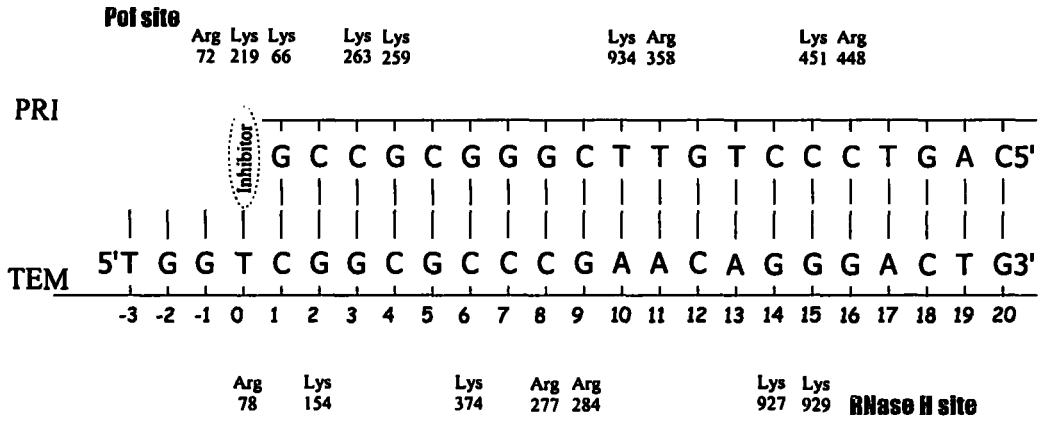


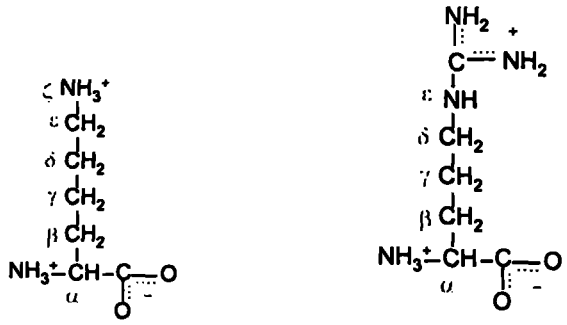
Figure 4.1: The schematic structure of double helix we used in the simulation systems no. 1-8.

the Langevin dynamics was used for a temperature control. The simulated systems were heated from 0 K to 310 K. The 310 K temperature is close to a normal human body temperature. The Langevin piston method was applied to reach an efficient pressure control with target pressure set to 1 atm. MD runs of length 2.5 ns were produced. Examples of NAMD keywords we used to control MD simulations can be found in Appendix .

4.4 Analysis of Simulation Data

All MD trajectories were analyzed using the PTRAJ module from the AMBER software package ([13]). Interactions between HIV-1 RT and nucleic acids - DNA:DNA and incoming subtrates - mediated by Arginine and Lysine residues, conformational preferences of furanose rings, internucleotide linkages and Watson-Crick hydrogen bonding were analyzed in detail.

Pairs of residues in DNA:DNA were named as follows: the first pair which is a pair of the inhibitor from the primer chain (PRI) and a nucleotide from the virus template (TEM) is named as "Base Pair 1" (BP1). Other base pairs are numbered increasingly toward the ends of DNA:DNA as shown in Figure 4.1. Figure 4.1 also shows three unpaired nucleotides called as TEM -3, TEM -2 and TEM -1. Further, there are indicated positions of the HIV-1 RT Arginine and Lysine residues (positively chaged) which are in contact with DNA strands (with negatively charged backbones).



Lysine **Arginine**

Figure 4.2: The chemical structure of Lysine and Arginine. From the figure, N_ζ of lysine and guanidinium of arginine are at the top their structures.

4.4.1 RT Protein-Nucleic Acid Interactions

We studied the interactions between RT and DNA double helix by considering the distances of the salt bridges which formed between phosphate (P) of nucleosites and positively charged amino acid side chains. The positively charged amino acid side chains we chose in this measuring were N_ζ of lysine (Lys) and guanidinium of arginine (Arg).

From 1T05 structure, there are ten nearest lysine and six nearest arginie residues around the DNA double helix. Four of these lysines are at the positions of template (TEM) 2, 6, 14 and 15 and the other six lysines are at the positions of primer (PRI) 0, 1, 3, 4, 10 and 15 (see Figure 4.1). The arginine residues are at TEM 0, 8, 9 and PRI 0, 11, 16. The distances between these N_ζ of lysines, guanidinium and their nearest phosphates were calculated at each time step from the trajectory files.

4.4.2 Sugar Puckering

Sugar puckering is a method for studying the conformation of furanose ring [26]. The furanose ring structure constists of five atoms: four carbons and one oxygen, as shown in Figure 4.4. The conformation of this five-membered ring can be divided into two main forms: envelope (E) and twist (T) forms (Figure 4.3). In E form, four of the five atoms lie in a plane and one deviates from this plane. In T form, three atoms lie in the same plane and the other two lie on opposite sides of this plane. For more specific structure, atoms displaced from these three- or four- atoms planes and on the same side as $C_{5'}$ in

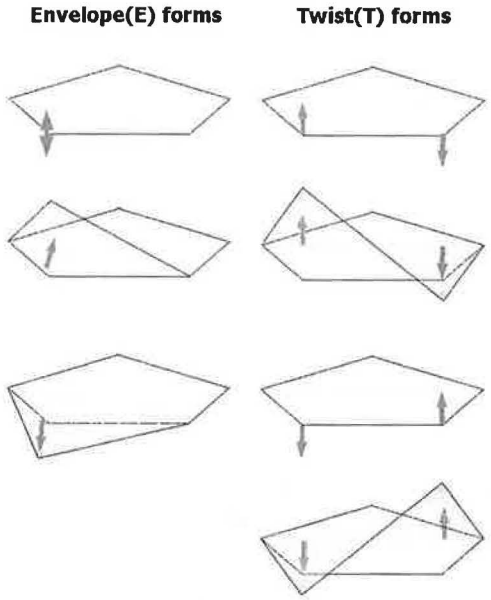


Figure 4.3: Puckering of furanose ring

Figure 4.4 are called *endo*; those on the opposite side are called *exo*[27].

Sugar puckering Modes

All posible conformations of furanose ring can be classified by using its torsion angles and the pseudorotation phase angle [26, 27]. In our case the pseudorotation phase angles P of nucleotides is calculated from the endocyclic sugar torsion as follows

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2 \cdot \nu_2 \cdot (\sin 36° + \sin 72°)} \tag{4.1}$$

where the torsion angles and their invloved atoms are shown in Table 4.4.2.

In our works the pseudorotation phase angles were calculated from every furanose rings in the DNA template and DNA primers. The PTRAJ command we used in this calculation was `pucker`.

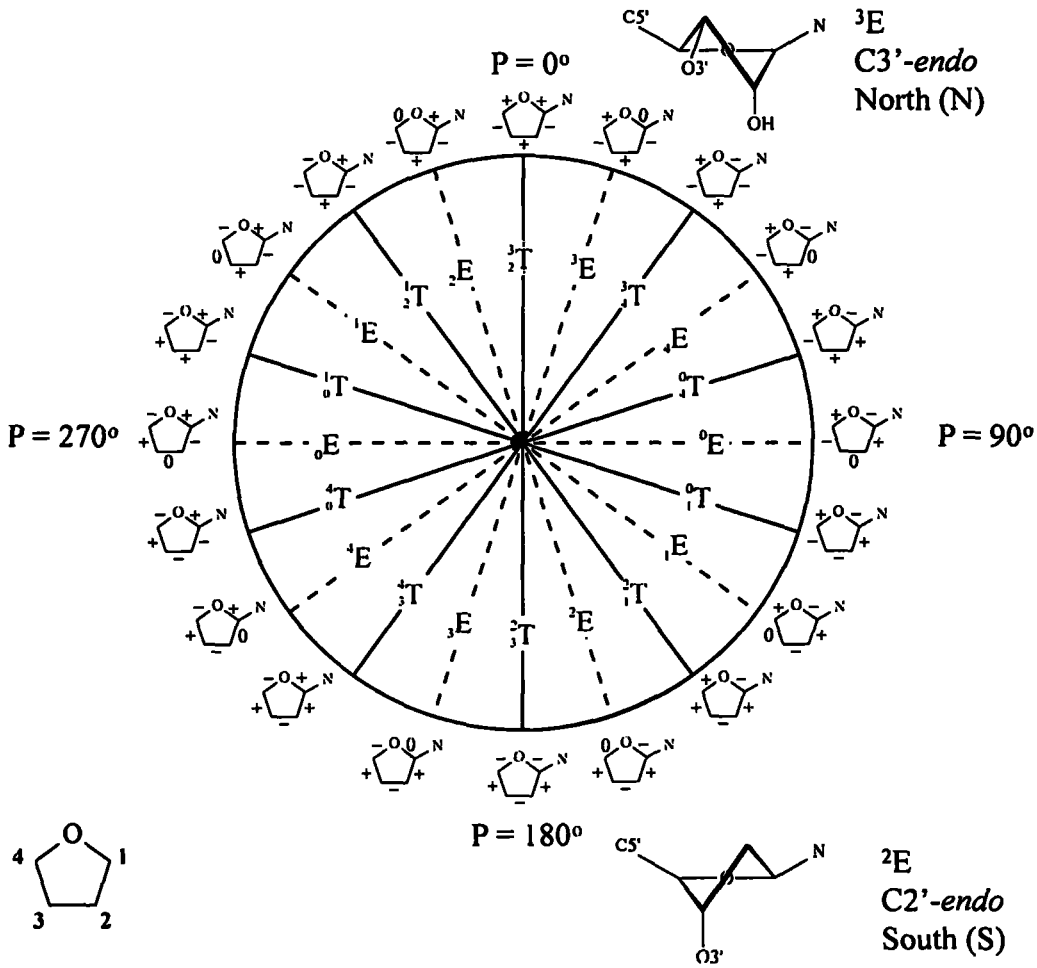


Figure 4.4: Pseudorotation cycle of the furanose ring in nucleosides.(*from* Eveline, *et al* [28].)

Torsion angle	Atoms involved
ν_0	$C_{4'} - O_{4'} - C_{1'} - C_{2'}$
ν_1	$O_{4'} - C_{1'} - C_{2'} - C_{3'}$
ν_2	$C_{1'} - C_{2'} - C_{3'} - C_{4'}$
ν_3	$C_{2'} - C_{3'} - C_{4'} - O_{4'}$
ν_4	$C_{3'} - C_{4'} - O_{4'} - C_{1'}$

Table 4.2: The definition of torsion angles of furanose ring.

4.4.3 Hydorgen bonding in Duplex Oligonucleotides

Hydrogen bond is the non-bonded interaction as described in the subsection 3.1.2. It is defined as any cohesive interaction X-H$\cdots$Y, where H carries a positive charge and Y a negative (partial or full) charge, and the charge on X is more negative than on H[27]. The conventional hydrogen bonds formed by the nucleotide bases involve N-H$\cdots$O and N-H$\cdots$N interactions.

In the complementary base pairing, purines interact with pyrimidines so that Adenine(A) pairs with Thymine(T) (or Uracil in RNA) and Guanine pairs with Cytosine(C) (Figure 4.5). These pairs are termed as Watson-Crick base pairs. We studied the strength of hydrogen bonds by measuring the distances between N-H and N in every base pair at each time step of the simulations.

4.4.4 Root mean square deviation (RMSD)

One of the quantities we use in studying the conformation and the stability of biomolecule structure during the simulations is the root mean square deviation (RMSD). The root mean square deviation of atoms in a molecule with respect to a reference structure can be computed as follows

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i |\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)|^2 \right]^{\frac{1}{2}} \tag{4.2}$$

where $M = \sum_{i=1}^{N} m_i$ is the total mass and $\mathbf{r}_i(t)$ is the position of atom i at time t. From the trajectory files, we calculated the RMSD of RT and DNA chains by using command `rms` in PTRAJ.

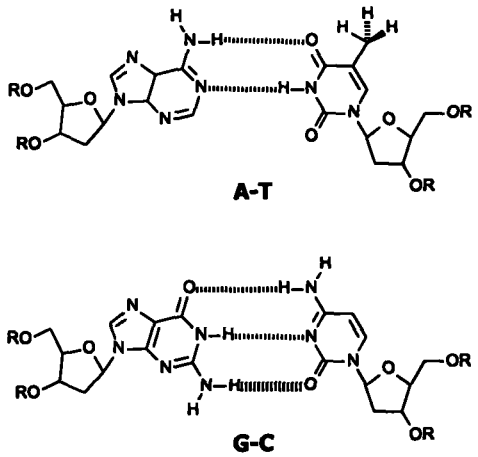


Figure 4.5: The A·T and G·C Watson-Click base pairs. The dashed lines represent hydrogen bonding interactions.

We did the RMSD calculation for the whole structures of RT and DNA chains. The RMSD of each part of RT and DNA chains were also carried out. We calculated the RMSD of fingers, thumb, palm, connection and RNase H parts in the RT structure. For the DNA chains, we divided them into four regions as the same as regions in the work of Stefan G.Sarafianos, *et al* [29] and then calculated the RMSD of each region.

4.4.5 Torsion angles of DNA backbone

The backbone of a DNA chain consists of a repeating unit of six single bonds as shown in Figure 4.6, namely, $P-O5'$, $O5'-C5'$, $C5'-C4'$, $C4'-C3'$, $C3'-O3'$ and $O3'-P$. The torsion angles about these bonds are denoted by the symbols α, β, γ, δ, ϵ, ζ, respectively. The involved atoms of each torsion angles are as shown in Figure 4.6. The torsion angles that we chose for studying and calculated were α and ζ.

4.4.6 Helical Analysis of DNA:DNA strands

From the trajectory files, we studied the DNA:DNA geometry by means of helical parameter which are used in the CURVES program(see Figure 4.7). The orientation of nucleotide base pairs were also studied in various forms of helical parameters.

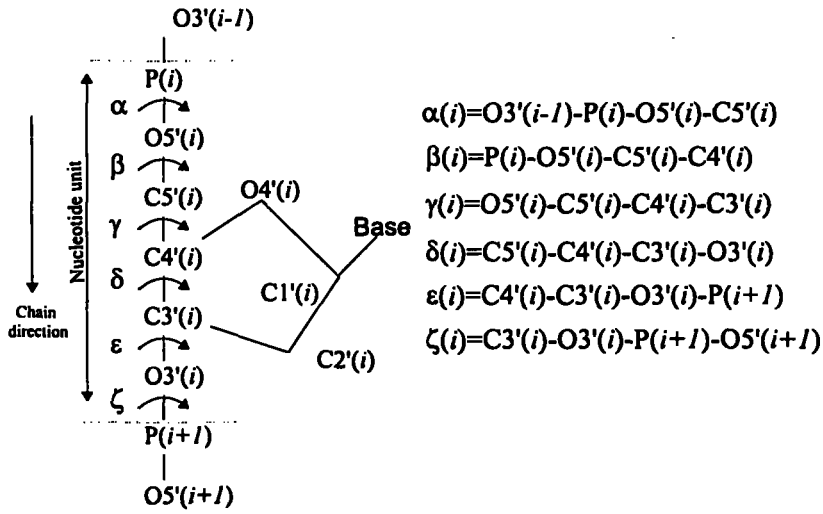


Figure 4.6: Torsion angles for backbone conformations of the i th nucleotide in polynucleotide chains

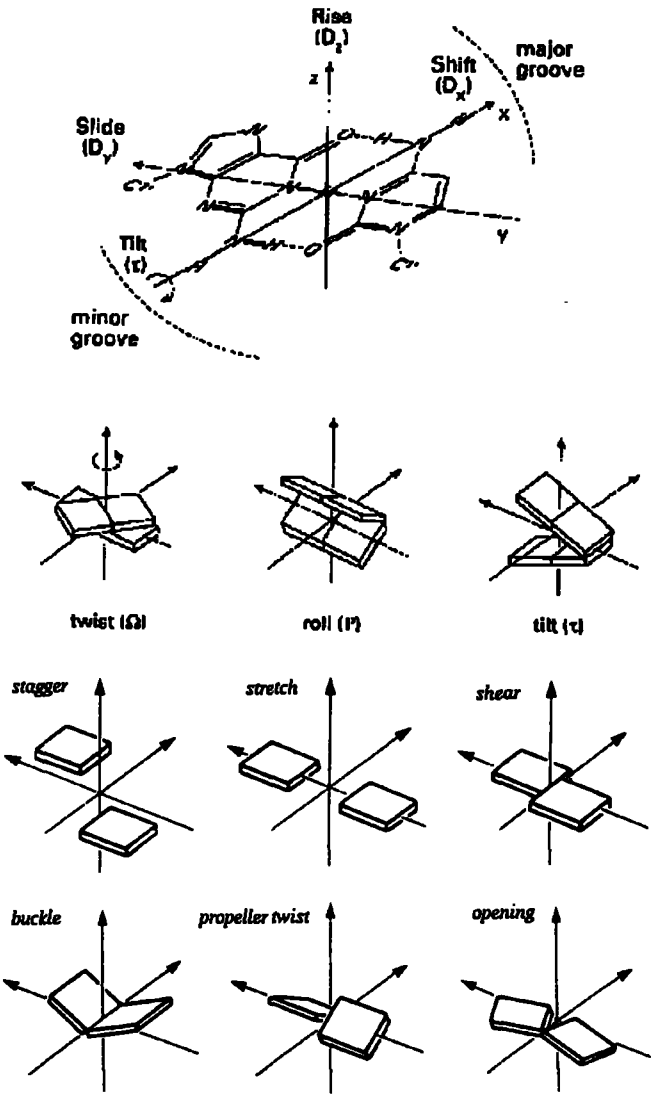


Figure 4.7: The nine inter base pair parameters (*from [31]*).

Results and Discussion

Chapter 5

Root Mean Square Deviations and Atomic Fluctuations

(see Figures C1-C11)

Time development of root mean square (RMS) deviations between the initial structures and the structures obtained from the 2.5 ns molecular dynamics trajectories were calculated for the HIV-RT/DNA:DNA complex as well as unliganded/free HIV-RT. Furthermore, an insight into a dynamic behavior of the structures was derived from atomic fluctuations computed for individual residues.

Comparison DNA:DNA 1-44 vs. HIV-RT 1-967 (see Figures C1 and C2)

The calculated structures remained close to the initial crystal structures during the entire simulations. RMS values of free HIV-RT were a little bit higher (4 Å) in comparison with bound HIV-RT (3 Å). It indicates a certain adaptation of the free HIV-RT structure after the DNA:DNA duplex was removed from the initial X-ray structure. In addition, RMS values uncovered higher overall conformational variability of the DNA:DNA duplex structure comparing to the HIV-RT enzyme.

Detail analysis given below shows RMS for different segments of the HIV-RT enzyme and DNA:DNA. The structure of the 24-mer-template/20-mer-primer double helical DNA:DNA complex was divided as follows. Region I contains 4 base pairs near the polymerase active site. Region II consists of the next 4 base pairs at the bend of the nucleic acid. The next 5 base pairs compose region III, followed by region IV close to the RNase H active site (see Figure 5.1).

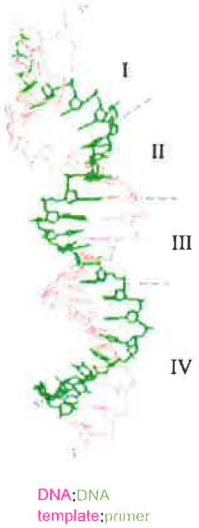


Figure 5.1: The DNA chains were divided into fours regions, namely, region I:BP 1-4, region II:BP 5-8, region III:BP 9-13 and region IV:BP 14-21.

Unpaired template DNA region res. no. 1-3 (see Figures C3-C6)

The most flexible part in all DNA:DNA structures seems to be the 5'-template overhang, which bent away from the duplex and extended across the face of the fingers subdomain. Nucleotides packed against the otherwise solvent-exposed residues Trp24, Pro25, Phe61, and Ile63.

DNA:DNA Region I BP 1-4 (see Figures C3-C6)

RMS values for DNA:DNA Region I indicates the troubles with adoption of the Adenosine substrate (structure no. 6) and primer chain terminating Tenofovir (structure no. 8) into the polymerase active site (as mentioned in detail below - see the Watson-Crick hydrogen bonding section)

DNA:DNA Region II BP 5-8 and Region III BP 9-13 (see Figures C3-C6)

Both central regions created the most rigid part of the nucleic acids structure. RMS values slightly exceeds 1 A.

DNA:DNA Region IV BP 14-21 (see Figures C3-C6)

Near the RNase H active site (there was the Mg++ ion bound by Asp443, Asp549) the position of the template backbone appeared to be close to the expected catalytic configuration for an RNA strand, even though a DNA duplex is not the correct substrate

for the ribonuclease activity. As a consequence, there were remarkable variations in RMS values for into the RNase H active poorly bound Region IV.

The p66 part of the HIV-RT enzyme was divided as follows: fingers, palm, thumb, connection, RNAse H domain.

HIV-RT p66 fingers domain - res. no. 1-85,118-155 (see Figures C7-C11)

The RMS values indicated the motion of the p66 fingers in the case of the structure no. 8 as a consequence of the troubles with adoption of the chain terminating Tenofovir. In addition, the fast adaptation of fingers of free HIV-RT was remarkable (see Figure 5.2). Interestingly, the lowest atomic fluctuations of the 1-85 segment of the p66 fingers domain were found rather in the case of poorly accommodated substrates rA (structure no. 4) and tenofovir (structure no. 8).

HIV-RT p66 thumb domain - res. no. 238-318 (see Figures C7-C11)

Again, fast adaptation of the p66 thumb domain in the case of free HIV-RT was remarkable.

HIV-RT p66 connection domain - res. no. 319-426 (see Figures C7-C11)

The connection domain seemed to be the most rigid part of the whole HIV-RT enzyme structure.

HIV-RT p66 RNase H domain - res. no. 427-554 (see Figures C7-C11)

Interestingly, the atomic fluctuations determined for the poorly bound RNase H domain were correlated with atomic fluctuations of the residues 800-900 in the p51 domain.

In summary, RMS values demonstrated the outright stability of the DNA:DNA and HIV-RT structures. The core site of their mutual contacts (DNA:DNA regions II. and III., HIV-RT connection domain) was found as the most rigid part of the complex. Terminal regions of p61 chain were clearly the most flexible parts of the enzyme structure. Detail analysis of atomic fluctuations showed, for example, that the α-helices were the most stable parts of the protein structure.

Figure 5.2: The comparison of the initial (above) and final (below) structure is shown for unliganded HIV-RT. The p66 thumb and fingers subdomains changed their conformations during the MD run.

Chapter 6

DNA:DNA geometry - helical parameters

(see Figures 6.1-6.5)

When the substrate nucleic acids are bound to enzymes, it is usual that they change their tertiary structures. Here, the DNA:DNA duplex bound along a groove (about 60 Åin length) stretching from the HIV-RT polymerase active site to the RNase H active site.

The DNA:DNA had an unusual A-like structure in the region of the polymerase active site (sugar puckering analysis is given below) but with a distinctly wider major groove and the smaller base pair tilt than standard A-conformation. RNA/RNA and RNA/DNA duplexes, which are expected to adopt an A-form conformation, are used as the template-primers during different stages of retroviral reverse transcription (initiation: tRNA/template-RNA, first DNA strand synthesis: primer-DNA/template-RNA). Because of HIV-RT must be able to use these template-primers, it may induce the DNA:DNA duplex to adopt a similar conformation.

Between BP5 and BP11, the DNA:DNA structures become gradually more B-like but the major groove remained unusually deep for B-DNA. The A- to B- form transition was accompanied by an overall bend of about 40°, centred near BP7 (see Figure 6.1). This bend is a hallmark of nucleic acids bound to a variety of polymerases. The bend is distributed over approximately four nucleotides and occurs in the vicinity of contacts with p66 thumb. Bending of the template-primer may have functional implications for RT catalysis, translocation, fidelity and/or processivity.

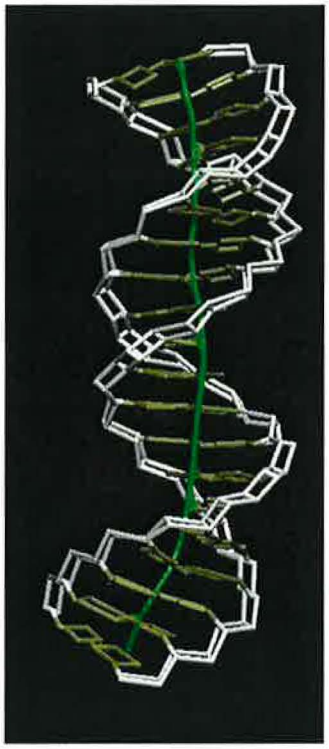


Figure 6.1: The bending of DNA:DNA structure in RT.

The variations of different structural parameters were analyzed in detail for DNA:DNA. Representative pattern of the duplex structures were obtained by averaging throughout the 2.500 ps MD trajectory. The helical parameters were then calculated separately for each respective structural motif (nucleobase, base pair or two neighboring base pairs) in the average structures by means of the CURVES program [30]. For the most of the helical parameters (the *Inclination*, *Tip*, *Shift*, *Slide*, *Rise*, *Tilt*, *Roll* and *Twist*) the systematic variations around the strands were found - again indicating adaptation of the double helical structure to fit properly into the enzyme substrate binding site (see Figure 6.2-6.5 where the values of helical parameters determined for system no.5 are shown as an example).

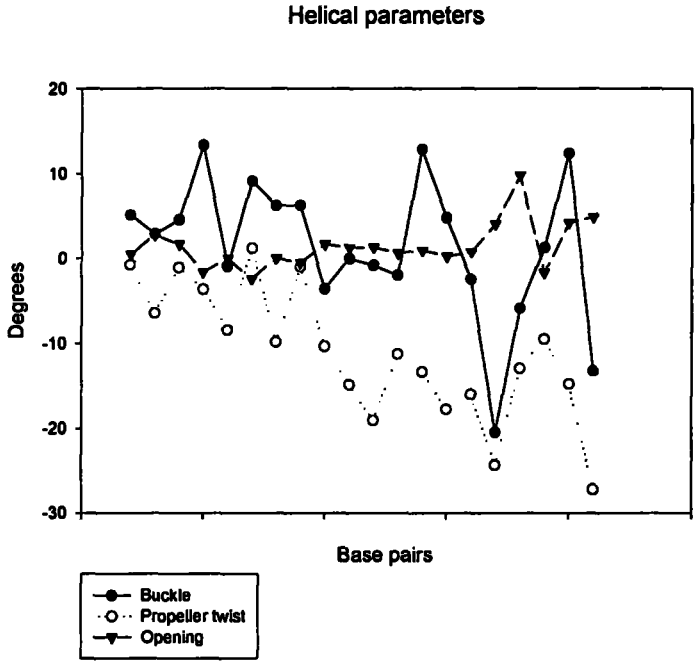


Figure 6.2: The base pair orientations by the means of helical parameters: *Buckle,Propeller Twist, Opening.*

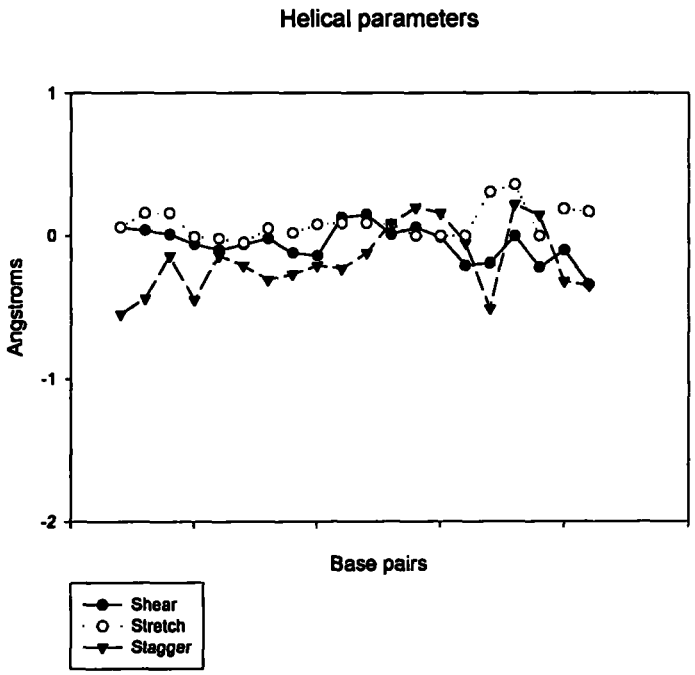


Figure 6.3: The base pair orientations by the means of helical parameters: *Shear, Stretch, Slagger.*

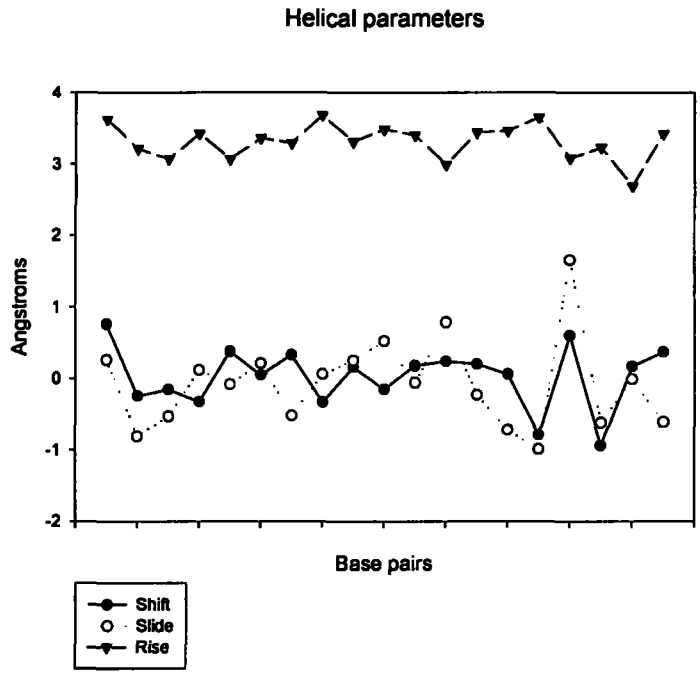


Figure 6.4: The base pair orientations by the means of helical parameters: *Shift, Slide, Rise.*

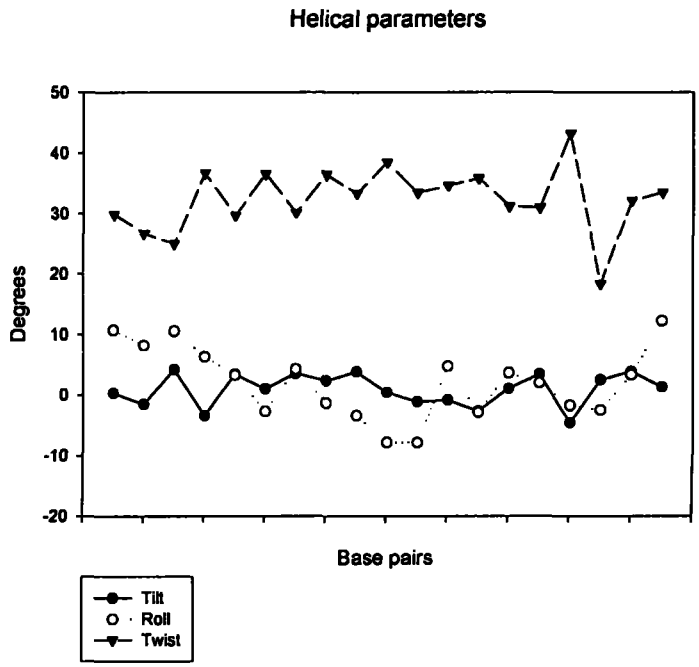


Figure 6.5: The base pair orientations by the means of helical parameters: *Tilt, Roll, Twist.*

Chapter 7

dsDNA - hydrogen bonding, sugar puckering, phosphodiester linkages

(see Figures D1-D9 for hydrogen bonding, E1-E18 for sugar puckering and F1-F18 for conformational preferences of phosphodiester linkages)

Structural consequences of DNA:DNA duplex binding to the HIV-RT enzyme were analyzed in subsequent paragraphs. There were found outright signs of rumpling by the enzyme - affection of puckering of deoxyribose moieties (*C3'*-endo conformations typical rather in the case of riboses), conformations of phosphodiester linkages (several of them were not found in usual -g-g conformations). On the other hand, the Watson-Crick hydrogen bonds seem to be unaffected with the exception of the first and second base pairs (BP1, BP2) situated in the polymerase active site.

Watson-Crick hydrogen bond connection (see Figures D1-D9)

Within the 2.5-ns MD run at 310 K, Watson-Crick hydrogen bond connections in DNA:DNA double-helical structures remained almost intact. The proton-acceptor distances of the Watson-Crick hydrogen bonds fluctuated prevailingly in the interval ranging from 1.7 Å to 2.5 Å. Even if some more remarkable fluctuation in distances of atoms participating in the Watson-Crick hydrogen bond net occurred, it showed the transient character and the base pairing was always reestablished.

It is a rather common observation that the ends of double helical structures are transiently frayed from time to time within a molecular dynamics run. However, the DNA:DNA end situated in the nearby of the HIV-RT RNase H domain (BP 21) was

stable in all structures.

On the other hand, the remarkable affection of the BP1 and/or BP2 hydrogen bond net was determined in the case of all structures with the rA residue in the polymerase active site (structure no. 4,6 I,6 II, see Figures D4, D6, D7). It indicates that the aromatic side chain of Tyr115 successfully prevented incorporation of a ribonucleoside by interfering with rA 2'-OH hydroxyl group. It secured a synthesis of pure DNA strands by the HIV-RT enzyme.

Interestingly, the adenine moiety of tenofovir-DP (structure no. 2, see Figure D2) seemed to form a non-canonical Watson-Crick base pair with the cognate template residue in the X-ray structure [8]. However, an unambiguous assignment of this hydrogen bonding pattern was beyond the resolution limit of the X-ray structure. On the other hand, biochemical data suggested there are no major differences in the mechanism of incorporation of tenofovir-DP and normal dNTP. It seems to be supported by MD simulations now, as tenofovir-DP as well as adefovir-DP (structures no. 1, no. 2) was found to form classical Watson-Crick base pairs.

In contrast, the incorporation of the chain terminating tenofovir residue into the primer strand led to the affection of either the BP2 (structure no. 8, see Figure D8) or BP1 (data not shown) hydrogen bonds. It seemed to be in agreement with X-ray data suggesting that unlike the case of the HIV-RT/DNA:DNA-ddAMP terminated complex, the end of the tenofovir-terminated primer was partially disordered and that tenofovir adopted at least two conformations [8]. In one conformation the adenine of tenofovir was in a position to stack with the base of the preceding dGMP residue of the primer strand but the tenofovir terminus does not properly base-pair with its template base. The complementary thymine on the template was not oriented for a Watson-Crick base-pairing. The second conformation of the tenofovir terminus in the binary complex had the adenine flipped out - 180° (this conformation is permitted by the absence of a complete sugar ring).

The biological role of the multiple conformations of tenofovir observed in the HIV-RT-DNA:DNA-tenofovir terminated structure (structure no.8) was unclear. It could be one from reasons of its enhanced therapeutical potency. The flexibility of a tenofovir-terminated primer might affect the partitioning of a tenofovir terminated primer between the N and P sites (before and after translocation sites) by making stable protein-DNA contacts in one of its alternate conformations. This could result in the direct inhibition of "retrotranslocation" to the N site or a failure to be oriented well enough for efficient

excision of tenofovir, if retrotranslocation does occur. It was also unclear whether the alternate conformation of tenofovir had an effect on the formation of a stable ternary complex of RT with a tenofovir-terminated template primer and an incoming dNTP.

Puckering of ribose and deoxyribose moieties (see Figures E1-E18)

The usual conformation of deoxyribose moieties in the B-DNA strand is $C2'$-endo. However, the deoxyribose moieties of both strands tended clearly to the $C3'$-endo conformation (typical for A-RNA) if situated in the nearby of the polymerase active site (residues TEM1-TEM3 and PRI1-PRI5). This was in agreement with the X-ray data [8]. The DNA had usually an A-like structure in the region of the polymerase active site. RNA/RNA and RNA/DNA duplexes, which are expected to adopt an A-form conformation, are used as template-primers during retroviral reverse transcription (initiation - tRNA/template-RNA, first strand synthesis - primer-DNA/template-RNA). Because of HIV-RT must be able to use these template-primers, it may induce the DNA/DNA duplex to adopt a similar conformation.

Moreover, the residues TEM 7, 15-17 occasionally tended to the atypical $C3'$-endo conformation too (it indicates a strong contact with the HIV-RT enzyme). The fluctuations of the pseudorotation angles in the remaining residues reached occasionally almost to $C3'$-endo, however, there was no remarkable re-puckering into this conformation.

Phosphodiester linkages (see Figures F1-F18)

The phosphodiester linkages preferred the -g-g conformation (in terms of the $C5' - O - P - O$ and $O - P - O - C3'$ torsion angles). However, in the case of PRI 13, 16, 17, 20 and TEM 15, 18, 21 residues transitions into the less obvious -gg position were found. It should be noted that roughly in this region, the remarkable variations of helical parameters as well as conformational preferences of deoxyriboses were found as described in detail before.

Chapter 8

Adefovir/Tenofovir conformational preferences

(see Figures G1 and G2)

The triphosphates of the incoming nucleotides wrapped around the Mg ion. A non-bridging oxygen from each of the phosphates (pro-Rp for α and β) contributed to the octahedral coordination of this metal; side chains from aspartates 110 and 185 completed its shell. Overall, the conformation and metal ligation of the triphosphate very closely resembled those of the dNTP in other polymerase replication complexes and its conformation was stable within a MD run. On the other hand, there could be expected some conformational variability consider the remaining parts of acyclic adefovir PMEApp or tenofovir PMPApp (structure no. 1 and no. 2, respectively). It was found, that only PMEApp underwent conformational transitions within a MD run (quantified using backbone torsion angles). Tenofovir-DP (PMPApp) carrying additional methyl group in the backbone was completely stable. Tenofovir PMPApp optimal binding into the polymerase active site could be one from reasons of its higher therapeutic potency.

Chapter 9

HIV-RT/DNA:DNA mutual interactions facilitated by ARG and LYS residues

(see Figures 9.1-9.8 and H1-H18)

The most extensive contacts between DNA:DNA and HIV-RT protein involved the sugar-phosphate backbone of the nucleic acid and the palm, thumb, and fingers subdomains of the HIV-RT p66 chain.

The only direct interactions with bases occured in the minor groove, at positions TEM-3 to TEM0. There were van de Waals contacts with Pro157 and Met184 and with Ile94 (base pairs TEM-2 and TEM-3), and hydrogen bonds between Tyr183 and TEM-1).

Contacts in the polymerase active site

The HIV-RT p66 palm contains the polymerase active site that is defined by a triad of aspartic acid residues at positions 110, 185, 186. These amino acids may bind the divalent cations that are required for catalysis. The incoming nucleotide/inhibitor pairs with the templating base TEM0. The base of the nucleotide/inhibitor stacked on the terminus of the primer strand almost as in a continuous DNA strand. The triphosphate moiety is coordinated by Lys65, Arg72 and Mg++ metal ion.

In the presented models, the guanidinium group of Arg72, which lay flat against the dNTP base, donated hydrogen bonds to the α-phosphate, and the amino group of Lys65,

Lys219 donated hydrogen bonds to the β or γ-phosphate. All of these side chains moved into position as a result of the HIV-RT p66 fingers closure. Lys65 and Lys219 seemed to be interchangeable with respect to interactions with β or γ-phopshate oxygens of the incoming nucleotide/inhibitor. Interestingly, the important Arg72 residue sandwiching incoming nucleotide/inhibitor was found to be more potently anchored to triphosphate moiety rather in the case of inhibitors than natural nucleotides (compare Figures H1,H2 vs. H5-H7).

Between nucleotides TEM0 and TEM3, the sugar phosphate backbone of the template strand contacted residues in the fingers and palm. These are largely van de Waals interactions, as might be expected for a contact along which the DNA backbone must move. However, there were found interactions facilitated by hydrogen bonds with Lys154 (more potent) and Arg78 residues too (see Figures H1-H9).

Remaining contacts

Between nucleotides PRI1 and PRI4 of the primer strand and TEM5 and TEM8 of the template strand, the minor groove faced the thumb. The primer strand contacted the loop between palm and thumb (so called primer grip). Lys374, Lys263 and Lys259 facilitate contact by donating hydrogen bonds. On the other hand, Arg277 and Arg284 were found as less potent binders (see Figures H10-H18).

Near the RNase H active site, the position of the template backbone appeared to be close to the expected catalytic configuration for an RNA strand, even though a DNA duplex is not the correct substrate for the ribonuclease activity. There were determined stable interactions with Lys934, Lys451, Lys929, Arg358 and partially with Arg448 (see Figures H10-H18).

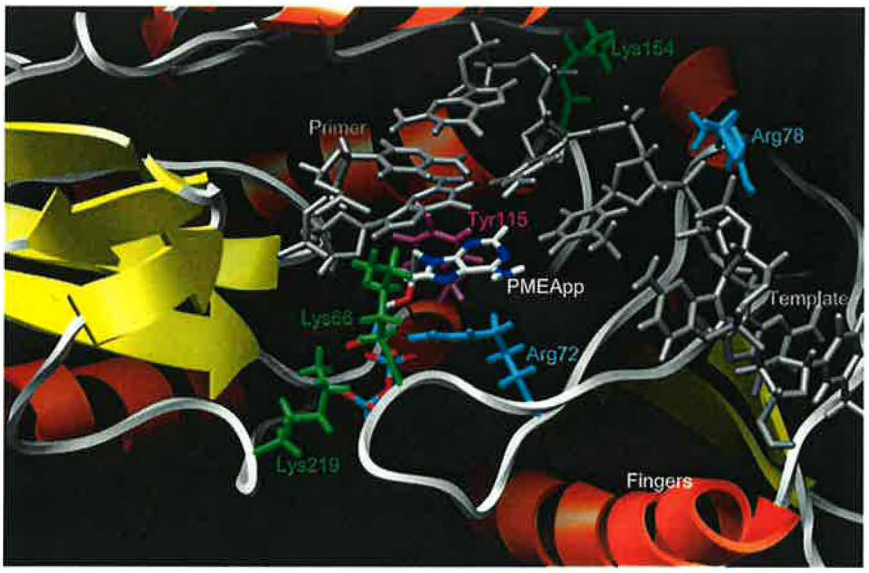


Figure 9.1: PMEApp at the polymerase active site in the structure no.1.

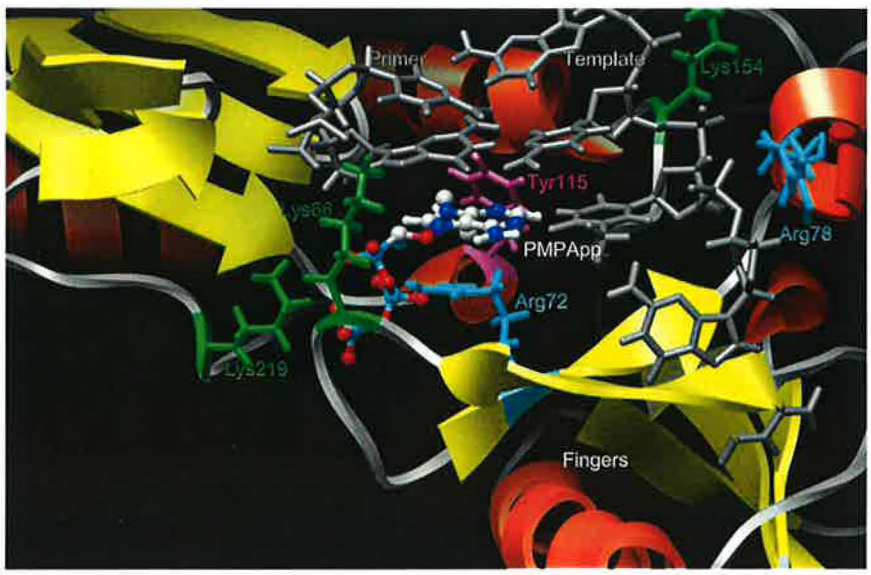


Figure 9.2: PMPApp at the polymerase active site in the structure no.2.

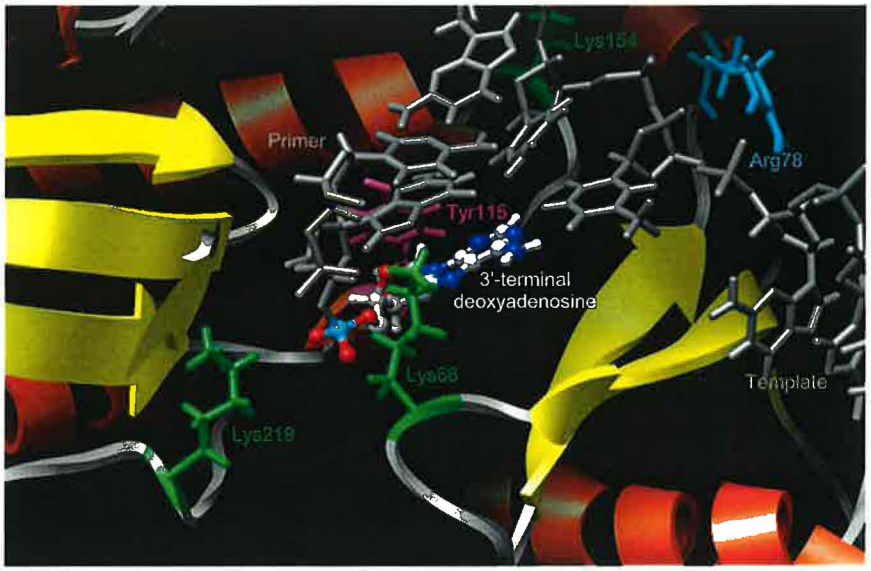


Figure 9.3: 3′-terminal deoxyadenosine at the polymerase active site in the structure no.3.

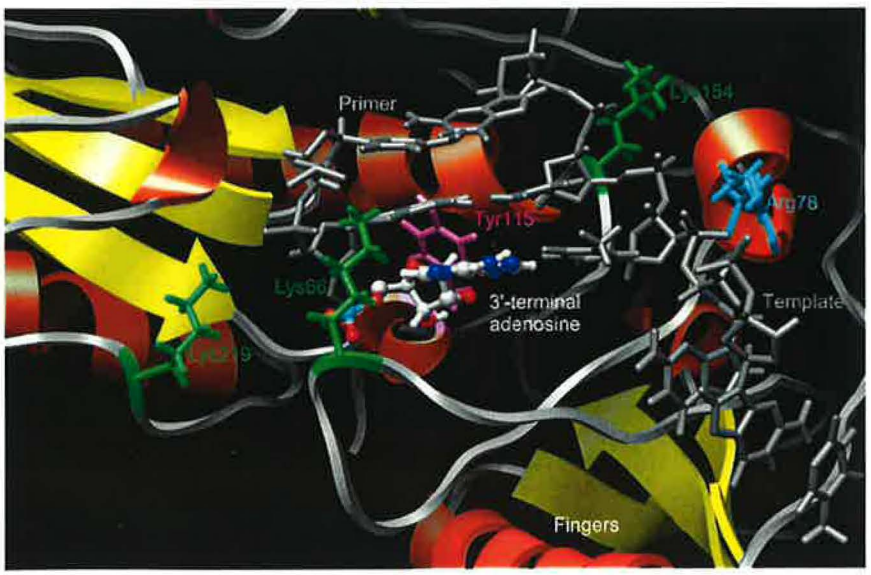


Figure 9.4: 3′-terminal adenosine at the polymerase active site in the structure no.4.

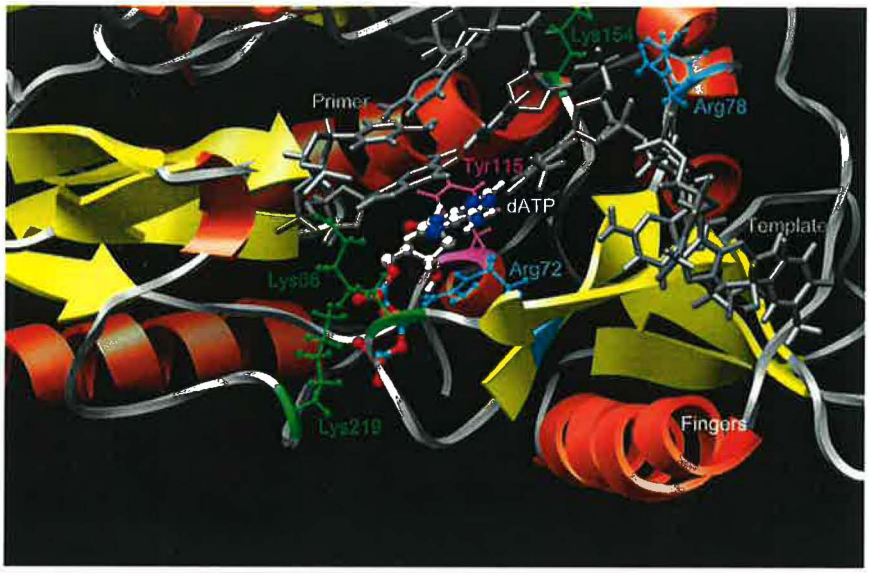


Figure 9.5: dATP at the polymerase active site in the structure no.5.

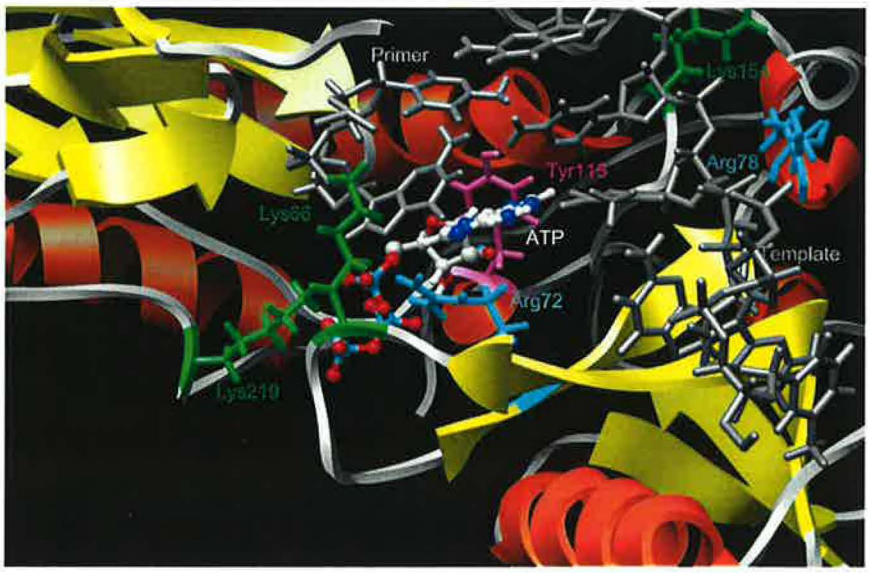


Figure 9.6: dATP at the polymerase active site in the structure no.6.

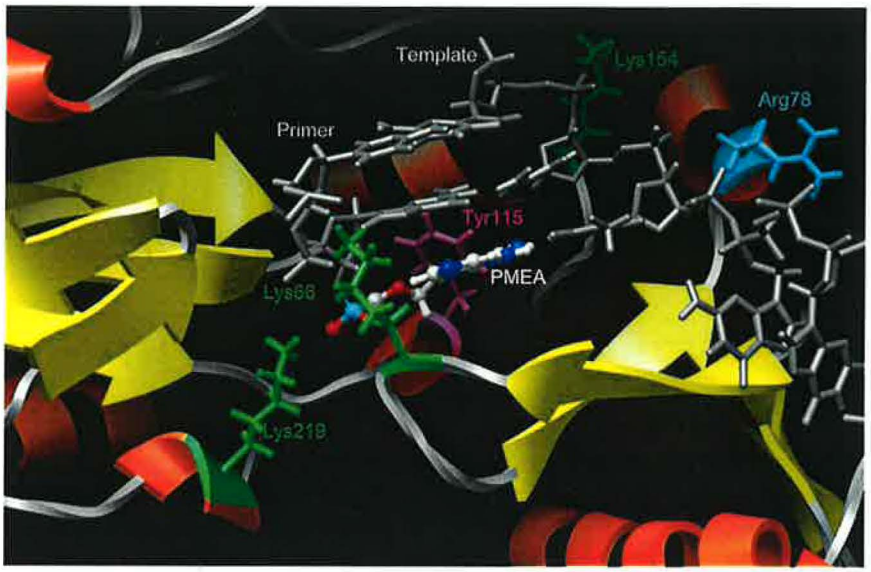


Figure 9.7: Adefovir at the polymerase active site in the structure no.7.

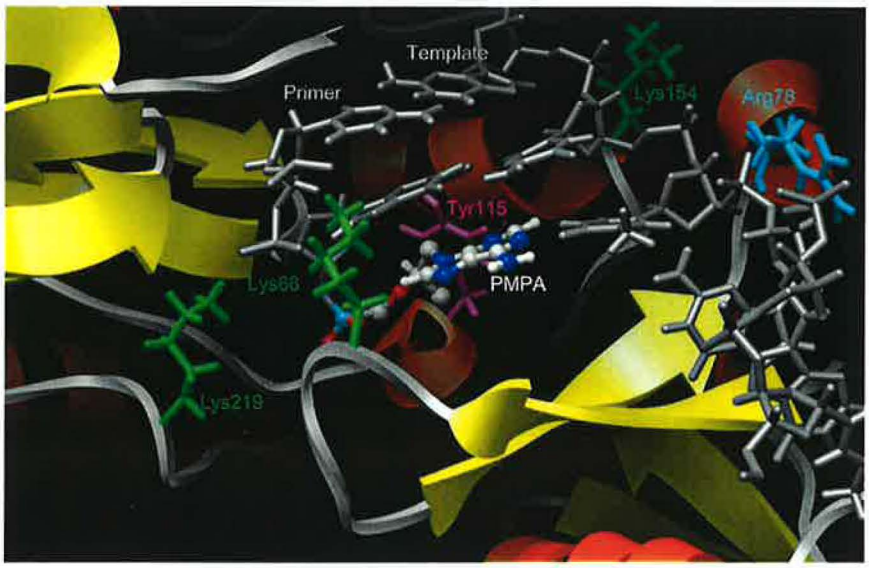


Figure 9.8: Tenofovir at the polymerase active site in the structure no.8.

Chapter 10

Water Solvent

Last but not least, 3D grid covering a simulation box (HIV-RT + DNA:DNA + solvent) was created by using PTRAJ module(Amber software package). Persistence of water molecules in the individual grid cells within a MD run was evaluated. There were grid cells with the highest water occupancy indicated by light blue color (see Figures 10.1-10.3).

It meant that the whole polymerase active site was surrounded by water molecules stabilized in defined positions. These water molecules were connected by a net of mutual hydrogen bonds stabilizing nucleotide/inhibitor binding into the polymerase active site.

Figure 10.1: HIV-RT + highest water occupancy

Figure 10.2: HIV-RT polymerase active site

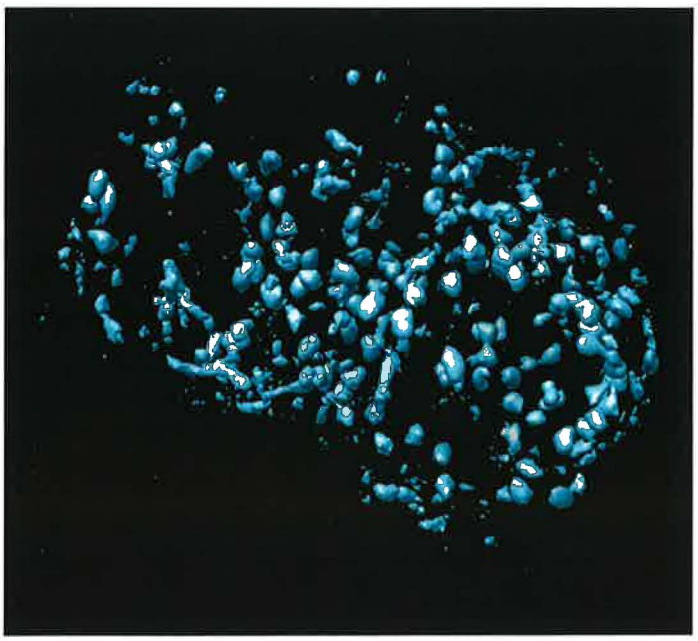


Figure 10.3: Highest water occupancy sites determined by MD simulation.

Chapter 11

Conclusions

The main objective of this work was to study the dynamics of the HIV-1 RT enzyme in complex with DNA:DNA and inhibitors (adefovir and tenofovir) by using MD simulations. We were interested in the subtle differences in binding of the nucleotides/inhibitors into the HIV-RT polymerase active site before and after they were incorporated into the primer DNA strand.

HIV-RT + DNA:DNA + nucleotide/inhibitor structures were solvated with water molecules. It led to simulate the systems consisting of about 130,000 atoms. These extremely time-consuming MD simulations were produced using large multiprocessor systems.

From analyzing of the MD trajectory files we found that the root mean square deviations(RMSD) values demonstrated the outright stability of the DNA:DNA and HIV-RT structures. The core site of their mutual contacts (DNA:DNA regions II. and III., HIV-RT connection domain) was found as the most rigid part of the complex. The terminal regions of p61 chain were clearly the most flexible parts of the enzyme structure. Detail analysis of atomic fluctuations showed, for example, that the α-helices were the most stable parts of the RT protein structure.

The structural consequences of DNA:DNA duplex binding to the HIV-RT enzyme were analyzed in detail. There were found the outright signs of rumpling by the enzyme - affection of puckering of deoxyribose moieties ($C3'$-endo conformations typical rather in the case of riboses), conformations of phosphodiester linkages (several of them were not found in usual -g-g conformations). On the other hand, the Watson-Crick hydrogen bonds seemed to be unaffected with the exception of the first and second base pairs

(BP1, BP2) situated in the polymerase active site. This was determined in the case of all structures with the rA residue in the polymerase active site (structure no. 4,6 I,6 II). It indicated that the aromatic side chain of Tyr115 successfully prevented incorporation of a ribonucleoside by interfering with rA 2'-OH hydroxyl group. It secured a synthesis of pure DNA strands by the HIV-RT enzyme. For the most of the helical parameters (the Inclination, Tip, Shift, Slide, Rise, Tilt, Roll and Twist) the systematic variations around the strands were found - again indicating adaptation of the double helical structure to fit properly into the enzyme substrate binding site.

It was found that only Adefovir diphosphate (PMEApp) underwent conformational transitions within a MD run (quantified using backbone torsion angles). The tenofovir diphosphate (PMPApp) carrying additional methyl group in the backbone was completely stable. The Tenofovir diphosphate (PMPApp) optimal binding into the polymerase active site could be one from reasons of its higher therapeutic potency.

Interestingly, Arg72 residue (sandwiching incoming nucleotide/inhibitor) was found to be potently anchored by hydrogen bonds to the triphosphate moiety of inhibitors rather than natural substrates.

The whole polymerase active site was surrounded by water molecules stabilized in defined positions. These water molecules were connected by a net of mutual hydrogen bonds stabilizing nucleotide/inhibitor binding into the polymerase active site.

Bibliography

[1] UNAIDS http://www.unaids.org/en/Regions_Countries/default.asp

[2] Laurent Berthiaume, Michel Tremblay, *Virus Life in Diagrams* 1998: CRC Press.

[3] UNAIDS http://www.unaids.org/en/HIV_data/

[4] C. Hoffmann, B. S. Kamps *HIV Medicine 2003* 2003: Flying Publisher.

[5] Nobel Foundation *The Nobel Prize in Physiology or Medicine 1975* http://nobelprize.org/nobel_prizes/medicine/laureates/1975/

[6] Kohlstaedt, *et al. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor* Science 1992: 256, 1783-1790.

[7] Brian G. Turner, Michael F. Summers *Structural Biology of HIV* J. Mol. Biol. 1999: 285, 1-32.

[8] Eddy Arnold *et al. Structure of HIV-1 RT-DNA complexes before and after incorporation of the anti-AIDS drug tenofovir* Nature Structural & Molecular Biology. 2004: 11, 469-474.

[9] OpenChemist.net *Overview of HIV Antiretroviral Compounds* http://openchemist.net/chemistry/show.php?id=hiv&story=comp001

[10] A. Holy, *et al. Antiretrovirus Activity of a Novel Class of Acyclic Pyrimidine Nucleoside Phosphonates* Antimicrob. Agents Chemother. 2002: 46, 2185-2193.

[11] FDA/Center for Drug Evaluation and Research *Microbiology review of VIREAD*™ http://www.fda.gov/cder/foi/label/2001/21356lbl.pdf

[12] FDA/Center for Drug Evaluation and Research *Microbiology review of HEPSERA*™ http://www.fda.gov/cder/foi/label/2002/21449lbl.pdf

[13] Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Chandra, S., Ingh, U., Weiner, P. & Kollman, P.A. AMBER (1991) University of California, San Franscico, CA.

[14] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. *Scalable molecular dynamics with NAMD* J. Comput. Chem. 2005: 26, 1781-1802.

[15] A.R. Leach *Molecular Modelling. Principles and Aplications* 2001: Pearson, Harlow.

[16] NAMD mannual http://www.ks.uiuc.edu/Research/namd/

[17] Berendsen, *et al. Molecular dynamics with coupling to an external bath* J. Chem. Phys. 1984: 81, 3684-3690.

[18] Frank Jensen *Introduction to Computational Chemistry* 2001: John Wiley&Sons.

[19] M. P. Allen and D. J. Tildesley *Computer Simulation of Liquids* 1991: Clarendon Press.

[20] Hans C. Andersen *Molecular dynamics simulations at constant pressure and/or temperature* J. Chem. Phys. 1980: 72, 2384-2393.

[21] Scott E. Feller, *et al. Constant pressure molecular dynamics simulation: The Langevin piston method* J. Chem. Phys. 1995: 103, 4613-4621.

[22] Glenn J. Martyna, Douglas J. Tobias and Michael L. Kiein *Constant pressure molecular dynamics algorithms* J. Chem. Phys. 1994: 101, 4177-4189.

[23] Daan Frenkel, Berend Smit *Understanding Molecular Simulation: From Algoirhtms to Applications* 2002: Academic Press.

[24] William L. Jorgensen, *et al. Comparison of simple potential functions for simulating liquid water* J. Chem. Phys. 1983: 79, 926-935.

[25] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. *UCSF Chimera - A Visualization System for Exploratory Research and Analysis* J. Comput. Chem. 2004: 25(13), 1605-1612.

[26] C. Altona and M. Sundaralingam *Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation* J. Chem. Soc. 1972: 94(23), 8205-8212.

[27] Wolfram Saenger *Principles of Nucleic Acid Structure* 1984: Springer-Verlag.

[28] Eveline Lescrinier, Matheus Froeyen and Piet Herdewijn *SURVEY AND SUMMARY: Difference in conformational diversity between nucleic acids with a six-membered 'sugar' unit and natural 'furanose' nucleic acids* Nucleic Acids Research. 2003: 31(12), 2975-2989.

[29] Stefan G.Sarafianos, *et al. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA* EMBO J. 2001: 20, 1449-1461.

[30] CURVES 5.1 http://www.ibpc.fr/UPR9080/Curindex.html

[31] Helical parameters http://www.imb-jena.de/Piet/help/helixparameter.html

A MD configuration parameters

The following codes are the NAMD parameters we used in doing the MD simulation of solvated RT (system no.9 in the table 4.1). These configurations were also applied to the other simulation systems.

```
amber on
parmfile rt_wat.parm
coordinates equil_rt_wat.coor #coordinates of equilibrated system
exclude scaled1-4
1-4scaling 0.833333
dielectric 1.0
velocities equil_rt_wat.vel
extendedsystem equil_rt_wat.xsc

#Long range interaction
switching off
cutoff 9.0
margin 0.0
stepspercycle 20

#SHAKE
rigidBonds all
rigidTolerance 0.00001
rigidIterations 250

#Ewald Method
PME on
PMETolerance 0.000001
```

```
PMEGridSizeX 145
PMEGridSizeY 92
PMEGridSizeZ 120

#Periodic Boundary Conditions
cellBasisVector1 141.9 0.0 0.0
cellBasisVector2 0.0 90.9 0.0
cellBasisVector3 0.0 0.0 116.6

#Temperature Control (applied Langevin Dynamics)
langevin on
langevindamping 1
langevintemp 310

#Pressure Control (Langevin Piston and Nose-Hoover)
langevinpiston on
langevinpistontarget 1.01325
langevinpistonperiod 200
langevinpistondecay 500
langevinpistontemp 310
usegrouppressure yes
useflexiblecell no
useconstantratio no

#Output
outputenergies 10
outputtiming 100
binaryoutput no
outputname run1_rt_wat
restartname run1_rt_wat_restr
restartfreq 100
binaryrestart yes
DCDfile run1_rt_wat.dcd
dcdfreq 100

numsteps 2500000
```

B Simple MD Program

```
C=============================================================
C MOLECULAR DYNAMICS
C THE LENNARD-JONES POTENTIAL IS TRUNCATED AT RCOFF AND NOT
C SMOOTHLY CONTINUED TO ZERO. INITIALLY THE NPART PARTICLES
C ARE PLACED ON AN FCC LATICE. THE VELOCITIES ARE DRAWN FROM
C A BOLTZMANN DISTRIBUTION WITH TEMPERATURE TREF.
C
C INPUT PARAMETERS ARE AS FOLLOWS
C
C NPART  NUMBER OF PARTICLES
C SIDE   SIDE LENGTH OF THE CUBICAL BOX IN SIGMA UNITS
C TREF   REDUCED TEMPERATURE
C ROCFF  CUTOFF OF THE POTENTIAL IN SIGMA UNITS
C H      BASIC TIME STEP
C IREP   VELOCITIES SCALING EVERY IREP'TH TIME STEP
C ISTOP  STOP OF SCALING OF THE VELOCITIES AT ISTOP
C TIMEMX NUMBER OF INTEGRATION STEPS
C ISEED  SEED FOR THE RANDOM NUMBER GENERATOR
C=============================================================
      REAL X(1:786),VH(1:786),F(1:786)
      REAL FY(1:256), FZ(1:256), Y(1:256) ,Z(1:256)
      REAL H,HSQ,HSQ2,TREF
      REAL  KX,KY,KZ
      INTEGER  CLOCK,TIMEMX, ld1
      EQUIVALENCE  (FY,F(257)), (FZ,F(513)), (Y,X(257)), (Z,X(513))

C DEFINITION OF THE SIMULATION PARAMETERS
```

```
NPART = 256
DEN = 0.83134
SIDE = 6.75284
TREF = 0.722
RCOFF = 2.5
H = 0.064
IREP = 50
ISTOP = 500
TIMEMX = 3
ISEED = 4711

c OUTPUT File
ld1=11
OPEN(unit = ld1, file = 'md.dat')

C SET THE OTHER PARAMETERS
       WRITE(6,*) 'MOLECULAR DYNAMICS SIMULATION PROGRAM'
       WRITE(6,*) '----------------------------------------'
       WRITE(6,*)
       WRITE(6,*) 'NUMBER OF PARTICLES IS ', NPART
       WRITE(6,*) 'SIDE LENGTH OF THE BOX IS ', SIDE
       WRITE(6,*) 'CUT OFF IS ', RCOFF
       WRITE(6,*) 'REDUCED TEMPERATURE IS ', TREF
       WRITE(6,*) 'BASIC TIME STEP IS ', H

       A = SIDE / 4.0
       SIDEH = SIDE * 0.5
       HSQ = H * H
       HSQ2 = HSQ * 0.5
       NPARTM = NPART - 1
       RCOFFS = RCOFF * RCOFF
       TSCALE = 16.0 / (1.0 * NPART - 1.0)
       VAVER = 1.13 * SQRT ( TREF / 24.0 )

      N3 = 3 * NPART
      IOF1 = NPART
```

```
      IOF2 = 2 * NPART
      CALL SRAND ( ISEED )

C THIS PART OF THE PROGRAM PREPARES THE INITIAL CONFIGURATION
C-------------------------------------------------------------

C SET UP FCC LATTICE FOR THE ATOMS INSIDE THE BOX
IJK = 0
DO 10 LG = 0,1
DO 10 I = 0,3
     DO 10 J = 0,3
    DO 10 K = 0,3
      IJK = IJK + 1
      X (IJK) = I * A + LG * A * 0.5
      Y(IJK) = J * A + LG * A * 0.5
      Z(IJK) = K * A

WRITE(LD1,*) IJK,X(IJK),Y(IJK),Z(IJK)

10     CONTINUE
       DO 15 LG = 1,2
       DO 15 I = 0,3
         DO 15 J =0,3
           DO 15 K = 0,3
       IJK = IJK +1
  X(IJK) = I * A + (2 - LG) * A * 0.5
       Y(IJK) = J * A + ( LG - 1 ) * A * 0.5
       Z(IJK) = K * A + A * 0.5
   WRITE(LD1,*) IJK,X(IJK),Y(IJK),Z(IJK)
 15      CONTINUE

C ASSIGN VELOCITIES DISTRIBUTED NORMALLY
       CALL MXWELL ( VH, N3 , H, TREF )
       DO 50 I = 1, N3
          F(I) = 0.0
 50    CONTINUE
```

```
DO 200 CLOCK = 1, TIMEMX

  DO 210 I = 1,N3
    X(I) = X(I) + VH(I) + F(I)
 210   CONTINUE

  DO 215 I = I,N3
    IF( X(I) .LT. 0 ) X(I) = X(I) + SIDE
    IF( X(I) .GT. SIDE ) X(I) = X(I) - SIDE
 215   CONTINUE

DO 220 I = 1,N3
  VH(I) = VH(I) + F(I)
 220 CONTINUE


C THIS PART COMPUTERS THE FORCES ON THE PARTICLES
VIR = 0.0
EPOT = 0.0
DO 225 I = 1, N3
  F(I) = 0.0
 225 CONTINUE

DO 270 I=1,NPART
  XI = X(I)
  YI = Y(I)
  ZI = Z(I)
  DO 270 J=I+1,NPART
    XX = XI - X(J)
YY = YI - Y(J)
ZZ = ZI - Z(J)
IF ( XX .LT. -SIDEH) XX = XX + SIDE
IF ( XX .GT. SIDEH) XX = XX -SIDE

IF ( YY .LT. -SIDEH) YY = YY + SIDE
```

```
IF ( YY .GT. SIDEH) YY = YY -SIDE

IF ( ZZ .LT. -SIDEH) ZZ = ZZ + SIDE
IF ( ZZ .GT. SIDEH) ZZ = ZZ -SIDE
RD = XX * XX + YY * YY + ZZ * ZZ
IF ( RD .GT. RCOFFS ) GOTO 270

WRITE(LD1,*) I,XX,YY,ZZ

EPOT = EPOT + RD ** ( -6.0) - RD ** (-3.0)
R148 = RD ** (-7.0) -0.5 * RD ** (-4.0)
VIR = VIR - RD * R148
KX = XX * R148
F(I) = F(I) + KX
F(J) = F(J) - KX
KY = YY * R148
F(I) = F(I) + KY
F(J) = F(J) - KY
KZ = ZZ * R148
F(I) = F(I) + KZ
F(J) = F(J) - KZ
 270 CONTINUE
DO 275 I = 1, N3
  F(I) = F(I) * HSQ2
 275 CONTINUE

DO 300 I = 1,N3
  VH(I) = VH(I) + F(I)
 300 CONTINUE

EKIN = 0.0
DO 305 I = 1,N3
  EKIN = EKIN + VH(I) * VH(I)
 305 CONTINUE
EKIN = EKIN / HSQ
```

```
  RP = ( COUNT / 256.0 ) * 100.0
  WRITE(6,6000) CLOCK, EK, EPOT, TEMP, PRES, VEL, RP
END IF

 200 CONTINUE
 6000 FORMAT ( 1I6,7F15.6)
CLOSE(LD1)
STOP
END
C -------------------------------------------------------------
SUBROUTINE MXWELL( VH,N3,H,TREF)
REAL VH(1:N3)
REAL U1,U2,V1,V2,S,R

NPART = N3 / 3
IOF1 = NPART
IOF2 = 2 * NPART
TSCALE = 16.0 / (1.0 * NPART - 1.0)

DO 10 I = 1,N3,2
 1   U1 = RAND()
  U2 = RAND()

  V1 = 2.0 * U1 - 1.0
  V2 = 2.0 * U2 - 1.0
  S = V1 * V1 + V2 * V2

  IF( S .GE. 1.0) GOTO 1
    R = -2.0 * ALOG(S) / S
VH(I) = V1 * SQRT(R)
VH(I+1) = V2 * SQRT(R)
 10 CONTINUE

  EKIN = 0.0
  SP = 0.0
DO 20 I = 1,NPART
```

```
  SP = SP + VH(I)
 20 CONTINUE
      SP = SP / NPART
DO 21 I = 1, NPART
  VH(I) = VH(I) - SP
  EKIN = EKIN + VH(I) * VH(I)
 21 CONTINUE

WRITE(6,*) 'TOTAL LINEAR MOMENTUM IN X DIRECTION IS ',SP
SP = 0.0
DO 22 I = IOF1 + 1, IOF2
  SP = SP + VH(I)
 22 CONTINUE
SP = SP / NPART

DO 23 I = IOF1+1,IOF2
  VH(I) =VH(I) -SP
  EKIN = EKIN + VH(I) * VH(I)
 23 CONTINUE
WRITE(6,*)'TOTAL LINEAR MOMENTUM IN Y DIRECTION IS ',SP
SP = 0.0

DO 24 I = IOF2 + 1 , N3
  SP = SP + VH(I)
 24 CONTINUE
SP = SP /NPART
DO 25 I = IOF2 + 1,N3
  VH(I) = VH(I) -SP
  EKIN = EKIN + VH(I) * VH(I)
 25 CONTINUE

WRITE(6,*)'TOTAL LINEAR MOMENTUM IN Z DIRECTION IS ',SP
WRITE(6,*)'VELOCITY ADJUSTMENT'

TS = TSCALE * EKIN
WRITE(6,*)'TEMPERATURE BEFORE SCALING IS ',TS
```

```
SC = TREF / TS
SC = SQRT(SC)

WRITE(6,*)'SCALE FACTOR IS ', SC
SC = SC * H
DO 30 I =1,N3
  VH(I) = VH(I) * SC
 30 CONTINUE
END
```

Apendix C:

Root Mean Square Deviations and Atomic Fluctuations

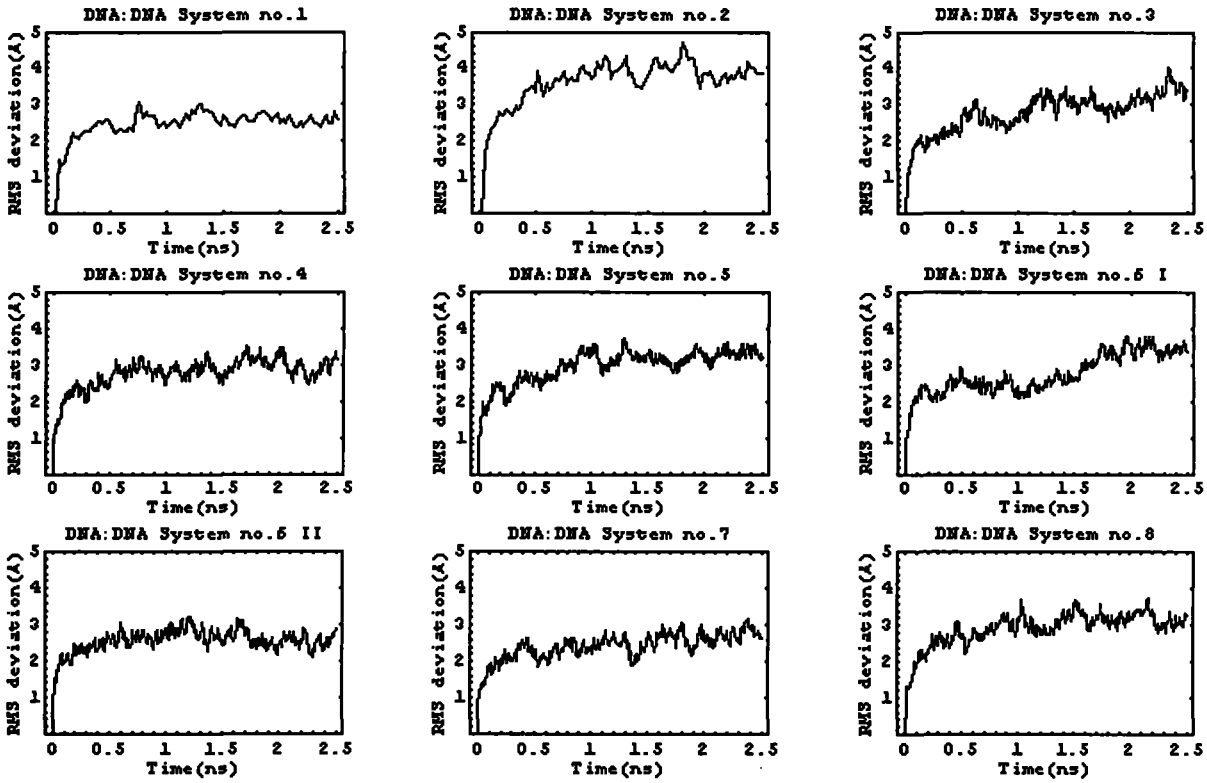


Figure C.1.: The root mean square deviation (RMSD) of the DNA:DNA chains in each simulated system as a function of time.

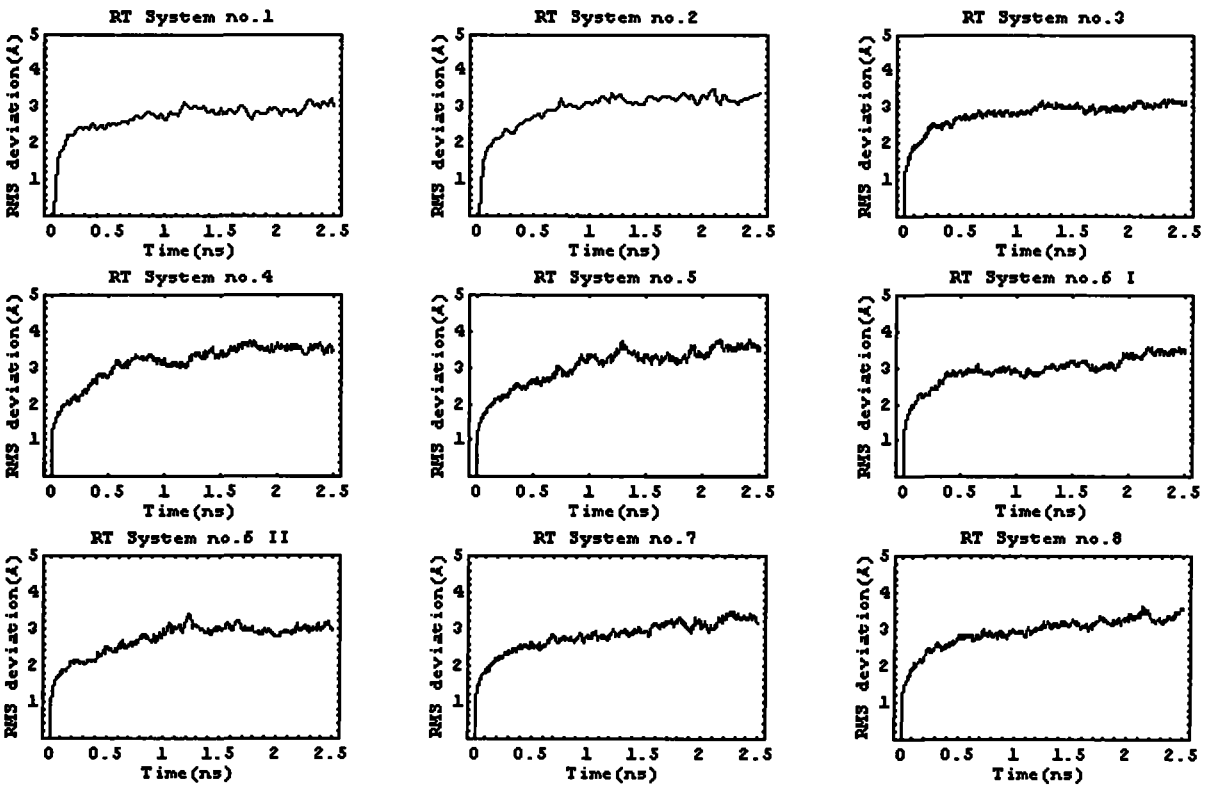


Figure C.2.: The root mean square deviation (RMSD) of the HIV-1 RT structure during the MD simulation.

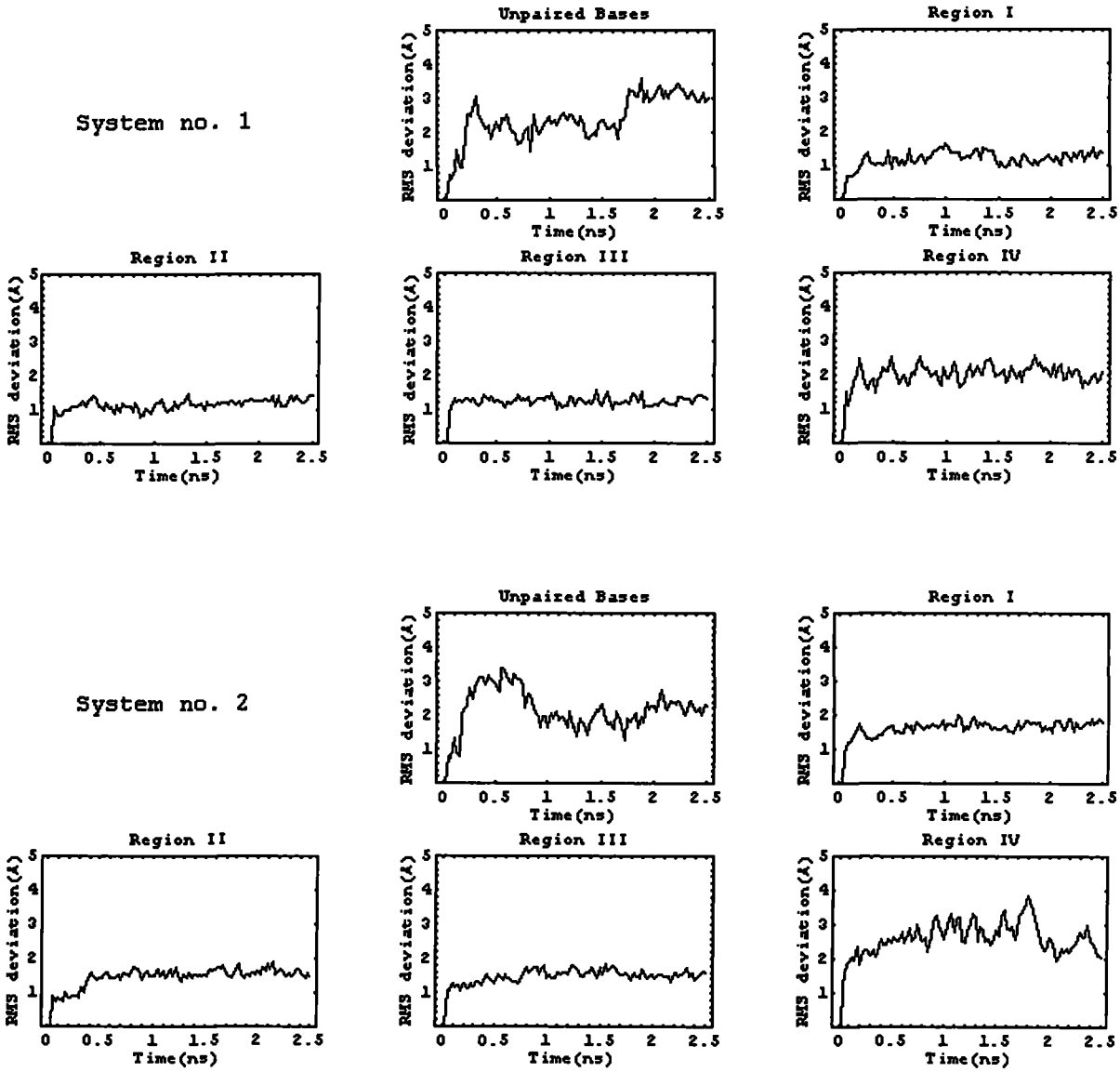


Figure C.3.: The root mean square deviation (RMSD) of each part of the DNA:DNA chains in the simulation system No.1-2. - HIV-1 RT and DNA:DNA complexed with the inhibitors PMEApp and PMPApp before their incorporation into the primer DNA chain.

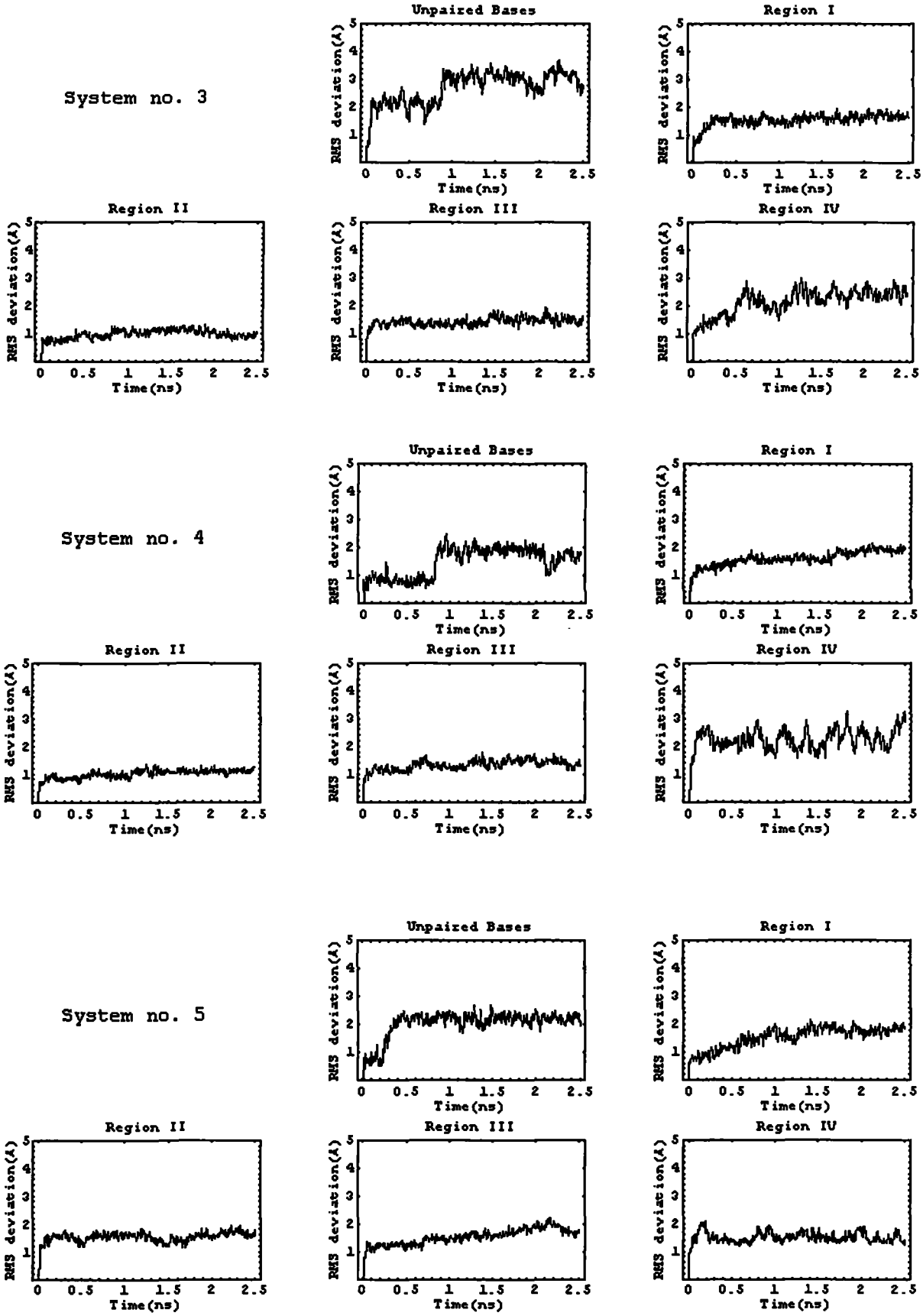


Figure C.4.: The root mean square deviation (RMSD) of each part of the DNA:DNA chains in the simulation systems No.3-5. – HIV-1 RT and DNA:DNA with the 3' terminal adenosine/deoxyadenosine at the primer DNA chain whilst the system no. 5 contains dATP near the 3' terminal of the primer DNA chain.

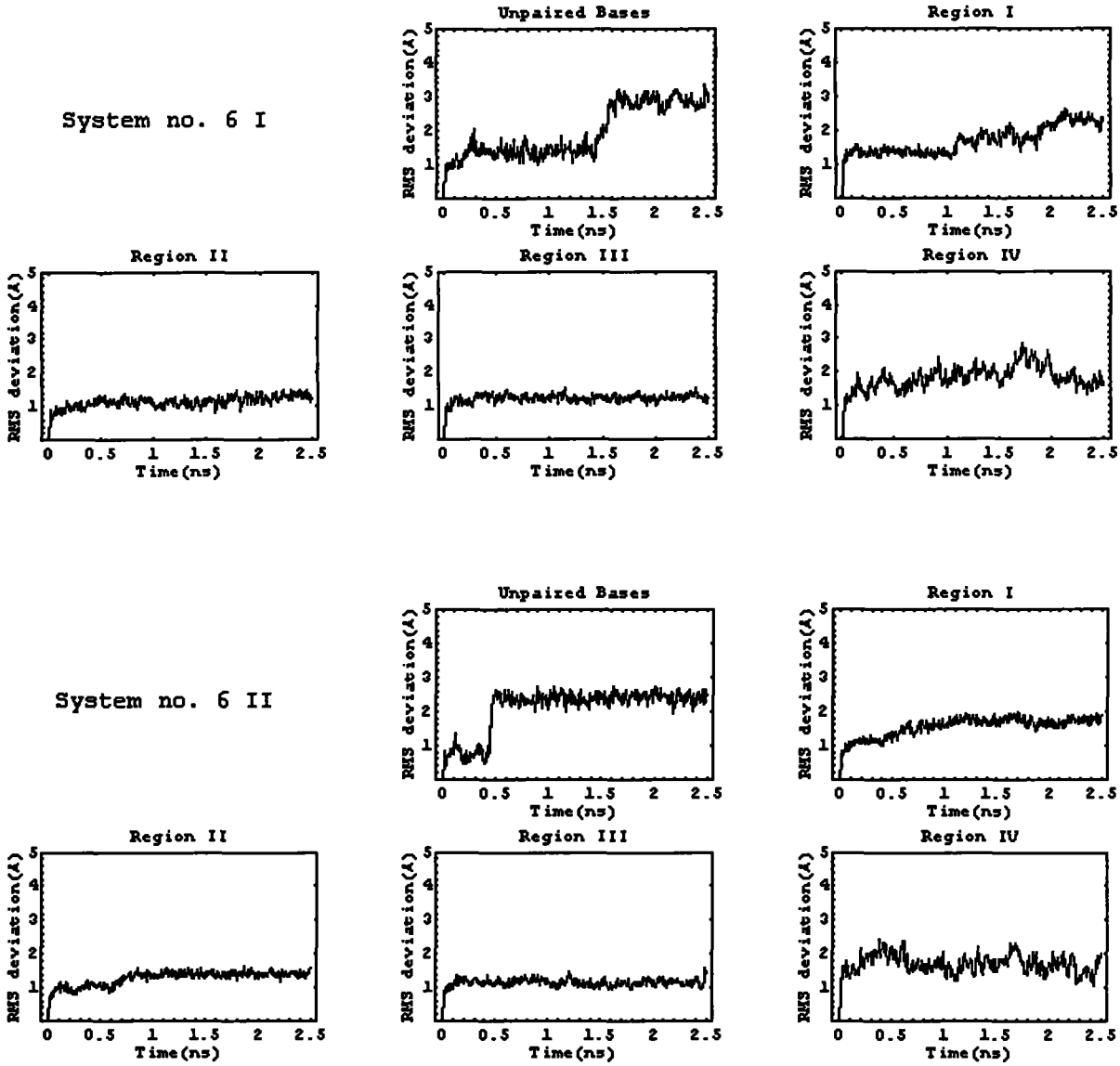


Figure C.5.: The root mean square deviation (RMSD) of each part of the DNA:DNA chains in the simulation system no.6 I and II . – HIV-1 RT, DNA:DNA and ATP near the 3' terminal of the primer DNA chain.

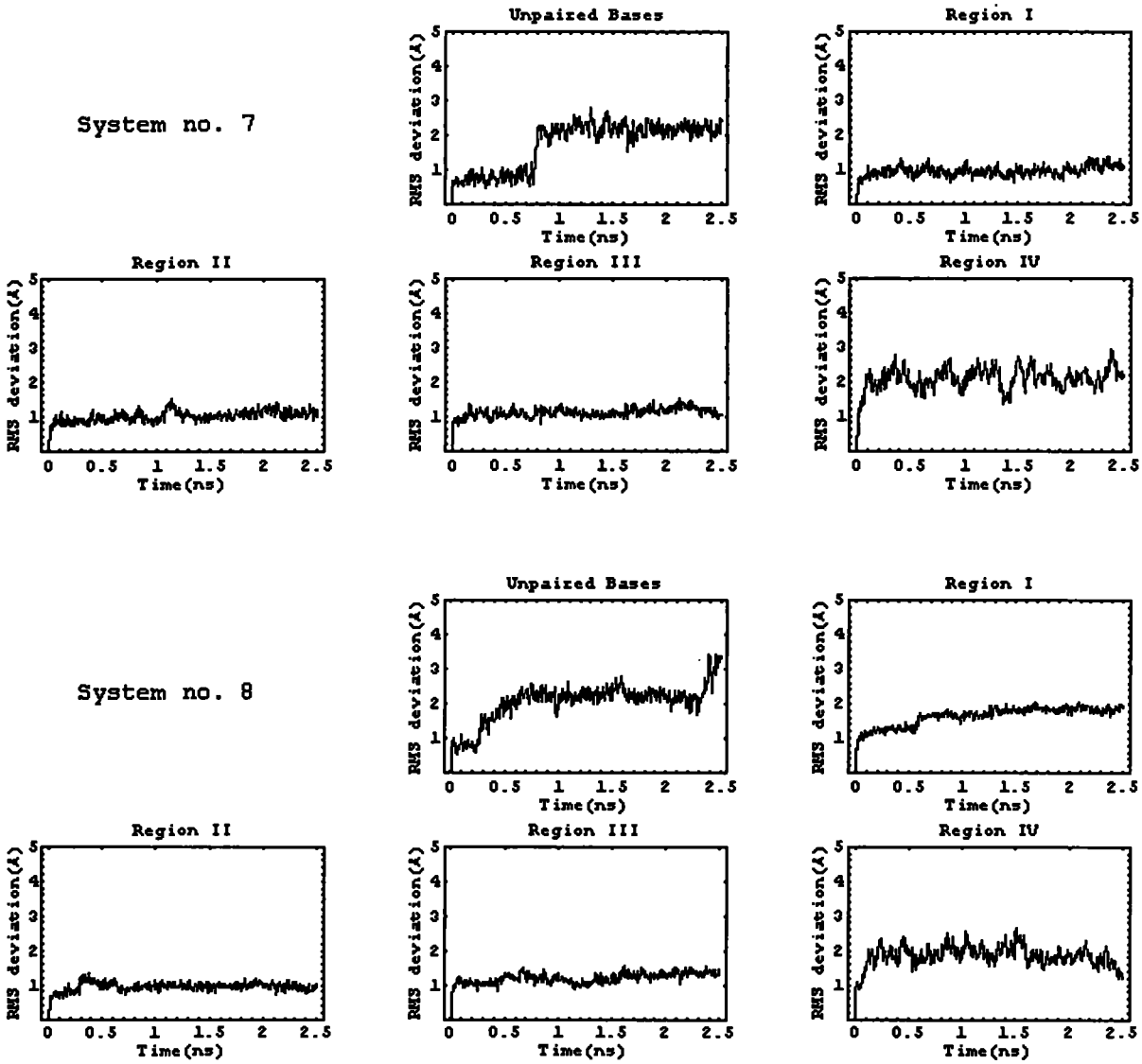


Figure C.6.: The root mean square deviation (RMSD) of each part of the DNA:DNA chains in the simulation system no.7-8 during the simulation. The systems no. 7 and 8 were the systems of HIV-1 RT and DNA:DNA complexed with the inhibitors (PMEA and PMPA, respectively) after their incorporation into the primer DNA chain.

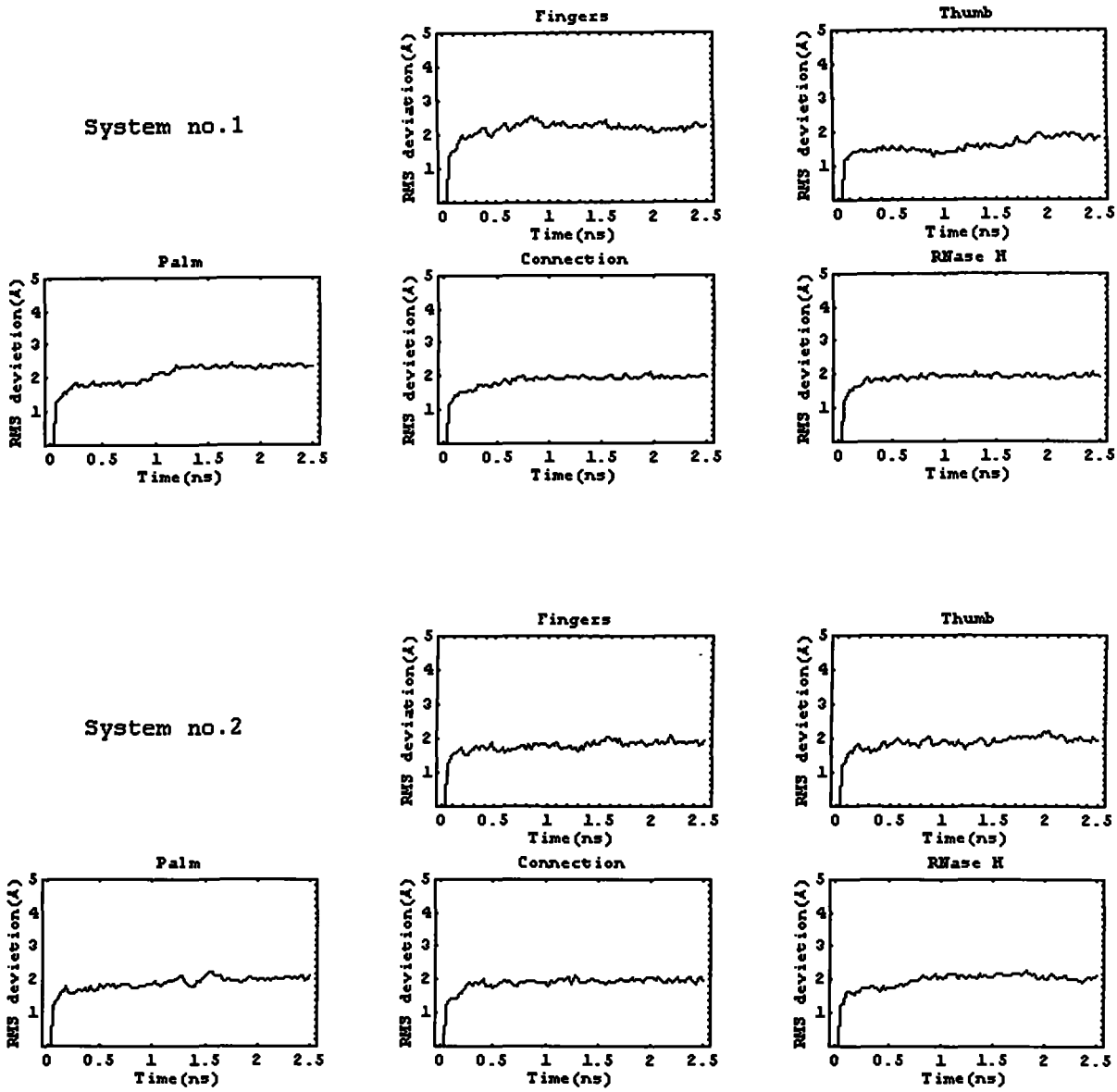


Figure C.7.: The root mean square deviation (RMSD) of each part of HIV-1 RT in the simulation system No.1-2. - HIV-1 RT and DNA:DNA complexed with the inhibitors PMEApp and PMPApp before their incorporation into the primer DNA chain.

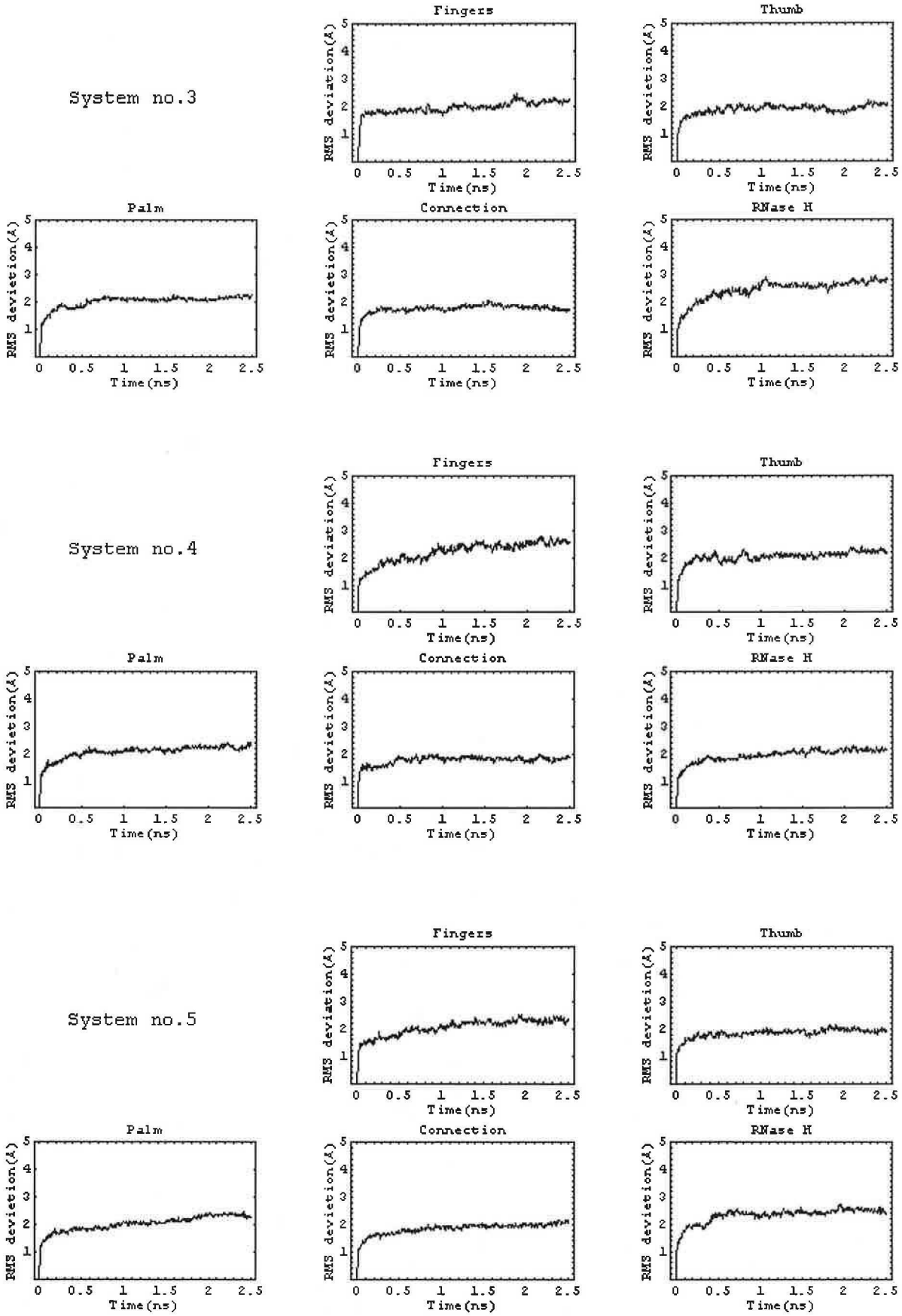


Figure C.8.: The root mean square deviation (RMSD) of each part of HIV-1 RT in the simulation systems No.3-5. – HIV-1 RT and DNA:DNA with the 3' terminal adenosine/deoxyadenosine at the primer DNA chain whilst the system no. 5 contains dATP near the 3' terminal of the primer DNA chain.

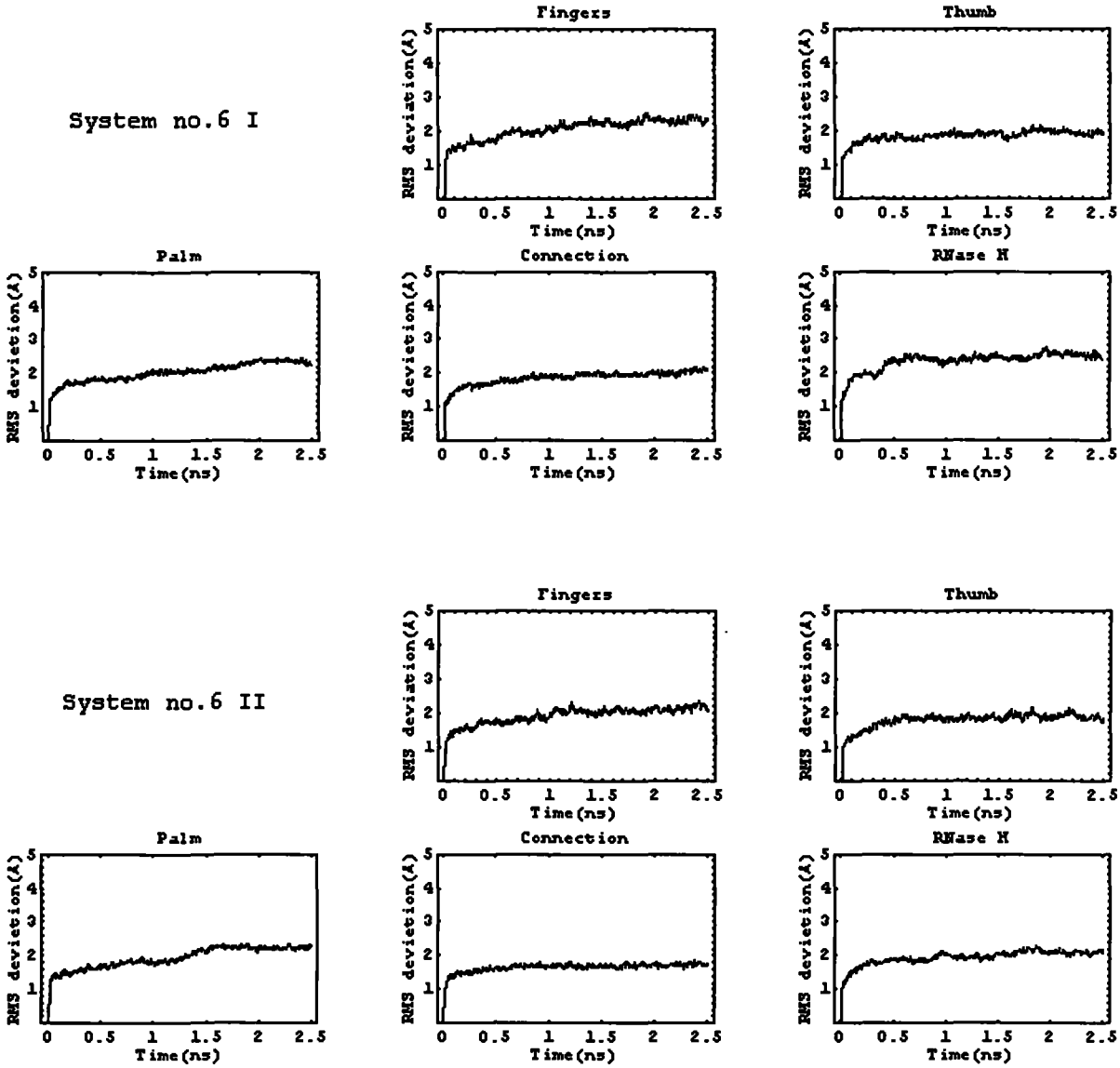


Figure C.9.: The root mean square deviation (RMSD) of each part of HIV-1 RT in the simulation system no.6 I and II . – HIV-1 RT, DNA:DNA and ATP near the 3' terminal of the primer DNA chain.

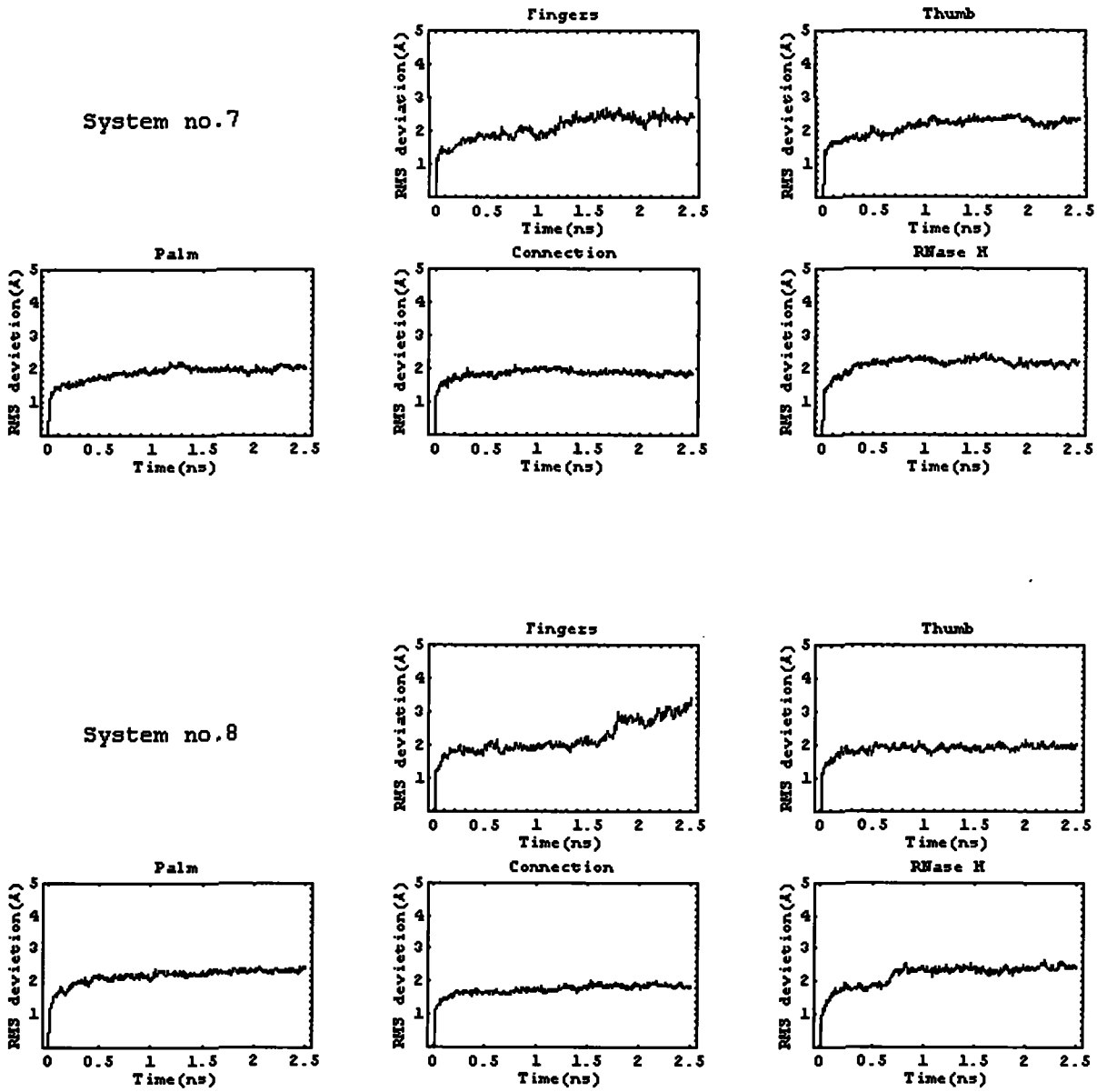


Figure C.10.: The root mean square deviation (RMSD) of HIV-1 RT in the simulation system no.7-8 during the simulation. The systems no. 7 and 8 were the systems of HIV-1 RT and DNA:DNA complexed with the inhibitors (PMEA and PMPA, respectively) after their incorporation into the primer DNA chain.

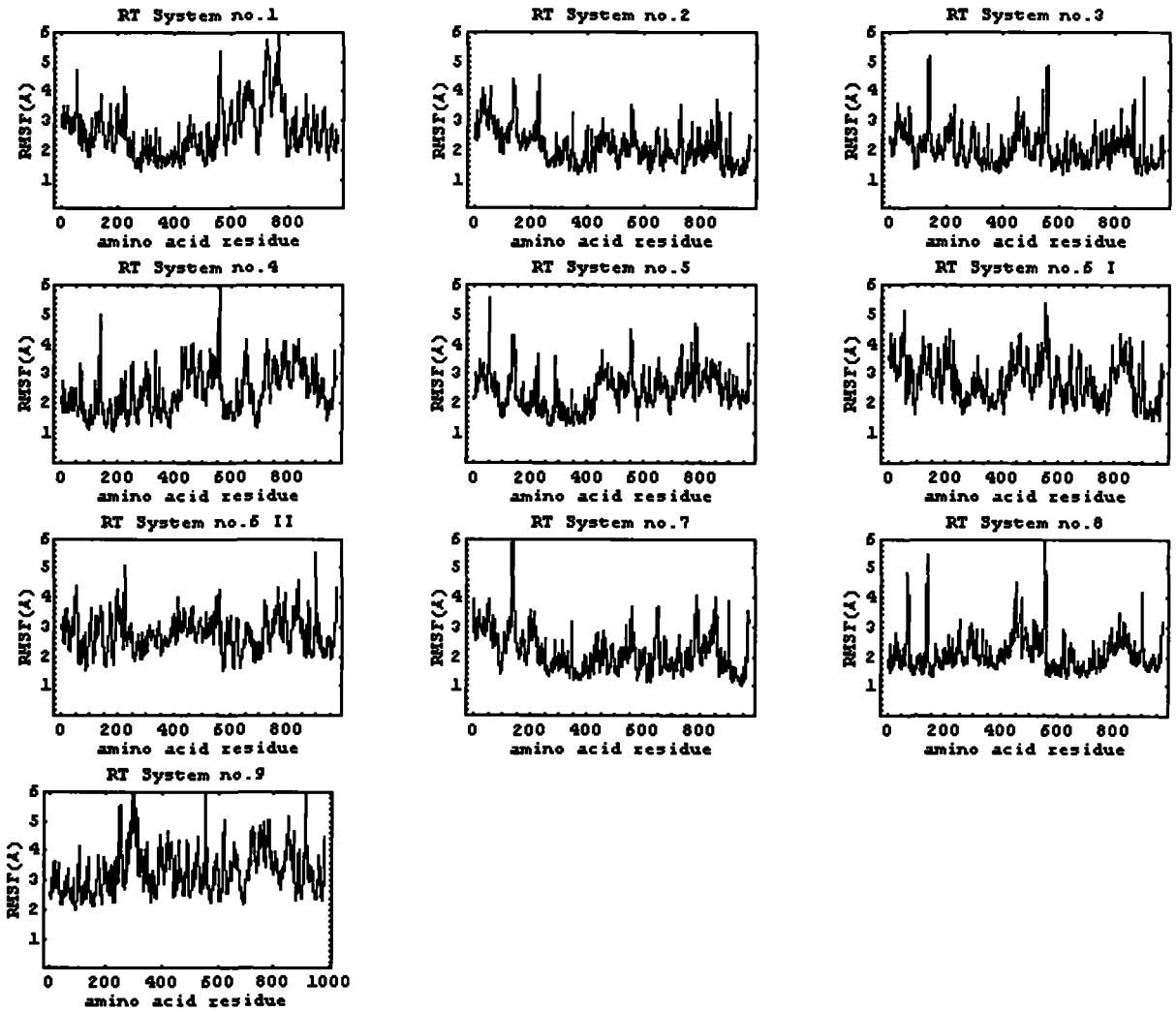
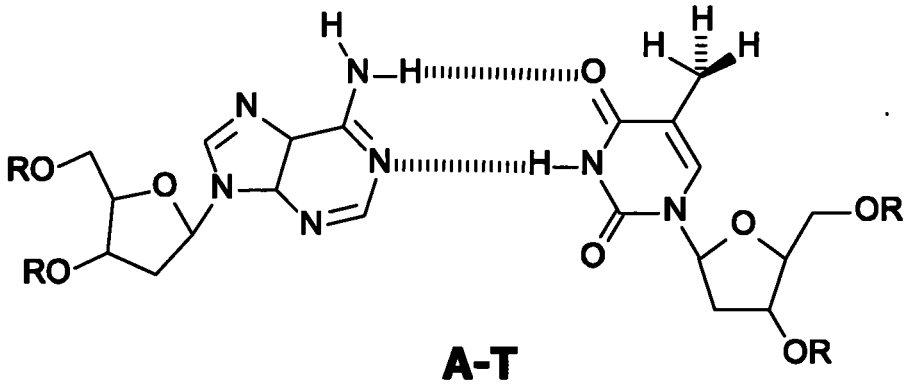


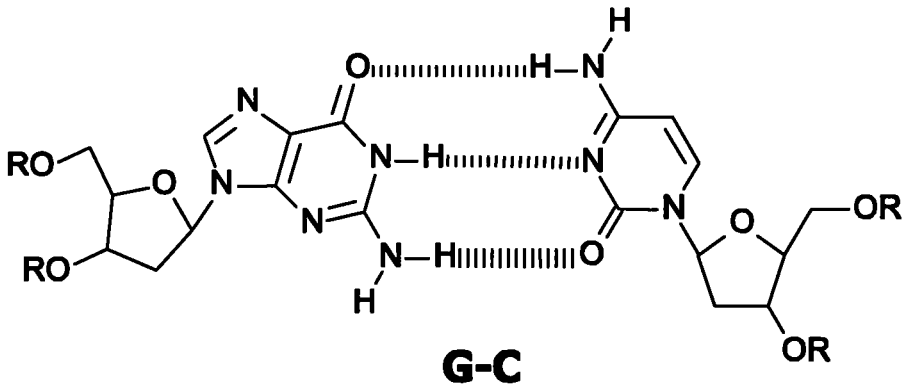
Figure C.11.: The root mean square fluctuation (RMSF) of HIV-1 RT amino acid residues from each simulation system. *Y-axis:* RMSF *X-axis:* the residue number of HIV-1 RT amino acids.

Apendix D:

Watson-Crick Hydrogen Bond Connection



A-T



G-C

System no.1: HIV-1 RT + DNA:DNA + PMEApp (before incorporation into the primer DNA)

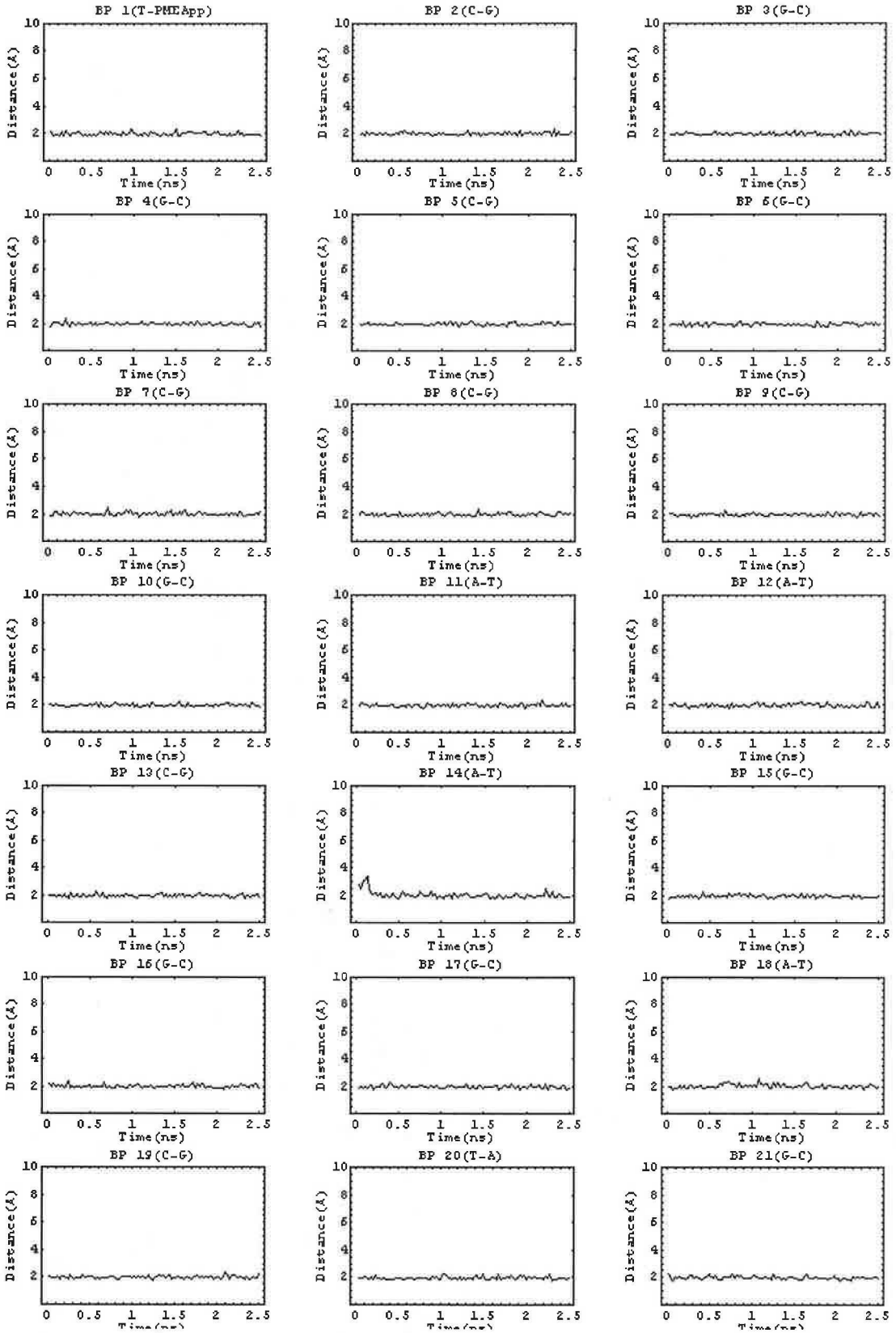


Figure D.1.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and Tenofovir diphosphate **(PMEApp)**. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

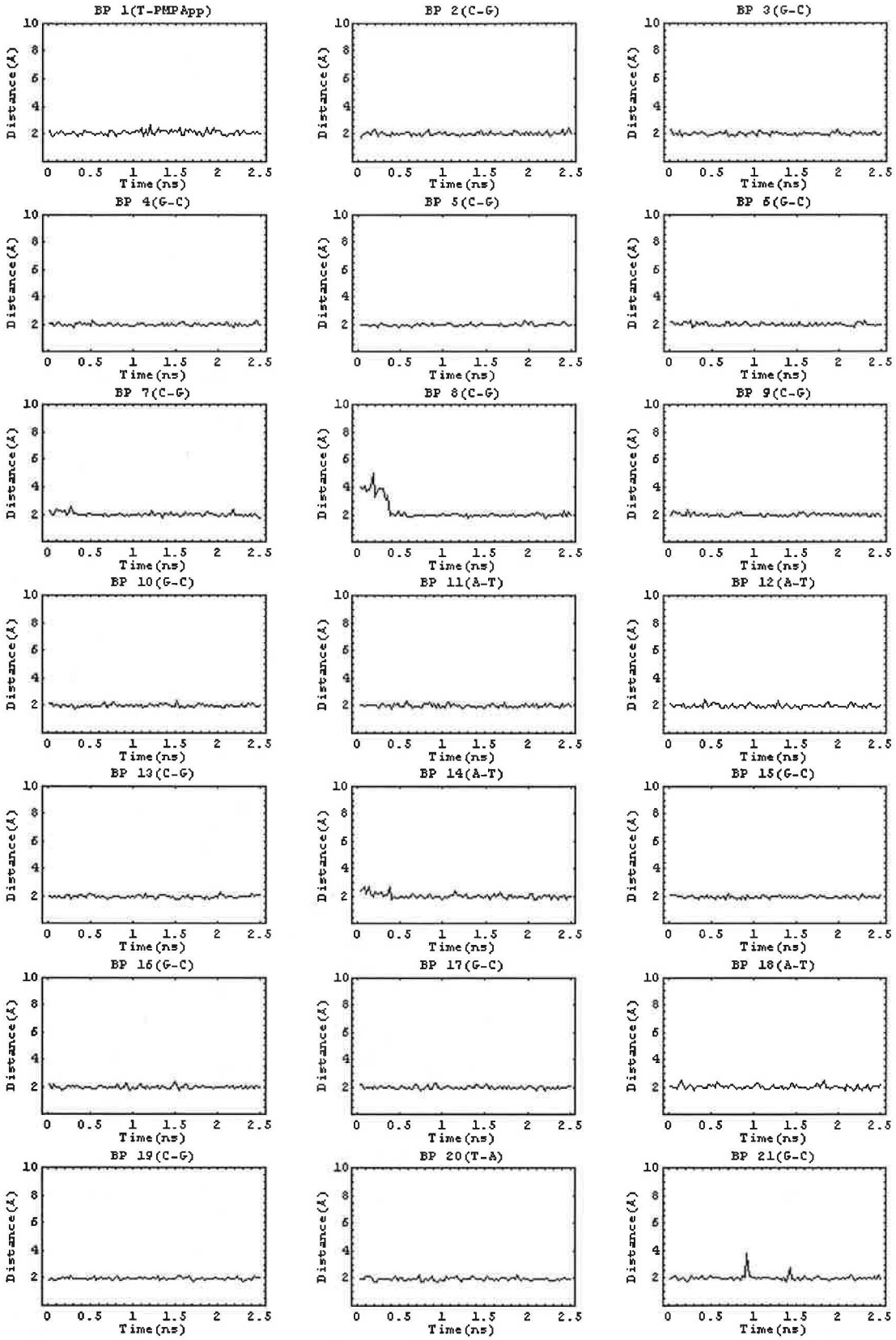


Figure D.2.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and Tenofovir diphosphate **(PMPApp)**. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

System no.3: HIV-1 RT + DNA:DNA + 3' Termial Deoxyadenosine

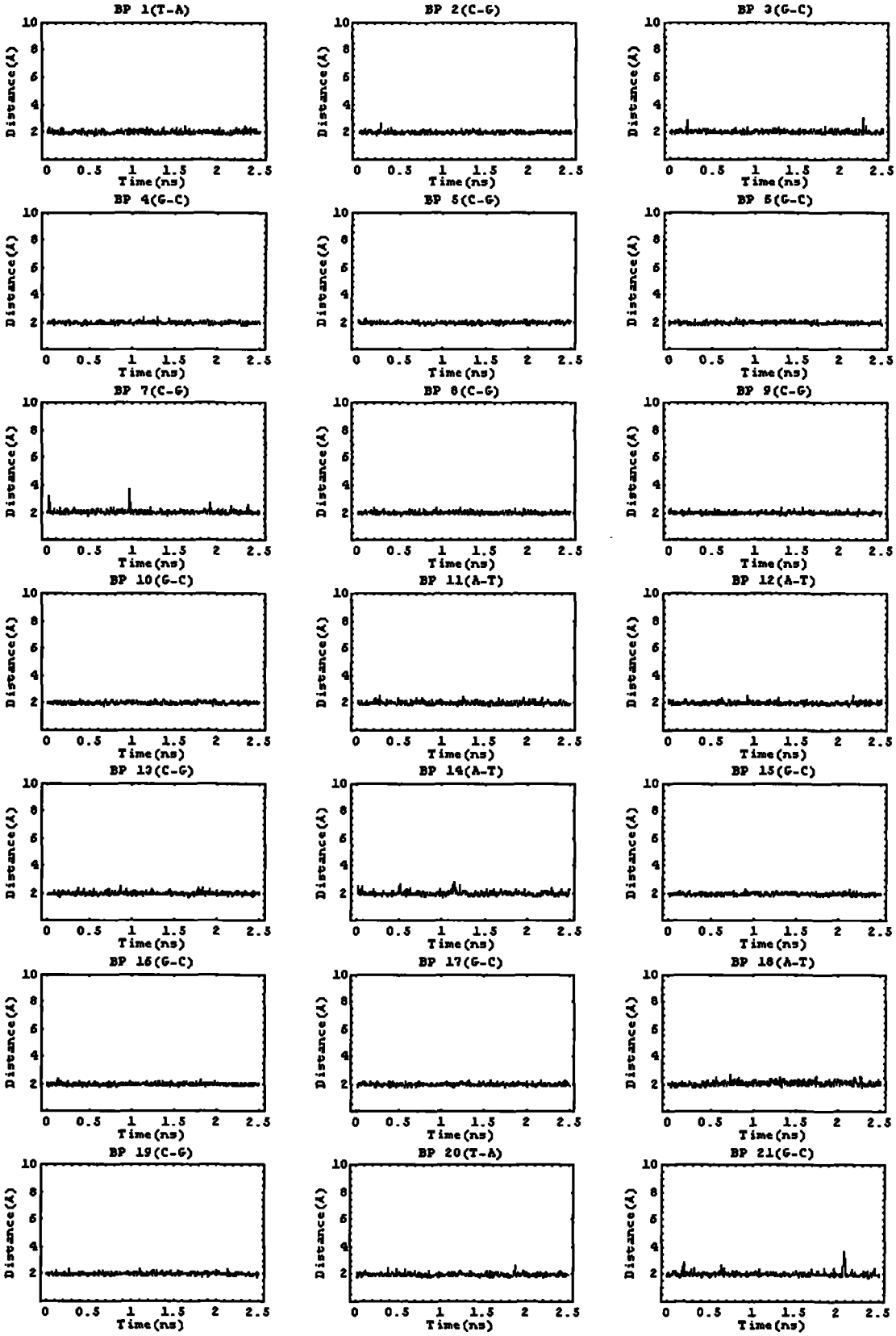


Figure D.3.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and **3'-Teminal Deoxyadenosine**. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

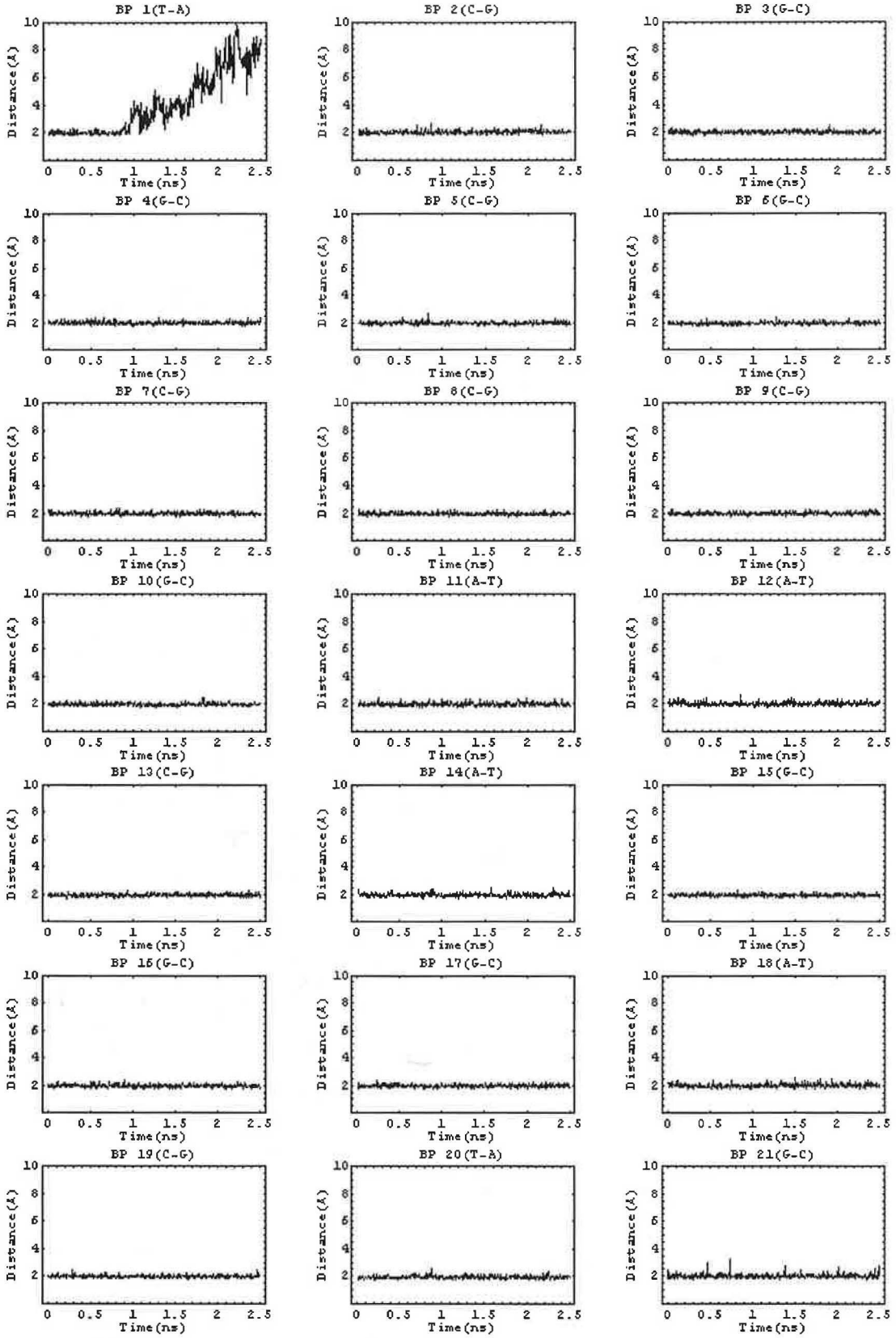


Figure D.4.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and the **3'-Teminal Adenosine**. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

System no.5: HIV-1 RT + DNA:DNA + dATP

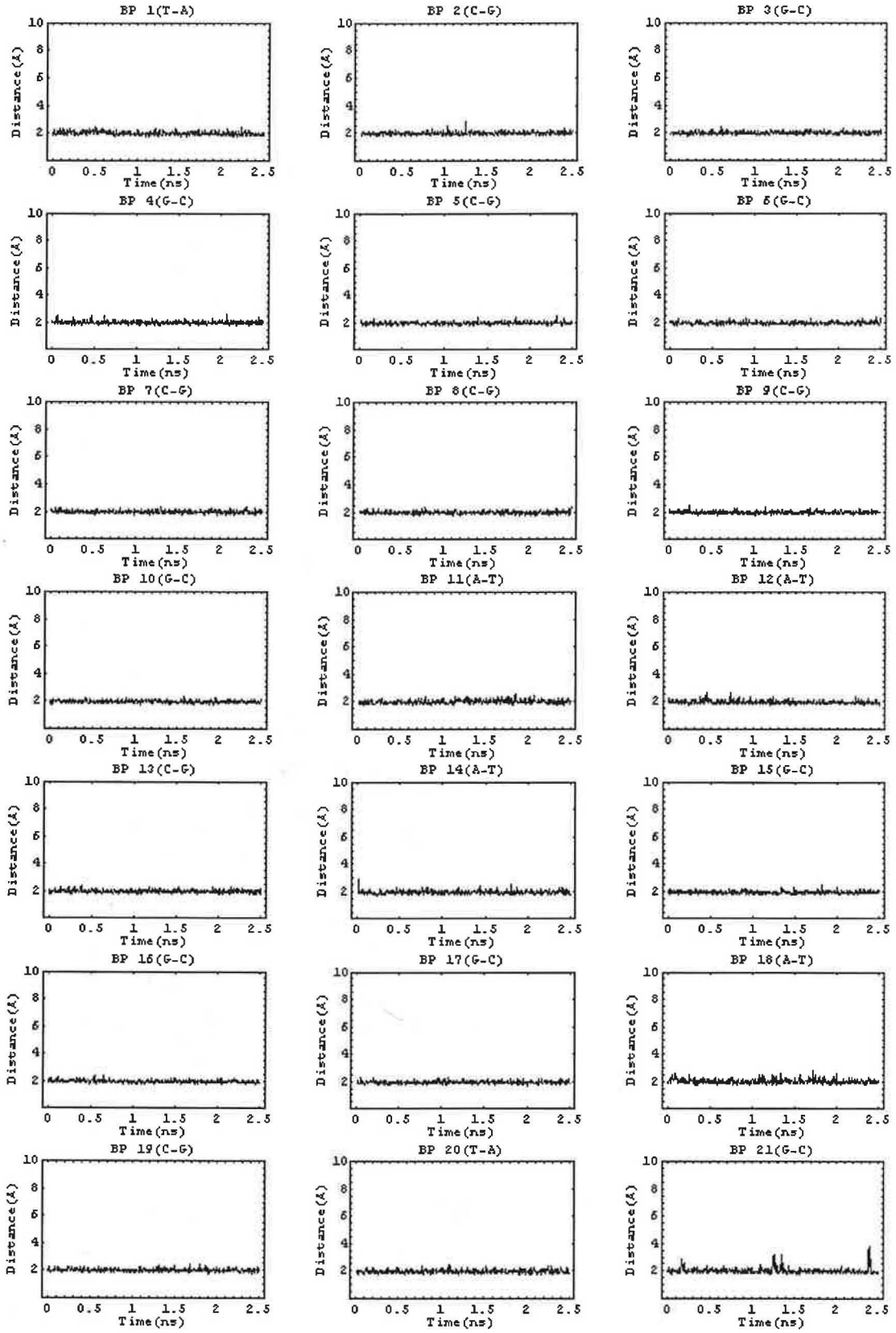


Figure D.5.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and Deoxyadenosine triphosphate **(dATP)**. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

System no.6 I: HIV-1 RT + DNA:DNA + ATP

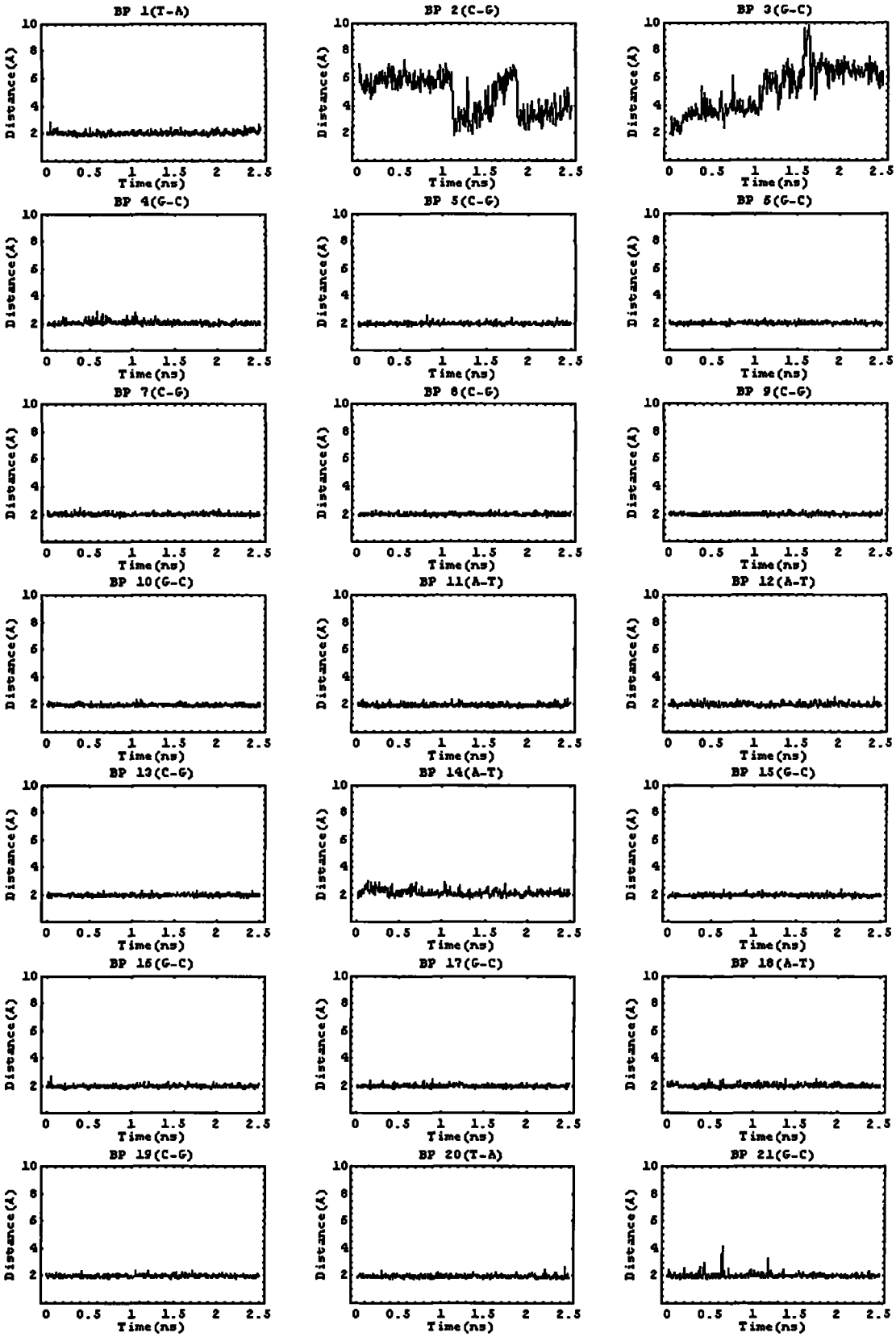


Figure D.6.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and Adenosine triphosphate (**ATP**). In the title of each graphs above, the first letter indicates the ucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

System no.6 II: HIV-1 RT + DNA:DNA + ATP

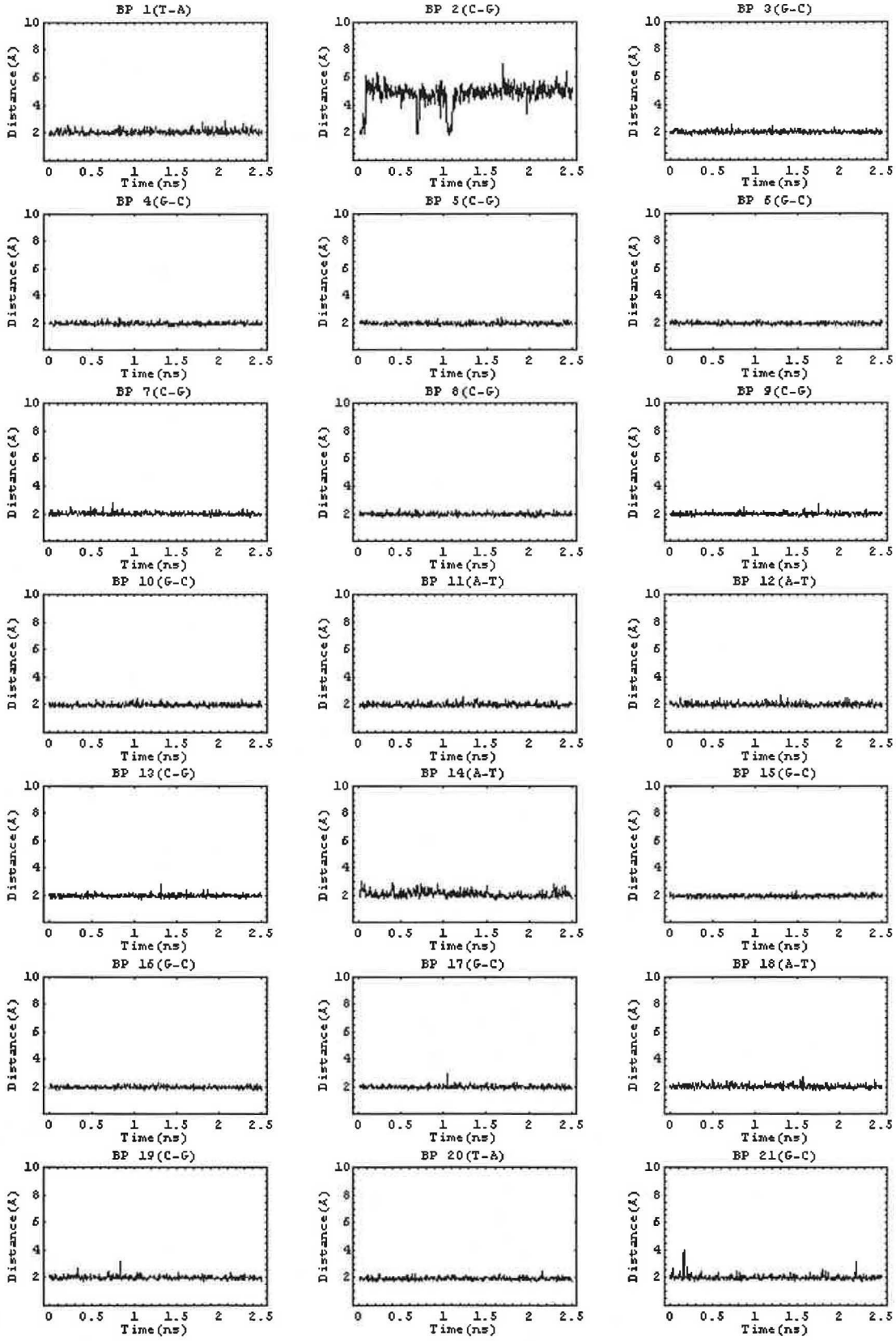


Figure D.7.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and Adenosine triphosphate **(ATP)**. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

System no.7: HIV-1 RT + DNA:DNA + PMEA (after incorporation into the primer DNA)

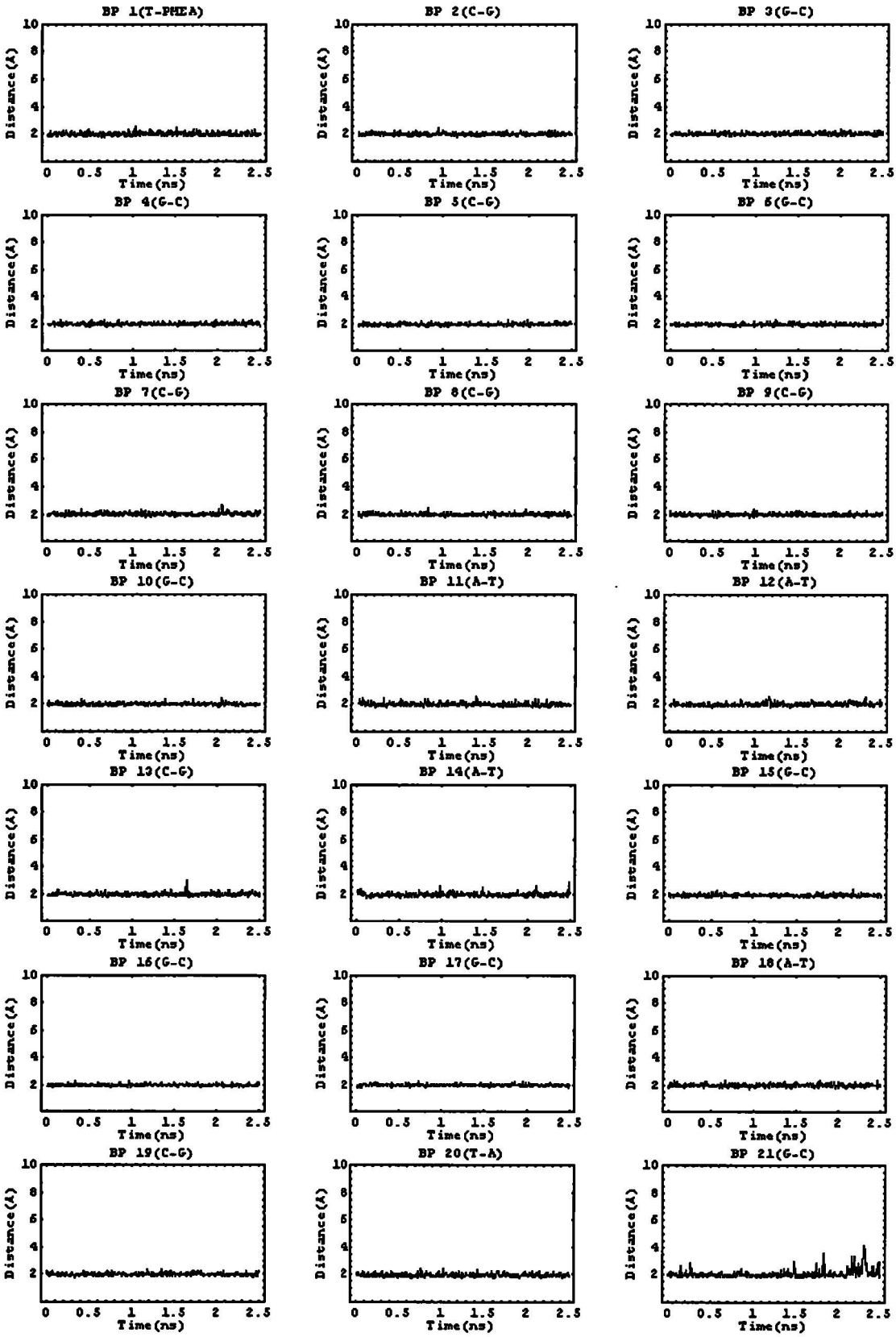


Figure D.8.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and **Adefovir** after incorporation with the primer chain. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

System no.8: HIV-1 RT + DNA:DNA + PMPA (after incorporation into the primer DNA)

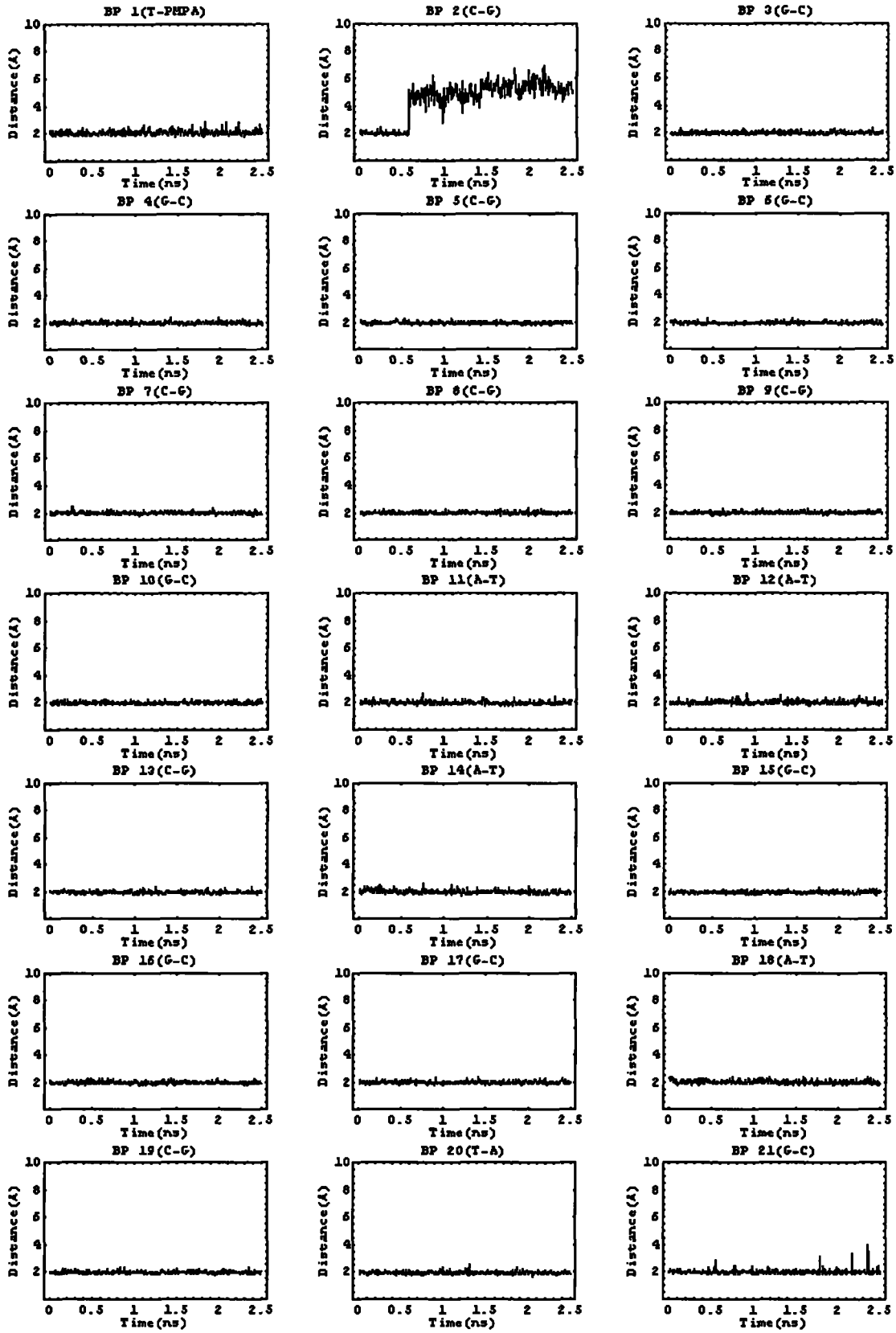
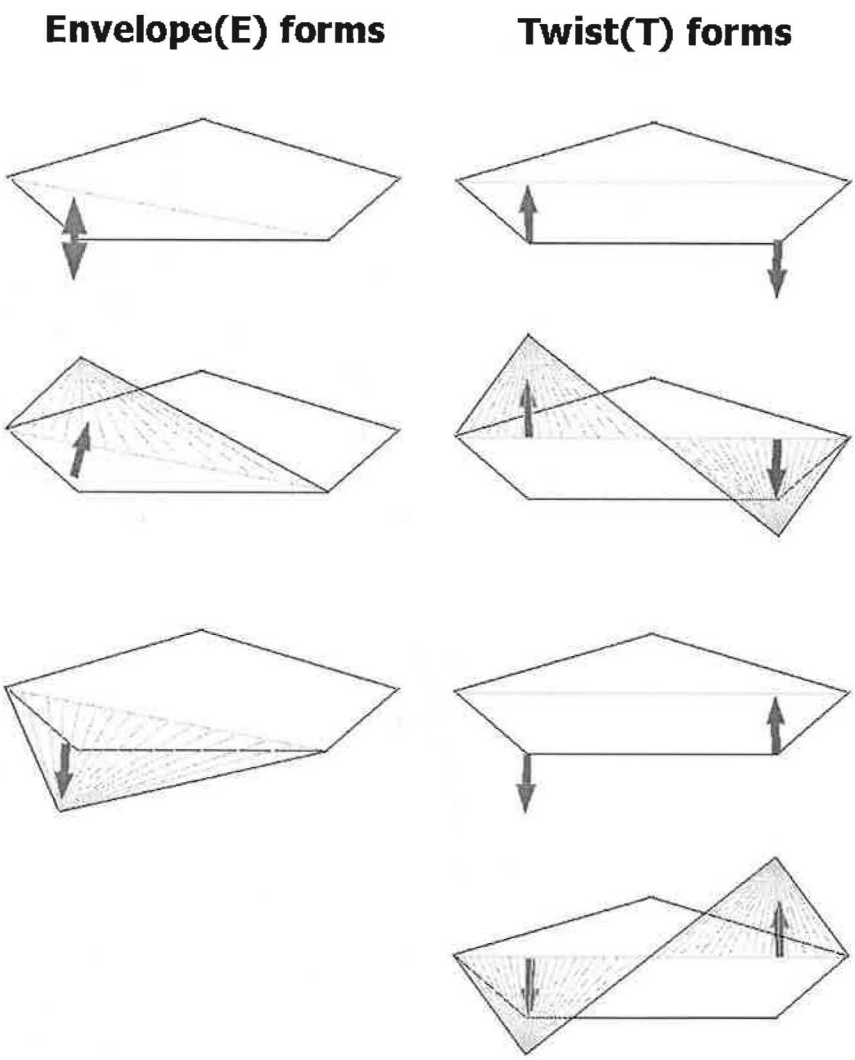


Figure D.9.: The distance between N-H (from Thymine (T) and Guanine (G)) and N (from Adenine (A) and Cytosine (C)) in each base pair during the simulation. The first pair (BP 1) is the pair of Thymine and **Tenofovir** after incorporation with the primer chain. In the title of each graphs above, the first letter indicates the nucleotide from the template chain and the second letter indicates the nucleotide from the primer chain.

Apendix E

Puckering of deoxyribose moieties



System no.1: HIV-1 RT + DNA:DNA + PMEApp (before incorporation into prim. DNA)

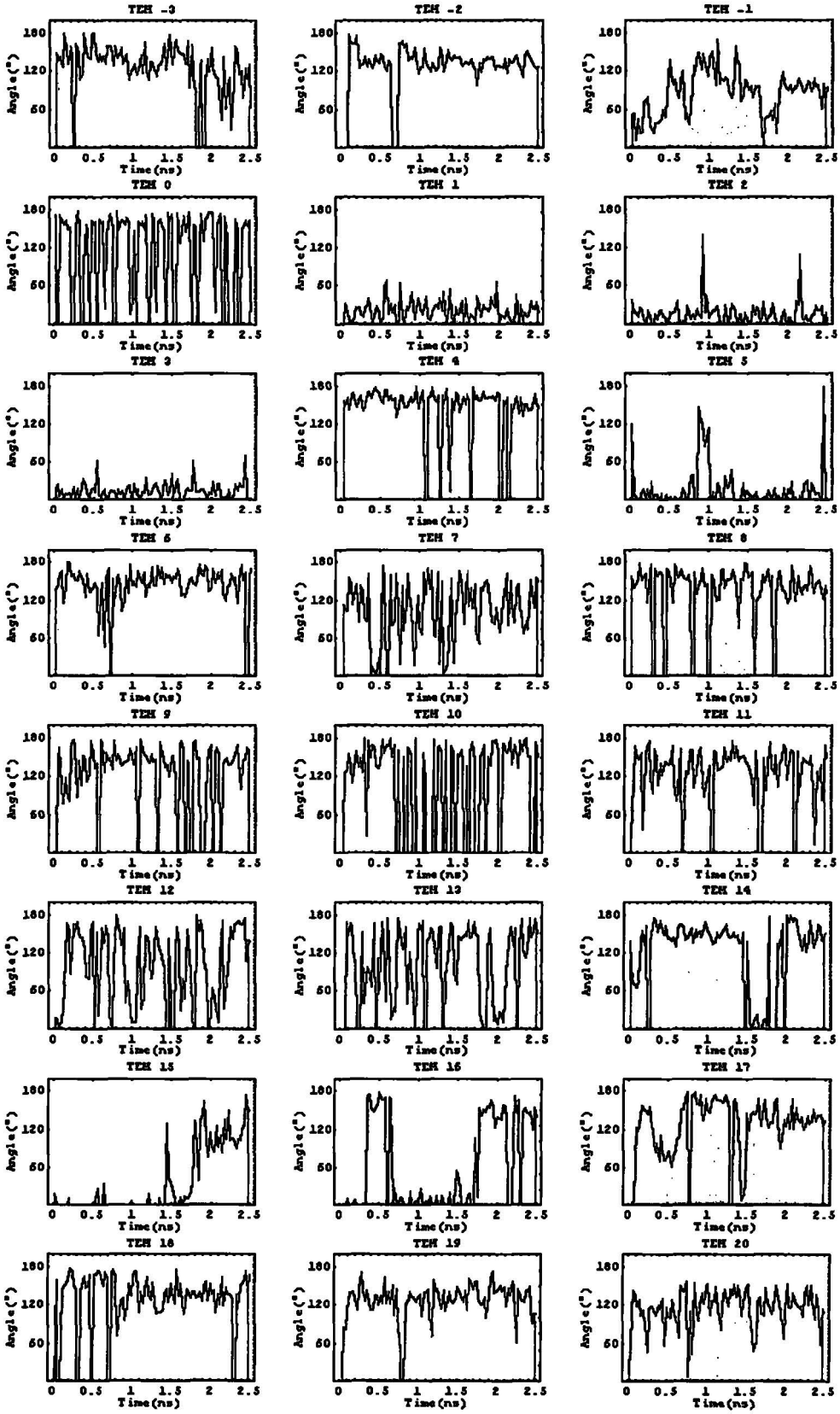


Figure E.1.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.1.

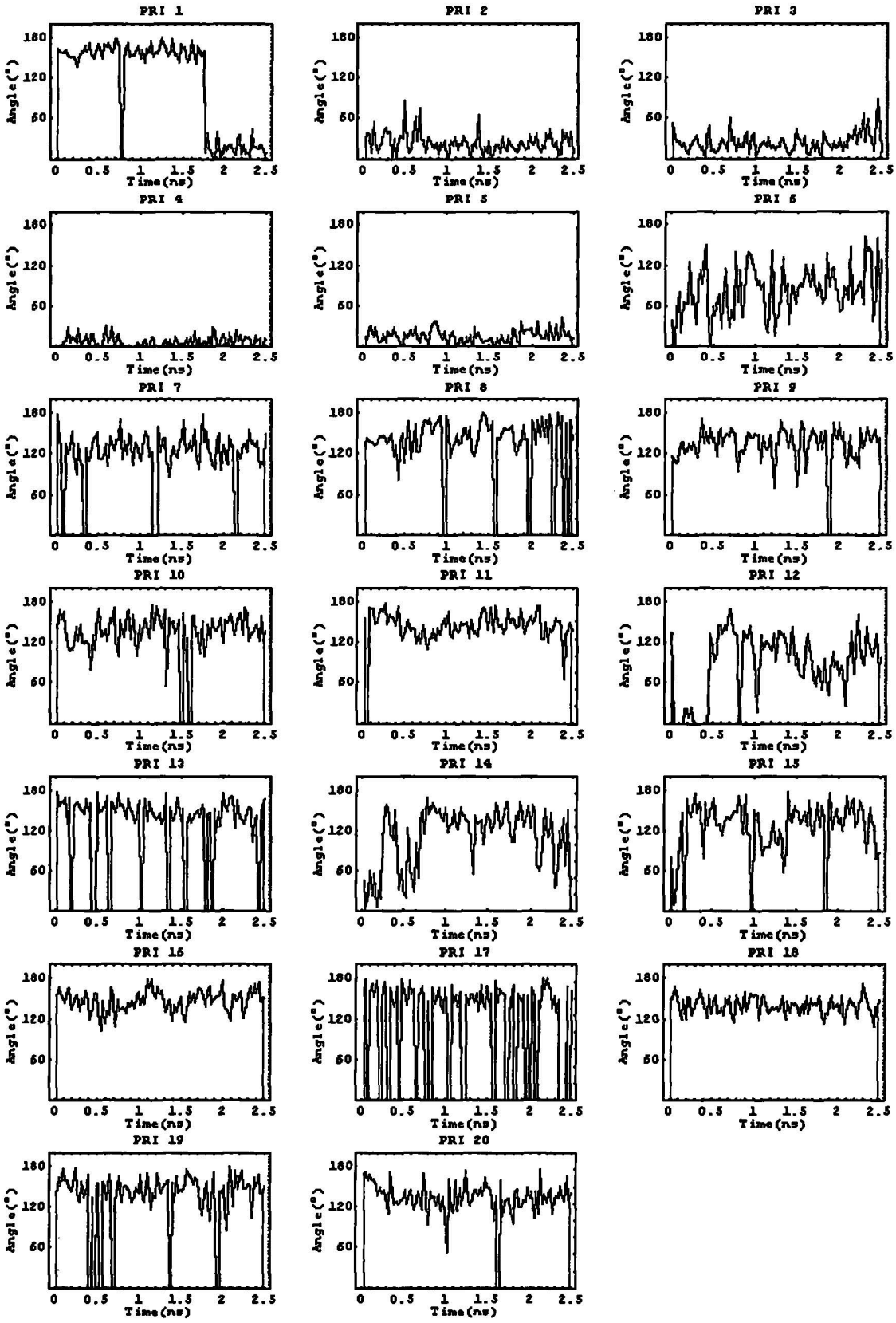


Figure E.2.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no1.

System no.2: HIV-1 RT + DNA:DNA + PMPApp (before incorporation into prim. DNA)

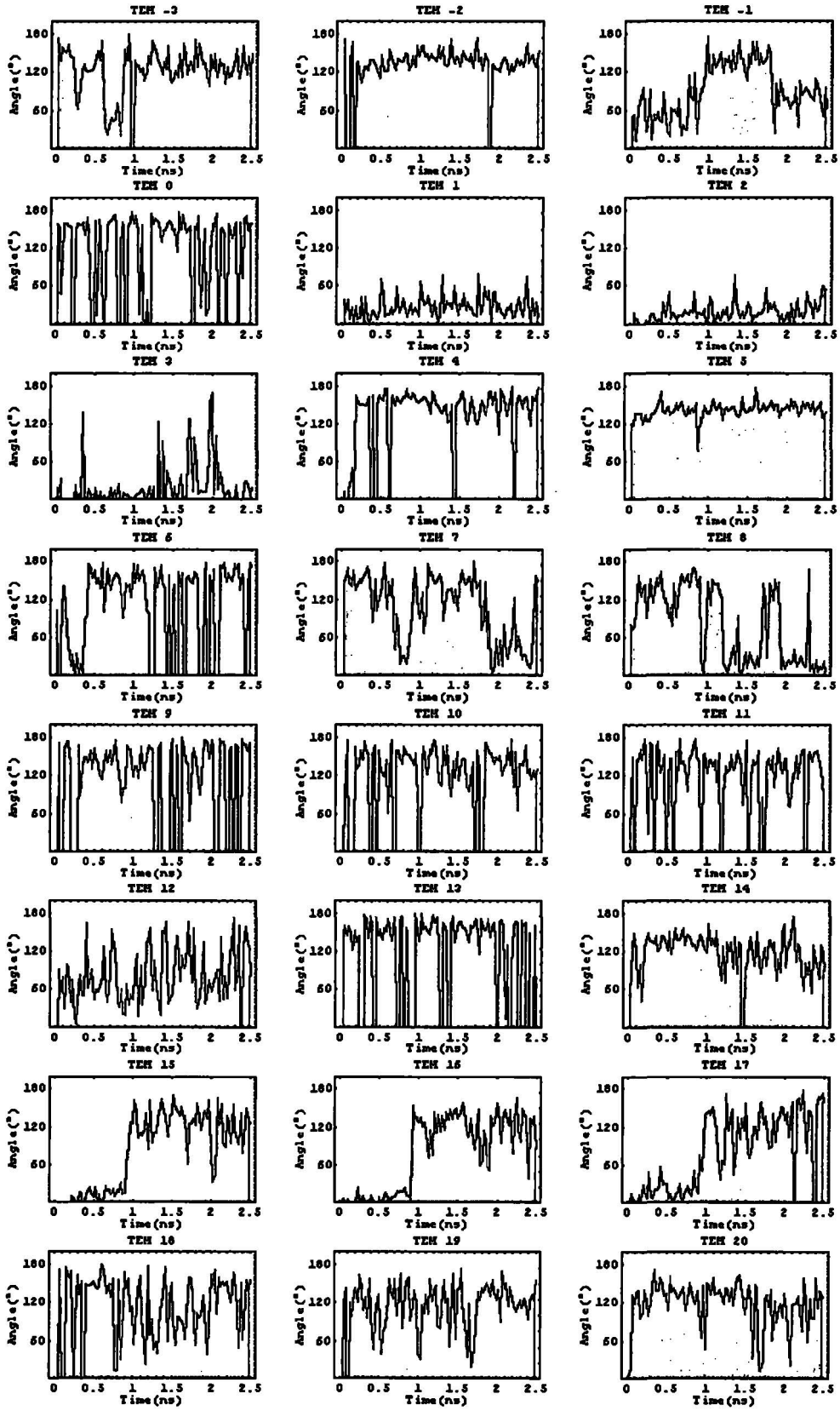


Figure E.3.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.2.

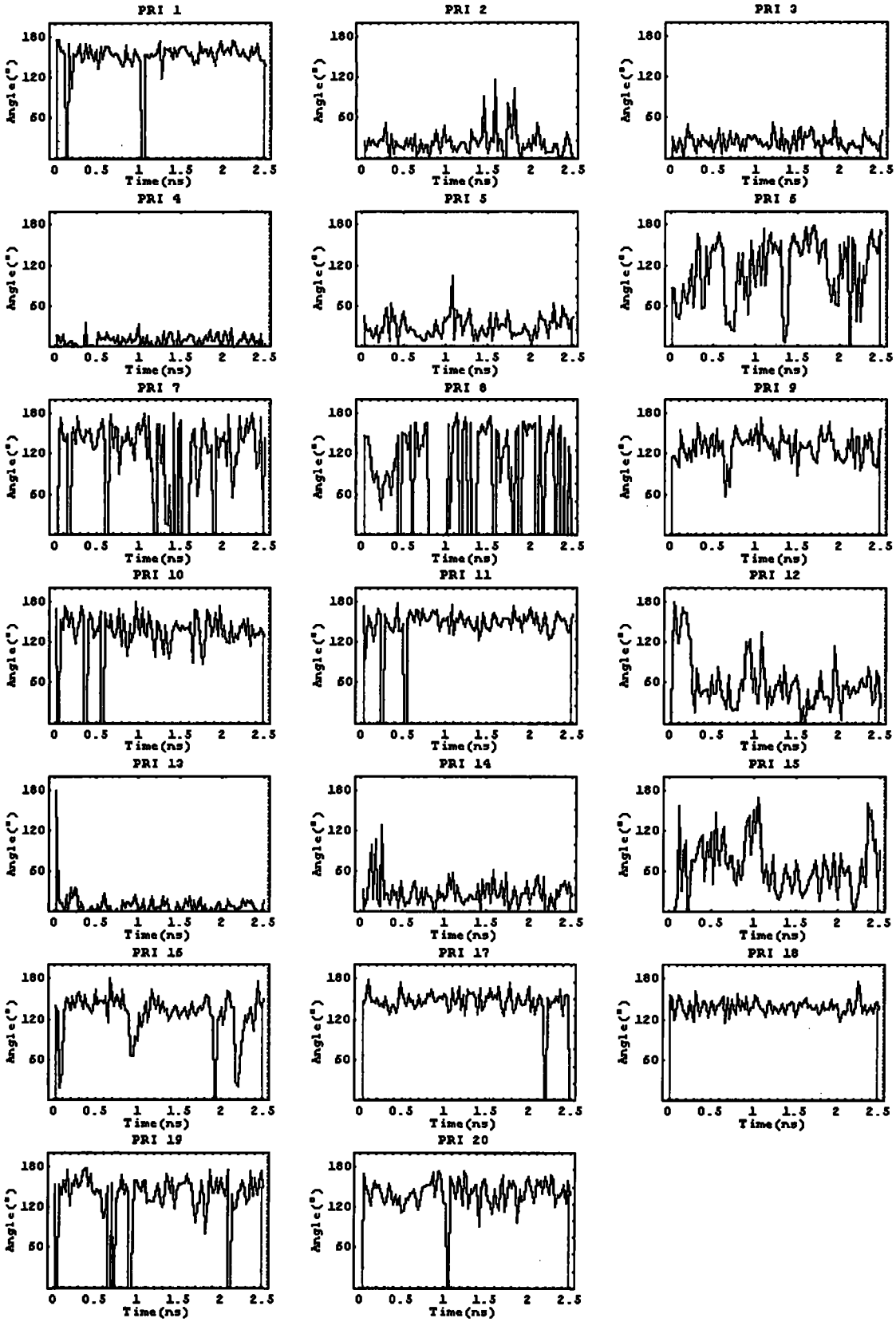


Figure E.4.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.2.

System no.3: HIV-1 RT + DNA:DNA + 3' Terminal Deoxyadenosine

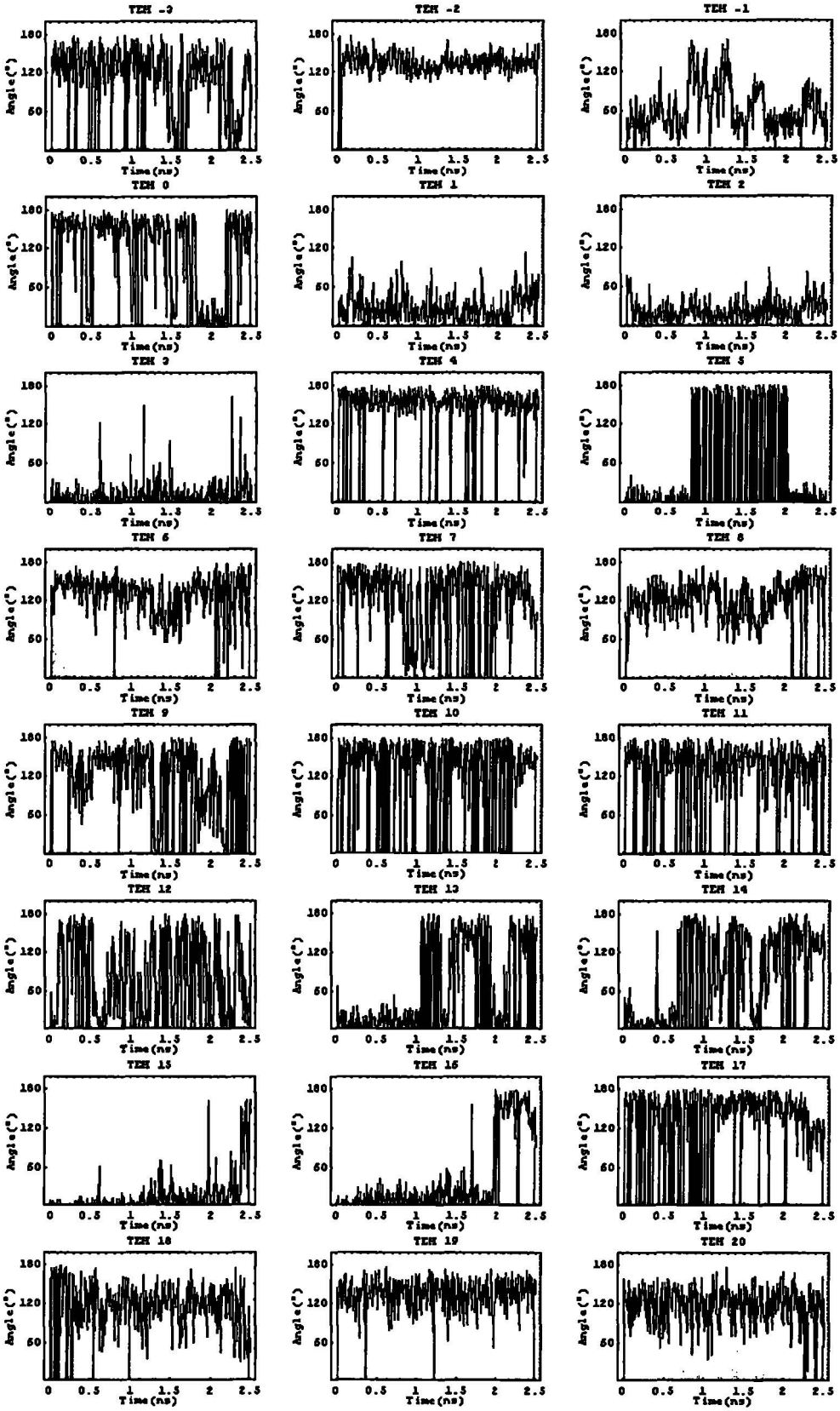


Figure E.5.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.3.

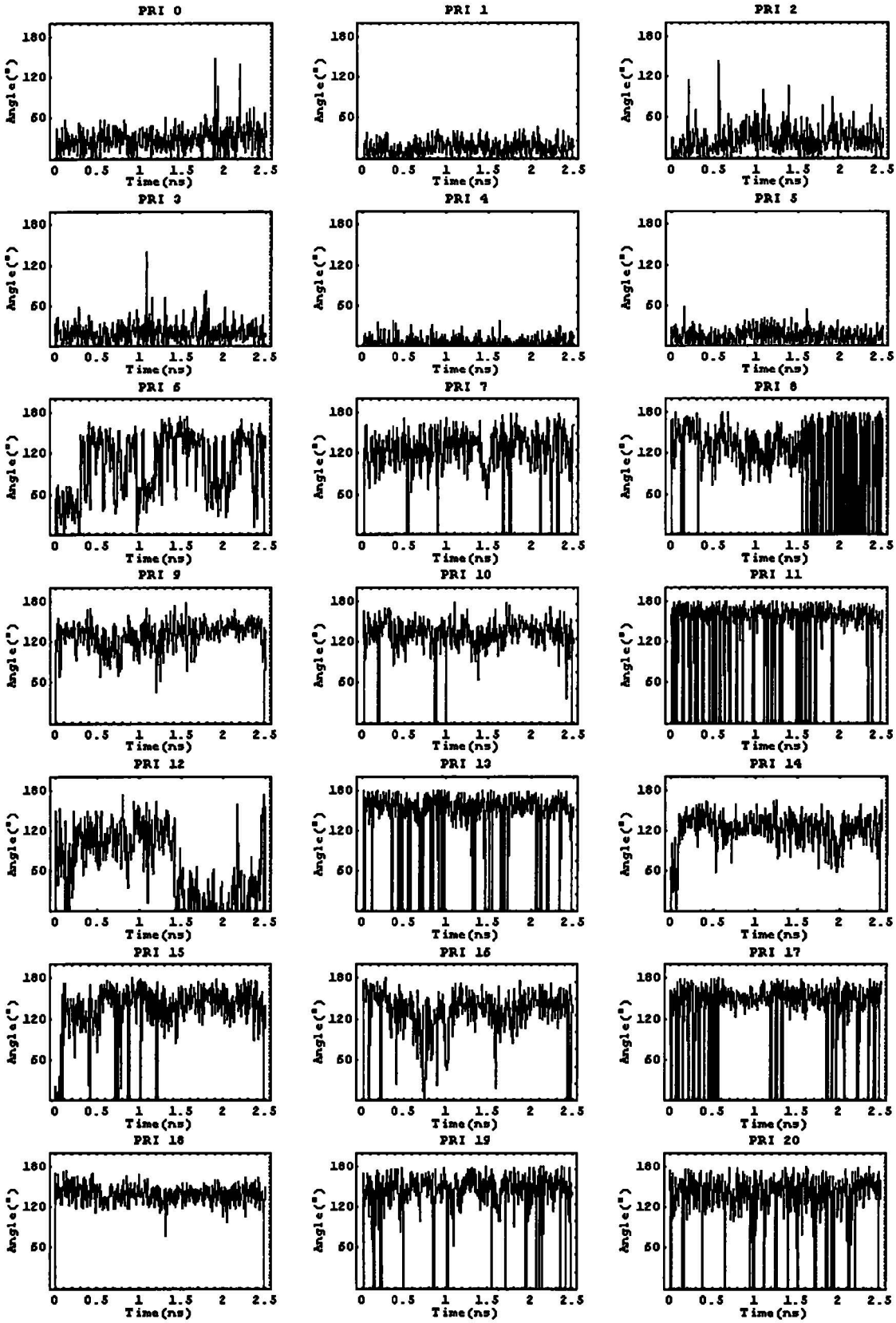


Figure E.6.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.3.

System no.4: HIV-1 RT + DNA:DNA + 3' Terminal Adenosine

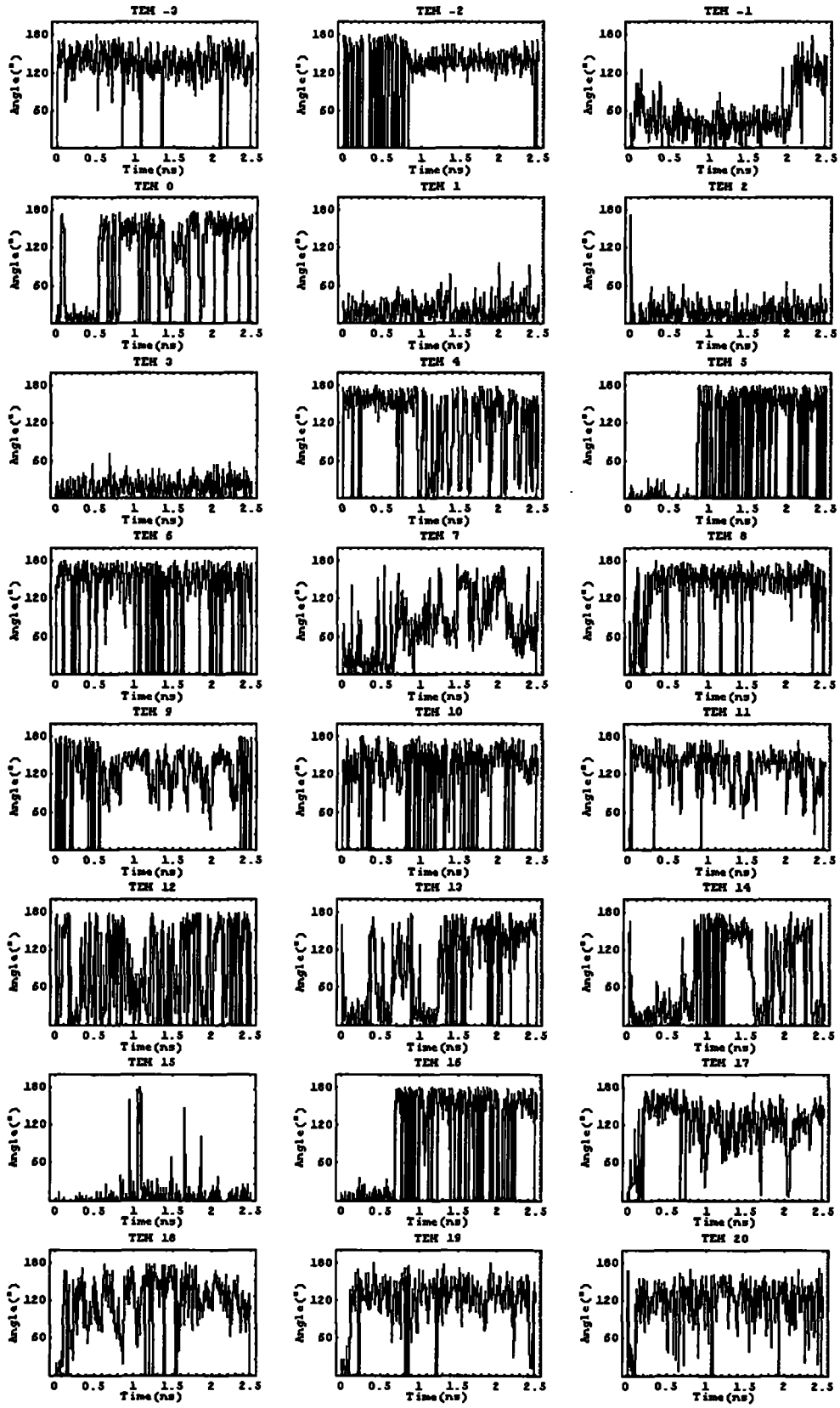


Figure E.7.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.4.

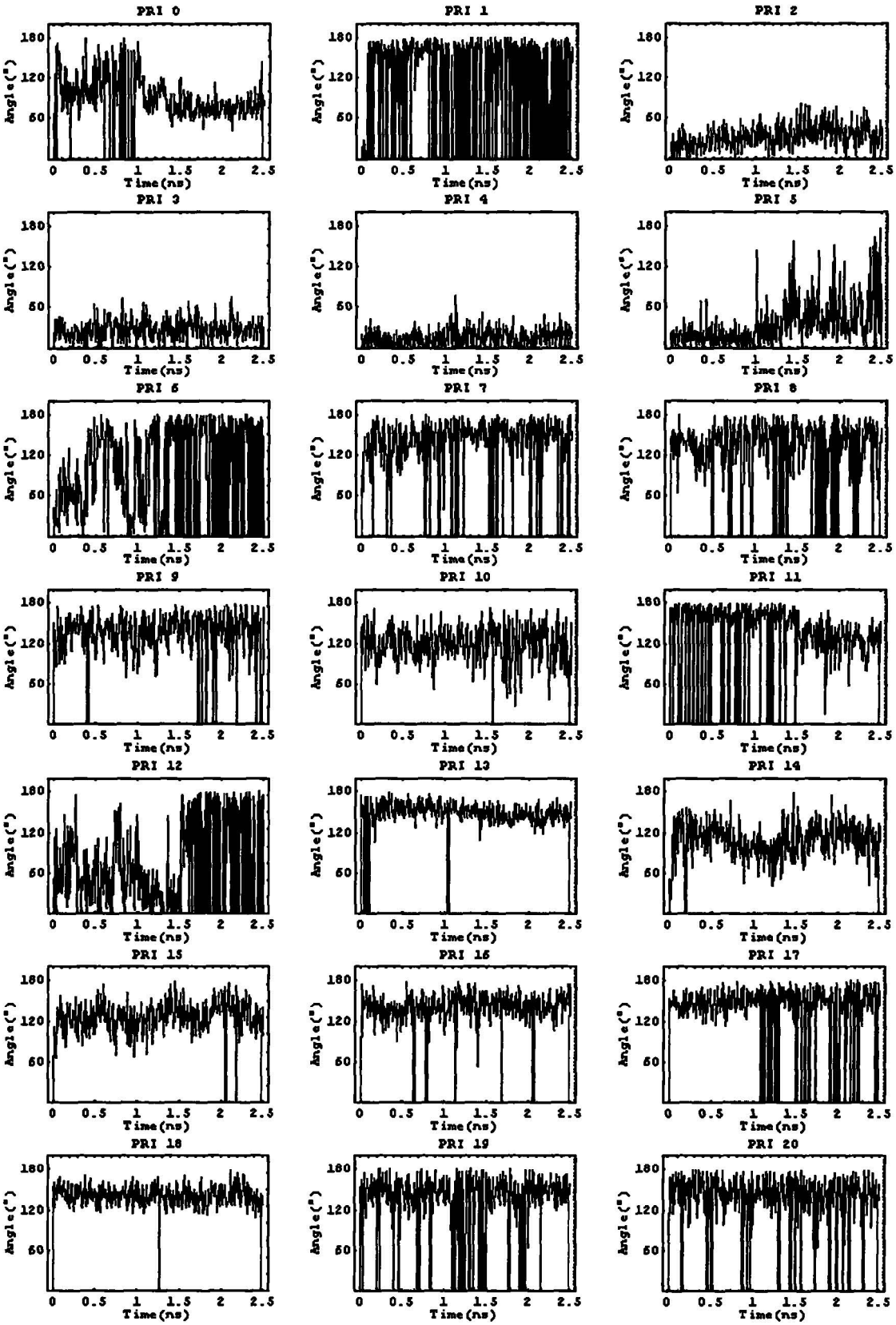


Figure E.8.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.4.

System no.5: HIV-1 RT + DNA:DNA + dATP

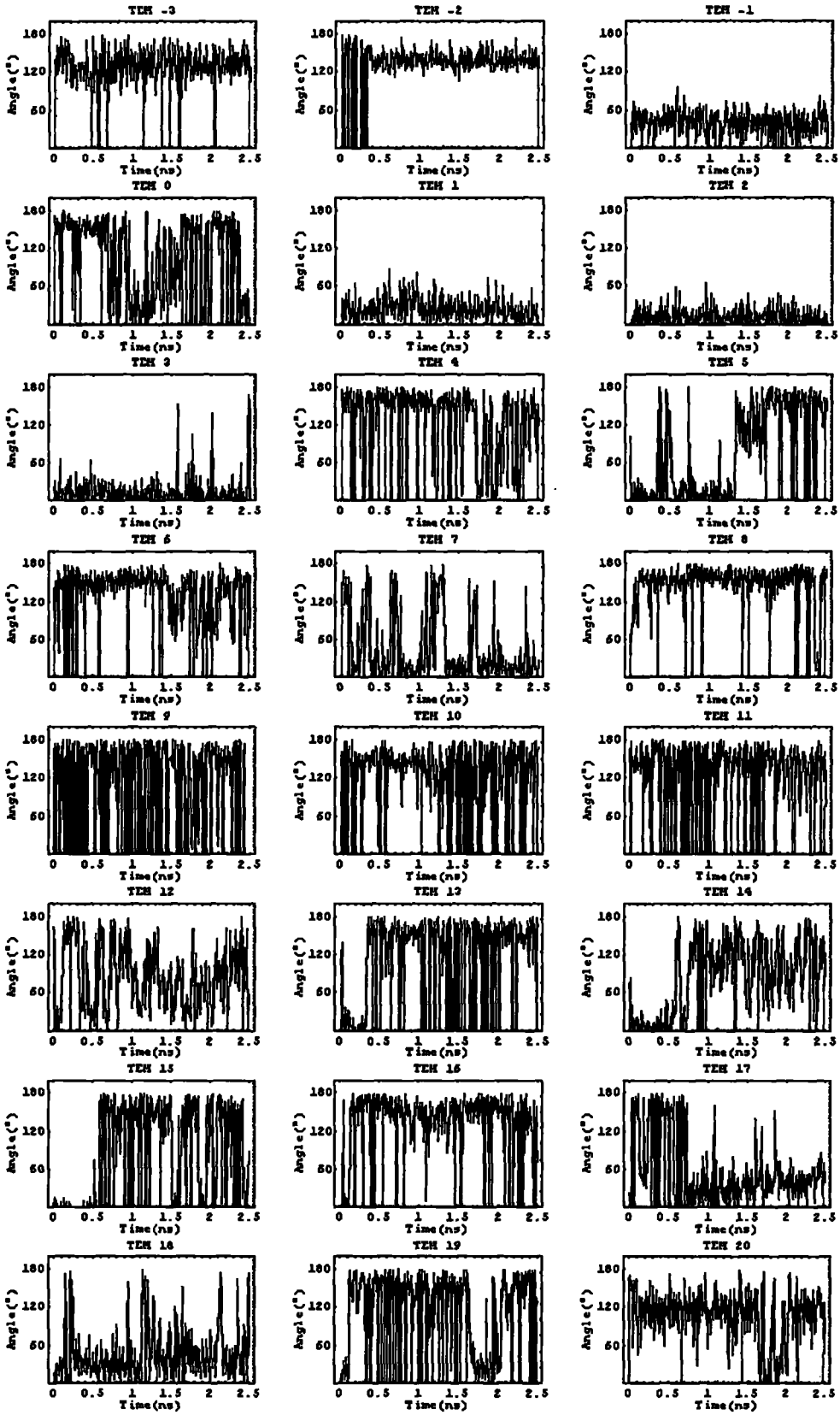


Figure E.9.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.5.

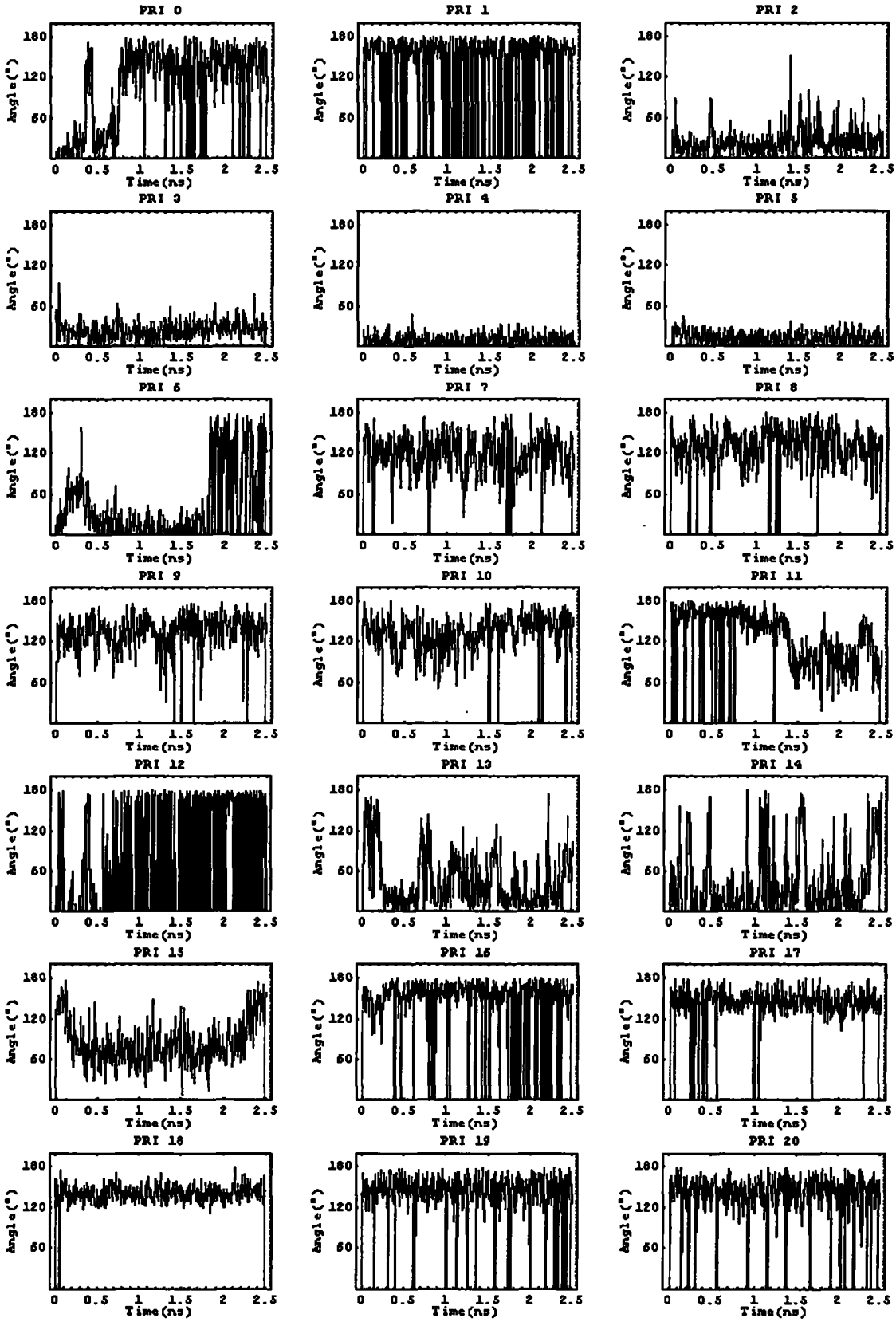


Figure E.10.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.5.

System no.6 I: HIV-1 RT + DNA:DNA + ATP

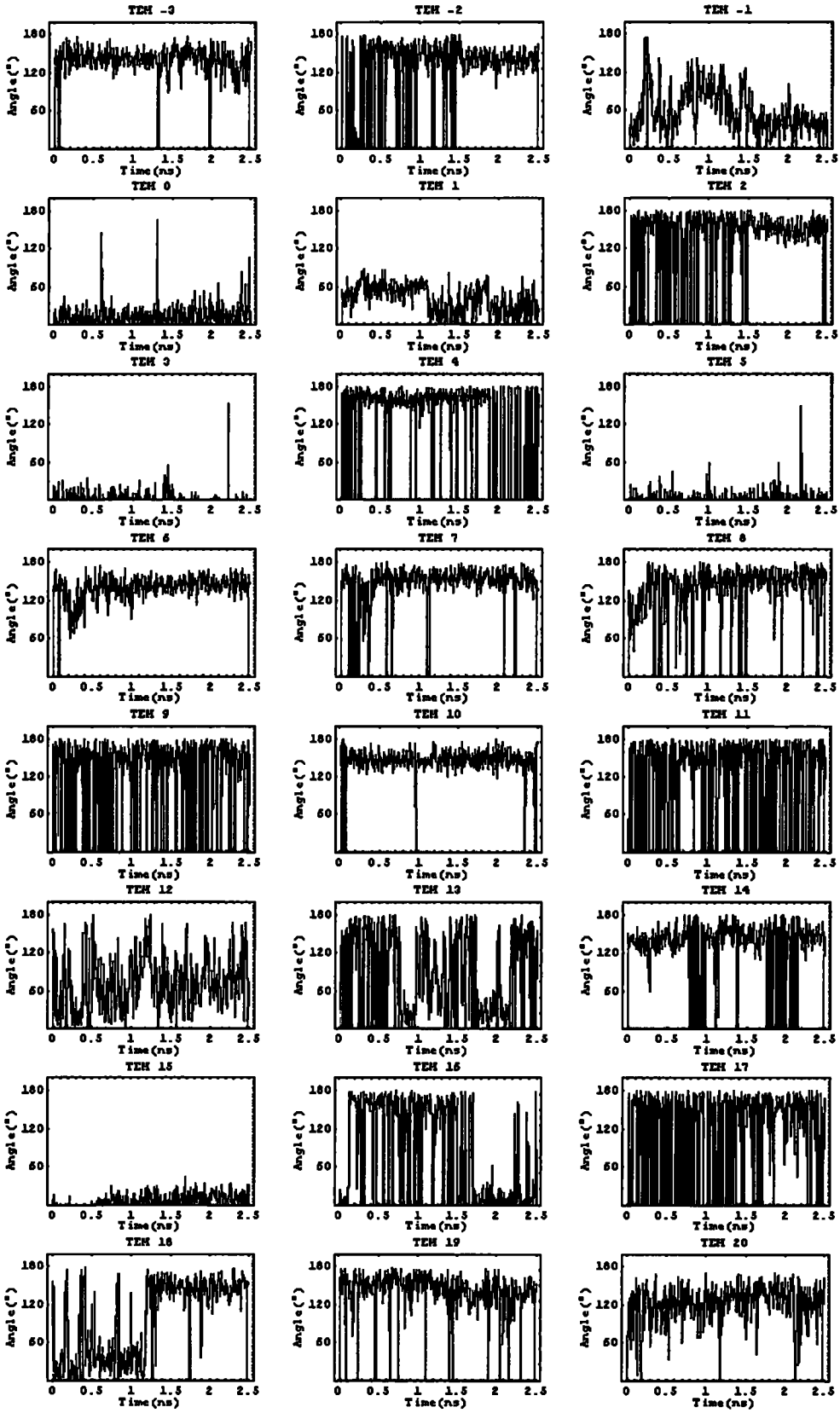


Figure E.11.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.6 I.

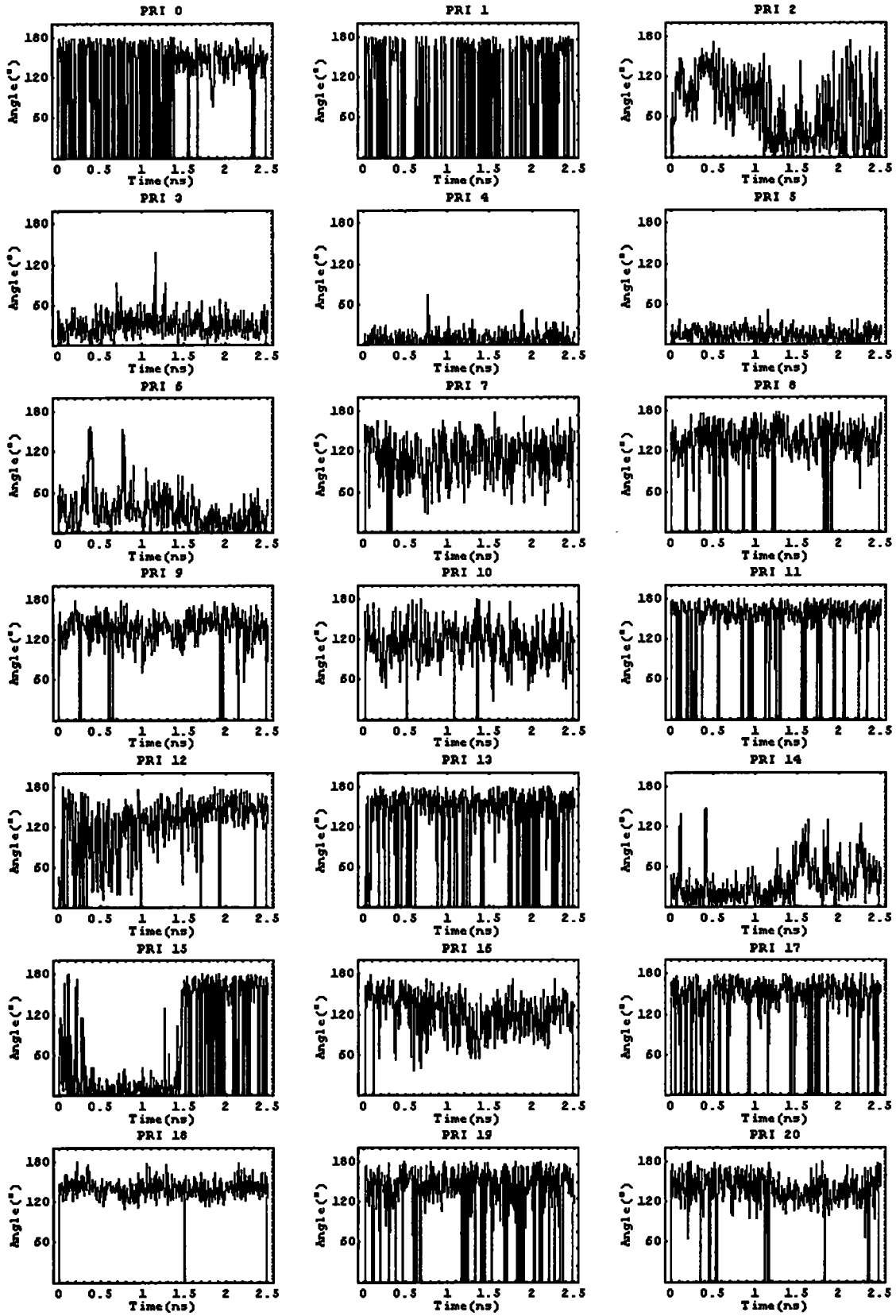


Figure E.12.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.6 I.

System no.6 II: HIV-1 RT + DNA:DNA + ATP

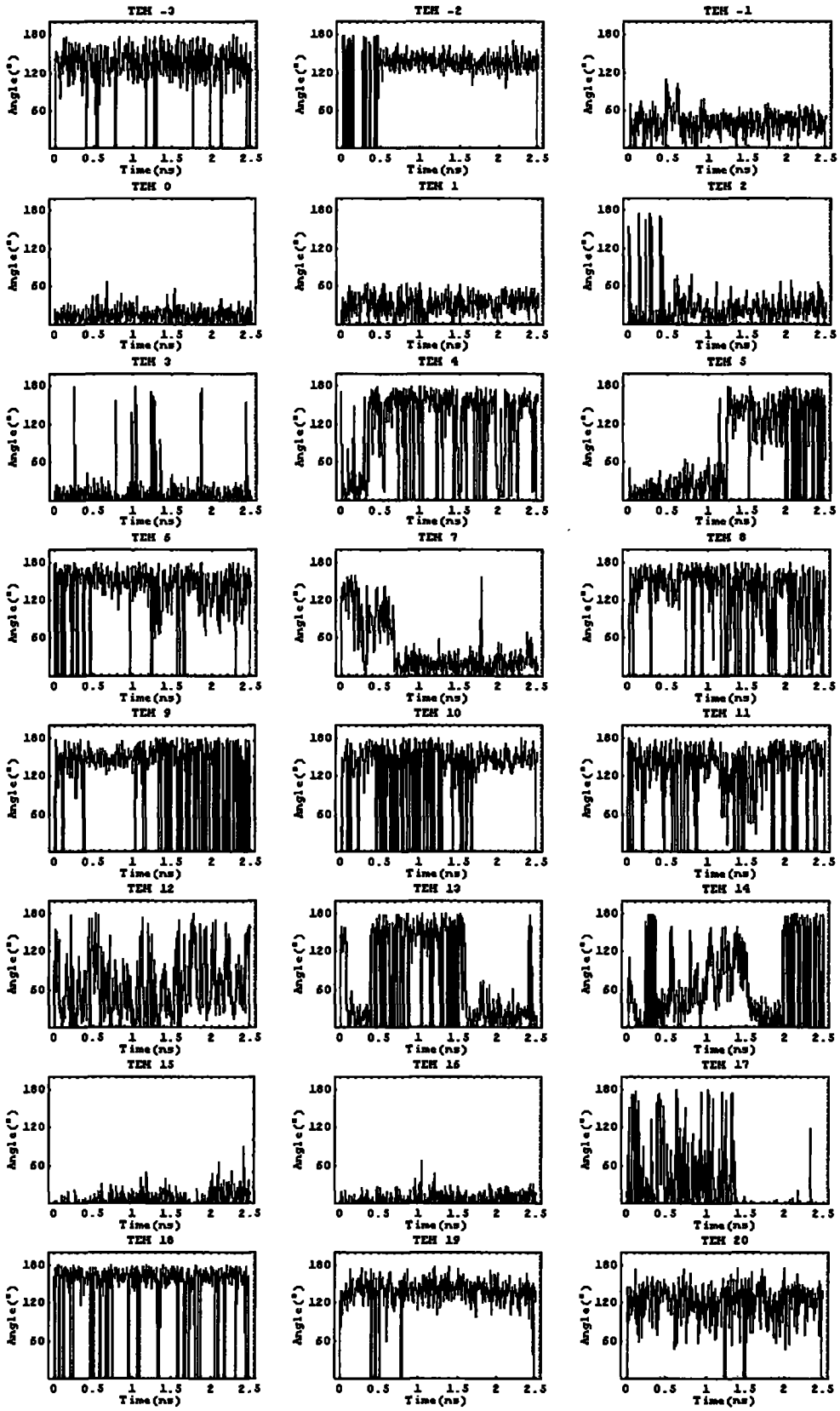


Figure E.13.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.6 II.

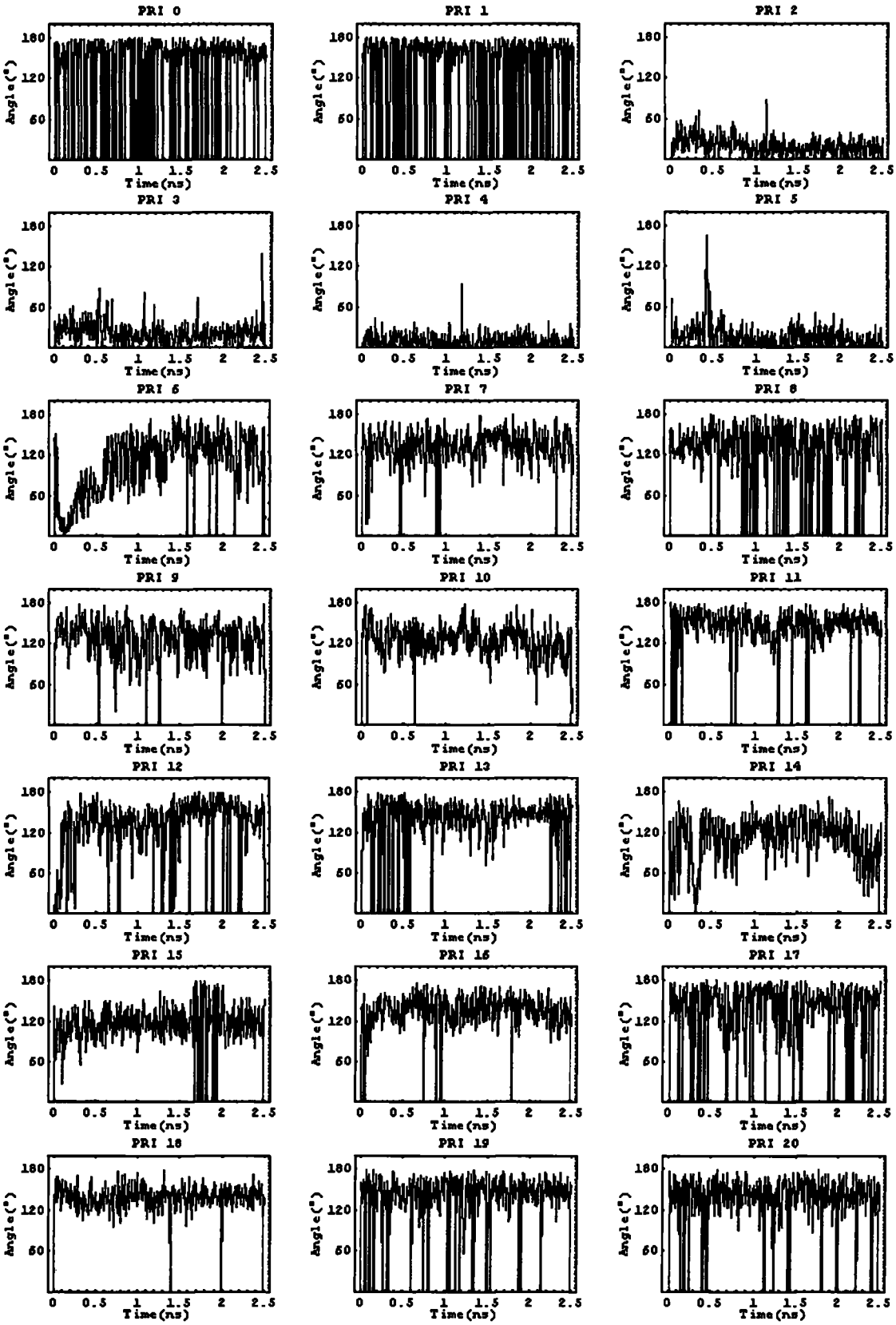


Figure E.14.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.6 II.

System no.7: HIV-1 RT + DNA:DNA + PMEA (after incorporation into primer DNA)

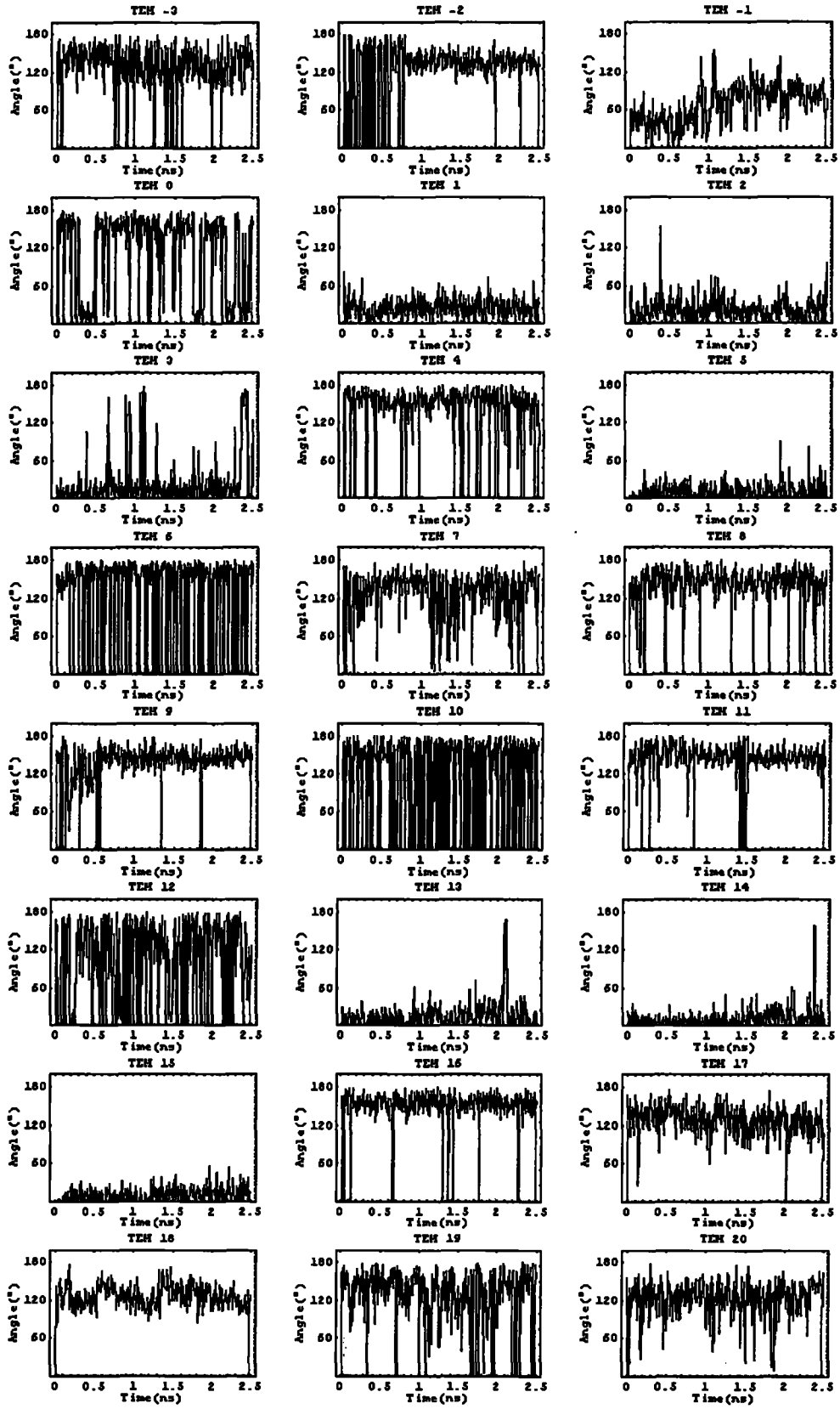


Figure E.15.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.7.

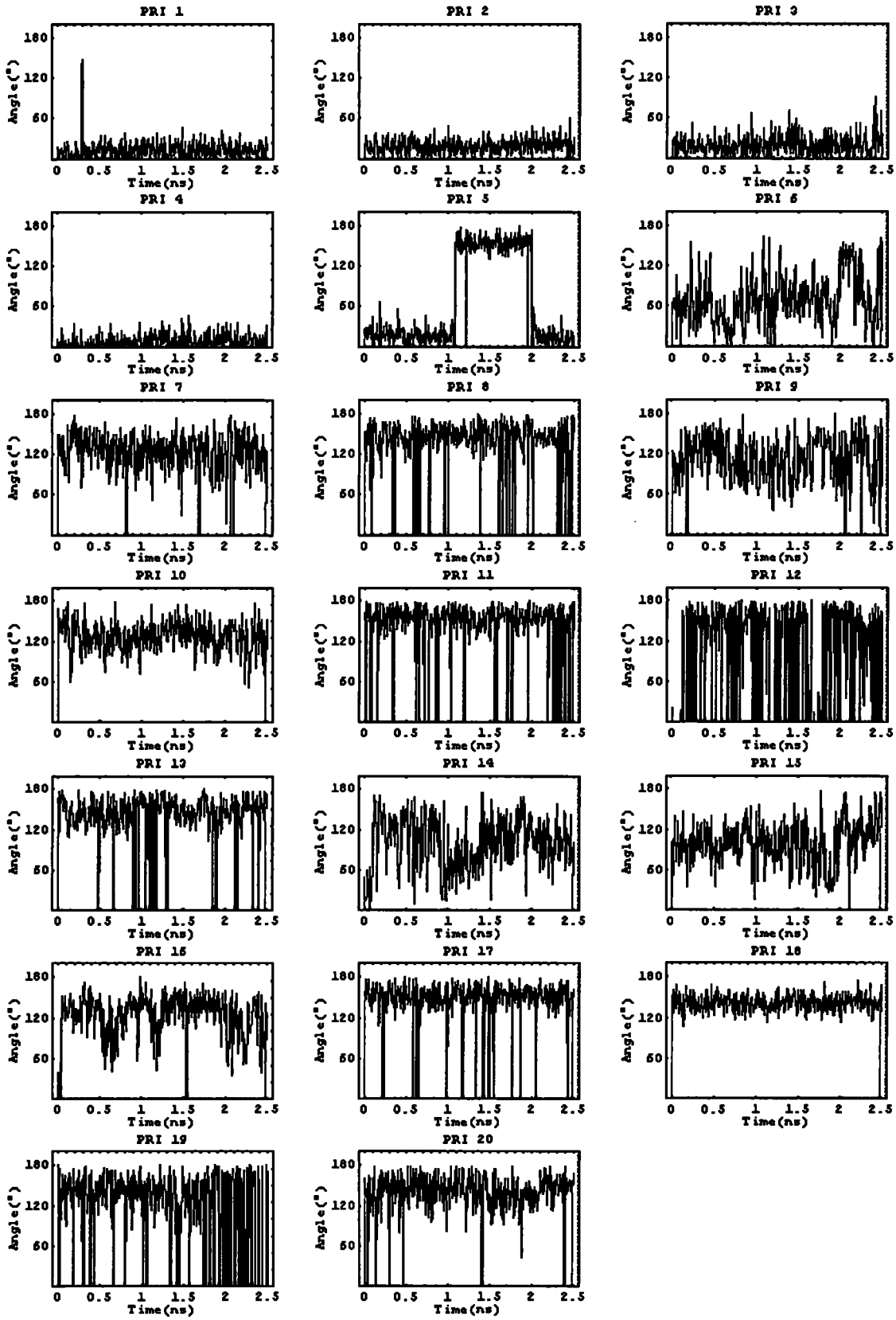


Figure E.16.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.7.

System no.8: HIV-1 RT + DNA:DNA + PMPA (after incorporation into prim. DNA)

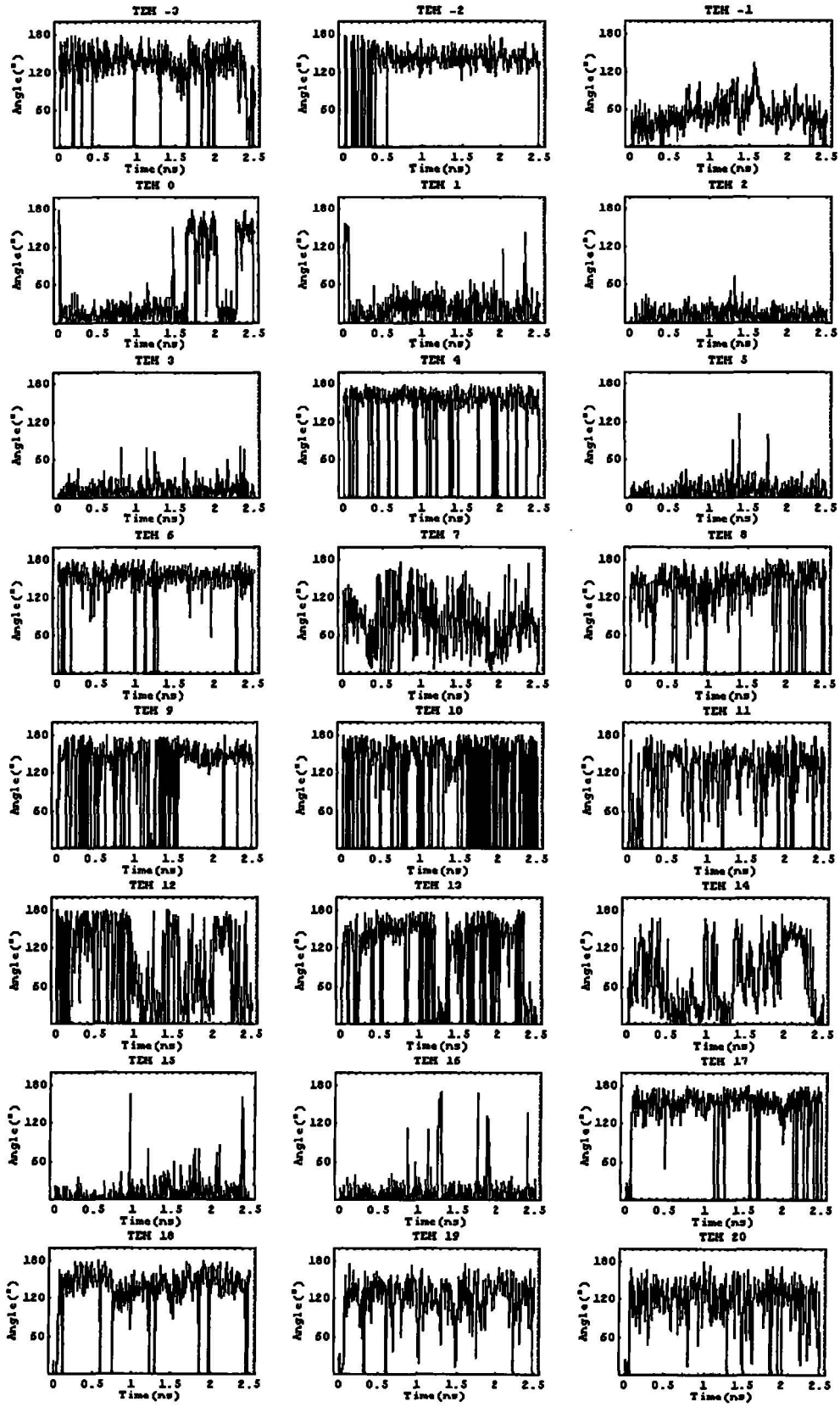


Figure E.17.: The pseudorotation angle of sugar rings from the nucleotides in the DNA template chain of the system no.8.

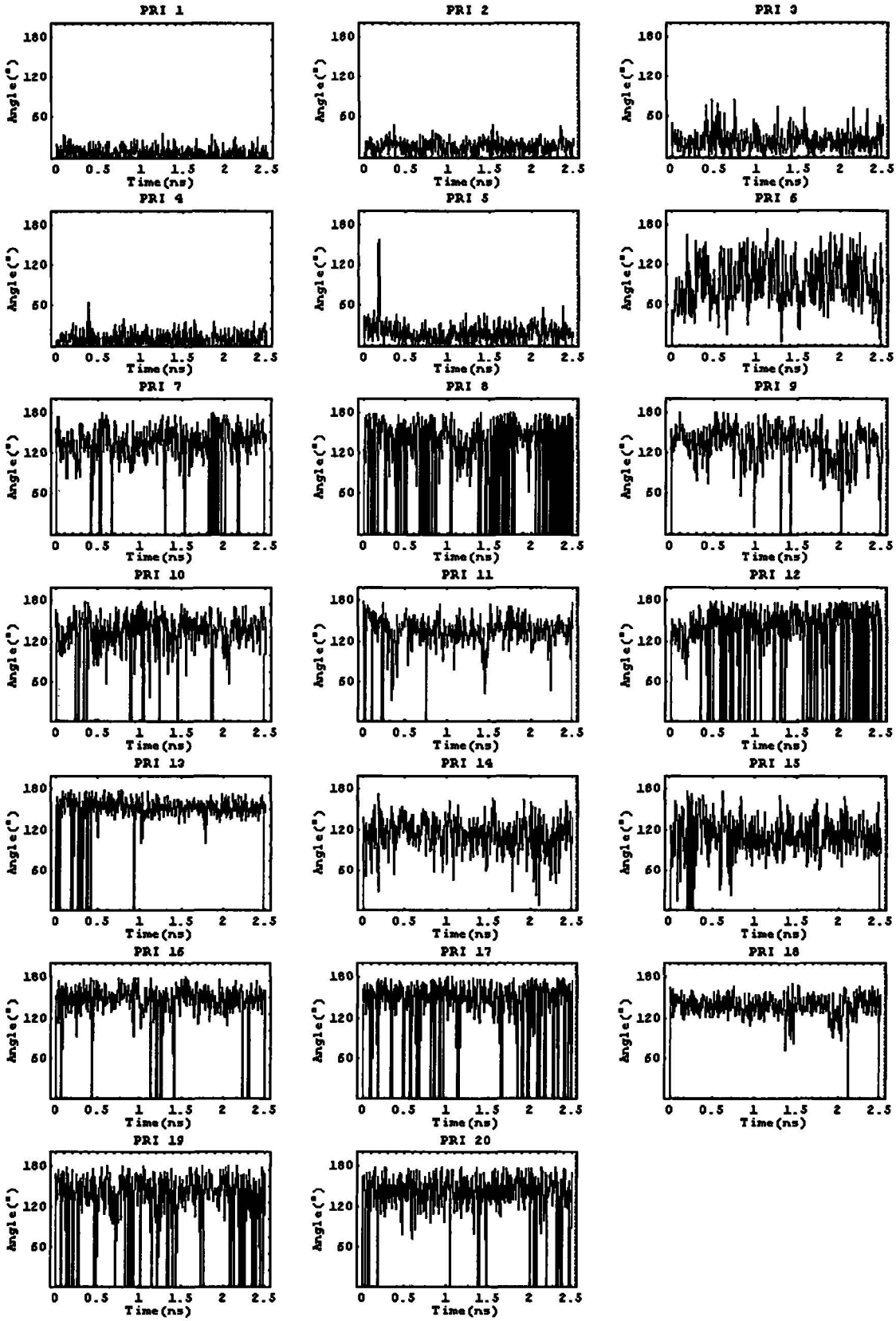
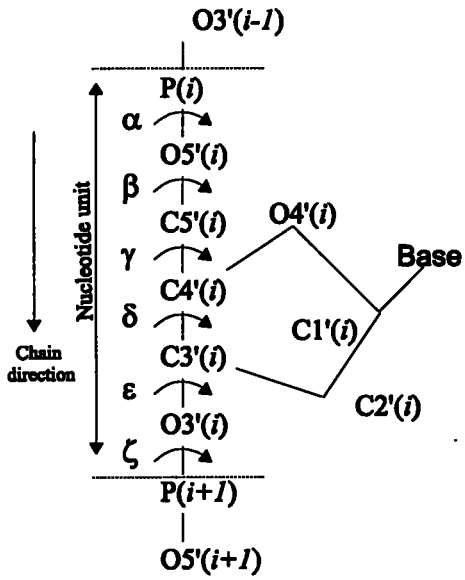


Figure E.18.: The pseudorotation angle of sugar rings from the nucleotides in the DNA primer chain of the system no.8.

Apendix F:

Phosphodiester linkages



α(*i*)=O3'(*i-1*)-P(*i*)-O5'(*i*)-C5'(*i*)

β(*i*)=P(*i*)-O5'(*i*)-C5'(*i*)-C4'(*i*)

γ(*i*)=O5'(*i*)-C5'(*i*)-C4'(*i*)-C3'(*i*)

δ(*i*)=C5'(*i*)-C4'(*i*)-C3'(*i*)-O3'(*i*)

ε(*i*)=C4'(*i*)-C3'(*i*)-O3'(*i*)-P(*i+1*)

ζ(*i*)=C3'(*i*)-O3'(*i*)-P(*i+1*)-O5'(*i+1*)

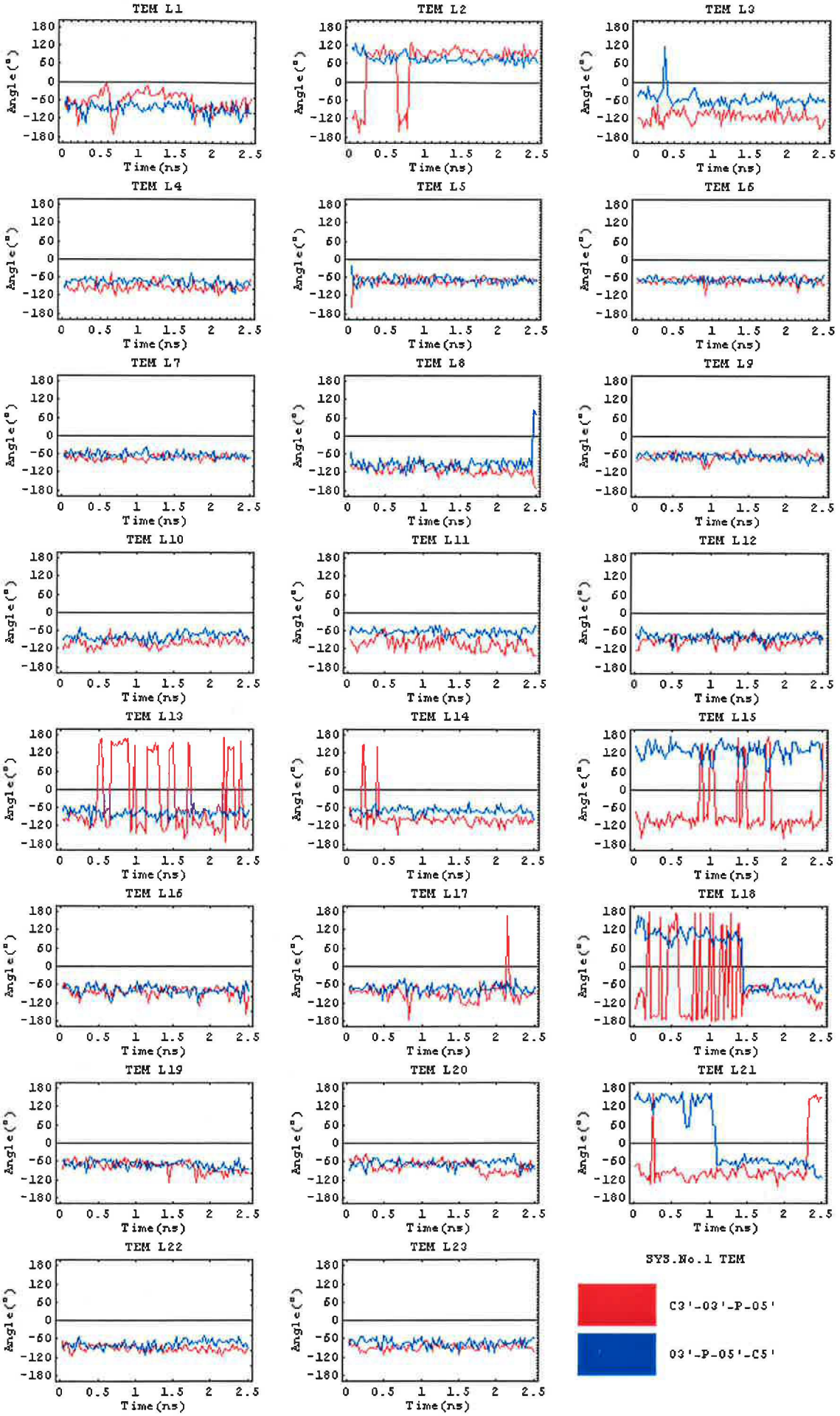


Figure F.1.: The torsion angles of the DNA template backbone in the simulation system no.1.

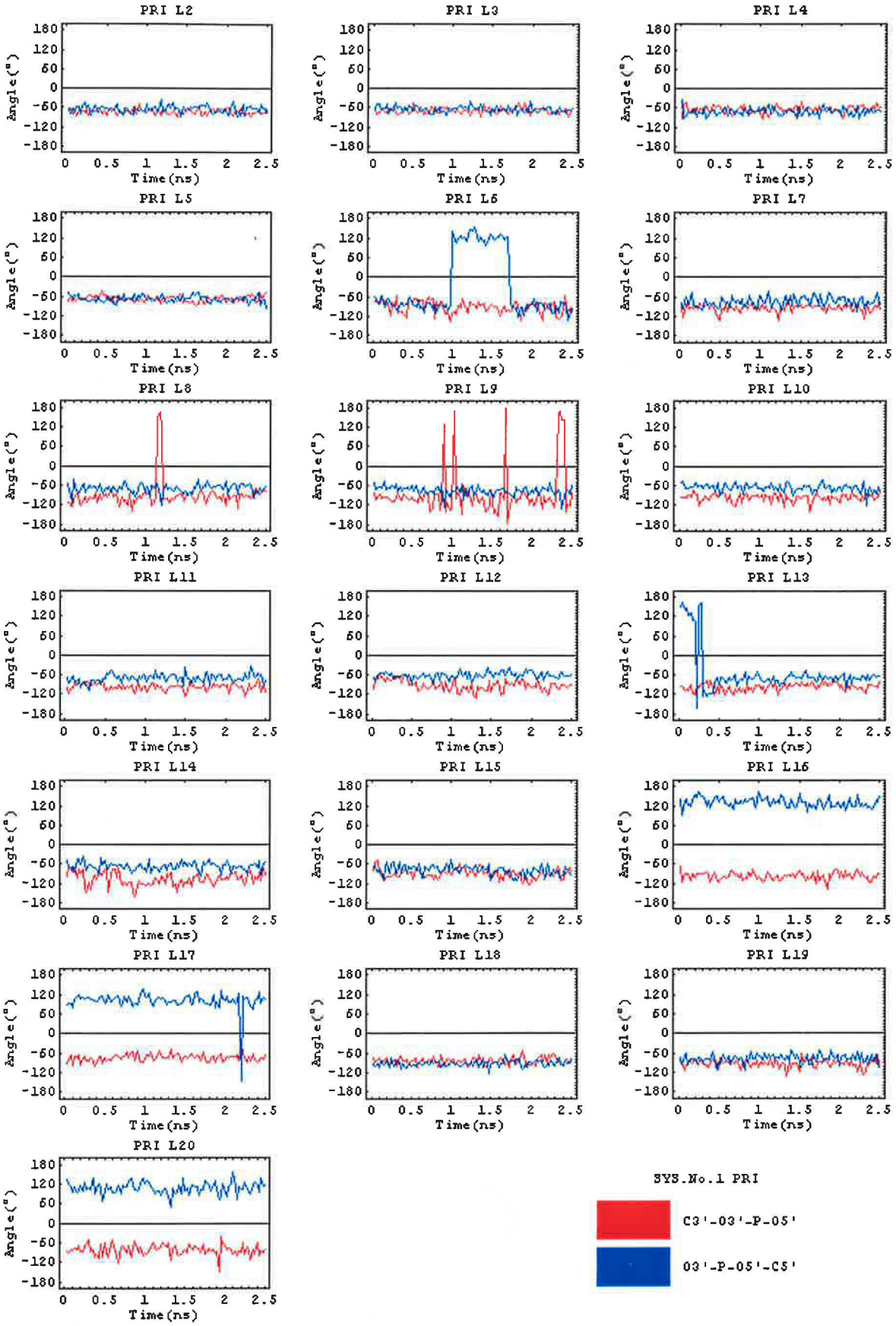


Figure F.2.: The torsion angles of the DNA primer backbone in the simulation system no.1.

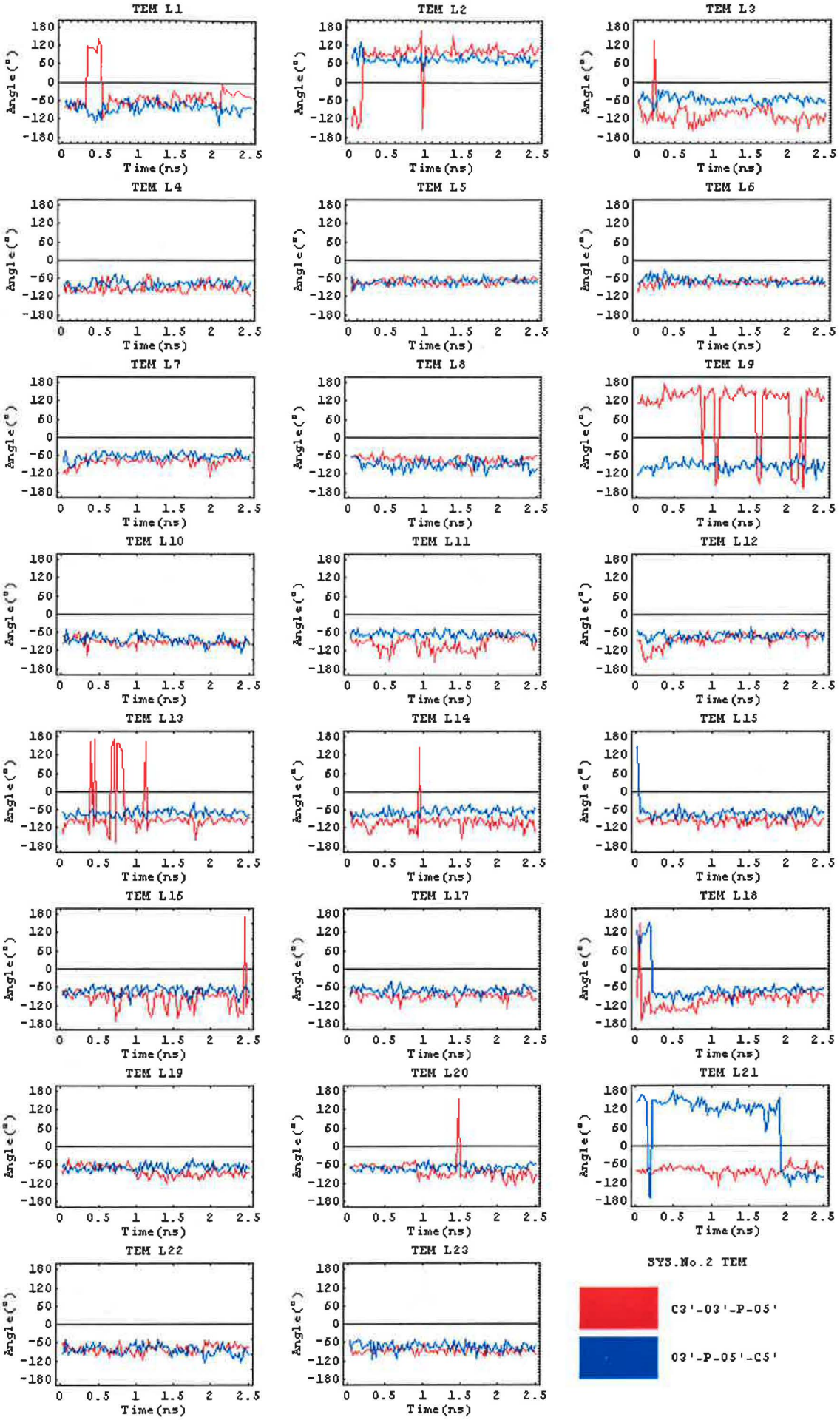


Figure F.3.: The torsion angles of the DNA template backbone in the simulation system no.2.

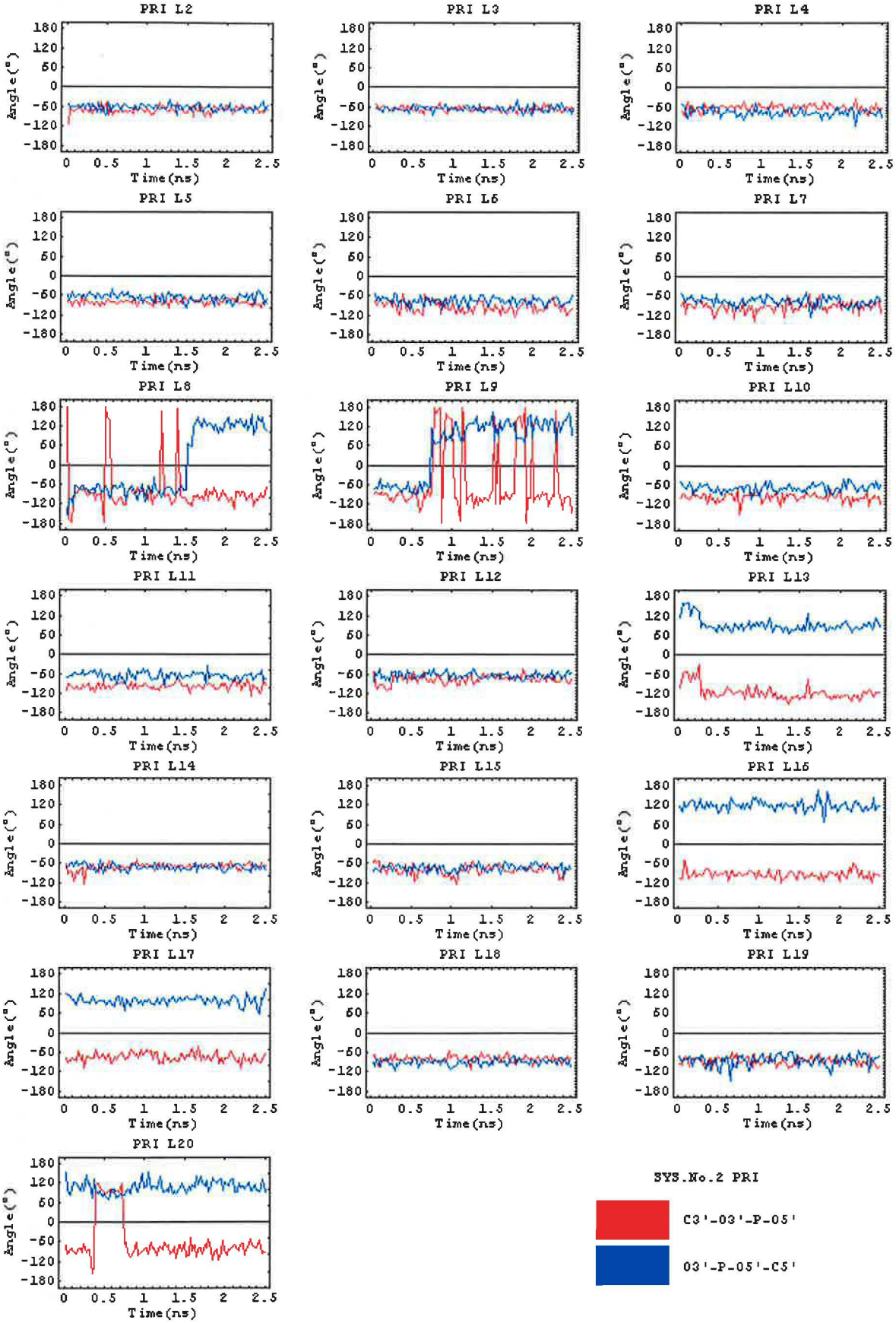


Figure F.4: The torsion angles of the DNA primer backbone in the simulation system no.2.

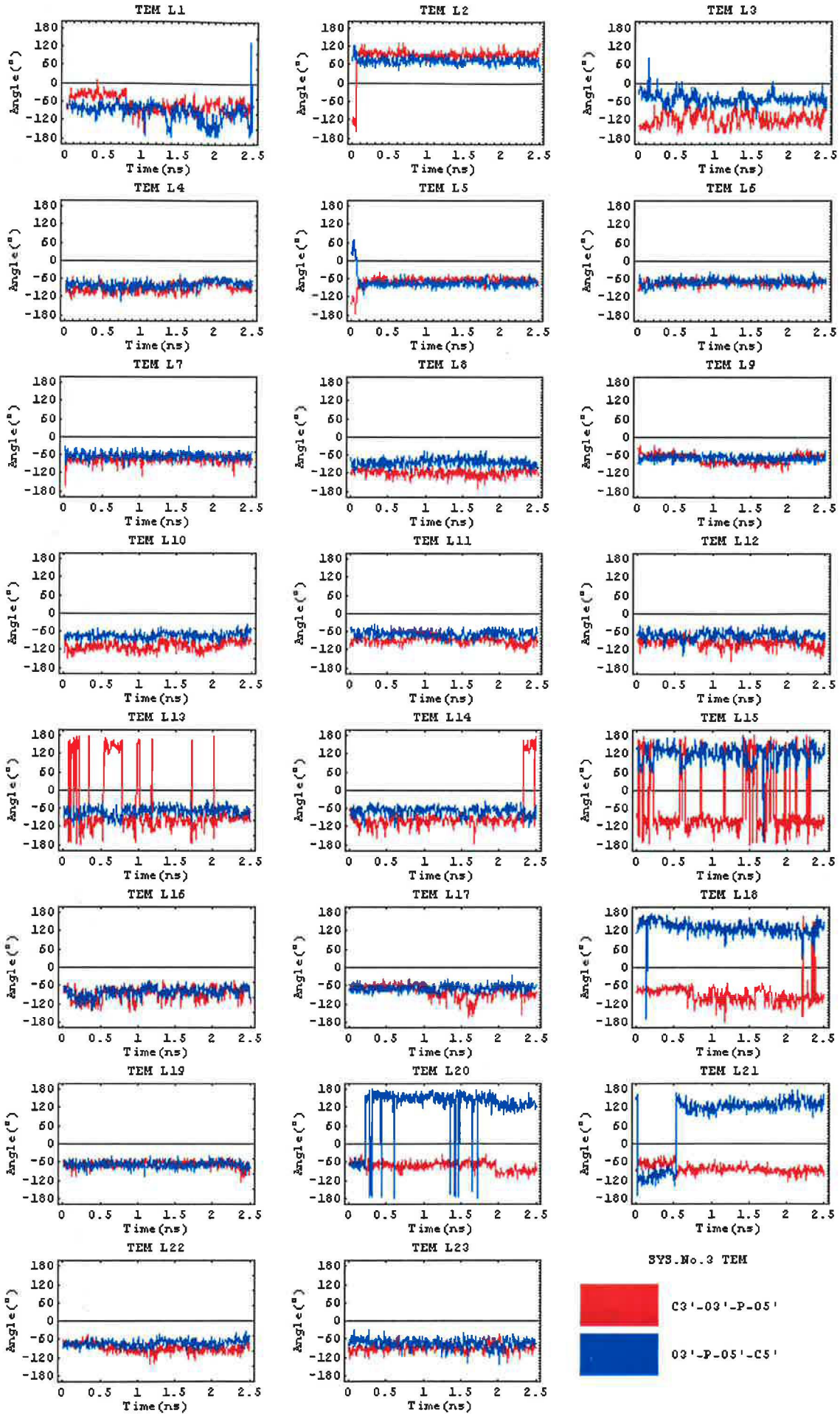


Figure F.5.: The torsion angles of the DNA template backbone in the simulation system no.3.

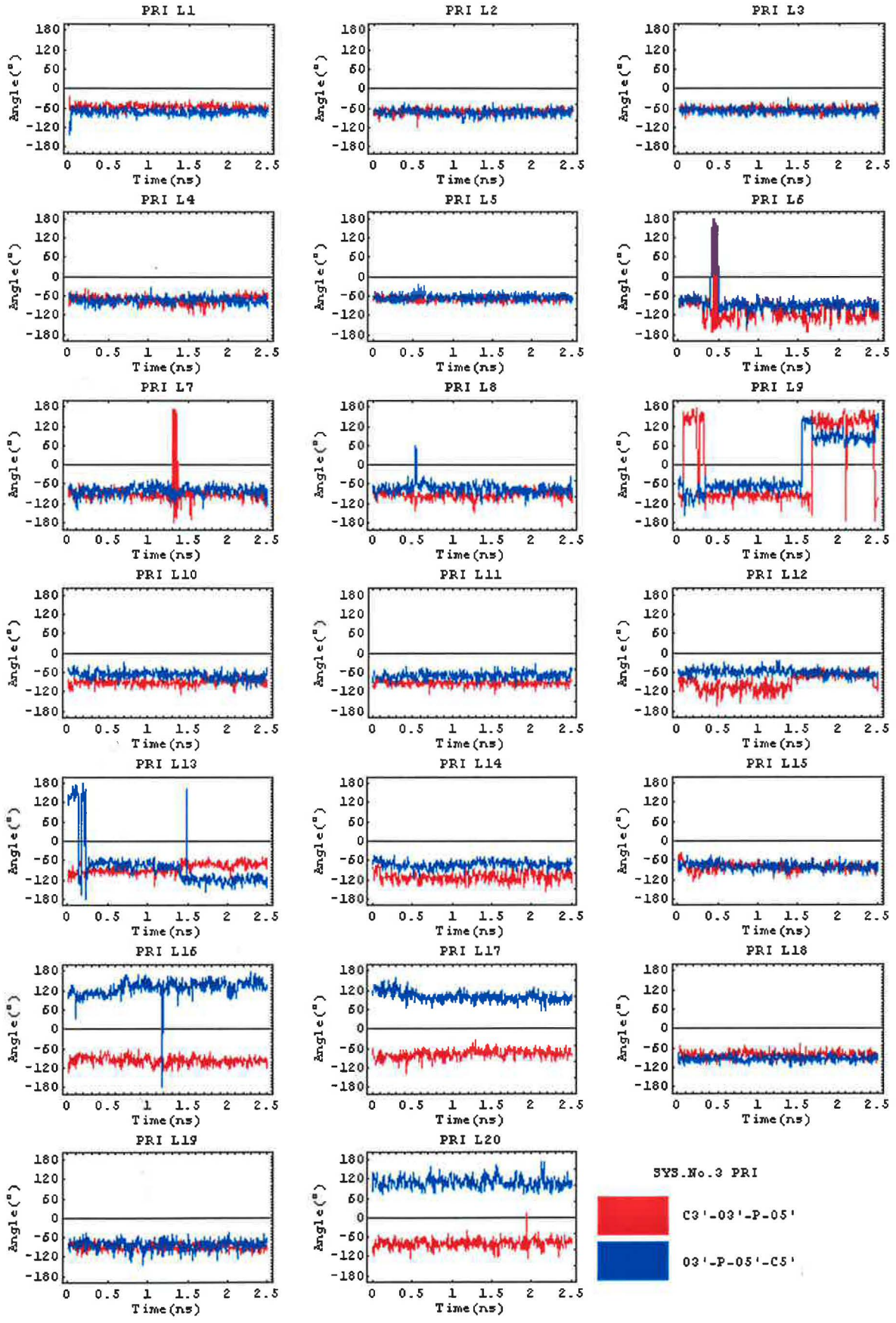


Figure F.6.: The torsion angles of the DNA primer backbone in the simulation system no.3.

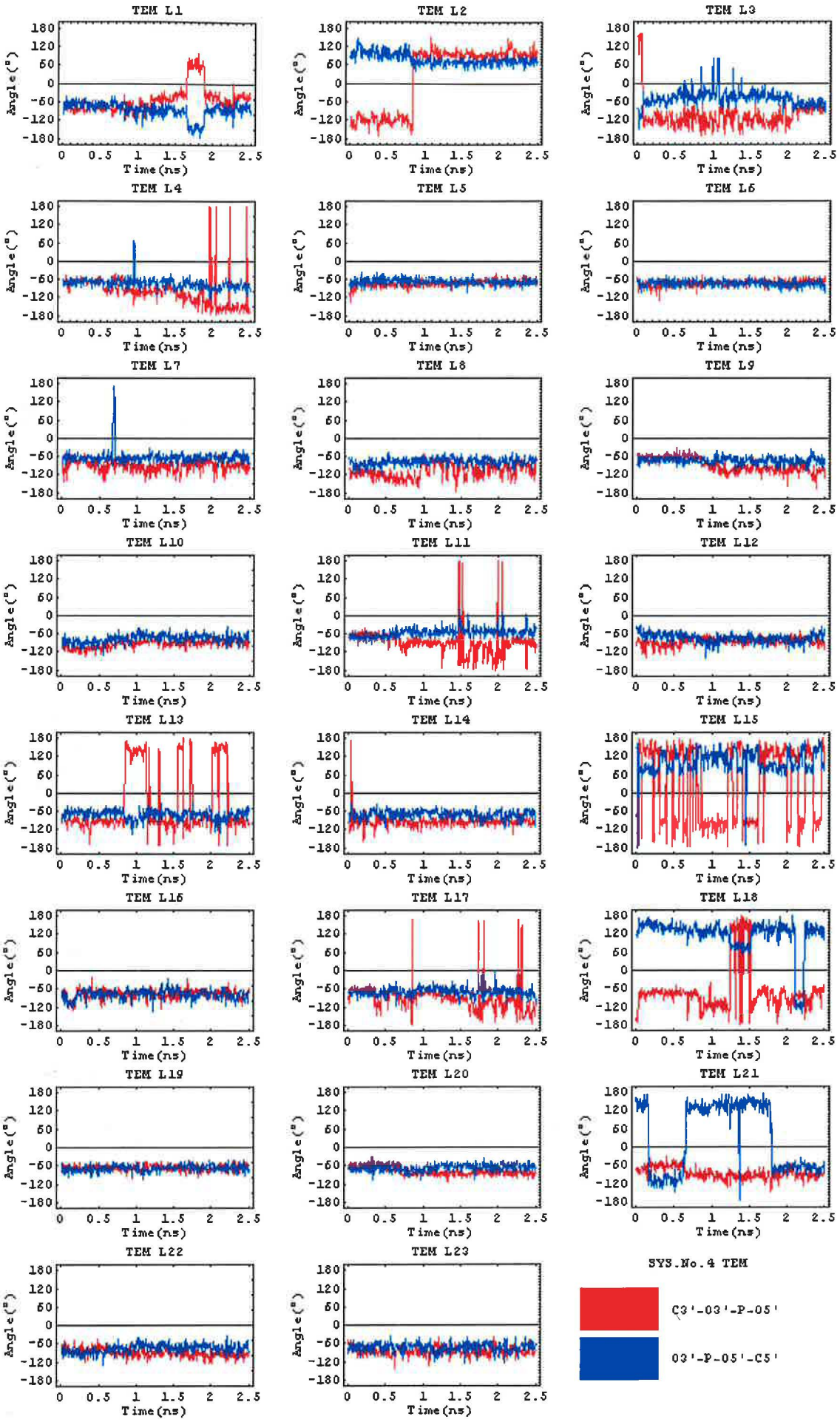


Figure F.7.: The torsion angles of the DNA template backbone in the simulation system no.4.

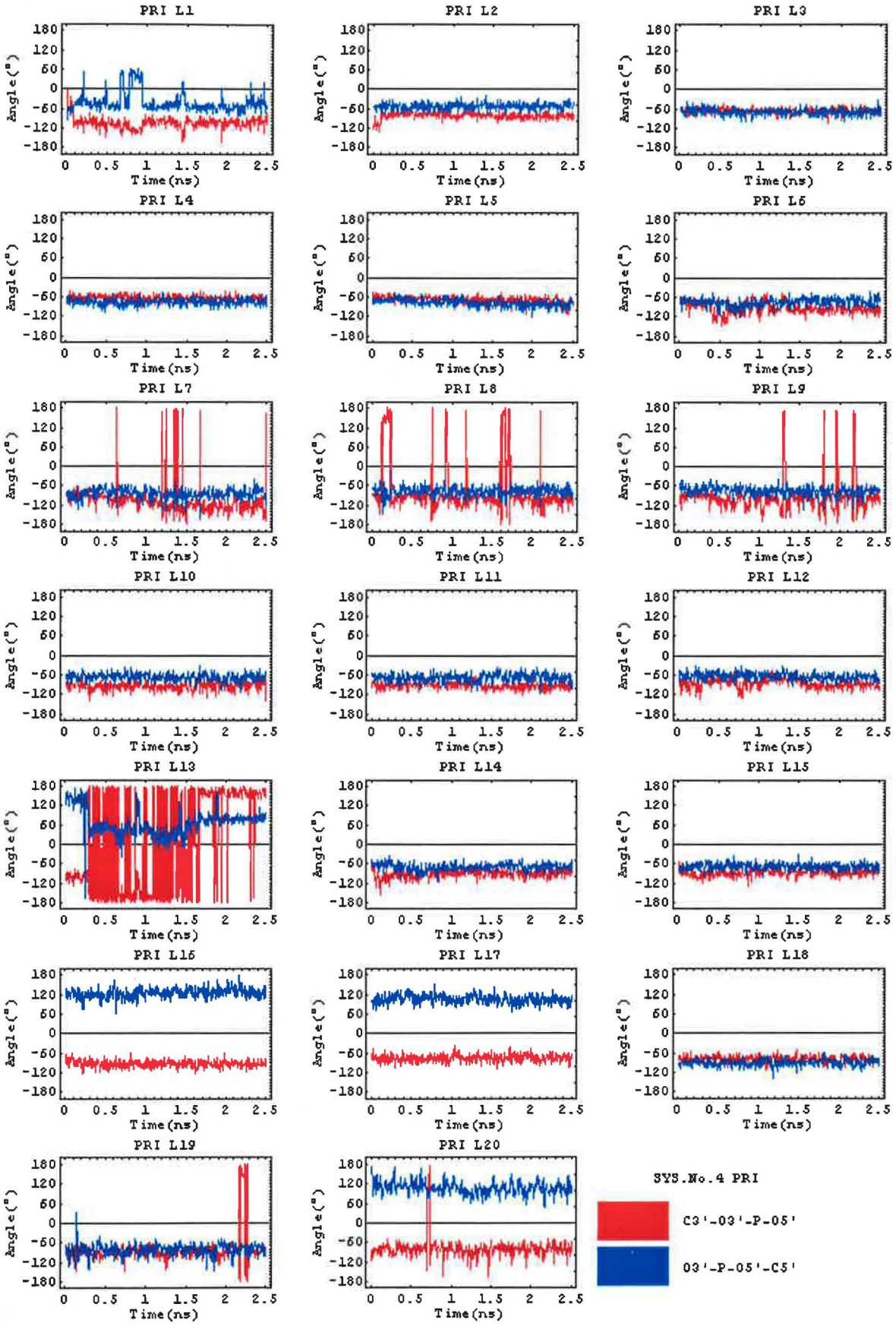


Figure F.8.: The torsion angles of the DNA primer backbone in the system no.4.

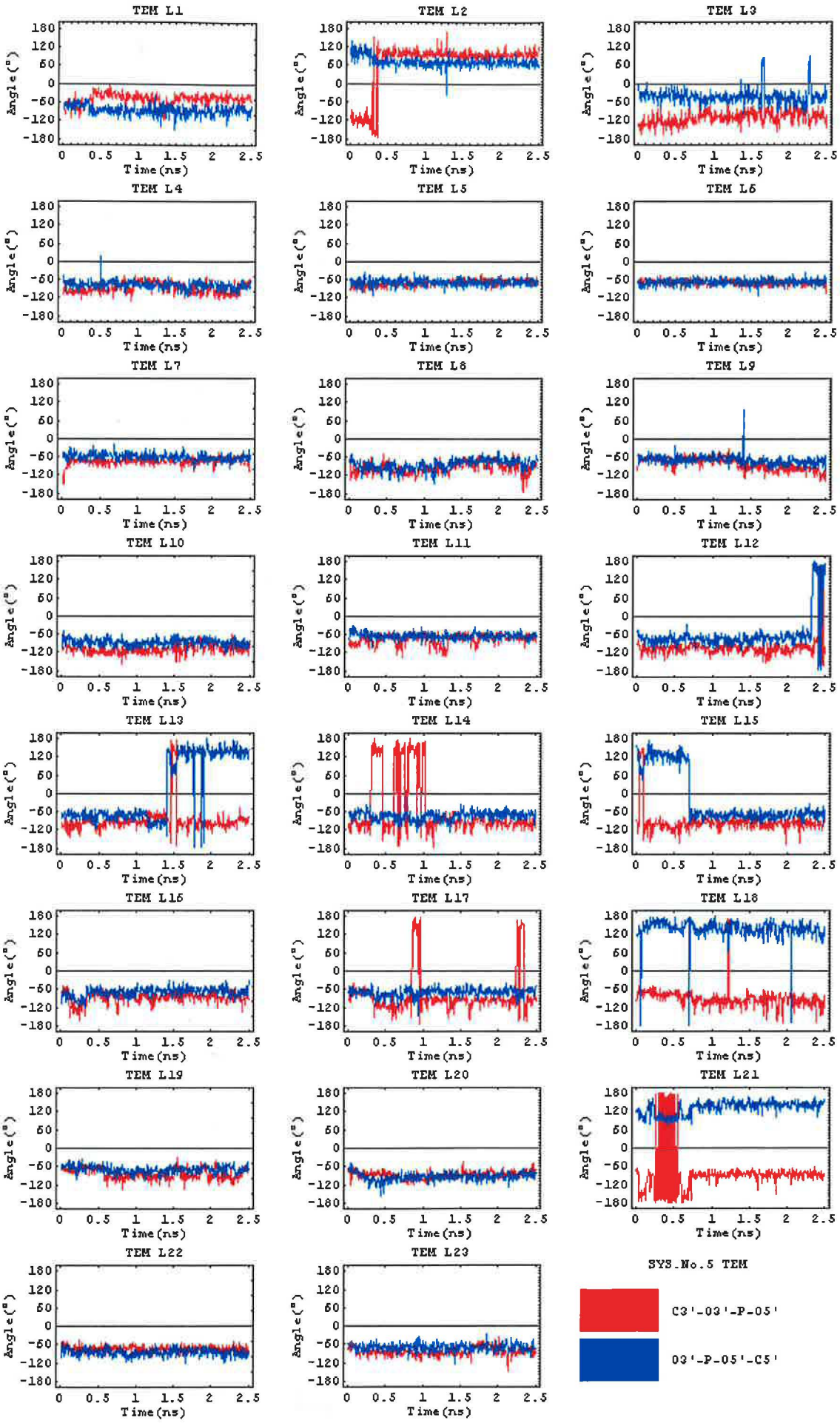


Figure F.9.: The torsion angles of the DNA template backbone in the simulation system no.5.

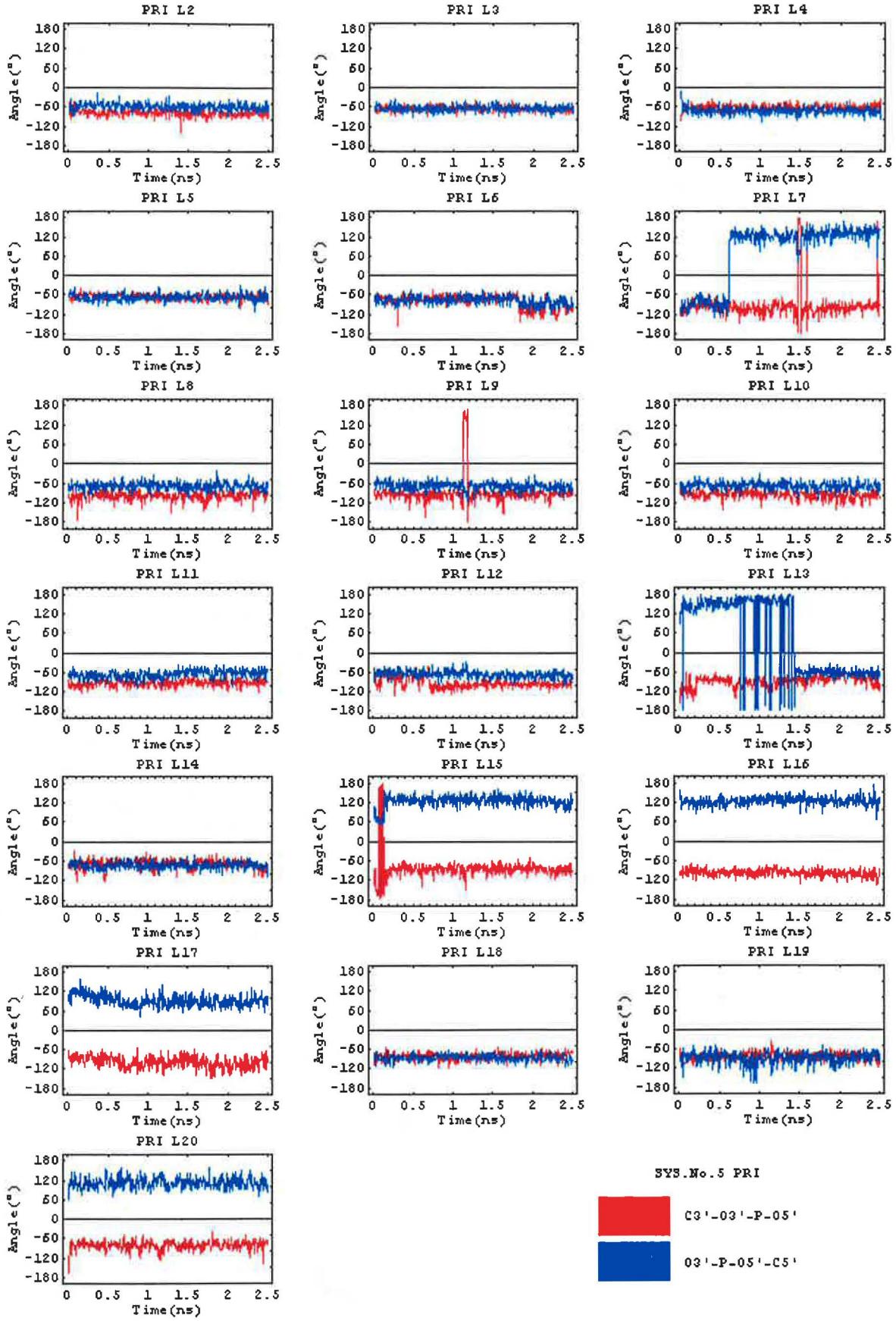


Figure F.10.: The torsion angles of the DNA primer backbone in the simulation system no.5.

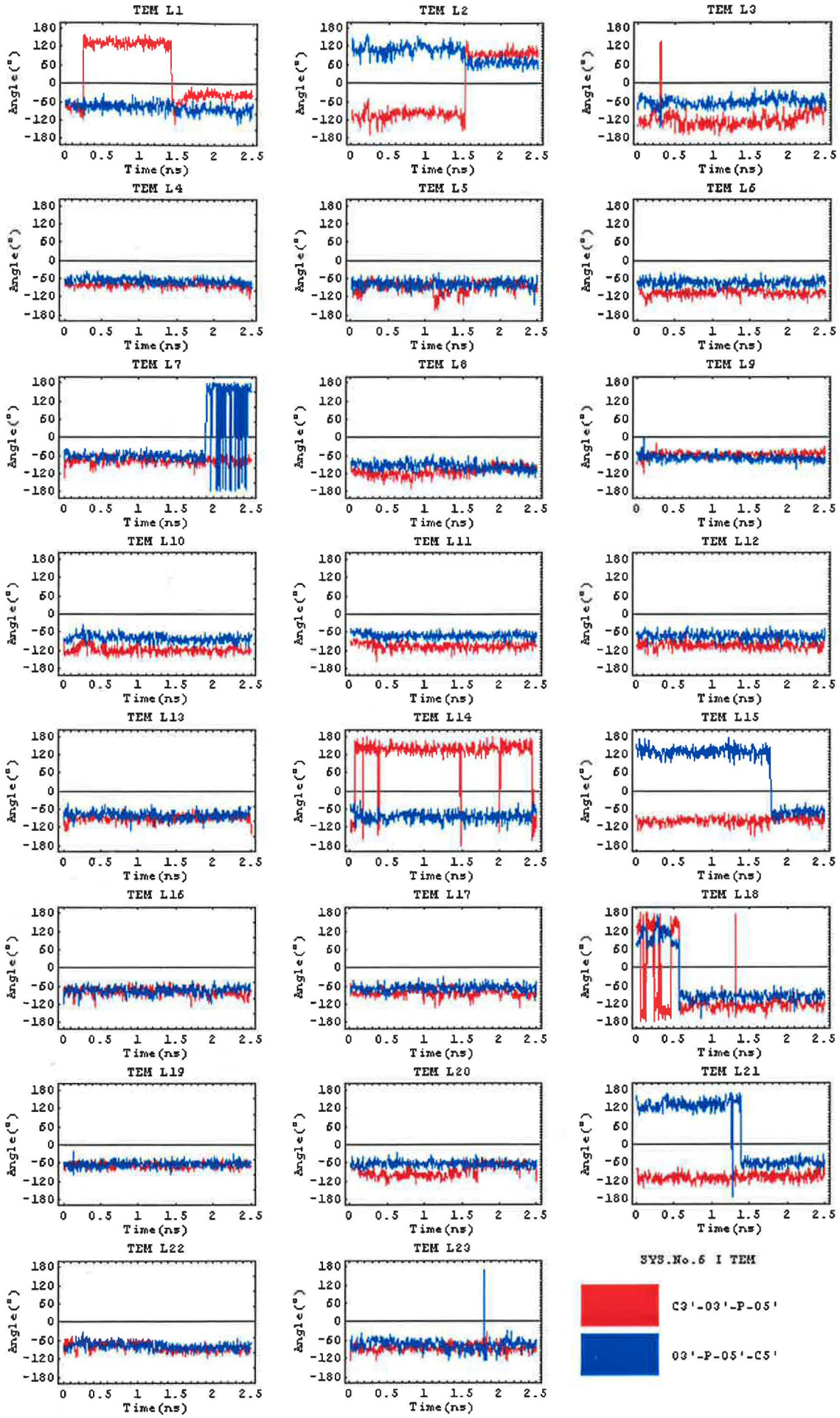


Figure F.11.: The torsion angles of the DNA template backbone in the simulation system no.6 I

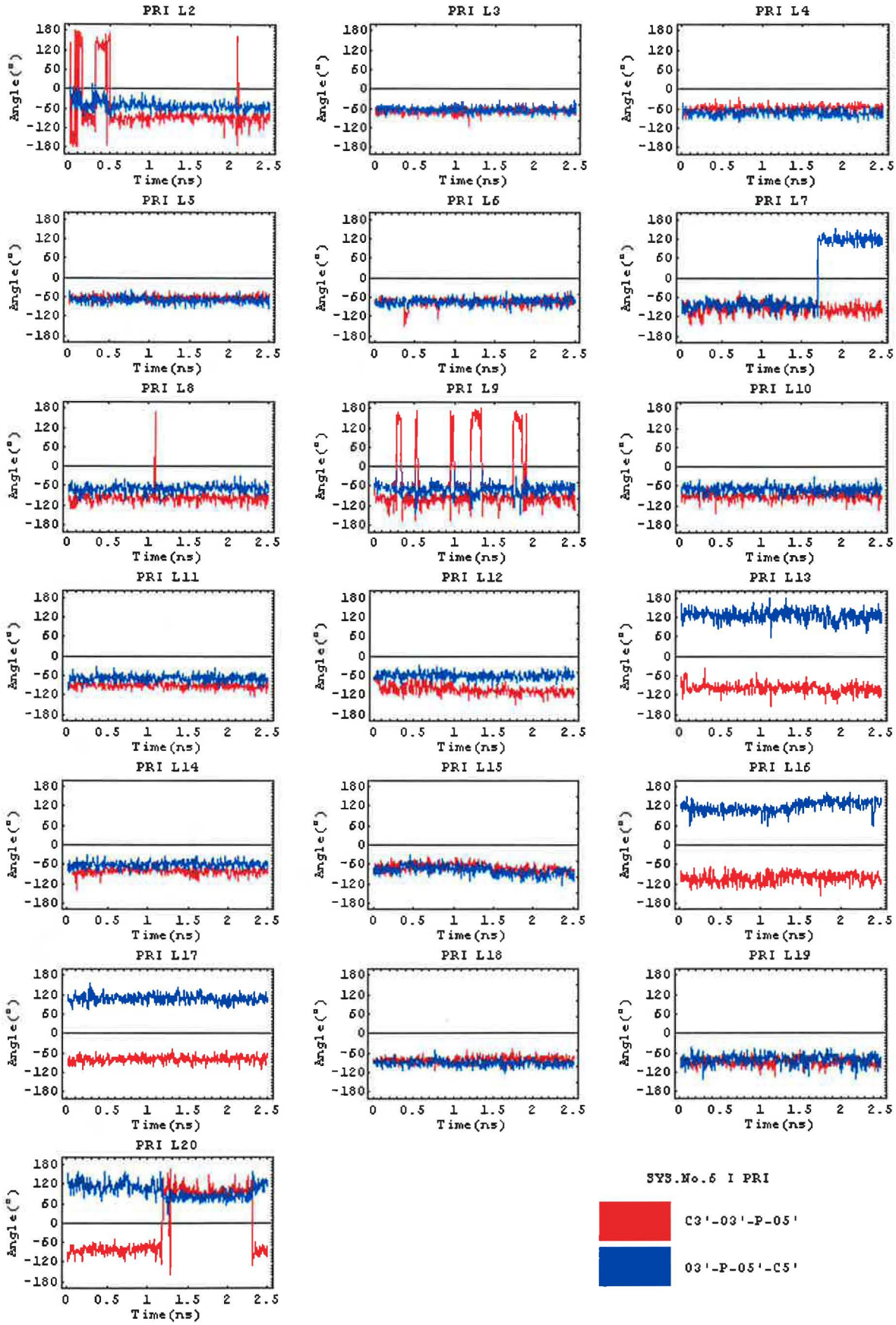


Figure F.12.: The torsion angles of the DNA primer backbone in the simulation system no.6 I.

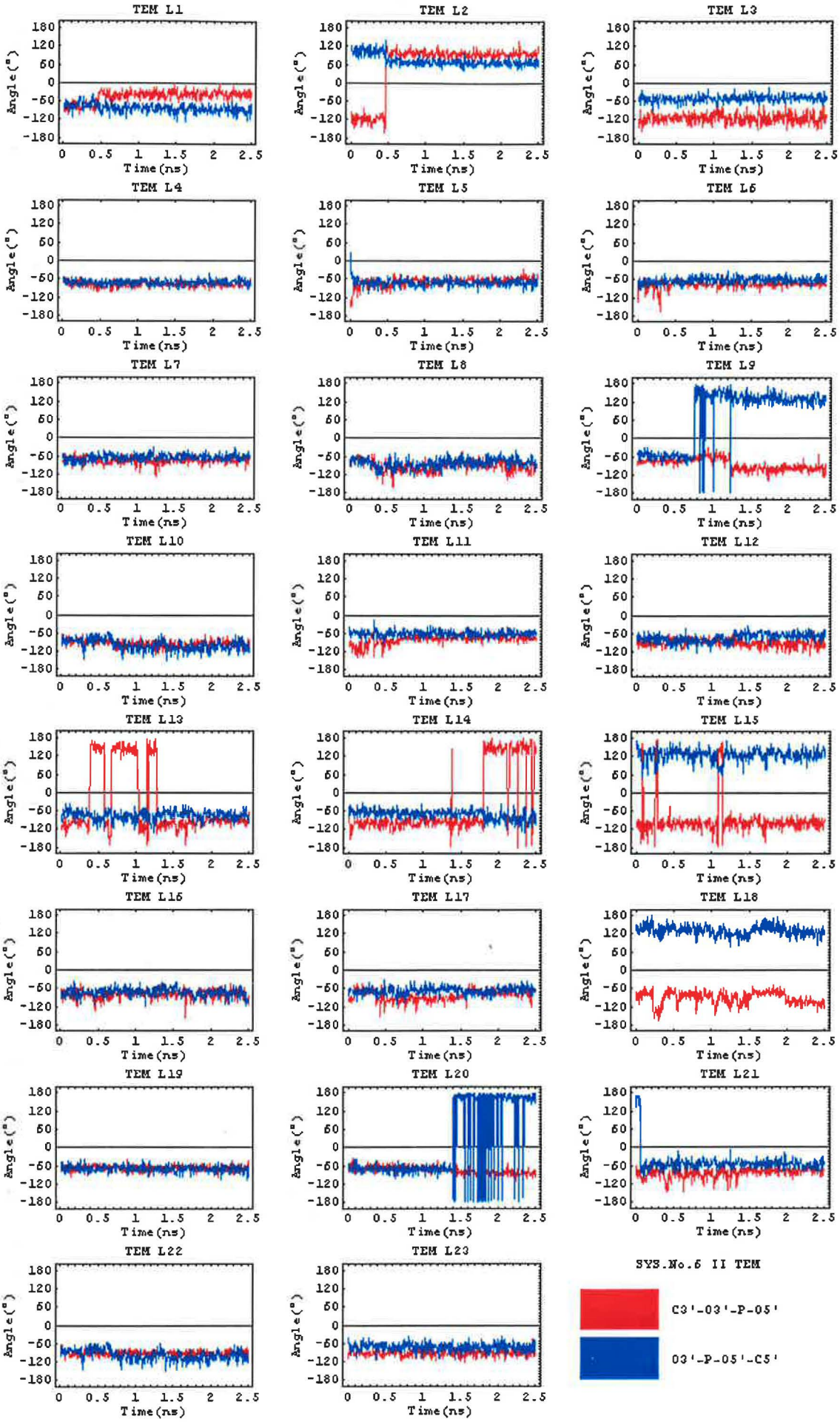


Figure F.13.: The torsion angles of the DNA template backbone in the simulation system no.6 II.

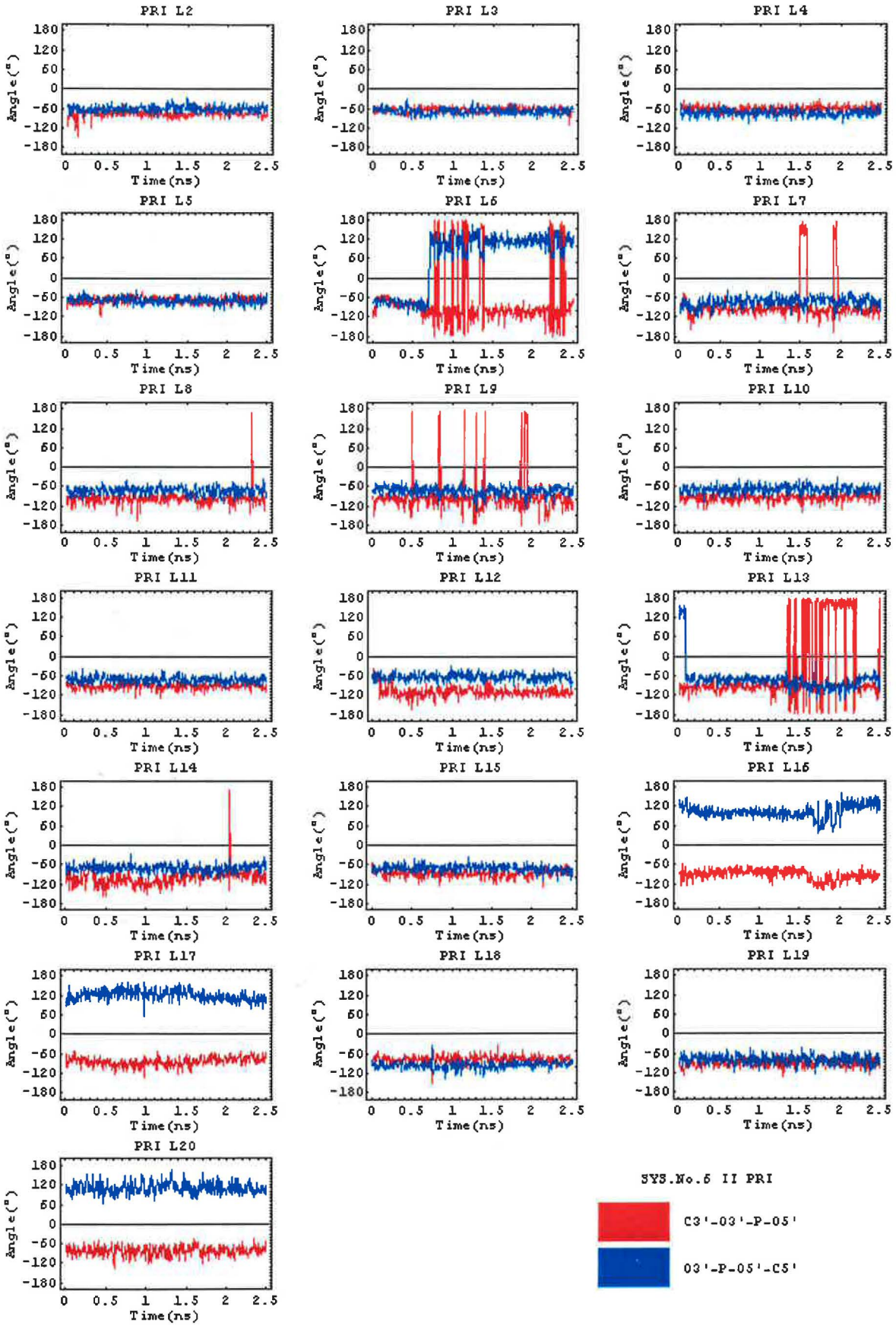


Figure F.14.: The torsion angles of the DNA primer backbone in the simulation system no.6 II.

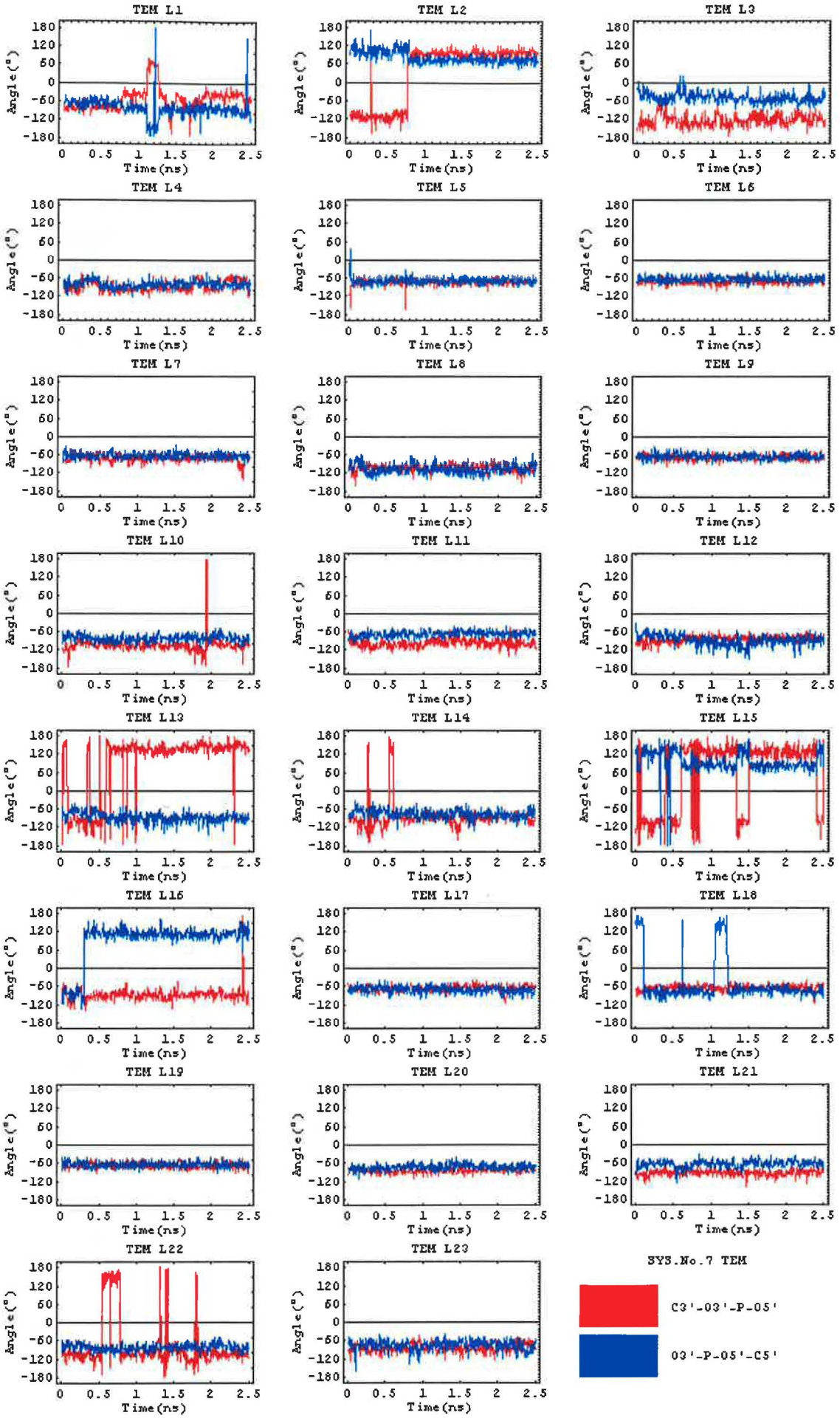


Figure F.15.: The torsion angles of the DNA template backbone in the simulation system no. 7.

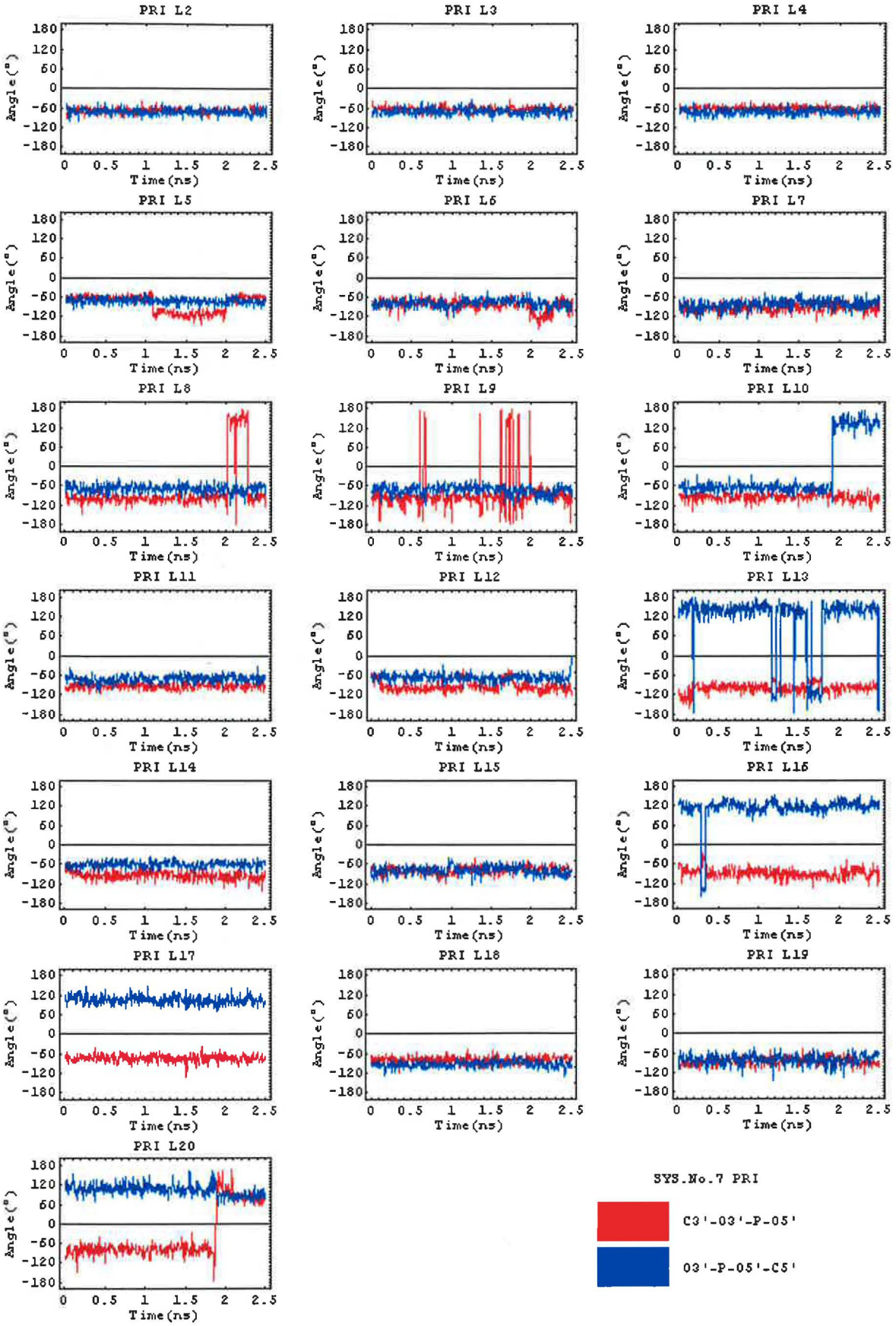


Figure F.16.: The torsion angles of the DNA primer backbone in the simulation system no.7.

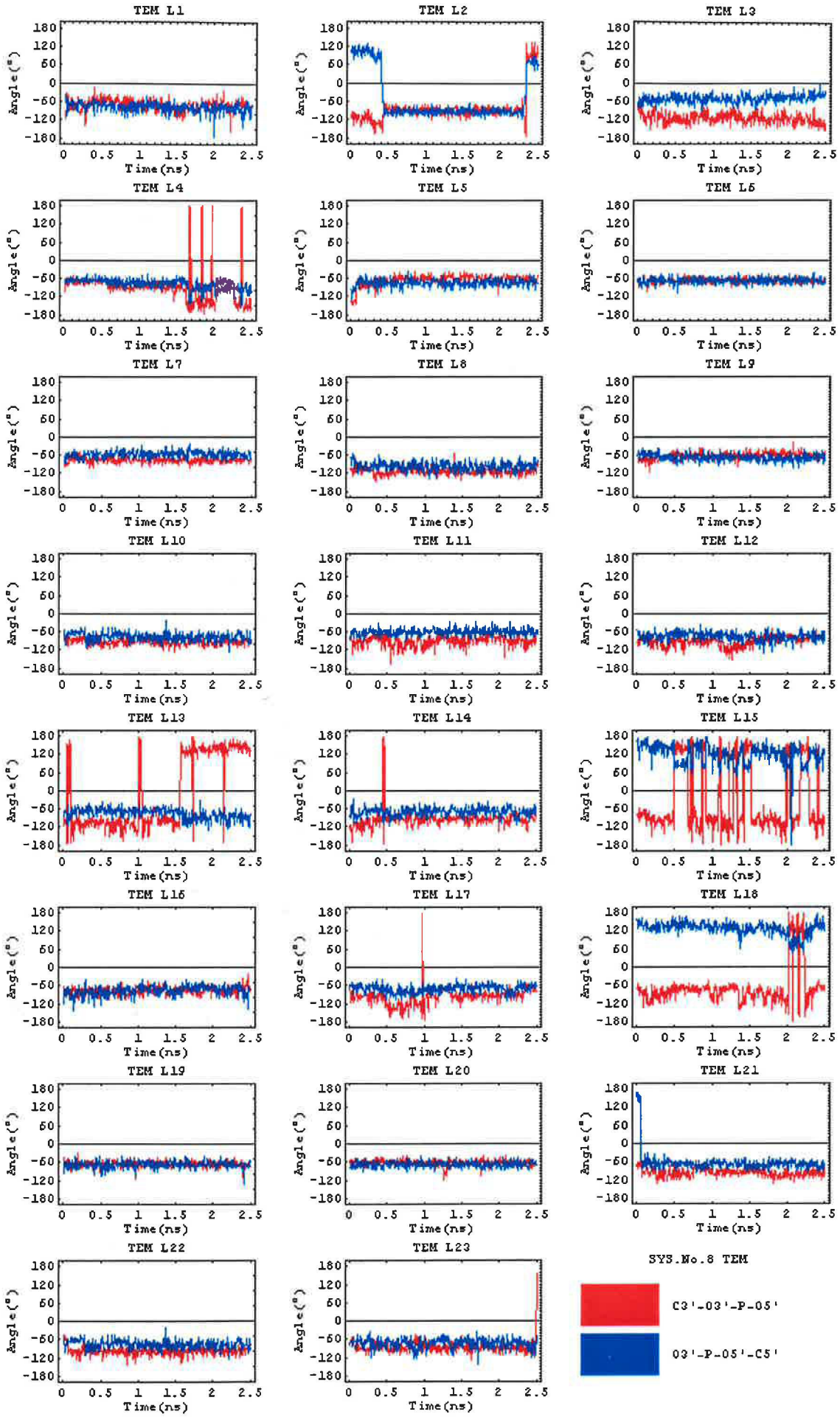


Figure F.17.: The torsion angles of the DNA template backbone in the simulatio system no. 8.

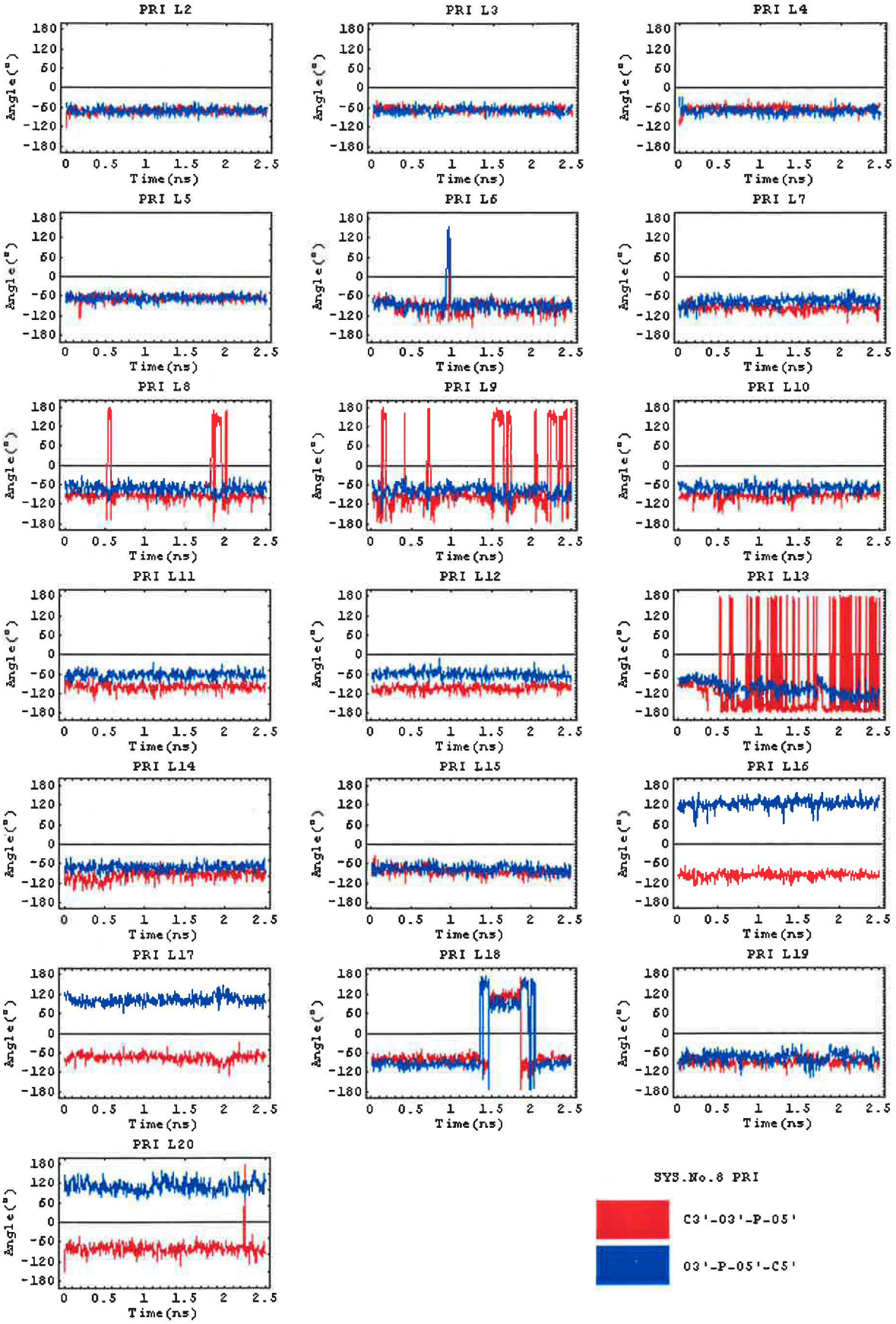
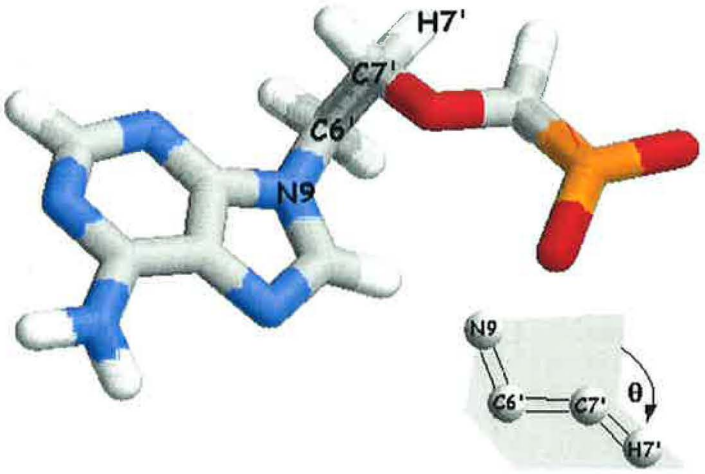


Figure F.18.: Torsion angles of the DNA primer backbone in the simulation system no.8.

Apendix G:

Adefovir/Tenofovir conformational preferences



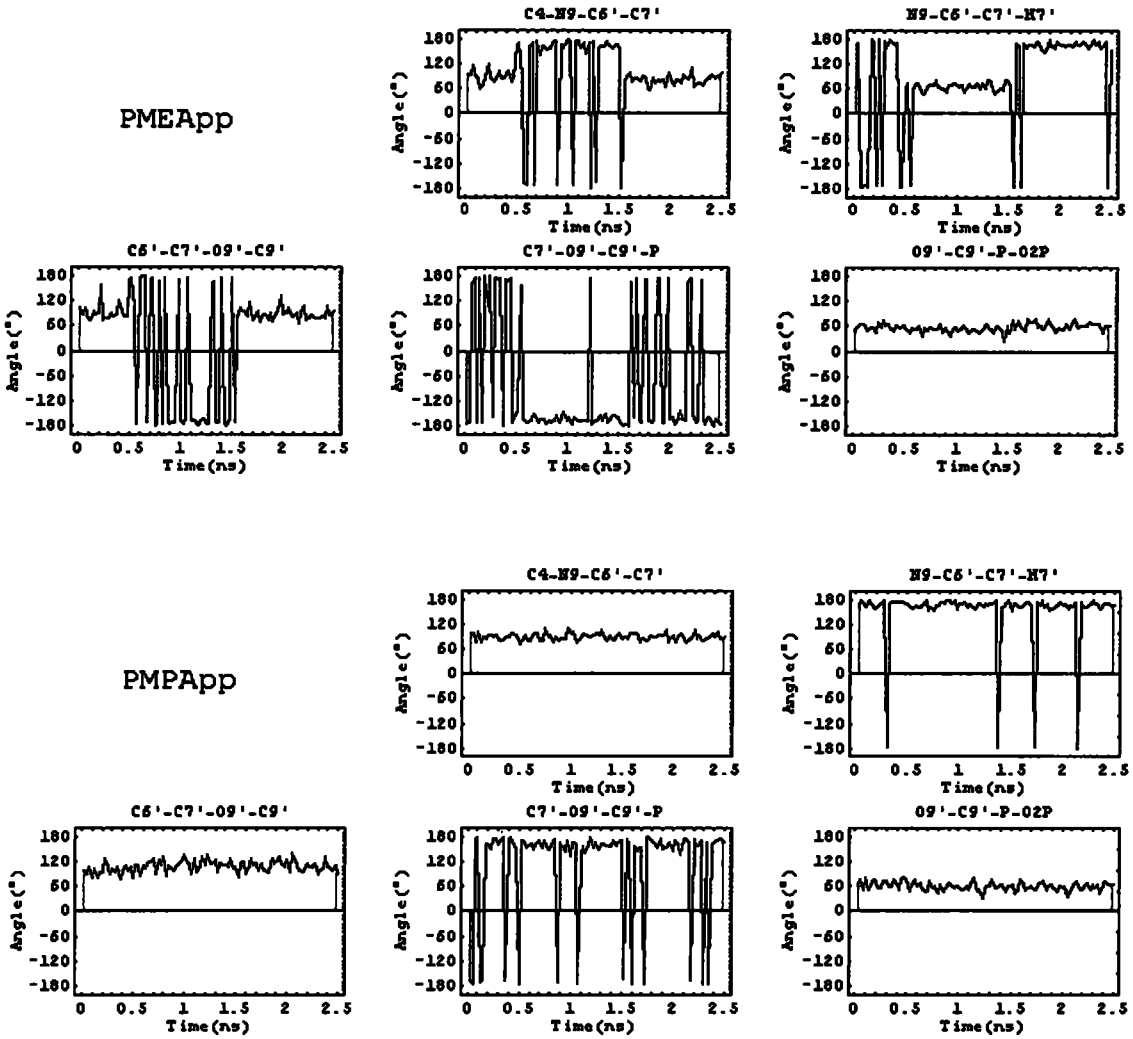


Figure G.1.: The torsion angles of Adefovir diphosphate (above) and Tenofovir diphosphate (below) during the simulation in the simulation system no.1 and 2. (inhibitors before incorporation into the primer DNA chain).

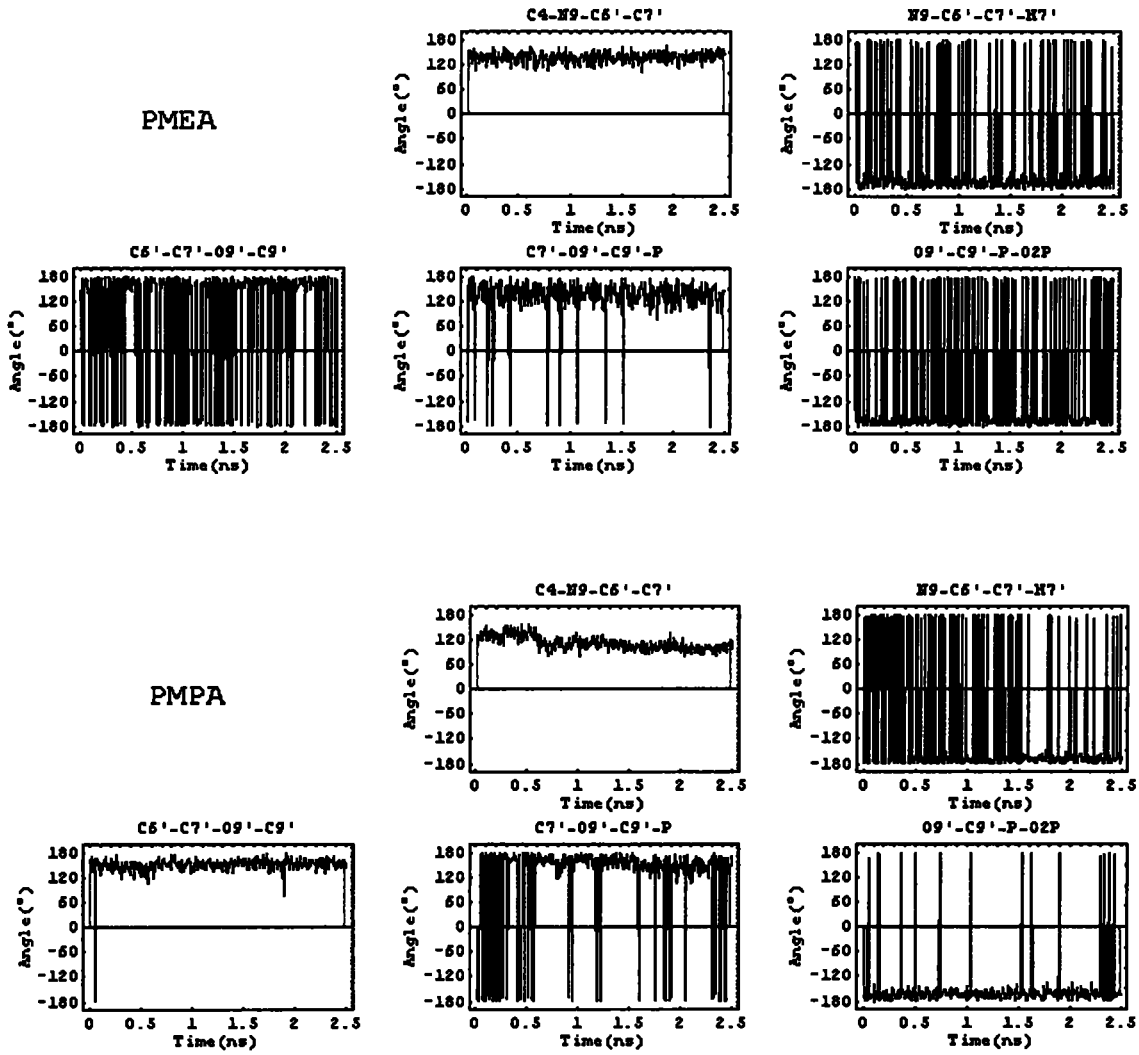
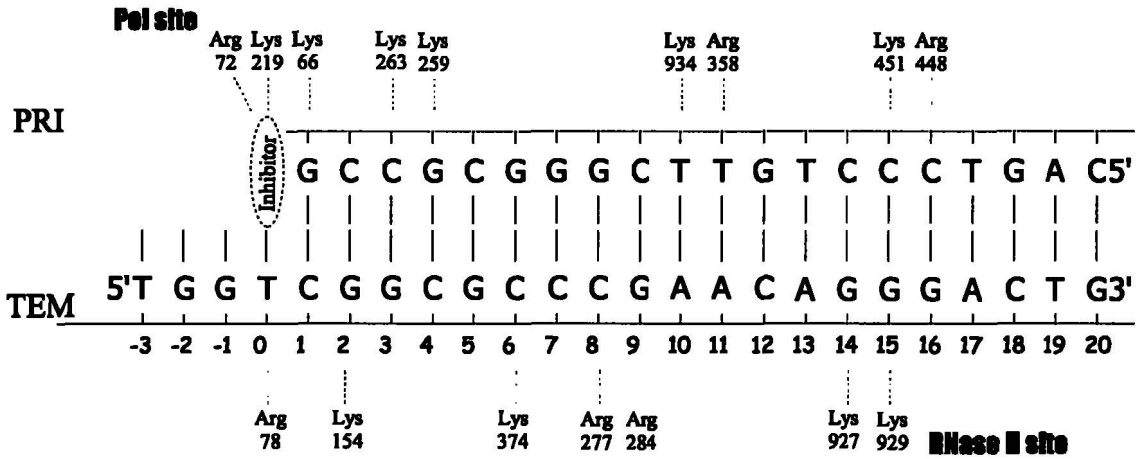


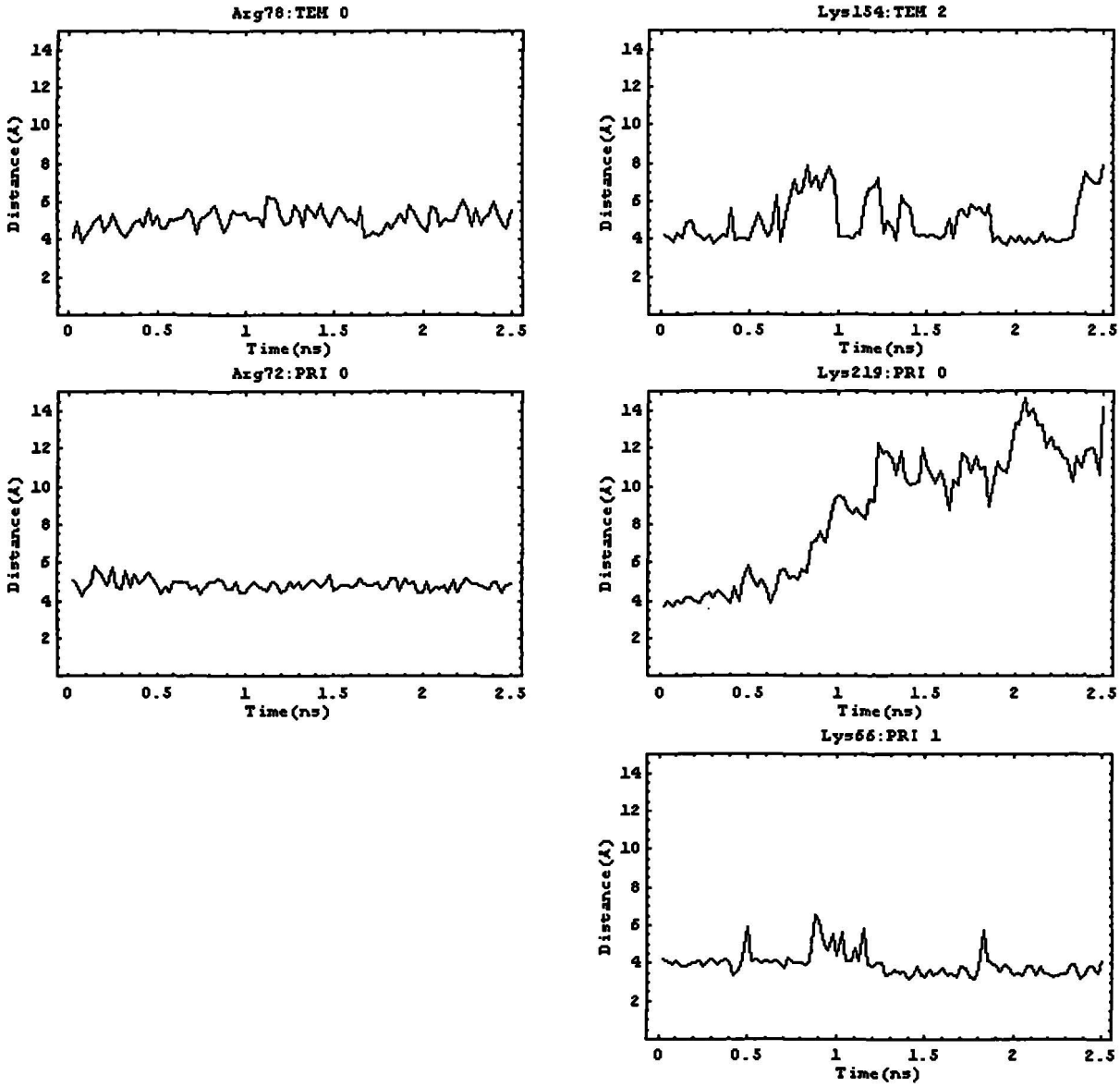
Figure G.2.: The torsion angles of Adefovir (above) and Tenofovir (below) during the simulation in the simulation system no.7 and 8, respectively (inhibitors after incorporation into the primer DNA chain).

Apendix H

HIV-1 RT / DNA:DNA mutual interactions facilitated by ARG and LYS residues



System no.1: HIV-1 RT + DNA:DNA + PMEApp (before incorporation into primer DNA)



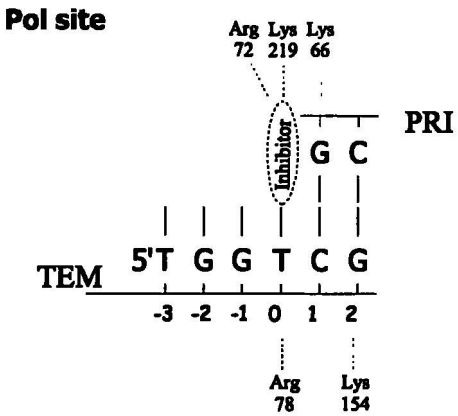
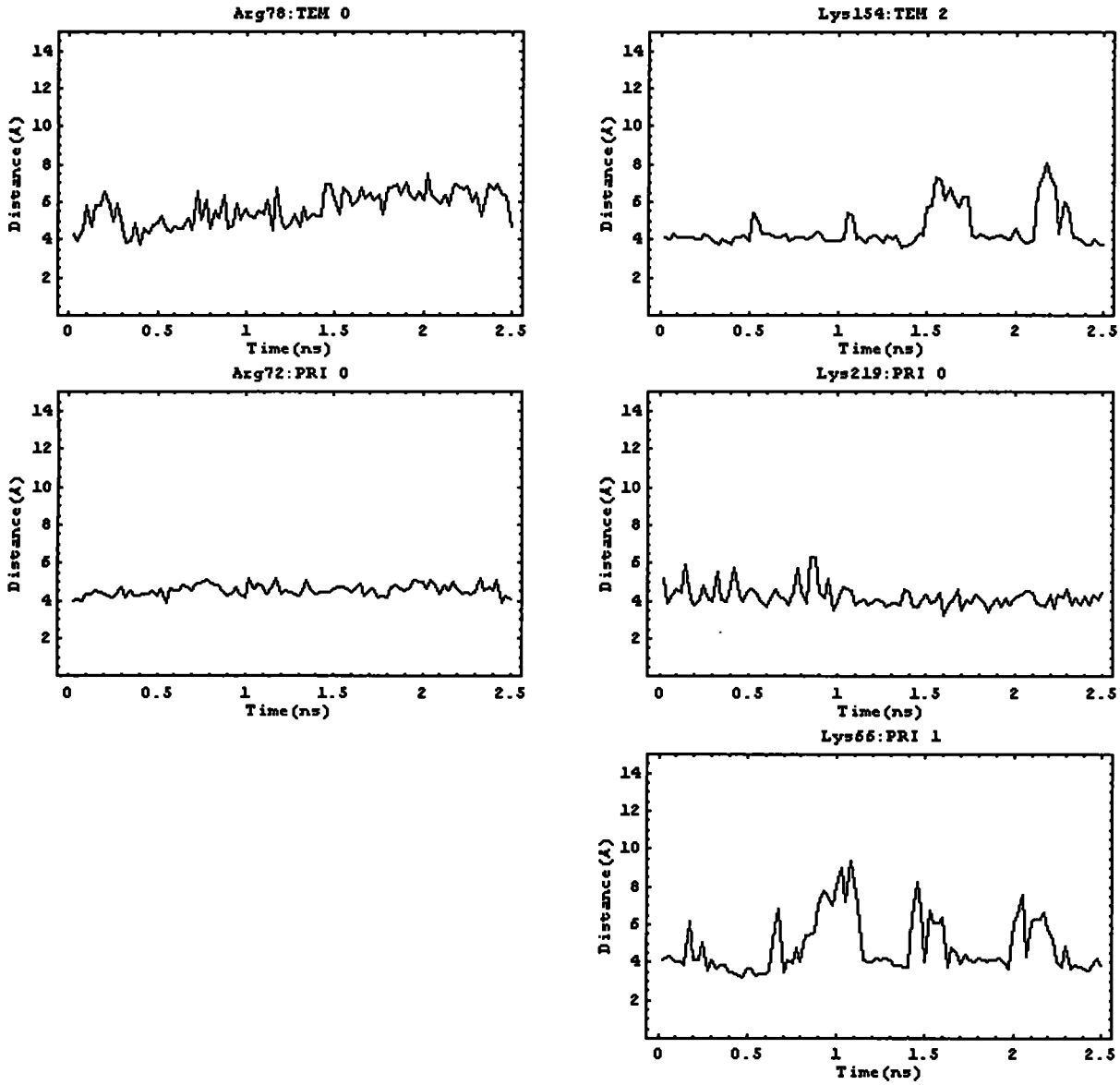


Figure H.1.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.2: HIV-1 RT + DNA:DNA + PMPApp (before incorporation into primer DNA)



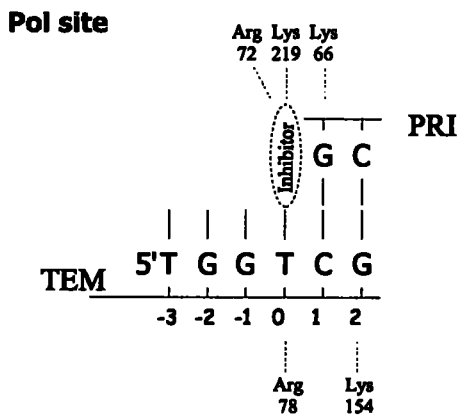
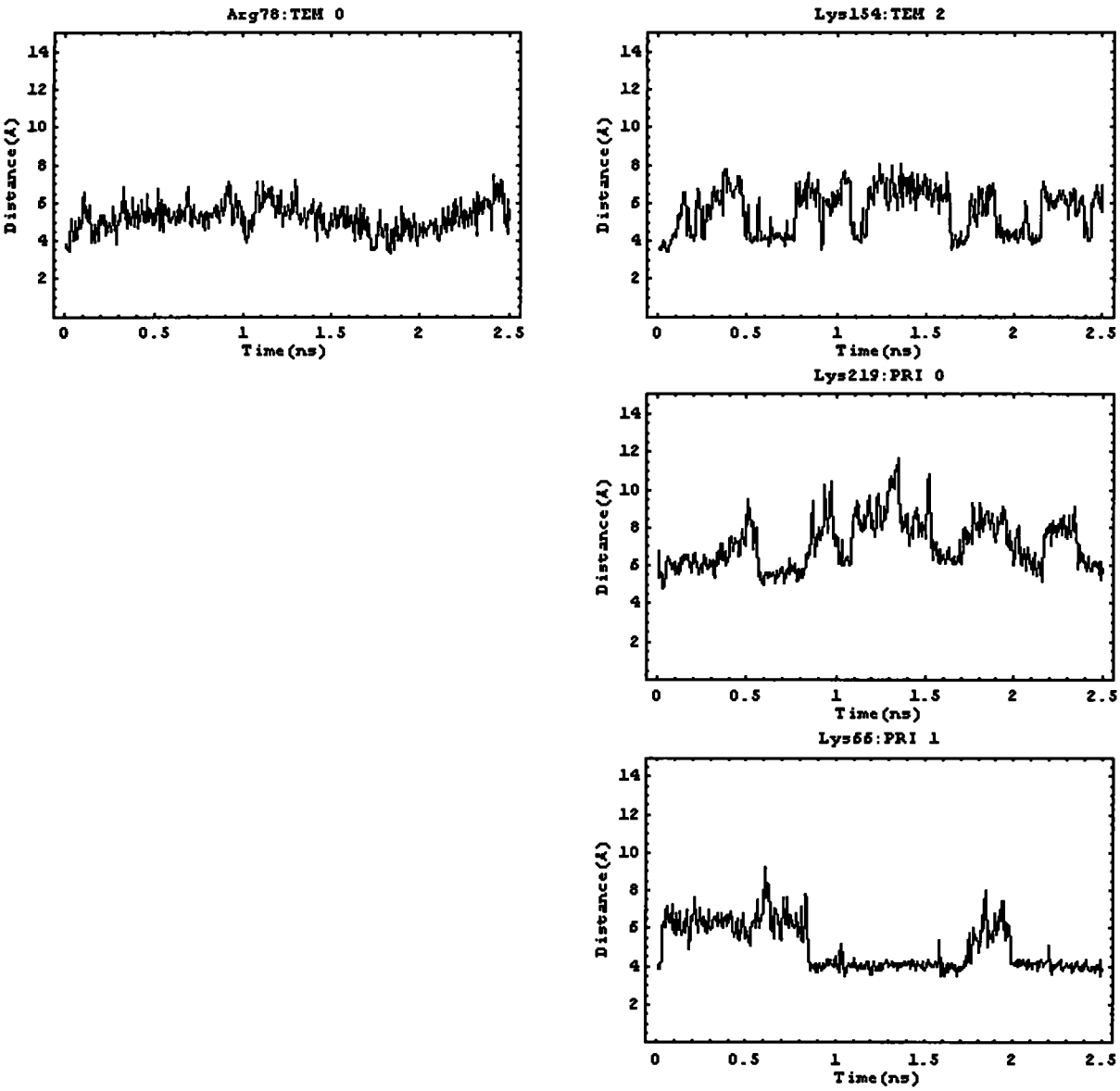


Figure H.2.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.3: HIV-1 RT + DNA:DNA + 3' Terminal Deoxyadenosine



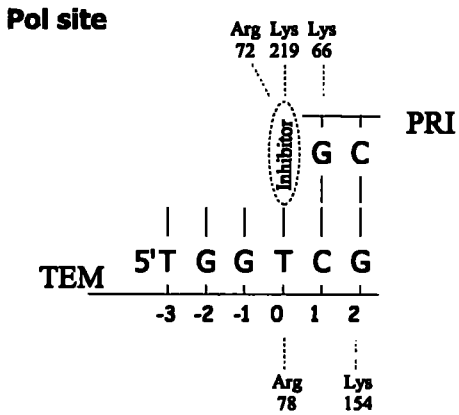
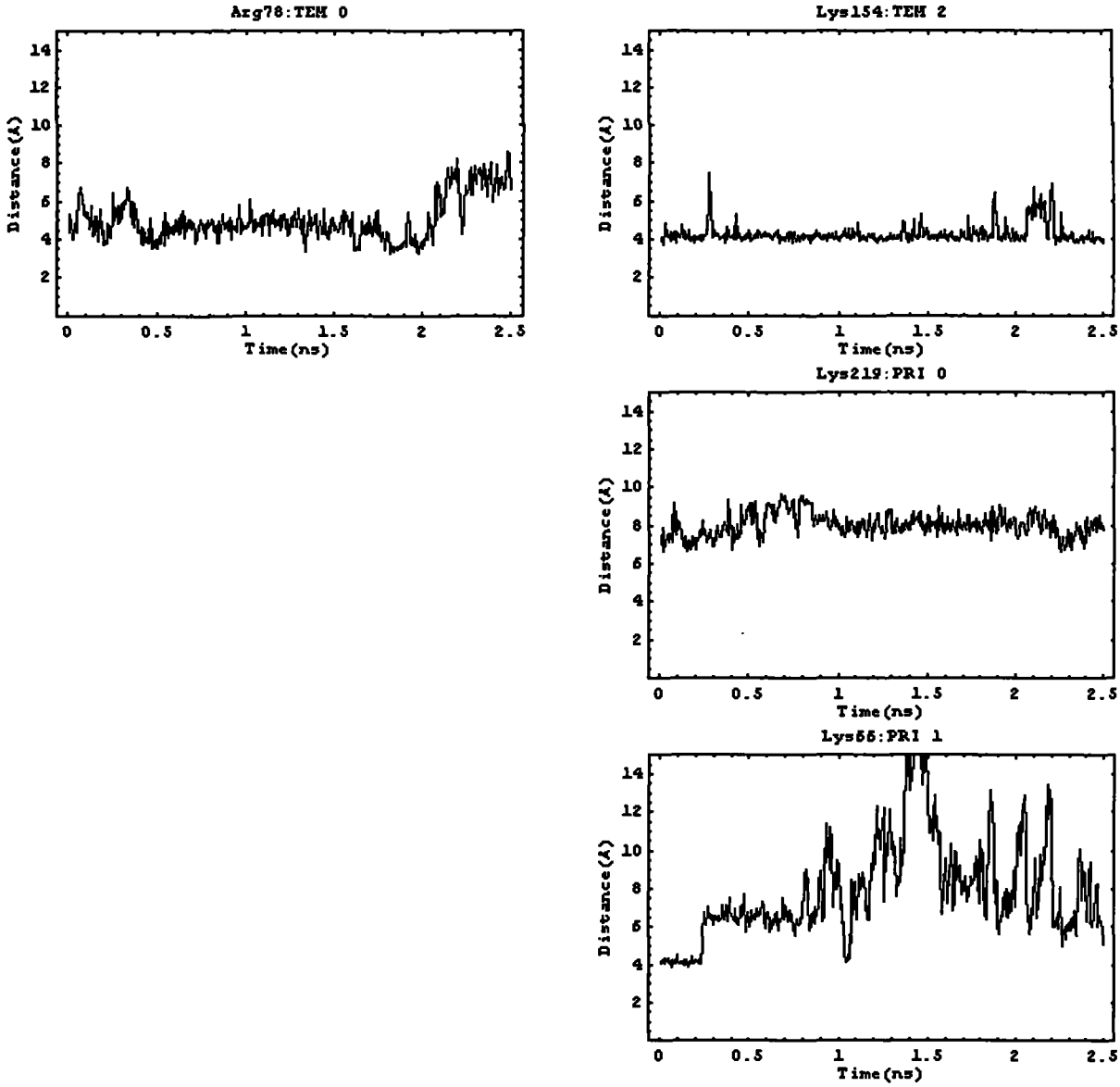


Figure H.3.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.4: HIV-1 RT + DNA:DNA + 3' Terminal Adenosine



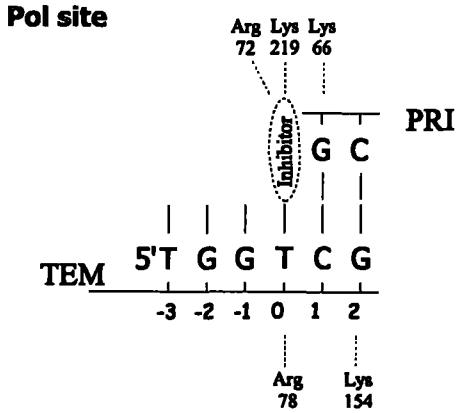
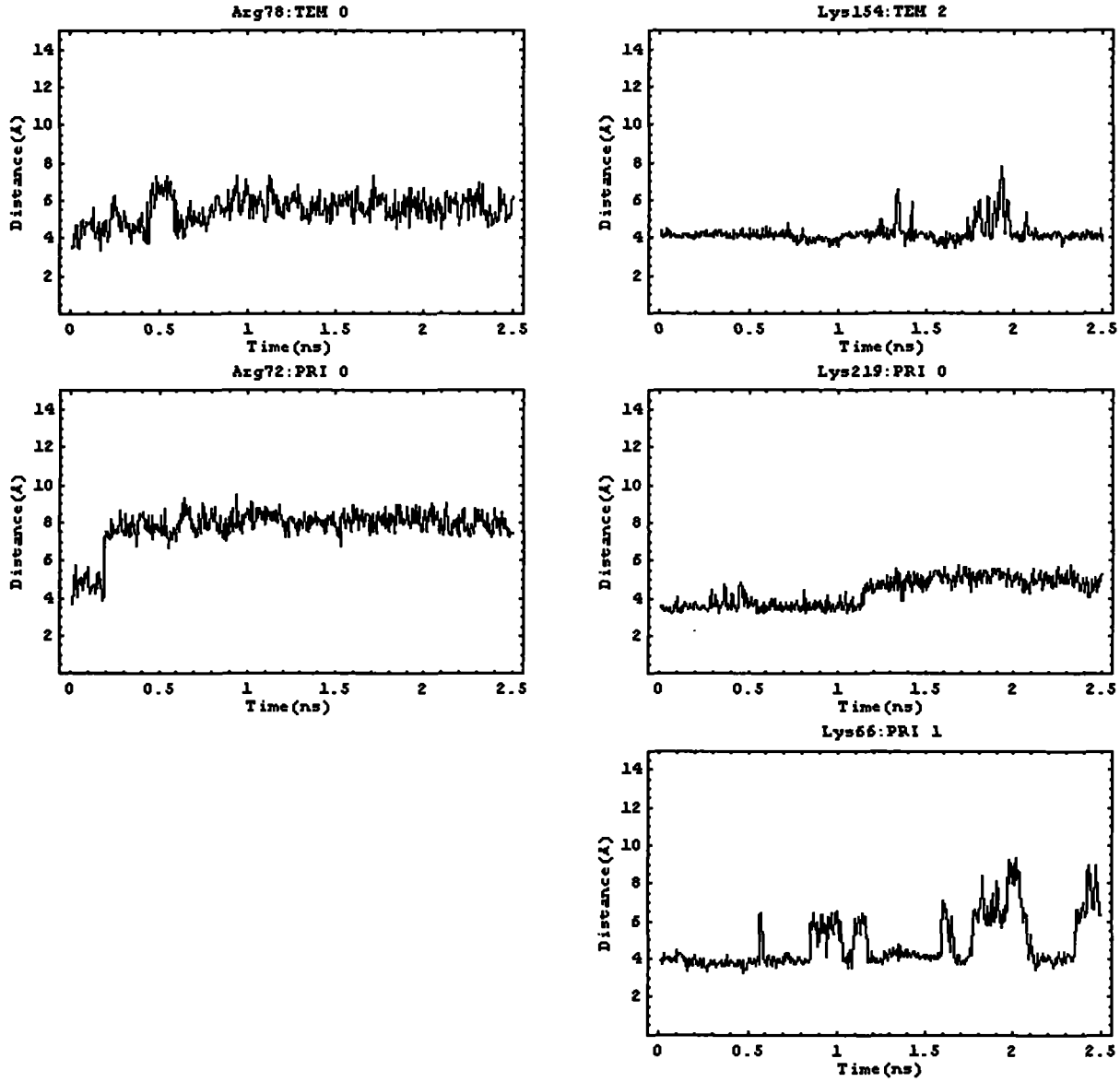


Figure H.4.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.5: HIV-1 RT + DNA:DNA + dATP



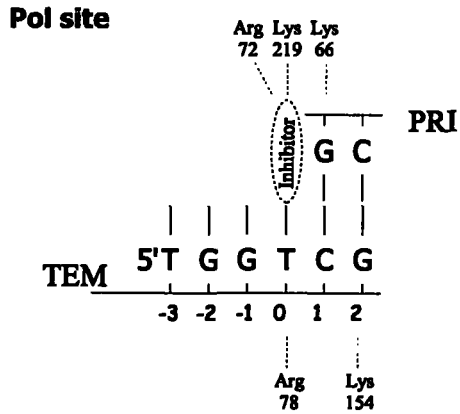
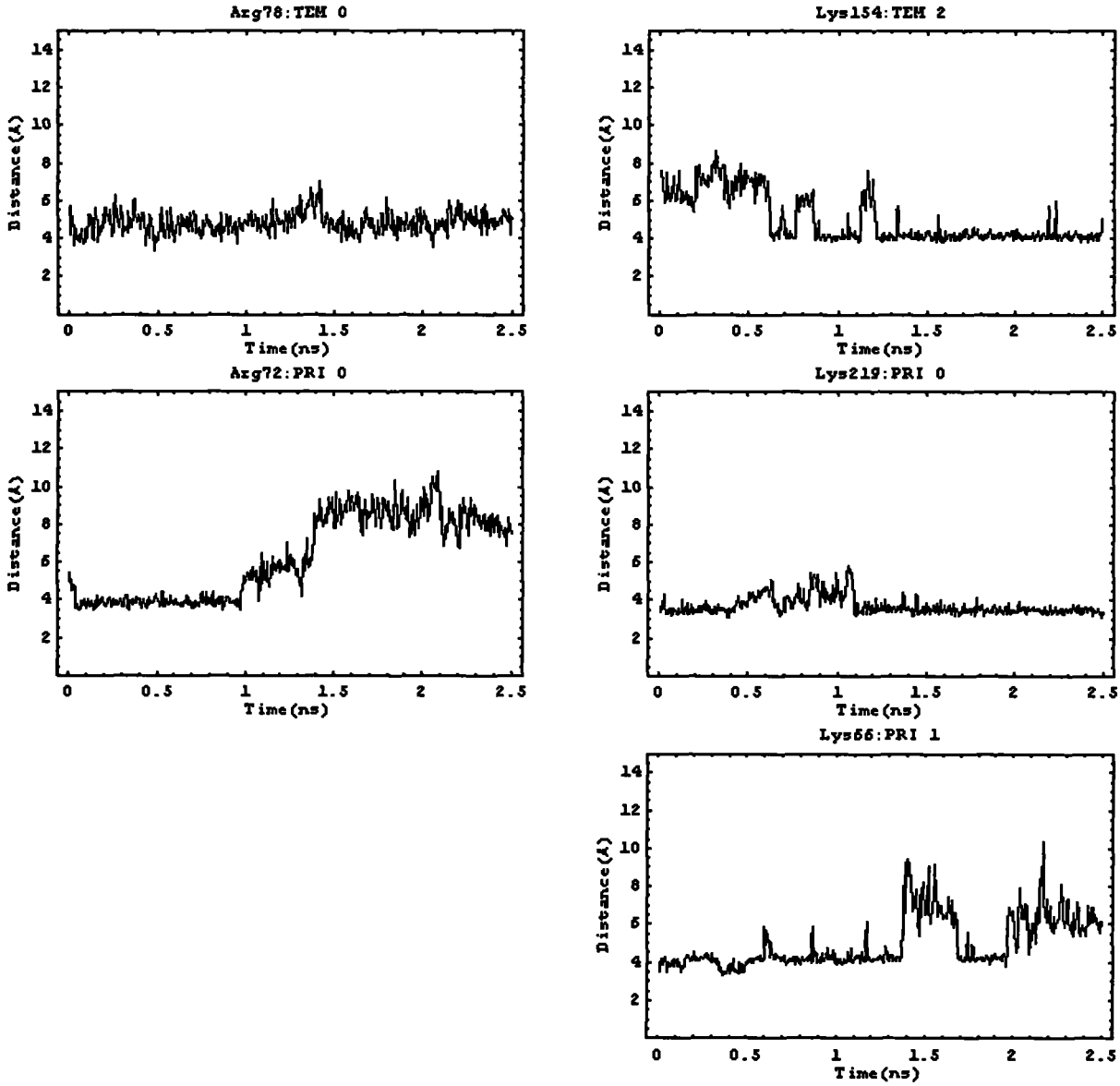


Figure H.5.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.6 I: HIV-1 RT + DNA:DNA + ATP



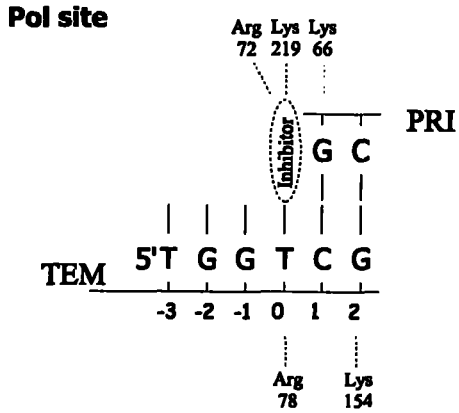
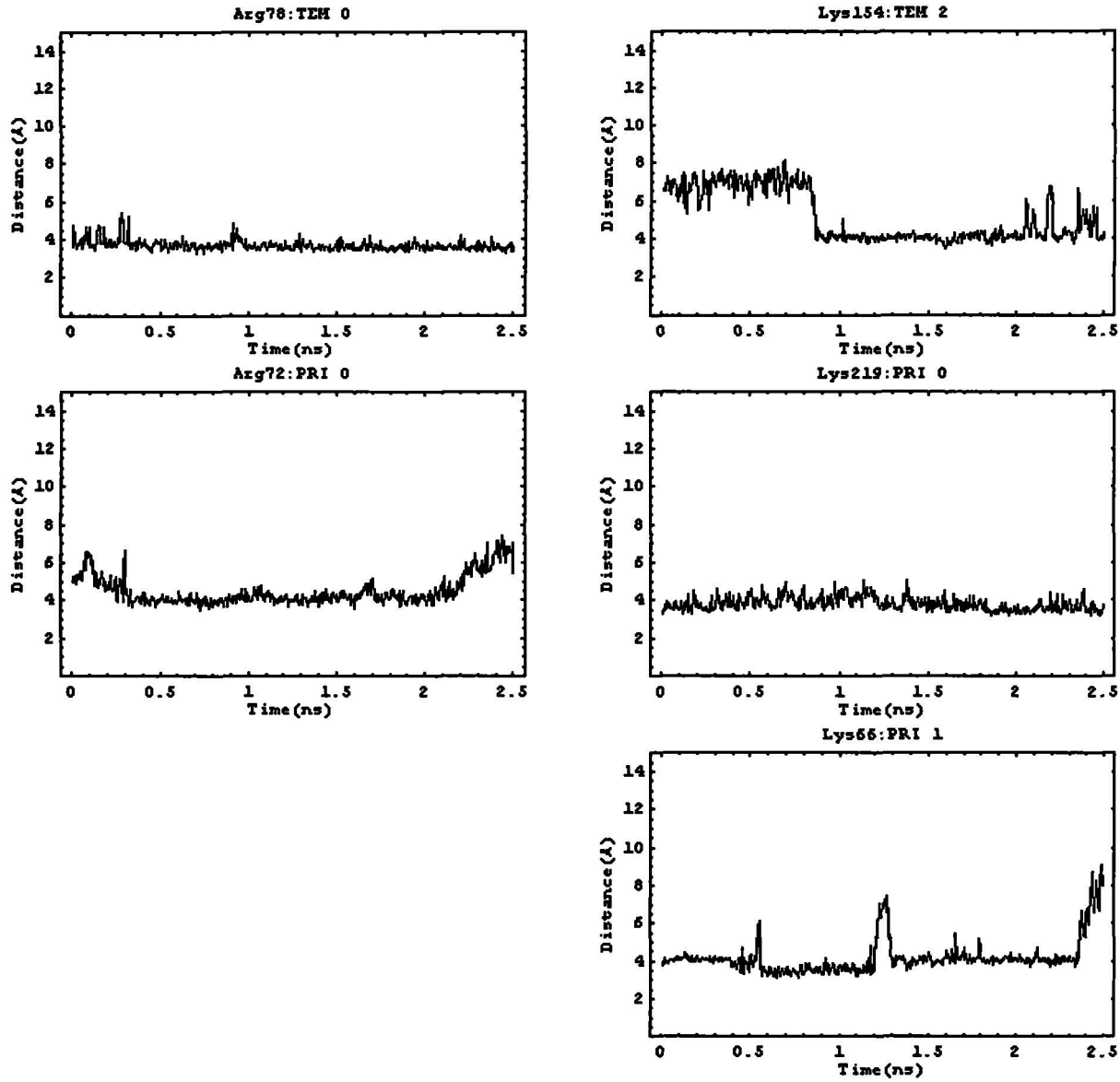


Figure H.6.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.



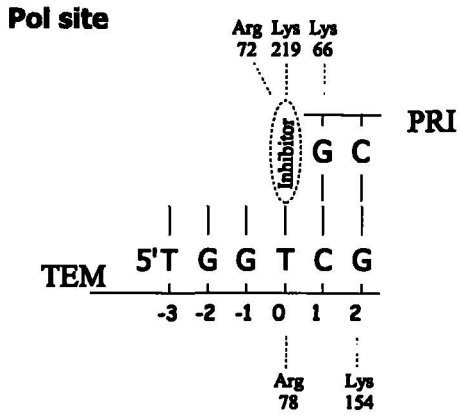
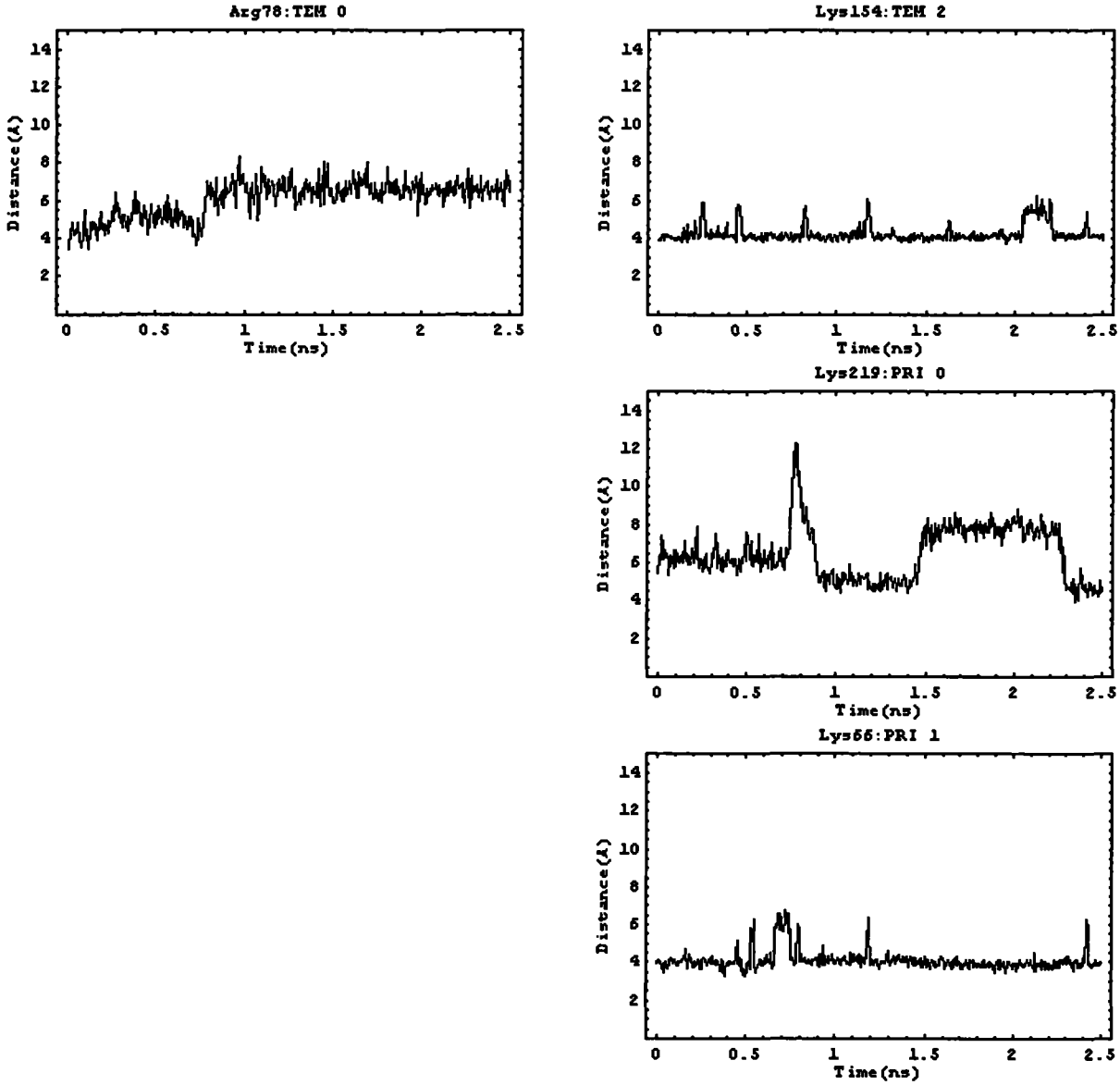


Figure H.7.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.7: HIV-1 RT + DNA:DNA + PMEA (after incorporation into primer DNA)



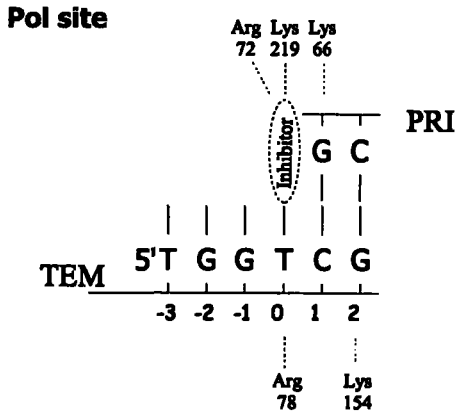
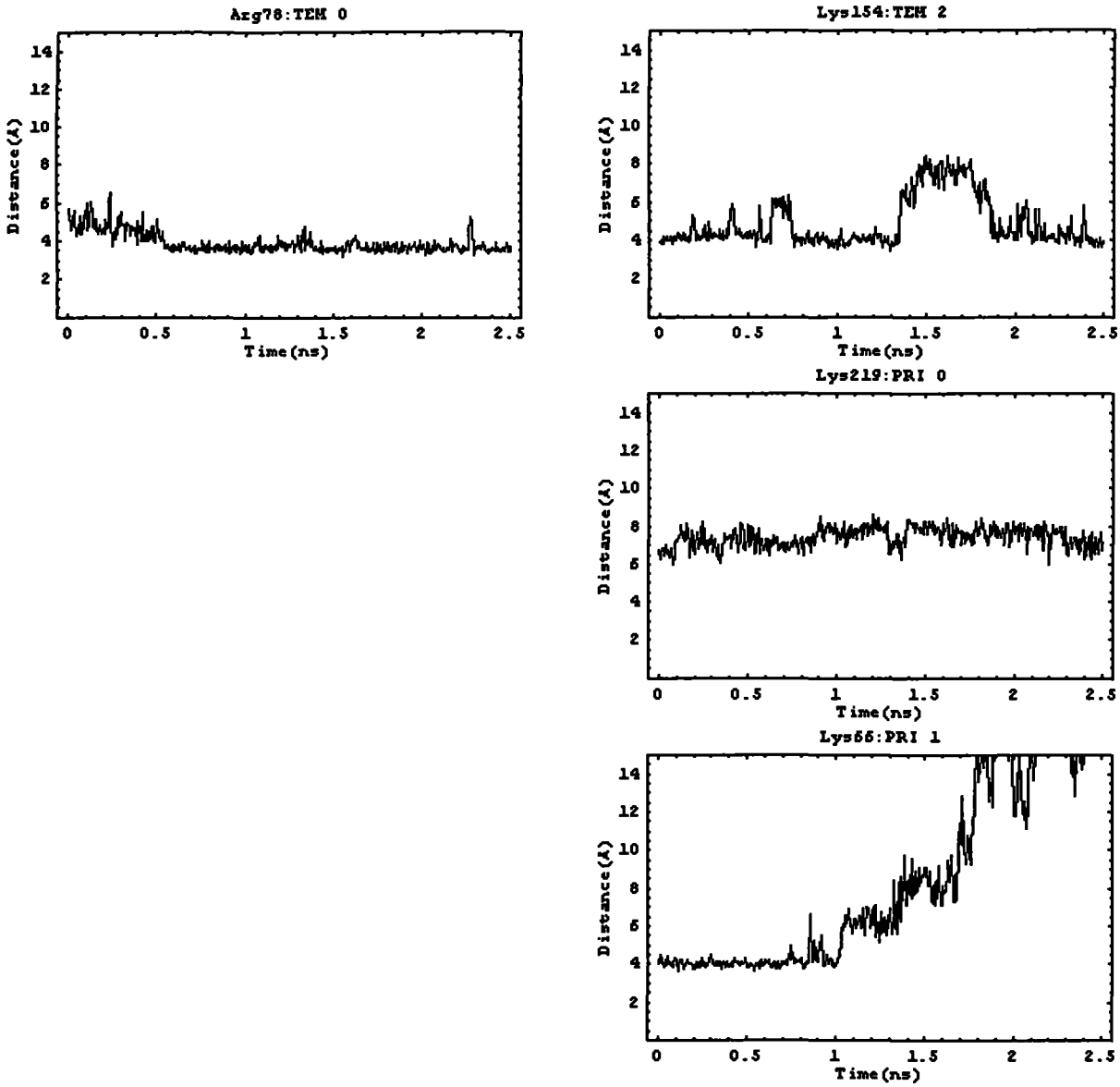


Figure H.8.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

System no.8: HIV-1 RT + DNA:DNA + PMPA (after incorporation into primer DNA)



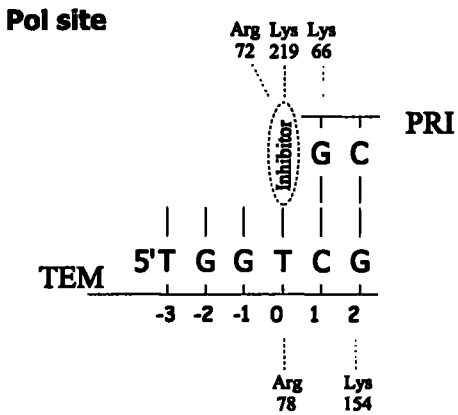


Figure H.9.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues in the HIV-1 RT polymerase active site.

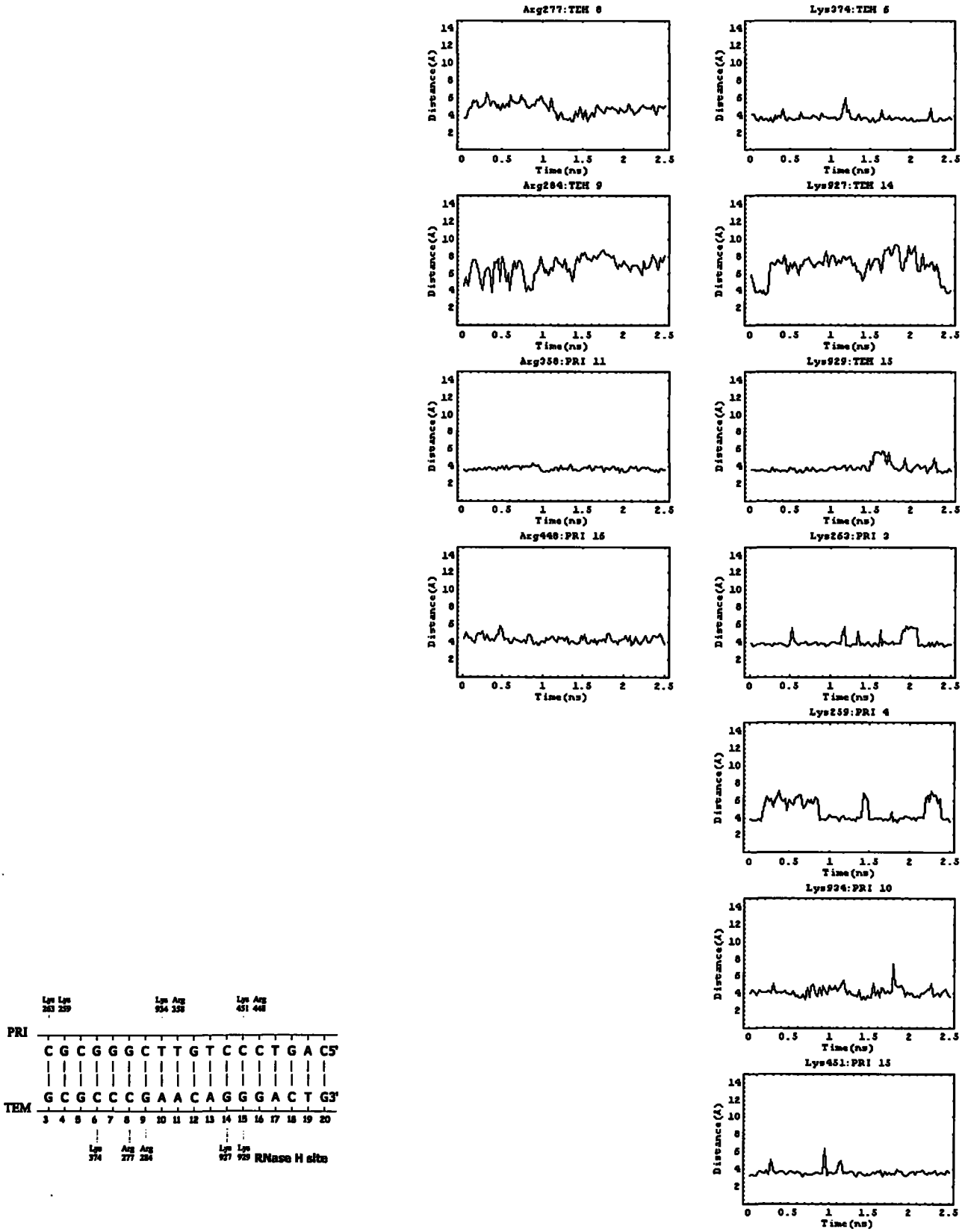


Figure H.10.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues.

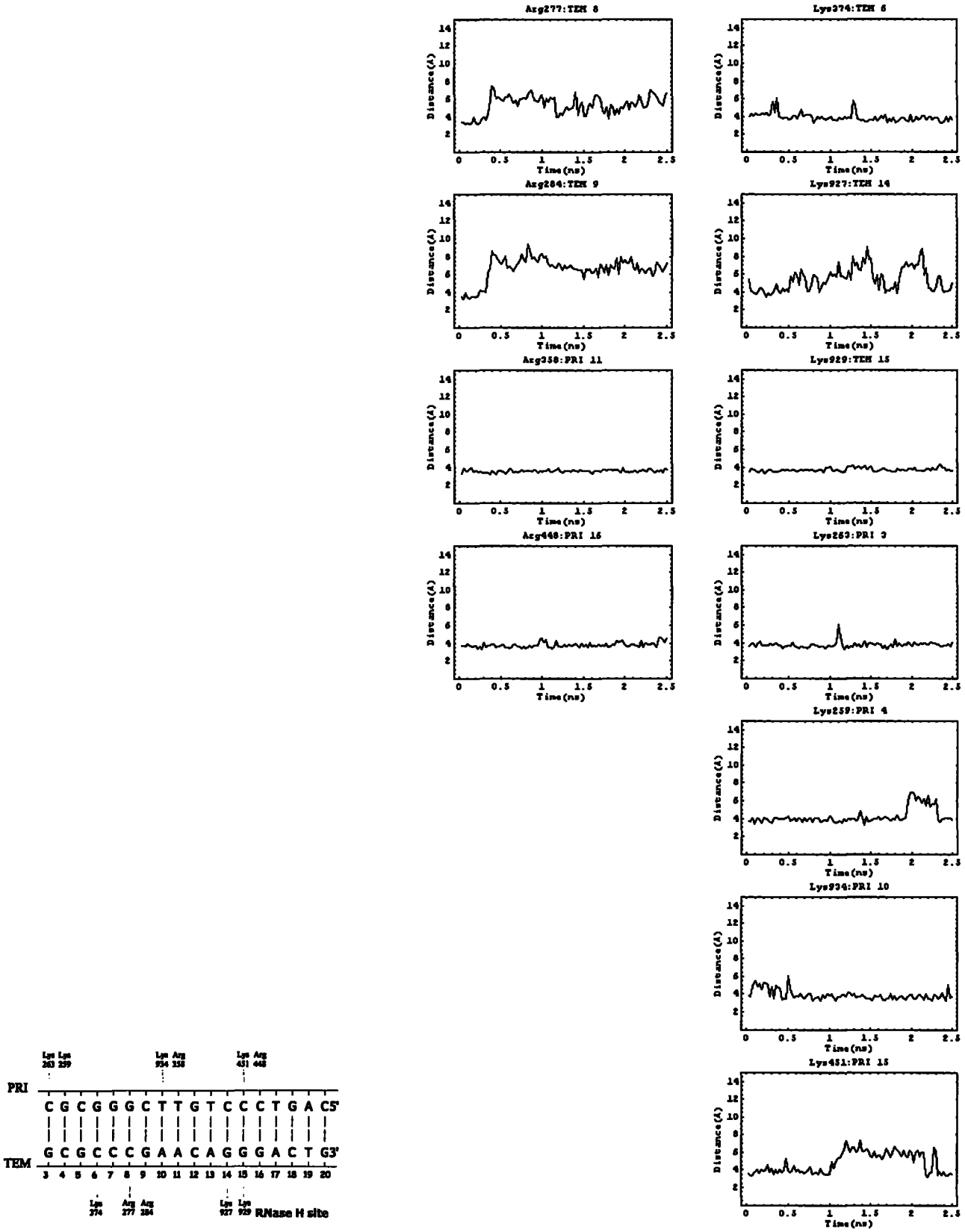


Figure H.11.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues.

System no.3: HIV-RT + DNA:DNA + 3' Terminal Deoxyadenosine

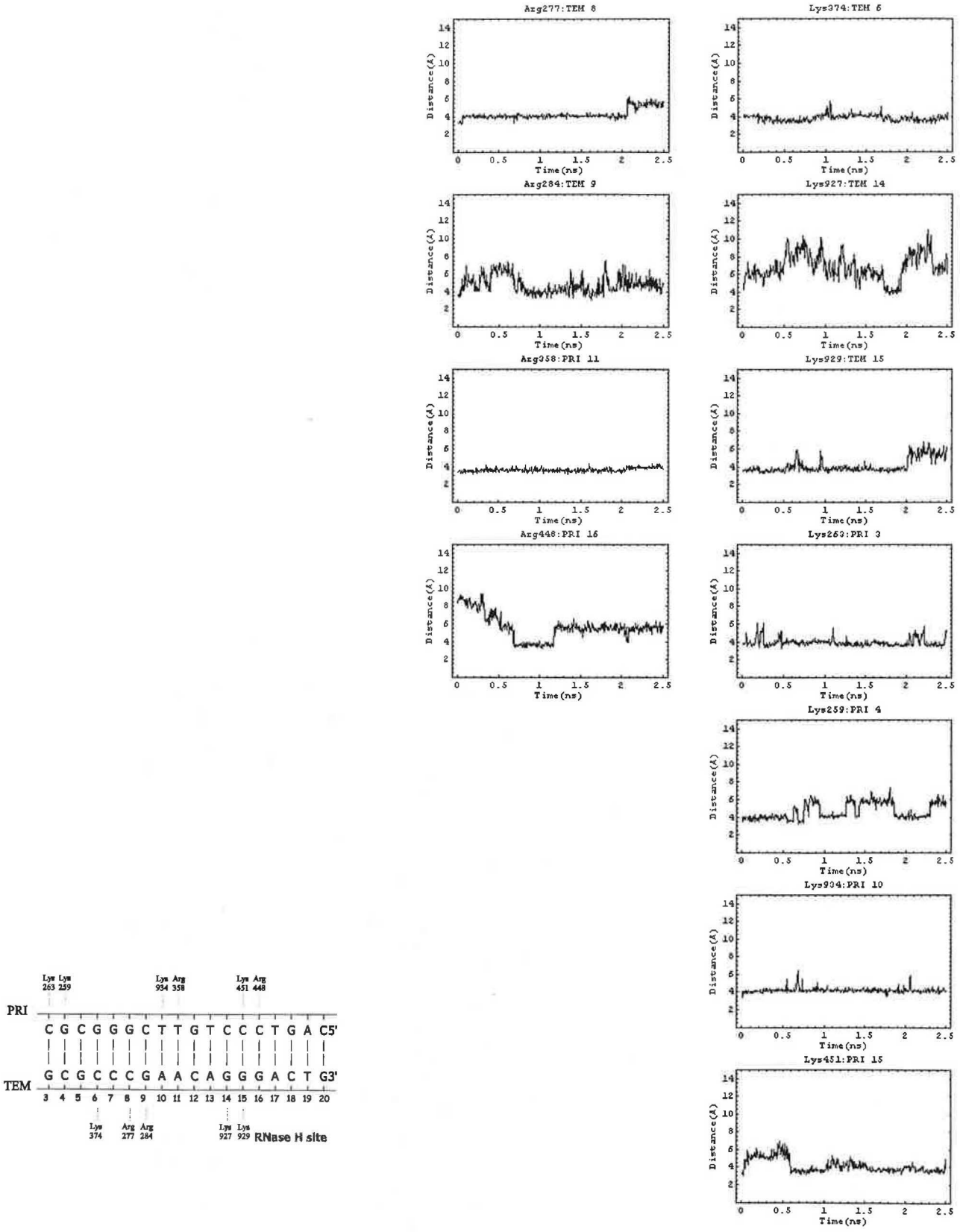


Figure H.12.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues.

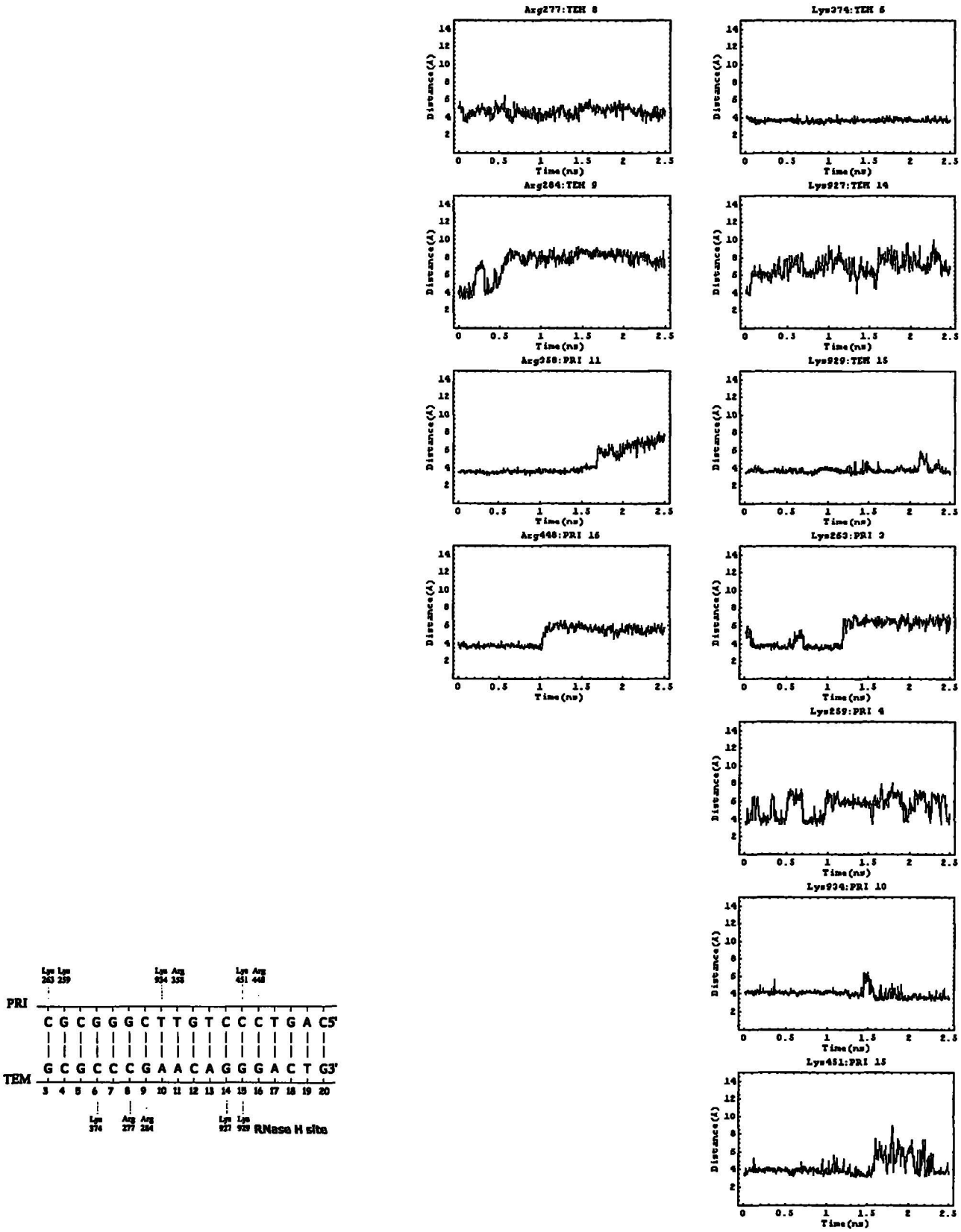
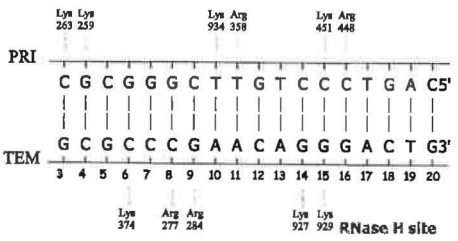


Figure H.13.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues.

System no.5: HIV-1 RT + DNA:DNA + dATP



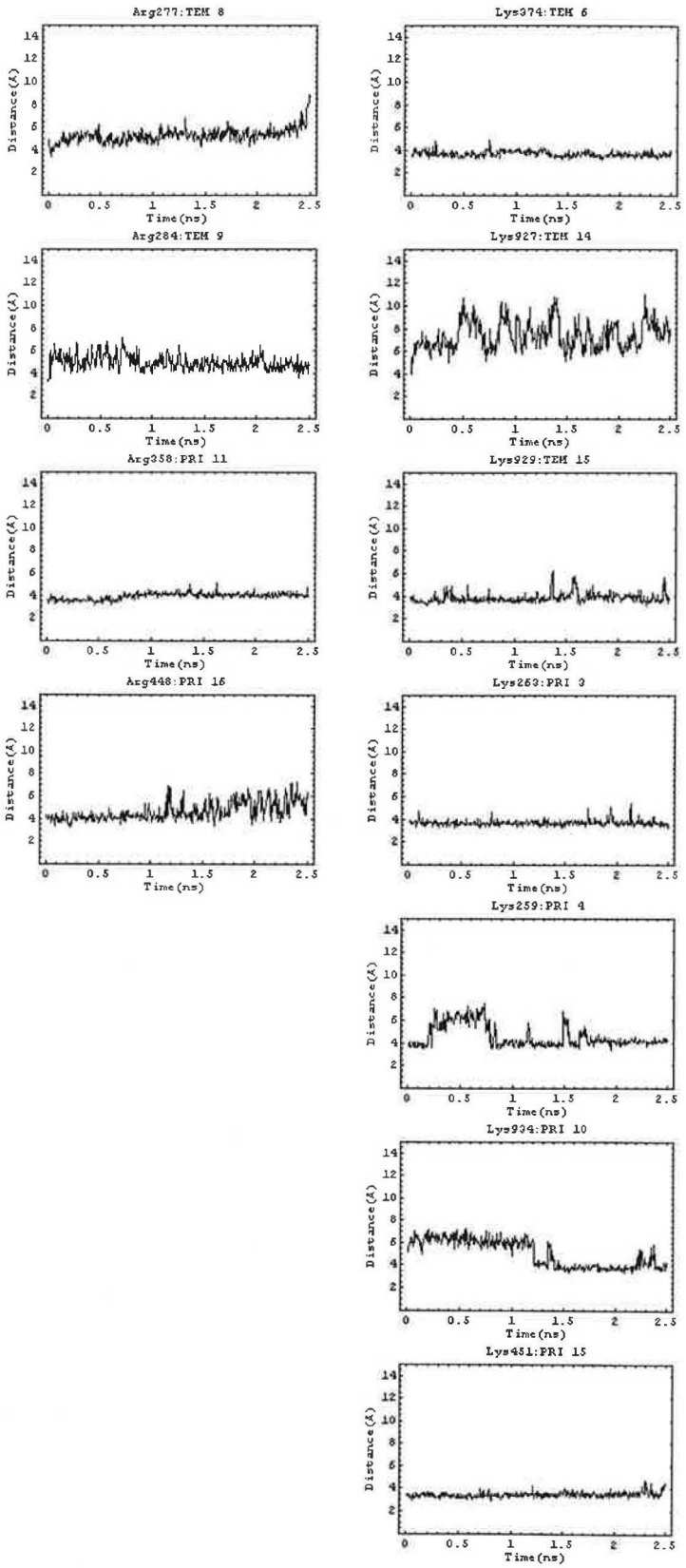


Figure H.14.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues.

System no.6 I: HIV-1 RT + DNA:DNA + ATP

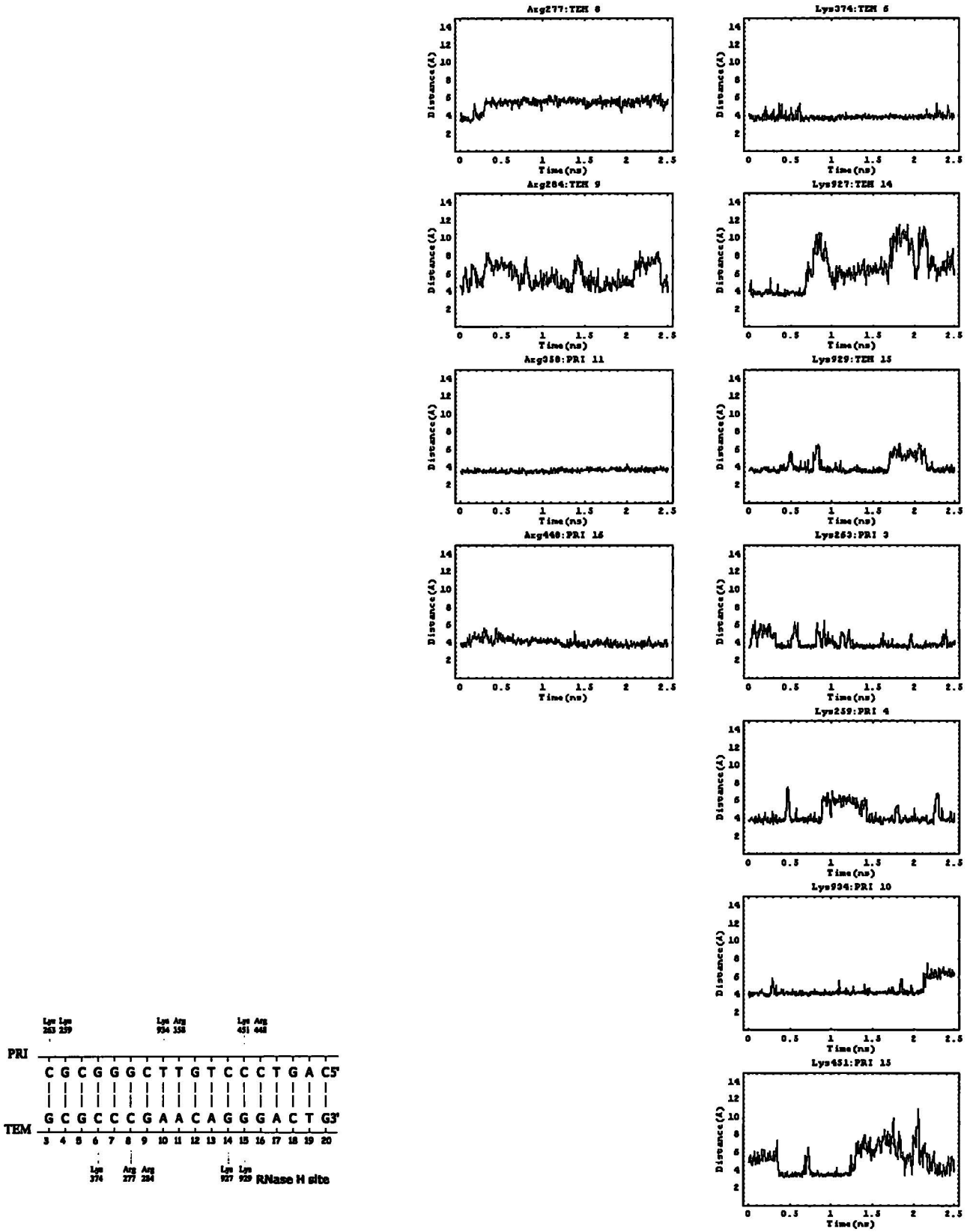


Figure H.15.: The distance between phosphate atom (P) of nucleotides and their nearest N_ζ atom of either Lysine or Arginine residues.

System no.6 II: HIV-1 RT + DNA:DNA + ATP

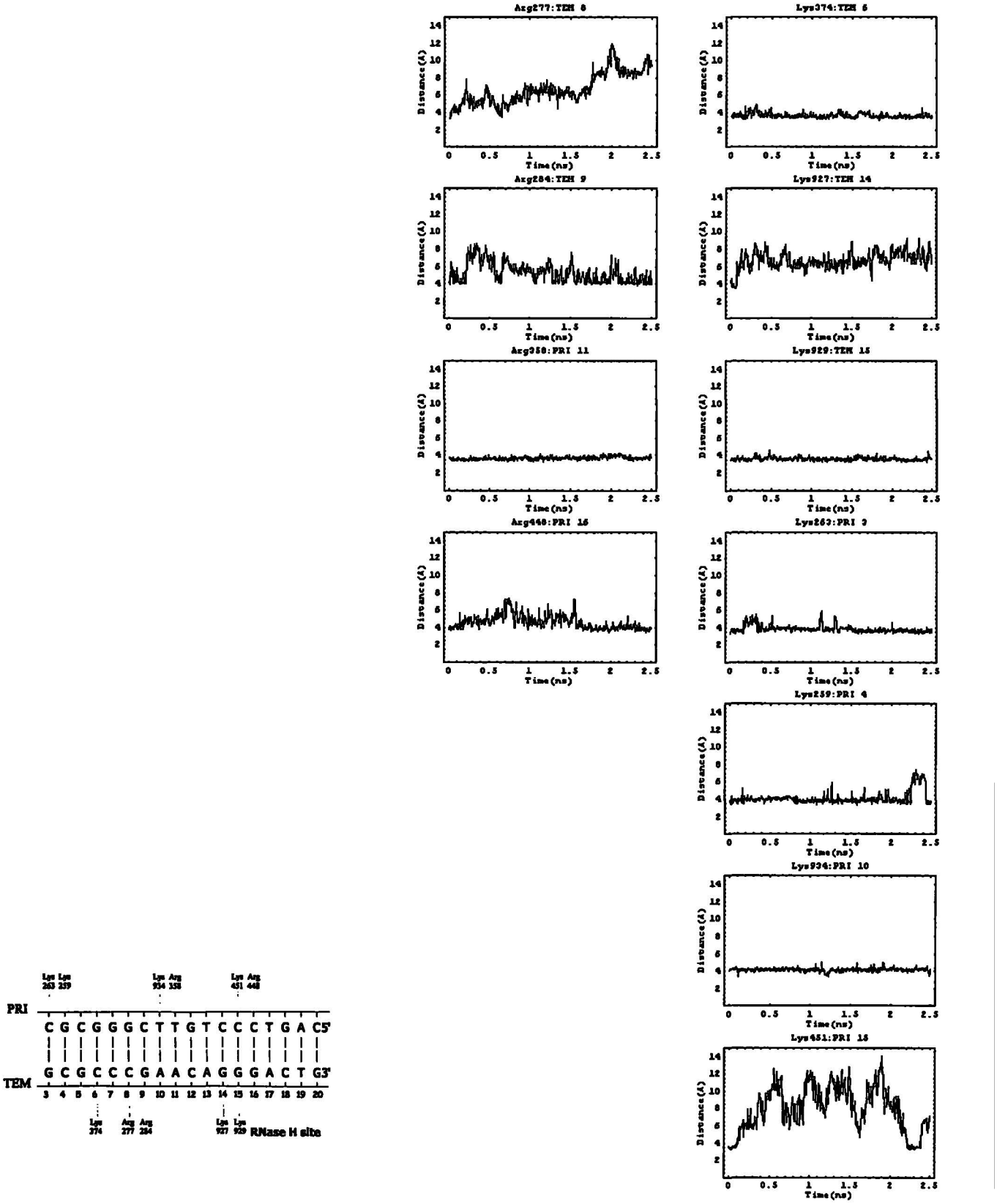


Figure H.16.: The distance between phosphate atom (P) of nucleotides and their neares N_ζ atom of either Lysine or Arginine residues.

System no.7: HIV-1 RT + DNA:DNA + PMEA (after incorporation into primer DNA)

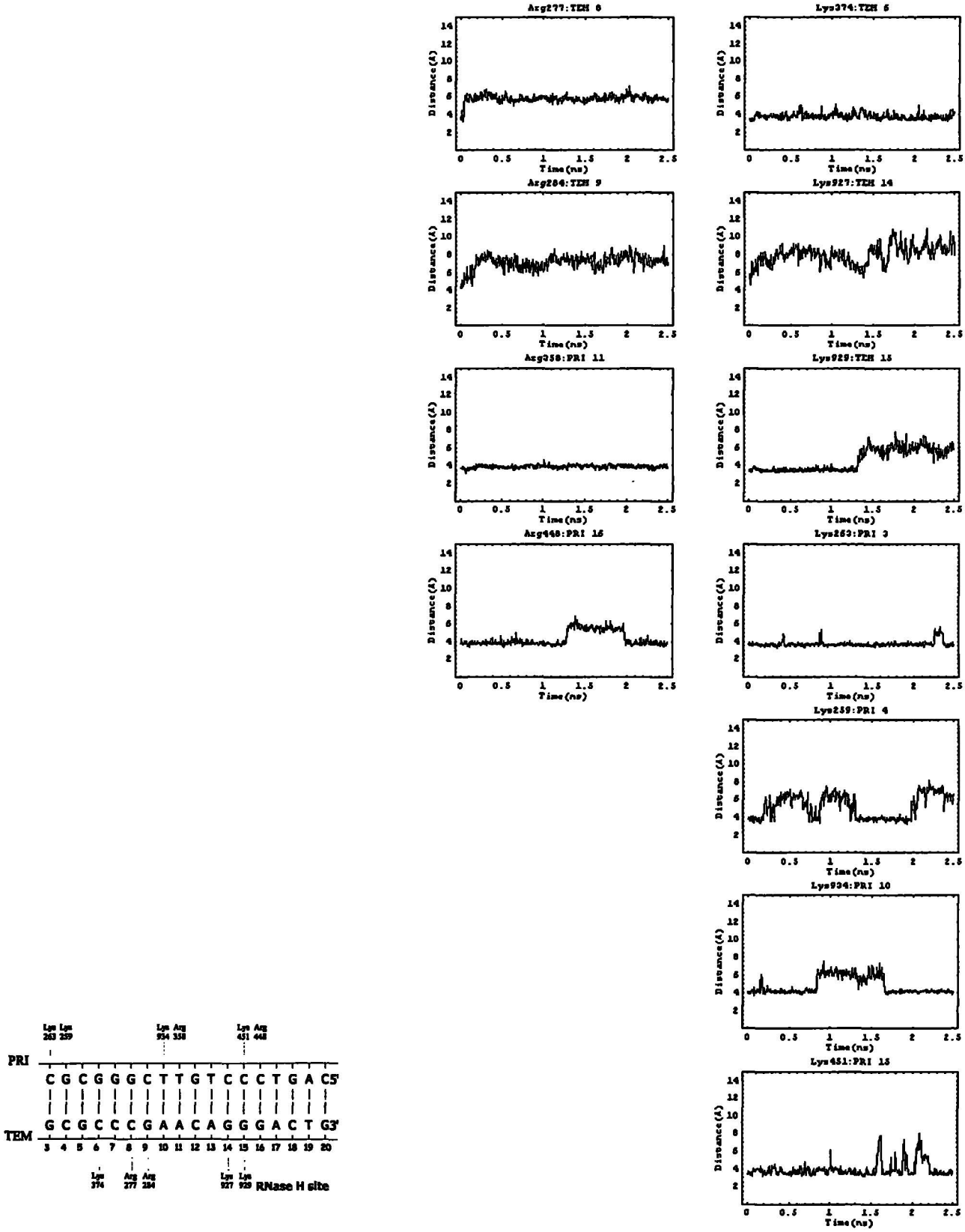


Figure H.17.: The distance between phosphate atom (P) of nucleotides and their neares N_ζ atom of either Lysine or Arginine residues.

System no.8: HIV-1 RT + DNA:DNA + PMPA (after incorporation into primer DNA)

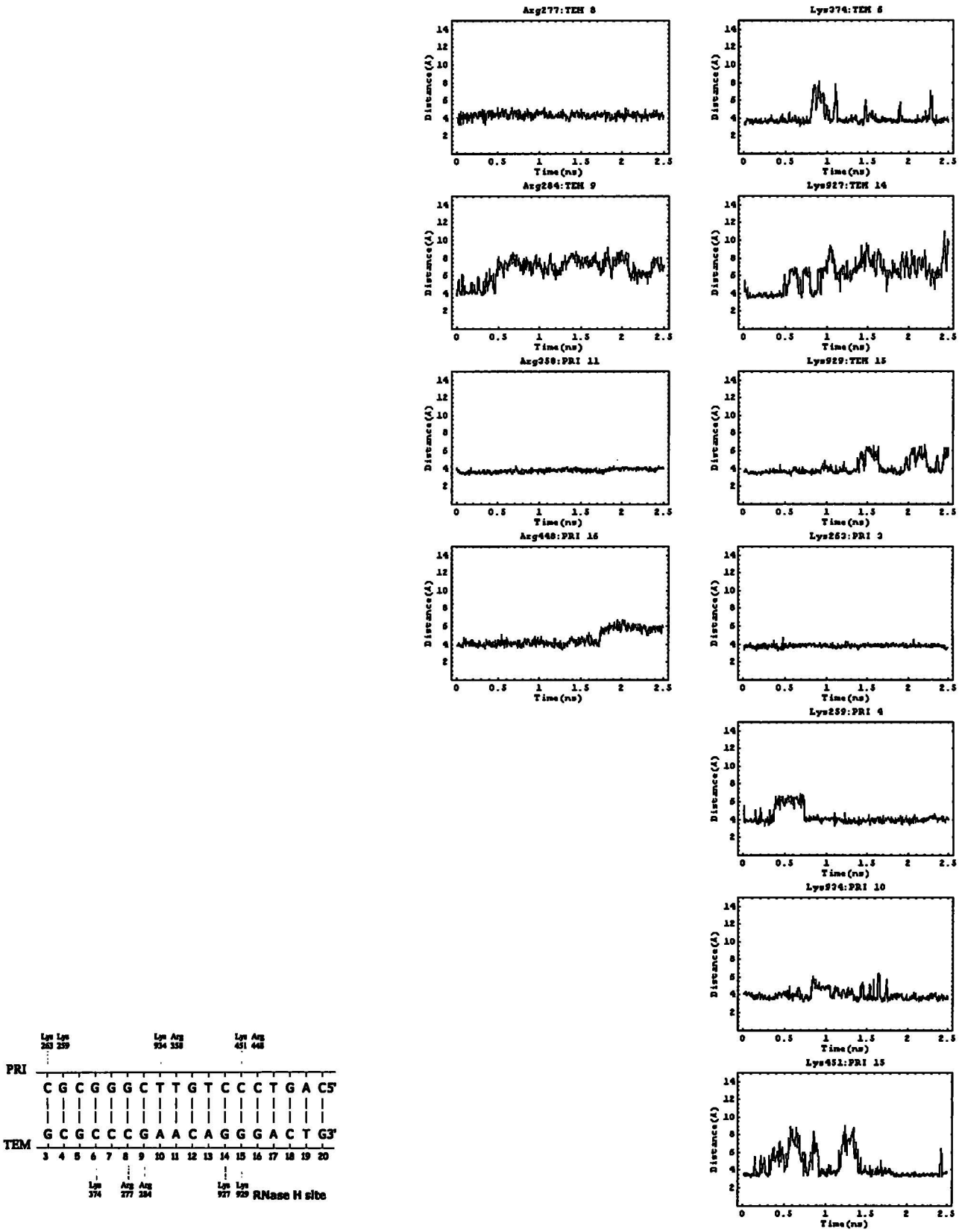


Figure H.18.: The distance between phosphate atom (P) of nucleotides and their neares
N_ζ atom of either Lysine or Arginine residues.