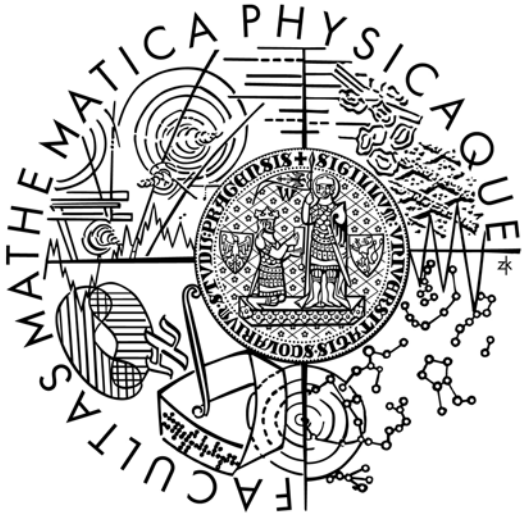
Universita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Daniel Šmít

Molekulárně dynamické simulace komplexů sestávajících z nukleových kyselin a proteinů

Fyzikální Ústav MFF UK

Vedoucí bakalářské práce: RNDr. Ivan Barvík, Ph.D.
Studijní program: Fyzika, Obecná fyzika

2008

Chtěl bych poděkovat svým rodičům za povzbuzování a podporu při studiu. Rovněž bych chtěl poděkovat vedoucímu práce RNDr. I. Barvíkovi, Ph.D. za trpělivost, vedení a cenné rady.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne Daniel Šmít

Contents

Název práce: Molekulárně-dynamické simulace komplexů sestávajících z nukleových kyselin a proteinů
Autor: Daniel Šmít
Katedra (ústav): Fyzikální ústav UK
Vedoucí bakalářské práce: RNDr. Ivan Barvík, Ph.D.
e-mail vedoucího: ibarvik@karlov.mff.cuni.cz

Abstrakt: Práce od základů seznamuje s biochemickými principy, strukturou nukleových kyselin a aminokyselin, stavbou proteinů a mechanismem jejich syntézy. Zvláštní zřetel je kladen na způsoby replikace nukleových kyselin různých virů a na možnosti terapie založené na interferenci s touto replikací. Dále jsou nastíněny i jiné moderní metody terapie spočívající v modulaci imunity či využití RNA interference. Struktura viru HCV je podrobně popsána. Replikační enzym viru, HCV RNA dependentní RNA polymeráza, je podrobně popsán spolu s předpokládaným průběhem jeho iniciace a polymerační aktivity. Dále jsou popsány základy molekulárně dynamických (MD) simulací, které jsou v rámci bakalářské práce demonstrovány na jednoduchém modelovém systému argonového clusteru. V druhé polovině práce jsou provedeny MD simulace inhibice HCV RdRp prostřednictvím látek PMEG, PMPG a HPMPG. Struktura a stabilita komplexu je detailně analyzována na atomární úrovni.

Klíčová slova: molekulární dynamika, nukleové kyseliny, HCV, HPMPG

Title: Molecular dynamics simulations of complexes consisting of proteins and nucleic acids
Author: Daniel Šmít
Department: Institute of Physics of Charles University
Supervisor: RNDr. Ivan Barvík, Ph.D.
Supervisor's e-mail address: ibarvik@karlov.mff.cuni.cz

Abstract: This work explains biochemical principles, structure of nucleic acids and amino acids, protein organization and mechanisms of proteosynthesis. Special attention is paid to means of replication of viral nucleic acids and to possibilities of therapy based upon interference with this replication. Further are explained foundations of other modern therapy methods, including immunity modulation agents and the usage of RNA interference. Structure of HCV virus is described to the higher extent. Replication enzyme of the virus, HCV RNA dependent RNA polymerase, is described in detail as well as its supposed initiation and polymerization activity process. The basics of molecular dynamics (MD) simulations are put forth, and demonstrated on the simple model system of the argon cluster. In second half of the work, MD simulations of inhibition of HCV RdRp through compounds PMEG, PMPG, and HPMPG are carried out. Structure and stability of complexes is in detail analyzed on the atomic level.

Key words: molecular dynamics, nucleic acids, HCV, HPMPG

Introduction

Hepatitis is term for an inflammation of liver with development of jaundice as a characteristic feature. We recognize six etiological agents - hepatic viruses A, B, C, D, E and G. Hepatitis C virus (HCV) is spread among 3%, or 170 million, of world population, of which only about 25% are having symptoms [1]. Current treatment with interferon-α (immunomodulation) and ribavirin (nucleotide inhibitor) has limited efficiency and numerous side effects. Only 40% of people exposed to HCV outbreak fully recover while rest is affected by chronic liver problems, with possibility of developing cirrhosis or even liver cancer.

Nucleotide inhibitors of HCV RNA dependent RNA polymerase (RdRp) have been explored in recent years with 2`-C-methylcytidine (NM107) and 4`-azidocytidine (R1479) reaching clinical tests [2]. Recent experimental studies show that a substance known for more than twenty years - (S)-HPMPA in a suitable prodrug form (to enhance cellular uptake) - might be able to fight HCV too.

The aim of this work is to explain (using molecular dynamics simulations) how acyclic phosphonate analogs of nucleotides (like PMEG, PMPG, (S)-HPMPG) could interfere with HCV replication machinery. And which of them would serve the best.

Chapter 1 – Nucleic acids and proteins

Two distinct nucleic acids are recognized: deoxyribonucleic acids (DNA, fig 1.1) and ribonucleic acids (RNA, fig 1.1) [3]. Both are polymeric chains composed of monomers called nucleotides. These nucleotides are built up of three parts. Sugars (riboses or deoxyriboses) and phosphate groups constitute backbones of nucleic acids. Further, nitrogenous bases are anchored to sugar phosphate backbones. Sugar and phosphate groups are joined by phosphodiester linkages formed between 3` and 5` carbon atoms of adjacent sugars (note that numbering of atoms is primed here in contrast to bases). Bases are attached to sugars at C1` by the N-glycosidic bond. The backbone of nucleic acids is hydrophilic. The pK value of phosphate groups is almost zero, so phosphate groups are negatively charged in neutral environment [4]. In contrast, nitrogenous bases are hydrophobic and rather non-soluble under neutral conditions.

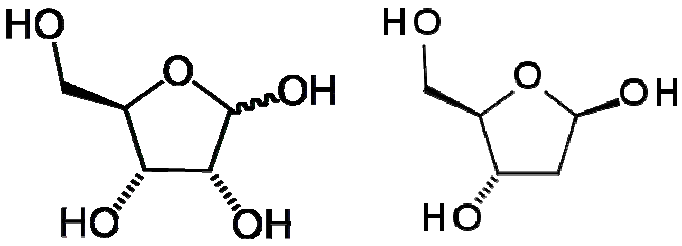


Figure 1.1: Ribose [5] and deoxyribose [6]

1.1 DNA

DNA is responsible for encoding of genetic information [3]. This is carried out by a unique sequence of bases in the polymer. DNA contains two pairs of bases; each pair consists of one pyrimidine and one (complementary) purine base, namely cytosine-guanine base pair and thymine-adenine base pair (see fig 1.2). DNA usually forms a double helix while complementary bases are connected by hydrogen bonds. Thus, the sequence of bases in one strand is fully determined by the sequence of bases in another strand (so called complementarity of bases). Two strands of DNA are antiparallel, and the most common B form structure consists of 10.5 base pairs per turn [4]. DNA helix is stabilized by hydrophobic stacking and hydrogen bonds.

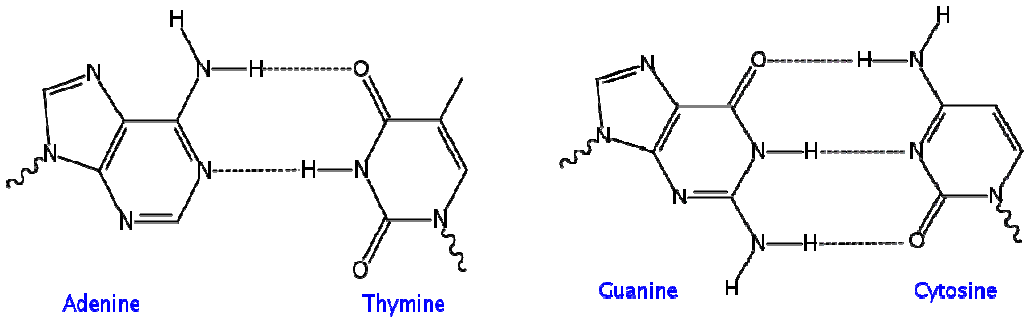


Figure 1.2: DNA base pairs with hydrogen bonds [7]

Hydrophilic backbones of duplexes are oriented outside in contact with water, while hydrophobic nitrogenous bases are stacked inside with rings roughly perpendicular to the helical axis. There are three hydrogen bonds in the C–G base pair and two hydrogen bonds in the A–T base pair [4].

Three canonical forms of DNA double helical structure are known (fig 1.3). The B form was mentioned above. The A form is common for waterless environment. It has 11 base pairs per turn and basis are inclined (20°) relating to the major axis. The Z form, unlike A and B, is left-hand rotated. It has 12 pairs per turn. Double helix, however, is not only possible spatial arrangement of DNA. Less common structures like "hairpins" or "cruciforms" can be observed. Helical structures consisting of three or four strands are well known too.

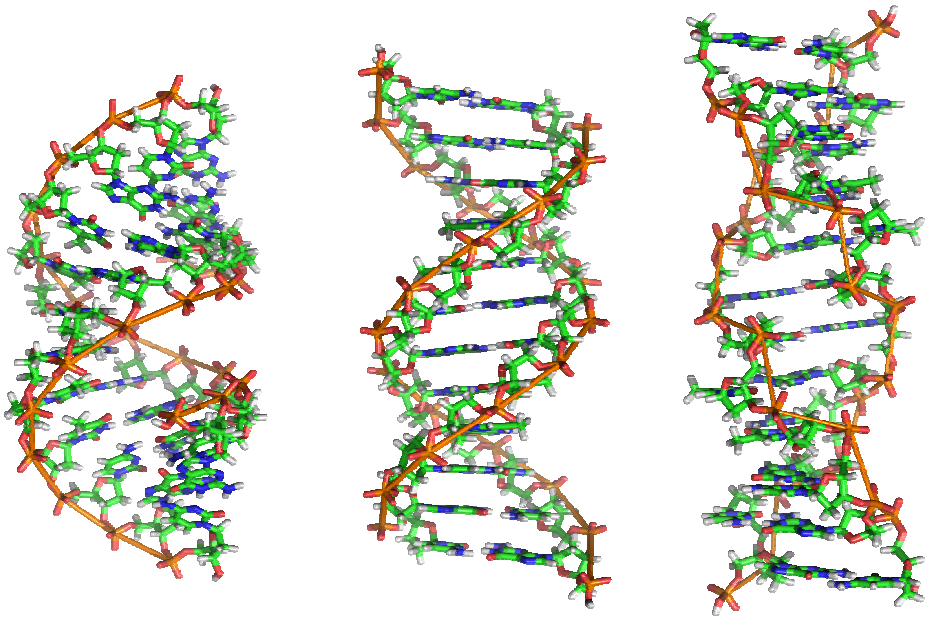


Figure 1.3: A, B, and Z DNA [7]

1.2 RNA

RNA carries out proteosynthesis. It contains riboses instead of deoxyriboses, and thymine is mostly replaced by uracil. Unlike DNA, three major types of RNA are recognized:

- **Messenger mRNA** is a transcript of nuclear DNA [3]. It carries the genetic information (base sequence) into cytoplasm. At cytoplasm, it's bound to the ribosome, and functional proteins are synthesized. The sequence of bases determines the sequence of amino acids. Each amino acid is encoded by three consecutive bases, called codon. If mRNA carries genetic information for one polypeptide, it is monocistronic (most common in eukaryotes), otherwise it's polycistronic. mRNA is always a little bit longer than necessary because it contains parts for a synthesis regulation.

- **Transfer tRNA** is much smaller than mRNA. It carries specific covalently bound amino acids to the ribosome. Amino acids are bound to the adenine residue. tRNA contains a specific counterpart for codon, called anticodon. As codons are paired with anticodons, tRNAs are ordered one by another on the

mRNA strand. Then, each amino acid establishes a peptide bond with the neighboring amino acid. This way is mRNA translated into a polypeptide chain (fig 2.1).

- **Ribosomal rRNA** is a structural component of ribosomes.

The arrangement of RNA is not plainly double helical, as base pairing takes place only in short segments [3]. Transcribed RNA is usually single stranded, and forms right handed helical structures. It can also pair with complementary regions of RNA and DNA. As a result, objects with many hairpins, internal loops, or bulges can exist. Resulting secondary structure is generally very complicated. Specific short base sequences found at hairpins are supposed to play an important role as starting points of RNA folding. In double helical conformation, A form is the most usual, Z form only under laboratory conditions, and B form was not yet observed [4].

1.3 Sugar Puckering

Five torsion angles among atoms constituting the pentose ring:

$$\begin{aligned} \nu_0 &:= C_4' - O_4' - C_1' - C_2' \\ \nu_1 &:= O_4' - C_1' - C_2' - C_3' \\ \nu_2 &:= C_1' - C_2' - C_3' - C_4' \\ \nu_3 &:= C_2' - C_3' - C_4' - O_4' \\ \nu_4 &:= C_3' - C_4' - O_4' - C_1' \end{aligned} \qquad (1.1)$$

They can be reduced to one pseudo-rotational angle P given as [8]:

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin 36° + \sin 72°)}. \qquad (1.2)$$

Further, for sugar puckering amplitude we have:

$$\nu_{\max} = \frac{\nu_2}{\cos P}. \qquad (1.3)$$

All possible sugar puckers are shown in figure 1.4. It was widely noticed [8], that B form DNA contains mainly C2'-endo pucker (144° < P < 190°) while A form prefers C3'-endo pucker (0° < P < 36°).

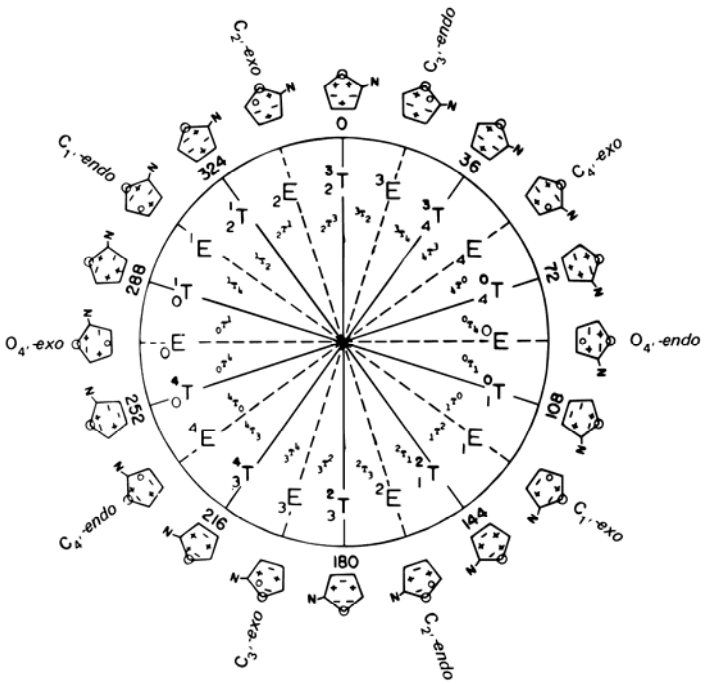


Figure 1.4: Sugar puckering [9]

1.4 Amino Acids

Until today, twenty common amino acids have been recognized. All of them are α-amino acids, it means they have carboxyl and amino groups both bound to the α-carbon [4] (fig 1.5). They have specific alkyl groups, which determine their physical and chemical properties (solubility etc.). Other modified amino acids also exist, but they do not build up proteins. According to the position of amino groups on the chiral centre, we designate L- (having the amino group oriented to the left in Fisher projection) and D- (to the right) amino acids. The L stereoisomer is the most common in natural proteins. Amino acids can be classified upon their alkyl groups as non-polar, polar, uncharged, positively or negatively charged, and aromatic.

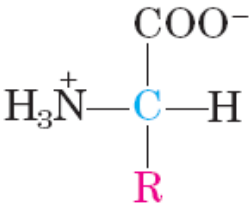


Figure 1.5: General structure of L-amino acid [4]

1.5 Proteins

Similarly to nucleic acids, proteins are also polymers [10]. They are made up by sequences of amino acids connected by a peptide bond, formed between N of the amino group and C of the carboxyl group. The formation of the peptide bond leads to the release of a water molecule as a by-product. The peptide bond is planar with oxygen and hydrogen atoms in a trans conformation. Peptide bonds are rigid due to

partial double bond character. Therefore, only rotations occur around bonds with the C^{α} atom. Angles of rotation are usually marked φ for C^{α} – N and ψ for C^{α} – C.

Four levels of the protein structure are distinguished [4]:

- **Primary structure** is a sequence of amino acids constituting the protein.

- **Secondary structure** is mainly a local spatial conformation. The α-helix (fig 1.6a), β-sheet (fig 1.6b, 1.6c) and β-turn are recognized. α-helix is s single chain helical structure stabilized by internal hydrogen bonds, with side chains outside of the helix [4]. Common in nature is helix with 3.6 amino acids per turn, long 5.44 Å, with φ = -60° and ψ = -45° to -50°. This structure is suitable for exploitation of hydrogen bonding as every residue n is bound to the residue $n + 4$. β-sheet, or pleated sheet, is a pleated plane of several chains put alongside and stabilized by hydrogen bonds between C=O and N-H groups [10]. It is called pleated because α carbons appear by turns above and below the plane of the sheet. It has to be pleated in order to make enough space for side chains. These side chains can be parallel, or antiparallel. β-turn is one of turns that connect helices and sheets in globular conformation of protein. This turn connects adjacent parts of antiparallel β-sheet, and is a 180° turn. Motif, or super secondary structure, is sometimes used to describe ordered structure of α-helices and β-sheets. In case no repeated units are present, protein conformation is called random coil [10].

- **Tertiary structure** describes spatial ordering of α-helices, β-sheets and β-turns on long range scale. It is considered to be induced by weak interactions among parts of the chain. Two groups of tertiary structures are fibrous and globular proteins. Fibrous proteins usually consist of single type of secondary structure. Globular proteins are made up of various secondary structures and usually play role of enzymes and other functional proteins [4]. Following super secondary structures may be found in globular proteins, they are stable conformations made up of secondary structure segments, as the most common are:β–α–β loop, α–α corner, connection of all-β motif (structure of pure β-conformations), connection between β-strands, β barrel (β–conformations alongside, twisted into a column) and twisted β-sheet (fig 1.6d).

- **Quaternary structure** is organization of multi-protein composed of several folded protomers (single polypeptide chains). Since they are composed of several protomeres, they usually posses certain symmetries (cyclic, dihedral etc.). Since the native structure of a protein is one of many conformations which could be realized, it seems strange that freshly synthesized proteins would go through random process of folding [4]. Two explanations were made. First, hierarchical model, states, that according to specific amino acids, secondary structures are made, and then, long range interactions lead to formation of the tertiary and quaternary structure. Other model, called molten globule, claims that first the protein collapses, and folding is carried out by hydrophobic interaction among residues. In reality, the process is most likely combination of both models. Some proteins are folded with help of other specialized proteins called chaperons.

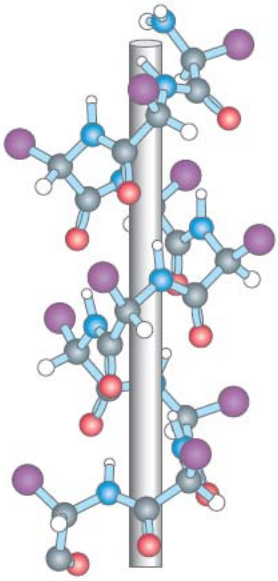


Figure 1.6a: α-helix [4]

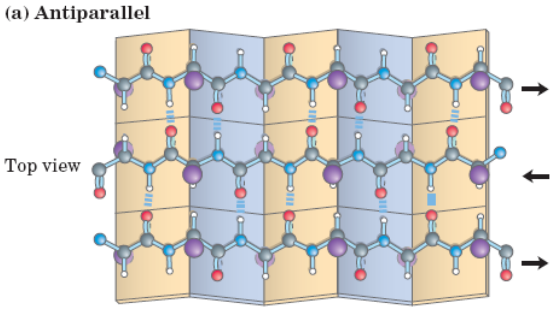


Figure 1.6b: β-sheet (antiparallel) [4]

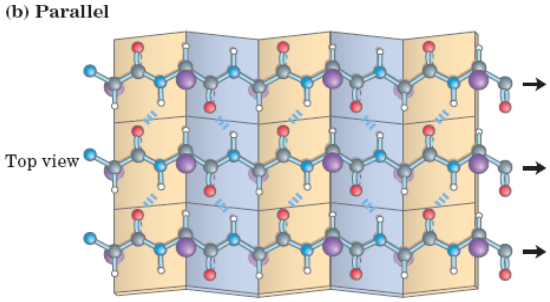


Figure 1.6c: β-sheet (parallel) [4]

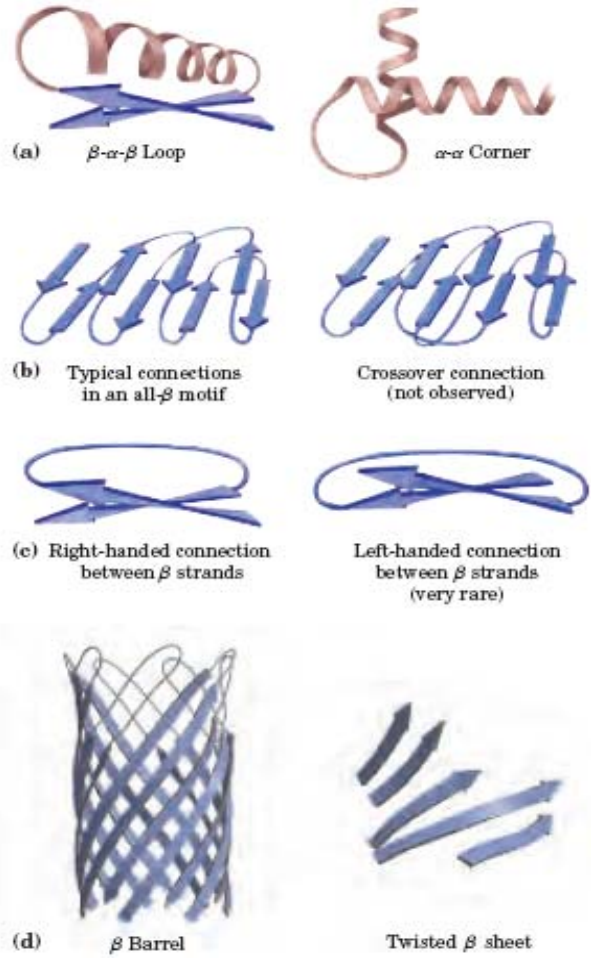


Figure 1.6d: Proteins – structural motifs [4]

Chapter 2

DNA replication, DNA transcription, and RNA translation

2.1 DNA replication

Biosynthesis of proteins starts at cell's nucleus, with a DNA [4]. Its self-replication ability is necessary for genetic information to be passed on. Self-replication is semiconservative, it means that one parent strand of double helix has complementary strand synthesized. Both parent strands are replicated simultaneously. The new DNA build-up proceeds in 5`→3` direction (so parent DNA is read in 3`→5` direction). Synthesis is carried out by DNA polymerase enzymes; it has two requirements, a template (parent) strand and a primer, a short complementary segment of RNA with open 3` end, so it can attack α-phosphate of dNTP. Primers are synthesized by a special enzyme. The replication is very accurate because of the Watson-Crick pairing, as well as specific active sites of polymerases. Several DNA polymerases were identified, making replication a very complex process. While 5`→3` replication proceeds continuously, 3`→5` direction is discontinuous and ends up with creation of 150-200 nucleotide fragments, called Okazaki fragments [10]. Okazaki fragments are connected by a replication complex.

2.2 DNA transcription

Transcription resembles replication in direction of synthesis, chemical mechanisms, and requirement for template [4]. On the other hand, only one DNA strand is transcribed at the time, primer is not necessary, and only some parts of template DNA are used. The responsible enzyme is called DNA-dependent RNA polymerase (DdRp). As today, we recognize RNA polymerase I-III (in eukaryotic cells). Enzyme requires DNA template, all four nucleotides and Zn^{2+}, Mg^{2+} ions. In analogy to DNA polymerase, it adds appropriate nucleotides to the 3` end of the strand; only Thymine base is replaced by Uracil base. Primer is not present, if transcription starts at the specific DNA sequence called promoter (it triggers transcription of specific segment, a gene). During the transcription, DNA is reversibly unwound. In the meantime, DNA – RNA double helix, about 8 base pairs long, is formed, but unwinds, as DNA recombines. The copied DNA strand is called template strand and the other one is called coding strand.

Nascent RNA product is called primary transcript, and after transcription, it is further processed. RNA is processed by proteins, and also by catalytic RNAs, ribozymes [4]. The primary transcript of mRNA usually contains sequence for one gene, yet might be non contiguous, with coding exons and non coding introns. Introns are removed (by so called splicing), so the pure exon sequence is made for a specific polypeptide. According to conditions, cleavage can sometimes produce various polypeptides from a single gene. Further, mRNA is closed by the 5` cap, which protects it against ribonuclease. Primary transcript of tRNA is also cleaved, sometimes purified of introns and has sugars and bases modified.

2.3 RNA translation

Proteosynthesis proceeds accordingly (fig 2.1): transcribed mRNA is eluted into cytoplasm, carrying information for synthesis of a certain protein [3]. In cytoplasm, it binds to several ribosomes (composed of about 65% of RNA and 35% of protein). tRNA brings amino acid and is bind to the mRNA according to codon – anticodon pairing. Since 20 amino acids are recognized, and code is carried out by four code letters (A, G, U, C), we have $4^3 = 64$ unique codons, so it means, several codons exist for one amino acid (the codon is then called degenerate). Some codons don't encode amino acids [3]. Codon AUG is a signal for the beginning (initiating codon) of the polypeptide (and is also coding a Met residue) and UAA, UAG and UGA are terminating codons, marking the end of the polypeptide.

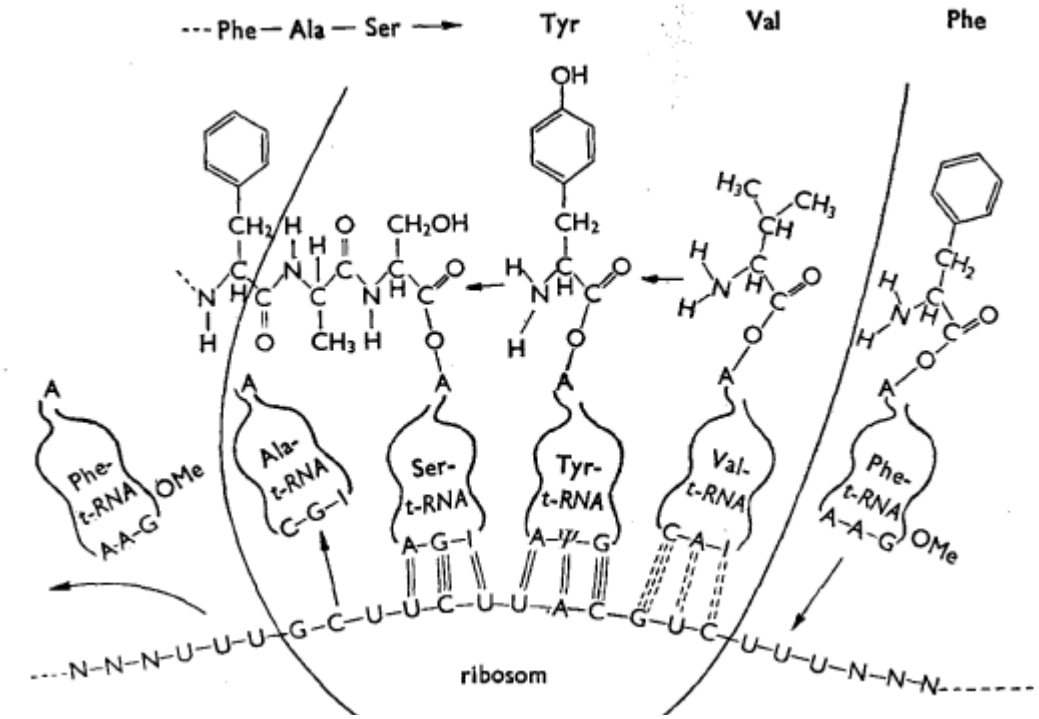


Figure 2.1: Illustration of RNA translation [3]

Chapter 3 - Polymerases

3.1 Eukaryotic RNA polymerase

Eukaryotic cells have three distinct nuclear RNA polymerases (I-III) with some common subunits. Each of them is specific in function, and is activated by a specific promoter. RNA polymerase I (abr. Pol I) is responsible for synthesis of pre-ribosomal RNA, which leads to various rRNAs [4]. Pol II is necessary to synthesize mRNA and some specialized RNAs. Many Pol II promoters contain the TATA box (sequence TATAAA) and an Inr (initiator) sequence. Pol III makes tRNA, one kind of rRNA and some smaller special RNAs. It is clear, that Pol II plays a key role in genetic expression. RNA polymerase II has 12 subunits (especially RBP1, RBP2, RBP3 and RBP11 are exhibiting high homology to their bacterial counterparts). Unlike bacteria, however, eukaryotic Pol II needs more inter-protein interactions, since it is more complex. These necessary proteins are called transcription factors; especially so-called general transcription factors (TFII) are needed at every promoter. Some proteins are present in preassembled complexes, simplifying the process of transcription.

3.2 HIV Reverse Transcriptase

In some RNA viruses, the transcription process can be inverted. It means that RNA is used as a template for DNA synthesis, which is carried out by an RNA-dependent DNA polymerase called reverse transcriptase [10]. The viral RNA strand enters a eukaryotic cell, and reverse transcriptase synthesizes complementary DNA. RNA part of the hybrid is then degraded; DNA double helix is formed and incorporated into the cell's genome. The genome is later transcribed by the cell along with viral DNA, so viral RNA is being recreated by the cell. Altogether, reverse transcriptase catalyses three processes: RNA-dependent DNA synthesis, RNA degradation and DNA-dependent DNA synthesis. Two different active sites of the reverse transcriptase enzyme carry out each catalyzed reaction (synthesis/degradation). Reverse transcriptase is most active on its own viral RNA; however, it can reverse many various RNAs to DNA. Viruses containing reverse transcriptase are called retroviruses. Host tRNA is used as a primer for DNA synthesis; it is paired on its 3` end with a sequence in the viral RNA. Reverse transcriptases miss proofreading mechanisms, common to eukaryotes; therefore many errors are made during the reversed transcription. As a result, viral evolution (mutations in viral RNA) is much faster and leads to new strains of retroviruses. Retroviruses cause for example cancer or AIDS. Especially HIV has high error rate, causing its fast evolution (approximately one error per transcription cycle), and thus complicating a vaccine development. Therefore, enzyme inhibitors are preferred to fight viral infections.

3.3 RdRp polymerase – RNA replicase

Some viruses containing just RNA genome are reproduces in host cells by RNA-dependent RNA polymerase (RNA replicase) [4]. Interestingly, this enzyme is not present in human cells. Mostly, the RNA replicase has four subunits. One of subunits has an active site for replication. The other three subunits help replicase to locate and bind 3` end of RNA.

Chapter 4 - HCV

4.1 HCV structure

HCV belongs to the group of single stranded RNA viruses with positive polarity (denoted as (+)ssRNA), flaviviridae family, genus hepacivirus, of which it is the only member [11]. Because of high mutation rates, 11 genotypes, subtypes and about a 100 strains are recognized. The most common type is 1 and 1b, responsible for 60% of infections worldwide [1]. HCV consists of Core with genetic material (RNA) presented in cytosol and covered by an icosahedral protein protective shell and lipid envelope of cellular origin [12]. RNA is made of 9600 bases, which can be translated into polyprotein of about 3000 amino acids. Several hundred bases belong to 3` and 5` terminal, which are not translated.

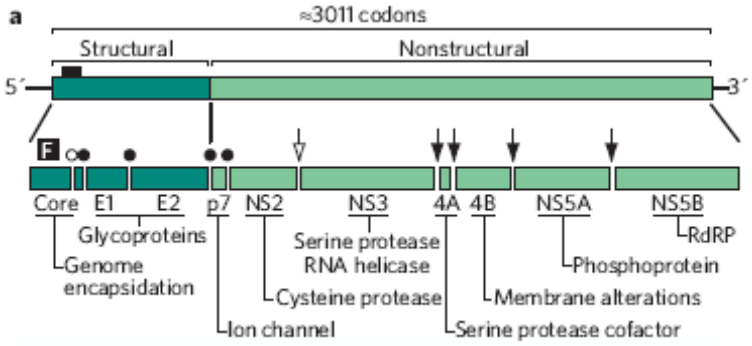


Figure 4.1: Structure of the HCV RNA [13]

4.2 HCV replication

HCV is quickly mutating at the rate approximately one nucleotide per cycle [13], effectively making the viral population heterogeneous. High viral replication rate after introduction, as well as veritable immunity reaction in liver, contributes to a high diversity. This diversity is to less extend in 3` and 5` NTR. Virus is also capable of high adaptation to various host cells. Translation process begins, as a single-stranded viral RNA is eluted into the cytoplasm and internal ribosome entry side (IRES) within the 5`-noncoding region (NCR) makes up an IRES-40S complex with a 40S subunit of cell's ribosome. A large polyprotein is produces and cleaved into 10 viral proteins (fig 4.1) [13]. One third (N-end) encodes viral structural proteins (core C and glycoproteins E1 and E2); the next part codes the membrane protein, p7. Rest of RNA encodes nonstructural (NS) proteins like NS2, NS3, NS4A, NS4B, NS5A and NS5B, which are important for viral intracellular life.

4.3 HCV RdRp

The main target of intended treatment of HCV is its RNA dependent RNA polymerase (RdRp, fig 4.1). It has a typical structure of polymerases resembling the right hand (Palm, Thumb, and Fingers domains). Conserved aspartic acids in the Palm domain bind Mg^{2+} ions. But unlike other polymerases, Thumb and Fingers are bound, and hence not allowed to change conformation [14]. Because of this, HCV RdRp is much more rigid than other polymerases.

About 30 Å away from the catalytic site is the surface binding site specific for an rGTP (ribose guanosine triphosphate). It is situated close to the connection of Fingertips and Thumb [14]. Nucleoside moiety of nucleotides is stabilized by contact with four residues from the Thumb domain (Pro495, Pro496, Val499 and Arg503) and two from the Fingers domain (Arg32 and Ser29). Triphosphate is less rigid and can have two conformations. The specific binding site only for rGTP suggests its importance as an initiating nucleotide.

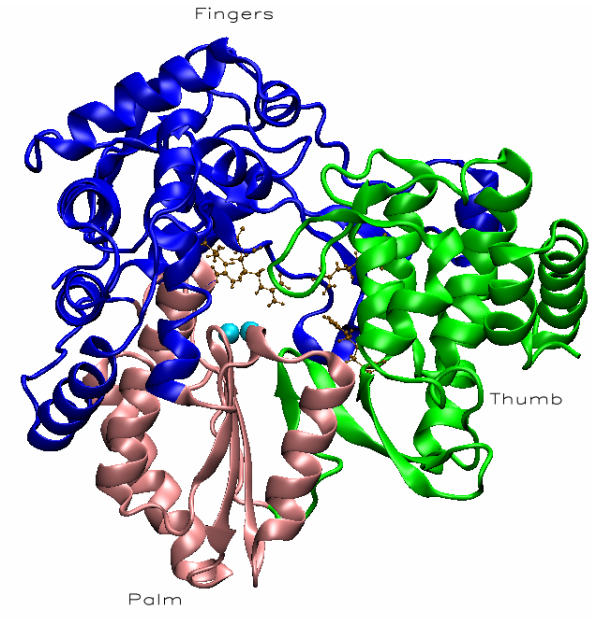


Figure 4.2: HCV RdRp "hand", Mg^{2+} - cyan, arginine (catalytic site) – ochre

4.4 HCV replication initiation complex

The key role in development of new HCV RdRp inhibitors plays understanding of the initiation of polymerization process. HCV is ssRNA virus. Nevertheless, HCV RdRp is highly similar to polymerase of dsRNA bacteriophage ϕ6. This suggests an evolutionary link between dsRNA viruses and flaviviridae, and justifies an assumption that ϕ6 polymerase initiation process (studied in [15]) will employ the same mechanism as initiation process in HCV RdRp. In X-ray study, oligonucleotide equivalent to the 3` end of the ϕ6 genome binds inside a basic tunnel leading to the active site in position and conformation equivalent to the position and conformation observed in template bound to HIV-1 RT [15]. Following this analogy, however, the template should go through the C-terminal domain. The same problem is observed in HCV RdRp. Because both polymerases can self-prime (de-novo initiation, they don't need a primer); supposedly, closing off the C-terminal may provide an instrument to form the initiation complex. Figure 4.3 describes the process of initiation of ϕ6-polymerase. There is an empty active site (I), with NTP (II) and with the template strand (III) [15]. First, the template strand and NTP are bound to the complex (IV), with first residue of template (T1) bound in the specific pocket (site S, it is specific to cytidine residue, conserved at the 3` end). GTP is base stacked with Y630 and forms Watson-Crick pair with cytidine (also conserved) of T2 (V) and then, template is pulled backward by electrostatic interaction (VI) allowing another GTP to be bound

to T1 (VII) and displacing the C-terminal domain, probably by an attraction of GTP phosphates to the Mn^{2+} (VIII). C-terminal is also more mobile and could flex enough to make enough space for the nascent strand to pass. This mechanism can be generalized to HCV.

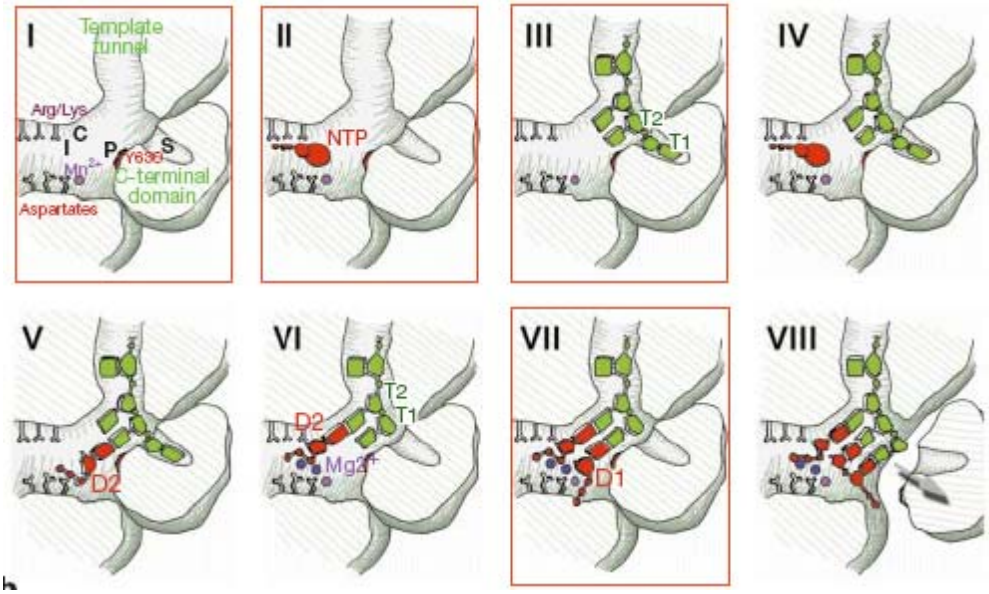


Figure 4.3: Initiation of ϕ6-polymerization [15]

4.5 HCV RdRp (NS5B RNA polymerase) inhibition

RdRp is not present in mammalian cells. Therefore, it is a suitable target for inhibition [16]. One way are nucleoside inhibitors. They are converted by host cells into nucleotides. During the viral replication, they are incorporated into the synthesized RNA and act as chain terminators. In the case of HCV, 2`-C-methyl nucleosides are mostly used. They confer antiviral activity irrespectively to the base [17]. One of them is NM283 - a prodrug of 2`-C-methyl-cytidine [16]. The methyl group added to C2` of ribose transforms a nucleotide substrate into a chain terminator. Large drop in viral load was observed with this cure, however, the level returned to the base value as soon as treatment was interrupted. It could be suitable for combined therapy. HCV can quickly acquire resistance, as it starts to discriminate among the drug and natural nucleotides by means of the mutation of Ser 282 for Thr. This mutation, however, leads to a drop of activity, down to 8 – 17% of the original polymerase [17].

It was also observed that HCV chemotherapeutic agents act synergistically with IFN-α (interferon alpha) and ribavirin to inhibit HCV replication [17]. 2`-C-Me-G combined with IFN-2α lead to increased inhibition, however, the same combination with 2`-C-Me-A was less effective.

Other possibility of treatment is non-nucleoside inhibition (NNI) using allosteric inhibitors. They prevent HCV RdRp from reaching an active conformation [16]. Several suitable binding sites for NNIs on the enzyme allow combination therapy. Inhibitors bind to the thumb cavity, which is crucial for “fingertip-thumb” interaction, thus forcing enzyme to remain in the inactive form. The mutation of Pro 495 for Ala or Leu can lead to resistance, but causes HVC RdRp to become futile.

4.6 Acyclic phosphonate analogs (S-HPMPA etc.)

Compound (S)-9-(3-hydroxy-2-phosphonylmethoxypropyl)adenine, or (S)-HPMPA was discovered by Czechoslovakian and Belgian scientists 22 years ago. The compound proved to be an effective and selective anti-DNA virus agent, treating various herpes viruses, varicella viruses, retroviruses (like HIV) and other viruses [18]. Unlike Nucleoside Inhibitors, however, Nucleotide Inhibitors have first phosphate group already present [19]. The monophosphate formation is often rate limiting, while subsequent phosphorylations are easily carried out by intracellular kinases. β-D-2`-Deoxy-2`-fluoro-2`-C-methylcytidine (PSI-6130) and its deaminated analog RO2433 were already studied. It was confirmed that first phosphorylation is limiting factor to antiviral activity [2]. When (S)-HPMPA is incorporated as a regular nucleotide into the viral DNA (e.g. HIV DNA produced by HIV RT), the absence of a sugar ring forbids other nucleotides to be added to the nascent DNA strand and the synthesis is aborted. The main advantage is that (S)-HPMPA has much bigger affinity for HIV RT than for human polymerases (it leads to a moderate toxicity). (S)-HPMPA can be also incorporated by HCV RdRp into nascent RNA strand. The hydroxyl group, however, makes (S)-HPMPA more hydrophilic and less suitable for a cell membrane transport. This issue can be solved by adding long carbonyl chains to the phosphate group. The chain easily penetrates to membranes, and the drug is transferred inside by flippase enzyme. The chain is then cut by cellular enzymes, and (S)-HPMPA, with high affinity to viral RdRp, can be incorporated into the new RNA strand.

The aim of molecular dynamics simulations presented here is to clarify, how acyclic phosphonate analogs fit into the active site of HCV RdRp. The fact, that (R)-HPMPA is anti-virally inactive [19] emphasizes the importance of proper molecule conformation. Interestingly, the HCV RdRd active site is closely similar to one of HIV RT, which has been intensively studied. In fact, HIV RT was the primary target for (S)-HPMPA.

4.7 HCV Protease (NS3-4A) inhibition

HCV protease is a dimeric protein, which is made of a NS3 protein and a smaller NS4A cofactor [16]. Besides viral protein cleavage, the enzyme is responsible for blocking of an antiviral response. Enzyme's substrate has to be a long polypeptide, with which it can form many weak interactions. Large polypeptides are, however, poorly convertible to drugs. In combination with a shallow substrate pocket, the task of an inhibitor preparation proved to be difficult. The first HCV protease inhibitor was BILN 2061 [16]. It successfully decreases the amount of viral load. Nevertheless, it failed in clinical trials because of a cardiac toxicity. Further, single mutations (Arg 155, Ala 156 or Asp 168) lead to resistance that developed in viral populations. Another inhibitor is VX-950. It is also product-derived inhibitor, stabilized in the active site by α-ketoamide, which forms a covalent bond with a catalytic serine [16]. This drug proves to be more effective and safe than BILN 2061, decreasing viral level rapidly, yet not eradicating the virus completely. Further, covalent bond doesn't prohibit virus from getting resistance. Again, replacing a single amino acid (Ala 156 for Ser, Thr or Val) is sufficient to make VX-950 ineffective.

Chapter 5 – Oligonucleotide based therapeutics

5.1 Antisense antiviral agents

Antisense oligonucleotides and ribozymes (RNA enzymes) could be used as therapeutic agents [16]. They bind to the mRNA and block translation of a disease-causing gene [20]. However, it is not easy to deliver large (highly charged) molecules to proper cells *in vivo*. Nevertheless, it can be carried out for liver, making HCV an ideal therapeutic target. HCV's IRES is well conserved among HCV populations, therefore it is a suitable target for antisense drugs [16]. Ribozyme RPI.13919 and antisense oligonucleotide ISIS-14803 were prepared, but with limited success.

5.2 RNA interference approach

RNA interference (RNAi) is a little bit more complex process starting with an endonuclease enzyme Dicer, which cleaves dsRNA (double-stranded RNA, usually viral) or pre-microRNA into 20-25 nucleotides long oligonucleotides, with few overhanging bases on each side. Products are called siRNAs, or miRNAs [21]. Dicer then induces creation of RISC - a complex consisting of guide siRNA and Argonaute (endonuclease) - which degrades the part of mRNA complementary to the guide siRNA strand. It is also supposed that process could be amplified by RdRp using Dicer-made siRNAs as a template, making secondary siRNAs distinct in structure [22]. miRNAs are single stranded, 21-23 nucleotides long, and complementary to parts of mRNA [23]. It is encoded by a non-coding RNA strand after transcription. In cytosol, it is processed by the Dicer into the miRISC. Further, so-called RITS complex doesn't act on RNA, but it alters histones and DNA by methylation or induction of heterochromatin formation (a tightly packed DNA with limited transcription [24]). RNAi method proved to be highly effective for inhibition of HCV *in vitro*. But it is not exploitable *in vivo* because of a delivery issue [16].

5.3 Immunomodulatory agents

To induce acute immunity reaction, antagonists of Toll-like receptors (TLRs) are made synthetically. Such receptors detect presence of pathogens like bacteria, viruses, and parasites through recognition of their molecular patterns. They are expressed by immune cells [16]. Ten TLRs are recognized, while TLR-3 (dsRNA), TLR-7 and TLR-8 (ssRNA) and TLR-9 (special DNA sequence, CpG) are sensitive to presence of viral nucleic acids. Their alert leads to inflammation and microbicidal reaction. TLR-7 and TLR-9 antagonists are expressed by B cells and plasmacytoid dendritic cells (PDCs), which can produce high levels of type 1 IFN (interferon, protein of immune system) and other secondary effects. Short, synthetic oligonucleotides containing CpG can be used to induce TLR-9 reaction and thus activate PDC. In the similar manner, TLR-7 reacts to compounds structurally related to nucleic acids.

Chapter 6 - Molecular Dynamics

6.1 Basics

Molecular dynamics method describes studied phenomenon on the particle basis (not necessarily literary particles, but some elementary parts of the system). Newton's motion equations are solved numerically [25]. The primary goal is to track trajectories of particles. If other parameters are needed, it is possible to calculate average values over the period of time for a trajectory in the phase space. Principles and techniques of molecular dynamics are demonstrated on the simple system of argon atoms, interacting only by weak interaction (see Appendix A), described by Lennard-Jones potential (σ and ε are constants):

$$E_{L-J} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right). \qquad (6.1)$$

Atoms were put into equilibrium positions (resembling vibrating crystal), while another atom hit the lattice. Newton's equations of motion were solved by the Verlet algorithm using Periodic Boundary Conditions.

6.2 Working area - Periodic Boundary Conditions

Basic is to determine a working area. In case of closed systems, system might be put into the working area as a whole [25]. More useful approach in modeling of large systems (like biomolecules surrounded by water molecules) is to use the sub-section of the studied system. First, the shape of the working area is determined (box, sphere, cylinder etc.) depending on the symmetry of the problem. In case of no symmetry, the best choice is a box-shaped working area. The size of the working area is derived from available hardware resources. Since long range ($\sim 1/r^2$) and short range ($\sim 1/r^7$) forces may be present, the size of the working area should be determined in a manner, that a particle in the centre of the working area is not influenced by particles outside of the working area. In other words, forces should be negligible at the half-length of the box's side. Then, boundary conditions must be set up [25]. If the working area is just a section of the whole system, periodic boundary conditions are usually introduced. It means that if a particle leaves the working area through one side, it enters again through the opposite side keeping its direction and speed. This works only for systems without a flux.

```
//periodic boundaries
if(pos == 2)
//gives margin for z axis, necessary for "hit scenario"
{//if(assembly[atom][pos] < 0)assembly[atom][pos] = assembly[atom][pos] + side2;
//lower periodic boundary, inactive during hit scenario
if(assembly[atom][pos] < 0)
{assembly[atom][2] = 0;
//base atoms are immobile in z direciton, because work area is only section near the
surface
assembly[atom][vel]= 0;}
if(assembly[atom][pos] > side2)assembly[atom][pos] = assembly[atom][pos] - side2;};
//upper periodic boundary
if(pos != 2)
//classical periodic boundaries for x, y axis
{if(assembly[atom][pos] < 0)assembly[atom][pos] = assembly[atom][pos] + side;
```

```
if(assembly[atom][pos] > side2)assembly[atom][pos] = assembly[atom][pos] - side;};};}
//end of periodic boundaries
```

Figure 6.1: Section of the md.cpp, regarding periodic boundaries; assembly – an array storing position and velocity for each atom, side(2) – length of side of work area (z axis)

6.3 Forces in Periodic Boundary Conditions

Another step in MD simulations is computation of imposed forces. For a certain particle, influence of all other particles should be taken into account. Some particles laying near the boundary of the working area are more influenced by the particles outside than inside [25]. This issue is solved by use of fictive systems periodically surrounding the central box. Following this, it is obvious that regardless the position of the particle, it is influenced by the same number of particles. This periodical approximation is effective only in periodic structures, or in working areas with rather large amount of particles.

```
void get_forces(float box_side,const int particles, double force[][3], double
pack[][6],float cut_off,float * potential)
//gathers the forces upon each particle, takes box side, number of particles, force
array and master array, cut-off distance, potential pointer
{float box_side2 = box_side + 2.0;   //box side for z in "hit scenario"
float limit(box_side/2.0);           //half of the box side
float limit2((box_side+2.0)/2.0);    //half of the box side for z axis in "hit
scenario"
float cut_off_sq=pow(cut_off,2);     //force limited reach
double distance[3] = {0,0,0};        //array stores distance of two particles
double distance_square(0);
double radial_force(0);              //radial part of force computation, to get force
in a direction, it is multiplied by a distance in that direction

double *force_pointer = &force[0][0];
populate_with_zeros(force_pointer,particles*3);     //old forces vector is scrapped

for(int up=0;up < particles;up++)           //goes through all particles
{for(int down=up+1;down < particles;down++) //goes through all particles higher than
first one
{for(int coor=0;coor < 3;coor++)            //goes through spatial coordinates
{distance[coor]=pack[up][coor]-pack[down][coor];

if(coor != 2)         //standard periodic force boundary conditions for x, y axis

{if (distance[coor] < (-1)*limit)distance[coor] = distance[coor] + box_side;
if (distance[coor] > limit)distance[coor] = distance[coor] - box_side;}

if(coor == 2)         //special rules for z axis in "hit scenario"
{if (distance[coor] < (-1)*limit2);
if (distance[coor] > limit2) distance[coor] = distance[coor] - box_side2;}}
distance_square = get_distance(distance);

if (distance_square <= cut_off_sq)   //cut-off condition
{*potential = *potential + pow(distance_square,(-6)) - pow(distance_square,(-3));
//calculates Lennard-Jones potential
radial_force = pow(distance_square,(-7)) - 0.5*pow(distance_square,(-4));
//calculates forces from potential
for(int coor=0;coor < 3;coor++)
{distance[coor]=distance[coor] * radial_force;      //changes distance to a force
force[up][coor]=force[up][coor] + distance[coor];   //updates 'up' particle summary
force by 'up-down' particle interaction
force[down][coor]=force[down][coor] - distance[coor];}}    //updates 'down' particle
summary force by 'up-down' particle interaction
else;}}
}
```

Figure 6.2: Section of the forces.h, regarding force computation

6.4 Numerical solution of Newton's equations of motion

Several numerical algorithms can be used to solve Newton's equations of motion for each atom *i*:

$$\vec{F}_i = m_i \vec{a}_i \quad \text{where } \vec{a} = \frac{d^2 \vec{r}}{dt^2}. \qquad (6.2)$$

Euler algorithm

Euler method is the simplest one, but also the least accurate. Accuracy is limited to the order of Δt:

$$\begin{aligned} \vec{r}_i^{\,k+1} &= \vec{r}_i^{\,k} + \vec{v}_i^{\,k+1}\Delta t + \frac{1}{2m_i}\vec{F}_i^{\,k}\Delta t^2 \\ \vec{v}_i^{\,k+1} &= \vec{v}_i^{\,k} + \frac{1}{m_i}\vec{F}_i^{\,k}\Delta t \\ \vec{F}_i^{\,k+1} &= \ldots \end{aligned} \qquad . \qquad (6.3)$$

Verlet algorithm

Verlet algorithm is more accurate than Euler method, yet, it has two major limitations [25]. New forces are computed before new velocities are gained. Therefore, it can't be used for forces dependent on the velocity (such as Lorentz force). Other problem is that it uses both new and old forces, so the memory needs are higher:

$$\begin{aligned} \vec{r}_i^{\,k+1} &= \vec{r}_i^{\,k} + \vec{v}_i^{\,k+1}\Delta t + \frac{1}{2m_i}\vec{F}_i^{\,k}\Delta t^2 \\ \vec{F}_i^{\,k+1} &= \ldots \\ \vec{v}_i^{\,k+1} &= \vec{v}_i^{\,k} + \frac{1}{2m_i}\left(\vec{F}_i^{\,k} + \vec{F}_i^{\,k+1}\right)\Delta t \end{aligned} \qquad . \qquad (6.4)$$

Leap-frog algorithm

Leap-frog method is of the same accuracy as the Verlet algorithm [25]. Interestingly, it needs initial velocities and positions in two different times. This problem is often solved by a computation of initial velocities by another algorithm:

$$\begin{aligned} \vec{r}_i^{\,k+1} &= \vec{r}_i^{\,k} + \vec{v}_i^{\,k+1/2}\Delta t + \frac{1}{2m_i}\vec{F}_i^{\,k}\Delta t^2 \\ \vec{F}_i^{\,k+1} &= \ldots \\ \vec{v}_i^{\,k+3/2} &= \vec{v}_i^{\,k+1/2} + \frac{1}{m_i}\vec{F}_i^{\,k+1}\Delta t \end{aligned} \qquad . \qquad (6.5)$$

Algorithms like Runge-Kutta or Predictor-Corrector are not very useful in the context of the MD. Their computation time is substantially higher (multiple force calculations for one time step are needed). Moreover, they don't provide stable trajectories in the phase space even for the simple harmonic oscillator.

```
//NEW POSITION COMPUTATION

for(int base_part=0;base_part<bottom;base_part++)   //set base particles to be rigid
in z-direction
{assembly[base[base_part]][5]=0;
forces[base[base_part]][2]=0;}

for(int atom=0;atom < n_particle;atom++)
{for(int pos=0;pos < 3;pos++)        //goes through coordinates
{int vel = pos+3;                    //coordinate corresponding velocity
assembly[atom][pos] = assembly[atom][pos] + assembly[atom][vel]*time_step +
forces[atom][pos]*time_step_sq*0.5;  //position update
assembly[atom][vel] = assembly[atom][vel] + forces[atom][pos]*time_step*0.5;
       //first part of velocity update (according to Verlet)
       .
//(periodic boundaries are applied, forces are computed)
       .
//FINISH VELOCITY UPDATE
for(int base_part=0;base_part<bottom;base_part++)   //set base particles to be rigid
in z-direction
{assembly[base[base_part]][5]=0;
forces[base[base_part]][2]=0;}

for(int atom=0;atom < n_particle;atom++)
{for(int vel=3;vel < 6;vel++)
assembly[atom][vel] = assembly[atom][vel] + forces[atom][vel-3]*time_step*0.5;};

//VELOCITY UPDATED
```

Figure 6.3: Section of the md.cpp file, regarding Verlet algorithm

6.5 Time step

Another issue is to determine a time step. It should be noted that processes faster than the time step can't be described [25]. Therefore, according to physical properties of the system, it is necessary to determine processes of interest with the highest frequency, and to set the time step accordingly. Sometimes, this is not possible because the chosen time step short enough for one process would make computation too long for another process of interest.

6.6 Initial conditions – initial positions, initial velocities

Initial positions of atoms are usually taken from X-ray experiments (even in the case of biomolecules). However, initial velocities can't be obtained from any experiment. Thus they must be generated by the statistical distribution [25]. Especially for small amount of particles, this might be inaccurate. In that case, the system is slowly thermalized to reach equilibrium.

```
void set_up(float box_side,const int particles,double pack[][6],float temp,float
temp_scale, long int *bottom) //sets up the particle lattice
{
long int particles_per_side_sq = particles_per_side*particles_per_side;
*bottom = 2*particles_per_side_sq;
double distance = minimal_potential_distance;       //distance among particles in
lattice

for(int toggle=0; toggle < 4; toggle++)             //construction of four parallel
lattices
{float one(0), two(0), three(0);                    //numbering of coordinates
long int x_coor(0), y_coor(0), z_coor(0);           //indexing the coordinates
if(toggle == 1)one = 0.5;
if(toggle == 2)two = 0.5;
if(toggle == 3)three = 0.5;
```

```
for(int dot=0;dot < particles;dot++)        //goes through 1/4 of particles and
constructs their array
{int drive(toggle*particles + dot);         //numbering particles in original array
(which has 4x more particles)
for(int bit=0;bit < 6;bit++)                //going through the positions and
velocities
{
switch(bit)
{
case 0: pack[drive][bit]=(x_coor+one+two)*distance; //assigns shifted coordinates, so
to make close lattice
break;
case 1: pack[drive][bit]=(y_coor+one+three)*distance;
break;
case 2: pack[drive][bit]=(z_coor+two+three)*distance;
break;
case 3:                                     //velocities are not updated here
case 4:
case 5:break;
};
};

x_coor = (x_coor++)%particles_per_side;     //skips to the next strand of
atoms when one is filled
if(x_coor == particles_per_side)x_coor=0;
y_coor = (((dot+1)/(particles_per_side))%particles_per_side);
if(y_coor == particles_per_side)y_coor=0;
z_coor = (dot+1)/particles_per_side_sq;
if(z_coor == particles_per_side)z_coor=0;
};
};
maxwell(pack,particles*4,temp,temp_scale);  //assigns Maxwell distributed
velocities
}
```

Figure 6.4: Section of the setup.h file, regarding initial lattice (quazicrystal) conditions

```
//assigns maxwell speed didstribution to the 3D cloud
//takes array of particles, number of particles, and temperature
void maxwell(double array[][6],int num_of_particles,float heat,float
temperature_scale)
{
srand ( time(NULL) ); //takes time as randomizer seed

double velocity_1(0),velocity_2(0),aid_1(0),aid_2(0),velocity_sq(0),radius(0);


for(int dot=0;dot < num_of_particles; dot = dot+2)  //randomizes velocities according
to maxwell
{
do {
aid_1 = (double)rand()/RAND_MAX;            //random number (0,1), evenly
distributed
aid_2 = (double)rand()/RAND_MAX;

velocity_1 = 2.0 * aid_1 - 1.0;
velocity_2 = 2.0 * aid_2 - 1.0;
velocity_sq = velocity_1*velocity_1+velocity_2*velocity_2;
}while(velocity_sq >= 1);                   //condition upon random numbers

radius = -2.0 * log(velocity_sq)/(velocity_sq);     //explicit formula to generate
Maxwell distribution from even distribution
int coordinate = (dot*3)/num_of_particles;
array[dot][coordinate+3]= velocity_1* sqrt(radius);        //Maxwell velocity for one
particle
array[dot+1][coordinate+3]= velocity_2* sqrt(radius);      //Maxwell velocity for
another particle
};

double kinetic_energy(0);           //initial overall energy
double momentum[3]={0,0,0};         //initial overall momentum

for(int direction=0;direction < 3;direction++)
```

```
{for(int dot=0;dot < num_of_particles;dot++)
momentum[direction]=array[dot][direction+3]+momentum[direction];
momentum[direction]=momentum[direction]/num_of_particles;
for(int dot=0;dot < num_of_particles;dot++)
{array[dot][direction+3]=array[dot][direction+3] - momentum[direction];
kinetic_energy = kinetic_energy + array[dot][direction+3]*array[dot][direction+3];};
};

scale_velocity(temperature_scale, heat, kinetic_energy, array,num_of_particles);
//call to the velocity scaling function

}
```

Figure 6.5: Section of the setup.h file, regarding Maxwell distribution of initial velocities

6.7 AMBER force field for biomolecules

In classical (non-quantum) mechanics, interactions among particles can be classified as bonded and non-bonded (electrostatic and van der Waals interactions). Bond related energies are given for covalently bound atoms, and can be calculated for example from the following formula:

$$E_{bond} = \sum_{bonds} \frac{k_b}{2}(l - l_0)^2 + \sum_{angles} \frac{k_a}{2}(\vartheta - \vartheta_0)^2 + \sum_{torsions} k_d\left(1 + \cos(n\varphi - \varphi_0)\right) + \sum_{improper} k_i(\omega - \omega_0)^2 \tag{6.6}$$

Energy terms are based upon harmonic oscillator model [19]. First and second term characterize bond's ability to stretch and bend according to the Hooke's law. The k is oscillator spring stiffness and parameters with "0" subscript correspond to the equilibrium. Charged atoms are subjects of Coulombic interactions:

$$E_{Coulomb} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}. \tag{6.7}$$

Further, van der Waals interactions, based on definite sizes of atoms, are given by Lennard-Jones potential:

$$E_{L-J} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right). \tag{6.8}$$

In MD simulations, parameters in energy terms mentioned above are set either empirically or from *ab initio* calculations.

6.8 Simple argon system output

Stored simulated trajectories of individual particles can be later used to examine a behavior of the system. Several screenshots from the argon simulation are shown in figure 6.6. A particle hits the surface of an argon lattice; the lattice has periodical boundaries in dimensions parallel with surface, and rigid atoms at the base of the perpendicular dimension. Qualitative conclusion can be made from the result: lattice

absorbs kinetic energy of impact particle and stores it in vibrations of itself. The impact particle, having lost its kinetic energy, is attracted by the weak interaction to the lattice. Vibrations and arrangement changes can be observed in the lattice.

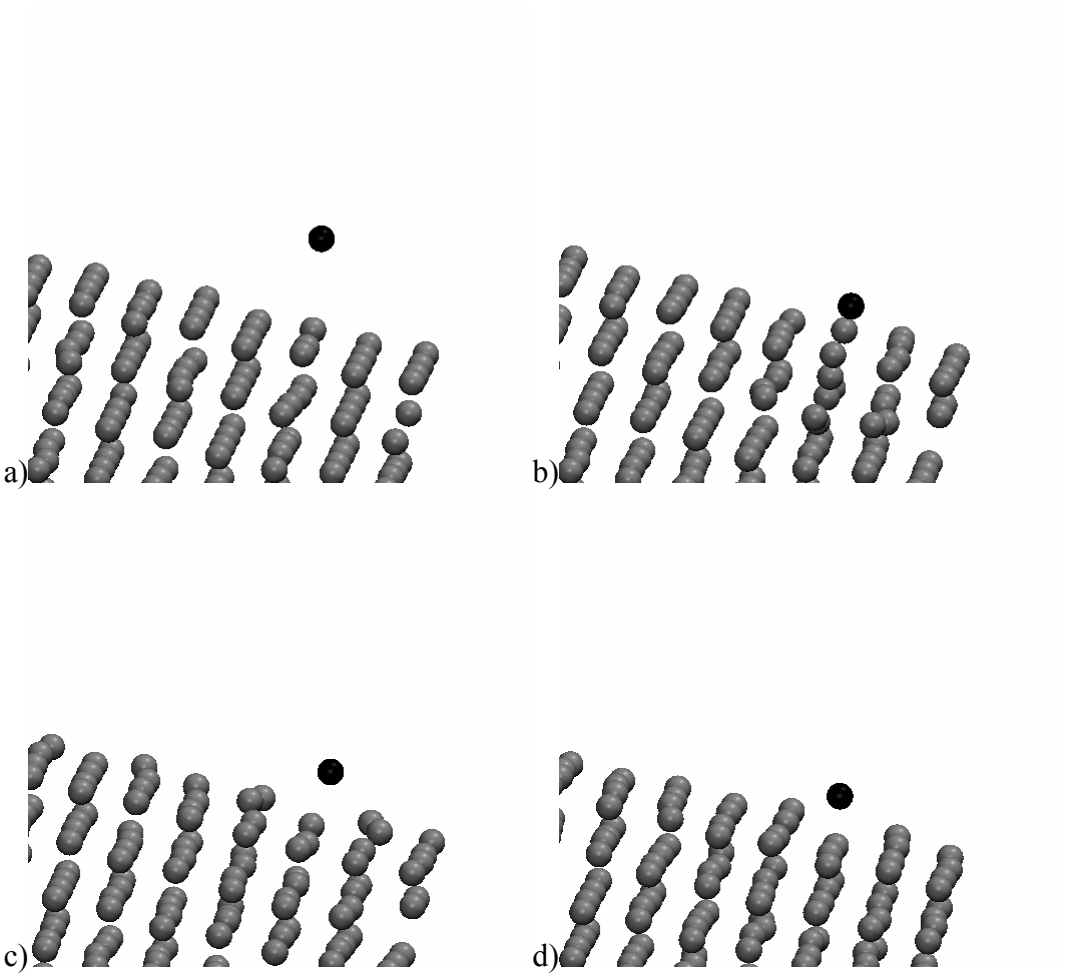


Figure 6.6: a) atom is closing on the intact cluster b) Lennard-Jones interaction disrupts the lattice at the impact c) disturbance is transmitted through the lattice as particle bounces off d) particle lost kinetic energy and is attracted by Lennard-Jones interaction to the lattice

Chapter 7 - Results

7.1 Simulated systems

As mentioned above, the HPMPA (9-(3-hydroxy-2-phosponyl-methoxy-propyl)adenine) structure is a nucleotide analog. The modification of HPMPA to HPMPG is reasonable, because first two residues of the initiation complex of HCV RdRp are necessarily guanines. In fact, HCV RdRp complex with PMEG (2-phospohonyl-methoxyethyl guanine) was initially examined to see interactions of the enzyme with the elementary molecule. Then, it was advanced to PMPG (9-(2-phospohonyl-methoxypropyl)guanine), which was run in two simulations with slightly different initial conditions. That was meant to point out randomness in MD simulations. Finally, a complex of HCV RdRp with HPMPG (9-(3-hydroxy-2-phosponyl-methoxy-propyl)guanine) was simulated. Structures of mentioned compounds are in figure 7.1. At the beginning of MD simulations, a template fragment consisting of four nucleotides (D7, D8, D9, and D10) was placed in the template tunnel. Two nucleotides (R2 and PMEG/PMPG/HPMPG) were positioned in the catalytic site (see fig 7.2a where thumb domain is hidden and fig 7.2b). Template DNA was used instead of proper RNA as it was taken from the X-ray structure of the bacteriophage polymerase. Hydrogen bonds established between template/nucleotides, two nucleotides, template/enzyme, and enzyme/nucleotides were analyzed in detail. Further, ionic bonds involving Mg2+, selected torsion angles and sugar puckering were examined.

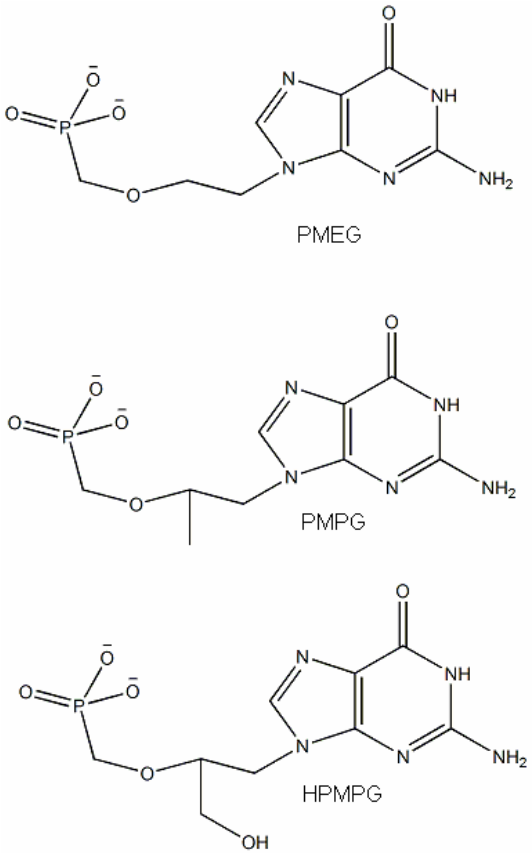


Figure 7.1: Structure of studied acyclic phosphonate analogs of nucleotides

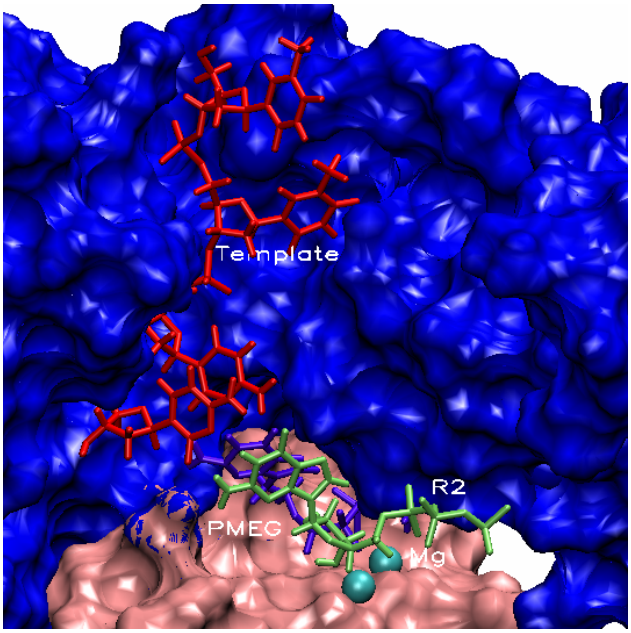


Figure 7.2a: Active site cavity of HCV RdRp with template; fingers domain – blue, palm domain – pink, template – red, PMEG – violet, R2 – lime, Mg^{2+} - cyan

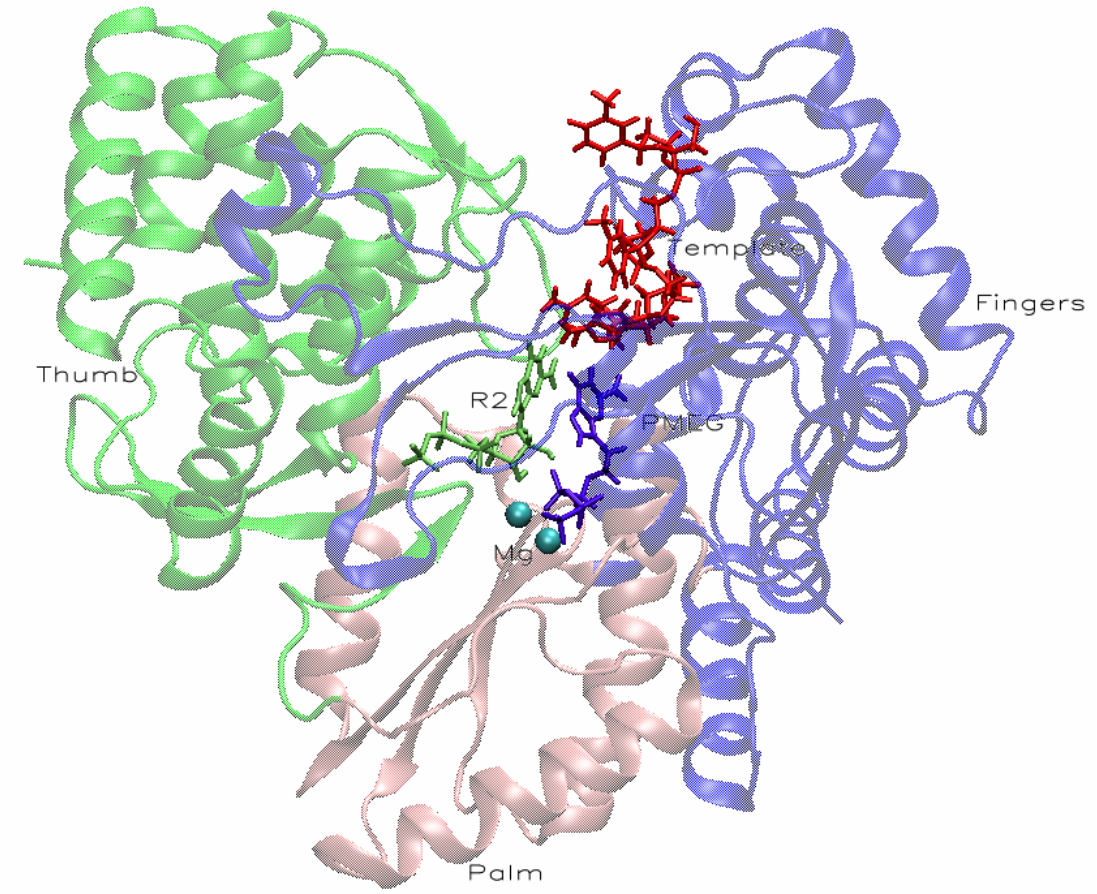


Figure 7.2b: Initiation complex of HCV RdRp; template – red, R2 – lime, PMEG – violet, Mg^{2+} – cyan, fingers domain – blue, palm domain – pink, thumb domain – green

7.2 PMEG

7.2.1 Hydrogen bonds between template and incoming nucleotides

Hydrogen bonds are illustrated in figures 7.3 and 7.4. Graphs show the hydrogen bond length observed during the MD simulation, and thus its stability (fig 7.5). Table 7.1 sums up some of the statistics regarding these bonds. Item 'Occupied' refers to amount of simulation time, during which the bond was realized. 'Distance' is an average distance of donor and acceptor atoms. 'Angle' is an angle between donor and acceptor at the point of hydrogen. Evaluating results, hydrogen bonding of template and nucleotides is stable over time. Table 7.1 reveals that binding with PMEG is negligibly weaker than the one with R2. This might be caused by lower stabilization effect induced by the enzyme, since the structural change is present in PMEG.

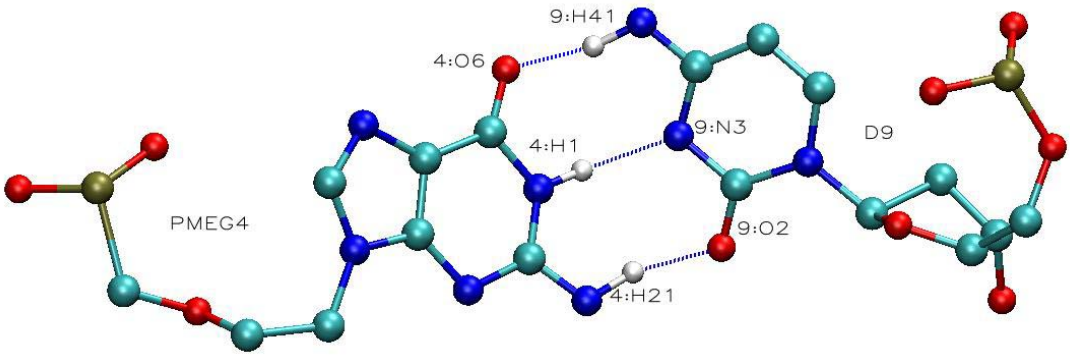


Figure 7.3: PMEG4 — D9; oxygen – red, phosphorus – gold, carbon – cyan, nitrogen –blue, hydrogen – white, hydrogen bond – blue-dashed line

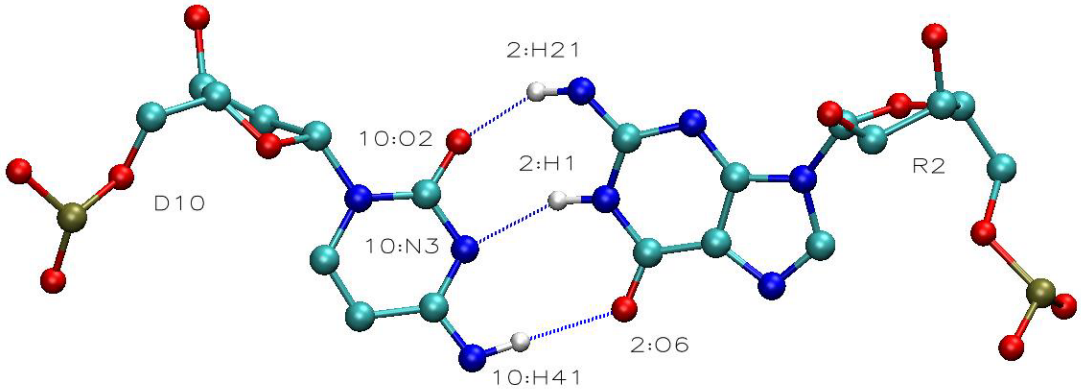


Figure 7.4: D10 — R2; oxygen – red, phosphorus – gold, carbon – cyan, nitrogen – blue, hydrogen – white, hydrogen bond – blue-dashed line

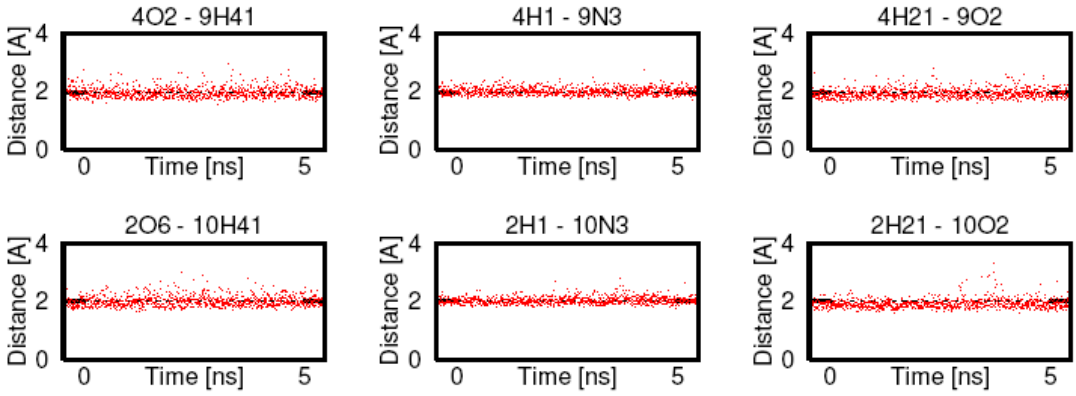


Figure 7.5: Length of hydrogen bonds of group 1

Table 7.1: Statistics of group 1 hydrogen bonds				
Donor	Hydrogen (Acceptor)	Occupied [%]	Distance [Å]	Angle [°]
10N3	2H1 (2N1)	100.00	3.000 (0.11)	19.60 (8.97)
10O2	2H21 (2N2)	99.90	2.888 (0.14)	18.27 (9.07)
2O6	10H41 (10N4)	99.80	2.983 (0.17)	16.90 (9.33)
9O2	4H21 (4N2)	99.70	2.909 (0.14)	20.02 (10.68)
10O2	2H1 (2N1)	99.20	3.366 (0.20)	43.86 (7.24)
9N3	4H1 (4N1)	98.70	2.916 (0.09)	31.72 (11.48)
4O6	9H41 (9N4)	94.30	2.830 (0.11)	35.49 (12.12)
9O2	4H1 (4N1)	93.10	3.606 (0.19)	47.61 (6.40)

7.2.2 Hydrogen bonds among nucleotides

Hydrogen bonds between R2 and PMEG were another field of study. Only two potential bindings belong to this group as shown in fig 7.6. This group is responsible for proper and stabile mutual distance of incoming nucleotides. According to results, it can be concluded that both nucleotides are bound to each other and remain in proximity at the catalytic site. 2H3T is highly mobile and practically not bound. Table 7.2 shows statistics of bonds. The situation would be worse, if rGTP and PMEG would switch places. Since later doesn't contain any hydroxyl group, it would hardly establish any stable arrangement, and complex would be highly volatile, regarding inter-nucleotide distances. This could cause steric interference and eventually abort replication.

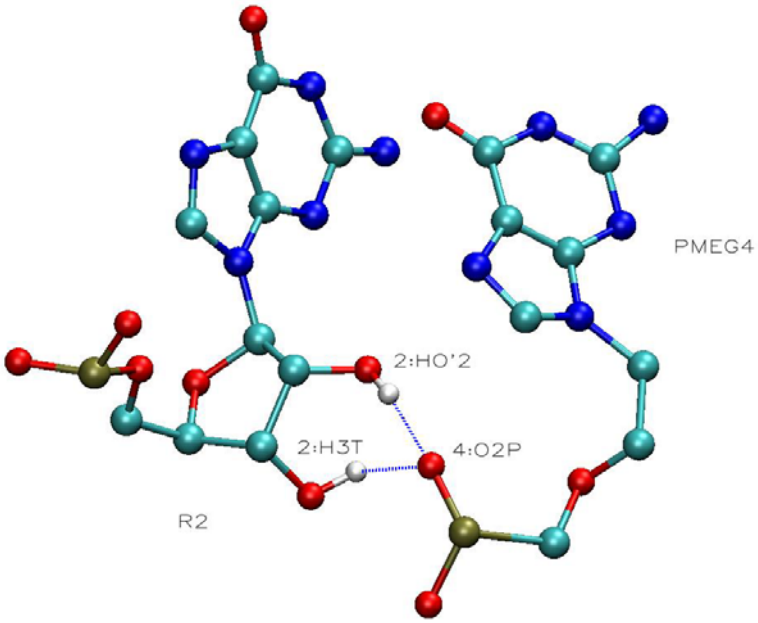


Figure 7.6: PMEG4 — R2; oxygen – red, phosphorus – gold, carbon – cyan, nitrogen – blue, hydrogen – white, hydrogen bond – blue-dashed line

Simulation reveals only one bond as efficient.

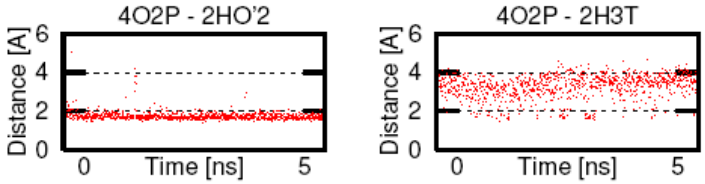


Figure 7.7: Length of hydrogen bonds of group 2

Table 7.2: Statistics of group 2 hydrogen bonds				
Donor	Hydrogen (Acceptor)	Occupied [%]	Distance [Å]	Angle [°]
4O2P	2HO`2 (2O2`)	98.90	2.653 (0.12)	17.11 (9.69)
4O2P	2H3T (2O3`)	9.00	2.711 (0.14)	30.24 (15.63)

7.2.3 Hydrogen bonds between template and enzyme

Stability of template within the enzyme structure is necessary for viral replication. Several closest residues were studied to determine their role in stabilization process. Figure 7.8 clearly explains which bonds were examined; later graphs (see Figure 7.9) show results. Ser106 has clearly stabilizing effect on the template. Little less stable are hydrogen bonds with Arg178, but because two bonds are close to each other, they also establish "crossed" interactions. Binding is carried out then by four hydrogen bonds (three of them are very persistent, as shown in table 7.3). This effect is making overall interaction of residues highly stabilizing, and is common for arginine side chains. Lys182 is a special case. Three hydrogen atoms are actually present on the Lys182 nitrogen (182NZ). Observed drop of distance (see Fig 7.9) is caused by conformational change. Hydrogen bond between nitrogen and carbon rotates, thus hydrogen bonding switches to another of hydrogen atoms (e.g. 182HZ3). The interaction is beneficial to overall stability. Gln456 and Lys108 bonds are getting more stable as simulation progresses, but their effect is limited. They may serve only as auxiliary buffers. In case of Lys108, instability is probably partially caused by the same "twist" effect, as in case of Lys182. Table 7.3 provides statistics of most occupied bonds. It is obvious that the most stabilized is the template backbone (especially in proximity of the terminal nucleotide which is anchored to R2). Otherwise, it seems that a slight interference (e.g. water flux) could rock the template strand and block the replication.

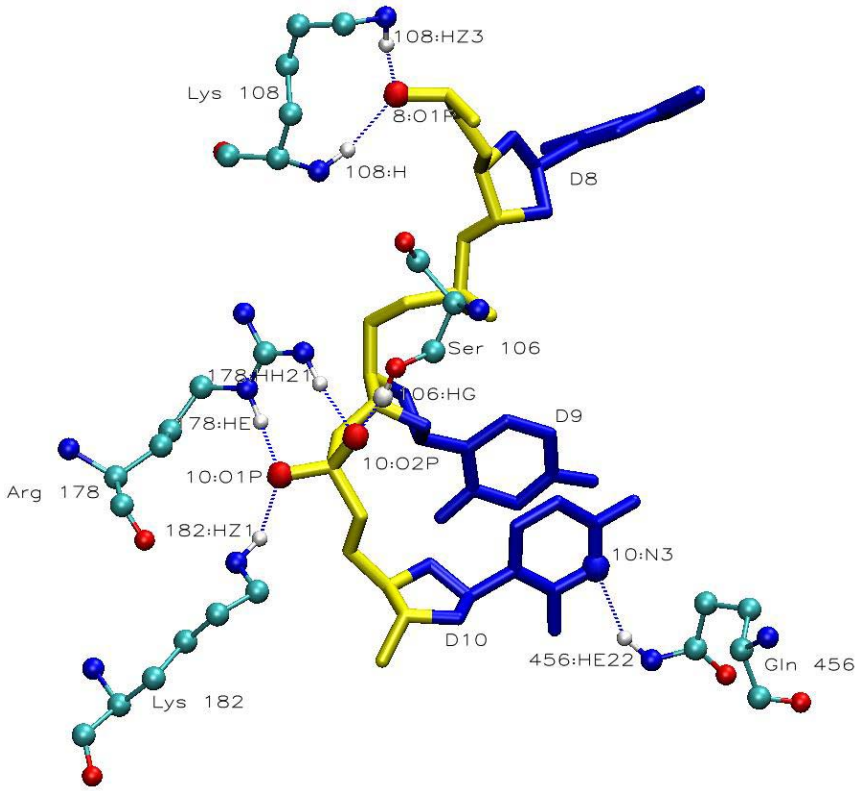


Figure 7.8: Enzyme — Template; oxygen – red, phosphorus – gold, carbon – cyan, nitrogen –blue, hydrogen – white, hydrogen bond – blue-dashed line, template backbone – yellow tube, template basis – blue tube

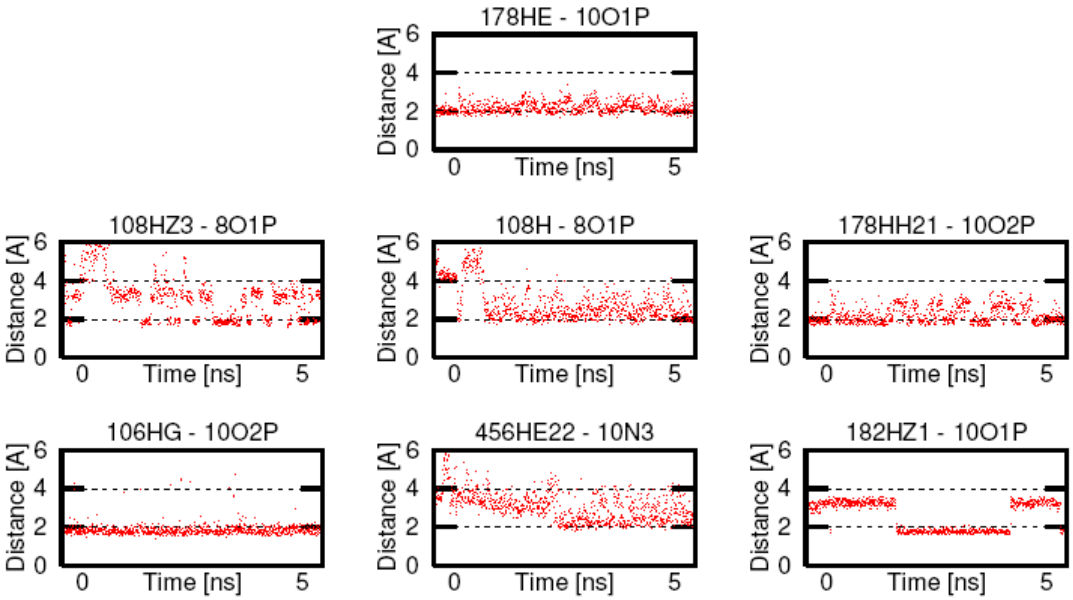


Figure 7.9: Length of hydrogen bonds of group 3

Table 7.3: Statistics of group 3 hydrogen bonds				
Donor	Hydrogen (Acceptor)	Occupied [%]	Distance [Å]	Angle [°]
10O2P	106HG (106OG)	98.50	2.727 (0.14)	20.56 (9.72)
10O2P	178HH21 (178NH2)	98.00	3.136 (0.35)	22.70 (13.05)
10O2P	178HE (178NE)	98.00	3.497 (0.23)	36.59 (14.27)
10O1	178HE (178NE)	94.90	3.052 (0.19)	31.96 (14.14)
8O1P	108H (108N)	71.20	3.233 (0.28)	30.39 (14.31)
10O1P	182HZ1 (182NZ)	46.30	2.719 (0.08)	22.92 (10.87)
10N3	456HE22 (456NE2)	34.50	3.180 (0.22)	36.02 (14.47)
8O1P	108HZ3 (108NZ)	34.20	2.832 (0.14)	22.61 (12.93)

7.2.4 Hydrogen bonds between enzyme and nucleotides

Proper positioning, conformation, and mutual distances of incoming nucleotides are crucial for successful initialization and elongation of the replication process. This is carried out by non-covalent interactions between nucleotide triphosphates and enzyme residues. The residues with the leading role in the process are depicted in Fig 7.10 and 7.11. Graphs in Fig 7.12 show mutual distances observed during the MD simulation. Table 7.4 provides statistics for hydrogen bonds mentioned in this section. Triphosphate tail is also considered because it is negatively charged, and thus suitable for interaction with enzyme's charged amino acids. It seems reasonable to ignore instabilities at the beginning of the simulation, since bonds conspicuously reached equilibrium over time, and remained stable for the rest of the simulation (e.g. Ser377 or Arg404). Bonds of Arg58 and Arg168 stabilize the PMEG's triphosphate tail. That is essential to establish phosphodiester bond with the previous nucleotide in the nascent strand. The same part for R2 was undertaken by hydrogen bonds with Arg404 and Arg396. This leads to conclusion, that two pair hydrogen bonds ("crossed bonds") from arginines are indispensable for the phosphate tail stabilization. The absence of ribose (in PMEG) is not problematic in this step, since this interactions influence triphosphate groups alone. Remarkable disturbance in bond length for Ser377 suggests that conformational changes (caused by alternations of torsion angles) lead to stable state over the period of time. Because this new state might be necessary for a proper work of HCV RdRp, we might conclude, that mutation of these residues could be unfavorable for enzyme efficiency. Hydrogen bond of Ser292 is very stable and is probably assisting the proper orientation of the guanine base of PMEG. This interaction should be the same for PMEG as for a proper nucleotide. On the other hand, Lysine residues 151 and 165 are very shaky suggesting rather inferior role of Lysine residues in the template stabilization.

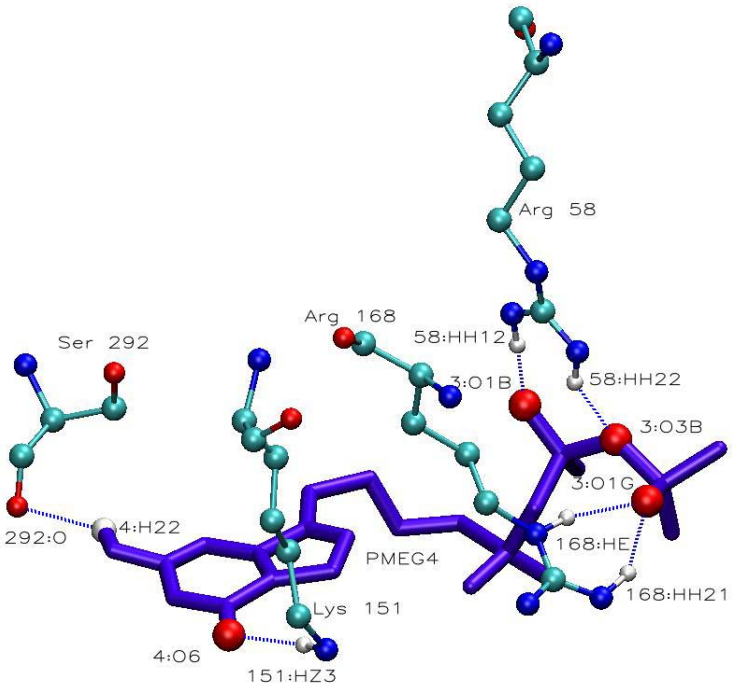


Figure 7.10: Enzyme — PMEG4; oxygen – red, carbon – cyan, nitrogen – blue, hydrogen – white, hydrogen bond – blue-dashed line, PMEG4 – violet tube

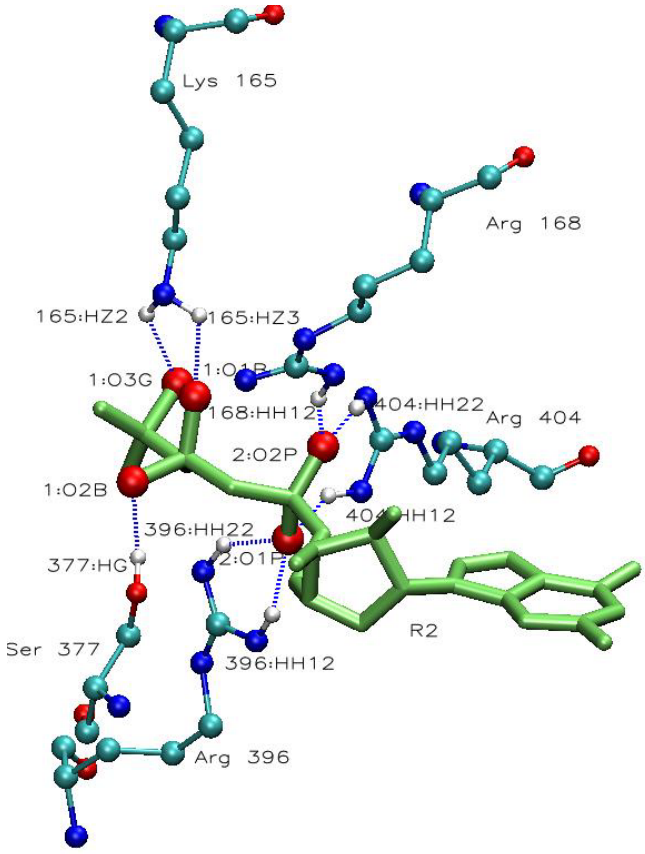


Figure 7.11: Enzyme — R2; oxygen – red, carbon – cyan, nitrogen – blue, hydrogen – white, hydrogen bond – blue-dashed line, R2 – lime tube

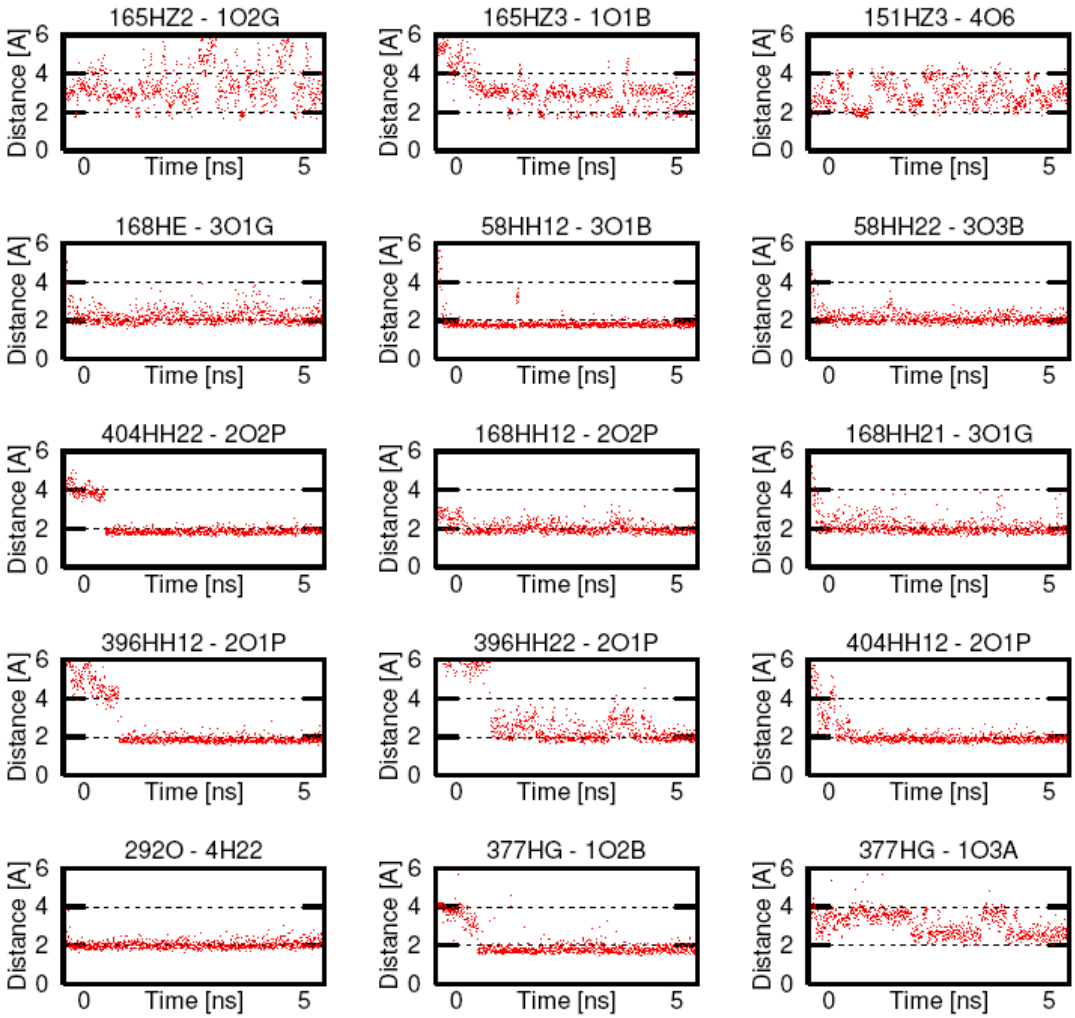


Figure 7.12: Length of hydrogen bonds of group 4

Table 7.4: Statistics of group 4 hydrogen bonds				
Donor	Hydrogen (Acceptor)	Occupied [%]	Distance [Å]	Angle [°]
2O2P	168HH12 (168NH1)	98.00	2.935 (0.22)	30.43 (12.50)
292O	4H22 (4N2)	97.00	2.897 (0.15)	34.93 (11.07)
3O1B	58HH12 (58NH1)	96.70	2.768 (0.11)	18.84 (9.51)
2O1P	404HH12 (404NH1)	91.70	2.902 (0.19)	20.63 (12.01)
3O1G	168HH21 (168NH2)	88.40	2.864 (0.17)	34.53 (11.60)
1O2B	377HG (377OG)	87.00	2.758 (0.25)	18.12 (12.05)
2O2P	404HH22 (404NH2)	84.00	2.823 (0.11)	18.63 (9.70)
3O1G	168HE (168NE)	83.20	2.937 (0.19)	39.48 (11.38)
2O1P	396HH12 (396NH1)	78.90	2.810 (0.13)	23.71 (10.30)
2O1P	396HH22 (396NH2)	73.80	3.056 (0.31)	38.00 (10.82)
2O1P	404HH22 (404NH2)	72.30	3.516 (0.34)	47.33 (12.24)
2O2P	404HH12 (404NH1)	71.70	3.510 (0.23)	50.62 (6.07)
3O3B	58HH12 (58NH1)	59.70	3.633 (0.25)	48.76 (7.24)
1O1B	168HH12 (168NH1)	58.70	3.359 (0.30)	49.30 (7.67)
1O3A	377HG (377OG)	40.60	3.359 (0.31)	48.18 (9.10)
1O3A	396HH22 (396NH2)	38.40	3.180 (0.35)	27.32 (12.81)

7.2.5 Ionic bonds involving Mg^{2+}

Presence of divalent ions in active sites of various polymerases, nucleases and transcriptases is considered to be an essential factor for its proper function. However, the exact role has not yet been determined. In this part, influence of Mg^{2+} ions on both enzyme and nucleotides was examined. Following interactions were suggested (see Fig. 7.13 and 7.14) and tried (see Fig. 7.15) during the MD simulation. It is clear, that Mg^{2+} ions play an important part in stabilization of the catalytic site. They tie acidic side chains of aspartic acid, so these chains can't interfere with hydroxyl groups on riboses of incoming nucleotides. Neither they can bind arginine side chains (for which they have high affinity), which are important for the triphosphate tails stabilization. As it can be noted, ionic interactions were stable during the whole MD simulation. Ionic interactions with nucleotides decreased mobility of the triphosphate tail of PMEG4, and bound the C3' hydroxyl group of R2 (residue to be added first), thus insuring its proper positioning. As it binds to C3' and not to C2' hydroxyl group, it does not discriminate between ribose and deoxyribose substrates. This feature, however, could lead to decreased incorporation of PMEG into the nascent strand, because it lacks necessary C3' hydroxyl group, and would be therefore in disadvantage, comparing to both rNTPs and dNTPs. Finally, blackout for the 5MG – 230OD2 interaction is caused by temporary shift of 230OD2 to 6MG, which was further supported by ionic bond of 5MG with 231O.

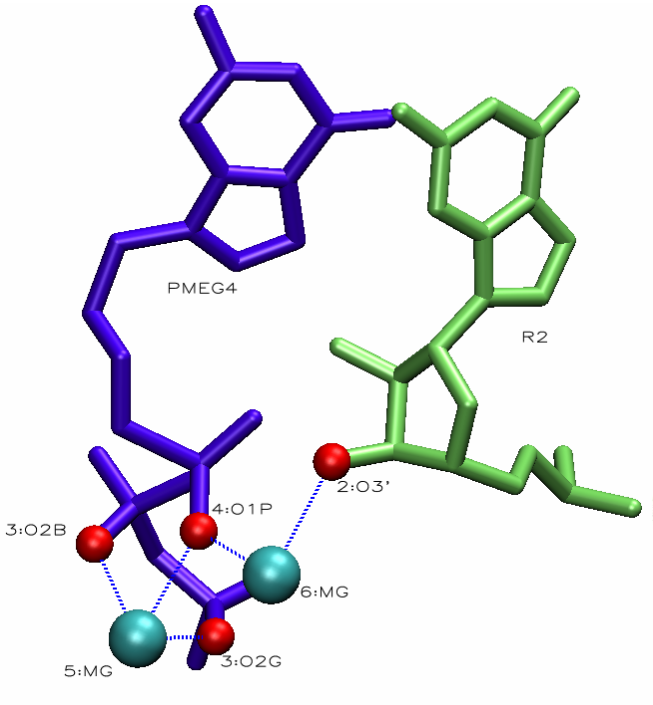


Figure 7.13: Nucleotide — Mg^{2+}; oxygen – red, magnesium – cyan, ionic bond – blue-dashed line, R2 – lime tube, PMEG4 – violet tube

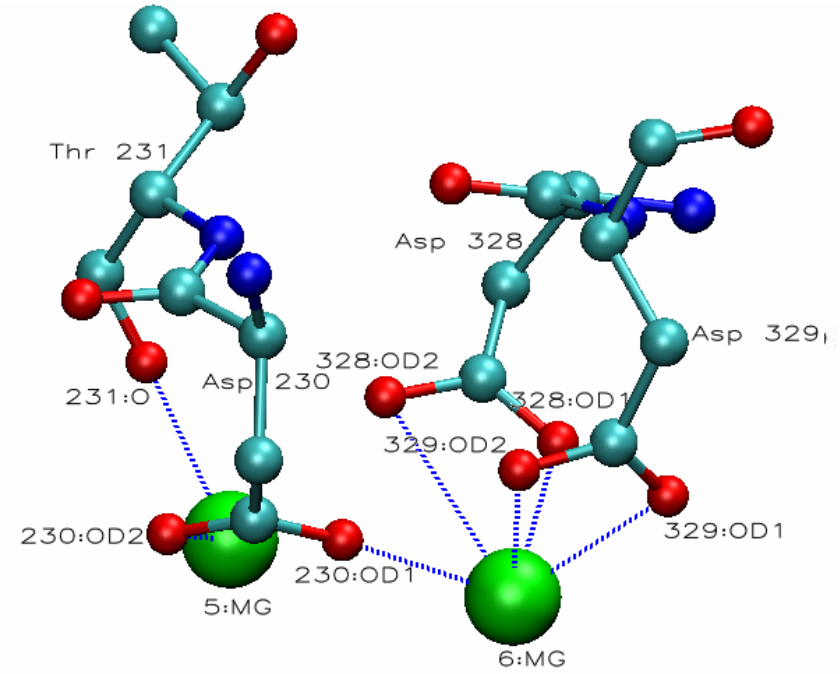


Figure 7.14: Enzyme — Mg^{2+}; oxygen – red, carbon – cyan, nitrogen – blue, hydrogen – white, magnesium ion – green, ionic bond – blue-dashed line

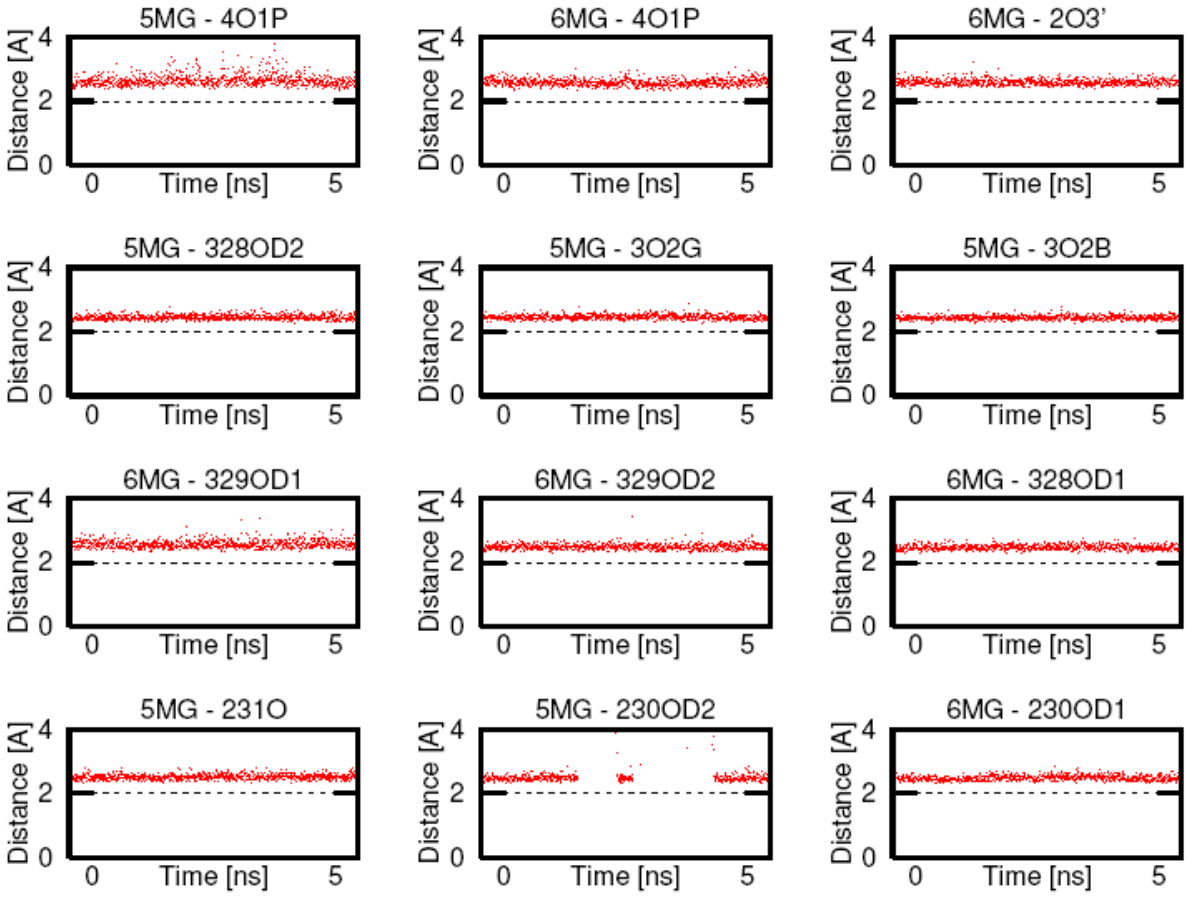


Figure 7.15: Length of ionic bonds

7.2.6 Torsion angles of backbone of PMEGt

Torsion angles of backbone of PMEG triphosphate were examined. Fig 7.16 explains the chosen naming. Following graphs show gathered results (see Fig 7.17). These graphs clearly show a high preservation of PMEG triphosphate conformation. It could be important for creation of the phospohodiester bond with a preceding nucleotide in the nascent strand.

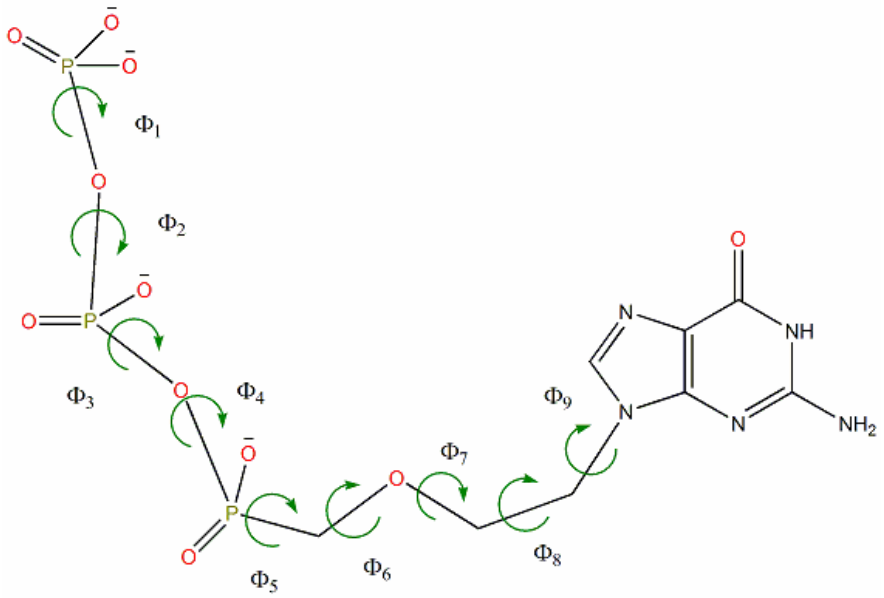


Figure 7.16: Φ_1 - O3G-PG-O3B-PB, Φ_2 - PG-O3B-PB-O3A, Φ_3 - O3B-PB-O3A-P, Φ_4 - PB-O3A-P-C9', Φ_5 - O3A-P-C9'-O9', Φ_6 - P-C9'-O9'-C7', Φ_7 - C9'-O9'-C7'-C6', Φ_8 - O9'-C7'-C6'-N9, Φ_9 - C7'-C6'-N9-C4

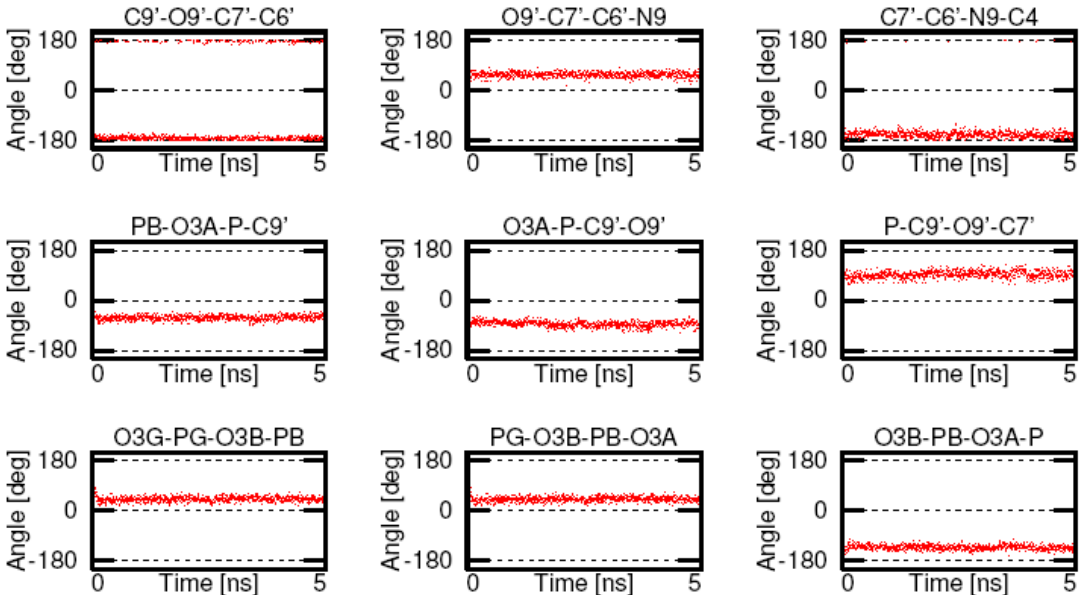


Figure 7.17: Torsion angles of PMEG triphosphate backbone

7.2.7 Sugar Puckering

The last step in evaluation of the PMEG MD simulation was sugar puckering of the ribose in rGTP as well as deoxyriboses in the template strand. Following graphs show results (see Fig. 7.18). R2 and D10 are displayed alongside as they are paired; rest of the dNs is displayed as they follow each other in the template strand. All nucleotides are in the C2'-endo configuration, typical for DNA in B form. Since R2 is not a part of any strand, it's deviation from more typical C3'-endo is not strange.

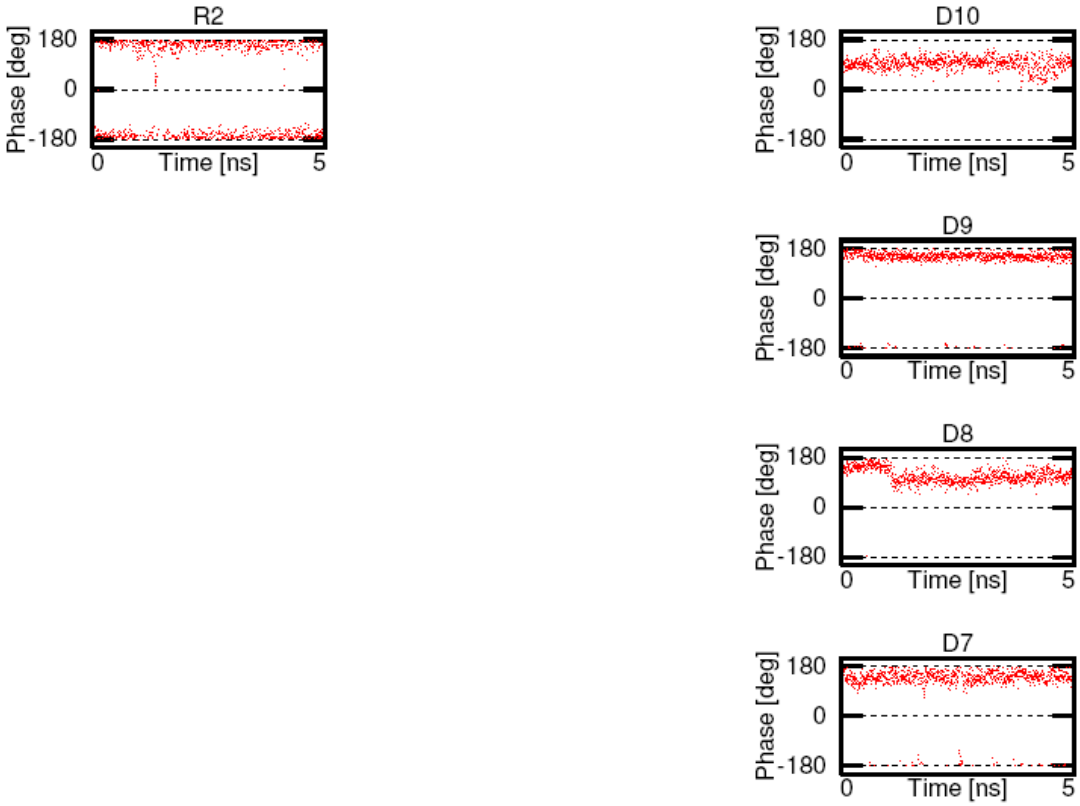


Figure 7.18: Sugar Puckering Pseudo-rotational angle

7.3 PMPG1

In this simulation, PMPG was used. It is only slightly altered from PMEG. The added chain is not charged. Results are basically very similar to PMEG.

7.3.1 Hydrogen bonds between template and incoming nucleotides

Change for PMPG has minimal influence upon hydrogen binding to the template D9.

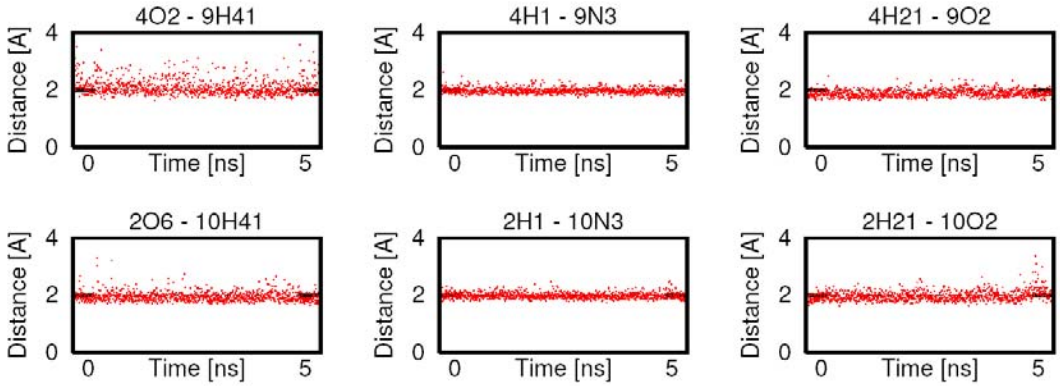


Figure 7.19: Length of hydrogen bonds of group 1

7.3.2 Hydrogen bonds among nucleotides

Change for PMPG, however, completely disrupted inter-nucleotide interactions.

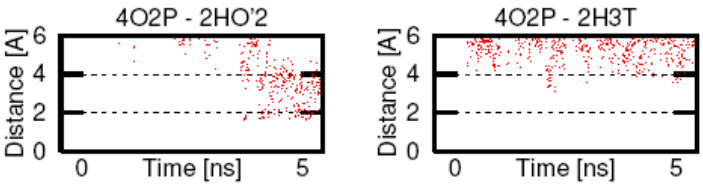


Figure 7.20: Length of hydrogen bonds of group 2

7.3.3 Hydrogen bonds between template and enzyme

Group 3 includes only interactions among the template and enzyme. Therefore, changing of PMEG for PMPG didn't affect the graphs (fig 7.21). Slight differences are caused mostly by randomness of MD simulations.

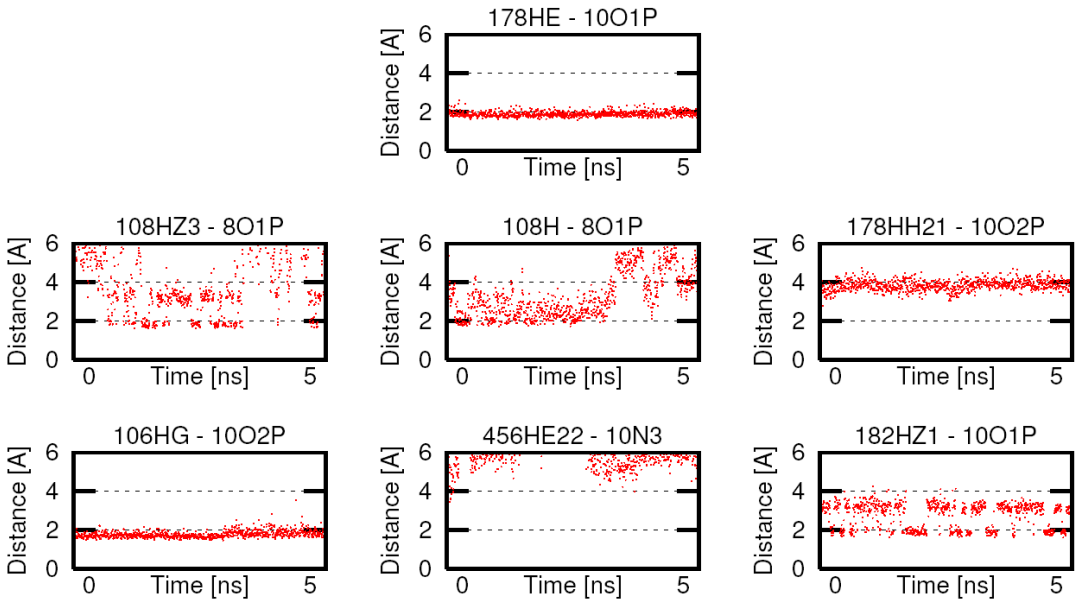


Figure 7.21: Length of hydrogen bonds of group 3

7.3.4 Hydrogen bonds between enzyme and nucleotides

This group of hydrogen bonds was most severely affected by alternation for PMPG (see fig 7.22). Most of the important bonds disappeared, especially bonds with Arg158, Arg396, and Arg58. Bonds with Arg404 as well as Ser377 seem to recover at the end of MD simulation. With regard to the new spatial arrangement, several new potential hydrogen bonds were proposed. While new bonds of Arg168 and Arg396 seem to be stable, Arg58 apparently couldn't find any suitable counterpart.

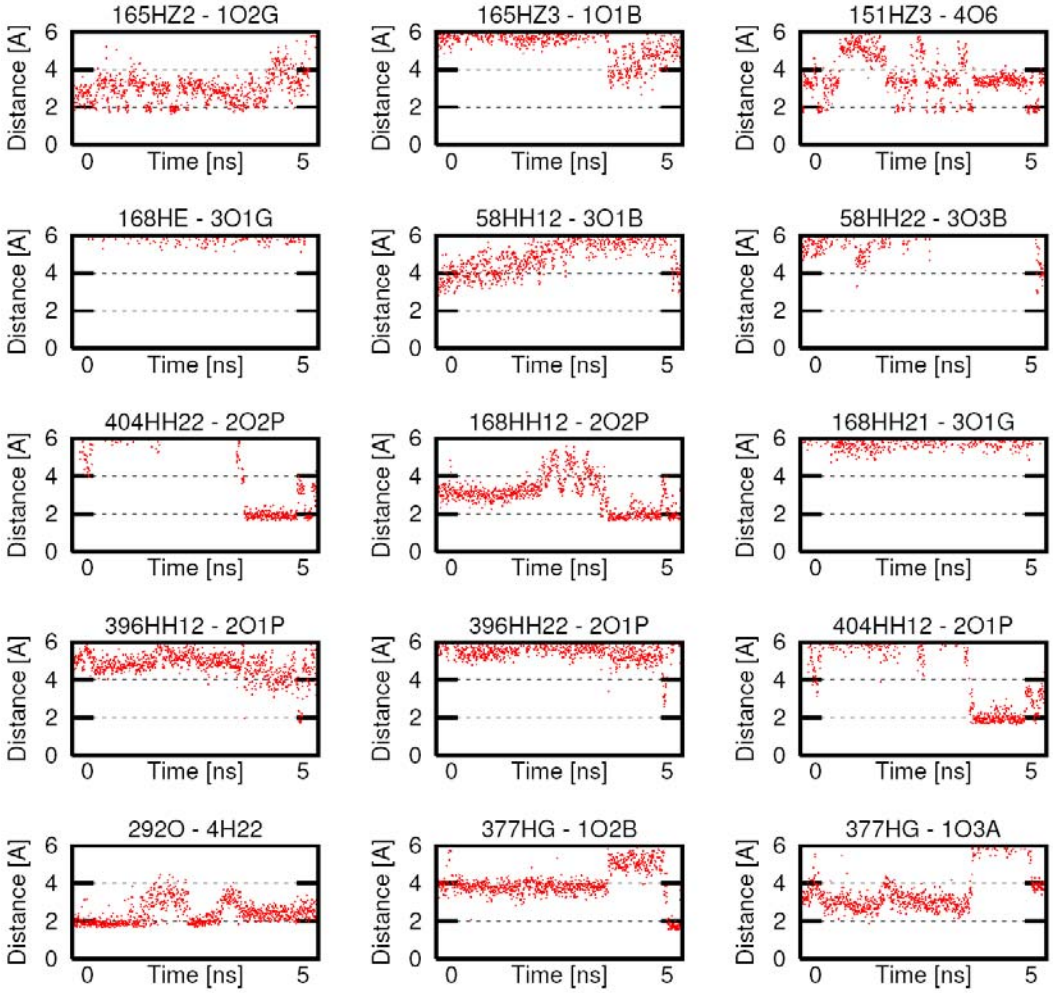


Figure 7.22a: Length of original hydrogen bonds of group 4

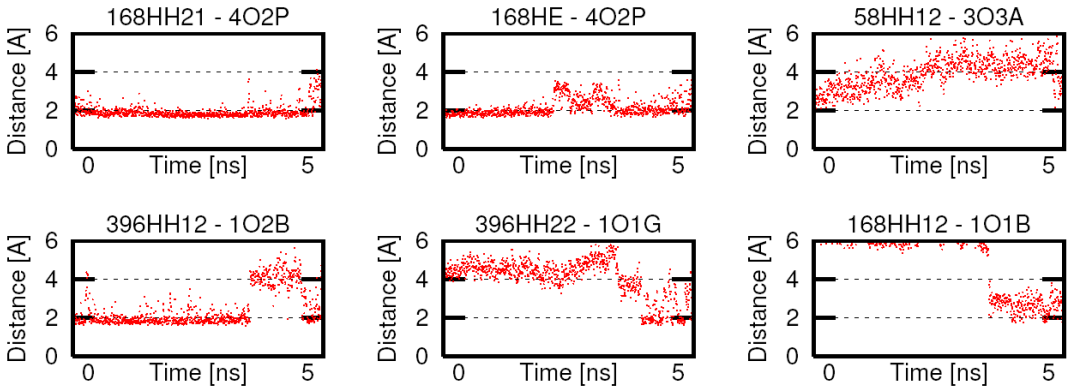


Figure 7.22b: Length of new hydrogen bonds of group 4

7.3.5 Ionic bonds involving Mg^{2+}

Behavior of ionic bonds with divalent magnesium ions remains the same. Two bonds fell apart, probably because of steric interference with PMPG. Nevertheless, the ionic interactions still stabilize enzyme, nucleotide, and PMPG.

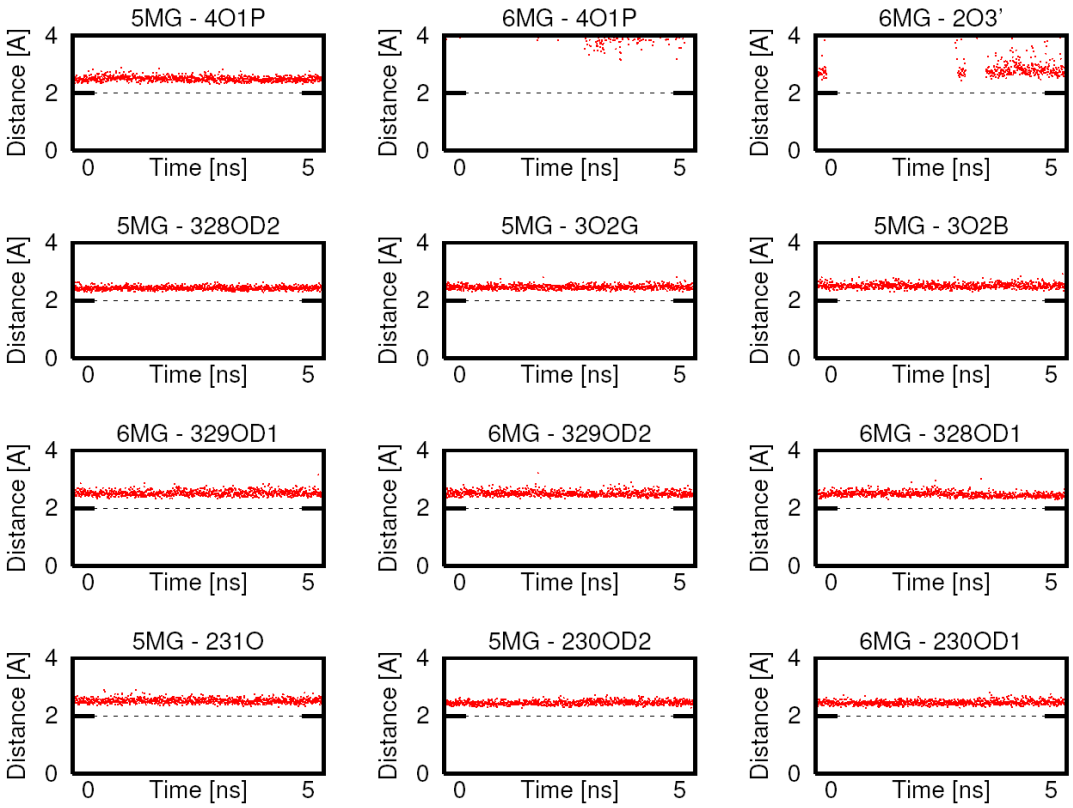


Figure 7.23: Length of ionic bonds of group 5

7.3.6 Torsion angles of backbone of PMPGt

In comparison with PMEGt, PMPGt undergoes conformational changes regarding its backbone. That is undoubtedly caused by interference of the additional methyl group with the enzyme.

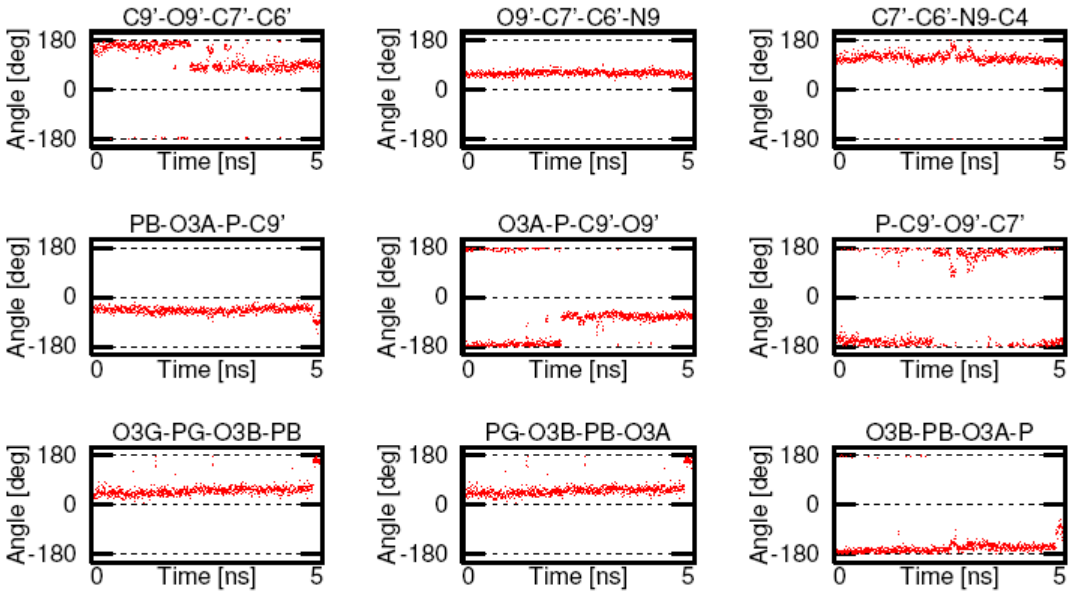


Figure 7.24: Torsion angles of PMPG triphosphate backbone

7.3.7 Sugar Puckering

Sugar puckering behavior is more turbulent than in the case of PMEG. While D7 and D8 pretty much hold the C2'-endo conformation, D10 is clearly in the C3'-endo conformation, and D9 is inconclusive. R2 is sharply shifting among the C2'-endo and C3'-endo conformations. The situation is probably caused by D9 and D10 acquiring the C3'-endo conformation and thus enforcing R2 to align with emerging A form structure (rather than B form observed in the case of PMEG).

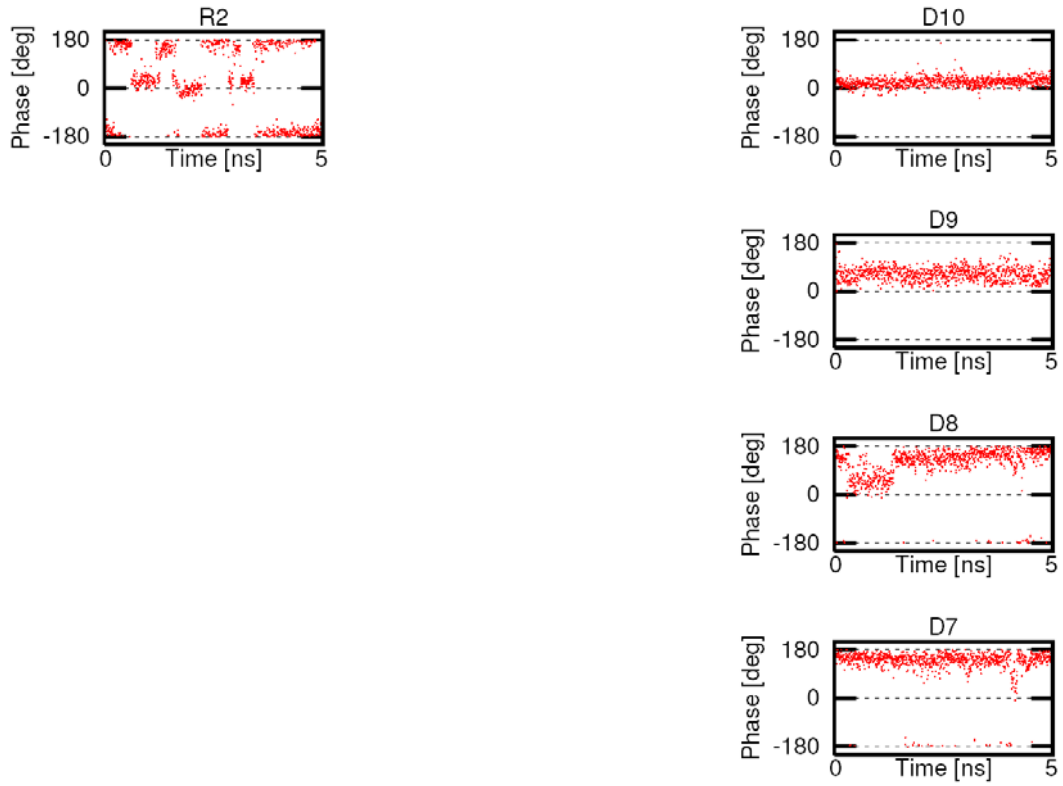


Figure 7.25: Sugar Puckering Pseudo-rotational angle

7.4 PMPG2

Another MD simulation was run with PMPG, using slightly different initial conditions. The graphs below are exactly the same as in the PMPG1 section.

7.4.1 Hydrogen bonds between template and incoming nucleotides

Hydrogen bonds among nucleotides and template remain very stable (fig 7.26).

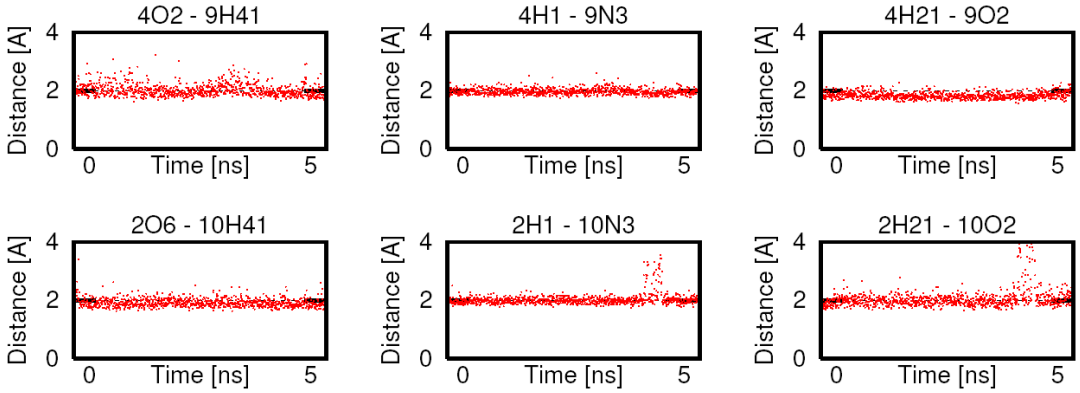


Figure 7.26: Length of hydrogen bonds of group 1

7.4.2 Hydrogen bonds among nucleotides

Stability of inter-nucleotide bonds expresses the same degree of disruption as for PMPG1 (fig 7.27).

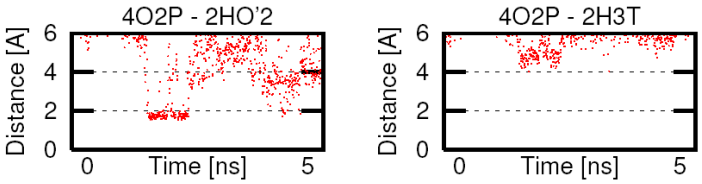


Figure 7.27: Length of hydrogen bonds of group 2

7.4.3 Hydrogen bonds between template and enzyme

Bonds that were tight in previous MD simulations remain fixed, but distances of less tightly bound atoms show more fluctuating behavior (fig 7.28).

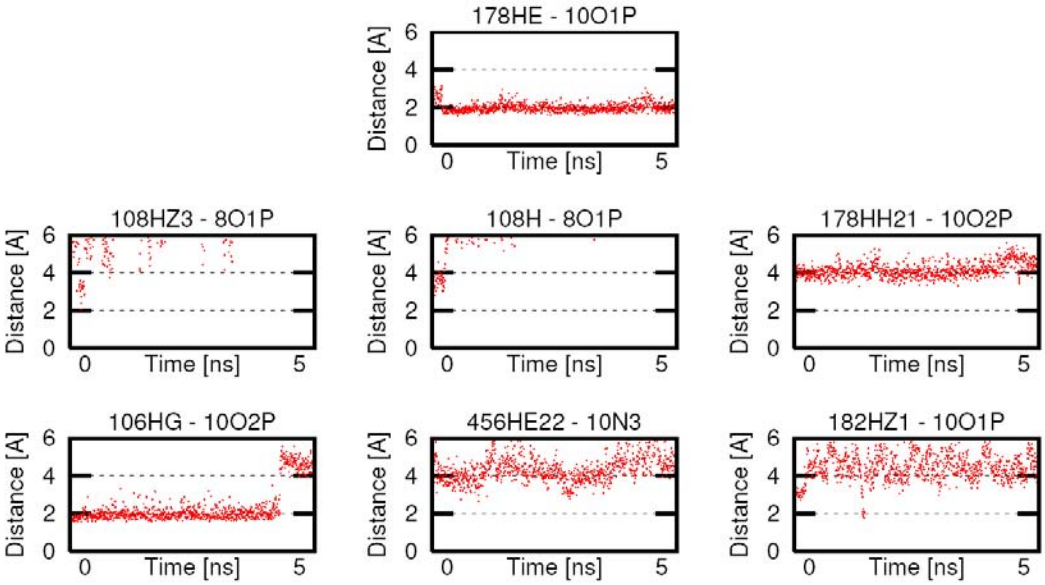


Figure 7.28: Length of hydrogen bonds of group 3

7.4.4 Hydrogen bonds between enzyme and nucleotides

The hydrogen bonds of this group are about the same as in the case of PMPG1 (fig 7.29).

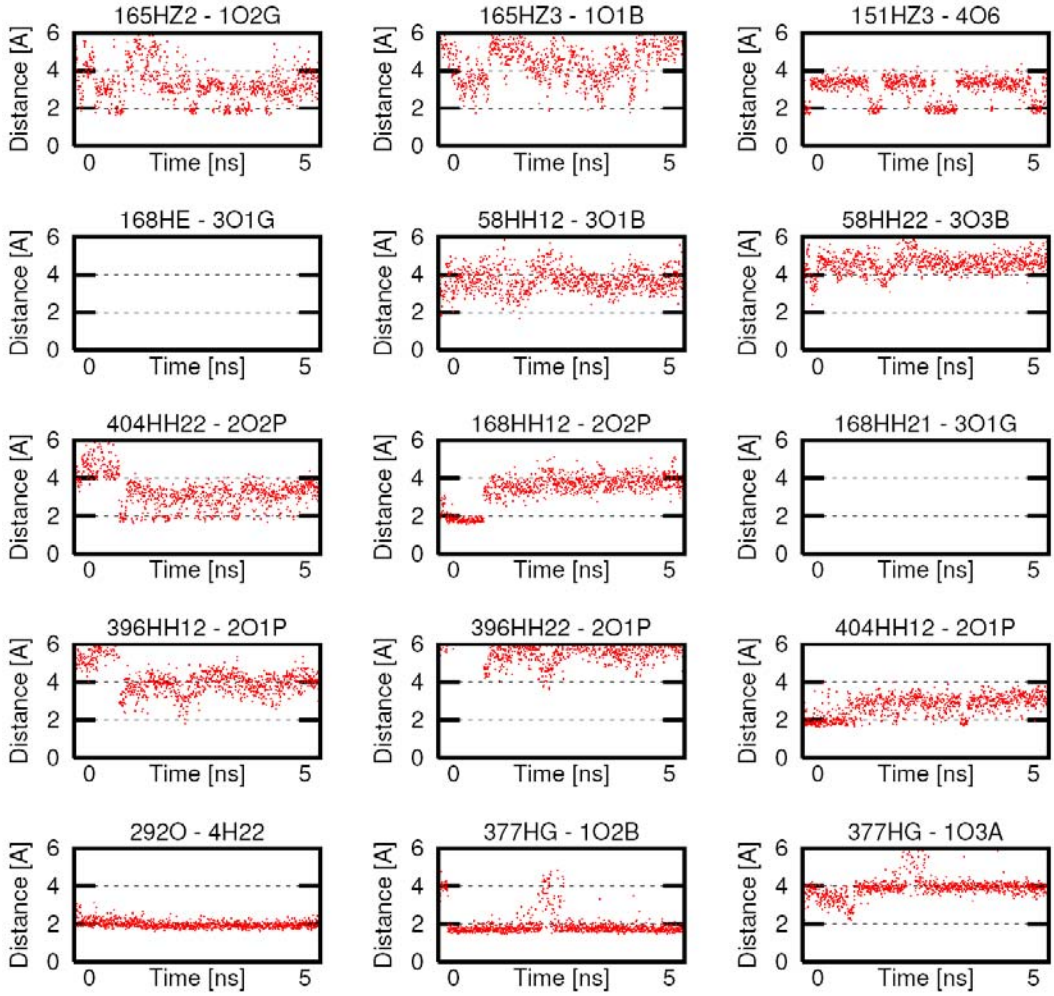


Figure 7.29a: Length of original bonds of group 4

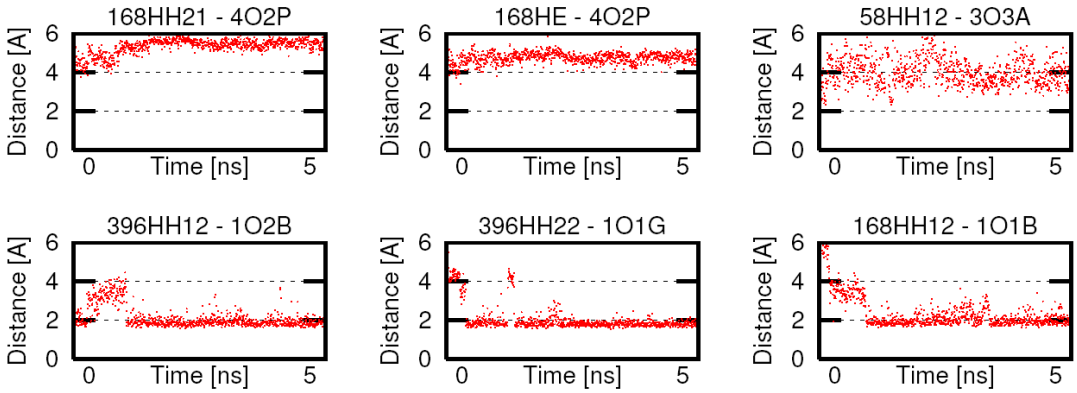


Figure 7.29b: Length of new hydrogen bonds of group 4

7.4.5 Ionic bonds involving Mg^{2+}

Ionic bonds express the same stability as in case of PMPG1 and PMEG (fig 7.30).

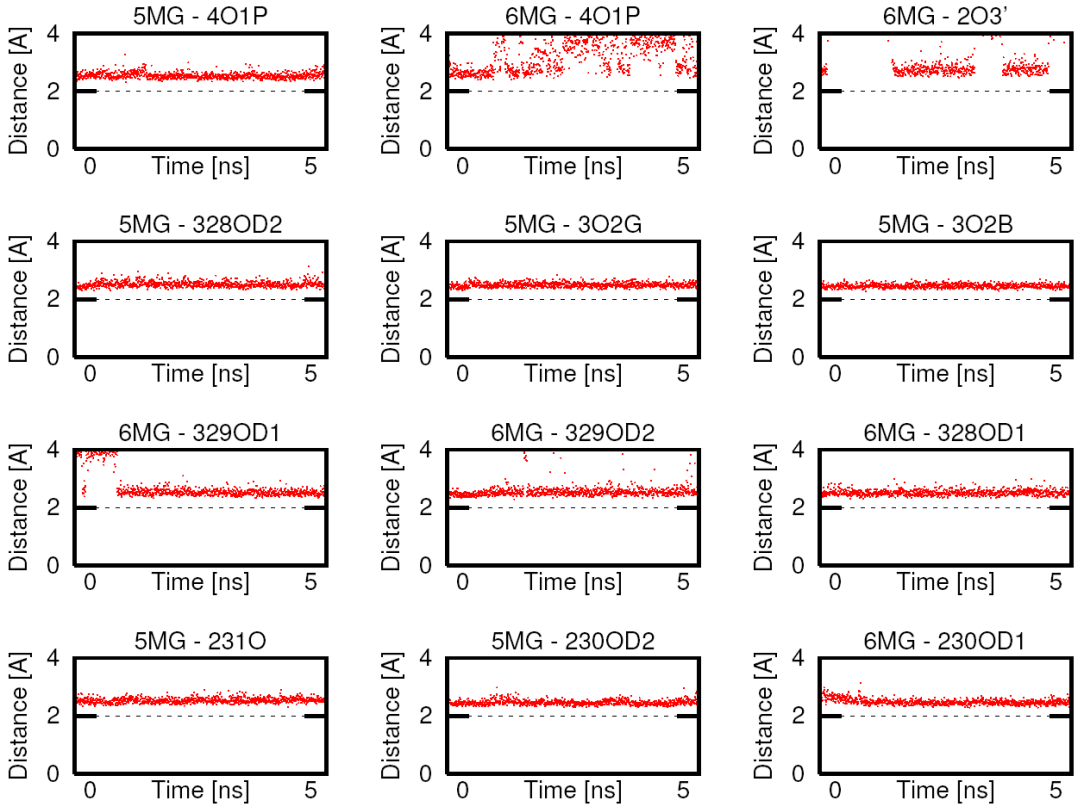


Figure 7.30: Length of ionic bonds of group 5

7.4.6 Torsion angles of backbone of PMPGt

Torsion angles go through temporary changes, similar to those of PMPG1 (fig 7.31).

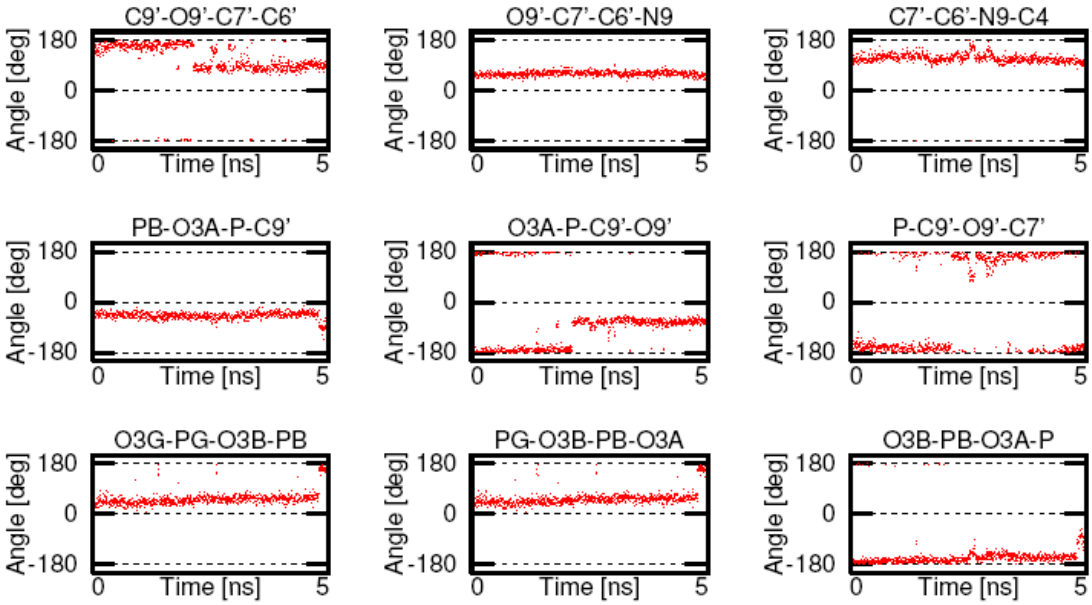


Figure 7.31: Torsion angles of PMPGt backbone

7.4.7 Sugar Puckering

D9 and D10 are clearly C3'-endo, while D7, D8 and R2 are assuming C2'-endo conformation. The conformation is more rigid in this simulation; sugars do not switch or balance on the borderline, like in PMPG1 simulation (fig 7.32).

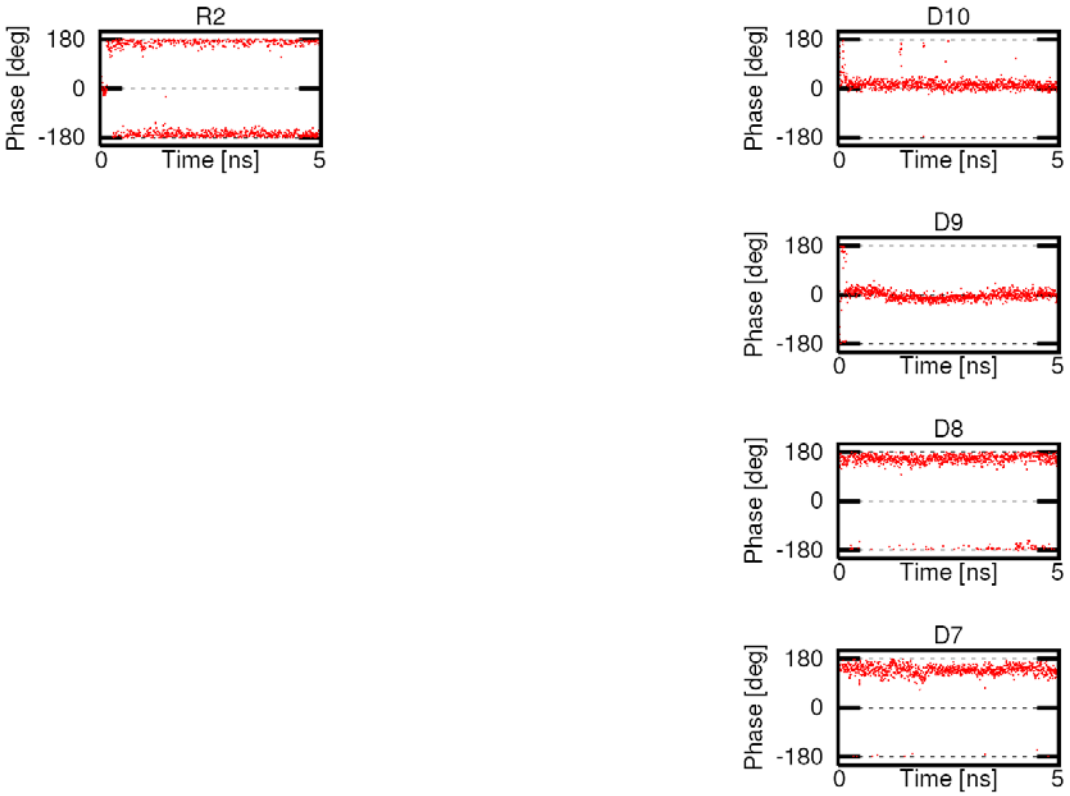


Figure 7.32: Sugar puckering Pseudo-rotational angle

7.5 HPMPG

Last simulation was run with HPMPG. Polar substitution – addition of hydroxyl group – gives HPMPG considerable difference from all previously studied compounds. This simulation was shorter than preceding simulations.

7.5.1 Hydrogen bonds between template and incoming nucleotides

Stability of the nucleotide – template Watson-Crick bonding was confirmed also by this simulation (fig 7.33). Hydroxyl group is too far away to induce any change.

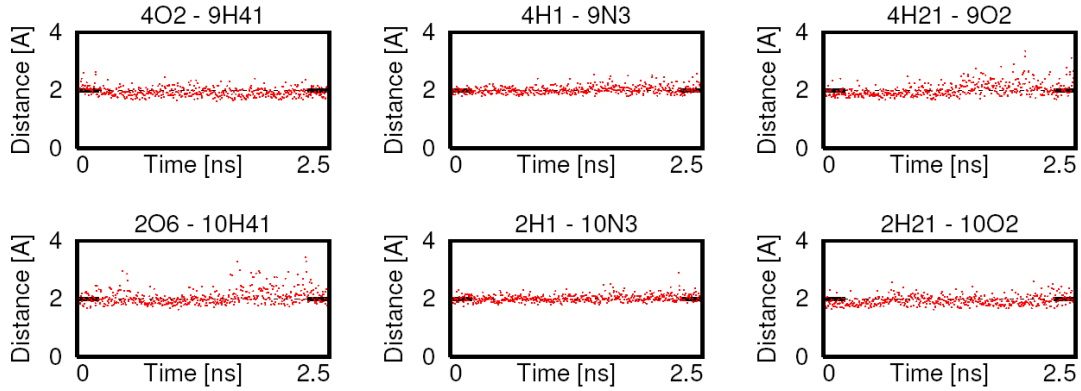


Figure 7.33: Length of hydrogen bonds of group 1

7.5.2 Hydrogen bonds among nucleotides

Inter-nucleotide interaction is again not very strong (fig 7.34). Extra hydroxyl group would indicate improvement, but it is oriented away from the R2. The hydroxyl group actually can't be bond to R2; the group is analogous to hydroxyl group of ribose at C2' in rN, which is used by enzyme to prefer rNTP over dNTP.

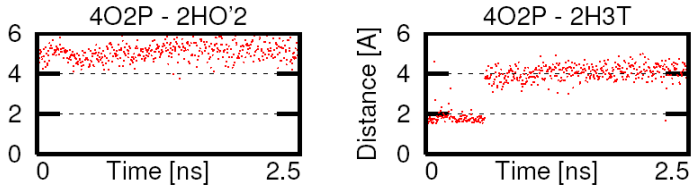


Figure 7.34: Length of hydrogen bounds of group 2

7.5.3 Hydrogen bonds between template and enzyme

This group of bonds has the same level of stability as in previous cases (fig 7.35), and is practically independent on the form of the nucleotide (or analog).

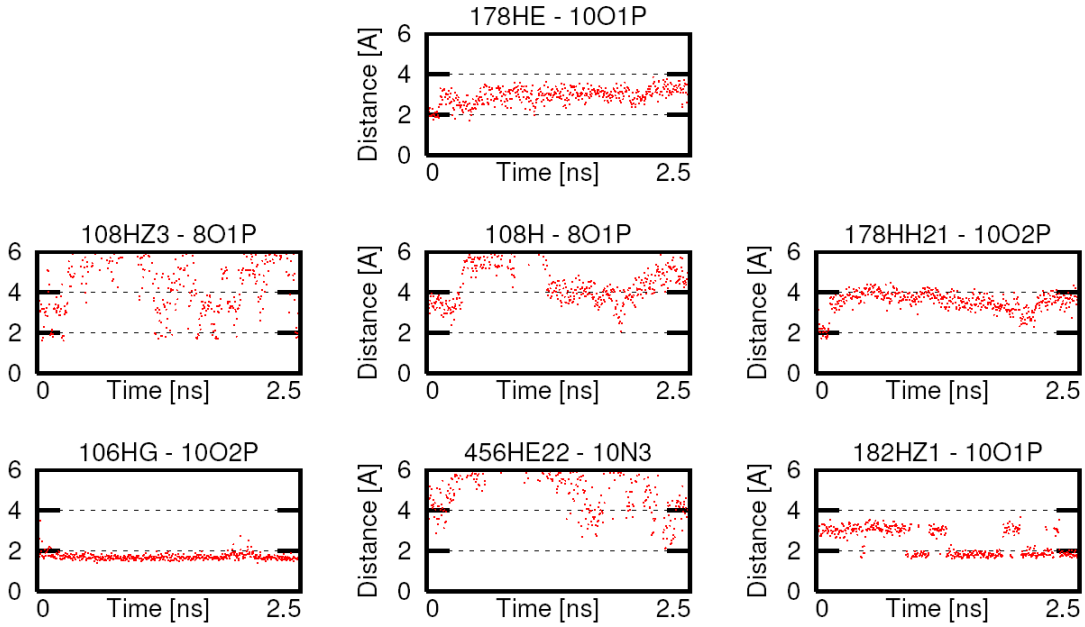


Figure 7.35: Length of hydrogen bonds of group 3

7.5.4 Hydrogen bonds between enzyme and nucleotides

The hydrogen bonds of this group are about the same as in case of PMPGs. The distances on the graphs (fig 7.37) are higher than for PMPG, but stable. That means, some hydrogen bond was established, probably with some close atom.

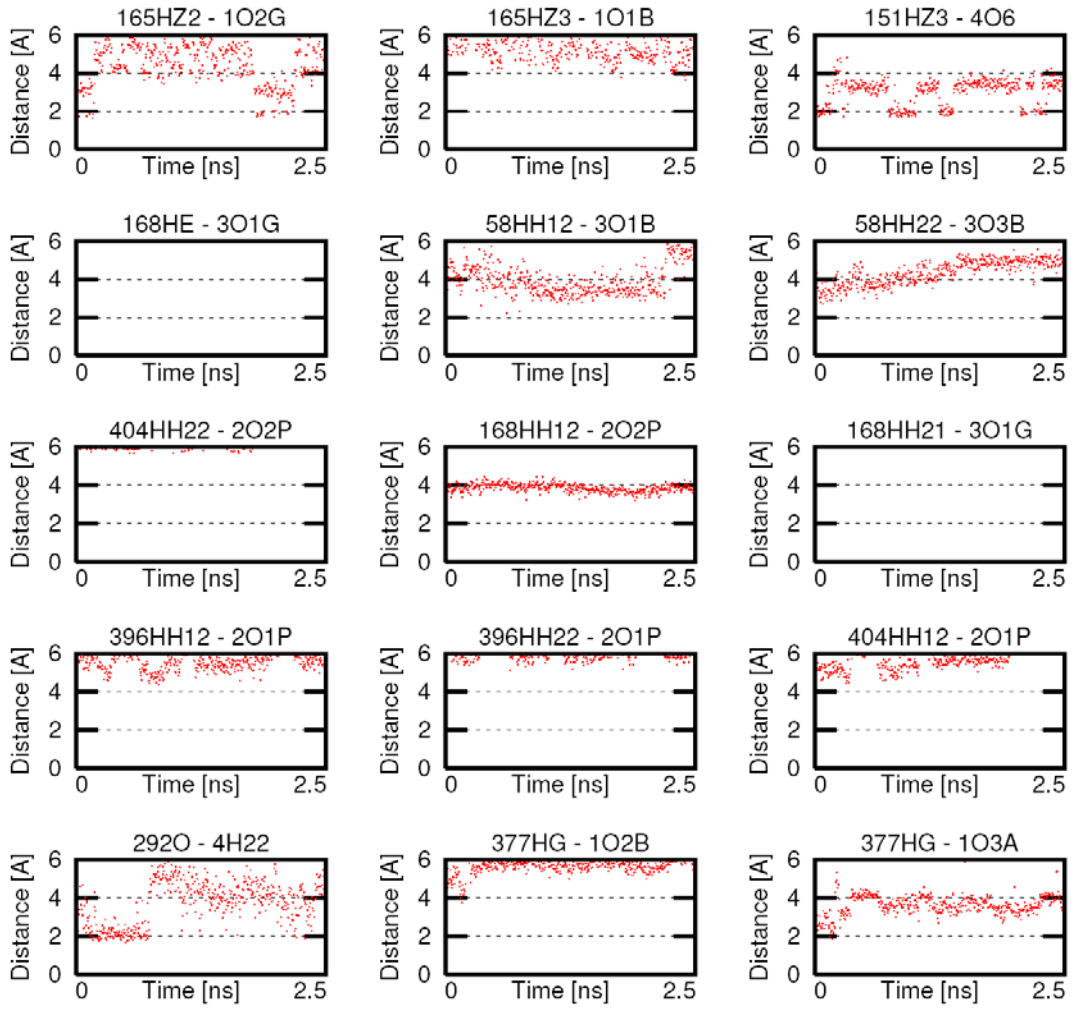


Figure 7.36: Length of original bonds of group 4

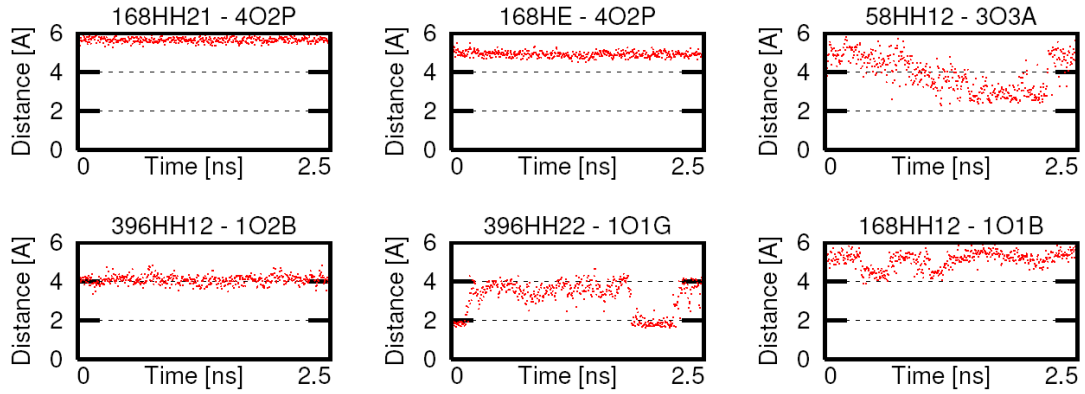


Figure 7.37: Length of new hydrogen bonds of group 4

7.5.5 Ionic bonds involving Mg^{2+}

Ionic bonds express steady stability. The hydroxyl group has no apparent effect on them (fig 7.38).

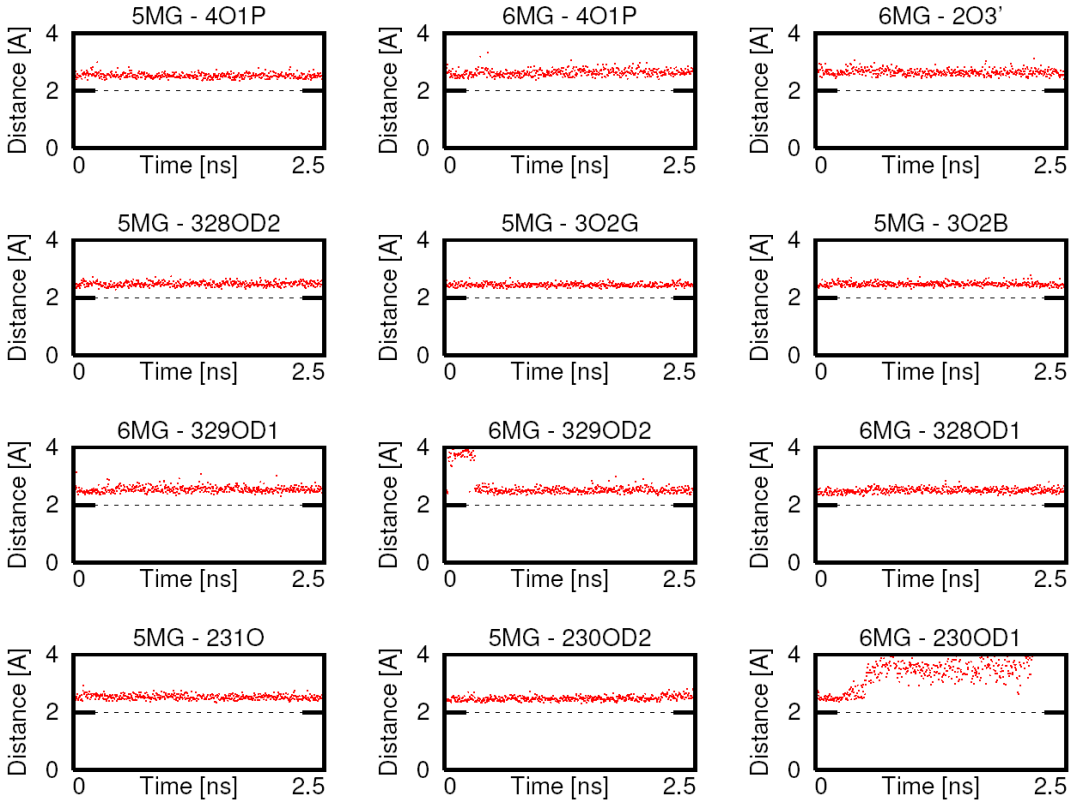


Figure 7.38: Length of ionic bonds of group 5

7.5.6 Torsion angles of backbone of HPMPGt

Torsion angles of HPMPGt are more stable than those of PMPG. Just one shift happens at the beginning of the simulation (fig 7.39). This is probably caused by hydroxyl group interaction with enzyme, which is discussed later.

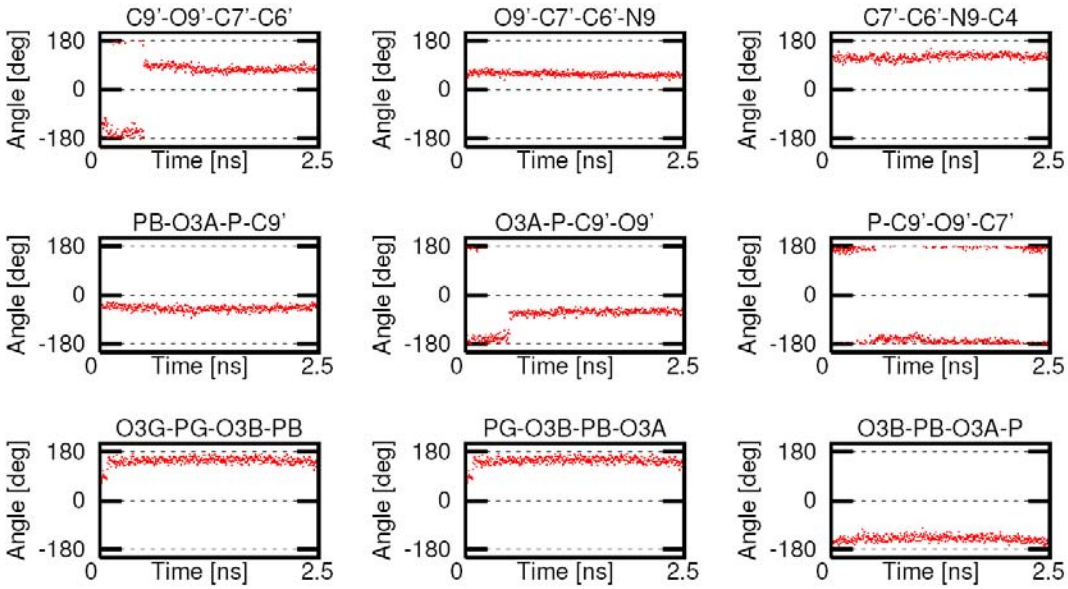


Figure 7.39: Torsion angles of HPMPGt backbone

7.5.7 Sugar Puckering

Nucleotide R2 oscillates between C2'-endo and C3'-endo; it can be result of hydroxyl group addition. D10 has clearly C3'-endo conformation as in the case of PMPGs, so it could be a result of sterical influence of HPMPG (PMPG). D9 and D8 are inconclusive and D7 is C2'-endo (fig 7.40).

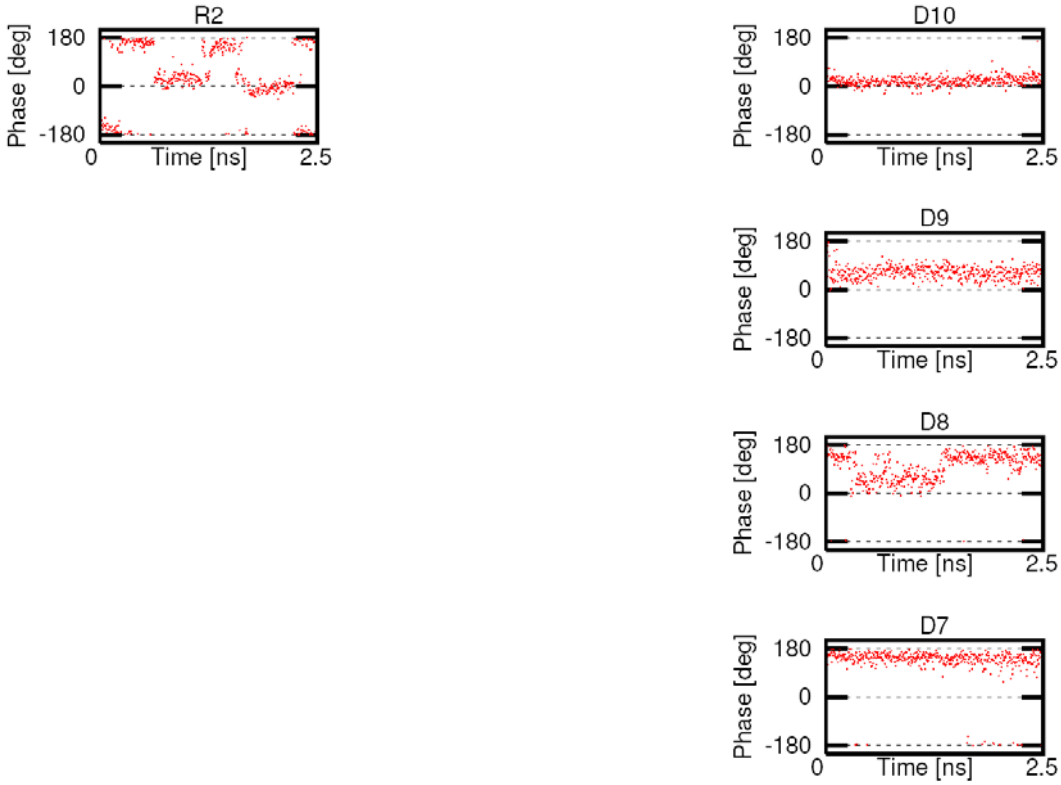


Figure 7.40: Sugar puckering Pseudo-rotational angle

7.5.8 Interaction of hydroxyl substituent

The hydroxyl substituent should interact with the enzyme in highly stabilizing manner, since it plays the same role, as the C2' hydroxyl group in rGTP, which should be a discrimination factor for HCV RdRp. The group forms two stable hydrogen bonds (fig 7.41a and fig 7.41b). As is clear from fig 7.42, during the first half of the MD simulation, the hydroxyl group forms bond with Asp235, and during the second half with the Leu169. Bonds are very stable, and mentioned residues are probably those allowing HCV RdRp to choose rNTP (and not dNTP).

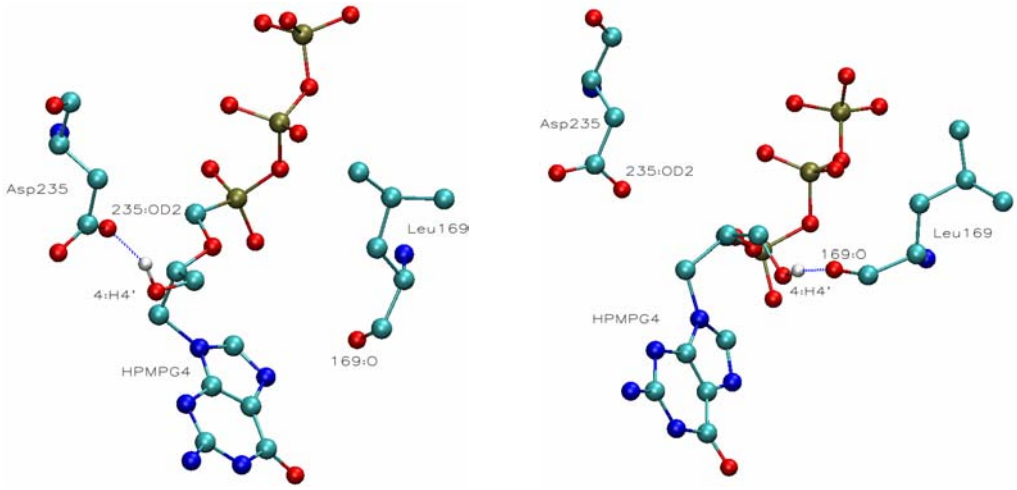


Figure 7.41a and 7.41b: Hydrogen bond of hydroxyl group with Asp235, Leu169 respectively; oxygen – red, phosphorus – gold, carbon – cyan, nitrogen – blue, hydrogen – white, hydrogen bond – blue-dashed line

Figure 7.42: Length of hydrogen bonds of hydroxyl group.

7.6 Comparison of simulations

In sections 7.3.4, 7.4.4, and 7.5.4, it seems that important arginine residues don't interact with the nucleotides. In fact, they mainly switch to another atoms, or change atoms over the period of simulation. Largest structural difference was among PMEG and HPMPG. At fig 7.43, their final positions are aligned. The picture shows that arginine residues changed their positions only very slightly. The alteration changed interacting atoms, but not the interaction with the triphosphate itself.

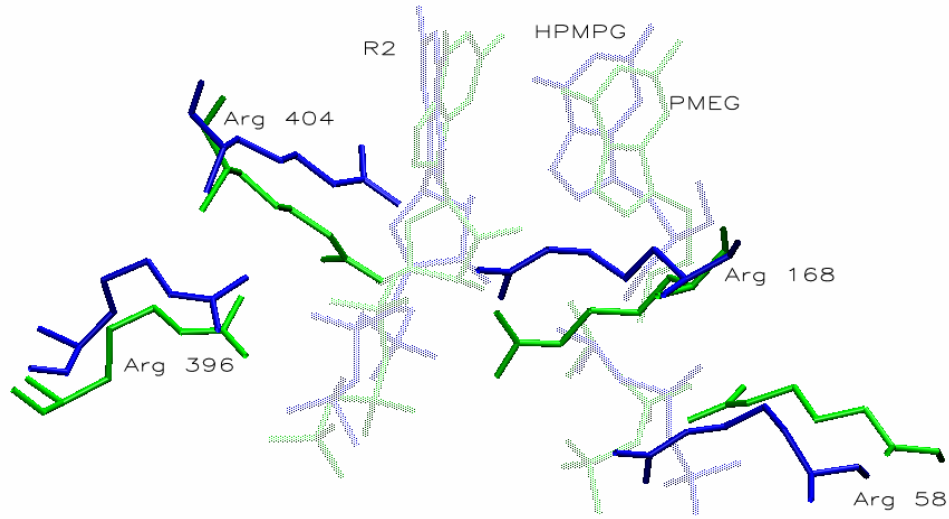


Figure 7.43: Alignment of final image of HPMPG and PMEG simulations

On the other hand, two MD runs with PMPG used the same nucleotide analog, only initial conditions were altered. Changes observed in the last frame of MD are therefore a result of the random processes within the MD simulation. As it can be seen in fig 7.44, difference is not much smaller than this in fig 7.43. This supports the claim that presence of the hydroxyl group doesn't influence stabilizing interactions of the triphosphate chain, and makes HPMPG more suitable candidate for HCV therapy than PMEG.

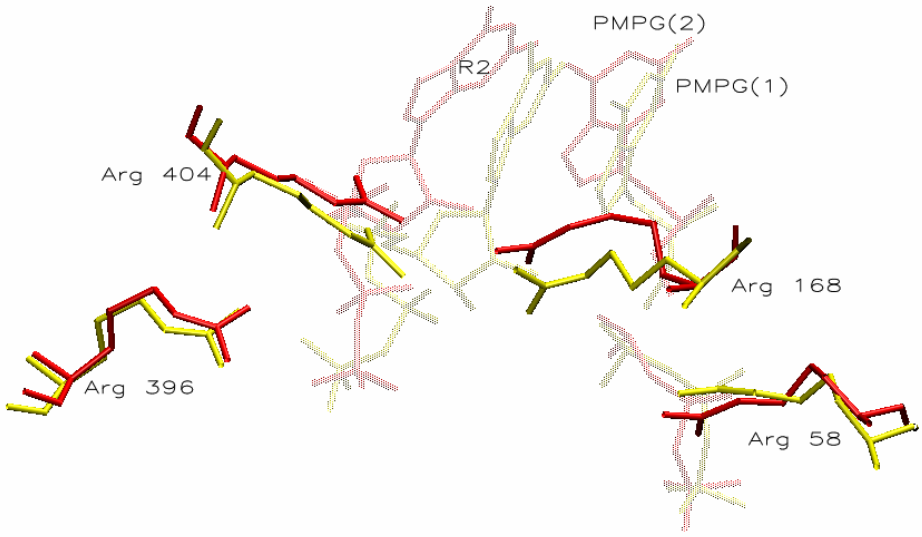


Figure 7.44: Alignment of final image of two PMPG simulations

Chapter 8 - Conclusions

Nucleic acids and proteins were described in first and second chapter, along with mechanisms of genetic code replication and its translation into the protein chain. In third chapter, important enzymes synthesizing RNA (polymerases) and DNA (reverse transcriptases) were discussed. The structure of HCV and its polymerase was scrutinized in chapter five and six. Further, principles of RdRp inhibition (both nucleotide and allosteric) and protease inhibition, oligonuclotide and immunomodulatory therapy were explained. In chapter six, basics of molecular dynamics simulation were presented by a program simulating a system consisting of argon atoms (interacting only through weak interactions).

In the rest of the work, MD simulations were used to determine possibilities of acyclic phosphonate analogs of nucleotides (PMEG, PMPG, and HPMPG) to inhibit HCV RdRp. Following ideas are based upon the data obtained from MD simulations. Watson-Crick base pairing of rNTP analogs and a template residue is stable independently on the C2' substitution of the analog. Inter-rNTP interactions of substrates are negligible (though little more stabile for the simplest molecule, PMEG). Enzyme's hydrogen bonds are largely directed towards negatively charged triphosphate tails. Arginine residues (namely Arg58, Arg168, Arg396, and Arg404) play the main role in substrate stabilization; they are highly adaptable to structural changes of substrates. Ionic interactions are independent on the analog used. Their main function is to bind aspartic acid residues, so they would not interfere with important arginines and a hydroxyl group of ribose. Magnesium ions themselves interact with the rGTP substrate hydroxyl group, and ensure its proper position for polymerization. The torsion angles of an analog backbone, which are good sign of stability, show that the simplest molecule (PMEG) is the most stable, while PMPG or HPMPG undergo shifts of their backbone conformations. Yet, the overall advantage of the HPMPG hydroxyl group is, that it makes persistent hydrogen bonds with the enzyme (Asp 235 and Leu 169). These residues tightly bind the analog and put it to the proper site. The residues probably pick rNTPs (over dNTPs) and enable HCV RdRp to recognize a proper substrate. The C2' hydroxyl group gives HPMPG appearance of the ribonucleotide. It should be noted both PMEG, and PMPG miss the hydroxyl group.

Summarized, it seems that HPMPG could be suitable as a nucleotide inhibitor, since it is stable enough and resembles ribose by the presence of the hydroxyl group. The guanine base is also advantageous because initiation of the polymerization process is carried out by rGTP. Therefore, HPMPG could stop polymerization at the very beginning.

Bibliography

[1] N. Previsani, D. Lavanchy: **Hepatitis C**, *WHO*, 2004

[2] H. Ma, W. R. Jiang, N. Robledo, V. Leveque, S. Ali, T. Lara-Jaime, M. Masjedizadeh, D. B. Smith, N. Cammack, K. Klumpp and J. Symos: **Characterization of the metabolic activation of Hepatitis C Virus Nucleoside Inhibitor β-D-2'-Deoxy-2'-fluoro-2'-C-methylcytidine (PSI-6130) and identification of a novel active 5'-Triphosphate Species**, Journal of Biological Chemistry, Vol. 282, 29812-20, 2007

[3] P. Karlson: **Základy Biochemie,** Academia, Praha, 1971

[4] D. L. Nelson, M. M. Cox: **Lehninger Principles of Biochemistry,** W.H. Freeman and Company, 4th edition, 2004

[5] http://en.wikipedia.org/wiki/Ribose

[6] http://en.wikipedia.org/wiki/Deoxyribose

[7] http://en.wikipedia.org/wiki/DNA

[8] D. Pavlová: **Molekulárně dynamické simulace analog nukleových kyselin a jejich komplexů s enzymem RNasou H,** Diplomová práce, Matematicko-fyzikální fakulta University Karlovy, Praha, 2007

[9] W. Saenger: **Principles of Nucleic Acid Structure,** Springer-Verlag, New York, 1984

[10] R. K. Murray, D. K. Granner, P. A. Mayes, V. W. Rodwell: **Harperova Biochemie,** H + H, Jinončany, 2002

[11] http://en.wikipedia.org/wiki/Flaviviridae

[12] http://en.wikipedia.org/wiki/Hepatitis_C_virus

[13] B. D. Lindenbach, Ch. M. Rice: **Unravelling hepatitis C virus replication from genome to function,** Nature, Vol. 436, 933 – 938, 2005

[14] S. Bressanelli, L. Tomei, F. A. Rey and R. De Francesco: **Structural analysis of the Hepatitis C Virus RNA Polymerase in complex with Ribonucleotides,** Journal of Virology, Vol. 76, 3482 – 3492, 2002

[15] S. J. Butcher, J. M. Grimes, E. V. Makeyev, D. H. Bamford, D. I. Stuart: **A mechanism for initiating RNA-dependent RNA polymerization, Nature,** Vol. 410, 235 – 240, 2001

[16] R. De Francesco, G. Migliaccio: **Challenges and successes in developing new therapies for hepatitis C,** Nature, Vol. 436, 953 – 960, 2005

[17] G. Migliaccio, J. E. Tomassini, S. S. Carroll, L. Tomei, S. Altamura, B. Bhat, L. Bartholomew, M. R. Bosserman, A. Ceccacci, L. F. Colwell, R. Cortese, R. De Francesco, A. B. Eldrup, K. L. Getty, X. S. Hou, R. L. LaFemina, S. W. Ludmerer, M. MasCoss, D. R. McMasters, M. W. Stahlhut, D. B. Olsen, D. J. Hazuda, and O. A. Flores et al: **Characterization of resistance to non-obligate chain-terminating ribonuleoside analogs that inhibit Hepatitis C Virus replication in Vitro,** Journal of Biological Chemistry, Vol. 278, 49164 – 49170, 2003

[18] Erik De Clercq, Antonin Holý, Ivan Rosenberg, Takashi Samýma, Jan Balzarini, and Prabhat C. Mandqal: **A novel selective broad-spectrum anti-DNA virus agent**, Nature Journal, Vol. 323, 464 – 467, 1986

[19] S. Saralamba: **Molecular dynamics simulation of HIV reverse transcriptase (RT) in complex with nucleic acids and its potential inhibitors,** Diplomová práce, Matematicko-fyzikální fakulta University Karlovy, Praha, 2006

[20] http://en.wikipedia.org/wiki/Antisense_therapy

[21] http://en.wikipedia.org/wiki/Dicer

[22] http://en.wikipedia.org/wiki/RNA_interference

[23] http://en.wikipedia.org/wiki/MiRNA

[24] http://en.wikipedia.org/wiki/Heterochromatin

[25] R. Hrach: **Počítačová Fyzika I,** PF UJEP, Ústí nad Labem, 2003

[26] L. Skála: **Kvantová teorie molekul,** Univerzita Karlova – Karolinum, Praha, 1995

Appendix A – Lennard-Jones potential

Attractive interaction of Lennard-Jones potential is decreasing as ~$1/r^6$. These forces can be derived from an idea of substituting particles by oscillating dipoles [26] (see fig A.1).

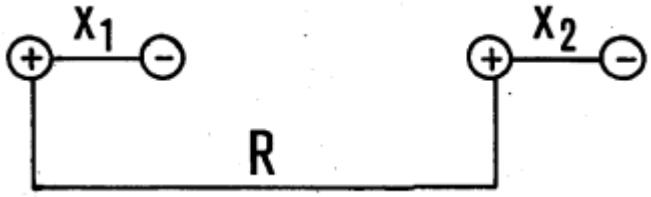


Figure A.1: Illustration of oscillating dipoles [26]

Schrödinger equation with dipole potential for the system can be solved as follows

$$\left(-\frac{h}{2m}\frac{d^2}{dx_1^2}-\frac{h}{2m}\frac{d^2}{dx_2^2}+U\right)\psi = E\psi$$
$$U=\frac{1}{2}kx_1^2+\frac{1}{2}kx_2^2+e^2\left(\frac{1}{R}+\frac{1}{R+x_2-x_1}-\frac{1}{R-x_1}-\frac{1}{R+x_2}\right) \quad \text{(A.1)}$$

Potential can be rewritten, supposing $|x_1|,|x_2| << R$ as

$$U=\frac{1}{2}kx_1^2+\frac{1}{2}kx_2^2-e^2\left(\frac{2x_1x_2}{R^3}\right). \quad \text{(A.2)}$$

Introducing substitution $z_1=\frac{1}{\sqrt{2}}(x_1+x_2), z_2=\frac{1}{\sqrt{2}}(x_1-x_2)$, $k_1=k-\frac{2e^2}{R^3}$, and $k_2=k+\frac{2e^2}{R^3}$, we obtain an equation of two independent lhos, with appropriate energy base state E_0

$$\left(-\frac{h}{2m}\frac{d^2}{dz_1^2}-\frac{h}{2m}\frac{d^2}{dz_2^2}+\frac{1}{2}k_1z_1^2+\frac{1}{2}k_2z_2^2\right)\psi = E\psi$$
$$E_0=\frac{1}{2}h(\omega_1+\omega_2)=\frac{1}{2}h\omega_0\left(\sqrt{1-\frac{2e^2}{kR^3}}+\sqrt{1+\frac{2e^2}{kR^3}}\right) \quad \text{(A.3)}$$

Energy can be further approximated by part of binomial series, which after addition of appropriate members yields base state of interacting oscillators

$$E_0 \approx h\omega_0\left(1-\frac{e^4}{2k^2R^6}\right). \quad \text{(A.4)}$$

Weak interaction for three dimensions is therefore

$$\Delta E_0(R)=-\frac{3}{4}\frac{h\omega_0e^4}{k^2R^6}. \quad \text{(A.5)}$$