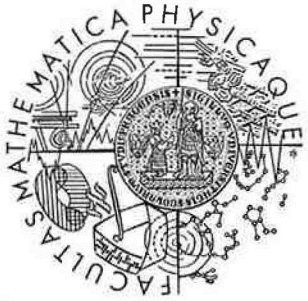
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Diana Pavlová

Molekulárně-dynamické simulace analog nukleových kyselin

a jejich komplexů s enzymem RNasou H

FÚ UK

Vedoucí diplomové práce: RNDr. Ivan Barvík, Ph.D.
Studijní program: Biofyzika

2007

Prohlašuji, že jsem svou diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 20. 4. 2007 Diana Pavlová

Contents

1 Introduction 1
- 1.1 Nucleic acids . . . 1
  - 1.1.1 Chemical structure of nucleic acids . . . 1
  - 1.1.2 Gene expression . . . 2
  - 1.1.3 DNA replication . . . 2
  - 1.1.4 Quantitative description of nucleic acid conformations . . . 4
- 1.2 RNase H enzyme . . . 9
  - 1.2.1 Function of RNase H enzyme . . . 9
  - 1.2.2 Structure and function of E. coli RNase H . . . 10
  - 1.2.3 Bacterial RNase H in complex with a substrate . . . 17
  - 1.2.4 Human RNase H . . . 19
- 1.3 Antisense technique . . . 21
  - 1.3.1 General information about the antisense technique . . . 21
  - 1.3.2 Modified oligonucleotides . . . 23
- 1.4 Aims and objectives . . . 29

2 Methods 31
- 2.1 Bioinformatics and Molecular Dynamics Simulations . . . 31
  - 2.1.1 Comparative model building . . . 31
  - 2.1.2 The method of molecular dynamics simulations . . . 33
  - 2.1.3 Thermodynamical properties of the simulated system . . . 37
  - 2.1.4 Running a MD simulation . . . 38
- 2.2 Software . . . 39
  - 2.2.1 Modeller . . . 39
  - 2.2.2 AMBER . . . 39
  - 2.2.3 NAMD . . . 41
  - 2.2.4 VMD and CHIMERA . . . 41

3 Results and discussion 42
- 3.1 Building up the human RNase H computer-based model . . . 42
  - 3.1.1 Model of the catalytic domain . . . 42
  - 3.1.2 Model of the RNA-binding domain . . . 44
- 3.2 Human RNase H bound to an unmodified substrate . . . 47

3.2.1 Initial model 47
3.2.2 Model refinement I: enzyme-substrate recognition 54
3.2.3 Model refinement II: active site stabilization - Mg ions parametrization 63
3.3 Models of RNase H bound to a modified substrate 68
3.3.1 The C3'-O-P-*C*-O-C5' modification 68
3.3.2 The C3'-O-*C*-P-O-C5' modification 76
3.4 E. coli RNase H with a substrate 83

4 Conclusions **87**

Název práce: Molekulárně-dynamické simulace analog nukleových kyselin
Autor: Diana Pavlova
Katedra: FÚ UK
Vedoucí diplomové práce: RNDr. Ivan Barvík, Ph.D.
e-mail vedoucího: ibarvik@karlov.mff.cuni.cz
Abstrakt: Cílem této diplomové práce bylo studium interakcí lidské RNasy H s přirozeným a modifikovaným substrátem. 3-D struktura enzymu byla vytvořena na základě homologního modelování a studována pomocí molekulární dynamiky. Bylo zjištěno, že interakce se účastní 10 reziduí: DNA-vazebná jsou Ser98, Trp90, Thr46, Asn105 a Gln48; RNA-vazebná Met77, Asn47, Glu51, Cys12 a Ser14. Vysoce konzervovaný DDE motiv tvoří karbonylová rezidua Asp139, Asp75, Glu51 a Asp10, nezbytná pro stabilizaci Mg iontů v aktivním místě. Dále byl zkoumán vliv modifikací internukleotidových spojek (C3'-O3'-CH2-P-O5'-C5' a C3'-O3'-P-CH2-O5'-C5') v DNA vlákně na interakci s enzymem. Detailně byly prozkoumány rozdíly mezi komplexem s lidskou/E. coli RNasou H. V předchozích studiích byla u E. coli pozorována nejvýraznější narušení vodíkových vazeb mezi rezidui Trp109, Tyr78, Lys104 a Asn105 a modifikovaným substrátem. V případě lidské RNasy H však v tomto místě váže substrát pouze Asn105. To znamená, že lidská RNasa H by mohla lépe tolerovat neisosterní fosfonátová residua ve srovnání se svým protějškem z E. coli.
Klíčová slova: molekulární dynamika, lidská RNasa H, modifikované oligonukleotidy

Title: Molecular dynamics studies - human RNase H in a complex with RNA/DNA duplex
Author: Diana Pavlová
Department: FUUK
Supervisor: RNDr. Ivan Barvík, Ph.D.
Supervisor's e-mail address: ibarvik@karlov.mff.cuni.cz
Abstract: The aim of this diploma thesis was to study the interaction between human RNase H enzyme and a natural and modified substrate. The 3-D structure was determined on the basis of homology modeling and studied by means of molecular dynamics. It was found out that there are 10 residues involved in the interaction with the substrate: DNA-binding residues are Ser98, Trp90, Thr46, Asn105 and Gln48; RNA-binding residues are Met77, Asn47, Glu51, Cys12 and Ser14. Following carbonyl residues (the highly conserved DDE motif) are substantial for a proper stabilization of Mg ions in the active site: Asp 139, Asp75, Glu51 and Asp10 that coordinates both Mg ions. Further, the impact of the internucleotide linkage modifications in the DNA strand (either C3'-O3'-CH2-P-O5'-C5' or C3'-O3'-P-CH2-O5'-C5') was investigated. Major differences in recognition of the substrate by E.coli/Human RNase H were analyzed in detail. In previous studies, the most remarkable disturbance of the H-bonds between the E. coli enzyme and a modified substrate was observed by the Trp109, Tyr78, Lys104 and Asn105 residues. In the case of human RNase H, only Asn105 recognizes the substrate here. It means that human RNase H could be more tolerant of nonisosteric phosphonate residues compared to the E.coli counterpart.
Keywords: molecular dynamis, human RNase H, modified oligonucleotides

List of Figures

1.1 The A-T and C-G Watson-Crick basepair [3] 2
1.2 The DNA replication [4] 3
1.3 The A, B and Z conformations of a helix [3] 5
1.4 An example of RNA hairpin motif 5
1.5 The definition of a torsion angle [7] 6
1.6 The C2'-endo and C3'-endo sugar pucker [7] 7
1.7 Properties of A, B and Z conformations [6] 8
1.8 The comparison of HIV, BH and E. coli RNase H: structure(A) and sequence alignment with conserved regions highlighted (B) [8] 9
1.9 A DNA/RNA hybrid complex model with E. coli RNase H (RNA-strand white, DNA-strand blue, three carboxylates Asp10, Glu48 and Asp70 shown) [17] 10
1.10 Active sites of RNases H and its relatives [18] 11
1.11 A model for the complex between E. coli RNase H1 and DNA/RNA hybrid: the solid region represents the basic protrusion [21] 12
1.12 The DNA/RNA hybrid is shown to bind to the surface of the enzyme along the line connecting Mg^{2+} ions and the basic protrusion [22] 13
1.13 The coordination of Mg^{2+} ions in several nucleases and polymerases [18] . . 14
1.14 Phosphoryl transfer reactions [18] 14
1.15 Three different proposed mechanisms of the catalysis [25] 15
1.16 A mechanism of the catalysis proposed by Haruki [27] 16
1.17 The active site of D132N and D192N in complex with RNA/DNA hybrid [8] 17
1.18 A comparison of metal ion coordination in the substrate (green) and product (orange) [28] 17
1.19 A schematic representation of the reaction steps proposed for RNase H [28] 18
1.20 The human RNase H1 sequence: black underlined RNA-binding domain, red underlined catalytic domain and the spacer region between them [31] . 20
1.21 Different antisense strategies in comparison with a conventional drug [35] . 21
1.22 The RNase H cleavage (A) and blocking of translation (B) [35] 22
1.23 The first(a) and the second(b) generation ASOs [35] 24
1.24 Differences in the interaction of RNase H with a MOE modified substrate based on its position [39] 25

1.25 Relative cleavage rates and cleavage sites for chimeric heteroduplex substrate [39] . . . 26
1.26 A model of the interaction of RNase H with the chimeric heteroduplex [39] 27
1.27 A model for the interaction of RNase H with the heteroduplex containing MOE modifications [39] . . . 27
1.28 The third generation ASOs [35] . . . 28
1.29 UOCHB modifications [50] . . . 29
1.30 Modifications studied in this work and a natural dinucleotide between them 30

2.1 An example of empirical similarity tables (BLOSUM40 table) [51] . . . 32
2.2 The course of Lennard-Jones potential [52] . . . 35
2.3 The concept of periodic boundaries [54] . . . 36
2.4 NAMD configuration file for MD . . . 41

3.1 The alignment of human and TT (1RIL) RNase H sequence . . . 43
3.2 Superimposed structures of human and TT (1RIL) RNase H . . . 43
3.3 The alignment of human and E. coli (1JL1) RNase H sequence . . . 44
3.4 Superimposed structures of human and E. coli RNase H with the DDE motif highlighted . . . 44
3.5 The alignment of the human and Saccharomyces RNase H RNA-binding domain sequence . . . 45
3.6 The model of the RNA-binding domain of human RNase H . . . 45
3.7 The DOPE-potential evaluation of the reliability of models (model red, template green) . . . 46
3.8 The structure of human RNase H in complex with the substrate (a) and residues interacting with substrate highlighted(b) . . . 49
3.9 The structure of human RNase H - different views, positioning of the substrate 50
3.10 The active site with Mg^{2+} ions (green) with nucleotides highlighted (a) and interacting residues highlighted (b) . . . 51
3.11 Residues interacting with the RNA-strand . . . 52
3.12 Residues interacting with the DNA-strand . . . 53
3.13 Model 1: Schema of the enzyme - substrate interaction . . . 57
3.14 Model 2: Schema of the enzyme - substrate interaction . . . 58
3.15 Model 3: Schema of the enzyme - substrate interaction . . . 59
3.16 Model 4: Schema of the enzyme - substrate interaction . . . 60
3.17 Model 5: Schema of the enzyme - substrate interaction . . . 61
3.18 Model 6: Schema of the enzyme - substrate interaction . . . 62
3.19 Schema of the active site and respective graphs for models 7,8 and 9 . . . 64
3.20 Model 7: Schema of the enzyme - substrate interaction . . . 65
3.21 Model 8: Schema of the enzyme - substrate interaction . . . 66
3.22 Model 9: Schema of the enzyme - substrate interaction . . . 67
3.23 Modification 1: Schema of the enzyme - substrate interaction . . . 71
3.24 Modification 2: Schema of the enzyme - substrate interaction . . . 72

3.25 Modification 3: Schema of the enzyme - substrate interaction 73
3.26 Modification 4: Schema of the enzyme - substrate interaction 74
3.27 Modification 5: Schema of the enzyme - substrate interaction 75
3.28 Modification 6: Schema of the enzyme - substrate interaction 78
3.29 Modification 7: Schema of the enzyme - substrate interaction 79
3.30 Modification 8: Schema of the enzyme - substrate interaction 80
3.31 Modification 9: Schema of the enzyme - substrate interaction 81
3.32 Modification 10: Schema of the enzyme - substrate interaction 82
3.33 Alignment of the E. coli and human RNase H with the DNA-binding (red) and RNA-binding (blue) residues and DDE motif (green) 84
3.34 Comparison of the human (green circle) and E. coli (yellow square) structures with DNA-binding residues 84
3.35 E.coli RNase H: structure and residues interacting with DNA-strand (red) and RNA-strand (black) 85
3.36 E. coli: Schema of the enzyme - substrate interaction 86

Acknowledgments

I would like to express my deepest gratitude to my supervisor, RNDr. Ivan Barvík, Ph.D., for his valuable advice, patient help and unceasing support. His deep insight in the subject was an inexhaustible source of understanding and knowledge for me. My personal thanks also to Sergej and my whole family for their loving encouragement. I am especially indebted to RNDr. Ivan Barvík Ph.D., my dad and Sergej for reading the manuscript, helpful remarks and suggestions.

Chapter 1

Introduction

1.1 Nucleic acids

1.1.1 Chemical structure of nucleic acids

Nucleic acids are biochemical macromolecules that carry genetic information. The term "nucleic acid" is related to their appearance predominately in cellular nuclei. Nucleic acids allow to preserve, use and transmit the genetic information. Therefore they are found in all living cells and viruses and they constitute the chemical basis of life.

A nucleic acid (DNA or RNA) is composed of nucleotide chains [1]. Each nucleotide consists of three components: a nitrogenous heterocyclic base, either a purine or a pyrimidine; a pentose sugar; and a phosphate group. The major difference between DNA and RNA is the sugar, 2'-deoxyribose in DNA and the alternative pentose sugar ribose in RNA. The nitrogenous bases found in the two nucleic acids are adenine (A), cytosine (C), guanine (G), and thymine (T, only in DNA) or uracil (U, only in RNA) [2]. These bases are classified into two types: adenine and guanine are fused five- and six-membered heterocyclic compounds called purines, while cytosine and thymine are six-membered rings called pyrimidines. The bases and the sugar form together the "nucleoside". By adding a phosphate group from H_3PO_4 a nucleotide is formed.

The DNA double helix is held together by hydrogen bonds between the bases attached to the two strands (see Figure 1.1 [3]). The purines form hydrogen bonds to pyrimidines, with A pairing with T, and C bonding only to G. The two types of base pairs form different numbers of hydrogen bonds: A-T two hydrogen bonds, and G-C forming three hydrogen bonds. DNA is copied into RNA by RNA polymerase enzymes that only work in the 5' to 3' direction. A DNA sequence is called "sense", if its sequence is copied by these enzymes and then translated into protein. The sequence on the opposite strand is complementary to the sense sequence and is therefore called the "antisense" sequence.

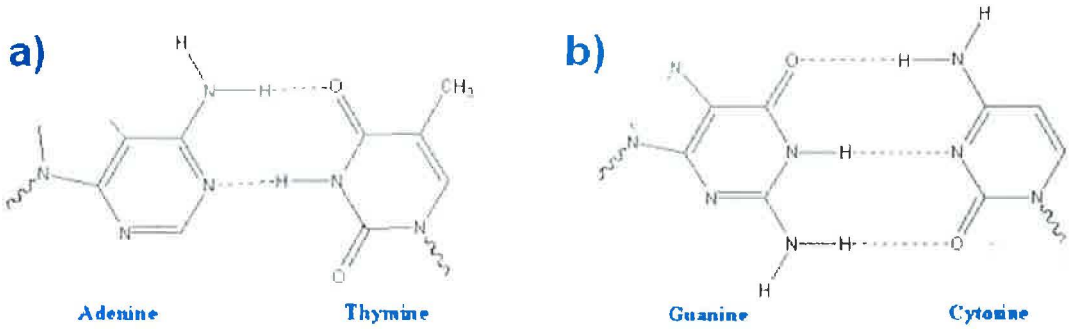


Figure 1.1: The A-T and C-G Watson-Crick basepair [3]

1.1.2 Gene expression

DNA contains the genetic information needed for the function, growth and reproduction of the cells. The pieces of DNA containing whole information for the synthesis of a given protein are called genes. Genetic information in genes is transmitted through complementary base pairing. When a cell needs to synthesize a protein encoded in the DNA, the DNA sequence of the gene is copied into a complementary messenger RNA sequence by RNA polymerase. This process is called transcription. The RNA copy (mRNA) may then serve for a production of a matching protein sequence by ribosomes in a process called translation. The relationship between the nucleotide sequences of genes and the amino-acid sequences of proteins is determined by solid rules. The genetic code is written using three-letter 'words' called codons formed from a sequence of three nucleotides, so that every three nucleotides encode one amino-acid. There are 4 bases in 3-letter combinations, so it results in 64 possible codons (4^3 combinations). In this way the twenty standard amino acids are encoded and most amino acids have therefore more than one possible codon. There are also three 'stop' codons which signal the end of the coding region (the TAA, TGA and TAG codons) [2]. The mRNA is decoded by a ribosome that reads the RNA sequence by base-pairing the messenger RNA to transfer RNA (tRNA). tRNA carries amino acids that are lined up to form a protein. The 3D structure comes up during the translation already.

1.1.3 DNA replication

The genetic information is just copied in a process called DNA replication during the process of cell replication. The process of DNA replication is in general common for all prokaryotic and eukaryotic organisms [1]. The two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA polymerase. As each DNA strand holds the same genetic information, both strands serve as templates for the reproduction of the opposite strand. The template strand is preserved in its entirety and the new strand is created from nucleotides. The two resulting strands are almost identical, the percentage of errors is very low. This is due to the fact that several mechanisms as

proof-reading and error-checking ensure a correct replication.

In cells, DNA replication occurs before the cell replication. In eukaryotes the timing of the replication is highly controlled and it occurs during the S phase of the cell cycle. DNA replication can be also preformed in laboratory in a process called PCR.

The structure formed by DNA during replication is called "replication fork" [2]. It comes up, when the enzyme helicase breaks the hydrogen bonds in the Watson-Crick pairs. Two single stranded regions are formed. The replication itself is performed by the enzyme DNA polymerase. This enzyme works, however, just in the 5'->3' direction. The DNA strand that is synthesized in this direction is called the leading strand. On this strand, DNA polymerase is able to synthesize DNA in a continuous manner using the free 3'OH group donated by a single RNA primer and continuous synthesis occurs in the direction in which the replication fork is moving.

The anti-parallel strand complementary to the 5'->3' template strand is known as the lagging strand as the DNA polymerase cannot build a strand in the 3'->5' direction. Thus, the lagging strand is synthesized in short segments known as Okazaki fragments (see Figure 1.2 [4]). On the lagging strand, primase builds short RNA primers along the DNA strand. DNA polymerase is then able to use the free 3'OH group on the RNA primer to synthesize DNA in the 5'->3' direction. The RNA fragments are then removed and new deoxyribonucleotides are added to fill the gaps where the RNA was present. DNA ligase is then able to ligate the deoxyribonucleotides together, completing the synthesis of the lagging strand.

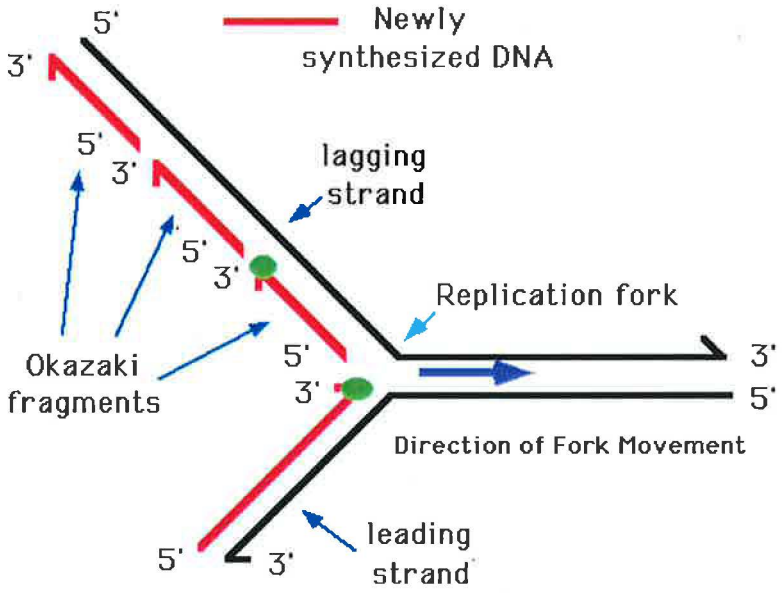


Figure 1.2: The DNA replication [4]

The removal of the RNA primers is possibly provided by the enzyme of RNase H. This

enzyme is thought to remove the RNA primers as it is able to cleave the RNA/DNA heteroduplex. It is found in all living organisms and in the eucaryotes it specifically cleaves the RNA strand in the distance of 7 bp from the 3'-DNA end.

1.1.4 Quantitative description of nucleic acid conformations

Various properties are measured that describe the conformation of the nucleic acids and of enzyme-substrate complexes. The geometry of the nucleic acids as substrate uses to play an important role in the substrate recognition. Therefore often it is extremely useful for a better understanding of the function of the enzyme to be able to describe, measure and quantify the properties of the protein and nucleic acids. Various values defined further are measured in our work: hydrogen bonds lengths, sugar puckering, torsion angles and RMSD.

Hydrogen bonds are principal in determining the 3D structure of nucleic bases and proteins. It is an attractive interaction between an electronegative atom and a hydrogen atom bonded to another electronegative atom. This electronegative atom typically is fluorine, oxygen, or nitrogen. Hydrogen bonds can occur between molecules (intermolecularly), or within different parts of a single molecule (intramolecularly). Typically, the hydrogen bond is stronger than van der Waals forces but weaker than covalent or ionic bonds. The length of hydrogen bonds depends on bond strength, temperature and pressure. The bond strength itself is dependent on the temperature, pressure, bond angle and the environment. The average length of a hydrogen bond in water is 1.97 Å.

Quantitative description of the conformation of nucleic acids by itself and in the complex with an enzyme (as e.g., RNase H, which is studied in this work) is important for a deeper analysis of the substrate recognition, binding conditions and the principle of the catalytical function of the enzyme.

A nucleic acid helix assumes different geometries [5]. The conformations observed so far for the DNA are: A-DNA, B-DNA, C-DNA, D-DNA, E-DNA, and Z-DNA (more about the properties of A, B and Z conformation in Figure 1.7 [6]). However, it seems that just A-DNA, B-DNA, and Z-DNA are found in nature. The 3D structure of these conformations is depicted in Figure 1.3 [3]. The B conformation described by Watson and Crick is believed to predominate in cells. In this conformation, the double helix makes one complete turn about its axis per approximately 10 bp of sequence length. The A conformation is characteristic mainly for the RNA molecule. A-DNA seems to appear especially in dehydrated samples (used e.g. in crystallography) and in hybrid DNA/RNA duplexes. The Z conformation is unlike others left-handed and it is typical for methylated segments used in the cells for regulatory purposes. RNA differs from DNA above all by the presence of a hydroxyl group at the 2'-position of the ribose sugar. The presence of this functional group enforces the C3'-endo sugar conformation, whereas the C2'-endo conformation is typical for the deoxyribose sugar in DNA. For this reason the RNA helix adopts the A-form geometry rather than the B-form, which is most commonly observed in DNA. Geometrically, a very deep and narrow major groove and a shallow and wide minor groove result. Generally, RNA is a single-stranded molecule and has a much shorter chain

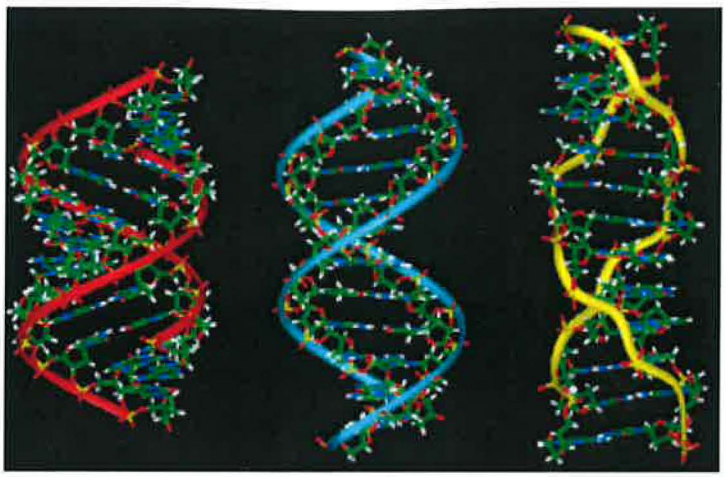


Figure 1.3: The A, B and Z conformations of a helix [3]

of nucleotides than DNA, but the hydroxyl groups make RNA less stable than DNA (it is more prone to hydrolysis). Therefore, several types of RNA contain a great deal of a secondary structure, which helps to improve their stability. A specific spatial conformation often comes up from the interaction between short complementary segments within the one RNA strand. In this way, a common spatial motif, the so called "hairpin", is constructed (see Figure 1.4). Most biologically active RNAs also may form double stranded helices. This structure is not just limited to long double-stranded helices but rather collections of short helices packed together into structures similar to proteins. In this way, RNAs can achieve chemical catalysis, like enzymes.

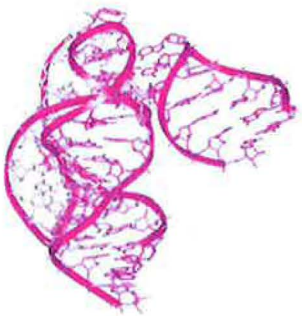


Figure 1.4: An example of RNA hairpin motif

The relative positions of atoms in a nucleic acid is described [5] above all by the torsion angles and the sugar puckering. The torsion angle is defined for four atoms (1-2-3-4) bound by chemical bonds. The torsion angle is the angle formed by the projection of bonds 1-2 and 3-4 in a plane perpendicular to the 2-3 bond direction as it is shown in Figure 1.5

[7]. The intervals of torsion angles are usually referred to as follows: clockwise 0 °-120 °

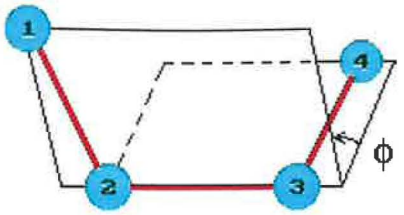


Figure 1.5: The definition of a torsion angle [7]

"gauche" (g), 120 °-240 ° "trans" (t), 240 °-360 ° "-gauche" (-g). By the inter-nucleotide conjunctions the angles describing the phosphate group are the most meaningful. Usually, the torsion angles between following atoms are measured: $C'^{(i-1)}_3$-$O'^{(i-1)}_3$-$P^{(i)}$-$O'^{(i)}_5$ and $O'^{(i-1)}_3$-$P^{(i)}$-$O'^{(i)}_5$-$C'^{(i)}_5$ (where i is the nucleotide sequence number). The state with minimal energy occurs when both the angles take up the -g-g conformation.

The sugar conformations [5] are calculated from the torsion angles about the five bonds in the ribose ring. The so called sugar puckering is a value that identifies the A and B conformations of the helices. It is given by a pseudo-rotational angle P defined by the following expression:

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin 36° + \sin 72°)}, \tag{1.1}$$

where following torsion angles are used:

$$\begin{aligned} \nu_0 &= \sphericalangle\, C'_4 - O'_4 - C'_1 - C'_2 \\ \nu_1 &= \sphericalangle\, O'_4 - C'_1 - C'_2 - C'_3 \\ \nu_2 &= \sphericalangle\, C'_1 - O'_2 - C'_3 - C'_4 \\ \nu_3 &= \sphericalangle\, C'_2 - O'_3 - C'_4 - O'_4 \\ \nu_4 &= \sphericalangle\, C'_3 - C'_4 - O'_4 - C'_1 \end{aligned} \tag{1.2}$$

The sugar puckering amplitude is then defined as:

$$\nu_{max} = \frac{\nu_2}{\cos P} \tag{1.3}$$

The B formation preferred by DNA [5] in physiological solution is given by the C2'-endo conformation, which is characterized by the value of pseudo-rotational angle $144\,° \leq P \leq 190\,°$. For the RNA, characteristic is the conformation A given by the C3'-endo sugar puckering, for which the values of pseudo-rotational angle are in the interval $0\,° \leq P \leq 36\,°$ (see Figure 1.6 [7]).

Not only for nucleic acids, but also for the whole studied system, evaluation of the root mean square deviation is important to assess the stability of a biomolecule and the volume of its conformational changes. The RMSD of atoms in the system is calculated

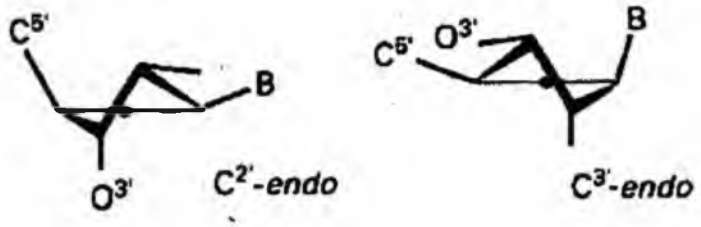


Figure 1.6: The C2'-endo and C3'-endo sugar pucker [7]

with respect to a reference structure (usually the initial one). It is defined in the following way:

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i |\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)|^2, \right]^{\frac{1}{2}} \tag{1.4}$$

where the M is the total mass and $\mathbf{r}_i(t)$ is the position of the atom i at time t.

Geometry attribute	A-form	B-form	Z-form
Helix sense	right-handed	right-handed	left-handed
Repeating unit	1 bp	1 bp	2 bp
Rotation/bp	33.6°	35.9°	60°/2
Mean bp/turn	10.7	10.0	12
Inclination of bp to axis	+19°	-1.2°	-9°
Rise/bp along axis	0.23 nm	0.332 nm	0.38 nm
Pitch/turn of helix	2.46 nm	3.32 nm	4.56 nm
Mean propeller twist	+18°	+16°	0°
Glycosyl angle	anti	anti	C: anti, G: syn
Sugar pucker	C3'-endo	C2'-endo	C: C2'-endo, G: C2'-exo
Diameter	2.55 nm	2.37 nm	1.84 nm

Figure 1.7: Properties of A, B and Z conformations [6]

1.2 RNase H enzyme

1.2.1 Function of RNase H enzyme

RNase H is a sequence-nonspecific nuclease and a member of the nucleotidyl-transferase superfamily. In all organisms, it is present with a highly conserved active site situated in the homologous catalytic domain (see Figure 1.8 [8]). RNase H specifically recognizes and cleaves the RNA strand in an RNA/DNA duplex. In a cell, RNase H serves first of all to remove RNA primers from Okazaki fragments. In unicellular organisms, the RNase H function is redundantly encoded, whereas for multicellular organisms, it is essential. It was also proved that single bacterial and eukaryotic cells often contain two RNases H, whereas single archaeal cell contain only one [9]. In retroviruses RNase H belongs to the reverse transcriptase complex [10] that converts a retroviral single-stranded RNA genome into double-stranded DNA.

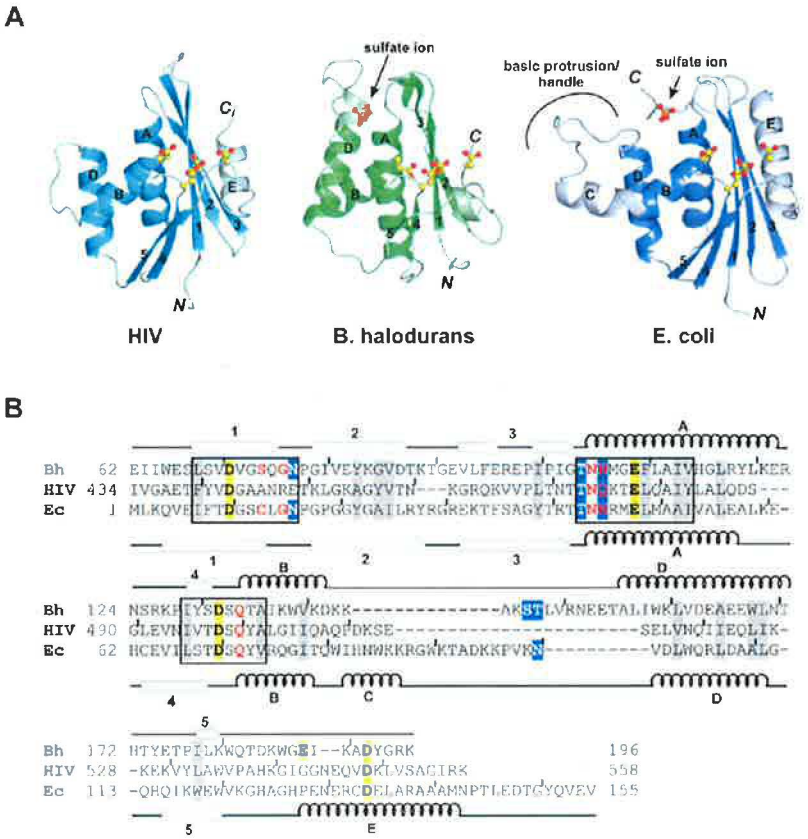


Figure 1.8: The comparison of HIV, BH and E. coli RNase H: structure(A) and sequence alignment with conserved regions highlighted (B) [8]

1.2.2 Structure and function of E. coli RNase H

The crystal structure of E. coli RNase H was first determined already in 1990 [11],[12] and therefore RNase H1 from Escherichia coli is the best characterized bacterial RNase H.

RNase H is an enzyme that specifically cleaves RNA hybridized to DNA. The enzyme, however, exhibits considerable affinity for RNA and DNA duplexes, and even single stranded oligonucleotides. Thus, RNA-DNA and RNA-RNA duplexes are bound 60-times more strongly [13].

The minimum size of an RNA-DNA hybrid for cleavage by the enzyme is four base pairs [14]. RNase H can promote the formation and cleavage of an RNA/DNA hybrid between an RNA strand and a base paired strand of a stable hairpin or duplex DNA.

Digestion of single strand RNA was observed exclusively within the 3' overhang region and not the 5' overhang region [15]. The number of single strand RNA residues cleaved by RNase H is influenced by the sequence of the single strand RNA immediately adjacent to the RNA-DNA duplex and appears to be a function of the stacking properties of the RNA residues adjacent to the RNA-DNA duplex.

The amino acid residues essential for the catalytic activity of ribonuclease H from Escherichia coli were first identified by site-directed mutagenesis. In this way Asp10, Glu48, and Asp70 were found [16] to be crucial for RNase H activity (Figure 1.9 [17]). These residues are fully conserved in the amino acid sequences of not only retroviral RTs but also plant viral and retrotransposon RTs (Figure 1.10 [18]).

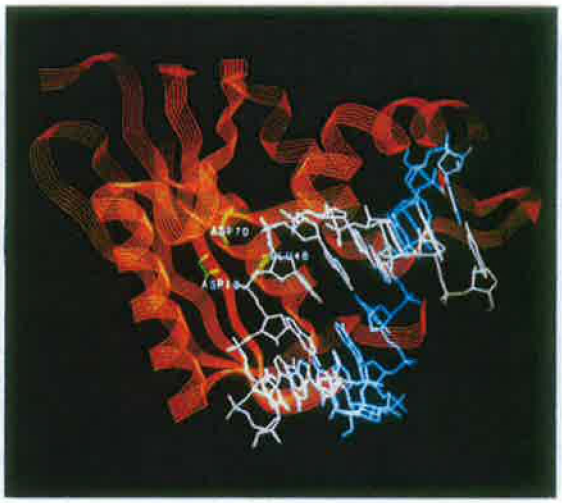


Figure 1.9: A DNA/RNA hybrid complex model with E. coli RNase H (RNA-strand white, DNA-strand blue, three carboxylates Asp10, Glu48 and Asp70 shown) [17]

Determination of the kinetic parameters [19] revealed that replacement of Asp134 by amino acid residues other than asparagines and histidine dramatically decreased the enzyme activity without seriously affecting the substrate binding. These results strongly

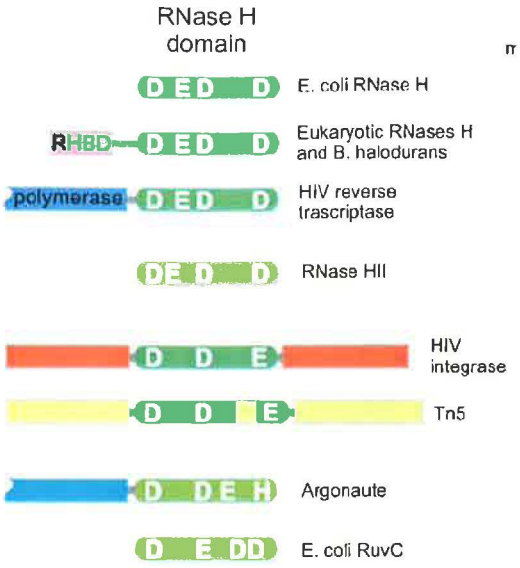


Figure 1.10: Active sites of RNases H and its relatives [18]

suggest that Asp134 does not contribute to the maintainance of the molecular architecture but the carboxyl oxygen at its position impacts catalysis.

Residues Cys13, Asn16, Asn44, Asn45, and Gln72 are not as essential [17] as residues Asp10, Glu48 or Asp70, but their replacement reduced substrate binding to some extent. Gln76 and Gln80 do not seem important for the enzyme's action.

The handle region (residues 84-99) in ribonuclease HI from Escherichia coli, which is rich in basic amino acids residues, was examined by scanning mutagenesis in another study [20]. It was suggested that the cluster of positive charges in the handle region is important for the effective binding of the substrate. Interestingly, the region of human immunodeficiency virus reverse transcriptase with homology to E.coli RNaseHI lacks the handle region, which may account for the poor RNase H activity of the domain when separated from the polymerase domain.

Another study [21] exploited a fact, that proteins are usually digested by protease at multiple sites. However, the susceptibilities to proteolytic digestion varies for different sites. For example, peptide bonds located within a loose conformation are usually more susceptible to proteolytic digestion than those located within a rigid conformation. If a given protein formed a stable complex with certain materials and the most susceptible sites to proteolytic digestion were located at the interface between these two molecules, this site would be protected from proteolytic digestion in the presence of these materials. Such a protein footprint analysis was used to identify the nucleic-acid-binding site of the mutant protein E.coli RNase H D10/E48R, because this mutant protein forms a stable complex with nucleic acids. Therefore, the mutant protein was digested with lysyl

endopeptidase, trypsin, or chymotrypsin in a limited manner in the presence or absence of poly(rA)/poly(dT). The protein footprint analyses indicated that the specific peptide bond probably between Tyr151 and Glu152 of the mutant protein was protected from the chymotryptic digestion in the presence of poly(rA)/poly(dT). Tyr151 is located in the groove-like depression of the protein molecule, which extends from the active site to the basic protrusion and has been suggested to be involved in substrate binding.

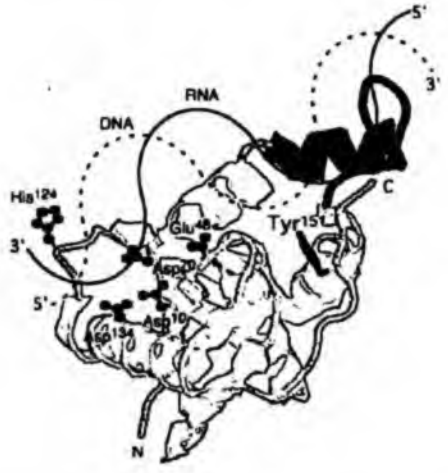


Figure 1.11: A model for the complex between E. coli RNase H1 and DNA/RNA hybrid: the solid region represents the basic protrusion [21]

In the complex model (Figure 1.12) proposed in [22], the scissile bond of the RNA backbone should be fixed near the Mg^{2+} - binding site. The 5'-phosphate group adjacent to the scissile bond then approaches the main-chain atoms Ser71 and Gln72. The 5'-phosphate group, two nucleotides away from the scissile bond, is close to the amino group of Lys122 and the amido group of Asn44. A glycine-rich sequence, Gly11-Ser12-Cys13-Leu14-Gly15-Asn16-Pro17-Gly18-Pro19-Gly20-Gly21-Tyr22-Gly23, forms a long, flexible loop, which is involved in the binding of DNA/RNA hybrid. This suggestion was confirmed by NMR measurements. The neighboring RNA backbone on the 3'-phosphate side is located near Ser12, Cys13, Leu14 and Gly15 in the glycine-rich flexible loop. The basic protrusion (see Figure 1.11 [21]) interacts in the model mainly with the RNA strand. The backbone phosphate groups, which are located 10 residues away from the scissile bond, face the main-chain imino group of Asp94. Some of the basic side-chains are interacting with the RNA backbone phosphate groups by electrostatic forces. On the other hand, the complementary DNA strand is located near Asn16 and Lys99 in the basic protrusion. Chains of Gln76, Gln80 and Trp81 form the kink between αII and αIII helices. His124 lying besides major groove site and Arg138 in αV helix is assumed to interact with the phosphate group that is two nucleotides away from His124 in the same backbone. Thus, roughly two turns of a DNA/RNA hybrid are sufficient to cover the entire interface, ranging from the amino end

of αV helix to the basic protrusion. The model seemed to be supported by experimental findings. Substitutions by site-directed mutagenesis of Ala for Cys13, Ans16, Asn44, Asn45 and Gln72 were found to reduce the affinity of RNase H for substrate. In addition, the replacement of Lys122 by Asn reduced the relative activity to 25%, and the replacements of His124 by Ala, and Arg138 by Cys, reduced activities to 2.5% and 60% of that of the wild-type, respectively.

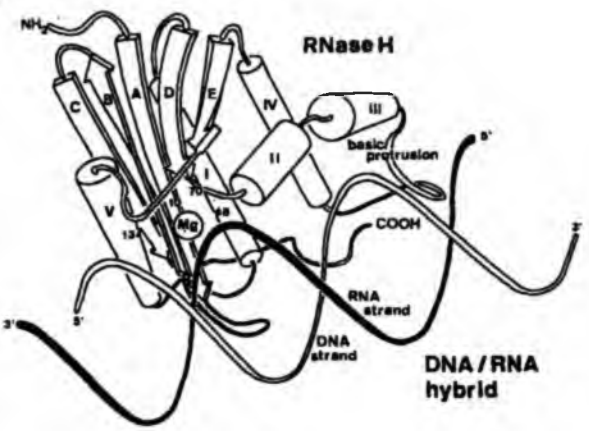


Figure 1.12: The DNA/RNA hybrid is shown to bind to the surface of the enzyme along the line connecting Mg^{2+} ions and the basic protrusion [22]

A model for the complex between E.coli RNase H1 and the DNA/RNA hybrid was refined [23] by molecular dynamics simulations. Many residues involved in binding of the DNA (Arg88, Asn84, Trp85, Trp104, Tyr73, Lys99, Asn100, Thr43 and Asn16) and RNA (Gln76, Gln72, Tyr73, Lys122, Glu48, Asn44 and Cys13) strand to the substrate-binding site of the RNase H enzyme were found. The most remarkable disturbance of the hydrogen bonding net was observed for structures with modified internucleotide linkages positioned in a way to interact with the Trp104, Tyr73, Lys99 and Asn100 residues (situated in the middle of the DNA binding site, where a cluster of Trp residues forms a rigid core of the protein structure).

Interestingly, active sites of enzymes either synthesizing or degrading nucleic acids are highly similar (see Figure 1.13 [18]). In all DNA and RNA polymerases and many nucleases and transposases, two Mg^{2+} ions are jointly coordinated by the nucleic acid substrate and catalytic residues of the enzyme [18].

In the Figure 1.14 [18] various phosphoryl transfer reactions related to the nucleic acid processes are illustrated: This figure shows nucleotide polymerization (A), nucleic acid hydrolysis (B), first cleavage of an exon-intron junction by a group I ribozyme (C) and by a group II ribozyme (D), strand transfer during transposition (E) and exon ligation during RNA splicing (F).

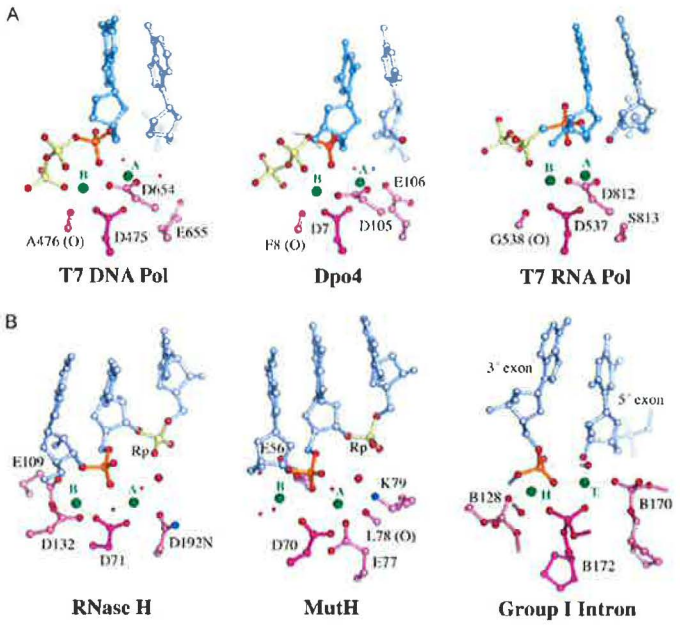


Figure 1.13: The coordination of Mg^{2+} ions in several nucleases and polymerases [18]

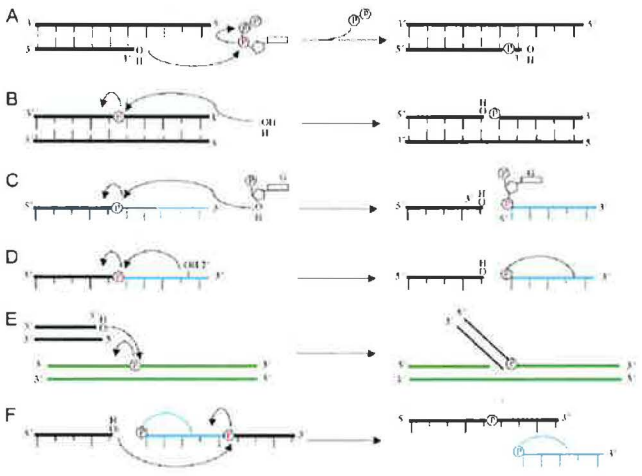


Figure 1.14: Phosphoryl transfer reactions [18]

In DNA-dependent RNA polymerases, reactions of RNA synthesis and degradation are performed by the same active center (in contrast to DNA polymerases in which they are separate) [24]. The active center involves a symmetrical pair of Mg^{2+} ions that switch roles in synthesis and degradation. One ion is retained permanently and the other is recruited ad hoc for each act of catalysis. The weakly bound Mg^{2+} is stabilized in the active center in different modes depending on the type of reaction.

Several catalytic mechanisms were proposed for E.coli RNase H1. In [25] three different possible catalytic mechanisms of E. coli RNase H1 were suggested (see Figure 1.15 [25]): with one Mg^{2+}, two Mn^{2+} and one Mn^{2+} ions in the active site. The two water molecules act as general acid and general base. Only in the two-Mn^{2+} ion model all four conserved residues are necessary for coordination of the ions. The schema (A) illustrates a possible involvement of Glu48 in general acid-base mechanism for the hydrolysis of the P-O3' bond by E. coli RNase H1. In a prior study [26] it is suggested that the Glu48 residue is not responsible for the binding of the ion, but required to anchor one of the water molecules (Water A). This study also lays out that the Mg^{2+} ion interacts with the 2'-hydroxyl group, instead of the phosphate group, stabilizing the transient intermediate.

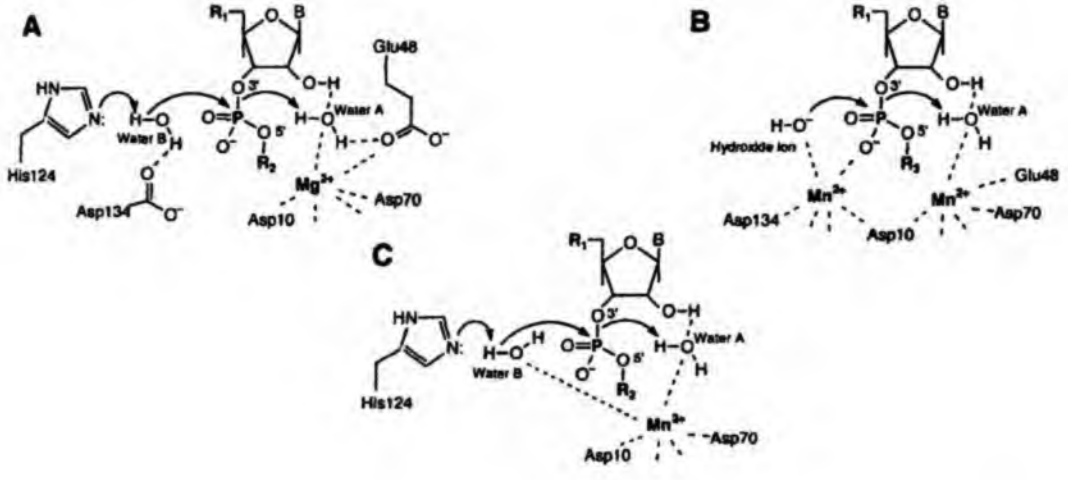


Figure 1.15: Three different proposed mechanisms of the catalysis [25]

Another study [27] proposes a possible involvement of the phosphate group next to the scissile phosphodiester bond in the 3'-direction (see Figure 1.16 [27]). Substitution of the pro-Rp-oxygen of the phosphate group with sulfur results in an uncharged double-bonded oxygen at the Sp-position. It significantly reduces the k_{cat} value of the wild-type RNase H, whereas it has little effect on that of the H124A mutant enzyme. These results suggest that this phosphate group functions cooperatively with His124 for catalysis. His124 may function as a general base. Thus, a possible role of the phosphate group is to enhance the catalytic efficiency by removing a proton from His124 [27]. The pKa of a nonbridging phosphate oxygen is too low to efficiently accept a proton from His124. The phosphate

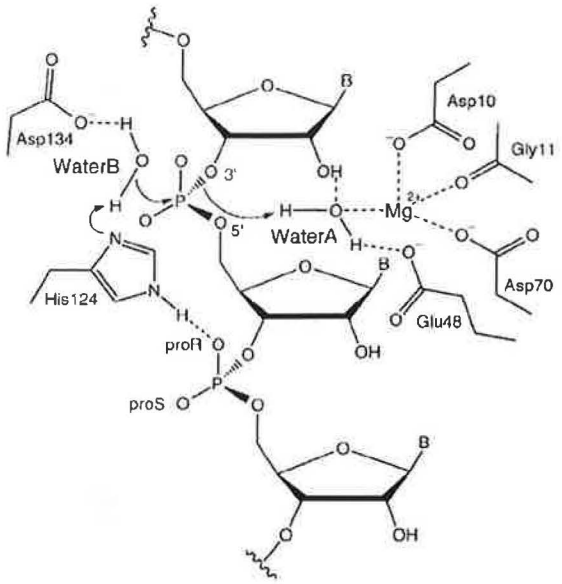


Figure 1.16: A mechanism of the catalysis proposed by Haruki [27]

group is also not likely to operate as a general base by directly deprotonating a water molecule for the same reason.

Thereof, it was put forward that the phosphate group contributes to orient His124 to the best position for the activation of the attacking hydroxide ion through the formation of a hydrogen bond between the imino proton of His124 and the pro-Rp-oxygen. This hydrogen bond could also elevate the pKa of His124, thereby facilitating the generation of the attacking hydroxide ion [27].

1.2.3 Bacterial RNase H in complex with a substrate

In 2006 the crystal structure of bacterial RNase H in catalytically deficient forms bound to an RNA/DNA structure was crystallized [8] and so it was possible to study the highly conserved active site (see Figure 1.17) during the interaction with the substrate.

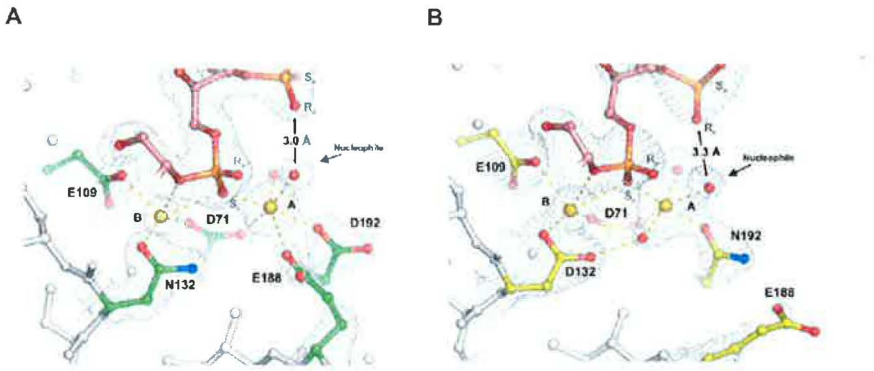


Figure 1.17: The active site of D132N and D192N in complex with RNA/DNA hybrid [8]

For the whole superfamily, the presence of two divalent metal cations Mg^{2+} or Mn^{2+} in the active site is characteristic. In nucleases, the two metal ions are asymmetrically coordinated and have distinct roles in activating the nucleophile and stabilizing the transition state. The crystal structures of Bacillus halodurans RNase H - RNA/DNA substrate complexes were obtained with resolution 1.5 - 2.2Å [8]. Further, the structures of reaction intermediate and product complexes [28] of RNase H at 1.65 - 1.85Å were recently reported, too (see Figure 1.18, oxygen atoms in red, nitrogen in blue).

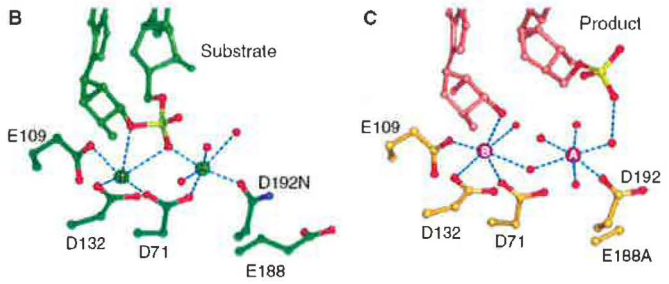


Figure 1.18: A comparison of metal ion coordination in the substrate (green) and product (orange) [28]

In two-metal catalysis, metal ion A (see Figure 1.19 [28]) has been proposed to activate the nucleophile and metal ion B to stabilize the transition state. The movement of the two metal ions (see Figure 1.19) suggests how they may facilitate RNA hydrolysis during

the catalytic process [28]. Firstly, metal ion A may assist in the nucleophilic attack by moving towards metal ion B and bringing the nucleophile close to the scissile phosphate. Secondly, metal ion B transforms from an irregular coordination in the substrate complex to a more regular geometry in the product complex. The exquisite sensitivity of Mg^{2+} to the coordination environment likely destabilizes the enzyme-substrate complex and reduces the energy barrier to form the product. Lastly, product release probably requires dissociation of metal ion A, which is inhibited by either high concentrations of divalent cations or mutation of an assisting protein residue.

One of the water molecules (see Figure 1.17) is 3.4Å away from the scissile phosphate and is perfectly positioned for an in-line nucleophilic attack [8]. This water molecule does not contact any protein residues, but it is only 3Å from the pro-Rp oxygen of the phosphate immediately neighboring the scissile bond. This pro-Rp oxygen most likely orients the water for the nucleophilic attack and may even serve as a general base for deprotonation and then shuttle proton to solvent.

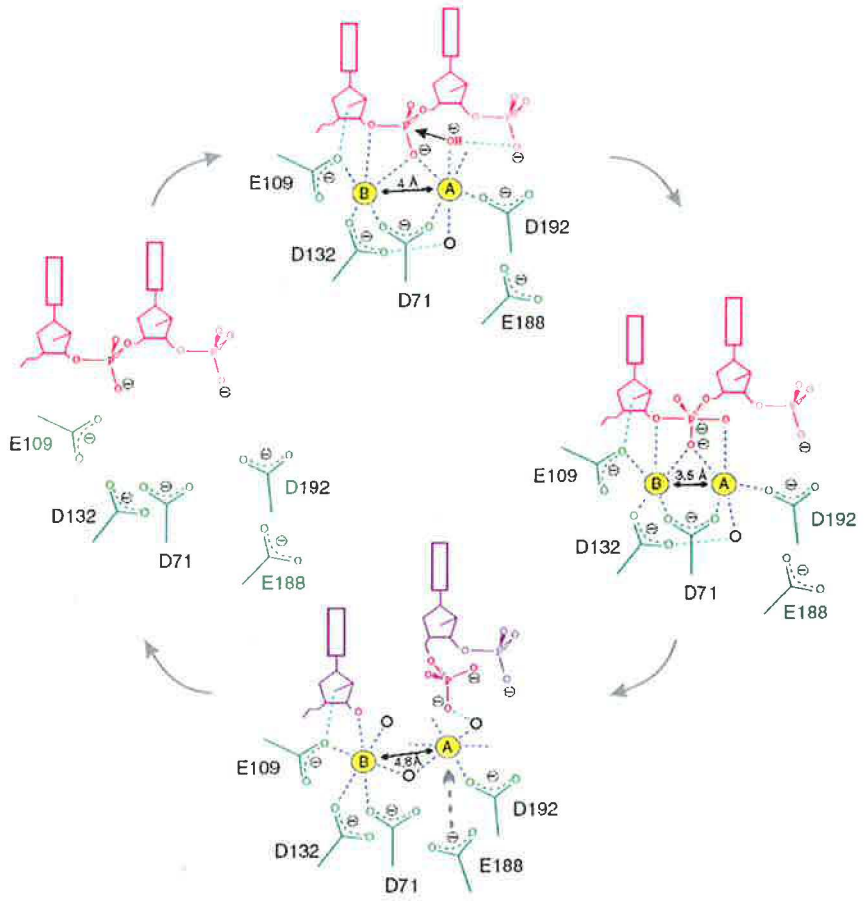


Figure 1.19: A schematic representation of the reaction steps proposed for RNase H [28]

By contrast, no metal ion was found at the active site in apoprotein crystals grown at the same Mg^{2+} concentration. It seems that both metal ions are essential and their position must be correct for a successful catalysis. The distance between the two ions is determined by the configuration of the highly conserved active site carboxylates (so called DDE motif) and substrate. The influence of the substrate on the metal-ion coordination ensures the catalytic specificity of RNases H. It was demonstrated, that the RNA/DNA hybrid in the complex with RNase H adopts a mixed A and B conformation. The RNA strand was found in A form with 3'-endo sugar puckers, and the DNA strand was found in the B-form. RNases H select and stabilize an RNA/DNA hybrid in the mixed conformation.

The DNA binding groove is upheld by a phosphate binding pocket. In the apoprotein structure sulfate ion is bound in it, which is then perfectly replaced by the DNA phosphate.

1.2.4 Human RNase H

There are various types of RNases H in different living organisms, characterized by the conserved active site but differing in many aspects. Two RNase H enzymes (H1 and H2) have been cloned [29] and shown to be expressed in human cells and tissues. To determine the role of RNase H1 the levels of the enzyme in human cells and mouse liver were varied and the correlation of those levels with the effects of a number of DNA-like antisense drugs (see Section 1.3) were determined. Results demonstrate that in human cells RNase H1 is responsible for most of the activity of DNA-like antisense drugs. Further, there are several additional previously undescribed RNases H in human cells that may participate in the effects of DNA-like antisense oligonucleotides. Therefore, human RNase H1 was chosen as the object of studies of this work. According to [30] the mRNA levels of RNase H1 and RNase H2 varied considerably in the human cell lines. The intracellular localization of the enzymes, assayed by green-fluorescent protein fusions, showed that RNase H1 was present throughout the whole cell for all cell types analyzed, whereas RNase H2 was restricted to the nucleus.

Human RNase H1 is a 286-amino acid protein and is expressed in all human cells. The structure of RNase H1 consists of a RNA-binding 73-amino acid region and a catalytic region with a DNA-binding domain and the active site [31]. The highly conserved catalytic region, which is essential for the activity, is connected to the RNA-binding domain by means of a 62-amino acid spacer region (see Figure 1.20 [31]). The catalytic domain is common for all types of RNases H, while the RNA-binding domain is found only in eucaryotes. The spacer region is required for a correct human RNase H activity, whereas the RNA-binding domain is not essential.

The catalytic region of human RNase H is highly homologous with RNases H from several organisms, as E. coli, yeast, chicken, mouse. The conserved 100-residue core-structure consists of a five stranded β-sheet and three α-helices. Human RNase H1 is active only under reduced conditions. Oxidation ablates the cleavage activity. The cysteine residues responsible for the observed redox-dependent activity of human RNase H1 were determined by site-directed mutagenesis to involve Cys147 and Cys148. Experimental data show that the formation of a disulfide bond between adjacent Cys147 and Cys148 residues

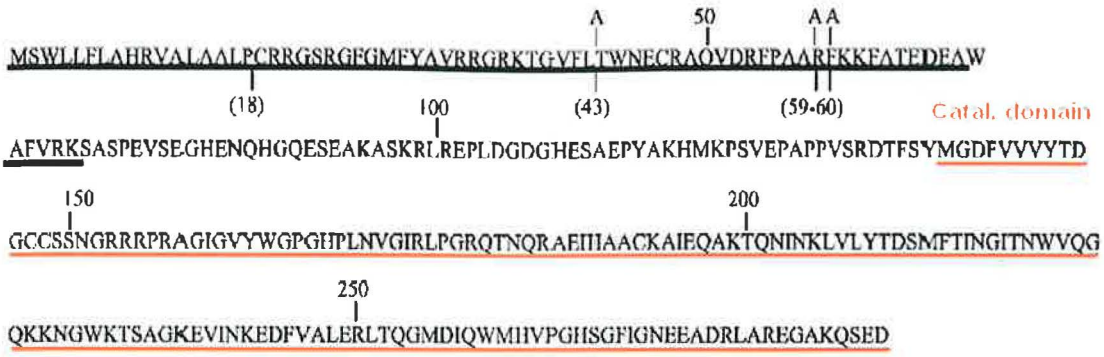


Figure 1.20: The human RNase H1 sequence: black underlined RNA-binding domain, red underlined catalytic domain and the spacer region between them [31]

results in an inactive enzyme conformation [32].

The RNA-binding domain of human RNase H1 is homologous with other eukaryotic RNases H1. It has been shown that this region is responsible for enhanced binding activity of the human RNase H and in particular for the strong positional preference for cleavage. Human RNase H1 cleaves the heteroduplex substrate between 8 and 12 nucleotides from the 3'DNA/5'RNA terminus, whereas e.g. the procaryotic E. coli RNase H with a lack of RNA-binding domain does not exhibit strong positional preference [31]. The positional preference of human RNase H upholds the theory about the biological role of RNase H, that is the removal of RNA primers during lagging strand DNA replication, as the length of the RNA primers is in average 7-14 nucleotides. Within the preferred cleavage site, the enzyme displays modest sequence preference with GU being a preferred dinucleotide [33]. The minimum RNA-DNA duplex length that supports cleavage is 6 base pairs, and the minimum RNA-DNA gap size that supports cleavage is 5 base pairs [34]. The RNA-binding domain probably interacts with the 3'DNA/5'RNA terminus of the heteroduplex. It was also shown [31] that the conserved Trp43, Lys59 and Lys60 residues constitute an extended nucleic binding surface responsible for the positional preference. Mutants, in which these residues have been substituted with alanine, exhibited a loss of the positional preference [31].

1.3 Antisense technique

1.3.1 General information about the antisense technique

The idea of the usage of oligodeoxynucleotides to inhibit viral replication came in 1978 from Zamecnik and Stephenson. The so called "antisense" strategy is based on the interaction of oligodeoxynucleotides with mRNA, in such a way, that its translation into protein is disabled [35]. This is also the most remarkable difference between the antisense method and conventional drugs, which bind directly to proteins modifying their function.

There are three main antisense strategies, differing in utilization of different tools as it is illustrated in Figure 1.21 [35]. First of them, antisense oligonucleotides, pair with their complementary mRNA. Second, ribozymes and DNA enzymes, that bind but can also cleave their target RNA. Finally, small interfering RNA (siRNA) molecules, that are bound by the RISC and induce degradation of the target mRNA.

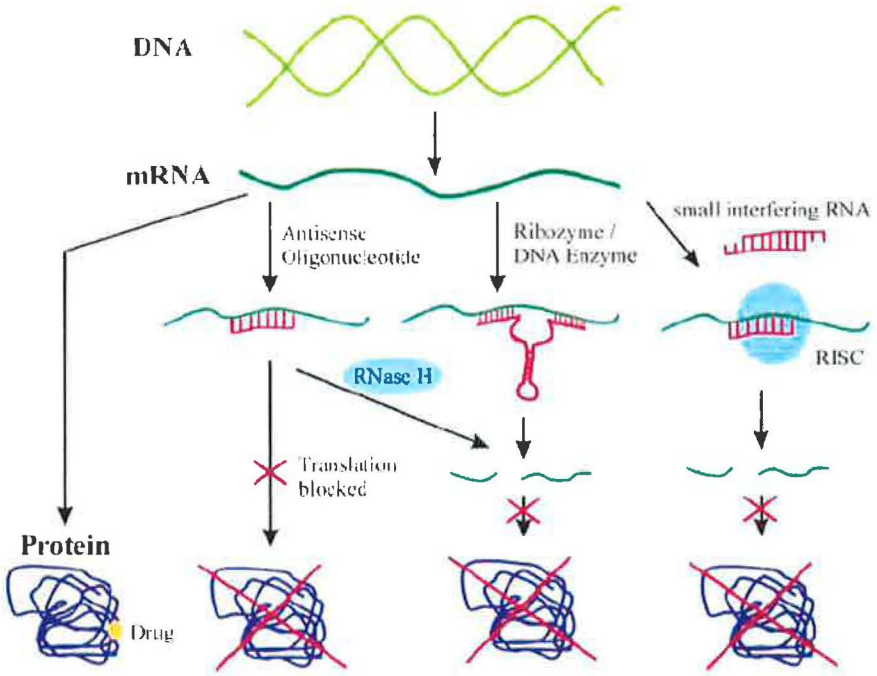


Figure 1.21: Different antisense strategies in comparison with a conventional drug [35]

In the present overview, mainly antisense-oligonucleotides (ASO) will be discussed. ASOs are single-stranded chemically modified DNA-like molecules, which consist of 15-20 nucleotides complementary to their target mRNA. They are designed to activate the enzyme RNase H, which cleaves the RNA part of the RNA:DNA heteroduplex (Figure 1.22 [35]). The target mRNA is cleaved and the ASO becomes free to bind another mRNA.

Those ASOs, which do not induce RNase H, may inhibit translation by steric blocade of the ribosome.

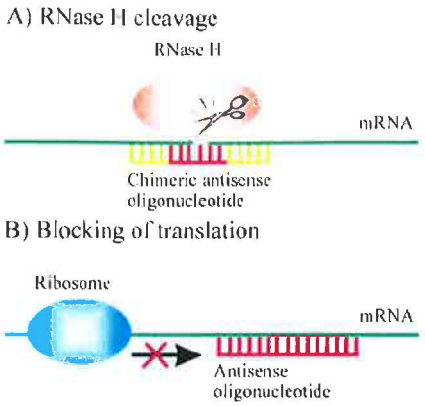


Figure 1.22: The RNase H cleavage (A) and blocking of translation (B) [35]

The method of gene silencing could have a broad range of therapeutic use in every case, where the disease is caused by the expression of a deleterious gene. This is the case of viral infections, inflammatory diseases or cancer growth. ASOs could be used also to correct aberrant splicing. Furthermore, they can serve as a powerful tool for target validation. The main advantages of this method are a broad applicability, direct utilization at low costs and a high specificity.

As it was mentioned above, antisense oligonucleotides could serve to target genes involved in cancer progression, which is especially useful in case of those not hittable by the small-body or antibody inhibition. Thanks to the improved understanding of the molecular mechanisms of cancer progression and therapeutic resistance, many new therapeutic gene targets arose. The most promising AS targets for the cancer therapy are those genes associated with tumor progression and therapeutic resistance. This comprises primarily genes that regulate apoptosis, proliferation and cell signaling [36]. As example can serve the BLC2 gene. It is an oncogene that enhances cell survival through inhibition of apoptosis. A similar effect have also Survivin or XIAP. Another examples are the HSPs, molecular chaperons. They are induced during stress responses and prevent precipitation and aggregation of various proteins. The expression of HSP27 is induced by hormone or chemotherapy and inhibits treatment-induced apoptosis through multiple mechanisms.

On the other hand, several problems appeared that obstruct a vast therapeutical usage of the antisense method, as the stability of ASOs in a living cell, efficient cellular uptake and toxicity. Various problems that arise in context of usage of the ASOs for therapeutical purposes will now be discussed.

There are several rules to be followed, when designing ASOs. The first task for a successful antisense approach is to identify accessible target sites on the mRNA. It is necessary to identify such sequences that are both unique to that particular gene and

accessible to hybridization. A significant homology with other mRNAs must be avoided. Long RNA molecules form complex secondary and tertiary structures and a great part of the sequence is not available on the surface of the molecule to be attacked by the ASOs. Computer-based structure models of long RNA molecules are unlikely to represent the real RNA structure inside a living cell. Therefore several strategies have been developed, as the usage of random or semirandom oligonucleotides libraries. ASOs must be synthesized in a form, which is not toxic or at least well tolerated in patients. ASOs containing CpG motifs should be avoided as they may stimulate immune responses in mammalian systems [35]. The CG dinucleotide is more frequently found in viral and bacterial DNA and therefore it is a marker for the immune system to signify infection.

Another uneasy problem is the cellular uptake of the molecules. In the in vivo experiments the delivery of ASOs into the cells is insufficient by naked DNA. The charged ASOs in this case have to cross a hydrophobic cell membrane. Therefore many methods have been developed, among others the use of delivery systems as liposomes and charged lipids. These complexes are usually internalized by endocytosis. Several macromolecular delivery systems have been developed recently that mediate an efficient ASOs distribution, as biodegradable polymers and ASO-binding nanoparticles. Another strategy, when targeting ASOs to specific tissues or organs, is the receptor mediated endocytosis. ASOs are conjugated to antibodies or ligands specifically recognized by a certain receptor, which mediates their cellular uptake.

1.3.2 Modified oligonucleotides

One of the major challenges for the antisense approach is the stabilization of ASOs. Natural oligodeoxynucleotides are rapidly degraded in living cells by nucleases. A great number of modifications have been introduced to overcome this problem. Usually, the modifications are classified in three groups. The first group, the first generation of modifications, comprises modifications, where one of the non-bridging oxygen atoms in the phosphodiester bond is replaced by sulfur (see Figure 1.23a [35]). These modified ASOs are known as phosporothioates. The introduction of this modification led to an enhancement of the nuclease resistivity of ASOs in vivo. In human serum unmodified oligonucleotides have a half-life cca 1 h, thanks to the modification it is increased to 9-10 h. Phosporothioates also form regular Watson-Crick base pairs and are able to activate RNase H, and they carry negative charges for cell delivery. Their major disadvantages are nonspecific interactions e.g., their binding to certain proteins, which may cause cellular toxicity. Despite this fact the first generation ASOs can be used for therapeutical purposes, and lower doses used for clinical trials in humans were generally well tolerated. Actually, the first (and till now the only one) antisense drug Vitravene, which was approved in 1998, is based on phosphorothioates. It serves for treatement of cytomegalovirus-induced retinitis in patients with AIDS. The group of second generation ASOs contains nucleotides with alkyl modifications at the 2' position of the ribose ring (Figure 1.23b). They are less toxic than phosphorothioates and have a slightly enhanced affinity toward the complementary RNAs. On the other hand, they are not able to induce RNase H activity. It was shown that for

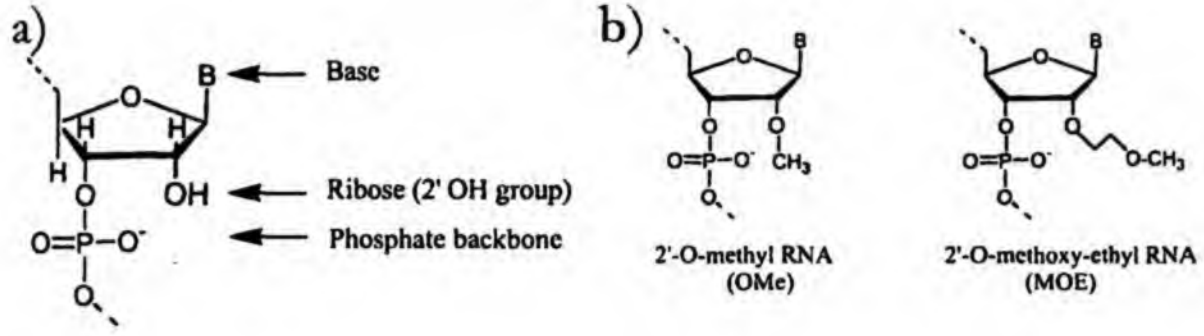


Figure 1.23: The first(a) and the second(b) generation ASOs [35]

efficient RNase H activity the correct width of the minor groove of the heteroduplex, its flexibility and availability of the 2'OH group is necessary. Modified heteroduplexes containing modifications predicted to mimic the sugar pucker and conformational flexibility of the deoxyribonucleotide exhibited cleavage rates comparable with those of the unmodified substrates [37]. The second generation ASOs do not fulfill these conditions. The antisense effect of these modified ASOs can therefore be due to a steric block of translation.

In order to increase antisense potency, however, the cleavage of target mRNA is desired. Therefore the "gapmer technology" was invented. It serves to enable the cleavage despite of the modification. The ASO nuclei are formed from DNA or phosphorothioate ASOs in the middle and there are second-generation ASO residues at the ends. The end blocks help to prevent nucleolytic degradation and the central part is sufficient for activating human RNase H. The gap requirements of RNase H1 within the sense strand were examined too. Results from these studies showed that RNase H1 requires at least five or six natural RNA residues within the sense RNA strand.

Evaluation of antisense activities of uniformly 2'-modified oligonucleotides revealed that these compounds were completely ineffective in inhibiting Ha-ras gene expression [38]. Nevertheless, activity was restored if the compound contained a stretch of at least five 2'-deoxy residues. Nuclease stability and antisense activity imparted by these sugar modifications in phosphodiester backbones correlated with the size of the 2'-alkoxy substituent (pentoxy > propoxy > methoxy > deoxy) [39]. The gapmer technology was found as fruitful in the case of methylphosphonate, phosphonoacetate and phosphonoformate oligonucleotides. Interestingly, the 5-4-5 methyl
phospohonates-phosphates-methylphosphonates construct, having a T_m of about 37°C, was at this temperature more than 4-times more effective at eliciting RNase H hydrolysis of mRNA than the natural congener of $T_m = 51$°C [40, 41].

The enzyme RNase H was found [42] to exhibit equal affinity for both the wild type (RNA:DNA) oligonucleotide substrate and heteroduplexes containing various 2'-sugar modifications, while the cleavage sites for these chemically modified substrates were without exceptions slower than for the wild type substrate. The introduction of a single positively

charged 2'-propoxyamine modification into the chimeric antisense oligonucleotide portion of the heteroduplex substrate resulted in both decreased binding affinity and a slower rate of catalysis by RNase H. The cleavage rates for heteroduplexes containing singe-base mismatch sequences within the chimeric oligonucleotide portion varied depending on the position of the mismatch but had no effect on the binding affinity of the enzyme.

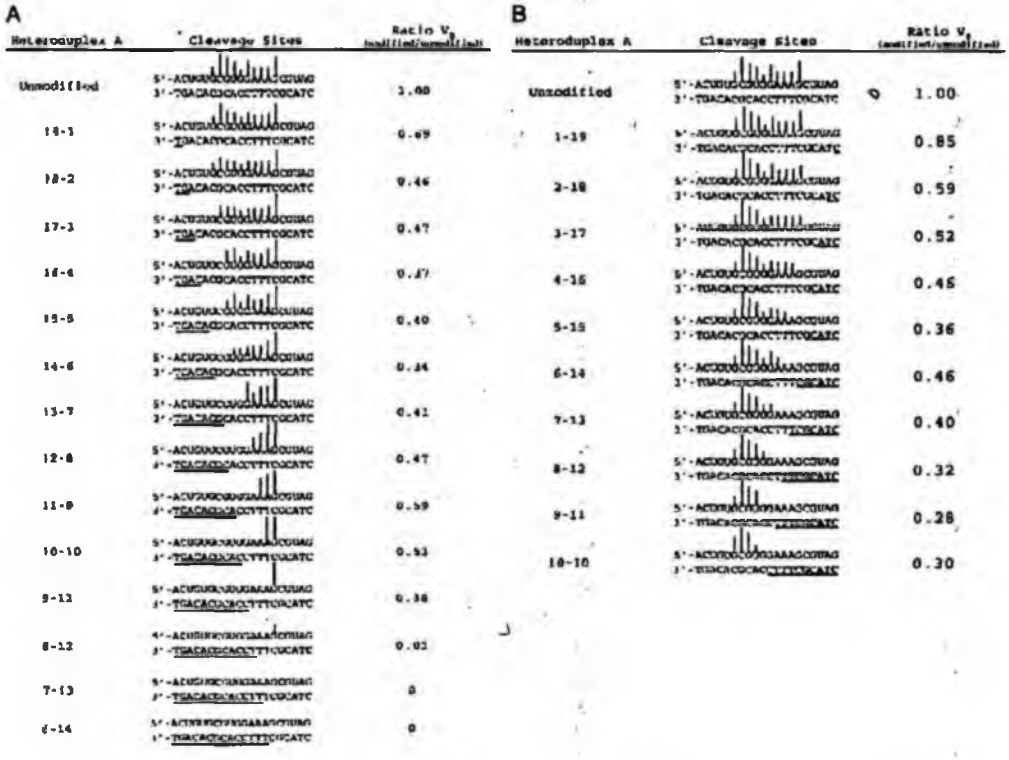


Figure 1.24: Differences in the interaction of RNase H with a MOE modified substrate based on its position [39]

The MOE substitutions at the 3' and 5' poles of the ASO resulted in fewer cleavage sites and slower cleavage rates compared with the unmodified substrates [43]. Figures 1.24 and 1.25 represent the results introduced in [43] about the dependence of initial RNase H1 cleavage rates on the position of the modified heteroduplexes. Lines indicate the position of the enzyme cleavages of the substrate: the length and thickness of the lines represent the intensity of the bends on the polyacylamide gel for each respective site. Underlined sequences indicate the position of the modifications, column (A) successive substitution of MOE at the 3' pole, column (B) at the 5' pole of the ASO. Ratio V_0 represents the initial cleavage rates. Similarly, Figure 1.25 represents results for chimeric ASO configurations of various heteroduplex sequences: (A) heteroduplexes containing unmodified, 5-10-5 and 25-10-8 chimeric ASOs, (B) heteroduplexes with chimeric and unmodified ASOs with similar 3'DNA termini. The results are schematically presented in Figure 1.26 [43]. Con-

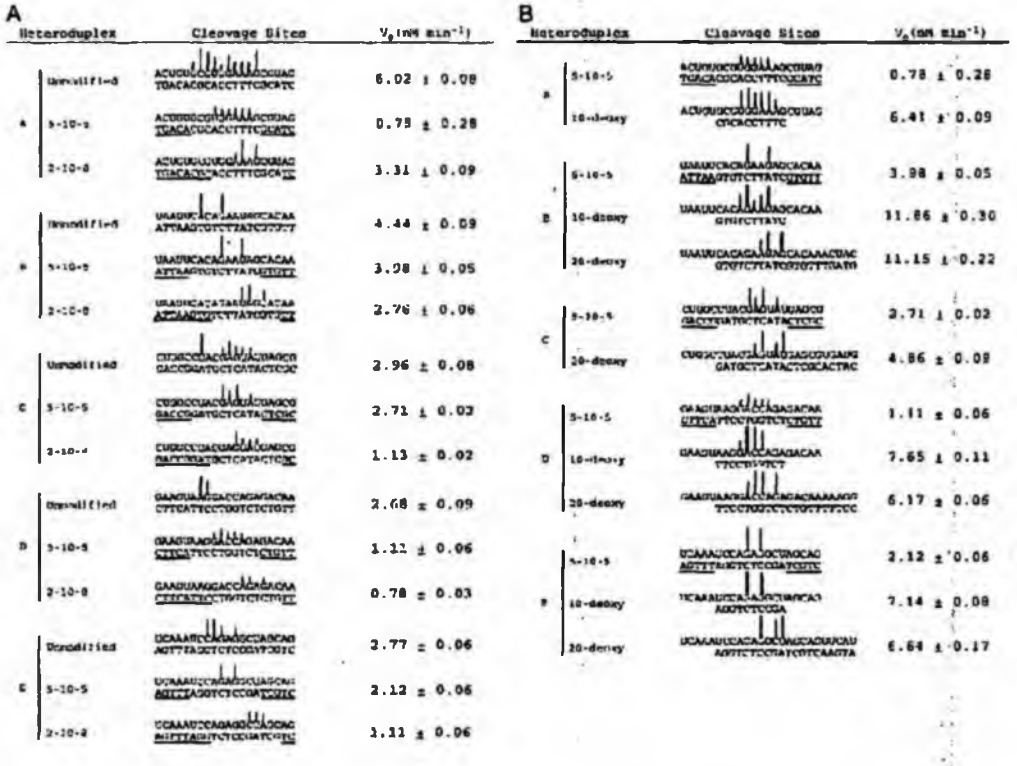


Figure 1.25: Relative cleavage rates and cleavage sites for chimeric heteroduplex substrate [39]

formational transmission of the MOE residues to the adjacent deoxyribonucleotides (A). The orange structures represent the C3'-endo A-geometry, whereas blue is the preferred geometry of the heteroduplex. The putative enzyme/nucleotide interactions (brown and navy structures) were studied with insertion of cleavage rate enhancing (yellow) and reducing (green) junction modifications (B). The arrow indicates the position of the scissile bond. (C) shows the schematic position of the RNA-binding (RNABD) and catalytic (cat) domains on the chimeric heteroduplex.

Furthermore, the heteroduplex substrates exhibited preferred cleavage sites that were cleaved 2- to 3-times faster than other sites in the substrate, and these sites exhibited the greatest influence on the initial cleavage rates. Figure 1.27 [43] illustrates the interaction of RNase H1 with the heteroduplex containing MOE modifications at the 3' (green box) and 5' (red box) poles of the ASO. Alteration in helical geometry or steric interference by the MOE subtitutions at the 3'-pole of the ASO (A) disrupts the binding interaction between the RNA-binding domain (RNA-BD) and the heteroduplex resulting in the observed ablation of of catalytic activity. MOE substitutions at the 5' pole of the ASO (B) are positioned adjacent to the catalytic domain (cat) of the enzyme.

A set of four hybrid duplexes formed by ASOs, each containing one C3'-endo locked

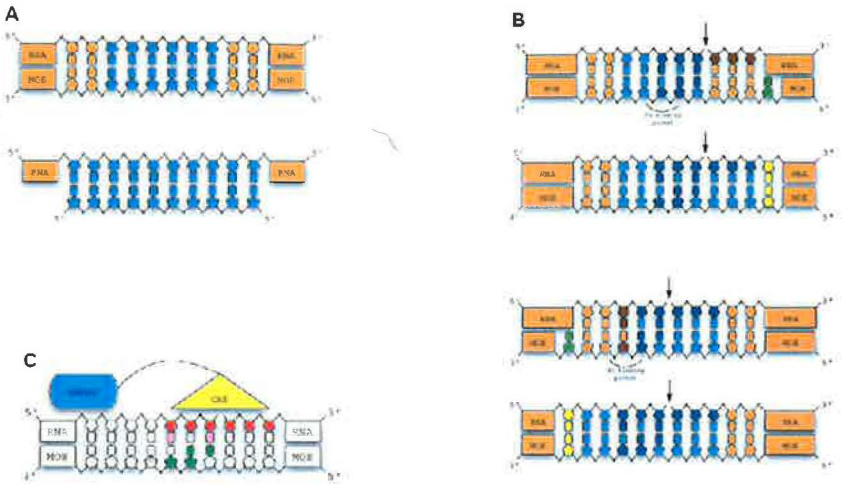


Figure 1.26: A model of the interaction of RNase H with the chimeric heteroduplex [39]

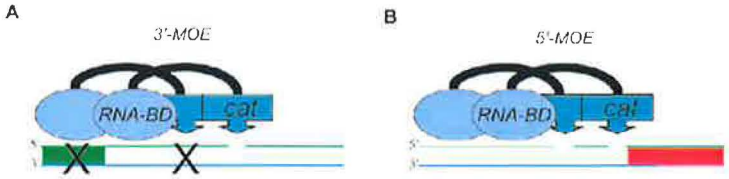


Figure 1.27: A model for the interaction of RNase H with the heteroduplex containing MOE modifications [39]

block at different sites and a matched 15-mer RNA, were subjected to the RNase H cleavage reaction [44], showing how the transmission of the local conformation owing to a single C3'-endo locked sugar steers the conformational changes of the neighboring nucleotides. Local structure perturbances upon introduction of a single C3'-endo constrained thymine spreads up to four nucleotides toward the 5'-end of the modification site. Further, conformationally biased and flexible modified nucleotides were positioned [45] at the junctions between the DNA and MOE residues of the chimeric substrates to modulate the effects of the MOE residues on human RNase H1 activity. The strong C3'-endo biased locked nucleic acid modification aggravated the negative effects of the MOE modifications resulting in slower human RNase H1 cleavage rates. Enhanced cleavage rates were observed for the eastern-biased 2'-ara-fluorothymidine and bulge inducing N-methylthymidine modifications positioned at the 5'-DNA/3'-MOE junction as well as the C3'-endo-biased 2'-methylthiothymidine and conformationally flexible tetrafluoroindole modifications.

The third generation ASOs (Figure 1.28 [35])are DNA and RNA analogs with modified

phosphate linkages or riboses and nucleotides with a completely different chemical moiety substituting the furanose ring. This group includes many different novel modifications and some of them seem to have a great potential as antisense molecules.

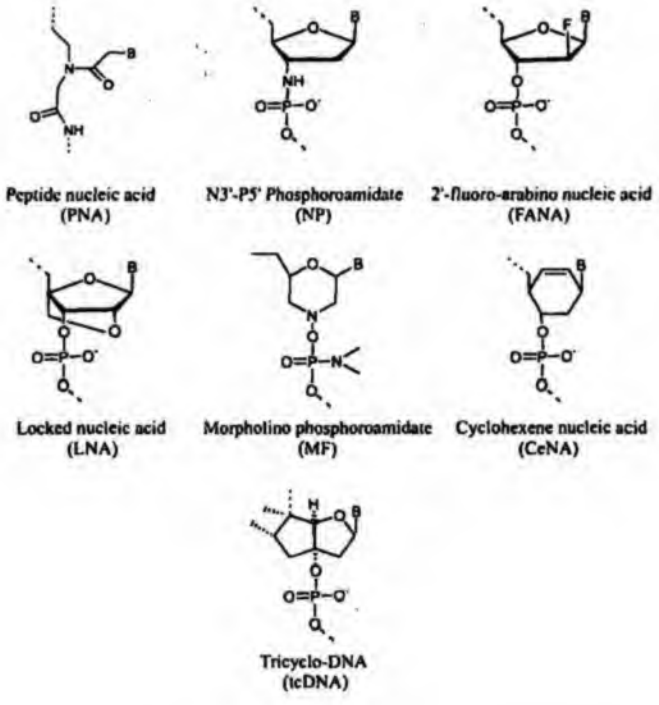


Figure 1.28: The third generation ASOs [35]

For example, boranophosphates support the RNase H cleavage of polyribonucleotides [46]. In [47] it was shown that antisense molecules with a weakened hybridization potential enhance the rate of hydrolysis. Similarly, the ability of modified antisense oligonucletides containing acyclic inter-residue units to support RNase H–promoted cleavage of complementary RNA was described. The butyl-modified 2'F-ANA ASOs described in [48] constitute the first examples of a nucleic acid species capable of eliciting high RNase H activity while possessing a highly flexible molecular architecture at predetermined sites along the ASO.

Other backbone modifications examined for the RNase H potency failed to recruit the enzyme to cleave the target RNA. Hybrid duplexes formed by methylphosphonates, posphoro-N-morpholidates, phosphoro-N-butylamidates, methylenemethylimines and N3'-P5' phosphoramidates were found to be resistant to the RNase H [49].

The recent advances in this field seem to be very promising, but still many experiments and clinical trials have to be done.

1.4 Aims and objectives

The propulsion of our work were experimentally proved differences in the RNase H activity induced by several types of modified oligonucleotides. The oligonucleotides were established in the Institute of organic chemistry and biochemistry of the AS CR: several types of isopolar modified oligothymidates and oligoadenylates (see Figure 1.29a) [50]. The modifications consisted in an insertion of an extra methylene group between one of the two oxygen atoms and the phosphorus atom in the phosphodiester linkage.

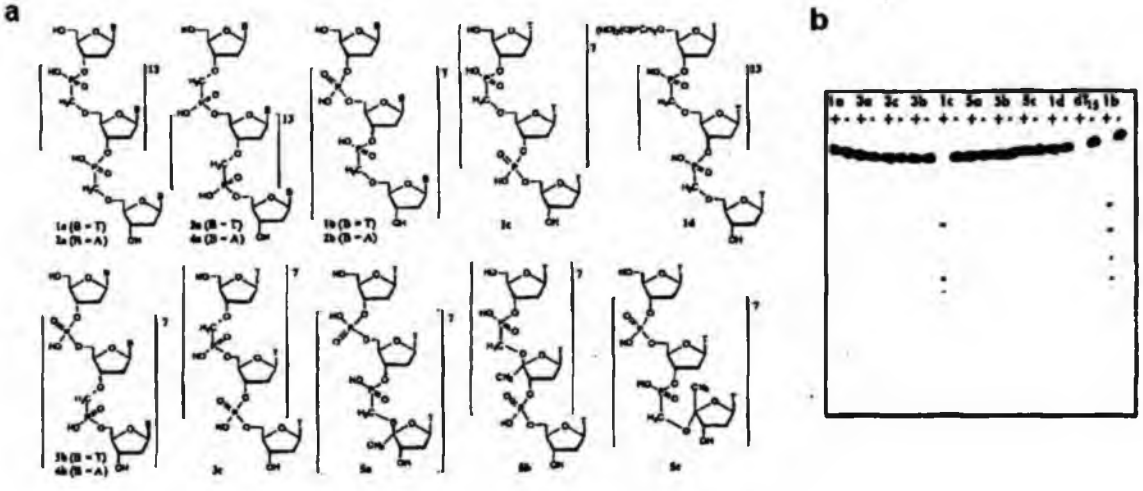


Figure 1.29: UOCHB modifications [50]

All the prepared modified thymine-containing oligonucleotides were examined for their ability to elicit E. coli RNase H activity. It was observed [50] that the combination of one sole type of the phosphate linkage (see oligonucleotides 1b and 1c in Figure 1.29b) was able to elicit the RNase H activity as much as a natural oligonucleotide. Considering the exceptional nuclease resistance [50] of the studied modifications, their protective effect on the neighboring phosphodiester linkages, hybridization properties of the modified oligonucleotides, the oligonucleotide constructs containing isopolar nonisosteric C3'–O-P-CH2-O-C5' linkage and a phosphodiester one were proposed for the use in antisense technology.

The 3D structure of the human RNase H was not yet published. Herewith, the understanding of the interaction of the human RNase H with novel modified antisense oligonucleotides would be very useful for improving the AS strategy. The aim of this work was to build up a model of human RNase H with a natural and two modified substrates and to study the enzyme-substrate interaction by means of MD. We studied the difference between the ability of the E. coli (examined in the past [23]) and human RNase H to bind the C3'-O-C-P-O-C5' and C3'-O-P-C-O-C5' (see Figure 1.30) internucleotide linkage modification.

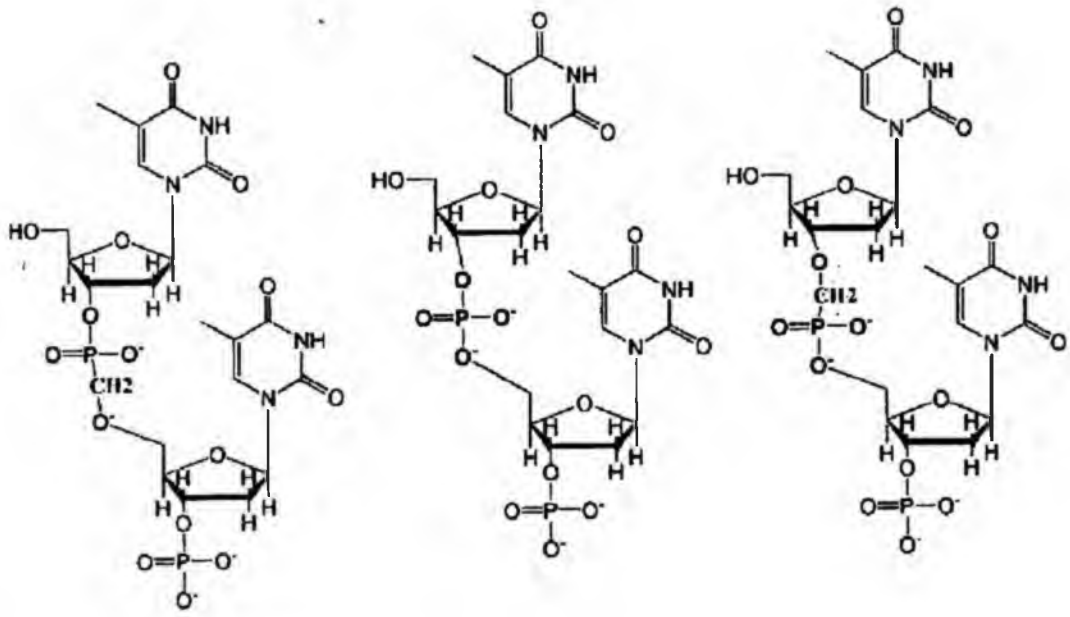


Figure 1.30: Modifications studied in this work and a natural dinucleotide between them

Chapter 2

Methods

2.1 Bioinformatics and Molecular Dynamics Simulations

2.1.1 Comparative model building

Nowadays, a large amount of data of the sequences of several thousand proteins is available due to the improvement of the genome sequencing knowledge. The number of novel sequenced cDNAs, however, increases much quicker than the number of solved 3D-structures. The three dimensional structure of a protein is determining for the protein function, knowledge of which is of the highest importance when planning or explaining experiments. Unfortunately, for the majority of protein sequences, no structural information is available yet, which can be caused by difficulties in obtaining sufficient protein or in crystallization.

The method of "knowledge based modeling" makes use of known 3D structures to predict the structure of a homologous protein or domain. Highly similar sequences that occur in proteins from different sources and sometimes even with diverse biological function generally are a sign of distinct structure similarity. We can expect a 1Å RMSD of the α-carbon coordinates for proteins with 50% residue identity.

Comparative modeling requires naturally at least one sequence of a known 3D-structure (template) with a high similarity to the target sequence. The template sequence is determined on the basis of alignment similarity. For this purposes, programs as FastA, BLAST or Modeller are used, so that the sequence alignments are computed among the target sequence and sequences of all the known structures, and those with the highest similarity are proposed for further use. The choice can be then restricted to e.g., at least 30% residue identity. The target sequence is then aligned with the chosen template or several templates.

The Needleman-Wunsch (and similarly the Smith-Waterman) sequence alignment algorithm [51] uses a dynamic programing algorithm to find an optimal local (global) alignment of two sequences, a and b. It is a recursive algorithm modified to store intermediate results. The algorithm is based on finding elements of a matrix H, where each element $H_{i,j}$ is the

optimal score for aligning sequence $a_1..a_i$ and $b_1..b_j$. Two similar amino acids receive a high score, two dissimilar a low score. The higher is the score of a path through the matrix, the better is the alignment. Each element of the H matrix is given by:

$$H_{i,j} = \max \begin{cases} H_{i-1,j-1} + S_{i,j} \\ H_{i-1,j} - d \\ H_{i,j-1} - d \end{cases} \tag{2.1}$$

The elements are computed from the $H_{1,1}$ element and then in the directions of increasing i and j. $S_{i,j}$ is the similarity score of comparison of the amino acids (the score table is given empirically e.g., BLOSUM40 in Figure 2.1 [51]) and d is the penalty for a single gap. When the last element of the matrix is achieved, the way back through the matrix H is found, which corresponds to the optimal sequence alignment. The obtained alignment is dependent on the chosen parameters, the similarity table and the value of the gap penalty, which can be changed if necessary to a get a more reliable alignment in a specific case.

```
A  5
R -2  9
N -1  0  8
D -1 -1  2  9
C -2 -3 -2 -2 16
Q  0  2  1 -1 -4  8
E -1 -1 -1  2 -2  2  7
G  1 -3  0 -2 -3 -2 -3  8
H -2  0  1  0 -4  0  0 -2 13
I -1 -3 -2 -4 -4 -3 -4 -4 -3  6
L -2 -2 -3 -3 -2 -2 -2 -4 -2  2  6
K -1  3  0  0 -3  1  1 -2 -1 -3 -2  6
M -1 -1 -2 -3 -3 -1 -2 -2  1  1  3 -1  7
F -3 -2 -3 -4 -2 -4 -3 -3 -2  1  2 -3  0  9
P -2 -3 -2 -2 -5 -2  0 -1 -2 -2 -4 -1 -2 -4 11
S  1 -1  1  0 -1  1  0  0 -1 -2 -3  0 -2 -2 -1  5
T  0 -2  0 -1 -1 -1 -1 -2 -2 -1 -1  0 -1 -1  0  2  6
W -3 -2 -4 -5 -6 -1 -2 -2 -5 -3 -1 -2 -2  1 -4 -5 -4 19
Y -2 -1 -2 -3 -4 -1 -2 -3  2  0  0 -1  1  4 -3 -2 -1  3  9
V  0 -2 -3 -3 -2 -3 -3 -4 -4  4  2 -2  1  0 -3 -1  1 -3 -1  5
B -1 -1  4  6 -2  0  1 -1  0 -3 -3  0 -3 -3 -2  0  0 -4 -3 -3  5
Z -1  0  0  1 -3  4  5 -2  0 -4 -2  1 -2 -4 -1  0 -1 -2 -2 -3  2  5
X  0 -1 -1 -1 -2 -1 -1 -1 -1 -1 -1 -1  0 -1 -2  0  0 -2 -1 -1 -1 -1 -1
   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  B  Z  X
```

Figure 2.1: An example of empirical similarity tables (BLOSUM40 table) [51]

Various parameters are evaluated to check the reliability of the resulting model. One of them, the DOPE potential used by the program Modeller, assesses the protein folding of the model. A potential value is assigned to every residue in dependence on its spatial vicinity with other residues. The value is based on empirical observations, and it quantifies the probability that a given residue would be found in the surrounding of the others in a protein in nature.

2.1.2 The method of molecular dynamics simulations

Molecular dynamics is used to describe the time evolution of a system of particles according to the laws of classical mechanics. A molecule is represented by a system of rigid balls connected by mechanical springs. For a continuous potential, the i-th particle is subjected to a force given by the following expression:

$$\mathbf{f}_i = -\frac{\partial U(\mathbf{r}_1, \mathbf{r}_2..\mathbf{r}_N)}{\partial \mathbf{r}_i}, \qquad i = 1, \ldots, N, \tag{2.2}$$

where N is the number of particles. The time evolution of the system is found by solving the equations of motion:

$$\ddot{\mathbf{r}}_i = \frac{\mathbf{f}_i}{m_i} \qquad i = 1, \ldots, N. \tag{2.3}$$

This is a system of $3N$ ordinary differential equations of the second order. A set of initial positions and velocities has to be given. As a result, trajectories of all particles within a given time interval are obtained. When resolving the equations numerically, the continuous argument (time t) is recorded in discrete points

$$t = t_0 + jh. \tag{2.4}$$

The time-step h has to be small enough to ensure the desired accuracy. The algorithm used in molecular dynamics calculations must fulfill the condition of time efficiency. Not only that there are many equations, but also the force, which is a function of the positions of all particles, has to be solved in each step.

The basic algorithm used for this purposes is the Verlet Algorithm:

$$\ddot{\mathbf{r}}_i(t) = \frac{\mathbf{r}_i(t-h) - 2\mathbf{r}_i(t) + \mathbf{r}_i(t+h)}{h^2} + O(h^2) \tag{2.5}$$

The expression is obtained from the Taylor expansion. Substituting (2.5) into (2.3) and neglecting $O(h^2)$ terms we get:

$$\mathbf{r}_i(t+h) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-h) + h^2\frac{\mathbf{f}_i(t)}{m_i} \tag{2.6}$$

This equation is a recurrent expression, and allows us to find new positions in time $(t+h)$, when the positions in times t and $(t-h)$ and forces in time t are known. Another group

of algorithms that may be used for molecular dynamics are the so called Gear integrator algorithms. This type of algorithm divides the integrational step in two parts, so that it is a kind of the predictor-corrector algorithm. There is a lot of other algorithms, such as the Leap Frog, commonly used to resolve the equations of motion in MD; some of them are derived from the two described above.

The equation (2.2) shows that a potential has to be known to evaluate the forces. The development of accurate potentials is an important object of research. The most commonly used interaction model is the Lennard-Jones pair potential:

$$\phi_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.7}$$

where ϵ is the depth of the potential well and σ is the distance at which the potential is 0. This potential (Figure 2.2 [52]) is attractive at large r, its minimum is at 1.122σ and at a shorter distance, it is repulsive. It reaches the value of 0 at $r = \sigma$ and then increases steeply.

The first term $\sim 1/r^{12}$, prevailing at short distances, models the repulsion between atoms when they are brought very close to each other. The attractive term $\sim 1/r^6$ dominates at long distances and ensures the cohesion of the particles. A $1/r^6$ attraction originates in Van der Waals forces, related to the dipole-dipole interactions of fluctuating dipoles. These are rather weak interactions, but they prevail in the bonding character of e.g. rare gases such as Ar or Kr. Therefore, these are the materials for which a LJ potential is the most appropriate to use.

The LJ potential has an infinite range. In practical applications, usually a cutoff radius R_c is given and the interactions between atoms separated by more than R_c are excluded from the computation. This results in simpler programs and tremendous savings of computer resources, because the number of atomic pairs separated by a distance r grows as r^2.

Potentials used for the molecular dynamics simulation have to take into account also the covalent bonds in the molecule. A more complicated potential is therefore required for proteins and nucleic acids. The basic force field of the following form used by Sander (molecular dynamics module of the AMBER software package) [53] is more convenient to preserve the nature of the molecules in condensed phases:

$$\begin{aligned}
U(\mathbf{R}) = {} & \textstyle\sum_{bonds} K_r(r - r_{eq})^2 && \textit{bond} \\
& + \textstyle\sum_{angles} K_\theta(\theta - \theta_{eq})^2 && \textit{angle} \\
& + \textstyle\sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) && \textit{dihedral} \\
& + \textstyle\sum_{i<j}^{atoms} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} && \textit{Van der Waals} \\
& + \textstyle\sum_{i<j}^{atoms} \frac{q_i q_j}{\epsilon R_{ij}} && \textit{electrostatic}
\end{aligned} \tag{2.8}$$

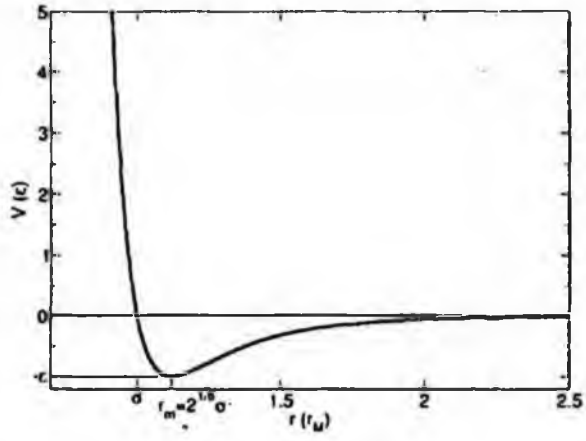


Figure 2.2: The course of Lennard-Jones potential [52]

The first three members of the above equation are the so called bonded interaction members. Via these members the covalent bonds in the molecule are considered. The mathematics of spring deformations is used here, and describes the stretching, bending and twisting of bonds. The two first members are based on the Hook's law. The first term represents the energy related to the geometry of covalently bonded atoms, the second term the energy related to the geometry of electron orbitals involved in covalent bonds, and the third term represents the energy of bond twisting. The Van der Vaals and electrostatic (given by Coulomb's law) interactions are so called non-bonded interactions. The Van der Vaals potential considers the dipole-dipole interaction between atoms, it has the form of the Lennard-Jones potential function.

The time required for a computer simulation depends on the integration time-step h. This is the time interval after which the potential is recalculated in every iteration. The time-step must be chosen as a compromise between the computational time and discretization errors. Typical time-steps for classical MD are chosen of the order 1 fs. There are algorithms as SHAKE and LINCS, which fix the fastest atoms (e.g. hydrogens) into place and therefore the time-step can be increased.

The simulation would not be realistic, if the system would be surrounded by a surface. Various artificial surface effects would take place and the particles near the surface would have non-homogeneous surroundings. A solution to this problem is to use **periodic boundary conditions** (PBC). When using PBC, particles are enclosed in a box, and we can imagine that this box is replicated to infinity by a translation in all the three cartesian directions, so that they completely fill the space (see Figure 2.3 [54]).

An important concept enabling the usage of this method is the so called "minimum image convention" [55]. According to it, for every particle from the central box just the nearest non-equivalent particles are taken into account. In this way for N particles in the box N(N-1)/2 interactions have to be computed.

In a MD simulation the **solvent** often plays an important role. There are two pos-

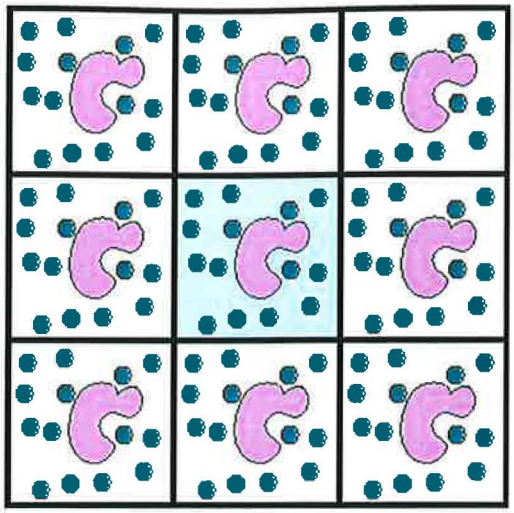


Figure 2.3: The concept of periodic boundaries [54]

sibilities how to handle it: by an explicit solvent or an implicit solvent. Explicit solvent particles (such as water types TIP3P and SPC/E) must be included in the calculation of the force field, while implicit solvents use a mean-field approach. The impact of explicit solvents on CPU-time can be very big, on the other hand the granularity and viscosity of explicit solvent is essential to reproduce certain properties of the solute molecules.

To simplify the MD calculation, the interaction between atoms whose distance exceeds a defined "**cutoff**" distance is neglected. We can easily omit short range interactions, which decline more slowly than $1/r^3$. This is the case of the Lennard-Jones potential that is used to describe the Van der Vaals force. On the other hand, long range interactions (such as e.g. coulombic ones, decreasing as $1/r$) cannot usually be neglected.

Coulombic interactions can be approximately calculated through the use of **Ewald summation method**. In this method, due to the use of the Fourier sum, the method implicitly assumes that the system under study is infinitely periodic (this is automatically fulfilled for the PBC). One repeating unit of this hypothetical periodic system is called a unit cell. One such cell is chosen as the "basic cell" for reference and it is surrounded by an infinite number of identical copies of itself. Thus, the long-range interaction energy is the sum of interaction energies between the charges of a central unit cell and all the charges of the lattice.

The basic cell contains $N/2$ each of positive and negative charges in some spatial arrangement. The entire system is neutral and contains an infinite number of charges at points $\mathbf{r}_{j+}$ and $\mathbf{r}_{j-}$, respectively. The total potential at the position $\mathbf{r}_i$ of a particle i is given by the difference of two infinite, diverging series:

$$\phi(\mathbf{r_i}) = q \sum_{j+=1}^{\infty} frac1|\mathbf{r}_i - \mathbf{r}_{j+}| - q \sum_{j-=1}^{\infty} frac1|\mathbf{r}_i - \mathbf{r}_{j-}| \tag{2.9}$$

Instead of evaluating the potential we first rewrite these charges as delta-like charge densities.

$$\rho(\mathbf{r_i}) = q \sum_{j+=1}^{\infty} \delta(\mathbf{r}_i - \mathbf{r}_{j+}) - q \sum_{j-=1}^{\infty} \delta(\mathbf{r}_i - \mathbf{r}_{j-}) \tag{2.10}$$

The delta-functions represent a lattice of point charges. Now we apply the Fourier transformation on the charge density. Fourier representation of a delta-function, however, requires infinitely many terms, so that the convergence problem is still not solved.

Therefore we use the following trick: To the given lattice we add Gaussian charge "clouds" of opposite sign:

$$\rho'(\mathbf{r}) = -q_j \left(\frac{\eta^2}{\pi}\right)^{3/2} e^{-\eta^2(\mathbf{r}-\mathbf{r}_j)^2}, \tag{2.11}$$

where η determines the width of the Gaussian distribution. At longer distances the Gaussian distribution is similar to the delta-functions, which it compensates. In this way the potential of this new lattice converges in **r**-space. The additional Gaussian charges are then neutralized by a second introduced lattice. This is a sum of Gaussian distributions, which converges in the **k**-space. Optimal convergence of both series is obtained by a suitable adjustment of the parameter η: a large η means a narrow Gaussian, which leads to a faster convergence in the **r**-space and to a slower one in the **k**-space.

2.1.3 Thermodynamical properties of the simulated system

The Newtonian equations of motion conserve the total energy of the system, which is characteristic for the microcanonical ensemble (NVE). This approach is often used in MD, but it is not a very realistic one. Usually also the temperature and pressure is conserved in the studied systems. The constant temperature in canonical ensemble (NVT) is maintained thanks to a thermal bath, with which the system may exchange temperature. One of the ways, how to modify the Newtonian equations in order to fulfill this condition, is to rescale the velocities in every step by multiplying them by a factor $\lambda = \sqrt{T/\tau}$, where T is the desired temperature and τ is the actual one. This method is unfortunately very rough and the trajectories differ notably from the Newtonian ones. Another approach is to simulate collisions with particles from an external heat bath with constant temperature. The scaling factor is given by the formula

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right), \tag{2.12}$$

where δt is the time step and τ is a free parameter, which represents the frequency of collisions.

The so called Langevin dynamics is another method of temperature control. The velocities are not modified, but friction and random forces are introduced. The interaction with

the heat bath is then involved in the equations of motion as another degree of freedom. These average interactions have a friction term, which lowers the kinetic energy, and a random component, which adds energy to the system:

$$m_i \frac{(d\mathbf{r}_i)^2}{dt} = U(x) + \Gamma_i(t) - \xi m_i \frac{(d\mathbf{r}_i)}{dt}, \tag{2.13}$$

where $U(x)$ is the potential energy (force field), ξ is the friction coefficient, and Γ is a random force derived form the Gaussian distribution, given by following:

$$\begin{aligned} &< \Gamma_i(t) >= 0 \\ &< \Gamma_i(t)\Gamma_i(t') >= 2k_B\xi T_0\delta(t-t') \end{aligned} \tag{2.14}$$

Similarly to the temperature, also the pressure can be controlled. In this case the volume of the system is not constant and therefore the description of the system is more complicated. By combining these two methods, it is possible to construct an isothermal-isobaric system (NTP), which is the most usual one in chemistry.

2.1.4 Running a MD simulation

Molecular dynamics is performed in several steps. First, the initial input files have to be prepared. **Input information** contains the coordinates of the simulated system and the molecular topology including force parameters. The atom coordinates of the 3D structure may be obtained from the PDB database or e.g. from a computer model. The topology file is then created using special programs from MD software packages as AMBER or CHARMM.

The next step is the **energy minimization** that reduces the potential energy of the system. This step avoids possible steric interferences caused by the random initial coordinates. Algorithms used are the steepest descent or conjugate gradient methods.

The temperature of the system is then linearly increased from the initial temperature of 0K to the desired value in the **heating** step. The initial distribution of velocities is given by the Maxwell-Boltzmann distribution.

Before performing the final molecular dynamics, the **equilibration** process takes place. It serves to establish a thermodynamic equilibrium in the studied system. A temperature control method is used.

The proper **MD simulation** is run to produce a MD trajectory to obtain data about the structural characteristics and dynamic development of the studied system.

These data serve for the **analysis** of the MD, where various physical quantities are calculated and the structural changes of the system in time are studied.

2.2 Software

2.2.1 Modeller

MODELLER is a computer program used in producing homology models of protein tertiary structures. It implements a technique inspired by nuclear magnetic resonance known as satisfaction of spatial restraints, by which a set of geometrical criteria are used to create a probability density function for the location of each atom in the protein. The method relies on an input sequence alignment between the target amino acid sequence to be modeled and a template protein, whose structure has been solved. The additional tools are de novo modeling of loops in protein structures, searching of sequence databases, comparison of protein structures, etc.

2.2.2 AMBER

Several force-field softwares have been developed for the purposes of molecular dynamics and the evaluation of the obtained data, e.g. AMBER, CHARMM, OPLS, GROMOS, etc. They differ slightly in the equations used and are suitable for various kinds of simulated systems. In the following passage, AMBER will be discussed in detail because it is the software used in this work.

AMBER is a group of force-fields for molecular dynamics of biomolecules. It also includes the molecular dynamics simulation package which uses this force fields. The programs are written in Fortran 90 and C and support Unix-like systems.

Information needed to run the dynamics is: the initial positions of all atoms acquired from X-ray crystallography, NMR or computer modeling in PDB format; topology (names of atoms, residues and bonds, charges) obtained also from PDB; force-field with several parameters (bond strength, torsion angles) from the AMBER databases; commands entered by the user.

The program **LEaP** serves for the preparation of the system for the following simulation. This program uses as an input the PDB file, and as a result, two files are generated: one of them contains the coordinates of the particles, the other one all the remaining topology information. The latter doesn't change during the simulation.

The basic program for running the simulation is called **sander**. First, all the atoms are iteratively relaxed according to the gradient of energy until the stationary thermodynamical state is reached. After the minimization step, the MD simulation follows.

For the analysis of the obtained trajectory and coordinates, the program **PTRAJ** is used. There are various quantities to be measured, as the bond length, hydrogen bonds, RMSD (mass weighted root mean square deviation to a reference structure), sugar puckering and torsion angle, etc.

```
amber on                    #AMBER force field
parmfile mod1.parm          #input parameter file
coordinates gil.coor        #input coordinates file
velocities gil.vel          #initial set of velocities
extendedSystem gil.xsc      #periodic cell parameters from the previous sim.
exclude scaled1-4           #which atoms are excluded from non-bonded..
                             #..interactions
1-4scaling 0.833333
binaryoutput no
outputname gil_rat1
outputEnergies   10
outputTiming     1000       #timesteps between timing output
xstFreq          5000       # how often to append..
dcdFreq          5000       # ..the syst. configuration to the XST/DCD file
dcdUnitCell      on

binaryrestart no
restartname gil_rat1_restart
restartfreq 5000

rigidBonds all              #usage of ShakeH: H-atom bond lengths
rigidTolerance 0.0005       #allowable bond-length error
rigidIterations 100         #max num. of iterations to constrain the bond-lengths
rigidDieOnError off
useSettle on

timestep         1
nonBondedFreq    2          #how often calc. short-range non-bonded interactions
fullElectFrequency 4        #number of timesteps between full elstat evaluations
stepsPerCycle    20         #how often calc. the elstat forces beyond the cutoff
switching        on         #use a smooth switching function to reach 0 (cutoff)
switchDist       8.5
cutoff           10
pairlistdist     11.5
wrapAll on                  #coordinates around periodic box wrapped
wrapWater on

PME on                      #PME algorithm (full elstat interactions)
PMEGridSizeX 96
PMEGridSizeY 96
PMEGridSizeZ 96
margin 2

langevin                    on    #Langevin temp. controle(damping/random forces)
langevinDamping             1     #Langevin coupling coefficient
langevinTemp                310   #temperature at which atoms will be adjusted
langevinHydrogen            no

langevinPiston              on          #Langevin-Piston pressure control
langevinPistonTarget        1.01325     #target pressure(bar)-athmosph. pressure
langevinPistonPeriod        200         #oscillation period
langevinPistonDecay         500
langevinPistonTemp          310
```

```
useGroupPressure        yes     # smaller fluctuations of pressure
useFlexibleCell         no      # allow dimensions of the box to fluctuate
useConstantRatio        no      # fix shape in x-y plane
firsttimestep           20000
run                     5000000    #number of steps to run
```

Figure 2.4: NAMD configuration file for MD

2.2.3 NAMD

NAMD (NAnoscale Molecular Dynamics) is a free-of-charge molecular dynamics code written using the Charm++ parallel programming model, noted for its parallel efficiency and often used to simulate large systems (millions of atoms). When running a MD simulation in NAMD, a force field, the input coordinates file and input topology file is needed and a NAMD configuration file. The NAMD configuration file specifies options and values that NAMD should use, such as the number of time-steps, initial temperature, etc. The options in this file specify the actual behavior of NAMD and serve to control the running of the simulation. Each line of the configuration file consists of a keyword identifying the option to be specified, and a value as a parameter of the option.

In this work, NAMD was used for running the MD simulations using the AMBER force field and AMBER was used for the initial preparation of the system (program LEap) and for the analysis of the trajectories (PTRAJ). An example of the NAMD configuration file can be found at the beginning of this section (Figure 2.4).

2.2.4 VMD and CHIMERA

The visualization of the results is made possible by various programs as **VMD** and **Chimera**. VMD is primarily developed as a tool for viewing and analyzing the results of molecular dynamics simulations, it also includes tools for working with volumetric data, sequence data, and arbitrary graphics objects. VMD is freely available including the source code.

UCSF Chimera or (Chimera) is an interactive molecular graphics program developed by the University of California, San Francisco. It is the successor of UCSF Midas and MidasPlus.

Chapter 3

Results and discussion

3.1 Building up the human RNase H computer-based model

Human enzyme RNase H consist of three regions: the catalytic domain containing the active site and the DNA-binding region, the RNA-binding region and a 62 amino acids sequence of the spacer region, which connects both of them. Deletion of the spacer region results in a significant loss of the RNase H activity. Furthermore, the binding affinity of this deletion mutant for the heteroduplex substrate was 2-times tighter than the wild type enzyme suggesting that this central 62-amino acid region does not contribute to the binding affinity of the enzyme for the substrate. The catalytic region is homologous with a broad variety of RNases H from various organisms. The RNA-binding region, however, is found only in eukaryotes. It is found on the amino terminus of human RNase H and its structure is consistent with a double-strand RNA-binding motif.

3.1.1 Model of the catalytic domain

As already mentioned, the 3D structure of human RNase H is not yet available. Therefore, a model based on known structures from different organisms was built. The program Modeller was used for this purpose. Several suitable structures were found according to the sequence similarity. Finally, two structures were chosen as possible templates for the catalytic domain: E. coli (1JL1) with resolution 1.3Å and Thermus thermophillus (1RIL) with resolution 2.8Å. The reliability of the two models was evaluated by the calculation of the DOPE potential (see Figure 3.7). In the model built according to the Thermus thermophillus (TT) structure there arose a gap (Figure 3.1) in the vicinity of the DNA-binding site, which is crucial for a correct activity. In the Figure 3.2, the superimposed structures are shown with the problematic place highlighted. In addition, the sequence alignment shows that the highly conserved DDE motif (marked with red) is not fully conserved in the template structure.

Therefore, at the end, the E. coli structure based model (Figure 3.4) was chosen, due

to the higher residue identity number, better resolution of the template structure and a suitable DOPE potential result. The sequence alignment shows that all the crucial residues are conserved (see Figure 3.3). According to the DOPE potential calculation (Figure 3.7), several regions (10-30, 30-70 and 125-150) were identified as potentially less reliable and so the RMSD of these regions was calculated as a part of the analysis of all models.

```
_aln.pos          10        20        30        40        50        60
1ril      -RKRVALFTDGACLGNPGPGGWAALLRFHAHEKLLSGGEACTTNNRMELKAAIEGLKALKEPCEVDLY
huma      MGDFVVVYTDGCCSSNGRRRPRAGIGVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNIN
_consrvd      *   *** *  *       *           *  *                  **

_aln.p    70        80        90        100       110       120       130
1ril      TDSHYLKKAETE----GWIEGWRKRGWRTAEGKPVKNRDLWEALLLAMAPHRVRFHFVKGHTGHPENE
huma      KLVLYTESMFTINGITNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNE
_consrvd      *    **       *  ** * ** *  ** * *     **             * ** *   **

_aln.pos  140       150
1ril      RVDREARRQAQSQAKT
huma      EADRLAREGAK-QSED
_consrvd   ** **  *  *
```

Figure 3.1: The alignment of human and TT (1RIL) RNase H sequence

Figure 3.2: Superimposed structures of human and TT (1RIL) RNase H

```
_aln.pos          10        20        30        40        51        60
Ecoli     -/KQVEIFTDGSCLGNPGPGGYGAILRYRGREKTFSAGYT---RTTNNRMELMAAIVALEALK--EHA
human     MGDFVVVYTDGCCSSNGRRRPRAGIGVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNIN
_consrvd     *   * * *  *         *  * *       *     * ** * *  **  * *  *

_aln.p    70    75  80        90        100       110       120       130
Ecoli     EVILSTDSQYVRQGITQWIHNWKKRGWKTADKKPVKNVDLWQRLDAALGQHQIKWEWVKGHAGHPENE
human     KLVLYTDSMFTINGITNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNE
_consrvd    * ***     *** *   *** ****   * * *       *        * *  * ** *   **

_aln.pos  139       150
Ecoli     RADELARAAAMNPTLEDTGYQVE
human     EADRLAREGAKQSED--------
_consrvd  ** ***  *
```

Figure 3.3: The alignment of human and E. coli (1JL1) RNase H sequence

Figure 3.4: Superimposed structures of human and E. coli RNase H with the DDE motif highlighted

3.1.2 Model of the RNA-binding domain

A model of the RNA-binding domain (Figures 3.5 and 3.6) was also created on grounds of a Saccharomyces cerevisiae structure (1QHK). This structure, however, was not used for further work because it was impossible to create the model of the spacer-region, which is found between the RNA-binding and the catalytic domain. This part of the enzyme is fortunately not crucial for its function. It was shown [56] that the RNA-binding domain is not required for RNase H activity, as the RNA-deletion mutants exhibited catalytic rates

2-times faster than the rate observed for wild-type enzyme. Comparison of the dissociation constant of human RNase H1 and the RNA-deletion mutant for the heteroduplex substrate indicates that the deletion of this region resulted in a 5-fold loss in binding affinity. Finally, comparison of the cleavage patterns exhibited by the mutant proteins with the cleavage pattern for the wild-type enzyme indicates that the dsRNA-binding domain is responsible for the observed strong positional preference for cleavage exhibited by human RNase H1.

```
_aln.pos          10         20         30         40
1qhkA      GNFYAVRKGRETGIYNTWNECKNQVDGYGGAIYKKFNSYEQAKSFLG
humrna     GMFYAVRRGRKTGVFLTWNECRAQVDRFPAARFKKFATEDEAWAFV-
 _consrvd * ***** ** **   *****  ***    *  ***     *  *
                          (43)           (59-60)
```

Figure 3.5: The alignment of the human and Saccharomyces RNase H RNA-binding domain sequence

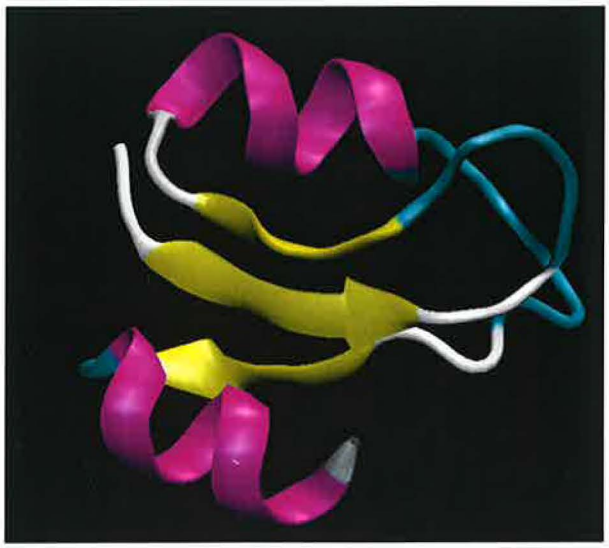


Figure 3.6: The model of the RNA-binding domain of human RNase H

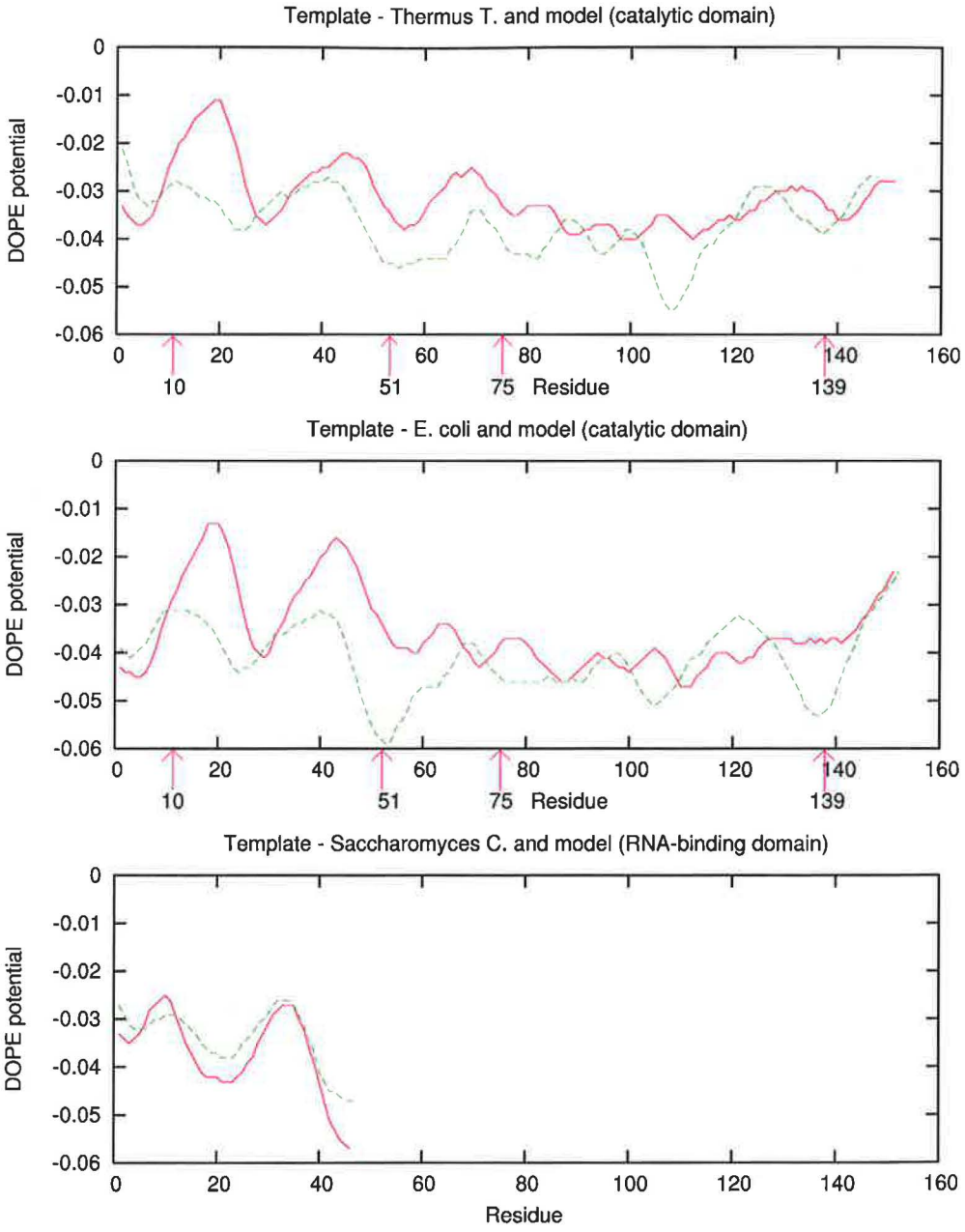


Figure 3.7: The DOPE-potential evaluation of the reliability of models (model red, template green)

3.2 Human RNase H bound to an unmodified substrate

3.2.1 Initial model

The catalytic domain of the human RNase H enzyme (Figures 3.35 and 3.9) was put together with a substrate on the base of analogy with the model derived in silico for the E. coli RNase H - RNA/DNA complex [23]. Structural alignment was made using the VMD software package. The DNA strand was positioned into a putative DNA binding pocket consisting mainly of tryptophan residues. The RNA strand interacts strongly with the active site of the enzyme. The mutual contacts in the active site (binding RNA, Mg ions and catalytic residues) were carefully established. Information from the Bacillus Halodurans RNase H X-ray structure was exploited [8].

Molecular dynamics simulations (duration 5 ns with some exceptions) were used to optimize mutual enzyme-substrate interactions (Model 1-6, see p. 54-62), the active site geometry (Model 7-9, see p. 63-67), to determine impacts of chemical modifications in the DNA backbone (Modification 1-10, see p. 68-82) and to find major differences in the recognition of a substrate by E.coli/Human RNase H (E. coli RNase H, see p. 83-85). Interestingly, the usage of fully solvated molecules was found as absolutely essential if subtle refinements of simulated systems were made. Re-solvation of biomolecules usually led to substantial disturbances of them (Model 4-5, see p. 60, 61). Several MD runs (Model 7-9, see p. 63-67) were devoted to optimization of force field parameters describing properties of magnesium ions. There is a significant deficiency of the force field description of divalent cations. There are substantial discrepancies considering the magnesium ion-ligand bond length (MD simulations 1.8Å, ab initio calculations 2.15Å, X-ray structures 1.8 - 2.6Å). Further, magnesium ions coordinate six ligands with strong directional preferences. Representations of ions as spheres bound to ligands by Van der Waals and Coulombic interactions is not fully adequate. Considering the above mentioned facts, force constants for Mg ions were optimized empirically to secure stability of the active site possessing the geometry known from the Bacillus Halodurans RNase H X-ray structure [8].

There are quantities measured and analyzed to determine the reliability of different models and potential impacts of chemical modifications made in the DNA backbone on the RNase H - RNA/DNA geometry:

A) **Root Mean Square Deviations (RMSD)** - a reasonable value of RMSD indicating a stable geometry should not exceed 3Å for the whole structure.

B) **Active site geometry** - following carbonyl residues are substantial for proper stabilization of Mg ions in the active site: Asp 139, Asp10, Asp75 and Glu51 (see Figure 3.10). These residues constitute the highly conserved DDE motif (or DDDE in this case) observed in all RNase H structures. Interestingly, the aspartic acid 10 coordinates both Mg ions. Potential disturbances of mutual contacts should serve as an indicator of undesired impacts of the chemical modifications made in the DNA backbone.

C) **Enzyme-substrate recognition** - interactions of nucleic acids with selected amino

acid residues. It was found out that there are 10 residues involved in the interaction with the substrate. DNA-binding residues are Ser98, Trp90, Thr46, Asn 105 and Gln48 (see Figure 3.12). RNA-binding residues are Met77, Asn47, Glu51, Cys12 and Ser14 (see Figure 3.11).

D) **Distances of atoms** involved in the Watson-Crick base-pairing.

E) **Conformational behavior of riboses and deoxyriboses** - sugar puckering - described using the pseudorotation angle. Common values of the pseudorotation angles are between $P = 100°$ and $P = 150°$ for the DNA strand in the heteroduplex (it is a little less than the value characteristic for the purely C2'-endo B-conformation) and between $P = 0°$ and $P = 40°$ for the RNA strand, which is a typical C3'-endo A-conformation.

F) **Conformational behavior of phosphodiester linkages** - the -g-g conformer (values from $-120°$ to $-180°$ in terms of the C3'-O3'-P-O5' and O3'-P-O5'-C5' torsion angles) is usual in the undisturbed nucleic acids backbone.

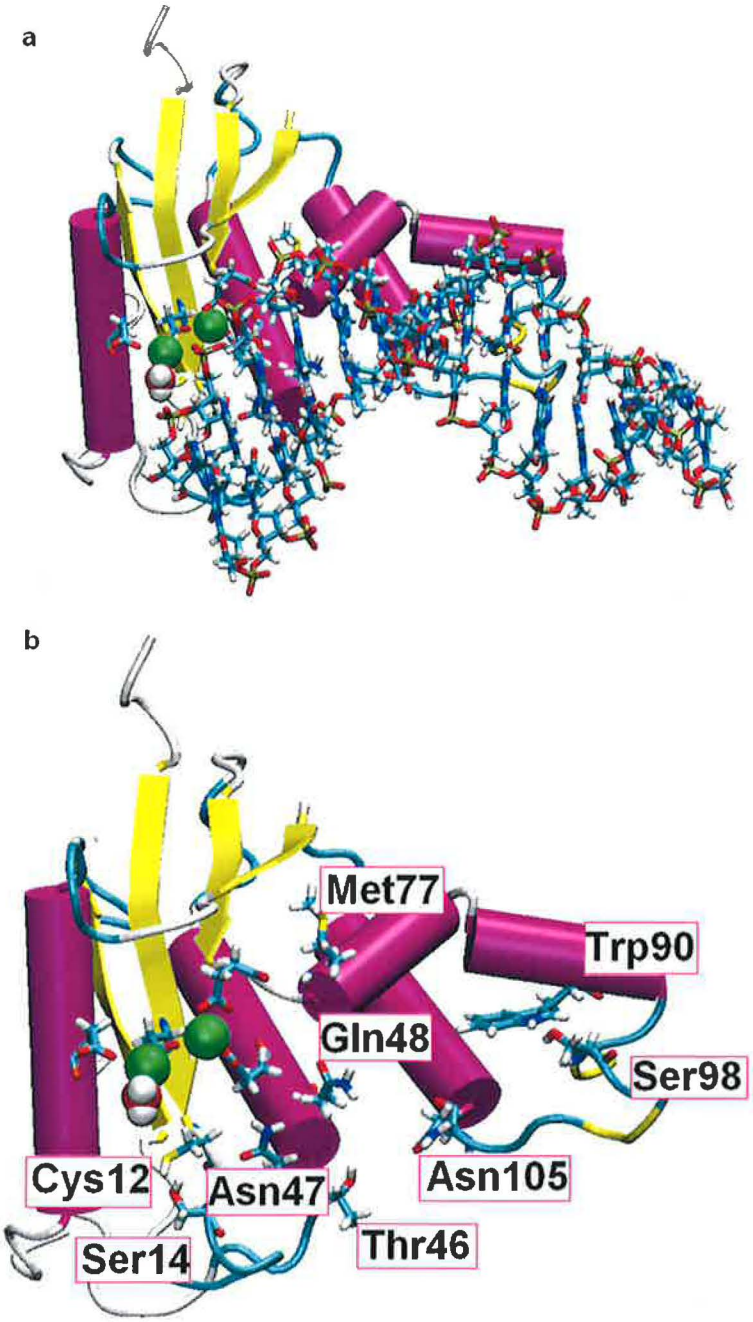


Figure 3.8: The structure of human RNase H in complex with the substrate (a) and residues interacting with substrate highlighted(b)

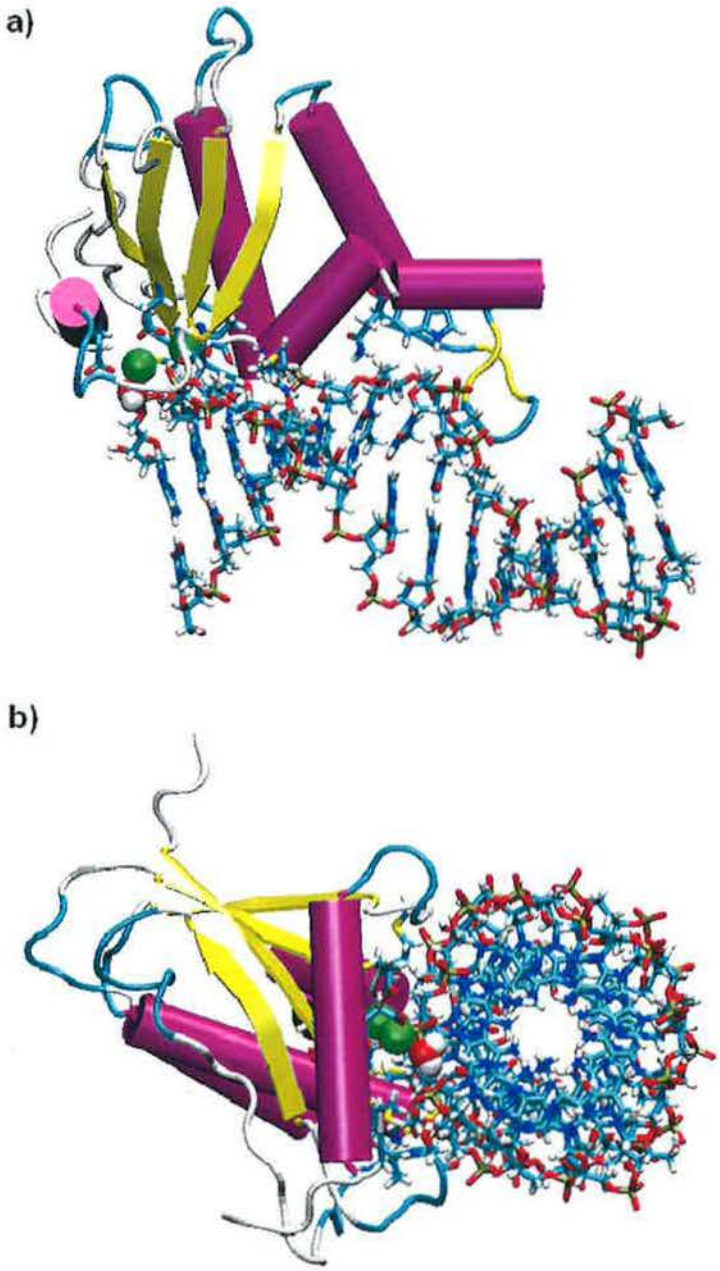


Figure 3.9: The structure of human RNase H - different views, positioning of the substrate

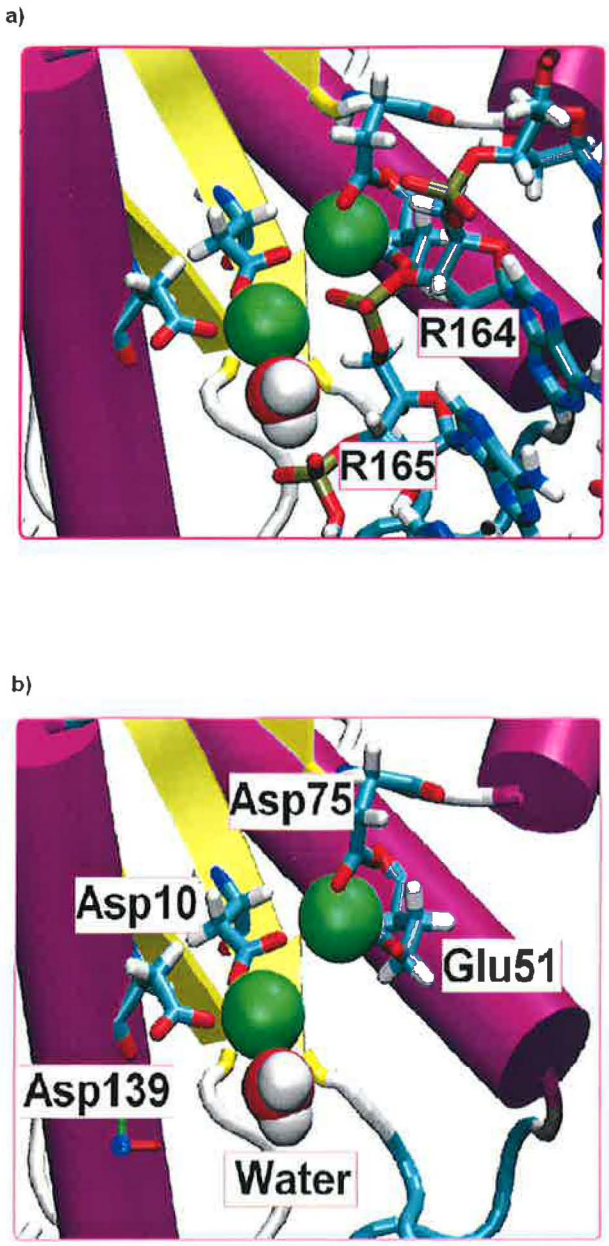


Figure 3.10: The active site with Mg^{2+} ions (green) with nucleotides highlighted (a) and interacting residues highlighted (b)

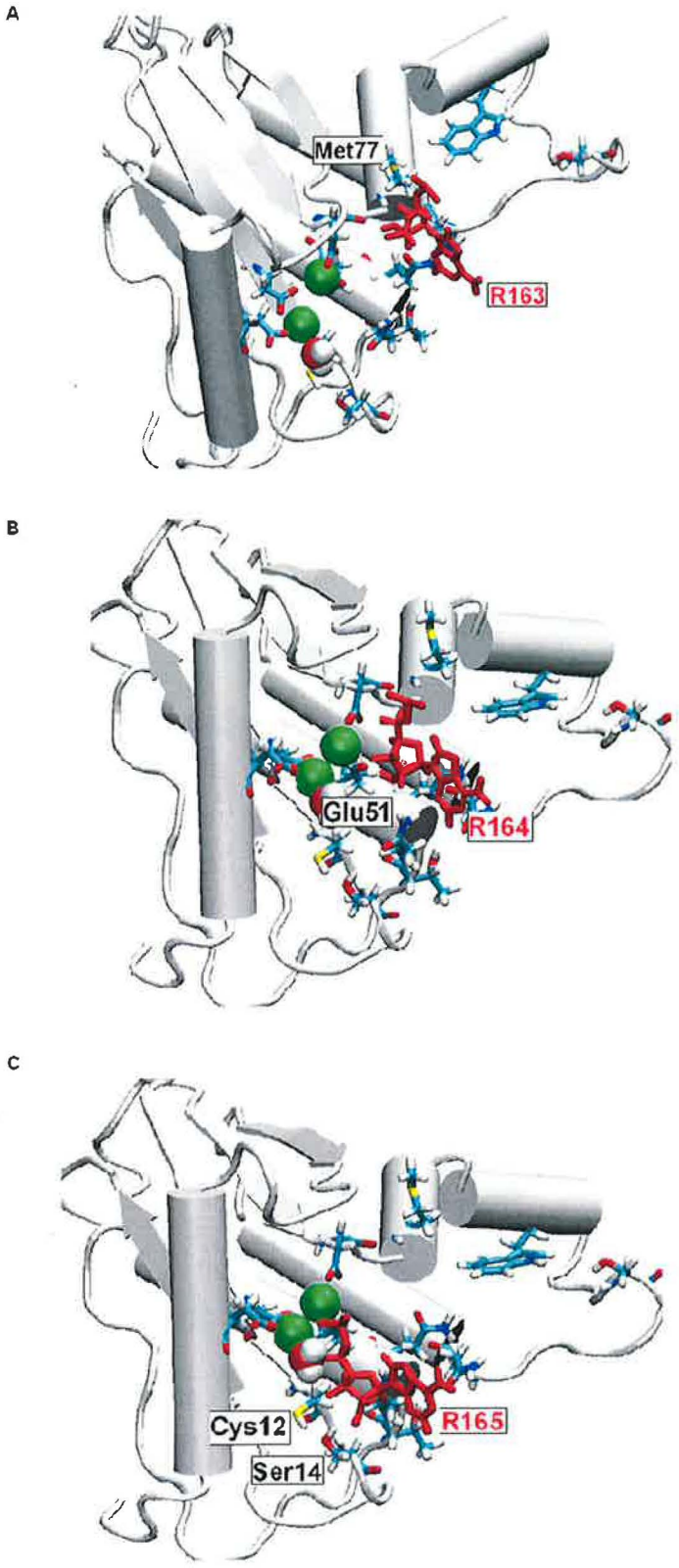


Figure 3.11: Residues interacting with the RNA-strand

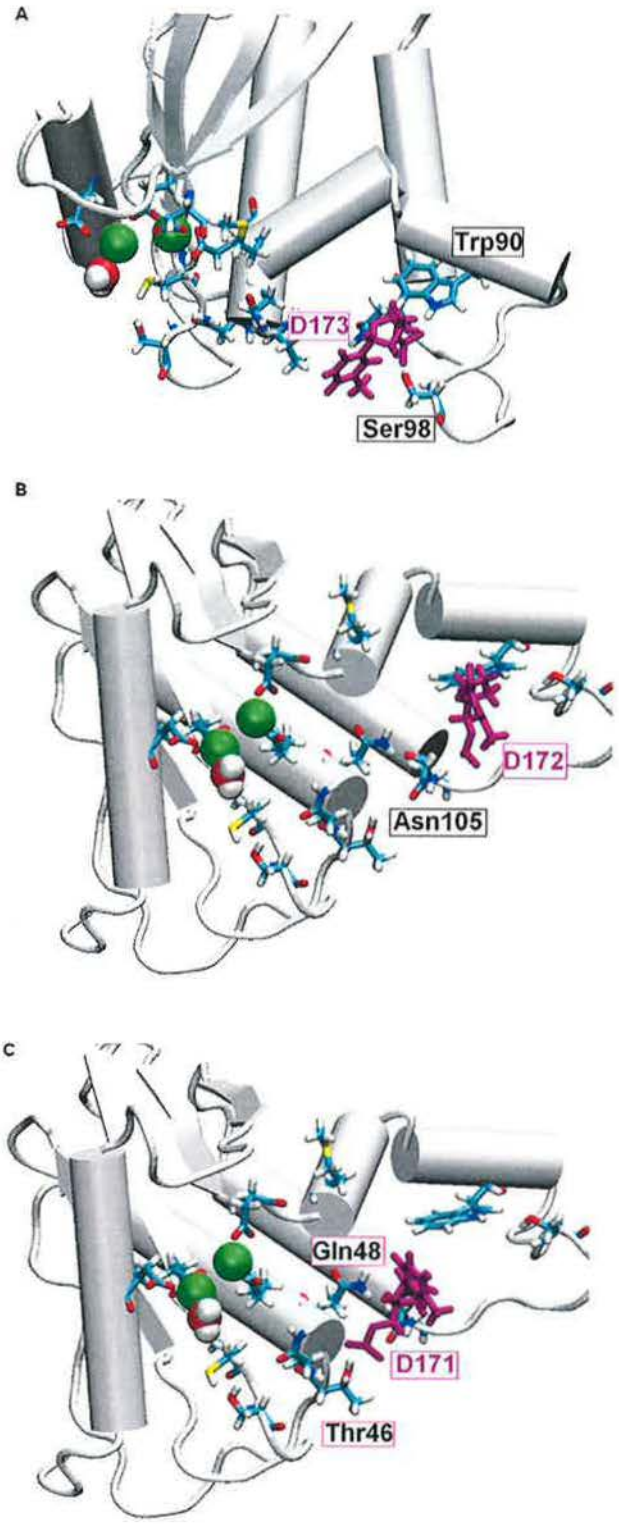


Figure 3.12: Residues interacting with the DNA-strand

3.2.2 Model refinement I: enzyme-substrate recognition

Model 1 (see Fig. 3.13 p. 57, Graphs A1-A8 Appendix A)

The Model 1 was influenced by a steric conflict between the Met77 residue and nucleic acids. Consequently, the DNA/RNA structure was substantially corrupted, and the RMSD of the nucleotide chains reached 6Å. Nevertheless, the RMSD for the whole structure did not exceed 3Å. It gives an evidence of a good overall stability of the model. The Watson-Crick hydrogen bonds connecting the R161-D171 nucleotides were affected by the Met77 obstruction. The neighboring R162-D170 and R160-D172 base pairs were cut across too. The riboses in this part of the nucleic acids structure adopted unusual conformations. Almost all of the enzyme-substrate hydrogen bonds were suspended (with the exception of contacts where the Trp90, Thr46 and Gln48 residues participated).

Model 2 (see Fig. 3.14 p. 58, Graphs A9-A16 Appendix A)

The position of the Met77 residue (see Figure 3.11) was corrected by hand using the VMD software package. The Watson-Crick hydrogen bonds were then reestablished . Nevertheless, the terminal R165-D167 base pair was separated within a MD run (but it is often observed in MD simulations that the ends of duplex structures are frayed). Further, the Watson-Crick hydrogen bond between the R163-D169 residues was turned off. This was caused by the fact that the R163 residue was involved in strong contacts in the active site, geometry of which was not fully optimized. Consequently, the whole RNA strand was distorted. This significant violation of the positioning of the RNA chain in the active site resulted in a rupture of almost all enzyme-substrate hydrogen bonds. The only one exception was the Trp90 hydrogen bond. It should be noted, that the Trp90 residue showed itself as an extremely strong anchor for the nucleic acids substrate also in other models.

The sugar puckering was unusual in the expected cases of outer nucleotides. The torsion angles in phosphate groups assumed the -g-g values. On the contrary, many of the O5'-P-O3'-C3' torsion angles in the RNA strand (situated in the nearby of the active site) took values of approximately 150° i.e. trans-conformation, which is another sign of the duplex deformation.

Model 3 (see Fig. 3.15 p. 59, Graphs A17-A24 Appendix A)

The case of the Model 2 illustrates the importance of a correct geometry of the active site. It was looked for such a geometry that would reasonably fulfill expectations based on the experimental knowledge. Two water molecules were positioned manually into the active site to avoid undesired direct contacts between nucleic acids and ions. One of the water molecules (see Figure 3.10) was found to bridge the Rp oxygen atom (from the R164-R165 linkage) and the Mg2 ion within a MD run. The water molecule has a functional importance, as it enables a creation of the hydroxide ion attacking the scissile bond.

Inner Watson-Crick hydrogen bonds were stabilized as a consequence of the rational positioning of water molecules into the active site. The hydrogen bonds established between enzyme side chains and the RNA strand were stable (except the usually weaker Ser14 binding). On the contrary, only the most potent binders remained anchored to the DNA strand. All the torsion angles but in the first phosphate group were in the -g-g conformation in both the RNA and DNA strand. The riboses of the first (R152) and the last two (R165 and R166) nucleotides were found in the B-conformation (similarly as in all previous models). In the DNA strand the values of pseudorotation angles for the D171 and D172 deoxynucleotides assumed values of cca. 50° (not belonging neither to A nor to B conformation). Namely these tightly bound nucleotides were involved in the enzyme-substrate interaction.

Model 4 (see Fig. 3.16 p. 60, Graphs A24-A32 Appendix A)

Further, subtle refinements of the DNA/RNA hybrid structure were made (namely a reduction of the RNA strand to 14-mer). Two water molecules were positioned in the active site (as in Model 3). The simulated system was surrounded by water molecules using the TLEAP module and a MD run was produced. However, it was recognized that a proper stabilization of the active site in Model 3 was a rather lucky chance.

Here, a carboxylic acid was retracted into the minor groove of the substrate. The measured values, however, attest that this probably was just a fluctuation. Especially many broken hydrogen bonds suggest that the interaction of the protein with the duplex was not sufficiently strong. The RMSD of residues involved in the interaction with the DNA strand exceeds 3.5Å.

Model 5 (see Fig. 3.17 p. 61, Graphs A32-A40 Appendix A)

Model 5 was derived from the final structure produced within the Model 3 MD run. Only two water molecules (rationally positioned into the active site) were taken and the whole structure was re-solvated. However, the simulated system was unstable again. Some of the enzyme-substrate hydrogen bonds came apart during the MD simulation, although at the beginning they seemed to be rather stable. This is the case of Asn105, Gln48 and Asn47. Big fluctuations can be observed in RMSD of both nucleic acids strands. The RMSD values surpasses 3Å.

Model 6 (see Fig. 3.18 p. 62, Graphs A41-A48 Appendix A)

The final structure produced within the Model 3 MD run was taken as initial again. However, all water molecules were included into the Model 6 now. Nucleic acids structure was refined (14 mer RNA strand). Moreover, force field constants for the Mg ion were modified to produce 2.5 Å ion-ligand distances. The model 6 was evaluated as the most reasonable one and therefore a MD simulation lasting for 10 ns was performed. The RMSD of the whole structure reaches 2.5Å. Even the regions found as the less reliable - using the DOPE potential (see Figure 3.7, p. 46) - do not exhibit RMSD values overreaching 3Å.

We can notice that distances of all the residues bound to the Mg ions are approximately 2.5Å. Both ions interact symmetrically with the oxygen atom from the R163-R164 phosphate group. The mutual distance between both ions is cca 4Å. All Watson-Crick hydrogen bonds are stable. All studied enzyme-nucleic acids hydrogen bonds arrive at the 2Å distance by the end of the simulation. Residues Ser98, Gln48 and Ser14 come close to the substrate only during the end of the simulation, whilst the rest of them are stable during the whole period of 10ns. The RNA binding Asn47 residue, on the other hand, shows a small withdrawal in the last 2ns of the MD simulation.

The pseudorotation angle values oscillate predominantly in the interval $0 - 50°$ for the RNA strand. The RNA strand adopts the usual A-conformation, with the exception of the terminal residue R152 inclining to the B-conformation. However, the terminal residues often manifest differences in the course of MD simulations (not declining the reliability of the model). The pseudorotation angles in the DNA strand take the values of $100 - 150°$, which are characteristic for the DNA-strand in the DNA/RNA heteroduplex. The 171 and 172 deoxynucleotides assume values $50 - 100°$ belonging neither to B nor to A conformation. The D172 residue was found initially even in the A-conformation with typical values of the pseudorotation angle $0 - 50°$. It should be noted, that the D171 and D172 residues are involved in the enzyme-substrate interaction, which affects their conformation. In contrast, the neighboring D170 and D173 deoxyriboses are forced to the purely C2'-endo B-conformation.

The phosphodiester linkages prefer the usual -g-g conformation. The only exceptions are the -gt conformers adopted by the D169-D170 and D170-D171 linkages situated in the DNA-binding region. The R164-R165 linkage (situated close to the active site) behaves similarly. Other disturbances were determined for the terminal R152-R153 and D167-D168 linkages. Additionally, the R158-R159 linkage was found in the -gt conformation in all modified models.

With respect to the obtained results, this model was used as a reliable reference for the further studies.

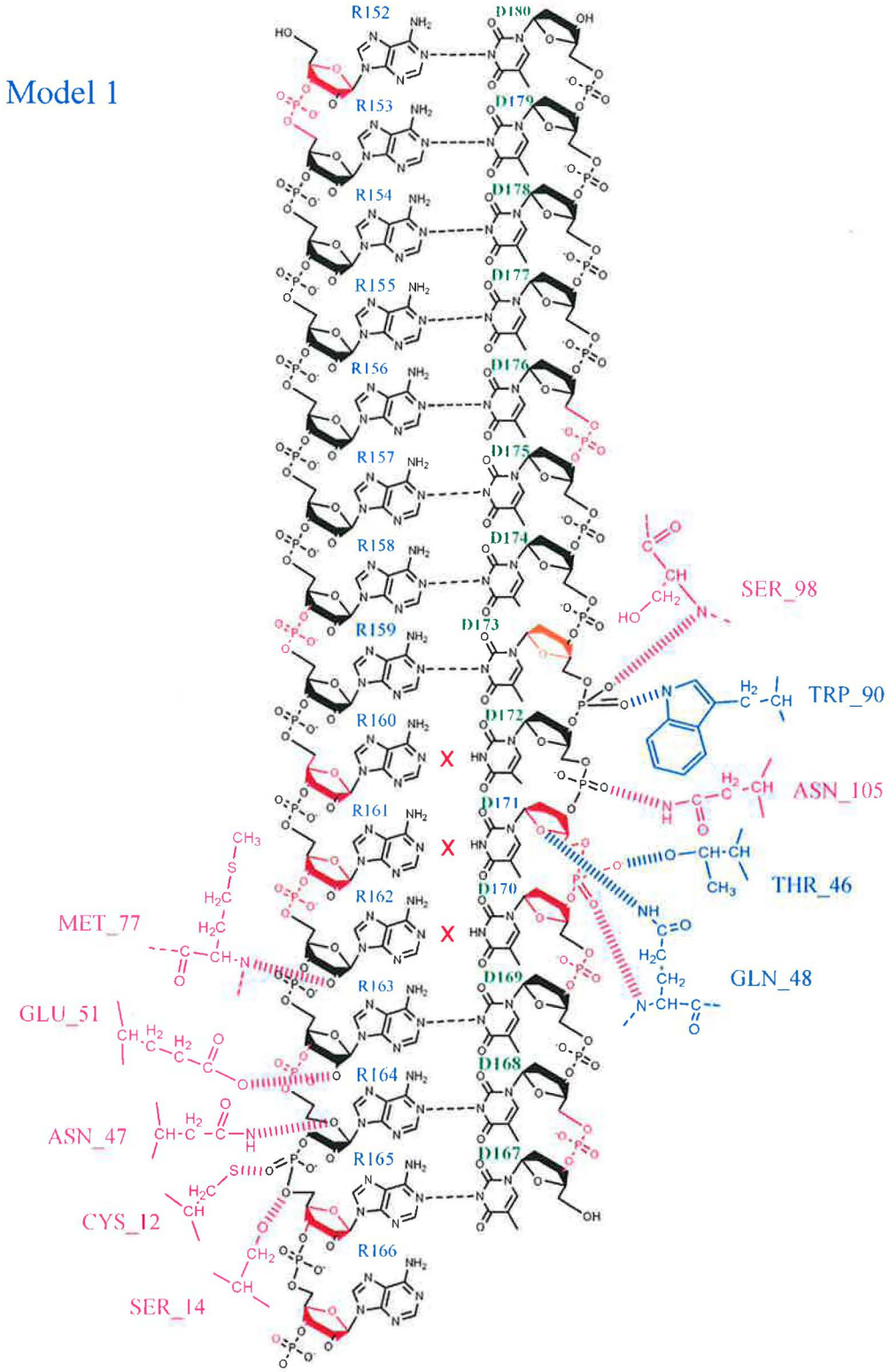


Figure 3.13: Model 1: Schema of the enzyme - substrate interaction

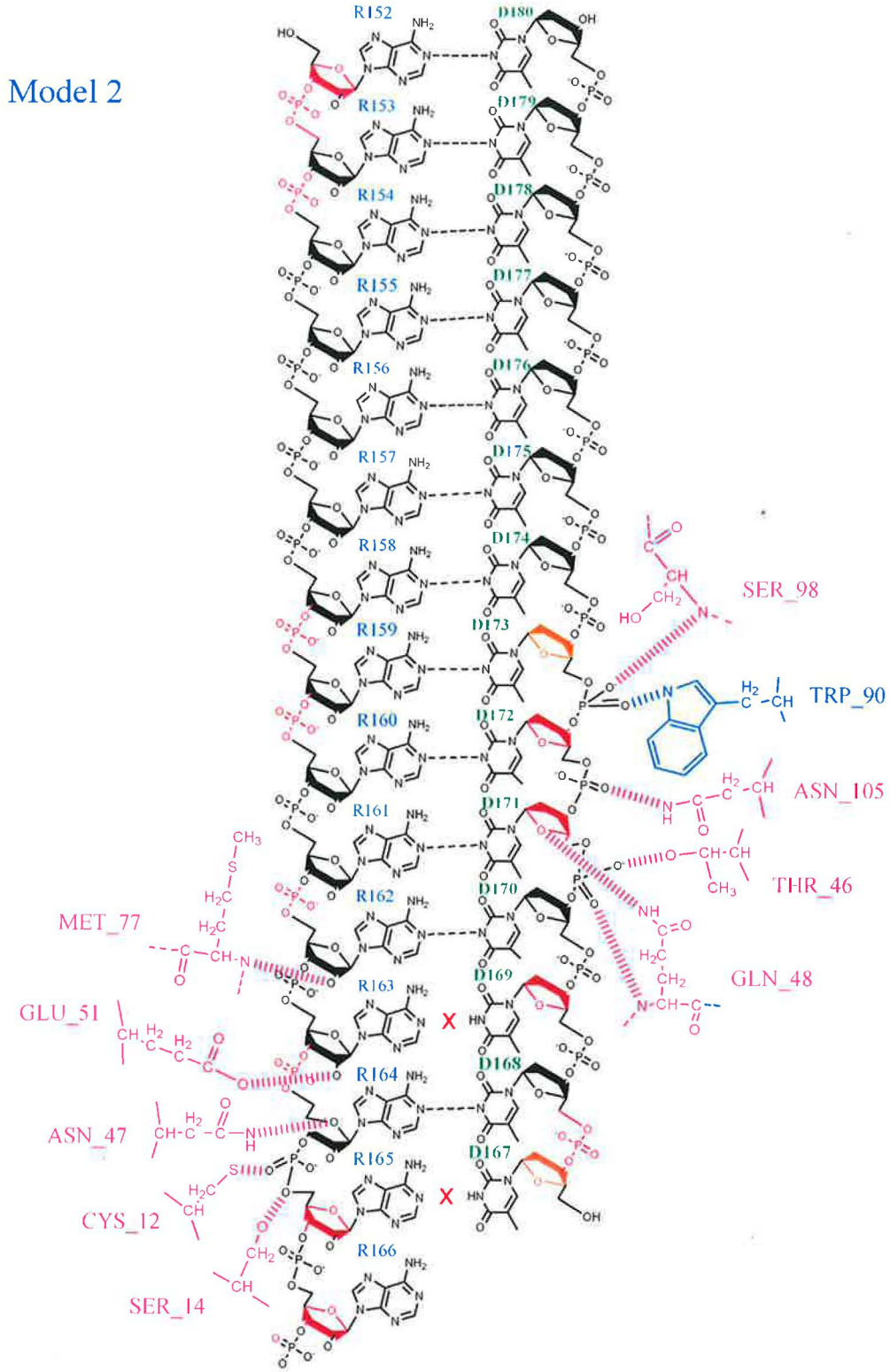


Figure 3.14: Model 2: Schema of the enzyme - substrate interaction

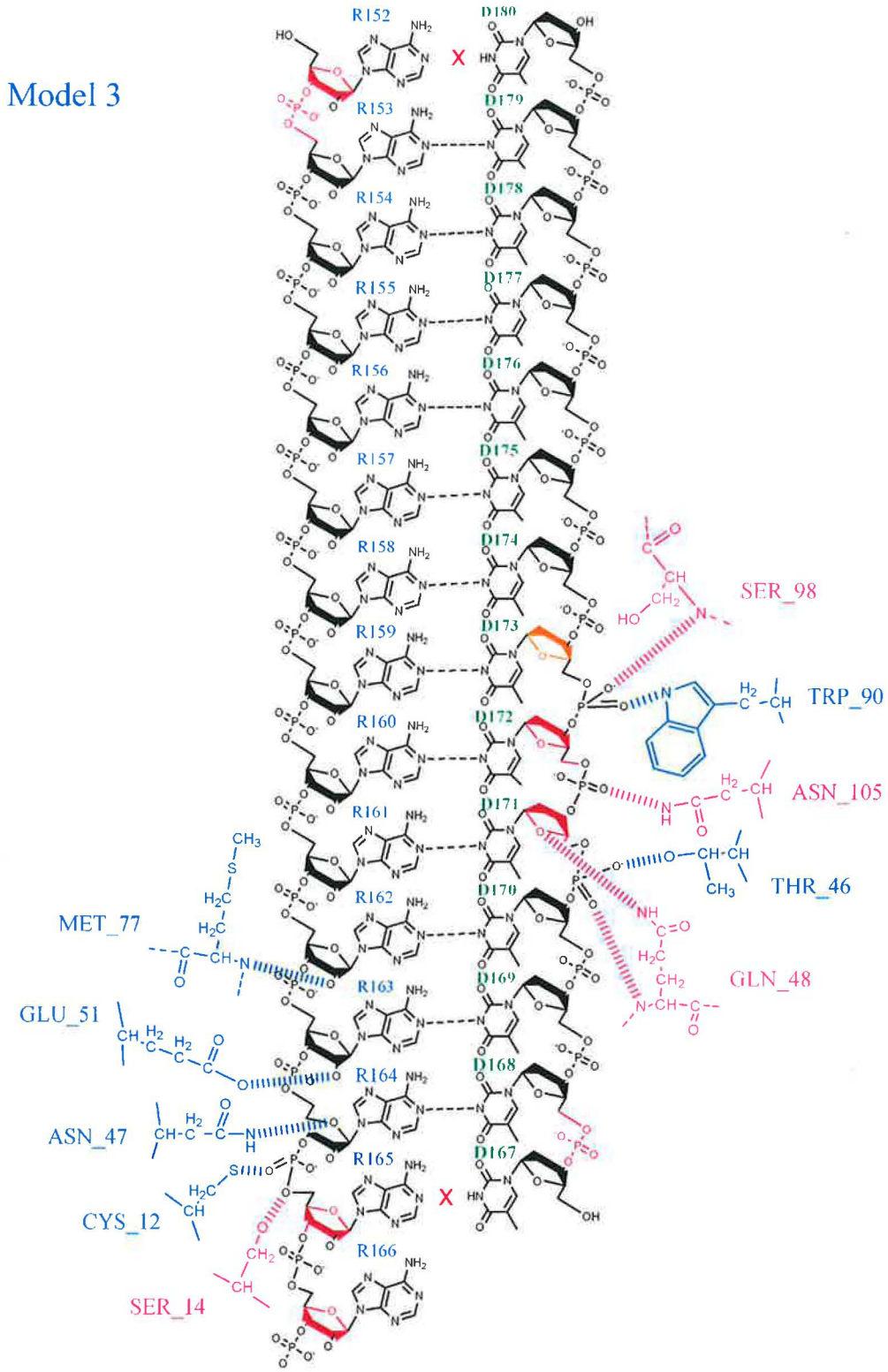


Figure 3.15: Model 3: Schema of the enzyme - substrate interaction

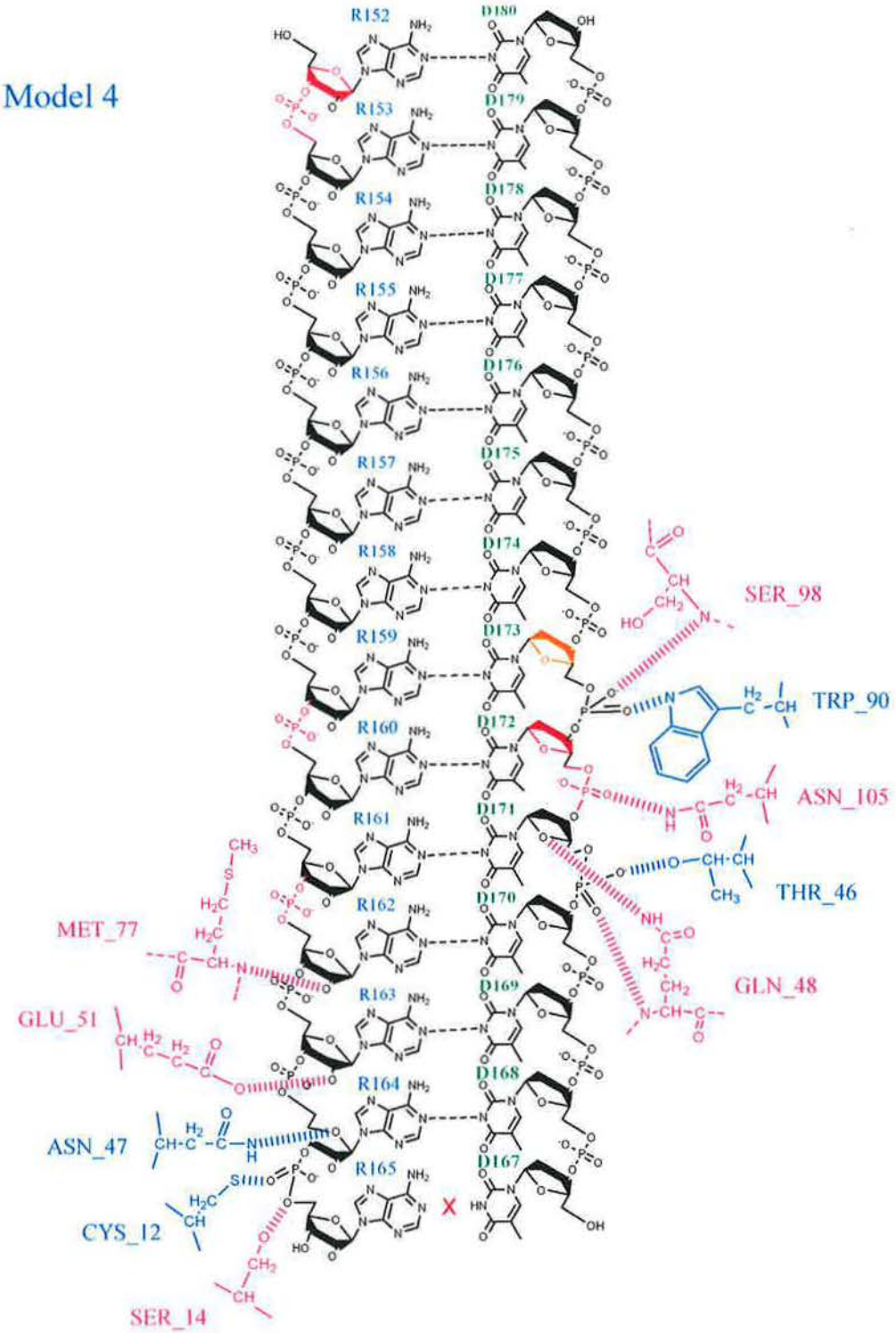


Figure 3.16: Model 4: Schema of the enzyme - substrate interaction

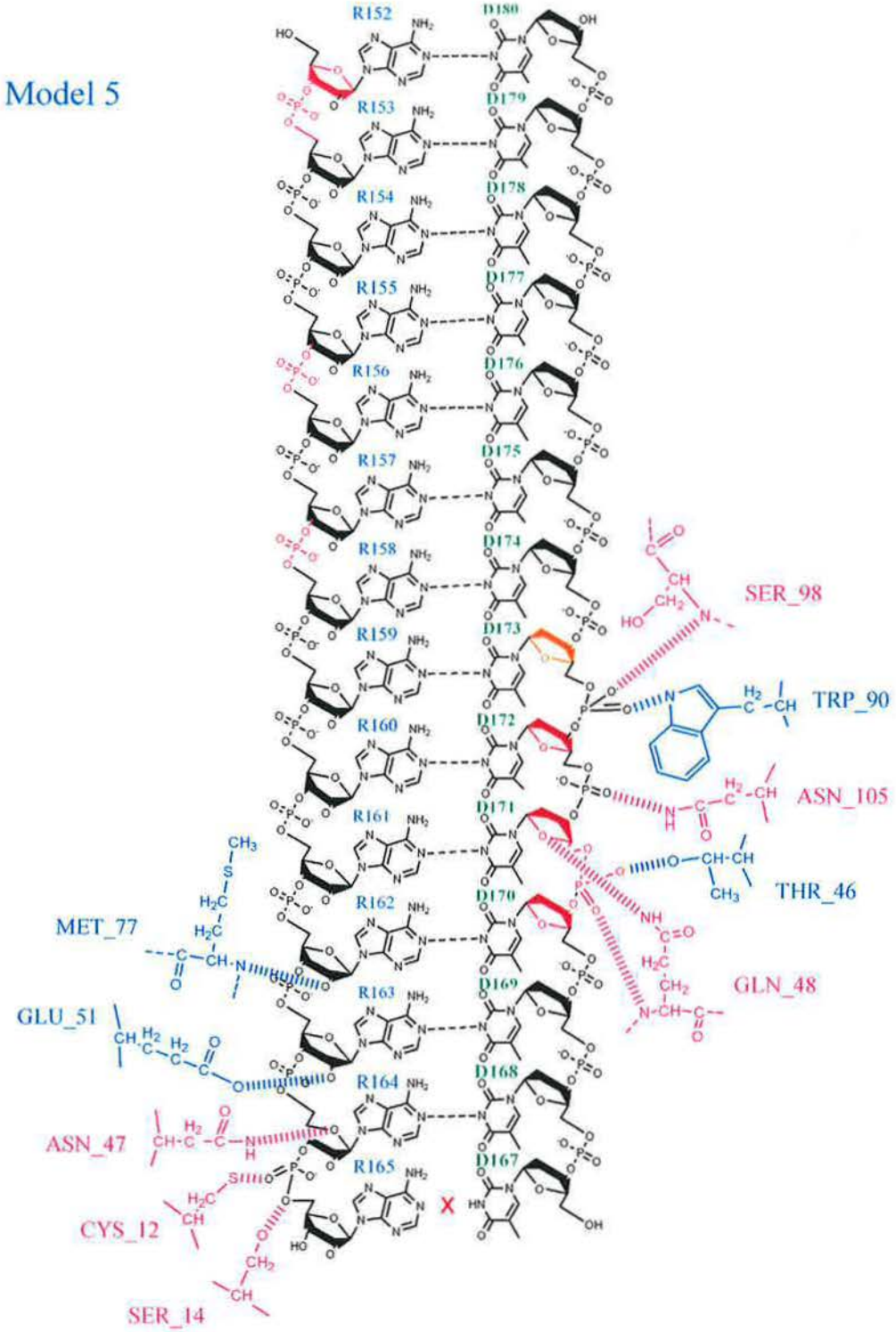


Figure 3.17: Model 5: Schema of the enzyme - substrate interaction

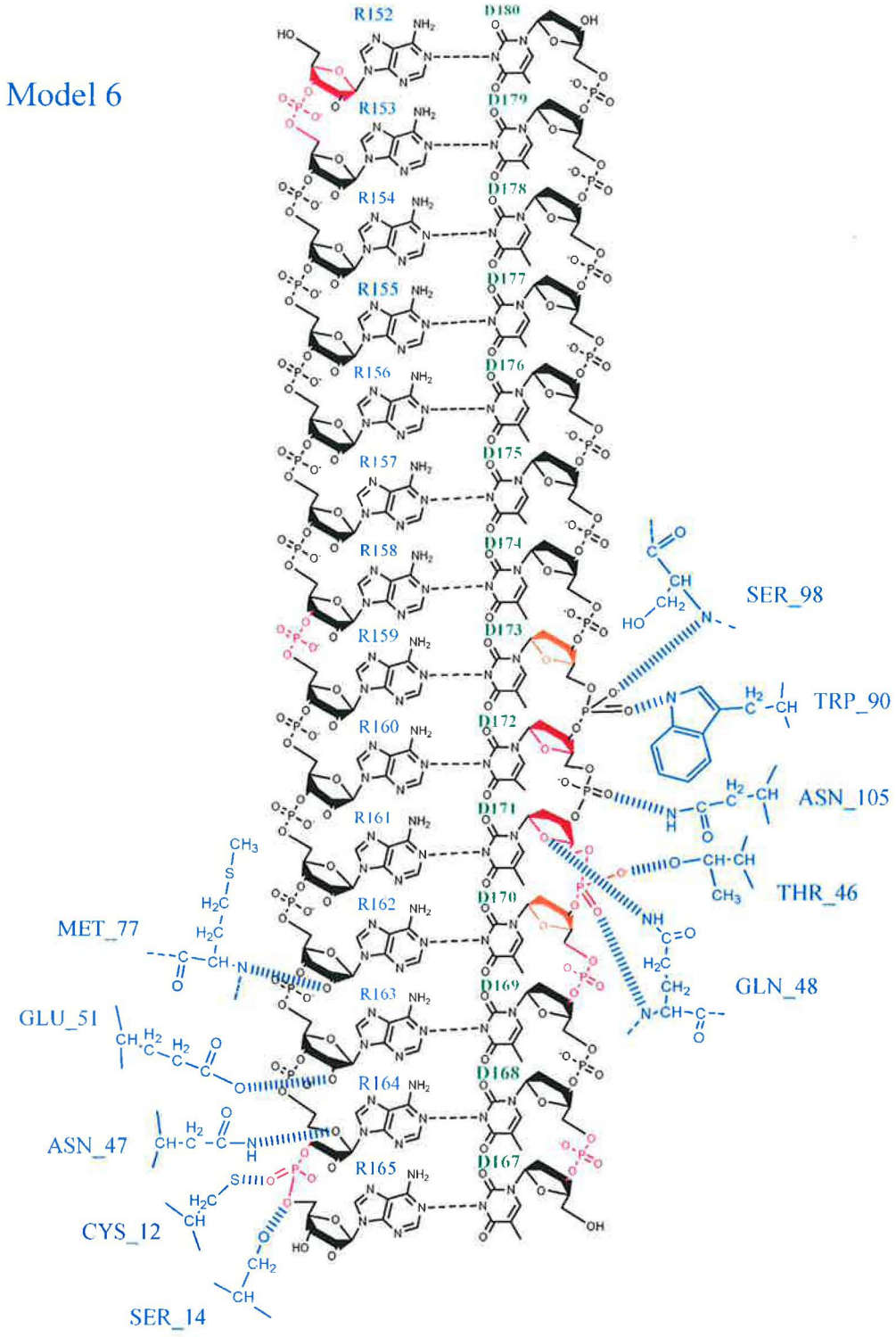


Figure 3.18: Model 6: Schema of the enzyme - substrate interaction

3.2.3 Model refinement II: active site stabilization - Mg ions parametrization

Several MD runs (Model 7-9) were devoted to the optimization of force field parameters describing properties of magnesium ions (see Figure 3.19). All MD runs were started from the conformation, which was reached by the Model 6 after a 5 ns MD run.

Model 7 (see Fig. 3.20 p. 65, Graphs B1-B8 Appendix B)

In the model 7 all the catalytic residues are closer to the ions, with mutual distances of approximately 1.8Å. There is observed an asymmetry, as the oxygen atom in the R163-R164 phosphate group is closer to the Mg2 ion. The enzyme-substrate hydrogen bonds, where the Ser98, Asn105 and Ser14 residues participate, are turned off. Moreover, the important hydrogen bond contact with the Trp90 residue disappeared. Disturbances in the torsion angle conformations (not observed in Model 6) were found in the R154-R155, R156-R157, R158-R159, R161-R162 and R162-R163 and D169-D170 phosphate groups. Newly also the D170 pseudorotation angle escapes from the interval of expected values.

Model 8 (see Fig. 3.21 p. 66, Graphs B9-B16 Appendix B)

The active site geometry was completely stable in the Model 8, where distances between the catalytic residues and ions reach about 2.5 Å(the same value as in Model 6). The overall RMSD of the whole structure is very low. All enzyme-substrate interactions began at the 2Å value, but several of them did not last during the whole MD simulation. Residues Asn105, Asn47 and Ser14 broke away. The Cys14 residue was situated in the distance of 2Å from the RNA strand with oscillations reaching to 6Å, which means an existing but not very stable hydrogen bond. Similarly the interaction of Gln48 with the DNA strand was reversibly switched off in some moments but reappeared afterwards. Sugar puckering and torsion angles conformational preferences are similar to the Model 6. There is a difference in the conformation of the R157 ribose, starting from the A-conformation but finishing in the B-conformation. As to the torsion angles, the -gt conformation is newly observed (compared to the Model 6) also at the R155-R156 and D168-D169 linkages. This model seems to be as suitable for a further usage as the Model 6.

Model 9 (see Fig. 3.22 p. 67, Graphs B17-B24 Appendix B)

All catalytic residues-ion distances in the active site assume lengths from 2.1 to 2.2Å. The active site geometry was found asymmetric and unstable. The oxygen atom in the R163-R164 phosphate group is closer to the Mg1 ion and there were found significant fluctuations considering the R164-Mg2 and Mg1-Mg2 distances. The RMSD of the nucleic acids is high, with maximum value outreaching 3.5Å. There were found several disturbances concerning the phosphodiester linkages. The interaction with the substrate, however, seems to be unaffected. The Asn105, Gln48 and Ser14 residues were unattached after few nanoseconds of the MD simulation and the Ser98 residue bound only weakly.

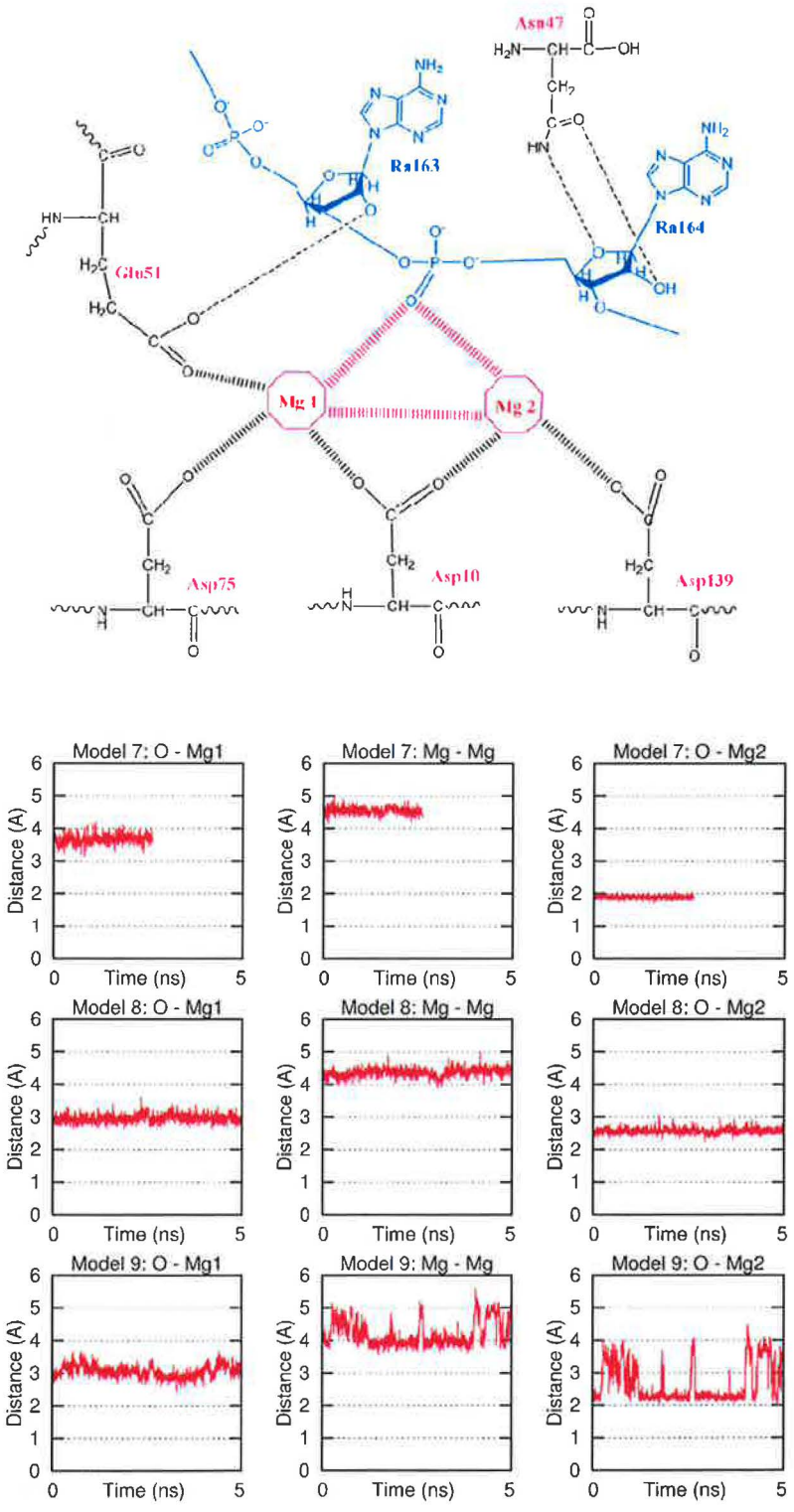


Figure 3.19: Schema of the active site and respective graphs for models 7,8 and 9

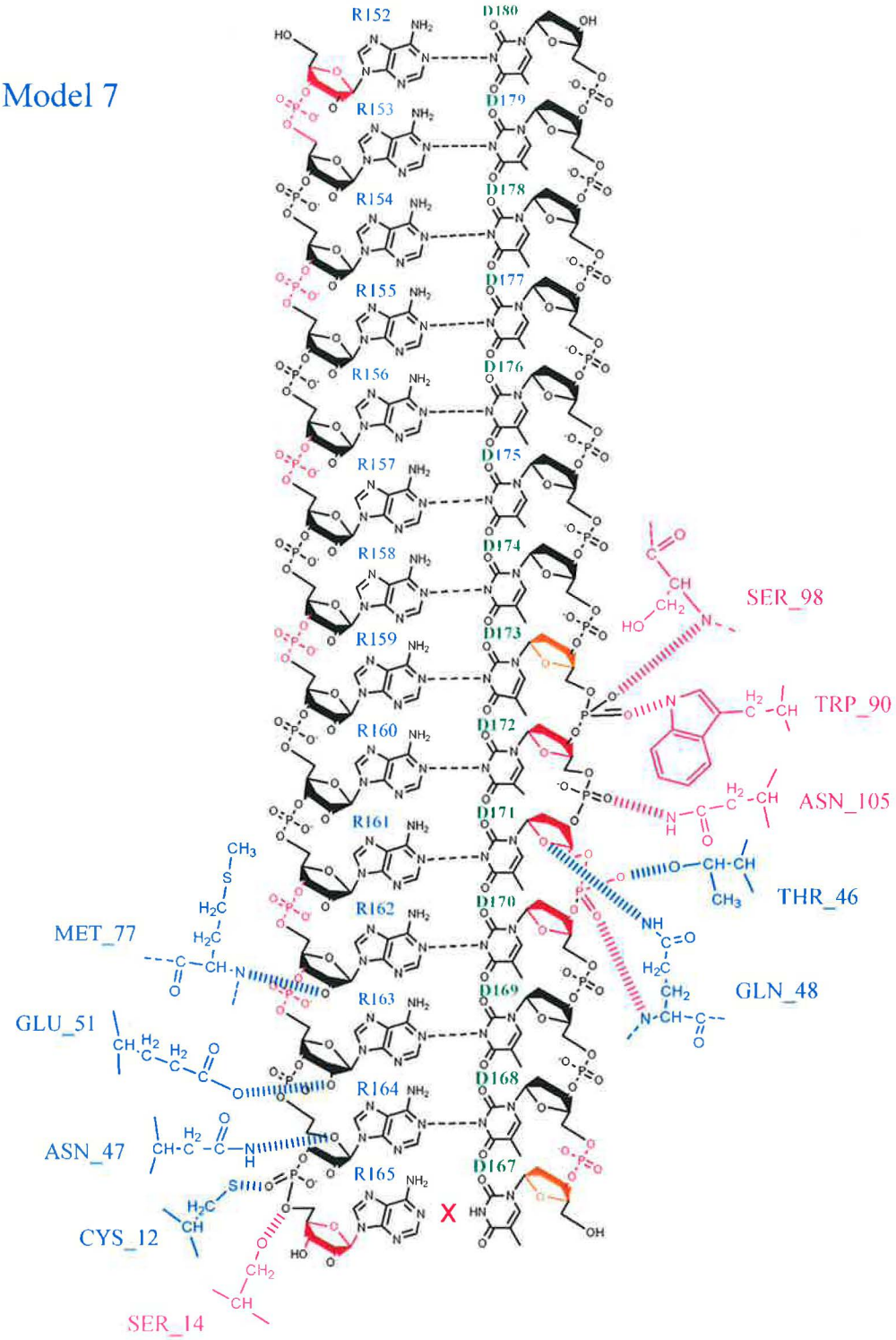


Figure 3.20: Model 7: Schema of the enzyme - substrate interaction

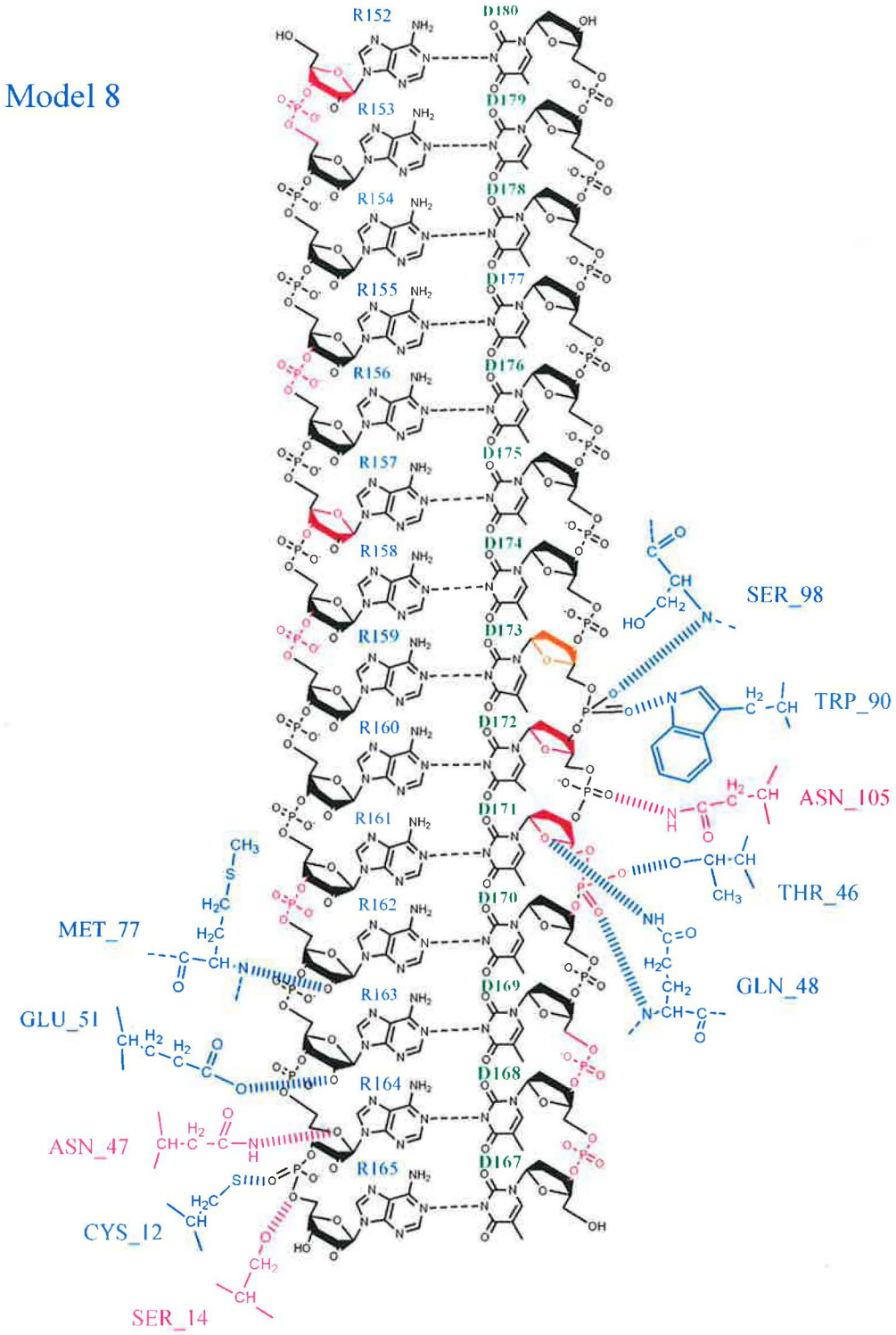


Figure 3.21: Model 8: Schema of the enzyme - substrate interaction

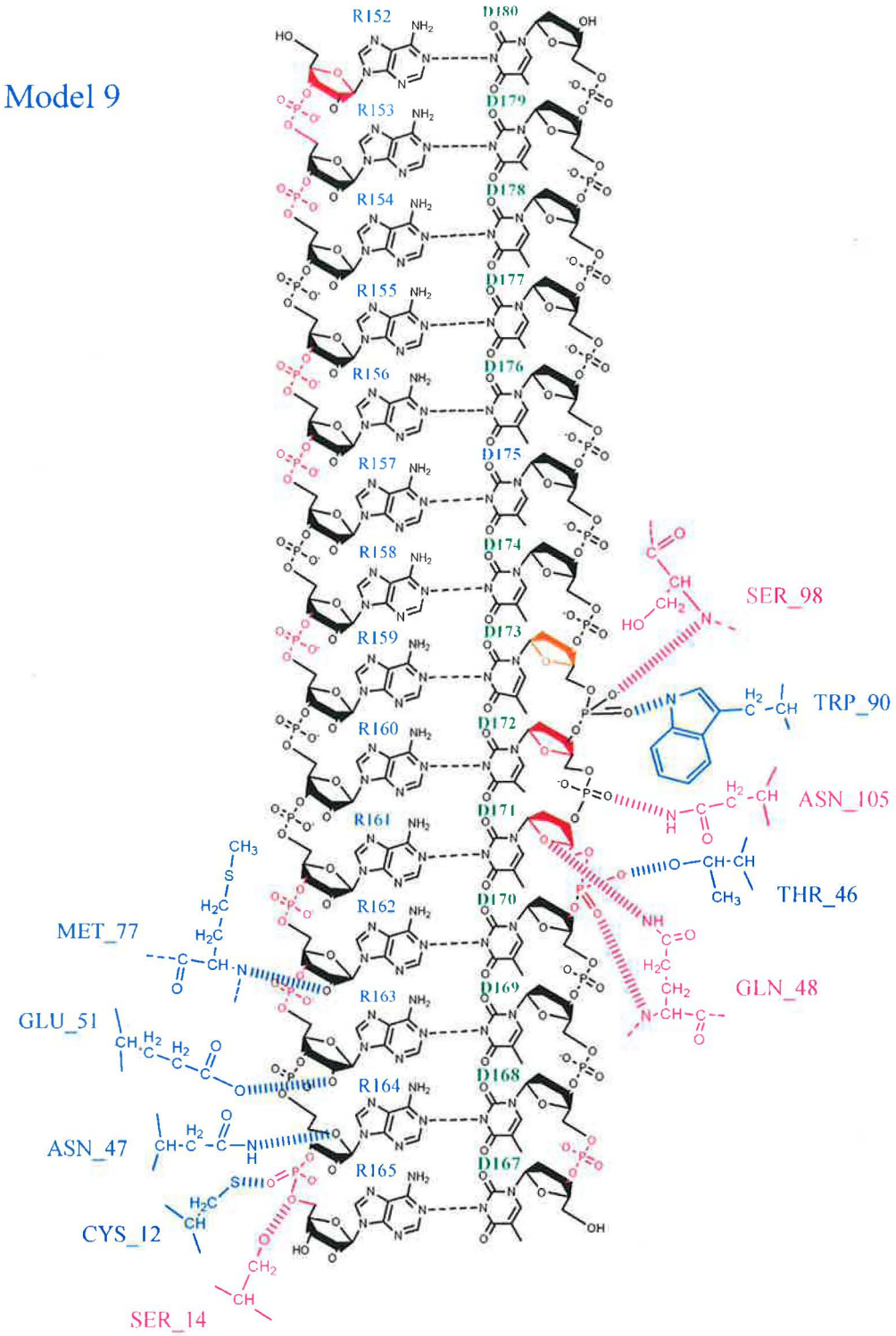


Figure 3.22: Model 9: Schema of the enzyme - substrate interaction

3.3 Models of RNase H bound to a modified substrate

Non-isosteric phosphonate internucleotide linkages were found to elicit the E. coli RNase H activity in vitro [50]. Subsequent MD simulations enabled to determine a putative position of the modified linkage leading to a substantial steric hindrance with the E. coli RNase H enzyme [23].

Recent MD simulations should shed light on the ability of human RNase H to recognize a substrate with prolongated non-isosteric phosphonate linkages incorporated into the nucleic acids backbone. All MD runs were started from the conformation, which was reached by the Model 6 after a 5-ns MD run.

3.3.1 The C3'-O-P-*C*-O-C5' modification

The C3'-O3'-P-C-O5'-C5' modification was constructed by insertion of a carbon atom into a phosphate group just by adding a new line into the initial PDB file. The carbon atom position was adjusted by hand using the VMD program. Two additional hydrogen atoms completing the methylene group were added automatically during the preparation stage using the TLEAP module.

Five models differing in positioning of the modified phosphonate linkage in the nucleic acids backbone were set up. Five residues close to the DNA-binding region of RNase H were covered up. All modified models were studied in the same way as their unmodified counterparts. The phosphonate linkage touched by the modification was characterized by three torsion angles (C3'-O3'-P-C, O3'-P-C-O5', P-C-O5'-C5').

Modification 1 (see Fig. 3.23 p. 71, Graphs C1-C8 Appendix C)

The modified linkage was positioned between the D171-D172 residues. It corresponds to the position, where the most severe steric conflicts were determined in the case of E. coli RNase H [23].

The highest RMSD values determined for the whole structure reached just 1.5Å (however, more than 3Å for the RNA strand alone). The active site geometry was remarkably stable. The distance separating both ions was slightly more than 4Å. The distance between the catalytic residues and ions (with the exception of Asp139(O1)-Mg2) was approximately 2.5Å (likewise in the Model 6). The same behavior was determined in the case of all modified structures mentioned below. The only subtle difference was found concerning the behavior of water molecules occupying hydroxide ion cavity. There are several MD runs, where the water molecule producing hydroxide ion was completely stable. On the other hand, there were found water molecules interchanging in the cavity within MD runs, too. Anyway, it seemed to be occupied by water molecules permanently. Therefore, no direct contacts between the rA164 phosphate group and the Mg ion (which could destabilize the duplex structure) were observed in subsequent MD runs. Some events, when water molecules were interchanging in the hydroxide ion cavity, were connected with the destabilization of the neighboring terminal base pair.

Almost all the studied residues anchored the substrate potently (with the exception of the Ser14 residue), whereas the Asn105 residue drew away during the simulation. This residue is situated in the vicinity of the modified internucleotide linkage. On the other hand, the initially distant Ser98 residue came closer. All the other measured enzyme-substrate hydrogen bonds are extraordinarily stable without any disturbances. They seem to be even more stable than in the unmodified structure.

The RNA/DNA duplex was stable - just the terminal R152-D180 base pair was separated. The sugar puckering of the RNA strand optimally corresponds with the C3'-endo A-conformation, just the terminal R152 and R165 residues exhibit a certain disturbances. The DNA strand was characterized by usual values of the pseudorotation angle falling into the interval of $100° - 150°$. There were some exceptions - the D171 residue adopted conformation characterized by values in the range of $50° - 100°$, the D172 residue assumed the C3'-endo conformation and the D170 residue preferred the pure C2'-endo conformation. The results roughly equal to the analysis of the unmodified Model 8. In contrast, differences may be found in the analysis of torsion angles. The modified D171-D172 linkage assumed the -g-g-g conformation during the major part of the MD simulation but terminated in the -gtt conformation.

Modification 2 (see Fig. 3.24 p. 72, Graphs C9-C16 Appendix C)

The modified linkage was positioned between the D172-D173 residues, where the important interaction with the Trp90 residue was found in unmodified models.

The RMSD values of this structure were very low, again, with the most pronounced dispersion observed for the DNA-strand. All hydrogen bonds between the enzyme and substrate were stable and strong. Small fluctuations were observed for Gln48, Asn47 and Ser14. The only residue, which broke away after 3ns, was the Asn105 residue. The sugar puckering was not affected by the modification too. The only exception was found in the case of the R157 ribose preferring the A-conformation (in contradiction to the initial structure). The torsion angles of the DNA strand were almost unchanged, but the modified linkage was found in the -gtt conformation. Also the R164-R165 internucleotide linkage was influenced by the interaction with the RNase H enzyme.

Modification 3 (see Fig. 3.25 p. 73, Graphs C17-C24 Appendix C)

The modified linkage was positioned between the D170 and D171 residues, where the phosphate binding pocket was proposed on the base of X-ray structures. The Asn105, Thr46, one of the Gln48 and Ser14 hydrogen bonds were turned off. Torsion angles and sugar puckering of the DNA strand were not affected by the modification. The ribose moiety of the R165 residue was slightly disturbed by the interaction with Ser14. A half of the phosphate linkages (much more than in the starting structure) were found in conformations different than -g-g.

Modification 4 (see Fig. 3.26 p. 74, Graphs C25-C32 Appendix C)

The modified linkage was positioned between the D169 and D170 residues in close proximity of the active site. The Asn105, one of the Gln48 bonds, Asn47 and Ser14 residues were excluded from the interaction with nucleic acids (similarly to the Modification 3). The terminal R165-D167 and R152-D180 base pairs were separated during the simulation. The conformations of the phosphate linkages in both strands were the same as in the initial structure. The only exception was found in the case of the modified linkage, where the -ggt conformation was preferred.

Modification 5 (see Fig. 3.27 p. 75, Graphs C33-C40 Appendix C)

The modified linkage was positioned between the D168 and D169 residues just against the scissile RNA linkage.

The only stable enzyme-substrate hydrogen bonds were those, which were established between Trp90 and D173, Thr46 and D171 and Glu51 and R163 residues. The Cys12 residue was bound to R165 weakly with fluctuations of interatomic distances reaching even 5Å. Other enzyme-substrate hydrogen bonds did not last for the whole 5ns. Ser98 was attracted to the D173 oxygen in the end of the simulation. This is a rather surprising finding, as none of the amino acids interacts directly with the modified nucleotide. The sugar puckering was the same as in Model 6. The R155 and R156 nucleotides switched to the C2'-endo conformation during the MD simulation. The terminal R165 residue was disturbed from the pure C3'-endo conformation. Internucleotide linkages were found predominantly in the -g-g conformation or in the same conformation as in Model 6 with the exception of the R153, R158 and D177 residues. In general, the parts of the duplex structure, which are distant from the binding site, are unusually strongly afflicted.

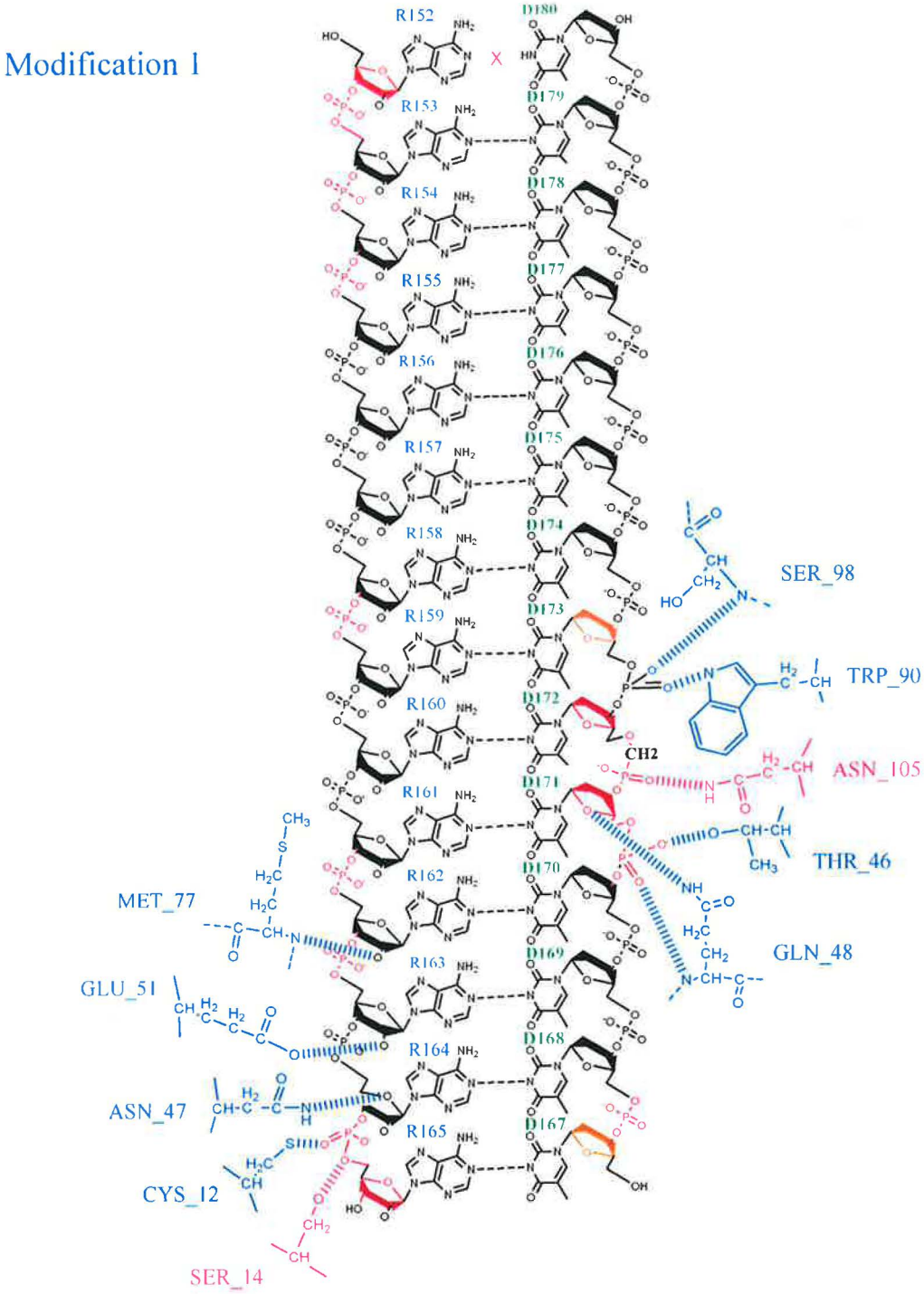


Figure 3.23: Modification 1: Schema of the enzyme - substrate interaction

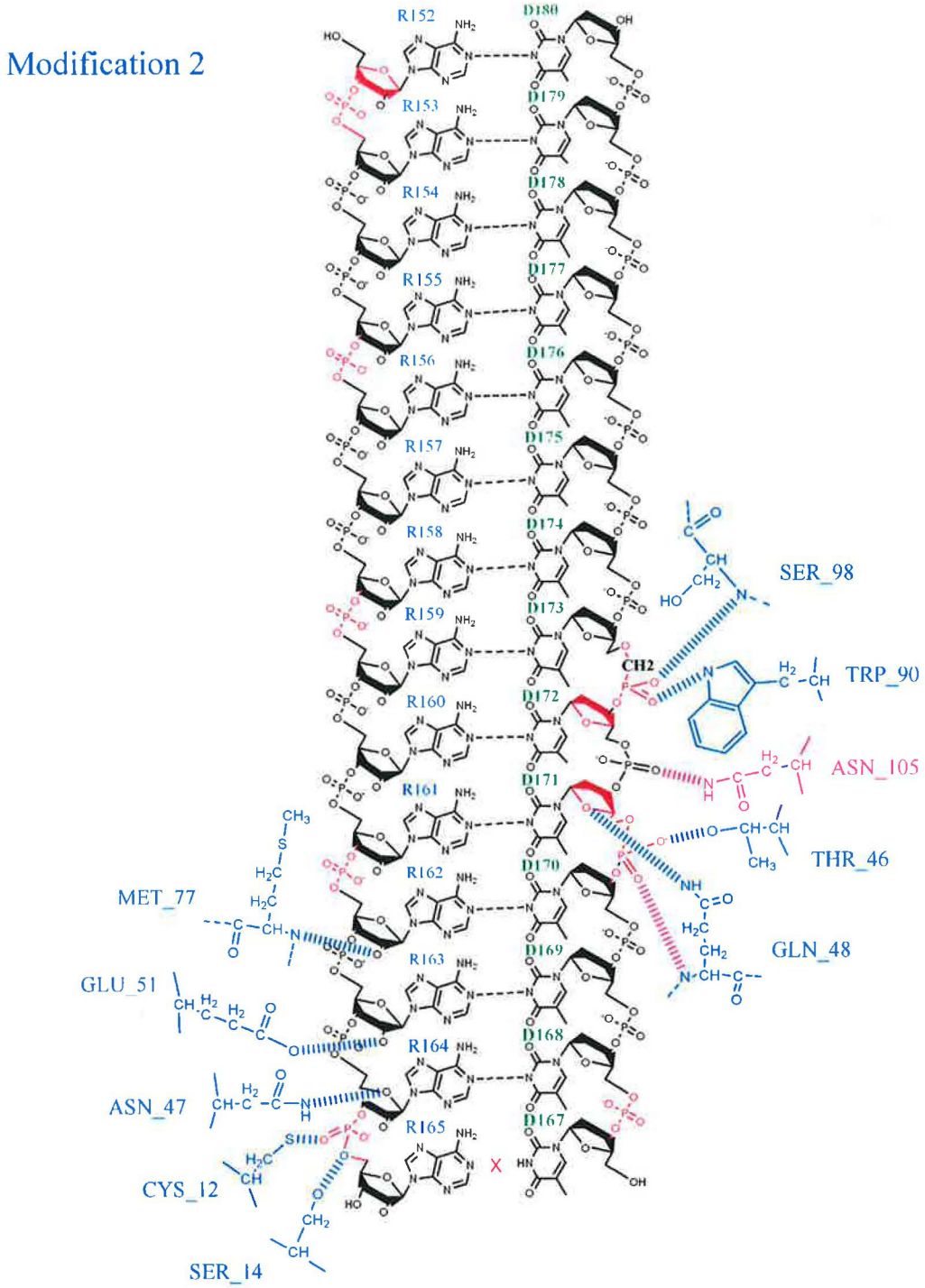


Figure 3.24: Modification 2: Schema of the enzyme - substrate interaction

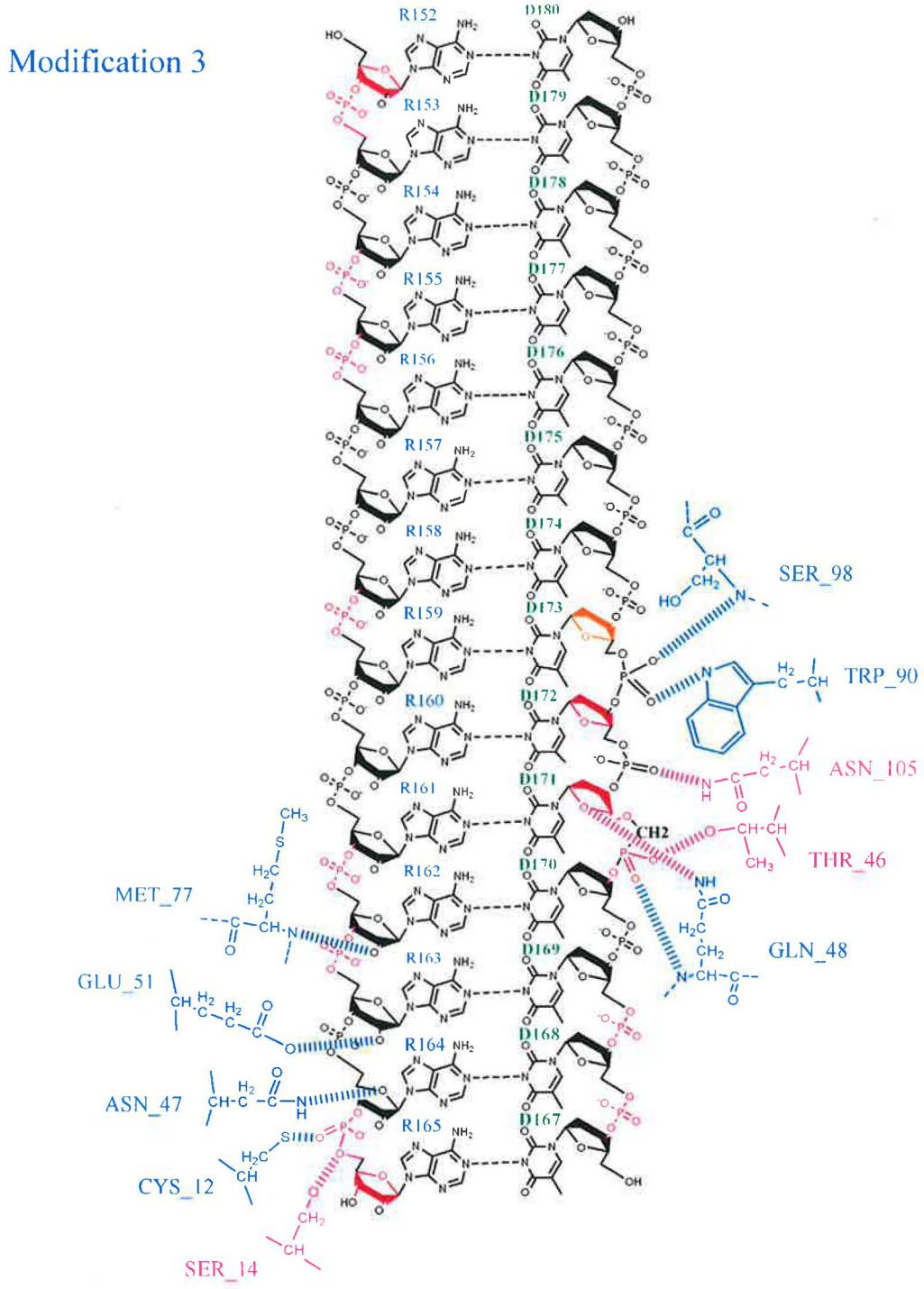


Figure 3.25: Modification 3: Schema of the enzyme - substrate interaction

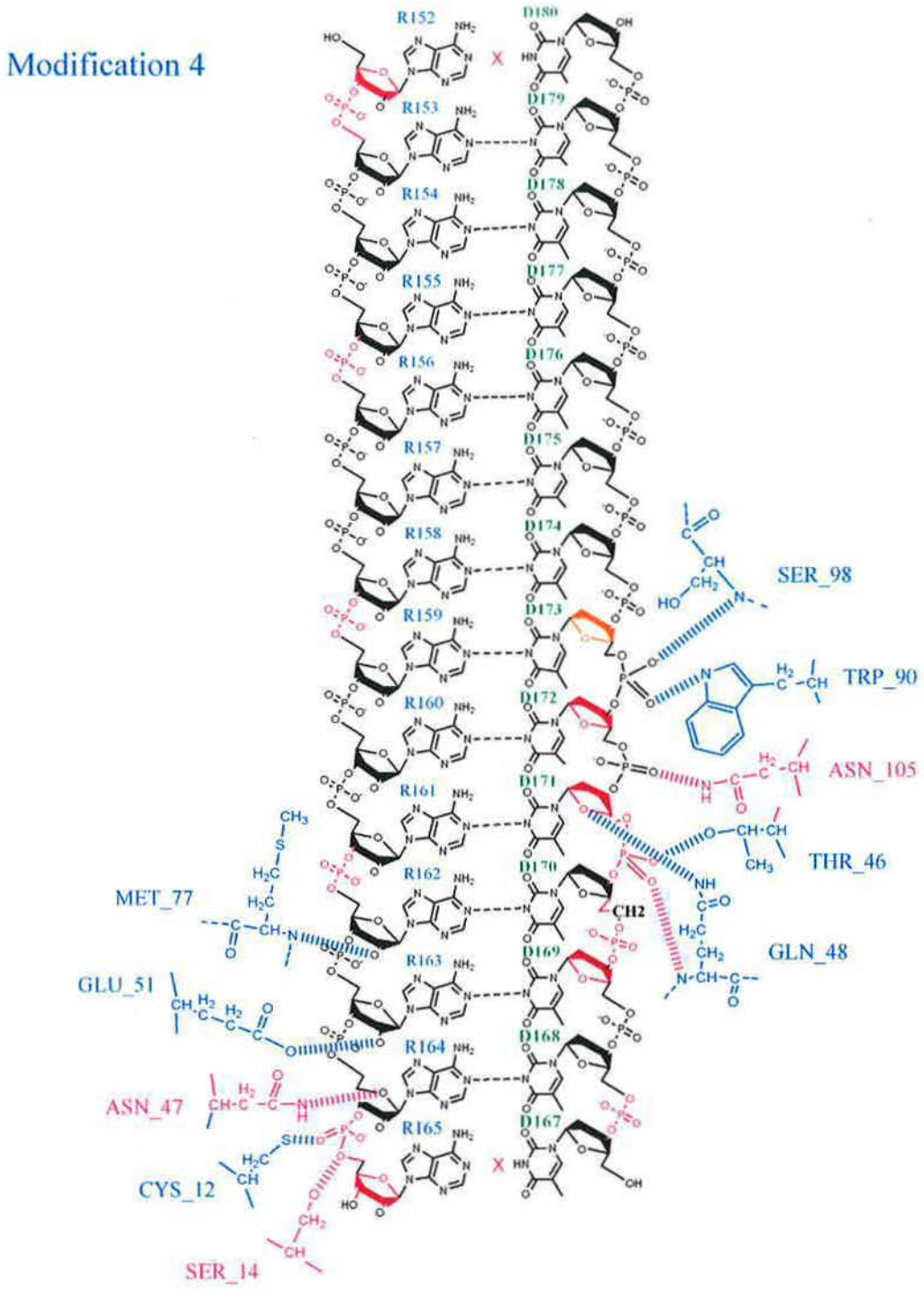


Figure 3.26: Modification 4: Schema of the enzyme - substrate interaction

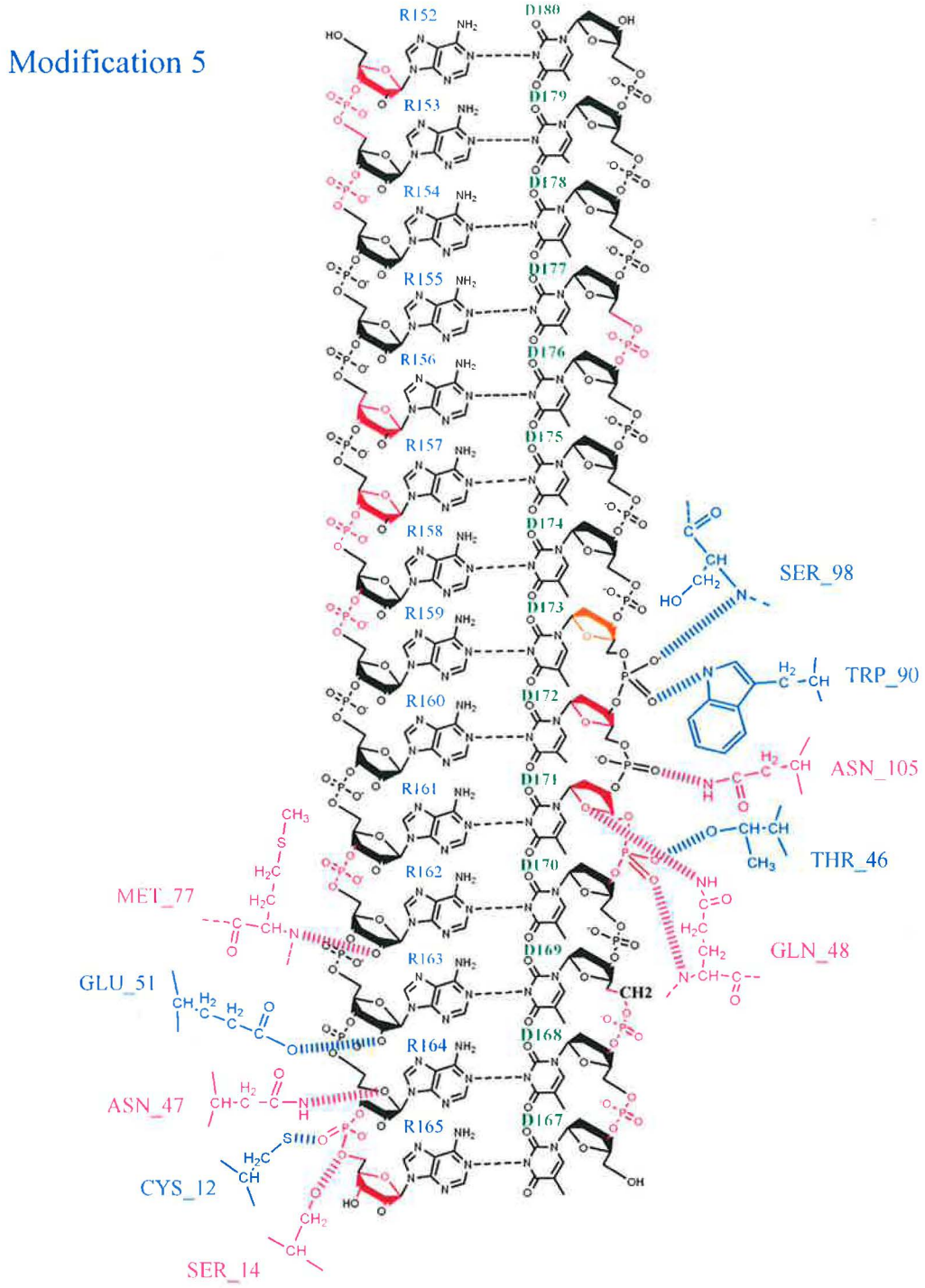


Figure 3.27: Modification 5: Schema of the enzyme - substrate interaction

3.3.2 The C3'-O-*C*-P-O-C5' modification

The methylene group was inserted in the position prior to the phosphate atom: C3'-O3'-C-P-O5'-C5' in the case of the second studied modification. The same five nucleotides (situated in the nearby of the DNA binding site of the human RNase H enzyme) were the object of interest.

Modification 6 (see Fig. 3.28 p. 78, Graphs D1-D8 Appendix D)

The modified linkage was positioned between the D171-D172 residues. It corresponds to the position, where the most severe steric conflicts were determined in the case of E. coli RNase H. The RMSD of the whole structure as well as of all separately studied segments was low and did not outreach 3Å. On the contrary, the interaction between the RNase H enzyme and DNA/RNA was softened. This is in concrete the interaction of Ser98-D173, Asn105-D172, Gln48-D171 and Asn47-R164. Also the residues Trp90 and Ser14 were stabilized at distances of cca 4Å in some periods of the simulation. Torsion angles of the RNA strand assumed values different from the starting model in the case of the R153-R154 and D168-D169 internucleotide linkages.

Modification 7 (see Fig. 3.29 p. 79, Graphs D9-D16 Appendix D)

The modified linkage was positioned between the D172-D173 residues, where the important interaction with the Trp90 residue was found in unmodified models. Similarly to the previous case, the Gln48, Asn105 and Ser14 residues were excluded from the enzyme-substrate interaction. Hydrogen bonds with the participation of the Asn47 and Met77 residues fluctuated more than usually observed. The sugar puckering of the D170 nucleotide changed during the MD simulation from the usual interval $100° - 150°$ to the interval $50° - 100°$. The torsion angles did not change their conformations in regard of the starting structure except the D168-D169 linkage, which was found in the preferable -g-g geometry now. The terminal R165-D167 base pair fell apart.

Modification 8 (see Fig. 3.30 p. 80, Graphs D17-D24 Appendix D)

The modified linkage was positioned between the D170 and D171 residues, where the phosphate binding pocket was proposed on the base of X-ray structures. MD simulation resulted in a structure with many stable enzyme-substrate hydrogen bonds (Trp90, Thr46, Asn47, Glu51 and Met77). However, the Ser98-D173 hydrogen bond contact seems to be substantially weaker. Remaining residues did not interact with the duplex. RNA nucleotides possessed expected conformations. Just the terminal R164 and R165 riboses fluctuated between several conformations. The terminal base pair R165-D167 was not bound. The torsion angles in the R161-R162, R162-R163 and R164-R165 internucleotide linkages, which are involved in the interaction with the RNase H enzyme were found in the -gt conformation.

Modification 9 (see Fig. 3.31 p. 81, Graphs D25-D32 Appendix D)

The modified linkage was positioned between the D169 and D170 residues in a close proximity of the active site. Surprisingly, the R156-D176 base pair was uncoupled. Also the pseudorotation angle for the D176 deoxyribose sank to 50°. The rest of the duplex structure did not change substantially.

The Gln48 residue seems to be bound for a period of time to the O4' atom of the deoxyribose ring of the D171 residue, whereas afterwards the interaction with the O1P atom in the phosphate internucleotide linkage prevailed. Also the interaction of Ser14 with the R165 nucleotide was initially stable and fell apart after a period of approximately 3ns. There was observed no hydrogen bond between the Asn105 and Ser98 residues and the duplex structure from the beginning of the simulation. The hydrogen bonds between Asn47, Ser14 and the RNA strand did not last either.

Modfication 10 (see Fig. 3.32 p. 82, Graphs D33-D40 Appendix D)

The modified linkage was positioned between the D168 and D169 residues just against the scissile RNA linkage. The enzyme-substrate hydrogen bonds were found as very unstable at the beginning of the simulation. The Ser98 and Trp90 residues were tightly bound to the DNA strand after a period of approximately 2ns. On the contrary, Gln48 and Asn47 broke away during the simulation. Only the Thr46, Met77 and Glu51 residues were anchored to the substrate during the whole simulated period. The RNA strand was rather unstable and some of the residues skip between -g-g and -gt conformation. The first R152-R153 internucleotide linkage is found in the gg conformation. The DNA strand seems to be unaffected.

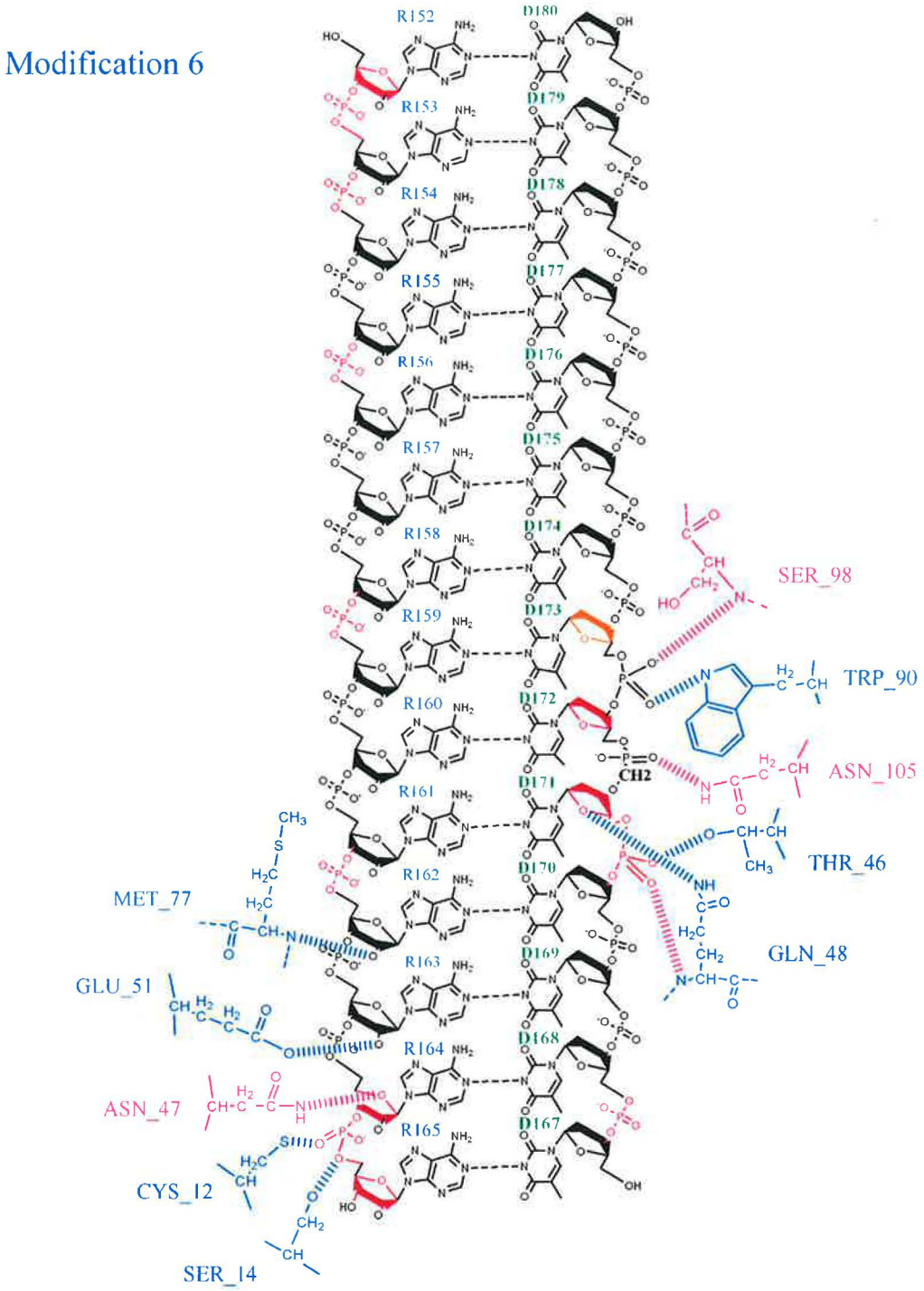


Figure 3.28: Modification 6: Schema of the enzyme - substrate interaction

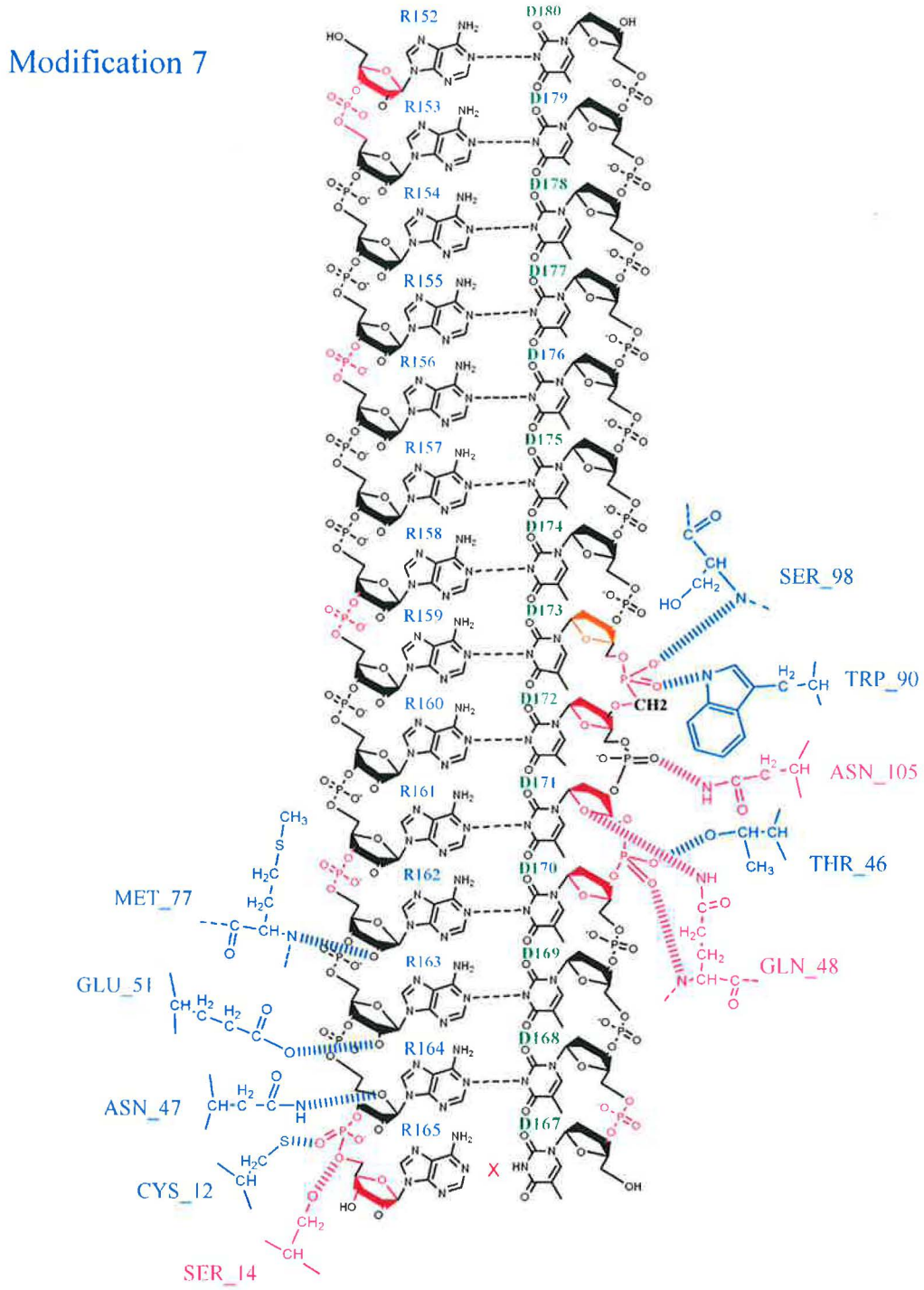


Figure 3.29: Modification 7: Schema of the enzyme - substrate interaction

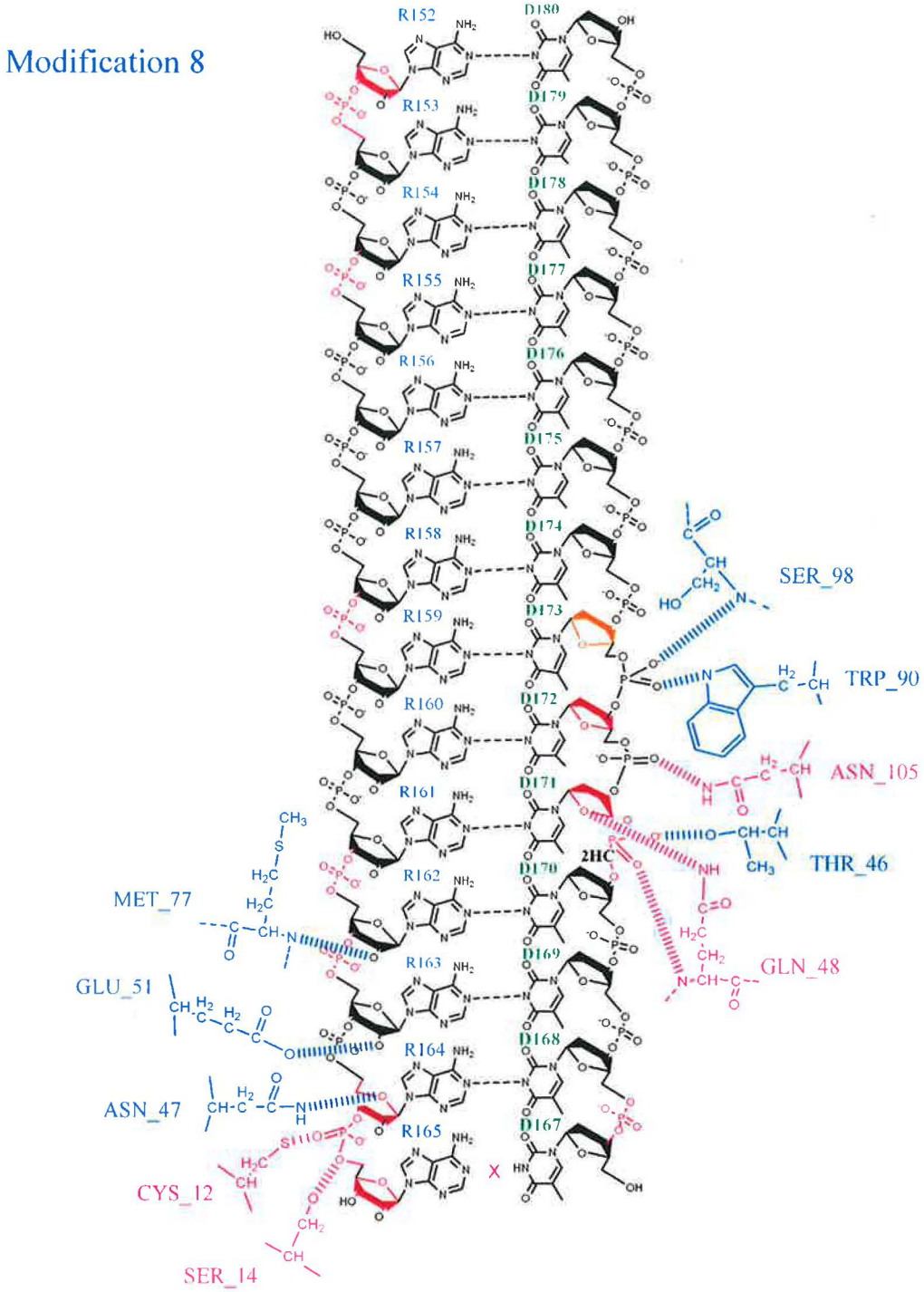


Figure 3.30: Modification 8: Schema of the enzyme - substrate interaction

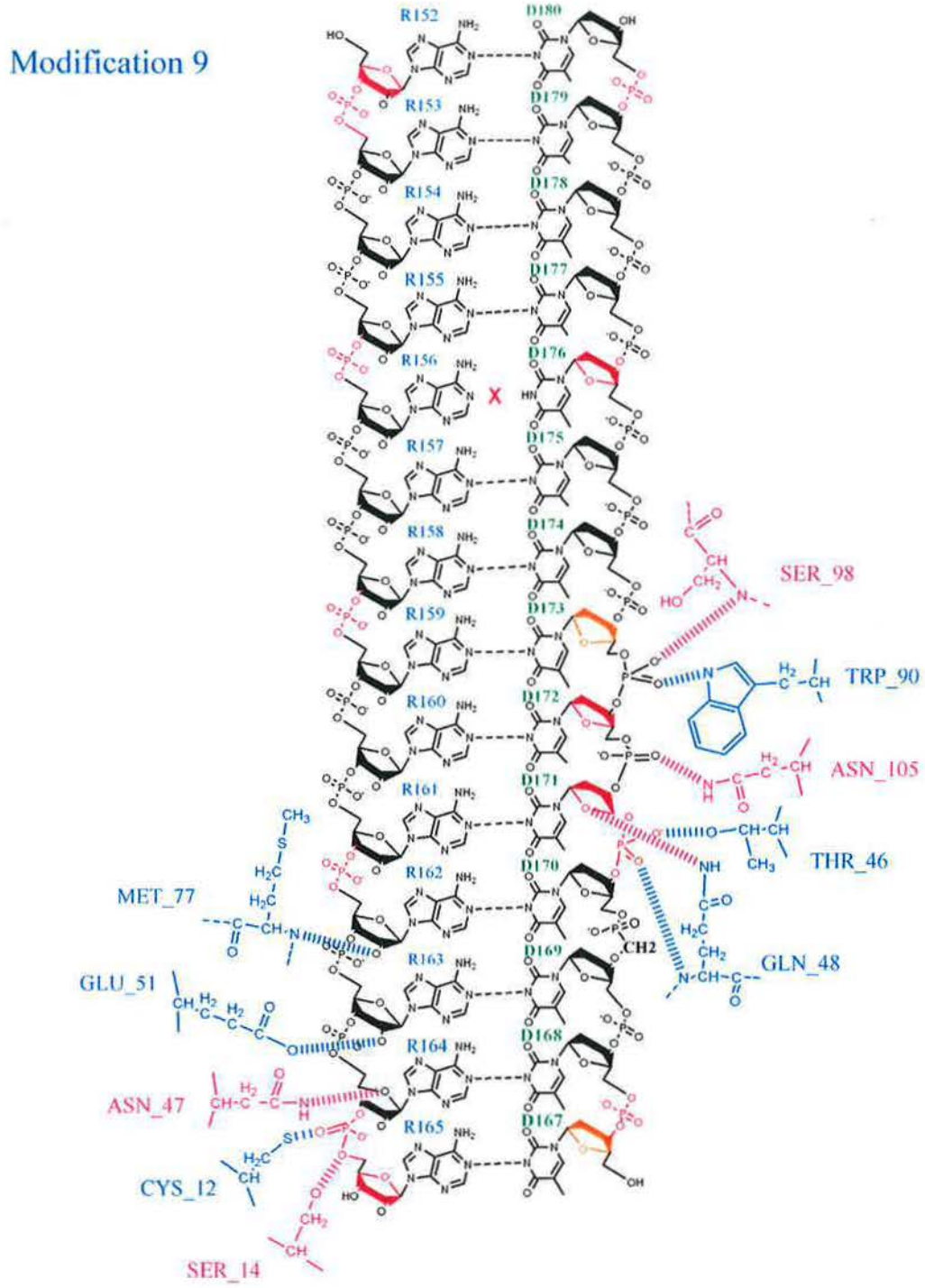


Figure 3.31: Modification 9: Schema of the enzyme - substrate interaction

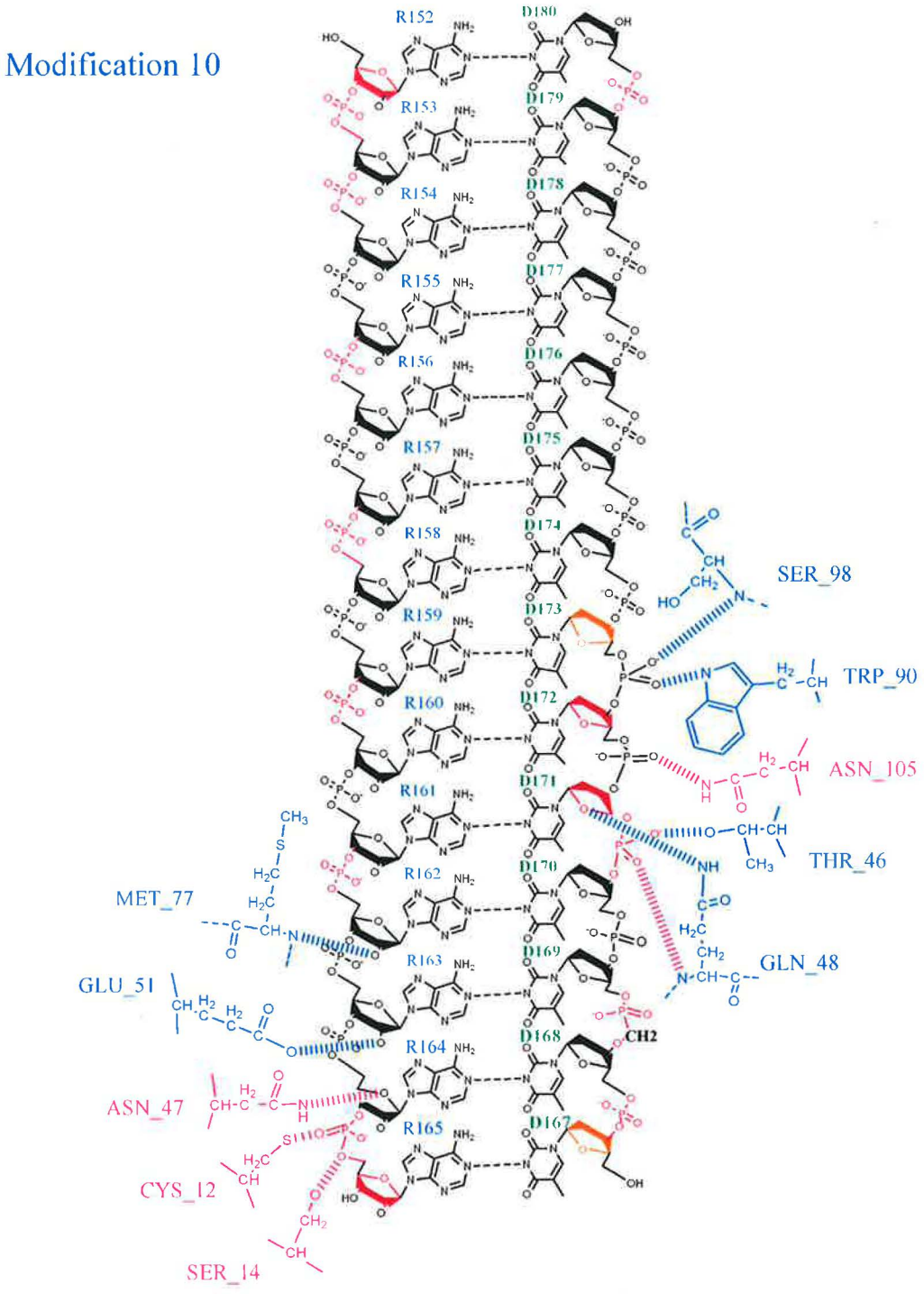


Figure 3.32: Modification 10: Schema of the enzyme - substrate interaction

3.4 E. coli RNase H with a substrate

Finally, an older MD run produced for E. coli RNase H in complex with a natural substrate (Figure 3.35)was analyzed using the methodology applied above. The graphs illustrating the analysis are to be found in Appendix E (E1-E8). The aim was to shed light on subtle differences in the recognition of nucleic acids substrates between both enzymes. Interestingly, there is just one Mg ion bound in the active site.

The RMSD of the model does not exceed 3Å. The Watson-Crick hydrogen bond connection of the DNA/RNA hybrid remained almost intact within the 5.4-ns MD run at 300 K. The proton-acceptor distances fluctuated prevailingly in the interval ranging from 1.7 to 2.5 Å. Heightened values determined for short periods for the terminal base pairs are a rather common observation as the ends of double helical structures are transiently frayed from time to time. However, in the case of central base pairs, it would indicate an adaptation to the shape of the substrate-binding site of the enzyme. The most remarkable affection of the hydrogen bond net was determined in the case of the R162-D171 and R163-D170 base pairs, where transient interruptions were observed for several tens of ps.

Many residues were involved in binding of the DNA (Arg93, Asn89, Trp90, Trp109, Tyr78, Lys104, Asn105, Thr48 and Asn21) and RNA (Gln81, Gln77, Tyr78, Lys127, Glu53, Asn49 and Cys18) strand to the substrate-binding site of the E.coli RNase H enzyme.

The average conformation of ribose moieties in the rA15 strand was 3'-endo. Despite of the fact, that fluctuations reached occasionally almost to 2'-endo, there was no remarkable re-puckering into this conformation. The deoxyriboses of the thymidine strand exhibited variable conformational behavior tending to the 3'-endo conformation (if situated in the nearby of the active site D172, D173, D174). This may be explained by the fact, that the hydrogen bonds with the enzyme take place here.

The phosphodiester linkages preferred the -g-g conformation (in terms of the C5'-O-P-O and O-P-O-C3' torsion angles). Occasional transitions into the -gt position were determined, too (as in the case of R157 and R161).

In comparison with the E. coli RNase H the number of H-bonds in the complex between human RNase H and substrate is notably smaller (see Figure 3.33, 3.34) and somewhat weaker. This is explained by the fact, that the catalytic domain of human RNase H is supported by the RNA-binding domain and the spacer region in nature.

```
 _aln.pos          10        20        30        40        50        60
Ecoli      -7KQVEIFTDGSCLGNPGPGGYGAILRYRGREKTFSAGYT---RTTNNRMELMAAIVALEALK--EHA
human      MGDFVVVYTDGCCSSNGRRRPRAGIGVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNIN
 _consrvd      *   *** *  *         *  * *       *     * ** * *  **  * *  *

 _aln.p    70        80        90       100       110       120       130
Ecoli      EVILSTDSQYVRQGITQWIHNWKKKGWKTADKKPVKNVDLWQRLDAALGQHQIKWEWVKGHAGHPENE
human      KLVLYTDSMFTINGITNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNE
 _consrvd    * ***     *** *   *** ****   * * *      *        * *  * ** *   **

 _aln.pos  140       150
Ecoli      RADELARAAAMNPTLEDTGYQVE
human      EADRLAREGAKQSED--------
 _consrvd  ** ***  *
```

Figure 3.33: Alignment of the E. coli and human RNase H with the DNA-binding (red) and RNA-binding (blue) residues and DDE motif (green)

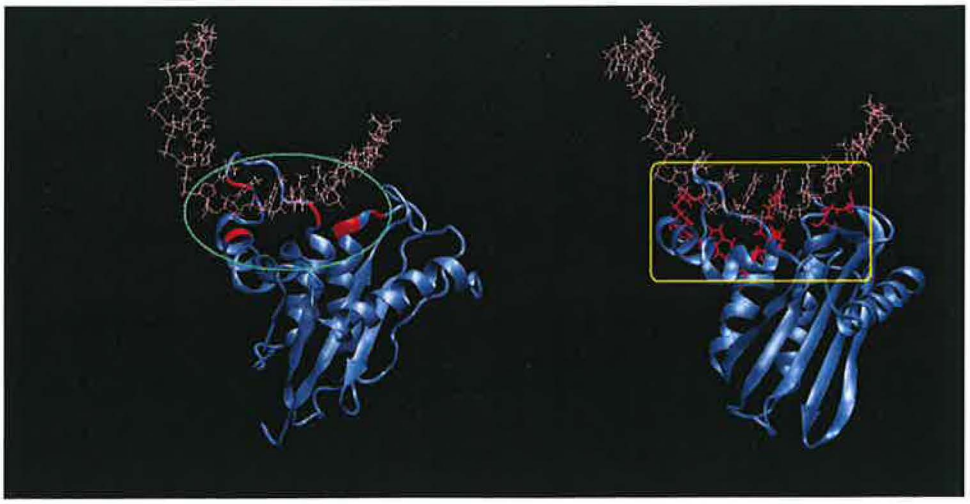


Figure 3.34: Comparison of the human (green circle) and E. coli (yellow square) structures with DNA-binding residues

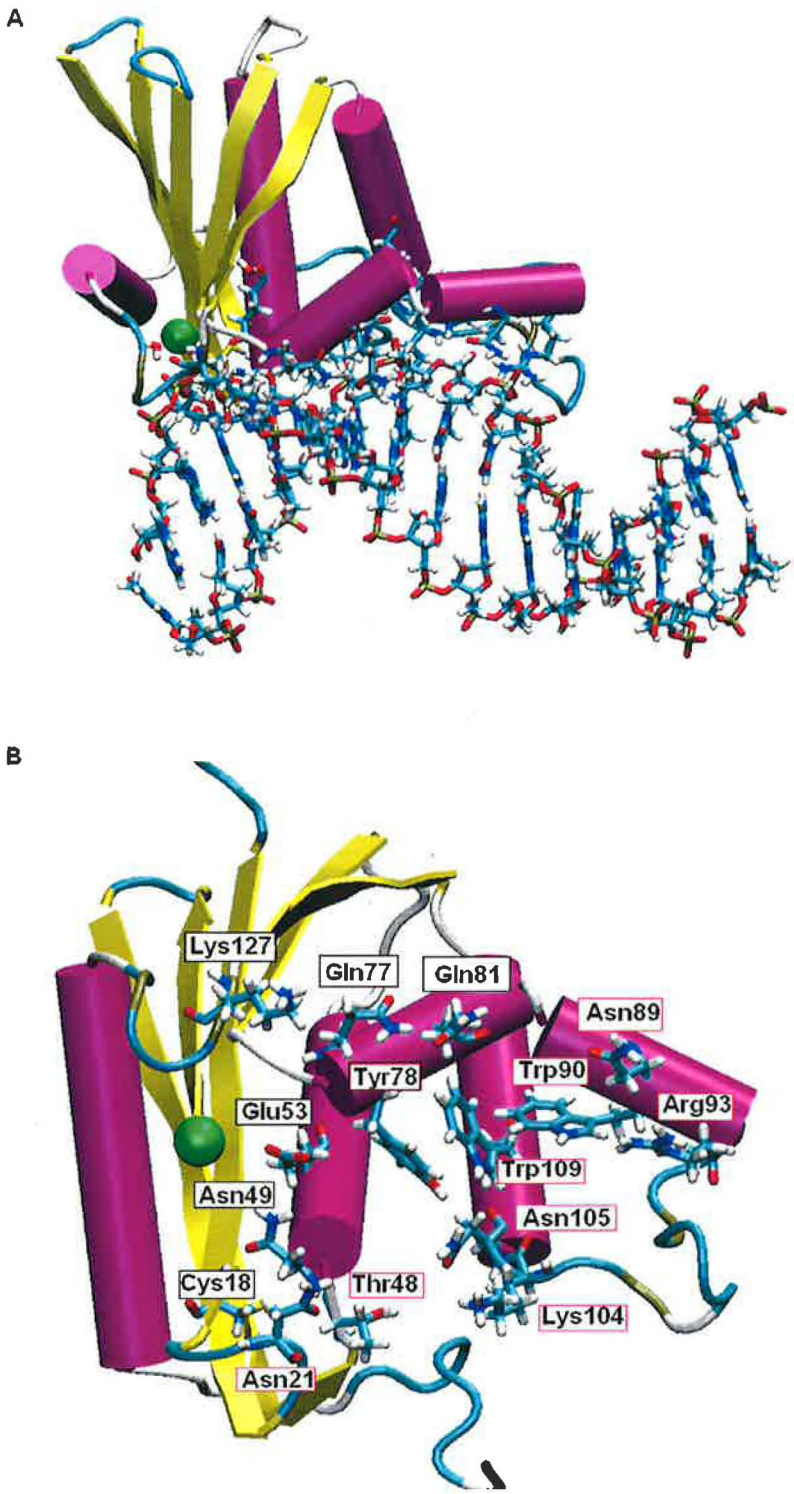


Figure 3.35: E.coli RNase H: structure and residues interacting with DNA-strand (red) and RNA-strand (black)

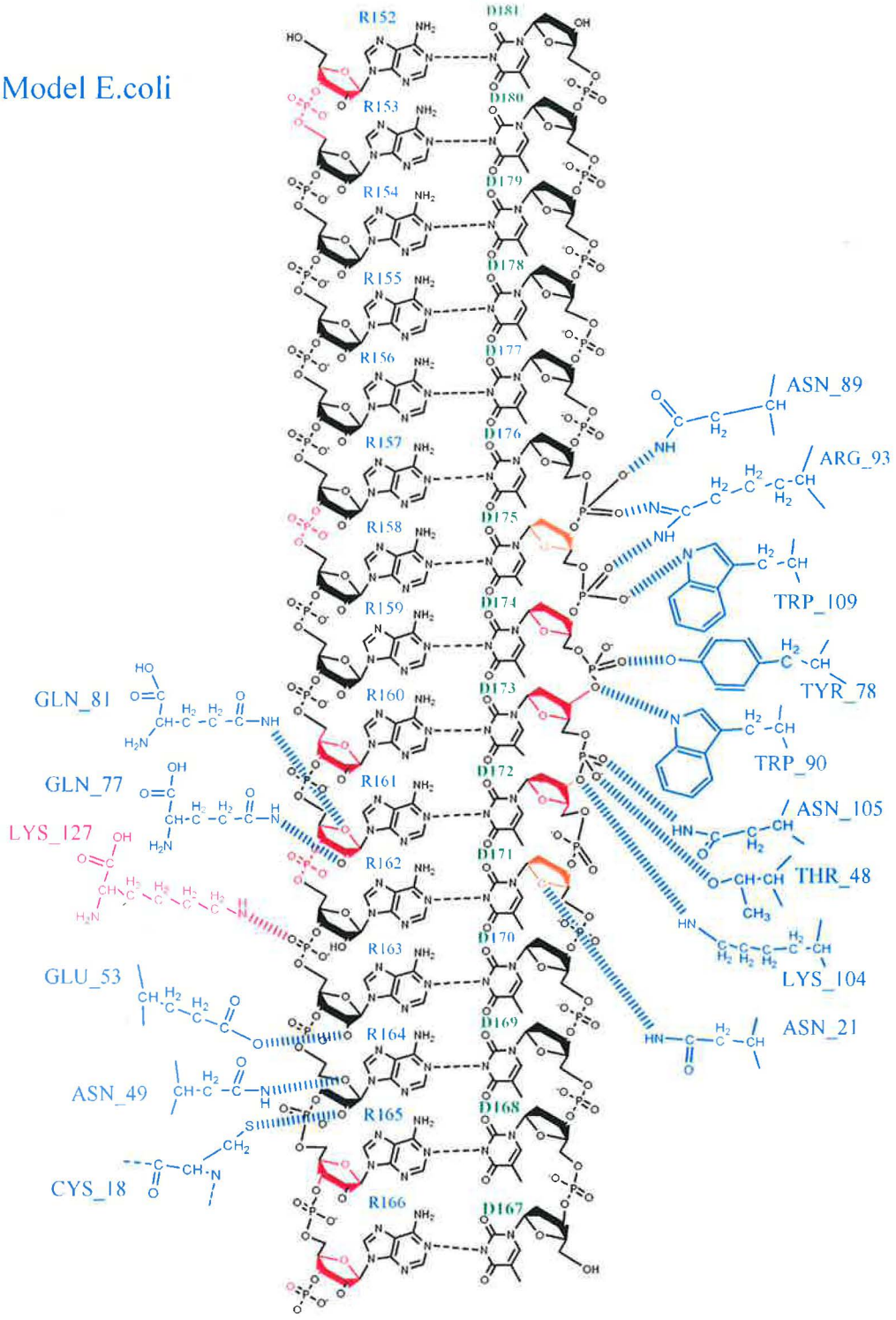


Figure 3.36: E. coli: Schema of the enzyme - substrate interaction

Chapter 4

Conclusions

The catalytic domain of the human RNase H enzyme was determined by means of homology modeling. Further, it was put together with a substrate on the base of analogy with the in silico model of a complex between E. coli RNase H and 15-mer RNA/DNA [23]. Molecular dynamics simulations were used to optimize mutual enzyme-substrate contacts (Model 1-6, see p. 54-62) and the active site geometry (Model 7-9, see p. 63-67). Further, the impact of the internucleotide linkage modifications in the DNA strand (either C3'-O3'-CH2-P-O5'-C5' or C3'-O3'-P-CH2-O5'-C5') was investigated (Modification 1-10, see p. 68-82). Major differences in recognition of the substrate by E.coli/Human RNase H were analyzed in detail (E. coli RNase H, see p. 83-86).

It was found out that there are 10 residues involved in the interaction with the substrate. DNA-binding residues are Ser98, Trp90, Thr46, Asn105 and Gln48. RNA-binding residues are Met77, Asn47, Glu51, Cys12 and Ser14. The DNA-binding residues contact phosphate groups from the D170-D173 deoxynucleotides (Ser98, Trp90/D173-D172, Asn105/D172-D171, Thr46, Gln48/D171-D170), which were the subject of chemical modifications. Further, the D170-D169 and D169-D168 linkages were modified to uncover potential destabilizations of the active site by highly flexible phosphonate linkers.

More or less significant disturbances of enzyme-substrate hydrogen bonds were observed predominantly in the nearby of modified linkages as expected. However, there is a just suspicion that several break-ups observed by the Trp90 residue were connected with the geometry of the distant active site. It seems, that either force constants describing properties of magnesium ions (Model 7: confront Figure B7 and B8 in Appendix B, Model 9: compare Figure B23 and B24 in Appendix B) or chemical modifications of linkages in the nearby of the active site (Modification 4: compare Figure C31 and C30 in Appendix C, Modification 5: compare Figure C39 and C38 in Appendix C, Modification 10: compare Figure D39 and D38 in Appendix D) could disturb a steady Trp90/D173-D172 hydrogen bond.

Ex post check of the in silico model derived here was made in April 2007. There was presented a preliminary X-ray structure on the NACON VII conference in Sheffield UK (M. Nowotny, W. Yang). As the most important DNA binding sites were proposed the Trp cavity (in particular the Trp90 residue) and the phosphate binding pocket consisting

of the Thr46, Asn105 and Arg43 residues stabilized by a net of mutual hydrogen bonds. It means that the only deficiency of the model derived here is the absence of the Arg43 residue involvement in interactions stabilizing the phosphate binding pocket. Interestingly, chemical modifications made on this position induce conformational transitions in the RNA strand touching the active site. Moreover, there is an occasional hydrogen bond contact between the backbone oxygen of the active site Asp75 residue and the R162 residue (counterpart of the D170 residue afflicted by the chemical modification).

Poorly contacted D171-D172 linkage is a surprise if previous findings made for E. coli RNase H are taken into the account (see below). Concerning the D169-D170 and D168-D169 linkages, which are not involved in enzyme-substrate contacts, a heightened conformational variability (more pronounced in the case of 3PC5 linkages) should be noted as it could lead to the disintegration of the active site geometry on longer time scales. There were determined large RMSD values in the case of 10-30, 30-70 enzyme segments (however, note that DOPE values offer another explanation here).

E. coli Rnase H consists of the catalytic domain only. Therefore, the recognition of the nucleic acids substrate is substantially stronger. Many residues were involved in the binding of the DNA (Arg93, Asn89, Trp90, Trp109, Tyr78, Lys104, Asn105, Thr48 and Asn21) and RNA (Gln81, Gln77, Tyr78, Lys127, Glu53, Asn49 and Cys18) strand. The most remarkable disturbance of the enzyme-substrate hydrogen bonding net was observed for structures, where modified internucleotide linkages were positioned in a way to interact with the Trp109, Tyr78, Lys104 and Asn105 residues (situated in the middle of the DNA binding site, where a cluster of Trp residues forms a rigid core of the protein structure) [23]. In the case of human RNase H, there were found aromatic residues not able to form hydrogen bonds instead of Trp109 and Tyr78. In fact, only Asn105 recognizes the substrate here. It means that human RNase H could be more tolerant of nonisosteric phosphonate residues compared to E.coli counterpart.

The hydrogen bond network in the active site (binding RNA, Mg ions and catalytic residues) was carefully established. Information from the Bacillus Halodurans RNase H structure was used [8]. Following carbonyl residues are substantial for proper stabilization of Mg ions in the active site: Asp 139, Asp10, Asp75 and Glu51. These residues constitute the highly conserved DDE motif (or DDDE in this case) observed in all RNase H structures. Interestingly, the aspartic acid 10 coordinates both Mg ions. Several MD runs (Model 7-9) were devoted to the optimization of force field parameters describing properties of magnesium ions to secure stability of interactions in the active site in agreement with Bacillus Halodurans X-ray structure [8].

There is a significant deficiency of the force field description of divalent cations [57]. These discrepancies are due to non-additive effects (neglected by the purely additive force fields) in the first ligand shell of Mg^{2+}, mainly inter-ligand polarization repulsion. In reality, the first-shell water molecules are heavily polarized by the ion, and their properties are very different from those of bulk water molecules. For example, they have the capacity to form strong, low-barrier hydrogen bonds with large infrared absorption shifts, where the hydrogen can easily jump between heteroatoms or may even be delocalized. Additionally, simulations are unable to provide a well-equilibrated distribution for multivalent ions.

Present force fields are biased towards direct (inner-shell) binding of Mg^{2+} to solute and require careful initial equilibration, since divalent cations consequently sample very poorly. There are substantial discrepancies considering the magnesium ion-ligand bond length (MD simulations 1.8Å, ab initio calculations 2.15Å, X-ray structures 1.8 - 2.6Å). Further, magnesium ions coordinate six ligands with strong directional preferences. Therefore, their representation as spheres bound to ligands by Van der Waals and Coulombic interactions is not fully adequate.

Regarding the above mentioned facts, force constants for Mg ions were optimized empirically to secure stability of the active site. Agreement with Bacillus Halodurans X-ray structure was chosen as a criterion for reliability of different models.

In the Model 7, all the catalytic residues are closer to the ions, with mutual distances of approximately 1.8 Å. There is observed an asymmetry, as the oxygen atom in the R163-R164 phosphate group is closer to the Mg2 ion. Moreover, the important hydrogen bond contact with the Trp90 residue disappeared. Similarly, the active site geometry was found asymmetric and unstable in the Model 9. There were found significant fluctuations considering the R164-Mg2 and Mg1-Mg2 distances. In contrast, the active site geometry was completely stable in the Model 8, where distances between the catalytic residues and ions reach about 2.5 Å. Therefore, this representation of Mg ions was used in further studies.

Properly coordinated water molecules at defined positions help to secure a stability of the active site, where strongly charged atoms are in close contacts (carbonyl side chains, phosphate groups and positively charged Mg ions). In particular, a water molecule bridging the oxygen atom from the R164 residue and the Mg ion was positioned manually prior MD runs were started, in a way to avoid undesired direct contacts between nucleic acids and ions. The water molecule has a functional importance, as it enables a creation of the hydroxide ion attacking a scissile bond. Interestingly, there are several MD runs, where the water producing the hydroxide ion was completely stable. On the other hand, there were found water molecules interchanging (2-3 within a MD run) in the hydroxide ion cavity too. Anyway, the cavity seemed to be occupied permanently by water molecules. It means that the recent model could be successfully used for more sophisticated QM/MD calculations.

It should be noted, however, that the two-ion mechanism proposed for nucleic acids cleavage is not fully accepted. It was derived on the base of X-ray structures of Bacillus Halodurans RNase H as well as of many polymerases, where two ions were found almost exclusively. Nevertheless, there are alternative proposals (based on functional data): the Asp134 residue could stabilize the attacking hydroxide, while the second Mg ion would be responsible rather for the inhibition of the RNase H enzyme. Therefore, the instabilities in the active site geometry (Model 7, Model 9 - smaller Mg ions) are not necessary artifacts. It could indicate inability of the unquiet Mg ion to stabilize the attacking hydroxide ion properly. It will be the subject of a further investigation.

The basic protrusion was found to be less important for the substrate binding than originally expected. Hydrogen bonds binding Lysine residues with nucleic acids were observed only occasionally. Maybe, the involvement of the basic protrusion is important rather

in the early stages of the enzyme-substrate recognition, when the active site geometry is not fully established. There was observed a disquiet considering the R159-D173 hydrogen bonds situated on the edge of the basic protrusion. Interestingly, certain chemical modifications (or even mismatches) positioned here influent positively catalytic abilities of RNase H enzyme. Further, the R156-D176 hydrogen bond was disturbed in the Model 8 and Modification 9 MD runs. It seems like a weak chain link for a hybrid DNA/RNA duplex distorted by interactions with the RNase H enzyme. Atypical conformations of deoxyriboses observed in the case of D170 and D171 could partially explain the ability of RNase H to bind DNA/DNA, DNA/RNA and RNA/RNA duplexes.

Bibliography

[1] P. Karlson. *Základy biochemie.* Academia, 1971.

[2] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Essential cell biology.* Garland Publishing, 1998.

[3] http://en.wikipedia.org/wiki/dna#physical_and_chemical_properties.

[4] http://www.physicsforums.com/showthread.php?t=33762.

[5] W. Saenger. *Principles of nucleic acid structure.* Springer-Verlag, 1984.

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

[7] I. Nezbeda, J. Kolafa, and M. Kotrla. *Úvod do počítačových simulací: Metody Monte Carlo a molekulární dynamiky.* Karolinum, 2003.

[8] M. Nowotny, S.A. Gaidamakov, Crouch R.J, and W. Yang. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell*, 121:1005–1016, 2005.

[9] N. Ohtani, M. Haruki, M. Morikawa, and S. Kanaya. Molecular diversities of RNases H. *Journal of Bioscience and Bioengineering*, 88(1):12–19, 1999.

[10] S.G Sarafianos, K. Das, Ch. Tantillo, A.D. Clark Jr, J. Ding, J.M. Whitcomb, P.L. Boyer, S.H. Hughes, and E. Arnold. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *The EMBO Journal*, 20:1449–1461, 2001.

[11] K. Katayanagi, M. Miyagawa, M. Matsushima, M. Ishikawa, S. Kanaya, M. Ikehara, T. Matsuzaki, and K. Morikawa. Three-dimensional structure of ribonuclease H from E. coli. *Nature*, 347:306–309, 1990.

[12] W. Yang, W.A. Hendrickson, R.J. Crouch, and Y. Satow. Structure of ribonuclease H phased at 2A resolution by MAD analysis of the selenomethionyl protein. *Science*, 249:1398–1405, 1990.

[13] L.V. Loukachevitz and M. Egli. Crystallization and preliminary X-ray analysis of Escherichia coli RNase HI-dsRNA complexes. *Acta Cryst.*, F63:84–88, 2007.

[14] J. Li and R.M. Wartelli. RNase H1 can catalyze RNA/DNA hybrid formation and cleavage with stable hairpin or duplex DNA oligomers. *Biochemistry*, 37:5154–5161, 1998.

[15] W.F. Lima and S.T. Crooke. Cleavage of single strand RNA adjacent to RNA-DNA duplex regions by escherichia coli RNase H1. *Journal of Biological Chemisty*, 272(44):27513–27516, 1997.

[16] S. Kanaya, A. Kohara, Y. Miura, A. Sekiguchi, S. Iwai, H. Inoue, E. Ohtsuka, and M. Ikehara. Identification of the amino acid residues involved in an active site of Escherichia coli ribonuclease H by site-directed mutagenesis. *Journal of Biological Chemisty*, 265(8):4615–4621, 1990.

[17] H. Nakamura, Y. Oda, S. Iwai, H. Inoue, E. Ohtsuka, S. Kanaya, S. Kimura, C. Katsuda, K Katayanagi, K. Morikawa, H. Miyashiro, and I. Ikehara. How does RNase H recognize a DNA.RNA hybrid? *Proc.Nat.Acad.Sci.USA*, 88:11535–11539, 1991.

[18] W. Yang, J.Y. Lee, and M. Nowotny. Making and breaking nucleic acids: Two-(Mg2+)ion catalysis and substrate specificity. *Molecular Cell*, 22:5–13, 2006.

[19] M. Haruki, E. Noguchi, Y.Y. Liu, M. Oobatake, M. Itaya, and S. Kanaya. Investigating the role of conserved residue Asp134 in Escherichia ribonuclease HI by site-directed random mutagenesis. *Eur. J. Biochem.*, 220:623–631, 1994.

[20] S. Kanaya, C. Katsuda-Nakai, and M. Ikehara. Importance of the positive charge cluster in Escherichia coli ribonuclease HI for the effective binding of the substrate. *Journal of Biological Chemistry*, 266(18):11621–11627, 1991.

[21] Y. Tsunaka, M. Haruki, M. Morikawa, and S. Kanaya. Strong nucleic acid binding to the Escherichia coli RNase HI mutant with two arginine residues at the active site. *Biochimica et Biophysica Acta*, 1547:135–142, 2001.

[22] K. Katayanagi, M. Miyagawa, M. Matsushima, M. Ishikawa, S. Kanaya, H. Nakamura, M. Ikehara, T. Matsuzaki, and K. Morikawa. Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution. *J.Mol.Biol.*, 223:1029–1052, 1992.

[23] I. Barvik Jr. E. coli RNase Hi and the phosphonate–DNA/RNA hybrid: Molecular dynamics simulations. *Nucleosides, Nucleotides & Nucleic Acids*, 24:435–441, 2005.

[24] V. Sosunov, E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov, and A. Goldfarb. Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerases. *The EMBO Journal*, 22(9):2234–2244, 2003.

[25] Y. Tsunaka, M. Haruki, M. Morikawa, M. Oobatake, and S. Kanaya. Dispensability of glutamic acid 48 and aparatic acid 134 for Mn2+-dependent activity of Escherichia coli ribonuclease HI. *Biochemistry*, 42:3366–3374, 2003.

[26] S. Kanaya, M. Oobatake, and Y. Liu. Thermal stability of Escherichia coli ribonuclease HI and its active site mutants in the presence and absence of Mg2+ ion - proposal of a novel catalytic role for glu48. *Journal of Biological CHemistry*, 271(51):32729–32736, 1996.

[27] M. Haruki, Y. Tsunaka, M. Morikawa, and S. Iwai, S. Kanaya. Catalysis by Escherichia coli ribonuclease HI is facilitated by a phosphate group of the substrate. *Biochemisty*, 39:13939–13944, 2000.

[28] M. Nowotny and W. Yang. Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *The EMBO Journal*, 25:1924–1933, 2006.

[29] W. Hu, W.F. Lima, H. Zhang, A. Fan, H. Sun, and S.T. Crooke. Determination of the role of the human RNase H1 in the pharmacology of DNA-like antisense drugs. *Journal of Biological Chemistry*, 279(17):17181–17189, 2004.

[30] A.L.M.A. tenAsbroek, M. vanGroenigen, M. Nooij, and F. Baas. The involvement of human ribonucleases H1 and H2 in the variation of response of cells to antisense phosphorothioate oligonucleotides. *Eour.J.Biochem.*, 269:583–592, 2002.

[31] W.F. Lima, H. Wu, J.G. Nichols, T.P. Prakash, V Ravikumar, and Crooke S.T. Human RNase H1 uses one tryptophan and two lysines to position the enzyme at the 3'-DNA/5'-RNA terminus of the heteroduplex substrate. *The Journal of Biological Chemistry*, 278:49860–49867, 2003.

[32] W.F. Lima, H. Wu, J.G. Nichols, S.M. Manalili, J.J. Drader, S.A. Hofstadler, and S.T. Crooke. Human RNase H1 activity is regulated by a unique redox switch formed between adjacent cysteines. *Journal of Biological Chemistry*, 278(17):14906–14912, 2003.

[33] H. Wu, W.F. Lima, and S.T. Crooke. Properties of cloned and expressed human RNase H1. *Journal of Biological Chemistry*, 274(40):28270–28278, 1999.

[34] P. Frank, C. Braunshoffer-Reiter, U. Wintersberger, R. Grimm, and W. Busen. Cloning of the cDNA encoding the large subunit of human RNase HI, a homologue of the prokaryotic RNase HII. *Proc. Natl. Acad. Sci. USA*, 95:12872–12877, 1998.

[35] J. Kurreck. Antisense technologies improvement through novel chemical modifications. *Eur.J.Biochem.*, 270:1628–1644, 2003.

[36] M.E. Gleave and B.P. Monia. Antisense therapy for cancer. *Nat Rev Cancer*, 5:468–479, 2005.

[37] W.F. Lima, J.G. Nichols, H. Wu, T.P. Prakash, M.T. Migawa, T.K. Wyrzykiewicz, B. Bhat, and S.T. Crooke. Structural requirements at the catalytic site of the heteroduplex substrate for human RNase H1 catalysis. *Journal of Biological Chemistry*, 279(35):36317–36326, 2004.

[38] B.P. Monia, E.A. Lesnik, C. Gonzales, W.F. Lima, D. McGee, C.J. Guinosso, A.M. Kawasaki, P.D. Cook, and S.M. Freier. Evaluation of 2'-modified oligonucleotides containing 2'-deoxy gaps as antisense inhibitors of gene expression. *Journal of Biological Chemistry*, 268(19):14514–14522, 1993.

[39] B.P. Monia, J.F. Johnston, H. Sasmot, and L.L Cummins. Nuclease resistance and antisense activity of modified oligonucleotides targeted to Ha-ras. *Journal of Biological Chemistry*, 271(24):14533–14540, 1996.

[40] C.M. Yamada, D.J. Dellinger, and M.H. Caruthers. Synthesis and biochemical evaluation of phosphonoformate oligodeoxyribonucleotides. *J.Am.Chem.Soc.*, 128:5251–5261, 2006.

[41] D.J. Dellinger, D.M. Sheehan, N.K. Christsen, J.G. Lindberg, and M.H. Caruthers. Solid-phase chemical synthesis of phosphonoacetate and thiophosphonoacetate oligodeoxynucleotides. *J.Am.Chem.Soc.*, 125:940–950, 2003.

[42] W.F. Lima and S.T. Crooke. Binding affinity and specificity of Escherichia coli RNase H1: Impact on the kinetics of catalysis of antisense oligonucleotide-RNA hybrids. *Biochemistry*, 36:390–398, 1997.

[43] W.F. Lima, J.B. Rose, J.G. Nichols, H. Wu, M.T. Migawa, T.K. Wyrzykiewicz, A.M. Siwkowski, and S.T. Crooke. Human RNase H1 discriminates between subtle variations in the structure of the heteroduplex substrate. *Molecular Pharmacology*, 71:83–91, 2007.

[44] P.I. Pradeepkumat, E. Zamaratski, A. Foldesi, and Chattopadhyaya. ransmission of the conformational information in the antisense/RNA hybrid duplex influences the pattern of the RNase H cleavage reaction. *Tetrahedron Letters*, 41:8601–8607, 2000.

[45] W.F. Lima, J.B. Rose, J.G. Nichols, H. Wu, M.T. Migawa, T.K. Wyrzykiewicz, G. Vasquez, E.E Swayze, and S.T. Crooke. The positional influence of the helical geometry of the heteroduplex substrate on human RNase H1 catalysis. *Molecular Pharmacology*, 71:73–82, 2007.

[46] V.K. Rait and B.R. Shaw. Boranophosphates support the RNase H cleavage of polyribonucleotides. *Antisense & Nucleic Acid Drug Development*, 9:53–60, 1999.

[47] X. Wang, M. Dobrikov, D. Sergueev, and B. R. Shaw. RNase H activation by stereoregular boranophosphate oligonucleotide. *Nucleosides, Nucleotides & Nucleic Acids*, 22(5-8):1151–1153, 2003.

[48] M.M. Mangos, K.L. Min, E. Viazovkina, A. Galarneu, M.I. Elzagheid, M.A. Parniak, and M.J Damha. Efficient RNase H-directed cleavage of RNA promoted by antisense DNA of 2'F-ANA constructs containing acyclic nucleotide inserts. *J.Am.Chem.Soc.*, 154:654–661, 2003.

[49] E. Zamaratski, P.I. Pradeepkumar, and J. Chattopadhyaya. A critical survey of the structure-function of the antisense oligo/RNA heteroduplex as substrate of RNase H. *J.Biochem.Biophys.Methods*, 48:189–208, 2001.

[50] D. Rejman, J. Snášel, R. Liboska, Z. Točík, O. Pačes, Š. Králíková, M. Rinnová, P. Koiš, and I. Rosenberg. Oligonucleotides with isopolar phosphonate internucleotide linkage: a new perspective for antisense compounds? *Nucleosides, nucleotides and nucleic acids*, 20(4-7):819–823, 2001.

[51] F. Autenrieth, B. Isralewitz, Z. Luthey-Schulten, A. Sethi, and T. Pogorelov. *Bioinformatics and sequence alignment.*

[52] http://de.wikipedia.org/wiki/lennard-jones-potential.

[53] http://amber.scripps.edu/.

[54] http://www.ccl.net/cca/documents/molecular-modeling/perio.gif.

[55] P. Jungwirth. *Klasická a kvantová molekulová dynamika.* lecture notes.

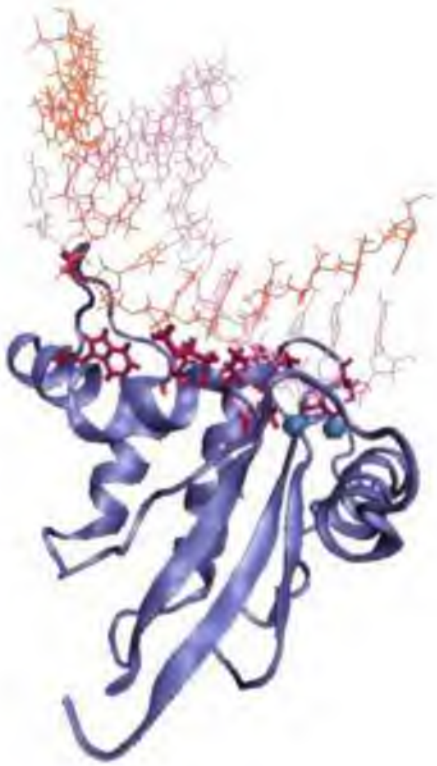
[56] H. Wu, W.F. Lima, and S.T. Crooke. Investigating the structure of human RNase H1 by site-directed mutagenesis. *Journal of Biological Chemistry*, 276(26):23547–23553, 2001.

[57] S.E. McDowell, N. Špačková, Šponer J., and N.G. Walter. Review - molecular dynamics simulation of RNA: An In Silico single molecule approach. *Biopolymers*, 85(2):169–184, 2006.

Appendix A

Graphs - human RNase H with an unmodified substrate: models 1-6

Model refinement I: enzyme-substrate recognition



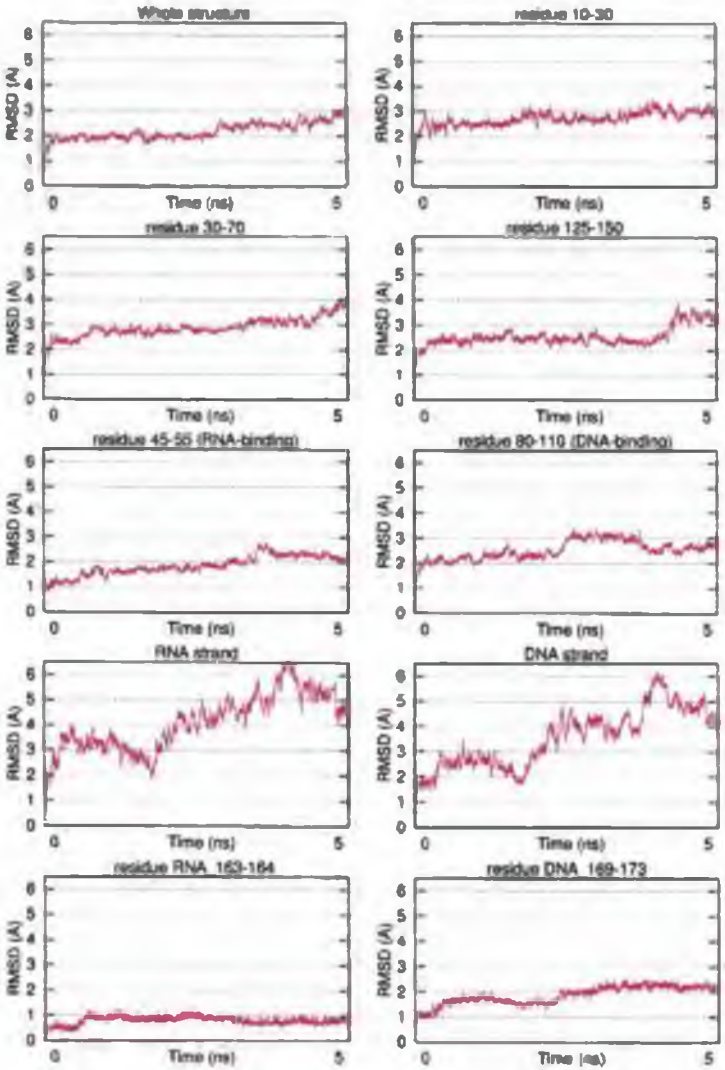


Figure A.1: RMSD of the whole structure and selected regions. Model 1

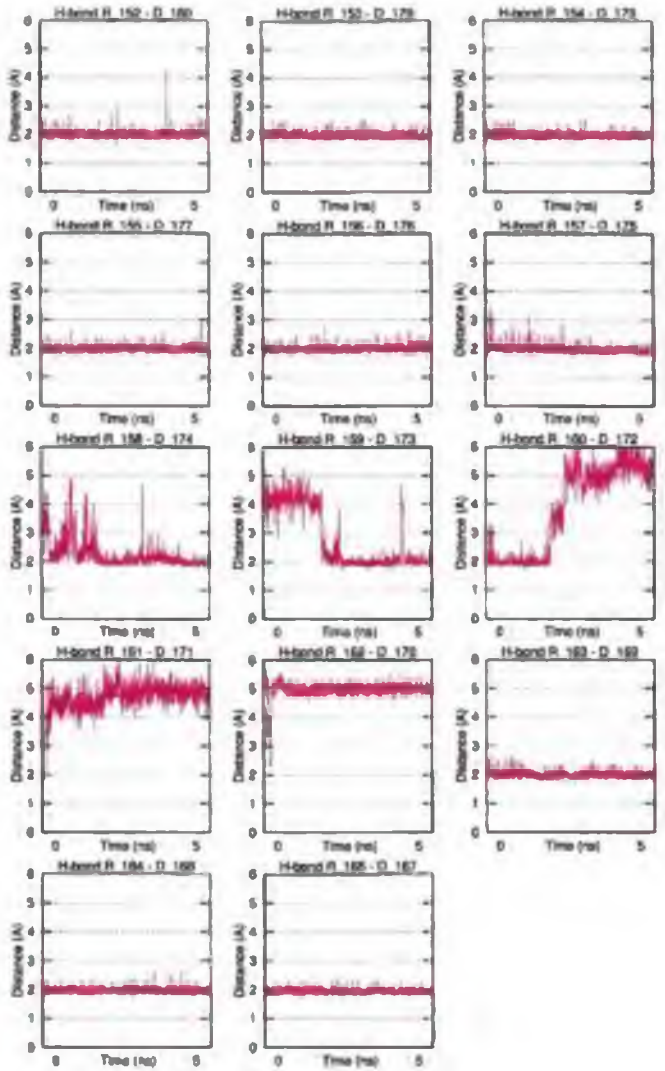


Figure A.2: The Watson-Crick hydrogen bonds. Model 1

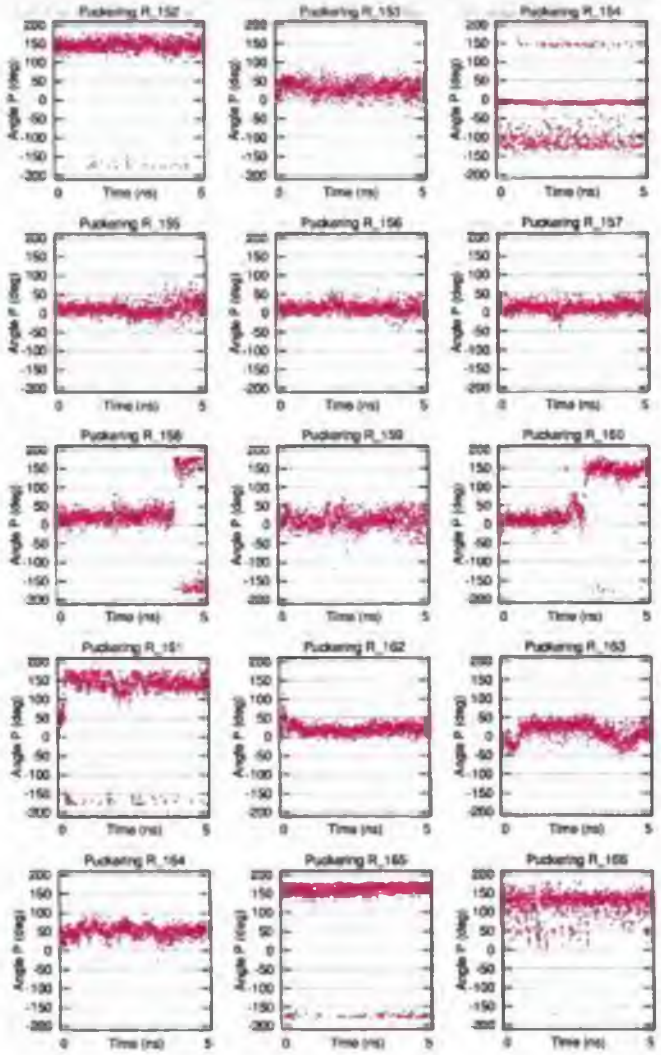


Figure A.3: Puckering of the RNA-strand. Model 1

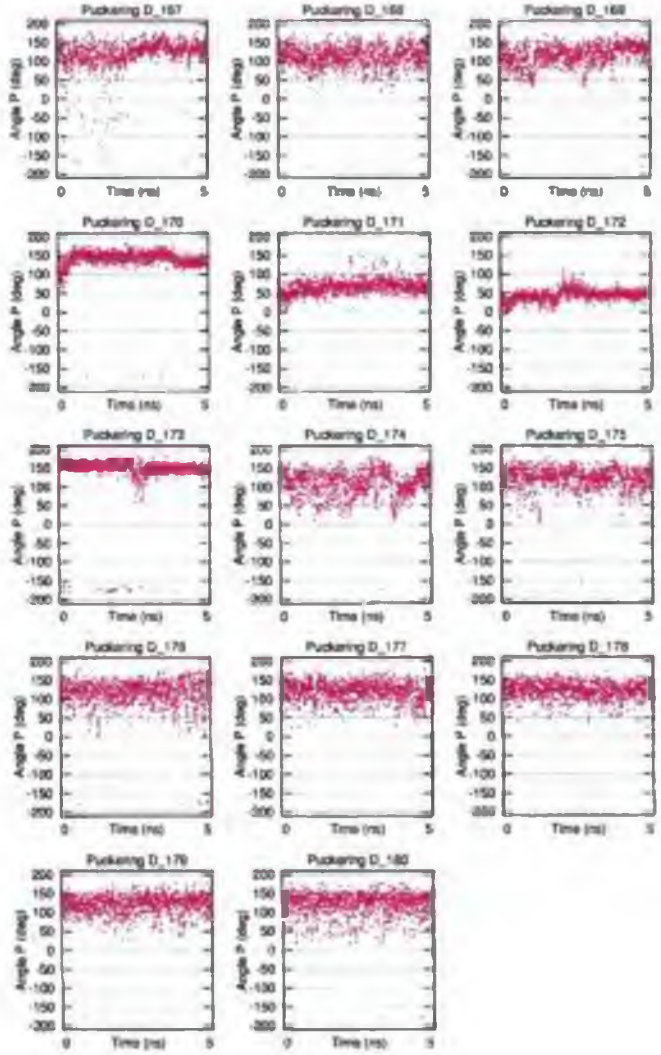


Figure A.4: Puckering of the DNA-strand. Model 1

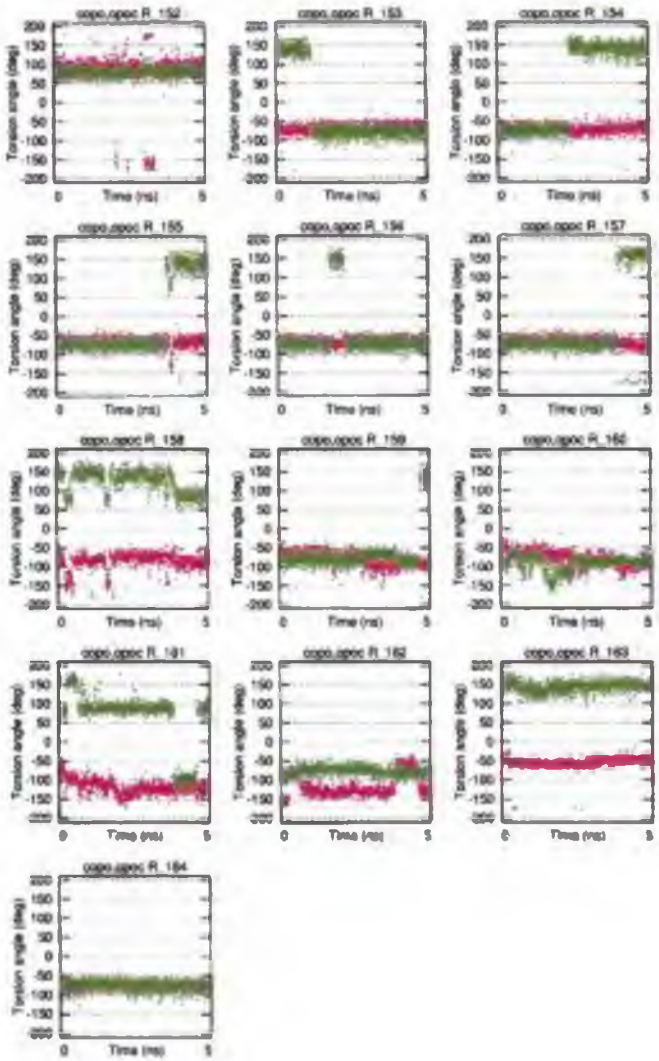


Figure A.5: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 1

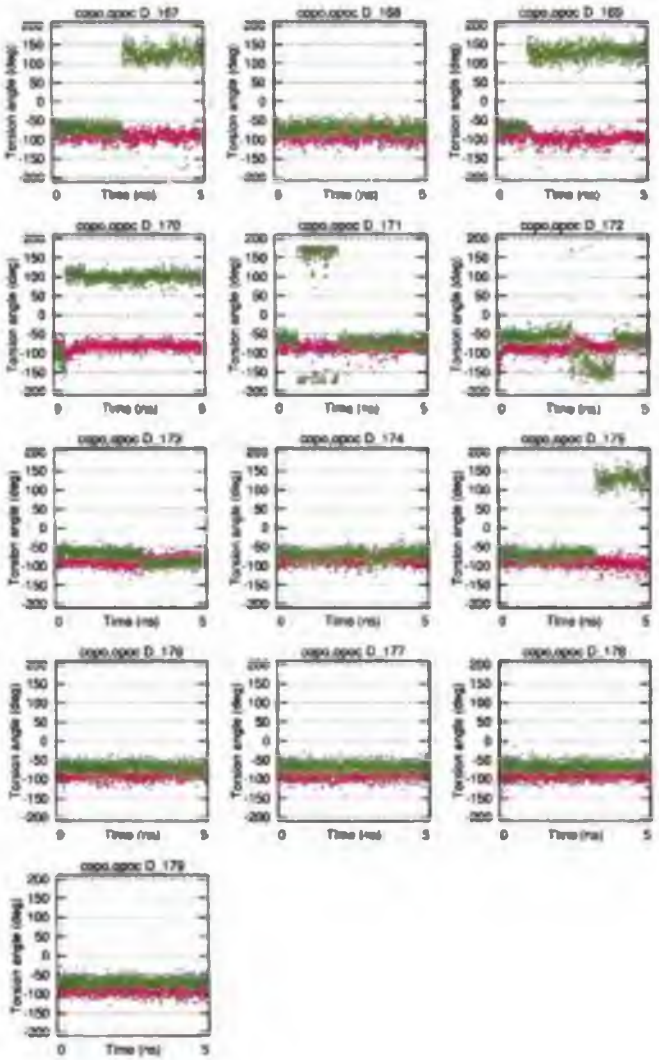


Figure A.6: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 1

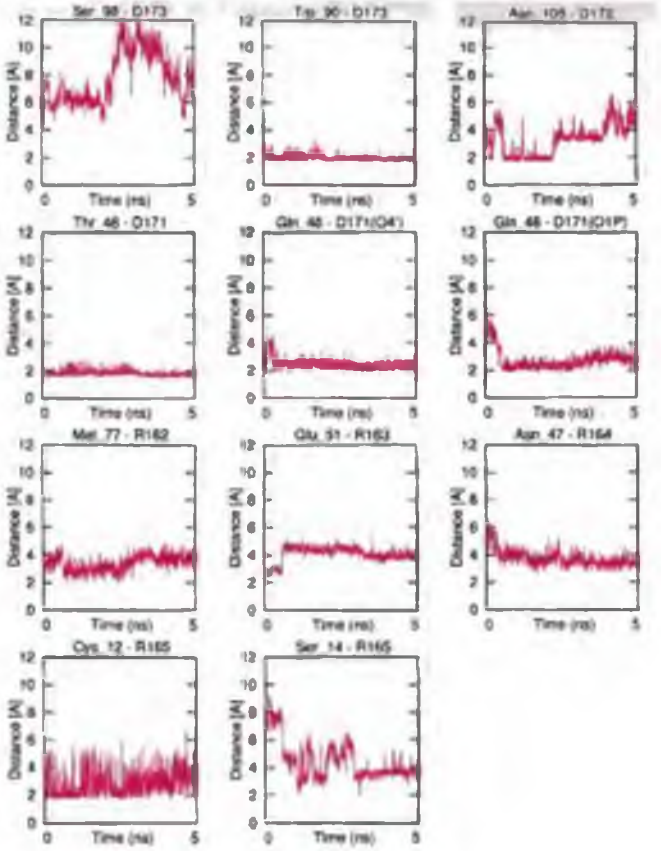


Figure A.7: The enzyme-substrate hydrogen bonds. Model 1

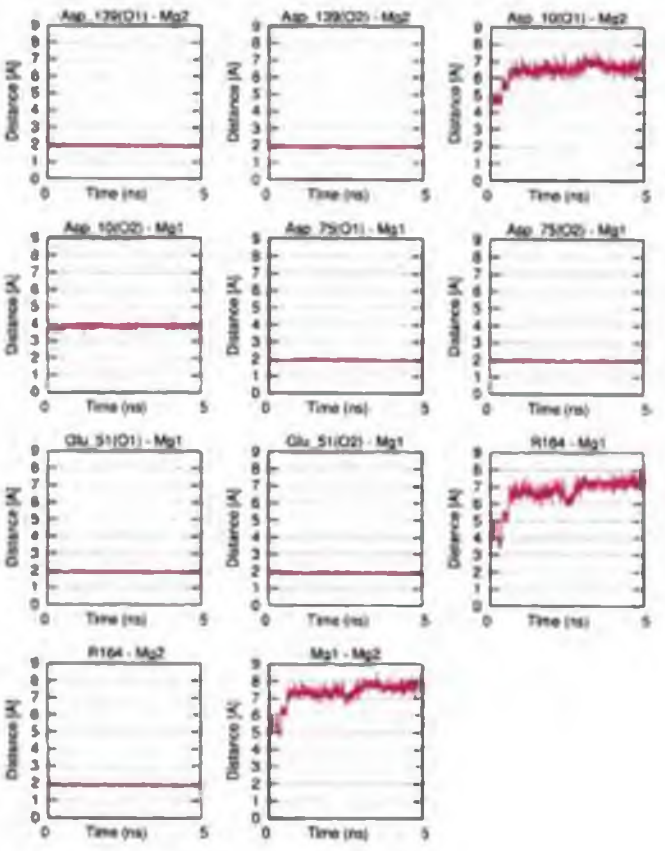


Figure A.8: Contacts between the Mg-ions and the DDE motif in the active site. Model 1

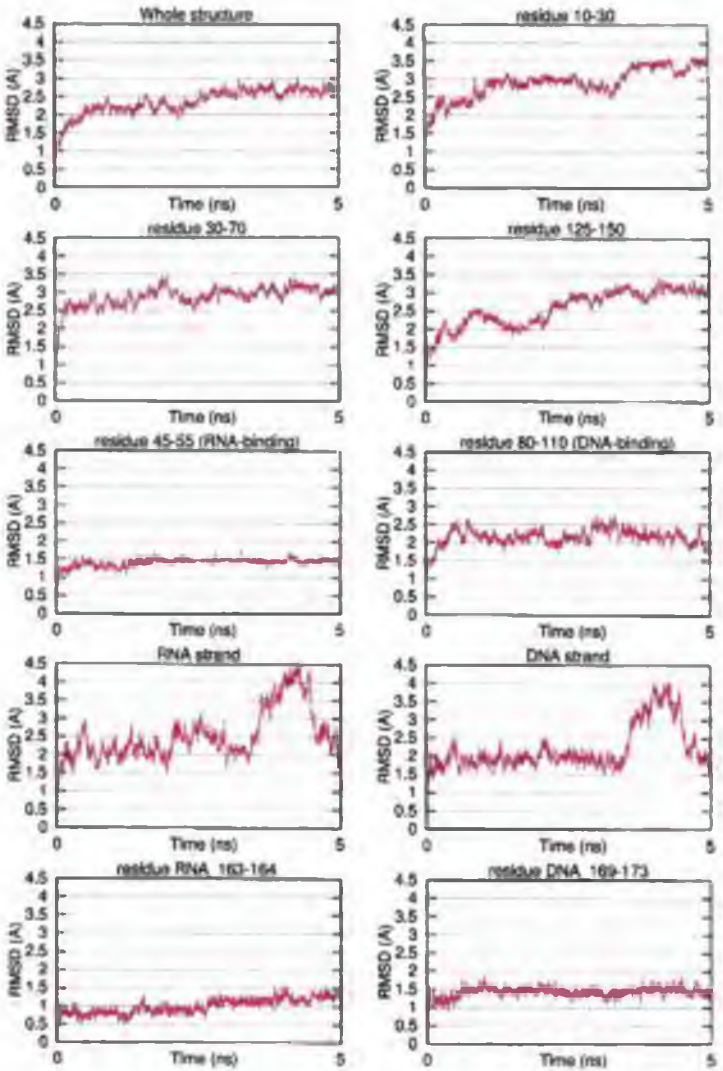


Figure A.9: RMSD of the whole structure and selected regions. Model 2

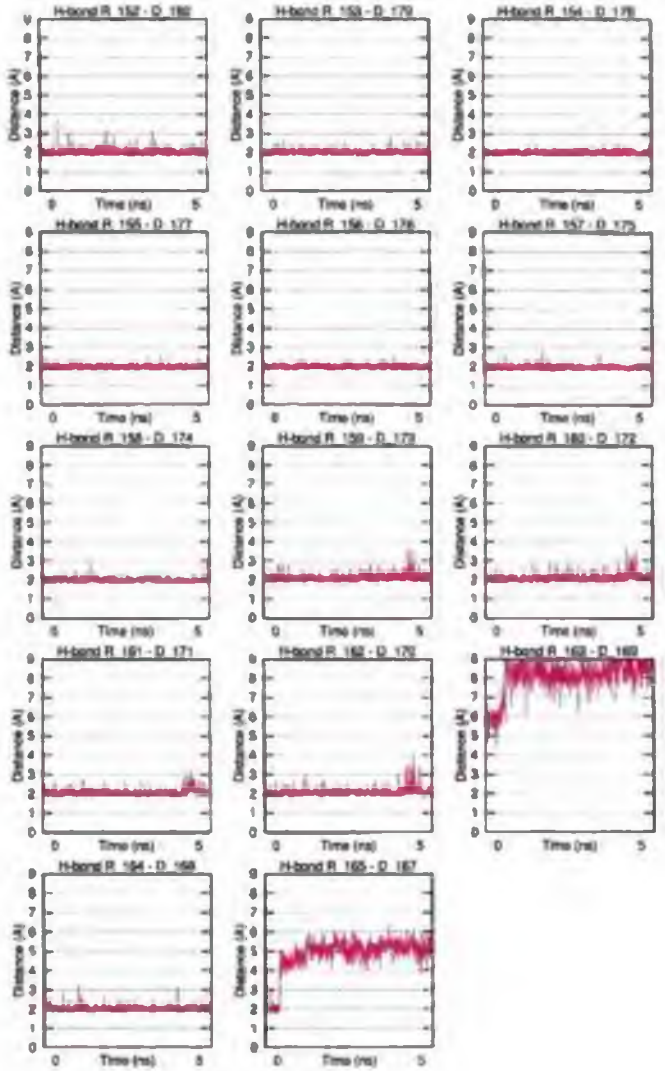


Figure A.10: The Watson-Crick hydrogen bonds. Model 2

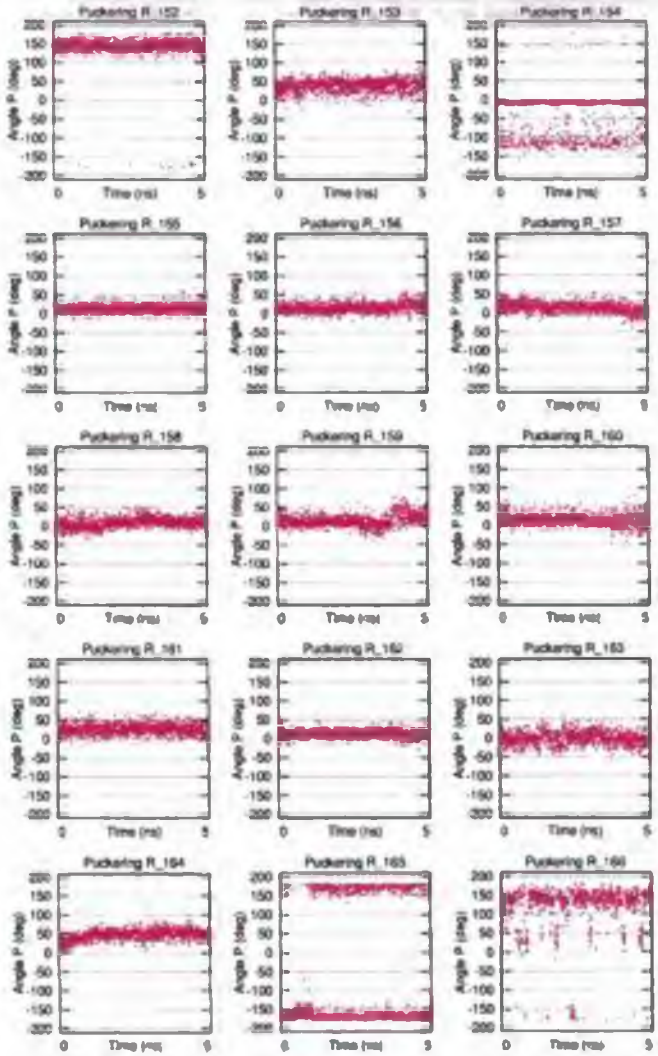


Figure A.11: Puckering of the RNA-strand. Model 2

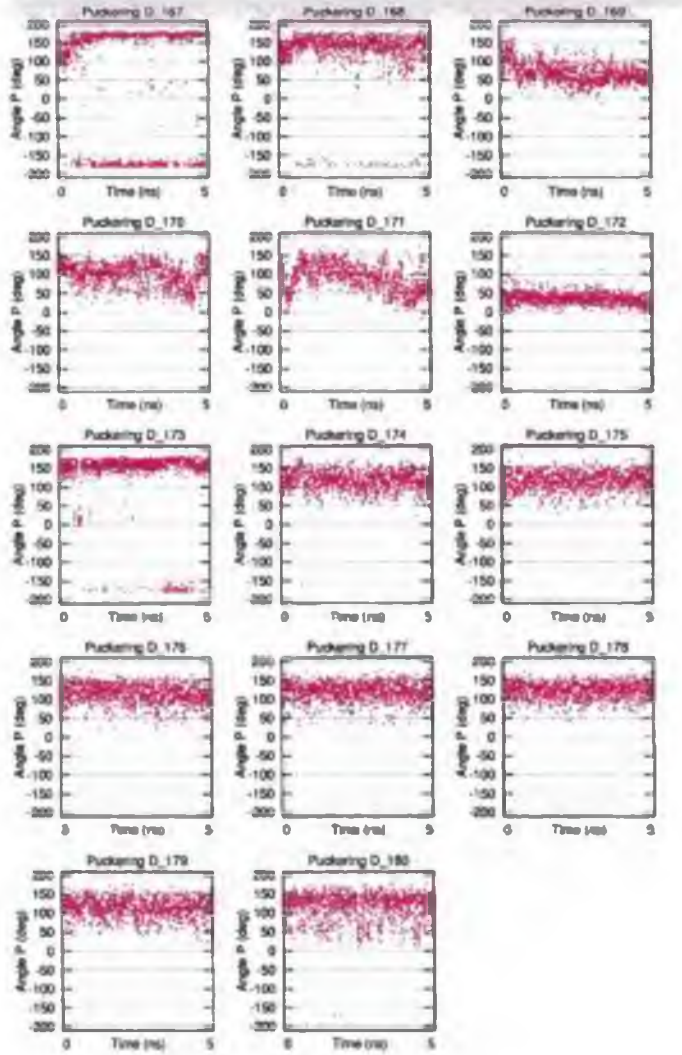


Figure A.12: Puckering of the DNA-strand. Model 2

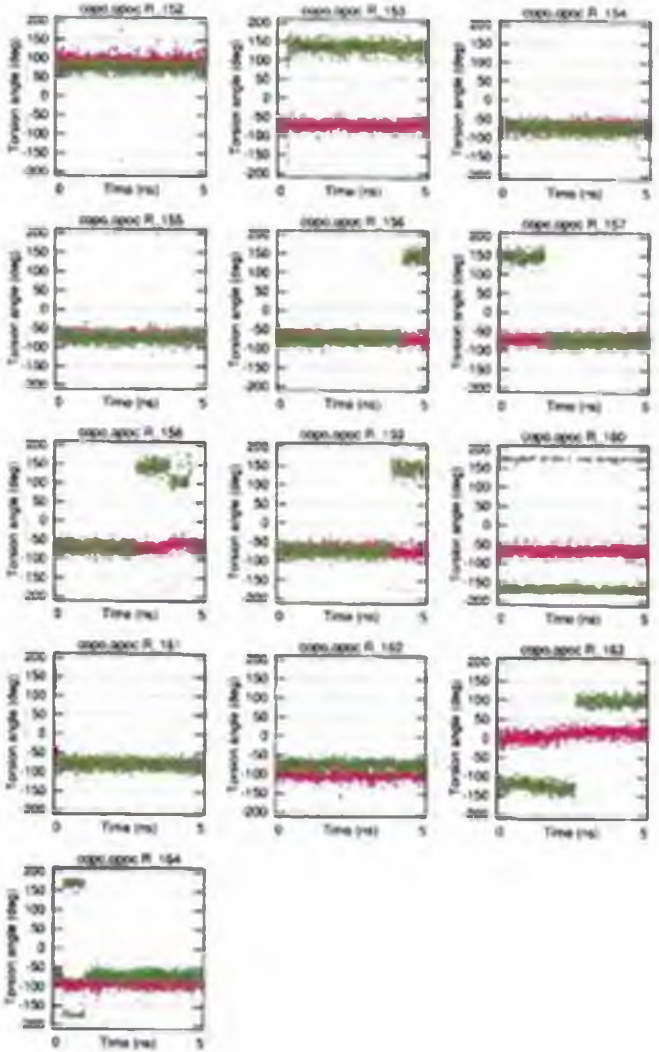


Figure A.13: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 2

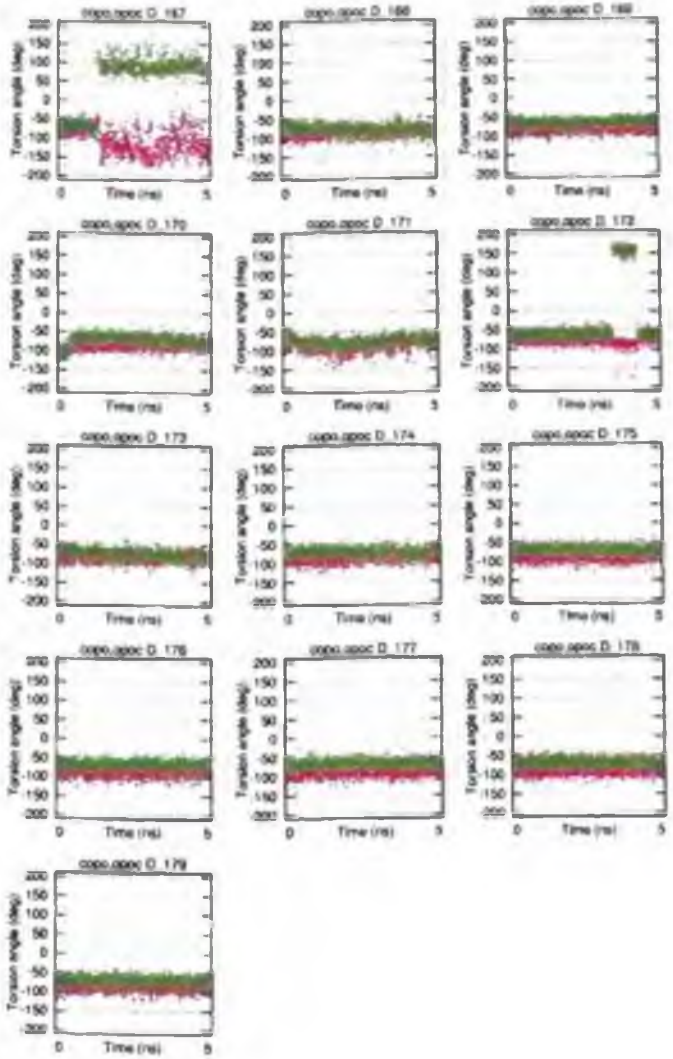


Figure A.14: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 2

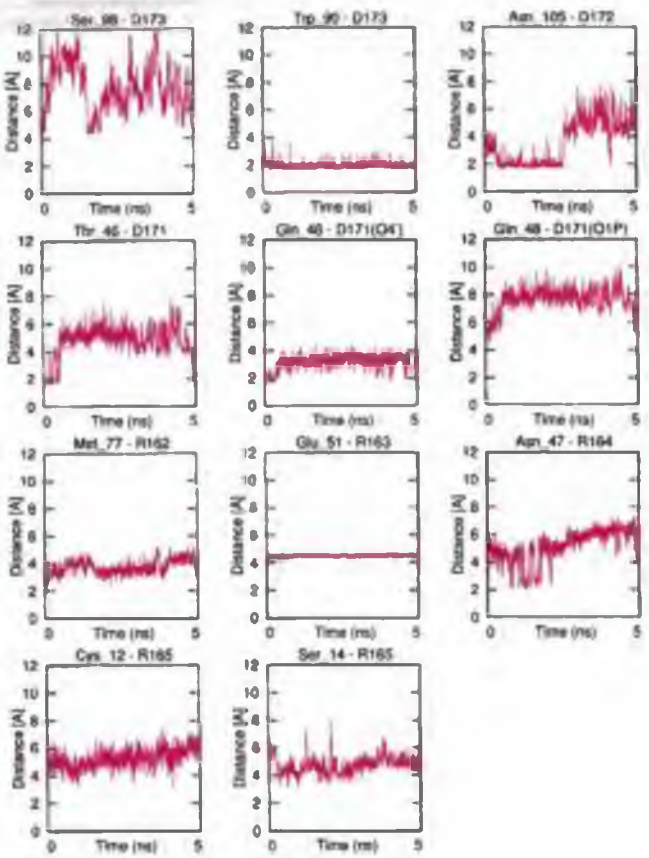


Figure A.15: The enzyme-substrate hydrogen bonds. Model 2

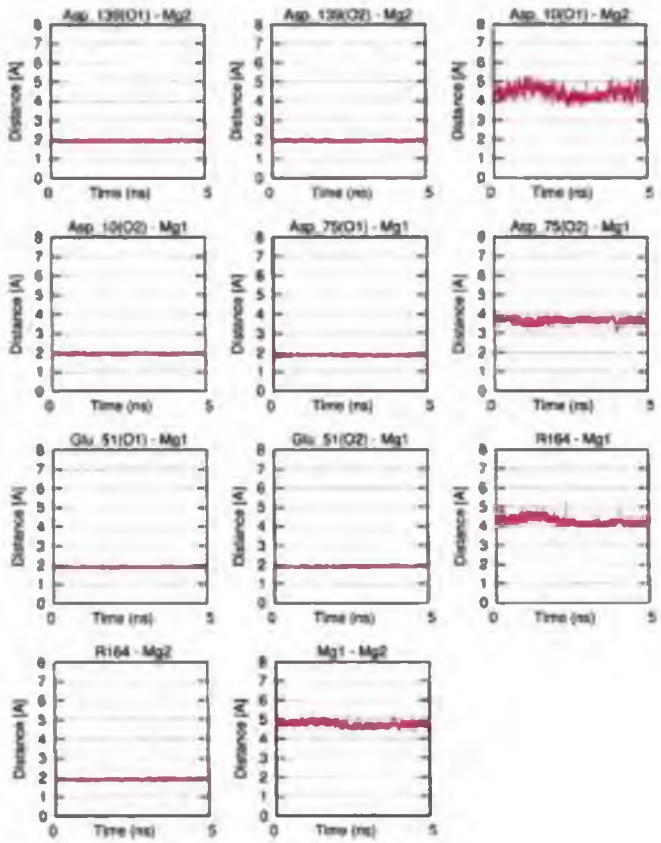


Figure A.16: Contacts between the Mg-ions and the DDE motif in the active site. Model 2

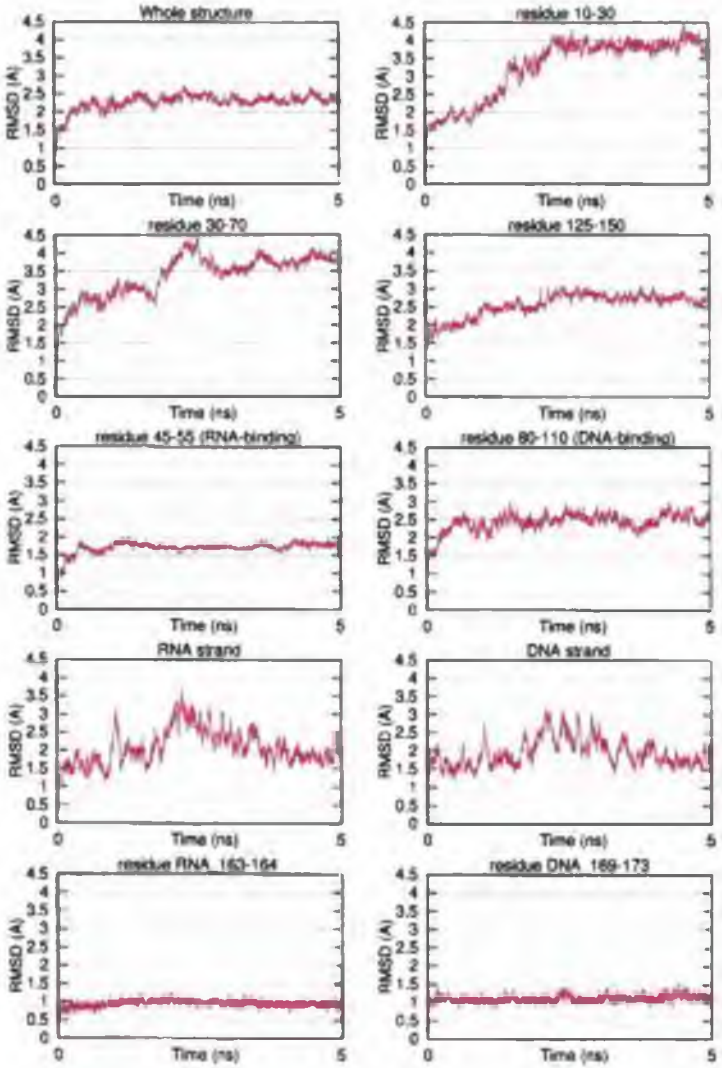


Figure A.17: RMSD of the whole structure and selected regions. Model 3

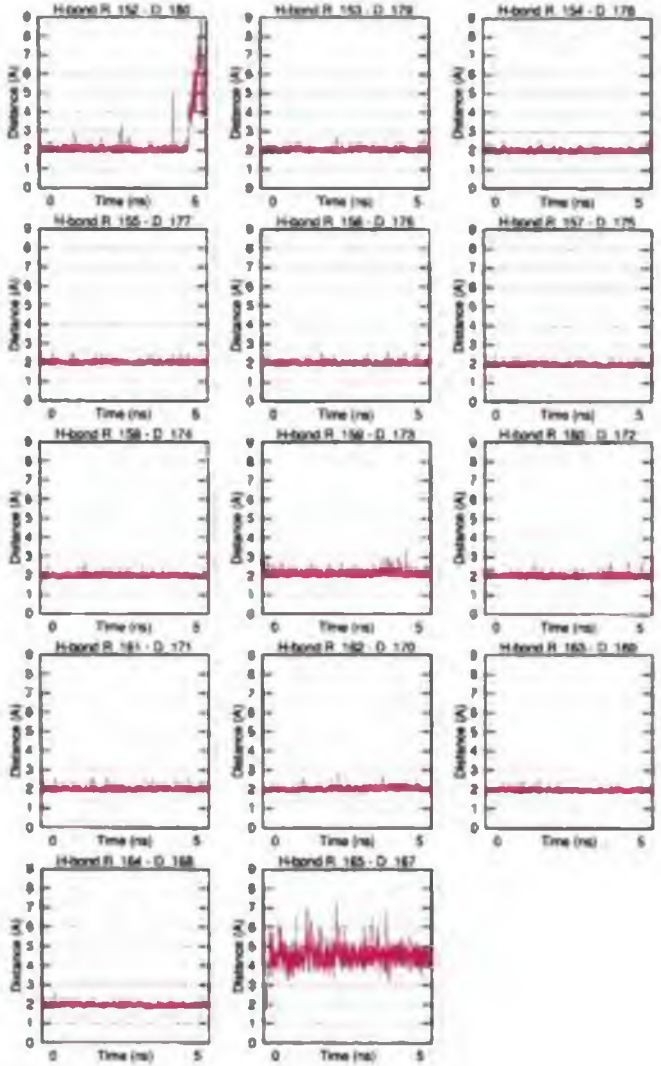


Figure A.18: The Watson-Crick hydrogen bonds. Model 3

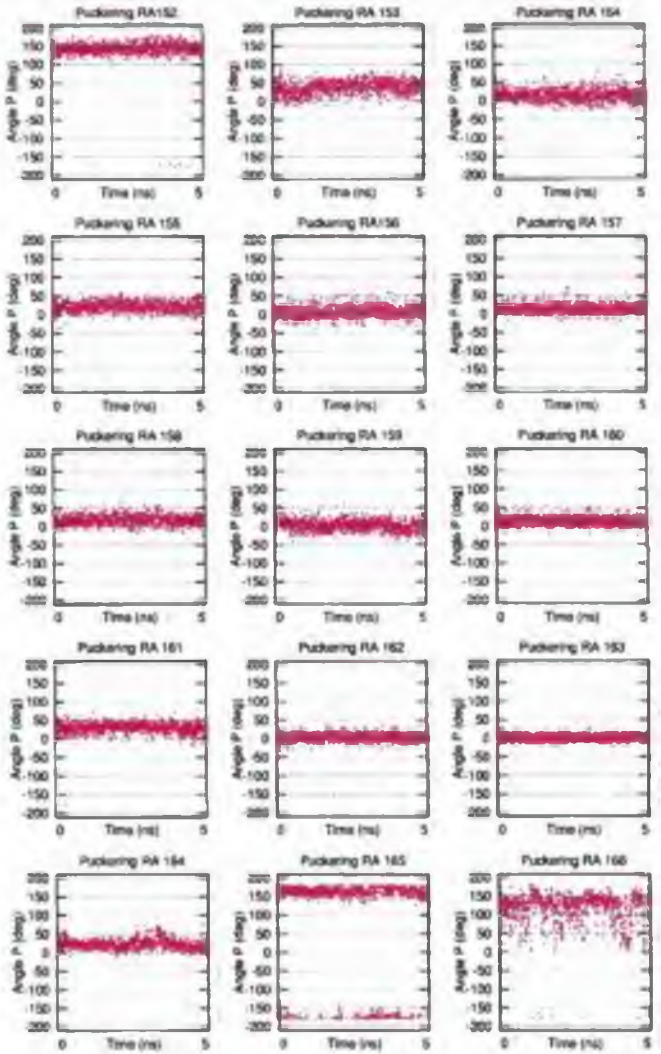


Figure A.19: Puckering of the RNA-strand. Model 3

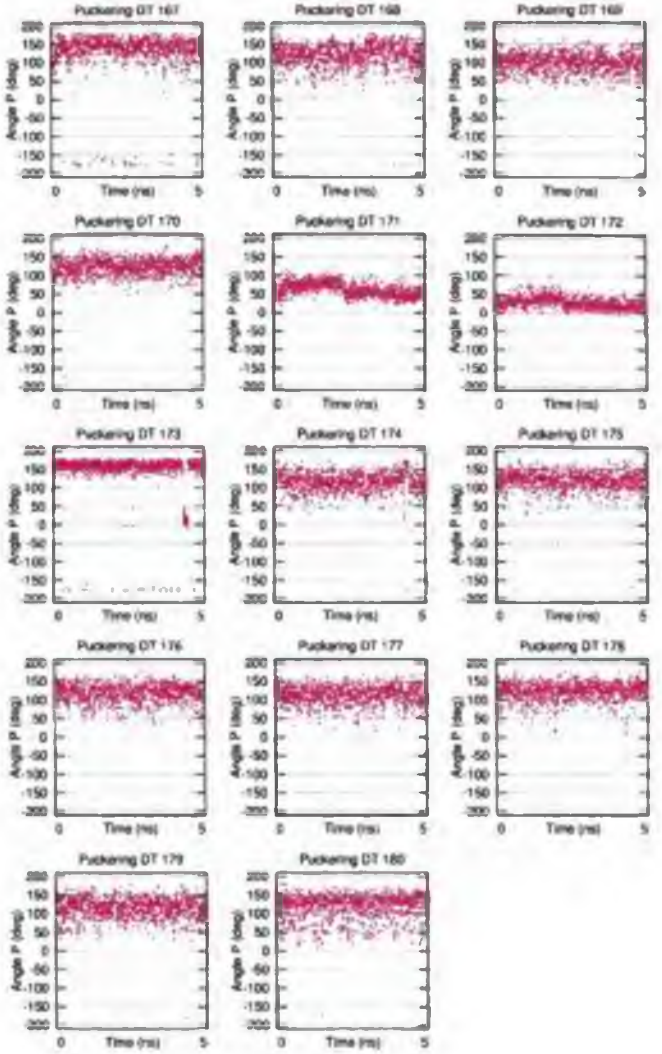


Figure A.20: Puckering of the DNA-strand. Model 3

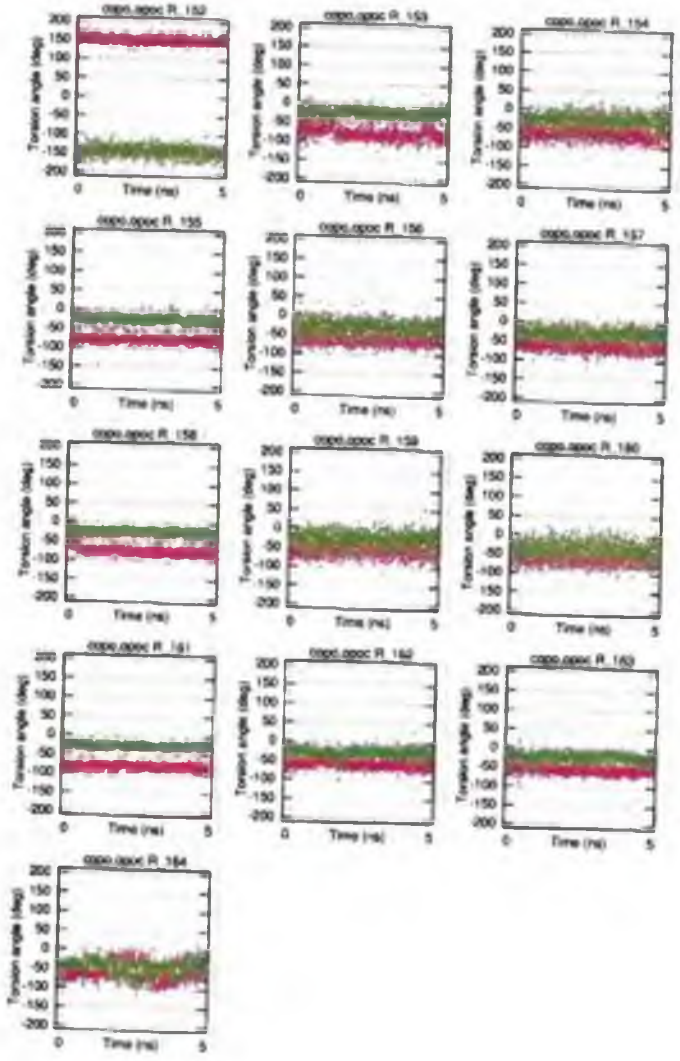


Figure A.21: The C3'-O3'-P-O5' (red) and O3'-P-O3'-C5' (green) torsion angles, the RNA-strand. Model 3

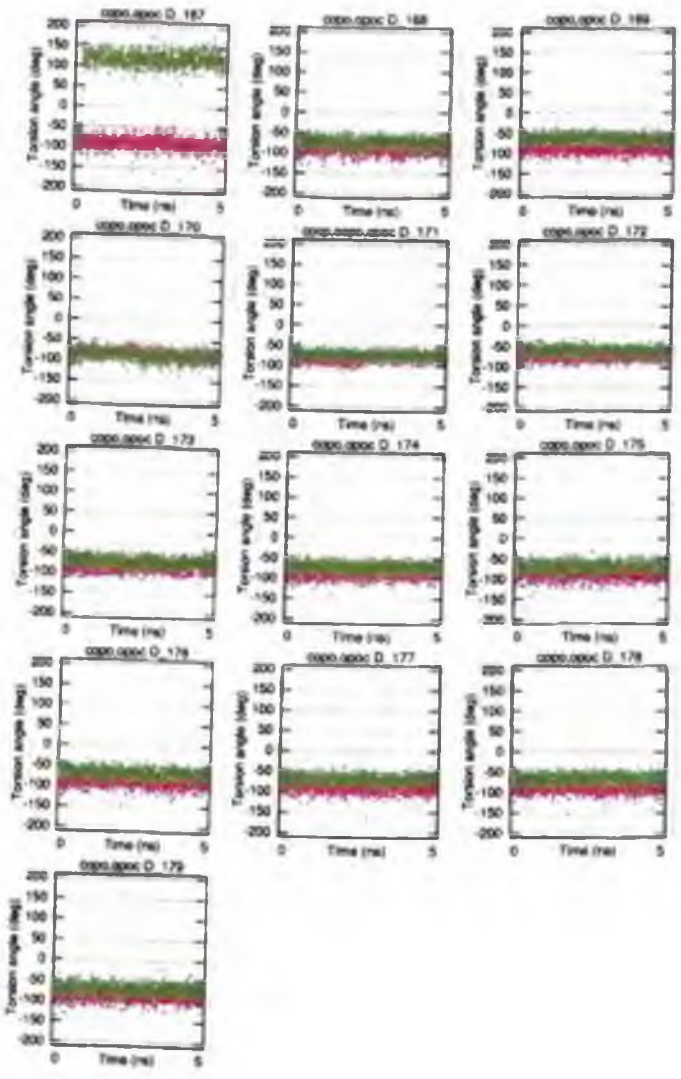


Figure A.22: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 3

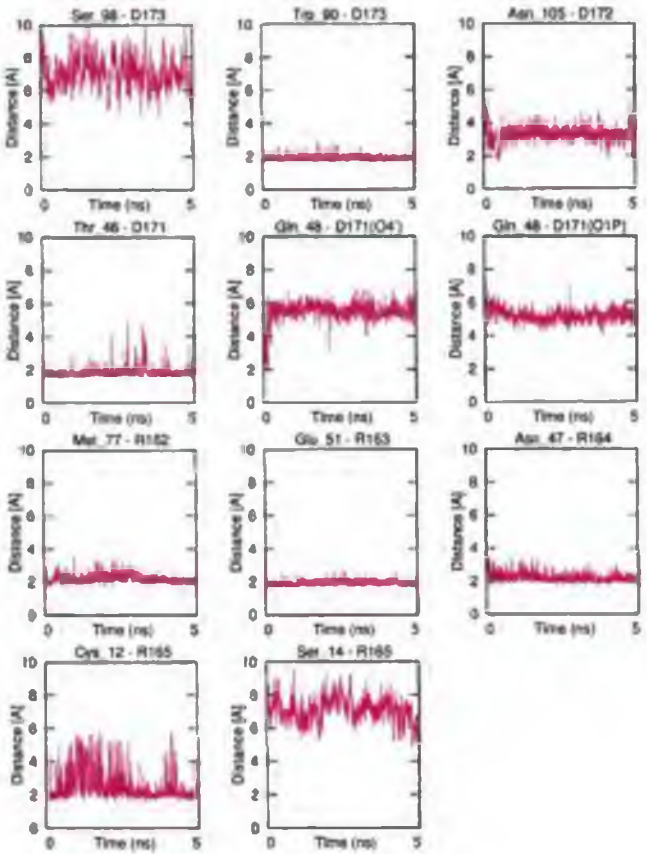


Figure A.23: The enzyme-substrate hydrogen bonds. Model 3

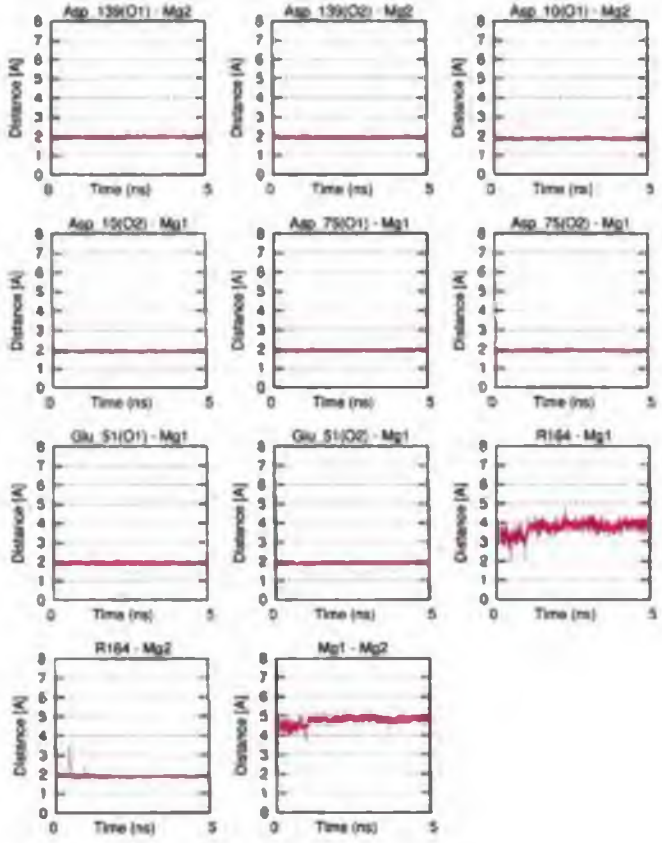


Figure A.24: Contacts between the Mg-ions and the DDE motif in the active site. Model 3

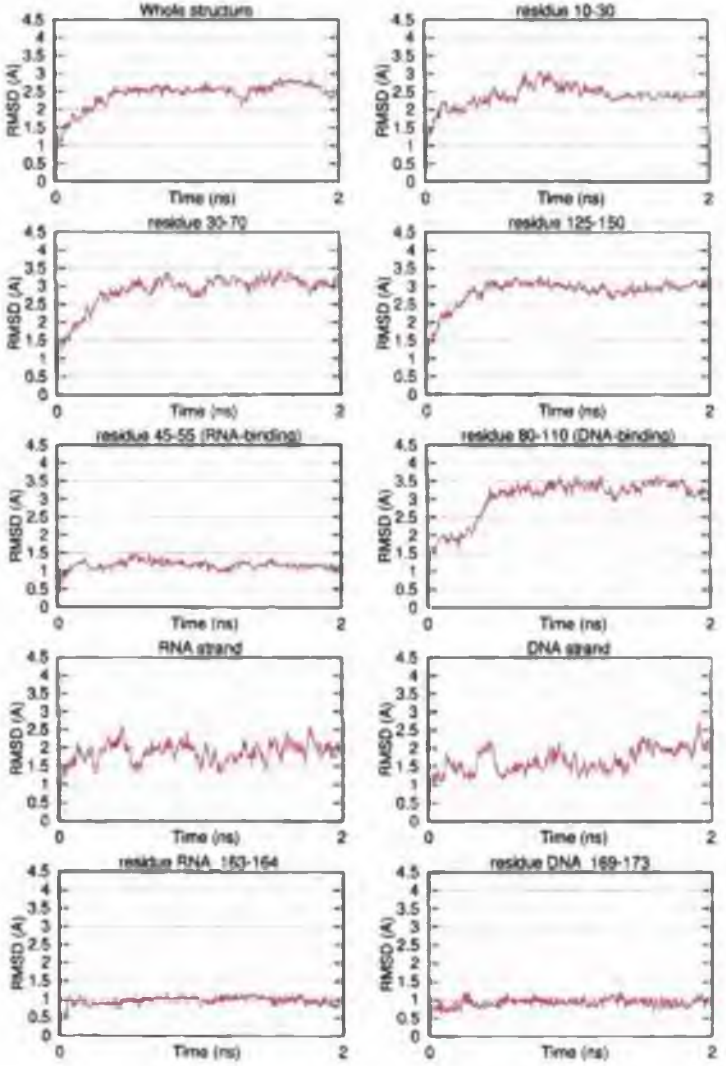


Figure A.25: RMSD of the whole structure and selected regions. Model 4

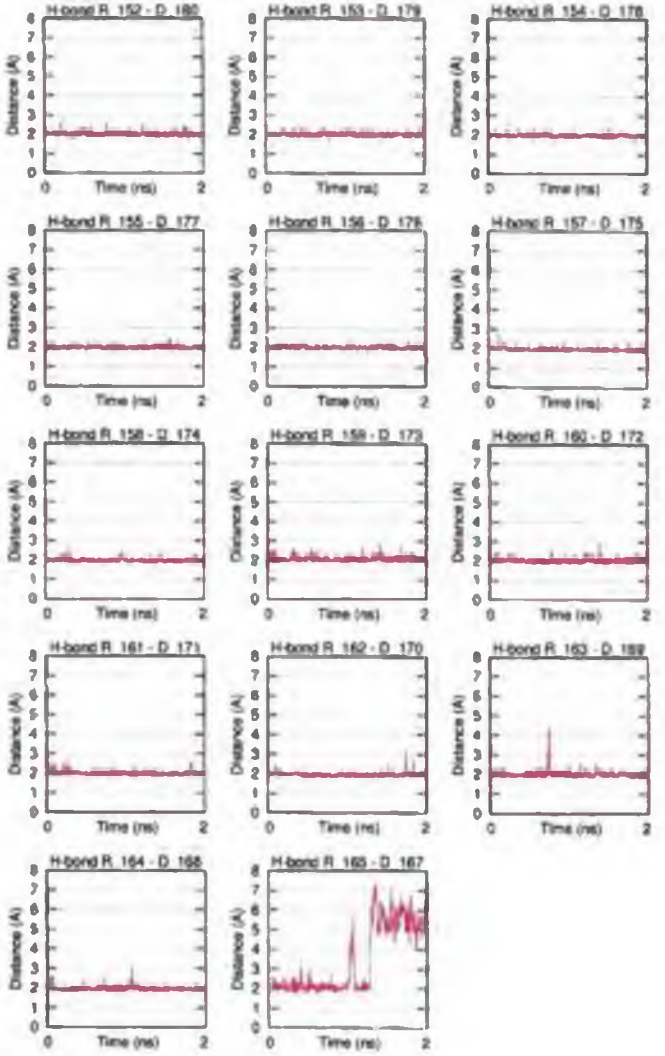


Figure A.26: The Watson-Crick hydrogen bonds. Model 4

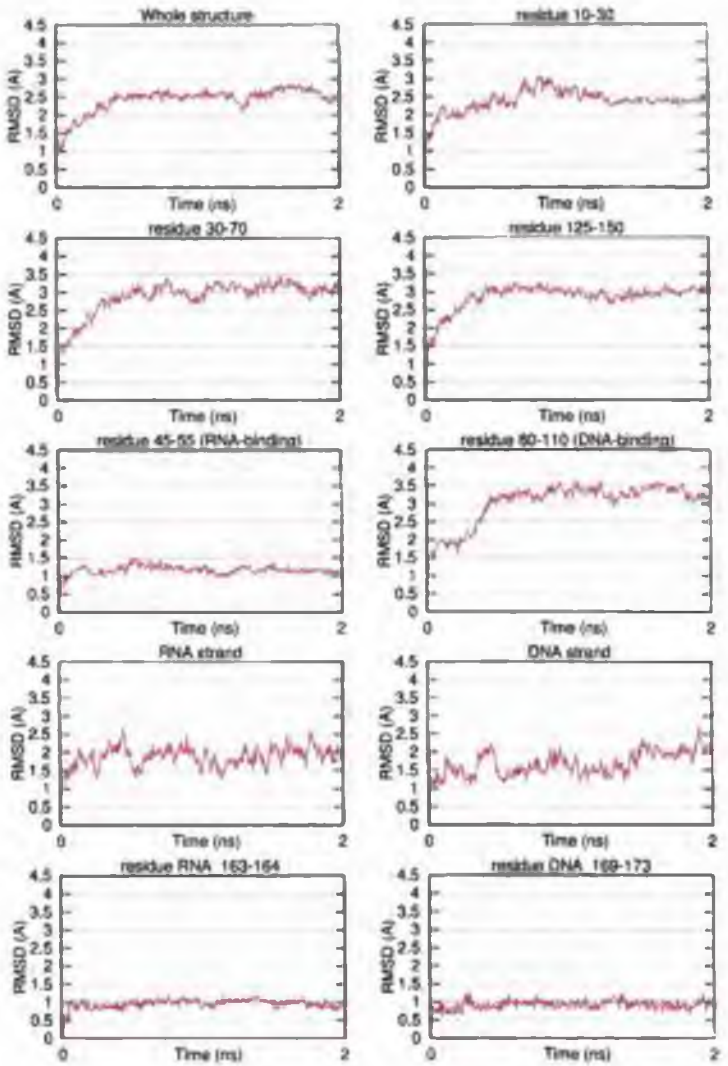


Figure A.27: Puckering of the RNA-strand. Model 4

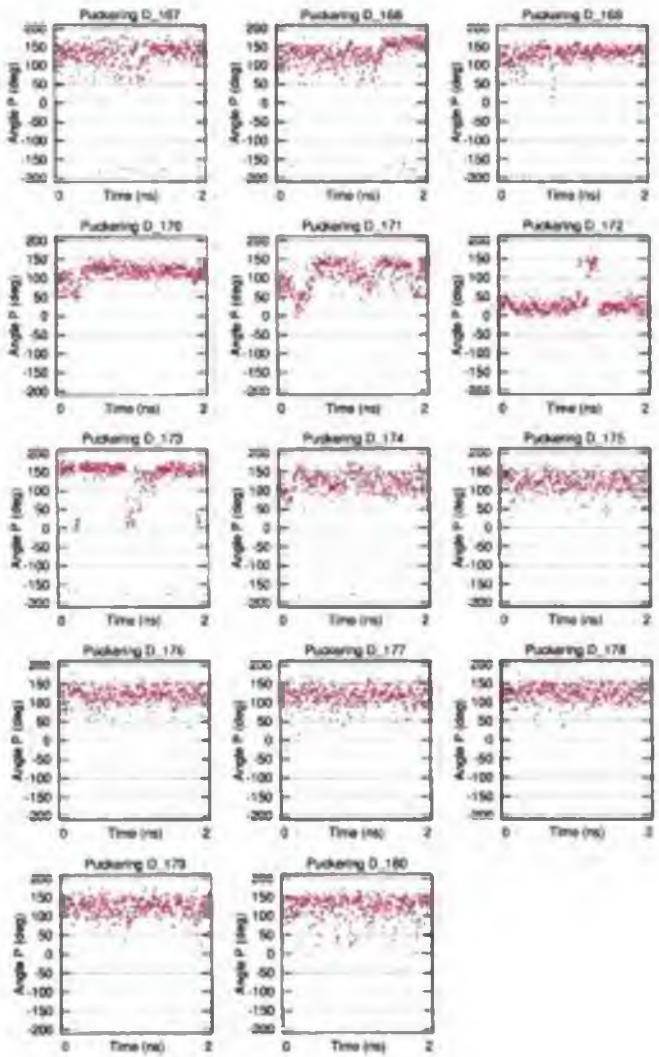


Figure A.28: Puckering of the DNA-strand. Model 4

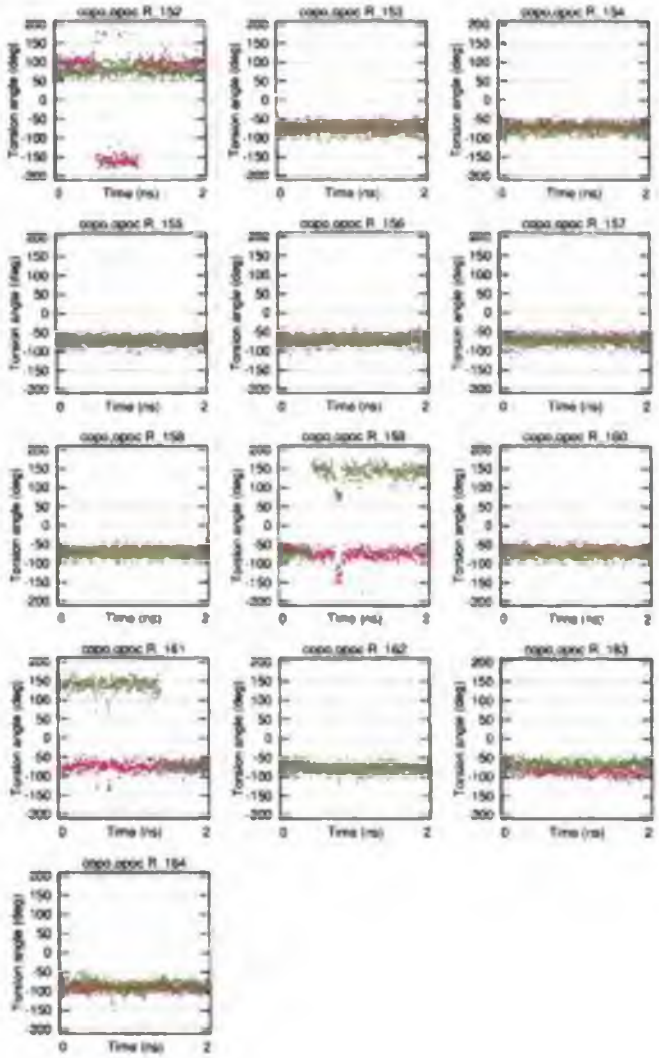


Figure A.29: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 4

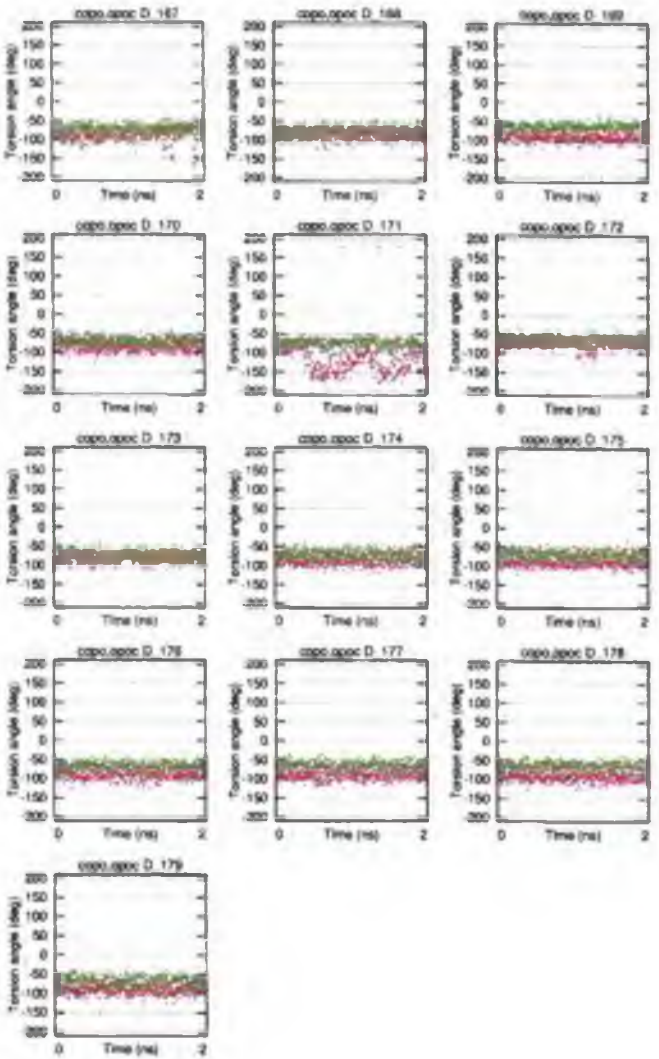


Figure A.30: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 4

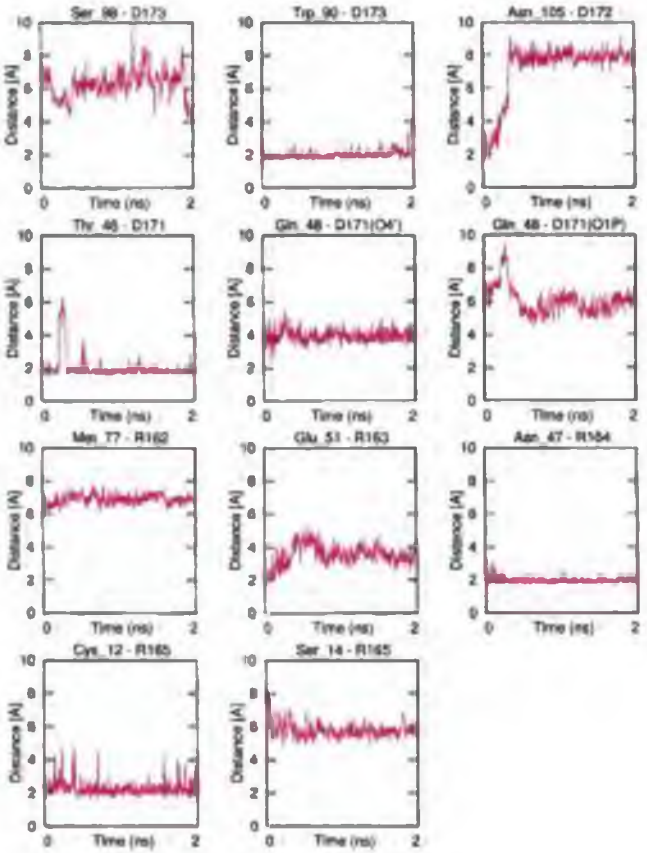


Figure A.31: The enzyme-substrate hydrogen bonds. Model 4

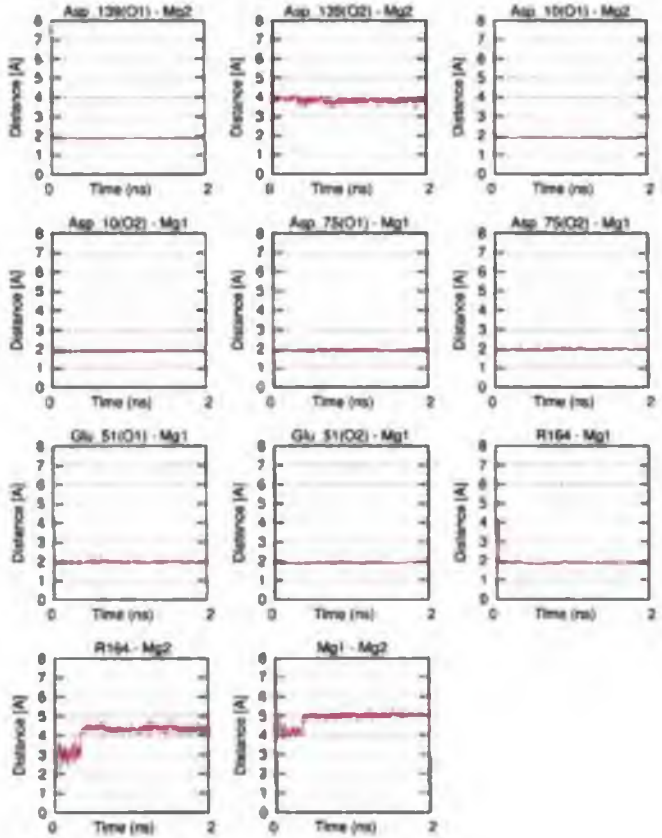


Figure A.32: Contacts between the Mg-ions and the DDE motif in the active site. Model 4

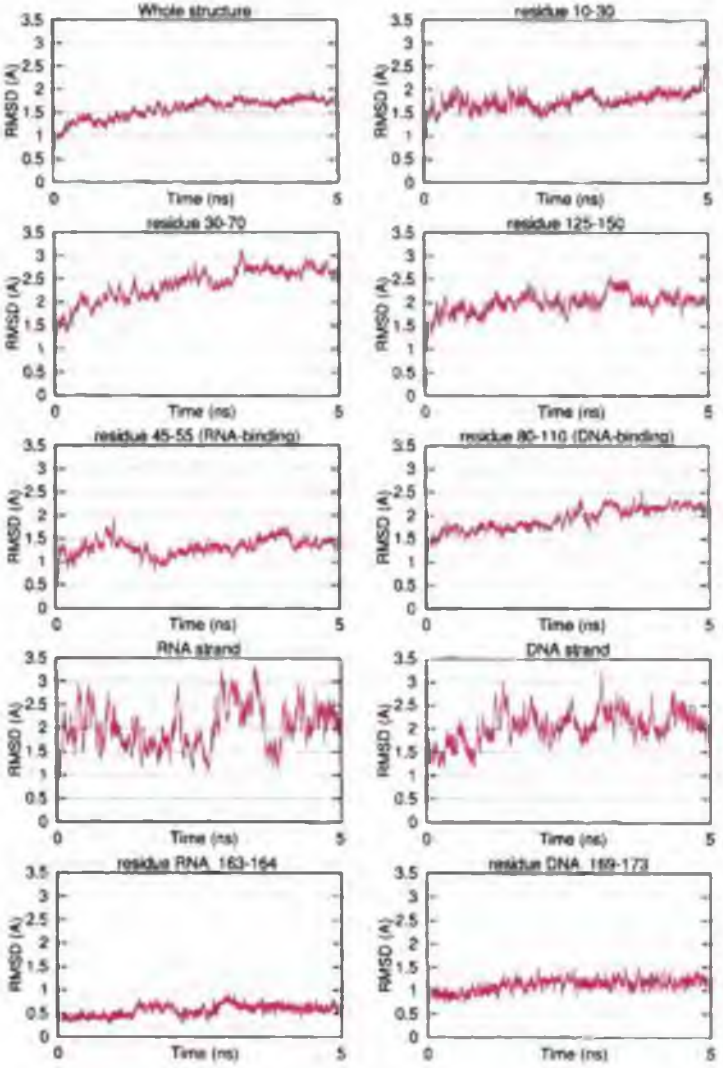


Figure A.33: RMSD of the whole structure and selected regions. Model 5

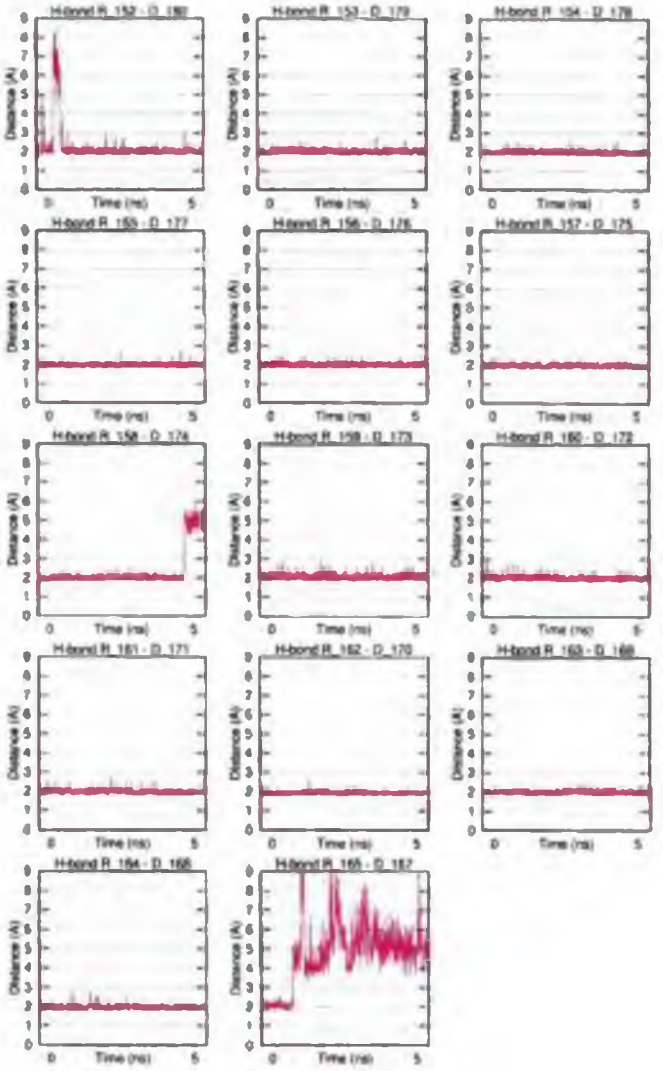


Figure A.34: The Watson-Crick hydrogen bonds. Model 5

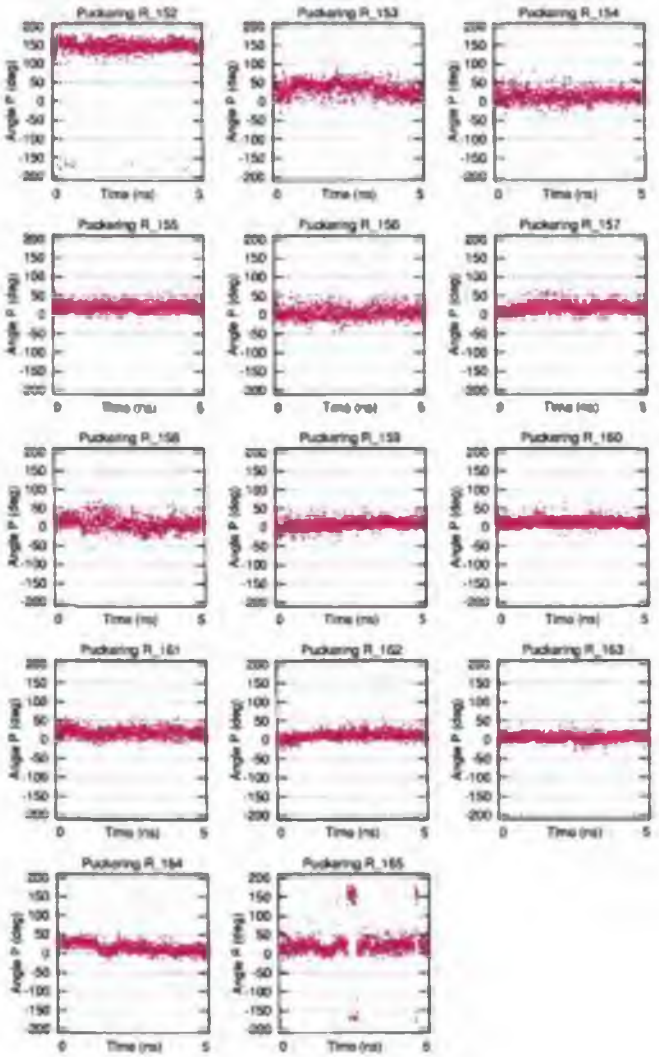


Figure A.35: Puckering of the RNA-strand. Model 5

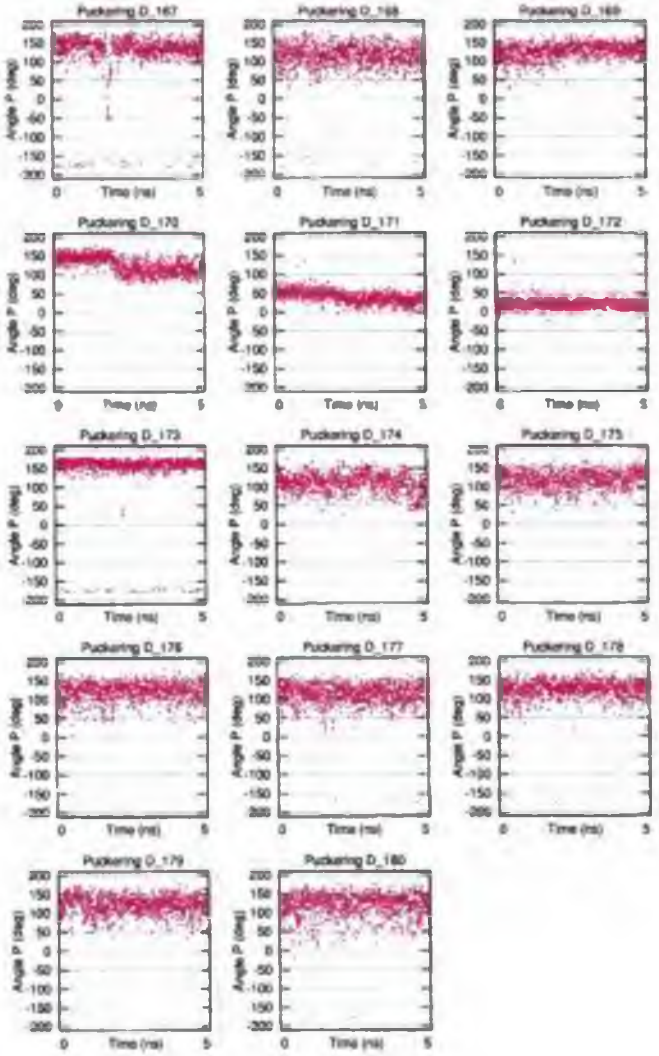


Figure A.36: Puckering of the DNA-strand. Model 5

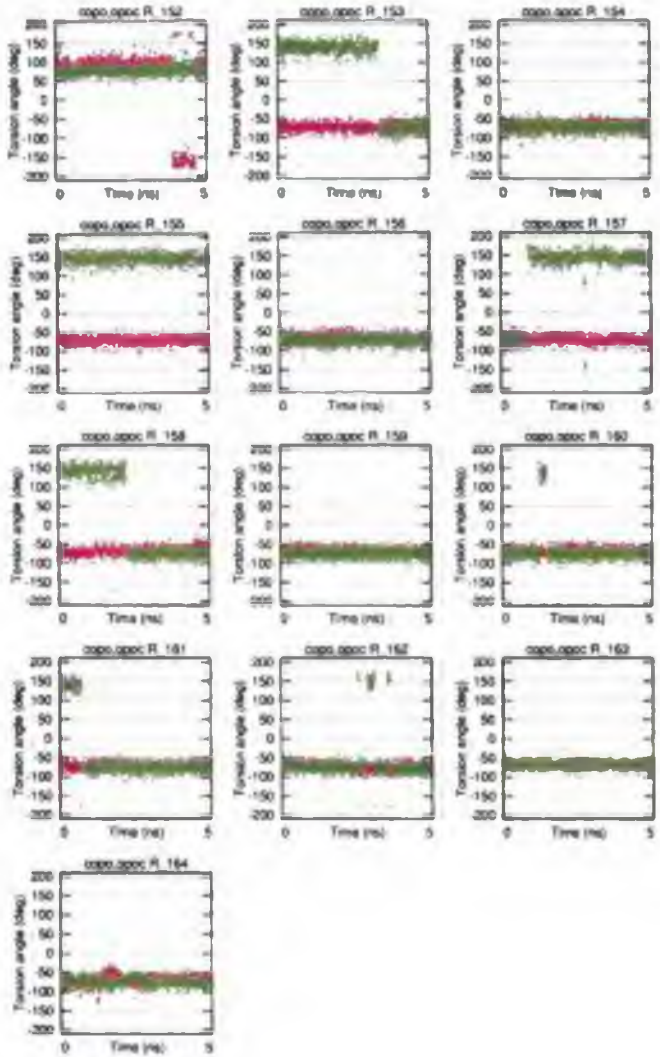


Figure A.37: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 5

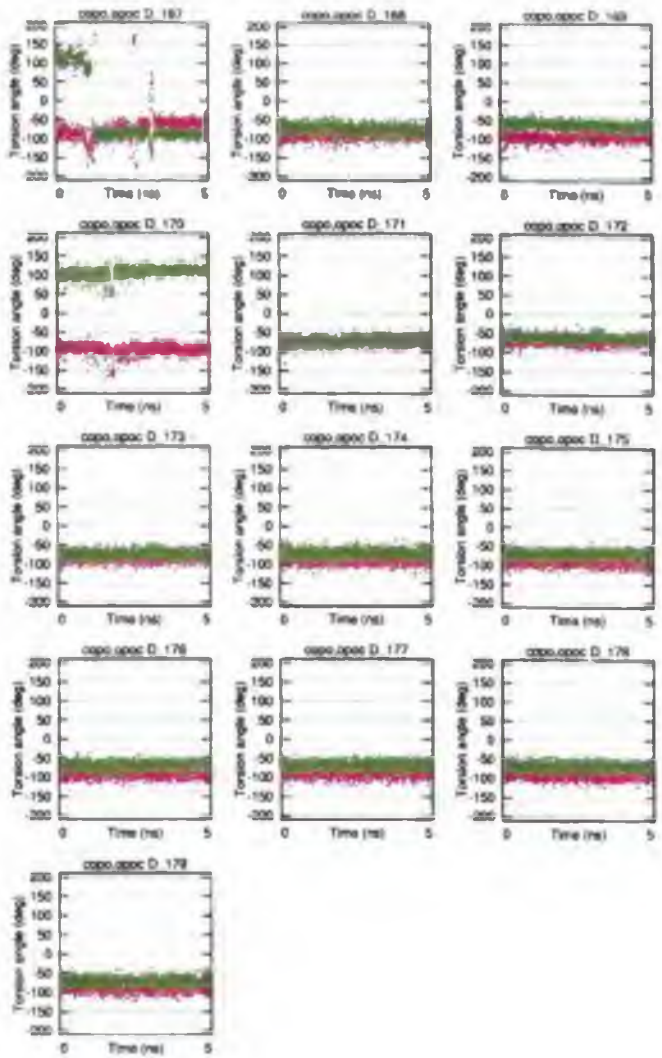


Figure A.38: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 5

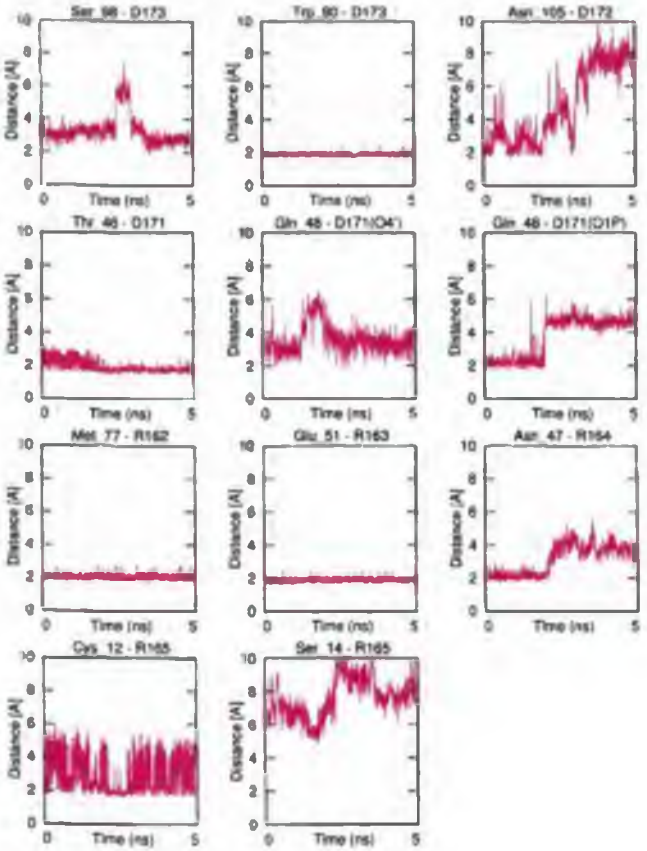


Figure A.39: The enzyme-substrate hydrogen bonds. Model 5

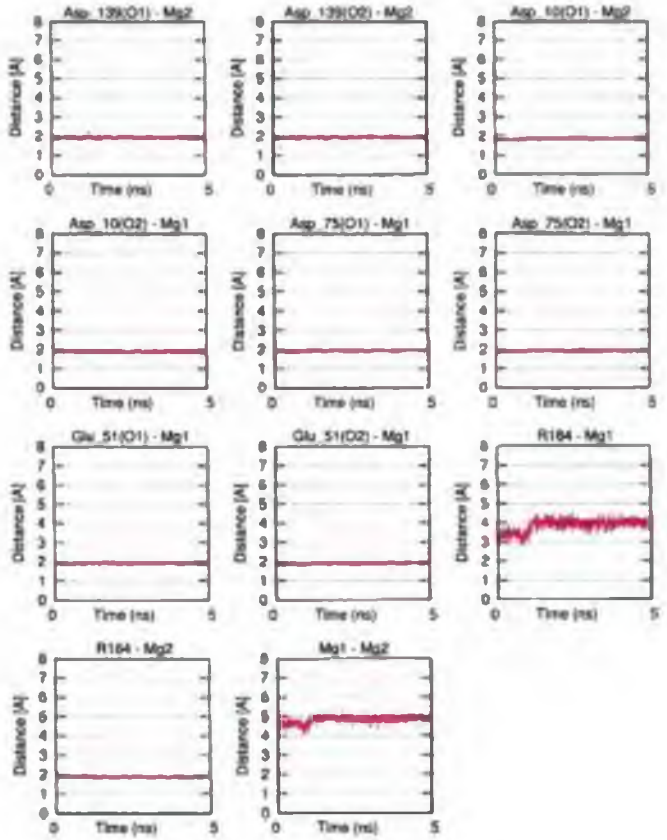


Figure A.40: Contacts between the Mg-ions and the DDE motif in the active site. Model 5

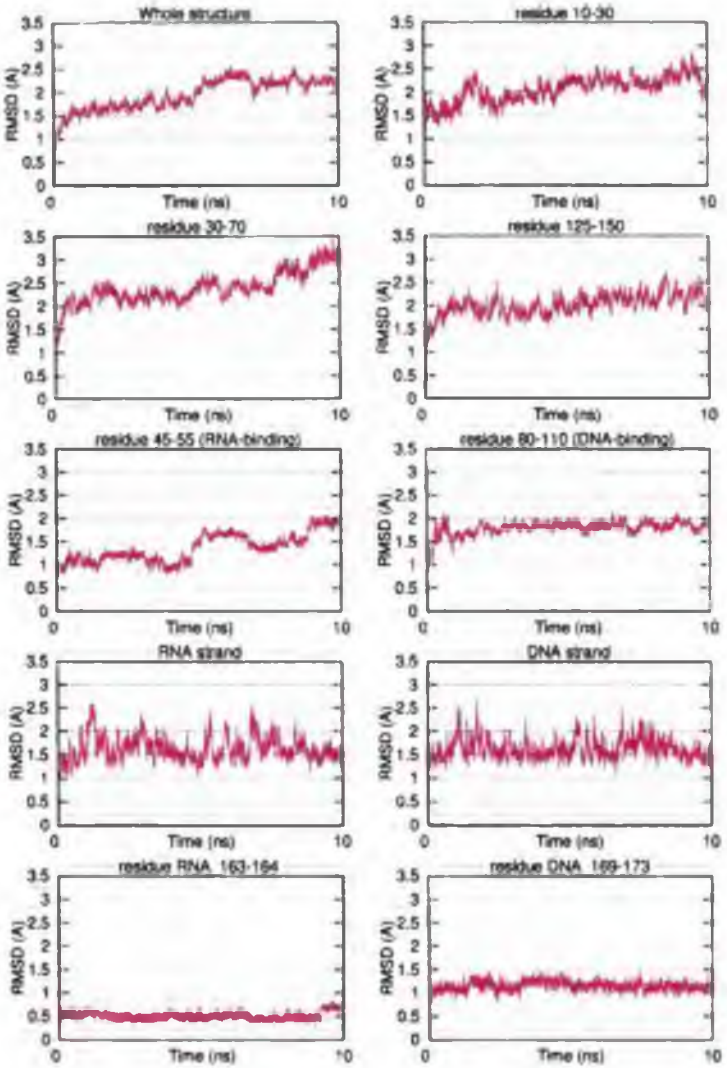


Figure A.41: RMSD of the whole structure and selected regions. Model 6

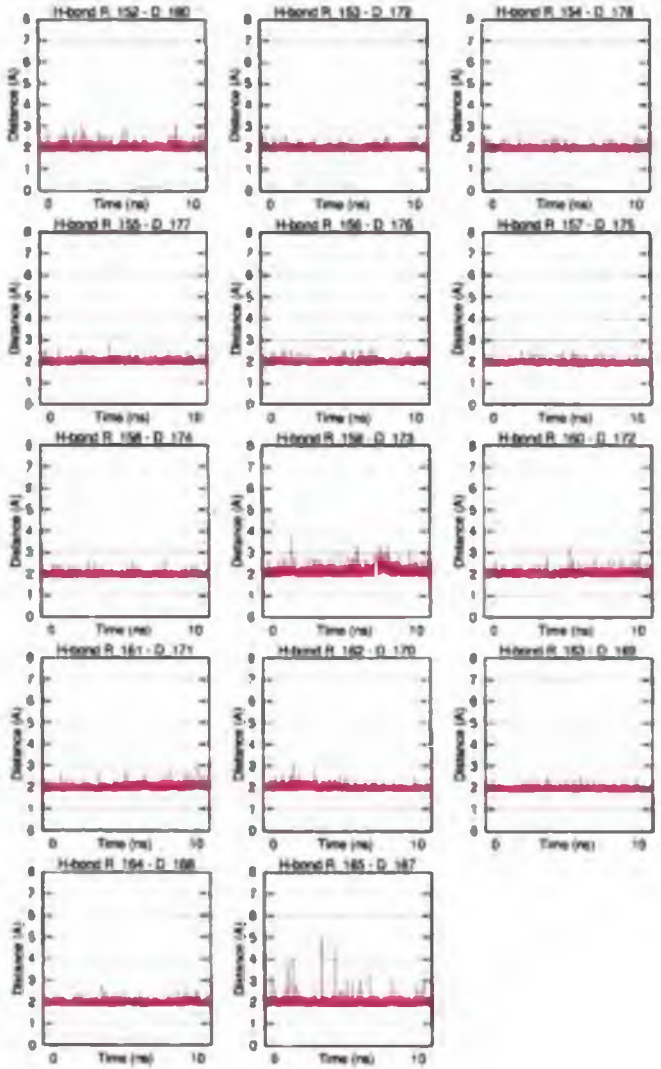


Figure A.42: The Watson-Crick hydrogen bonds. Model 6

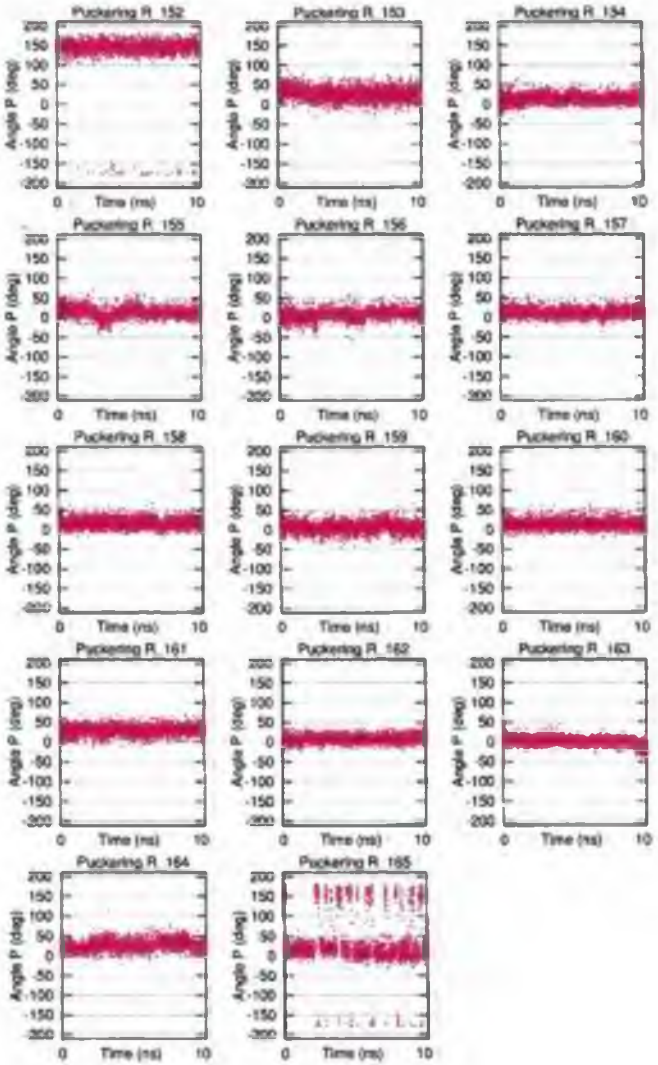


Figure A.43: Puckering of the RNA-strand. Model 6

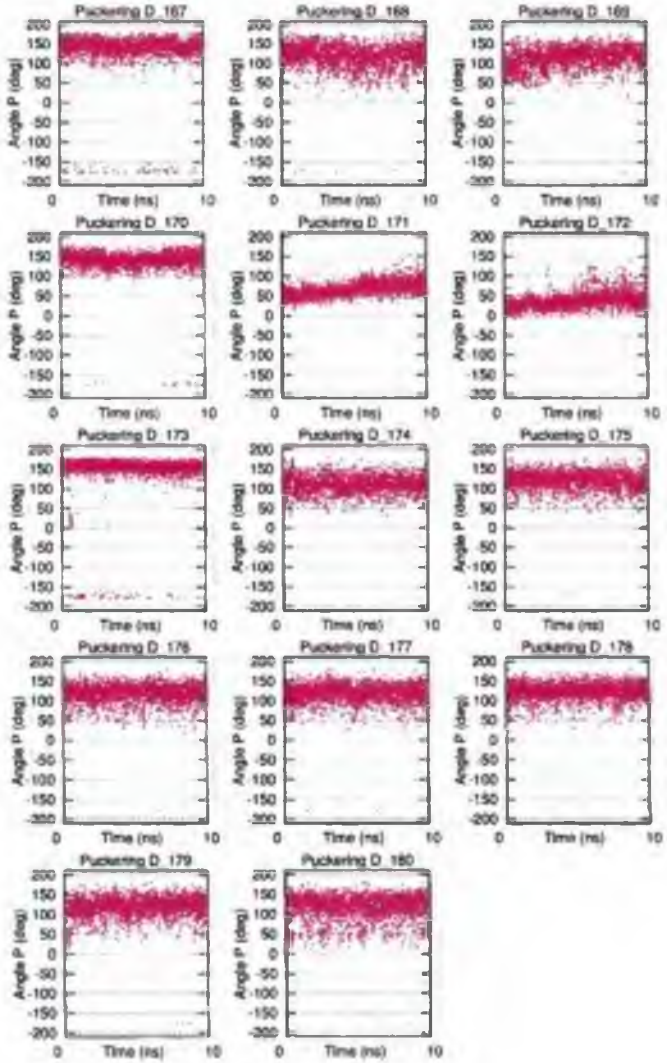


Figure A.44: Puckering of the DNA-strand. Model 6

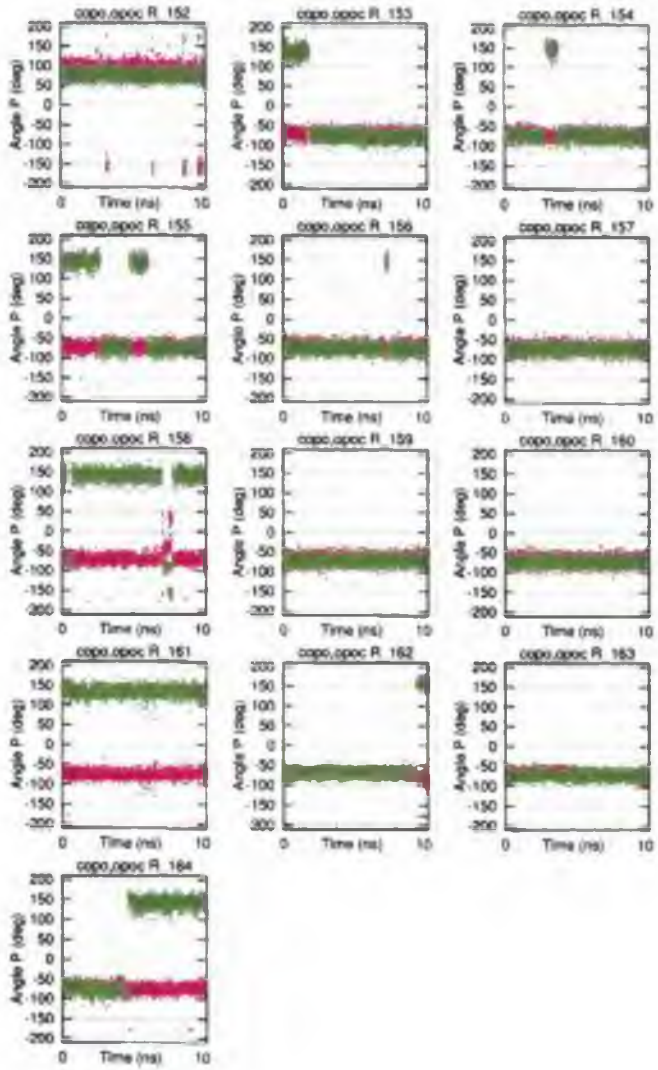


Figure A.45: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 6

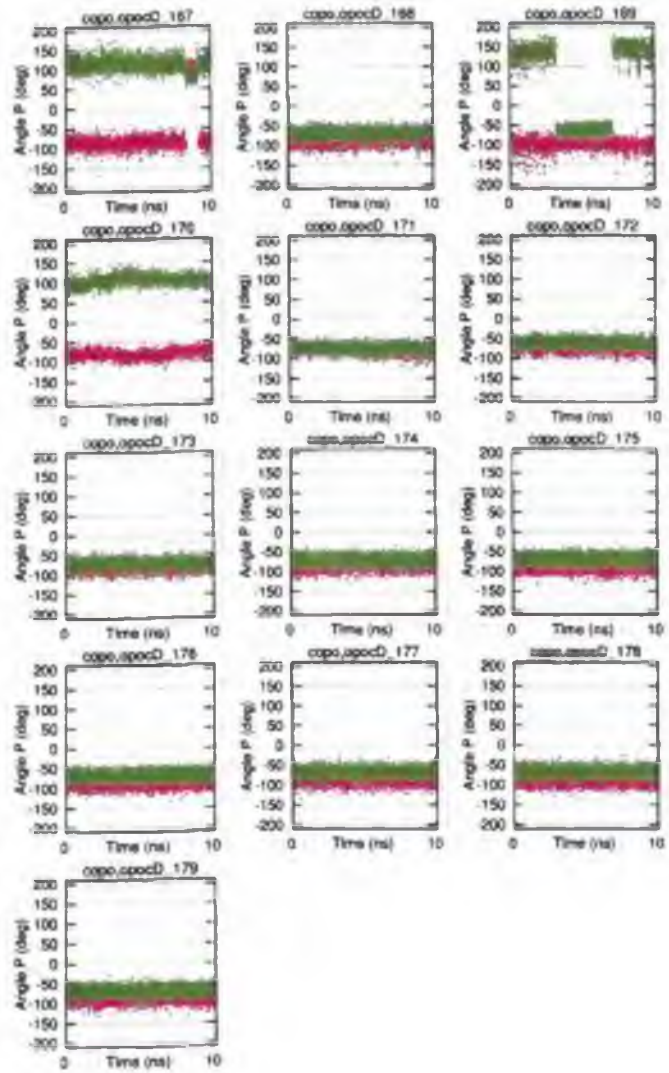


Figure A.46: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 6

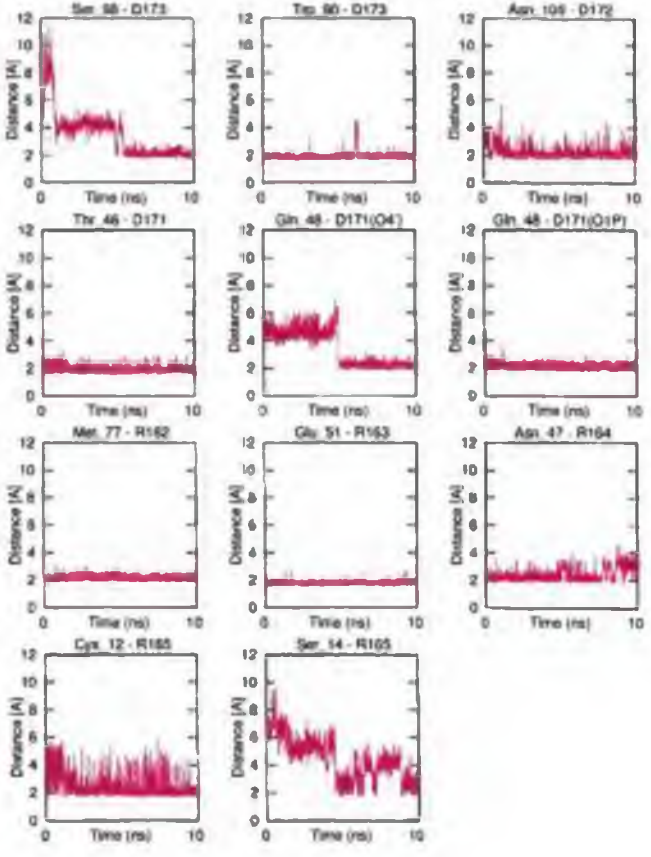


Figure A.47: The enzyme-substrate hydrogen bonds. Model 6

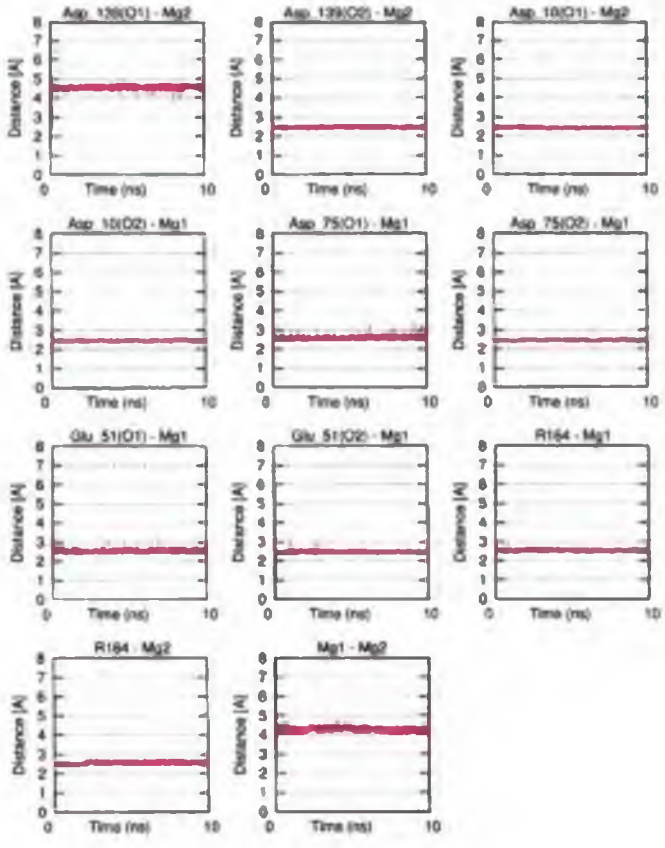


Figure A.48: Contacts between the Mg-ions and the DDE motif in the active site. Model 6

Appendix B

Graphs - human RNase H with an unmodified substrate: models 7-9

Model refinement II: active site stabilization - Mg ions parametrization

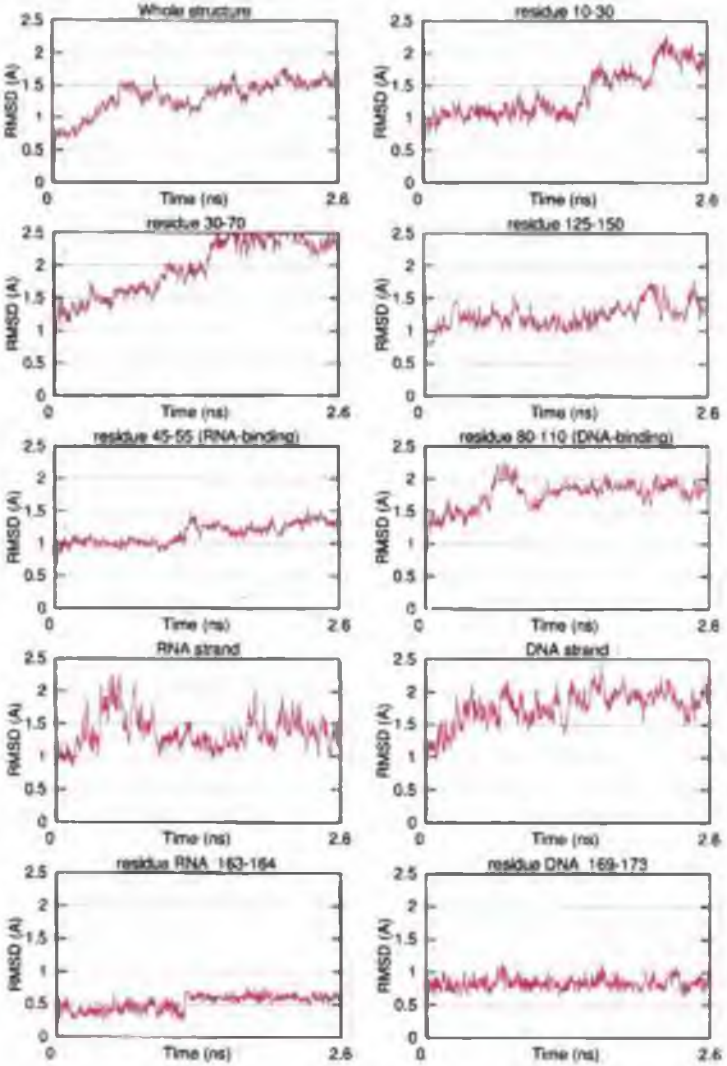


Figure B.1: RMSD of the whole structure and selected regions. Model 7

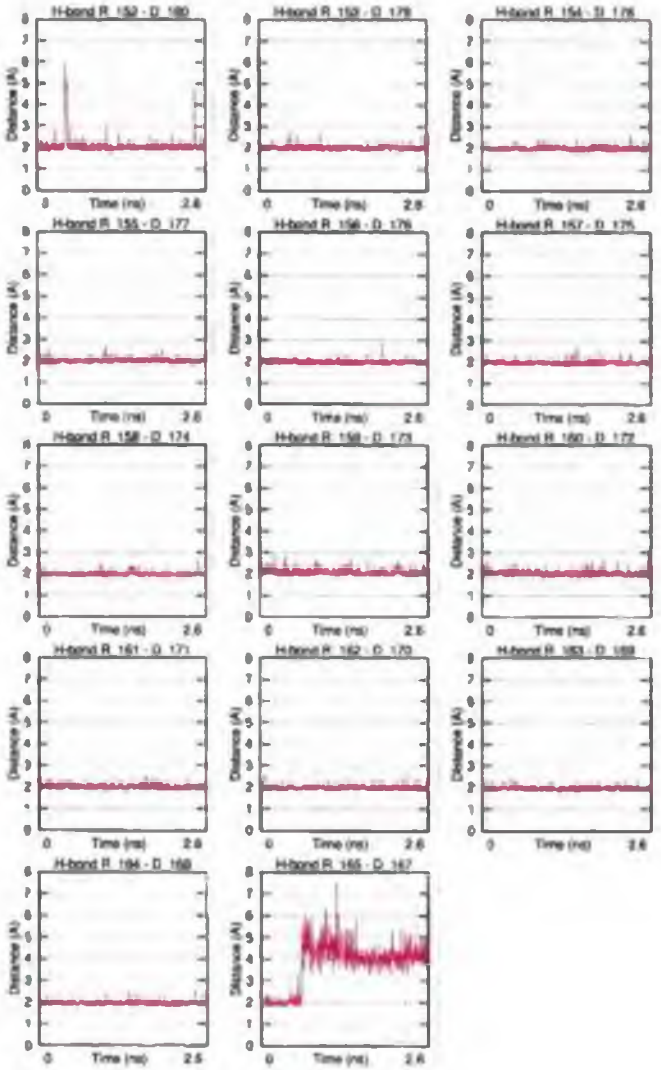


Figure B.2: The Watson-Crick hydrogen bonds. Model 7

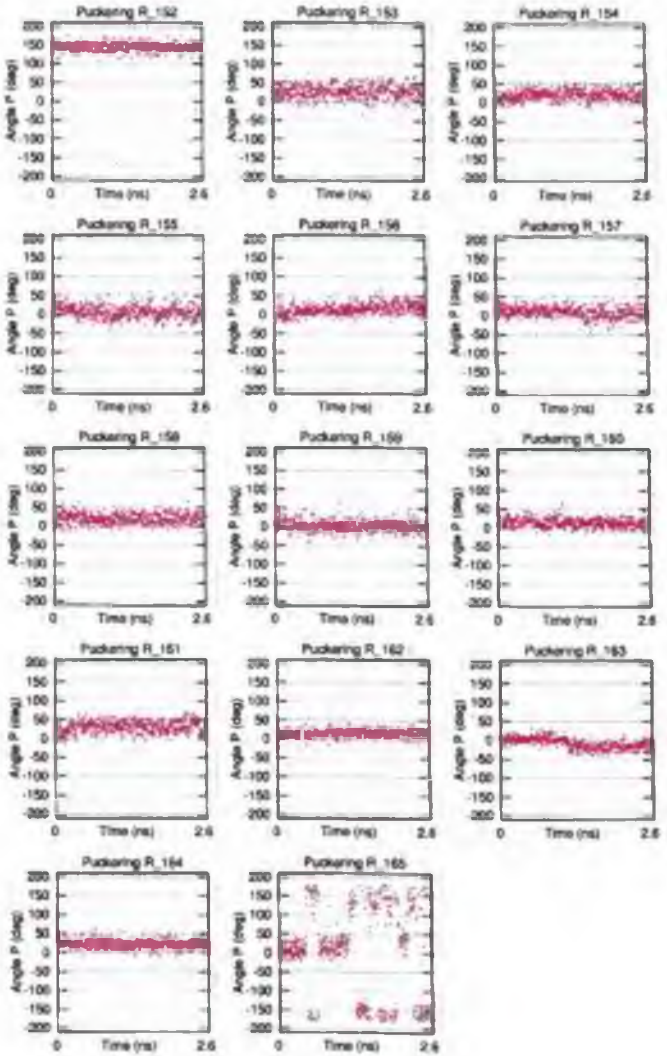


Figure B.3: Puckering of the RNA-strand. Model 7

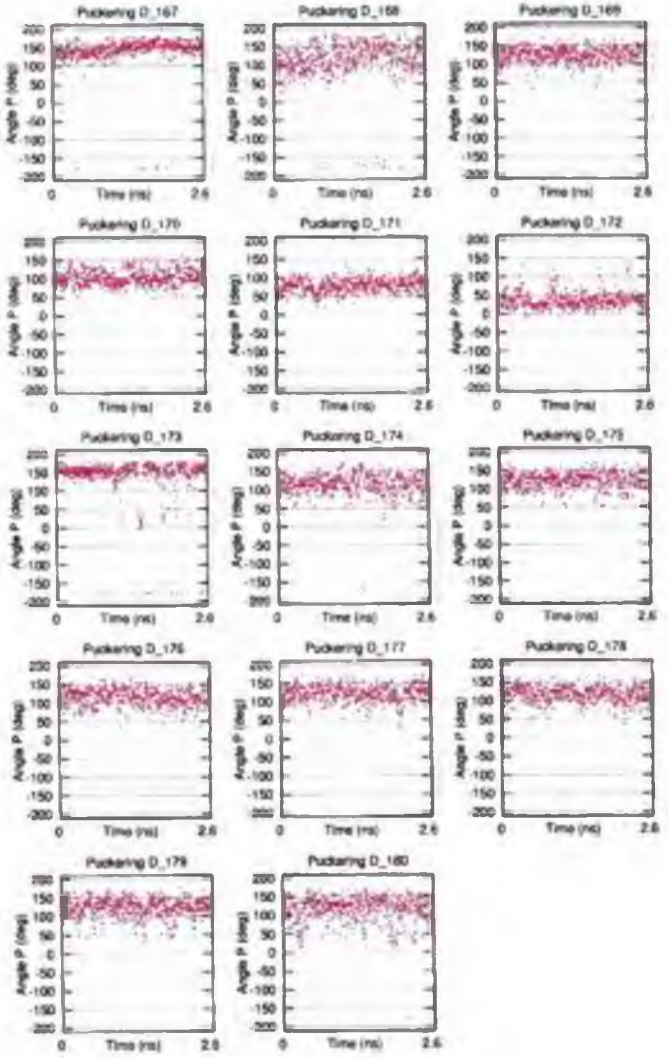


Figure B.4: Puckering of the DNA-strand. Model 7

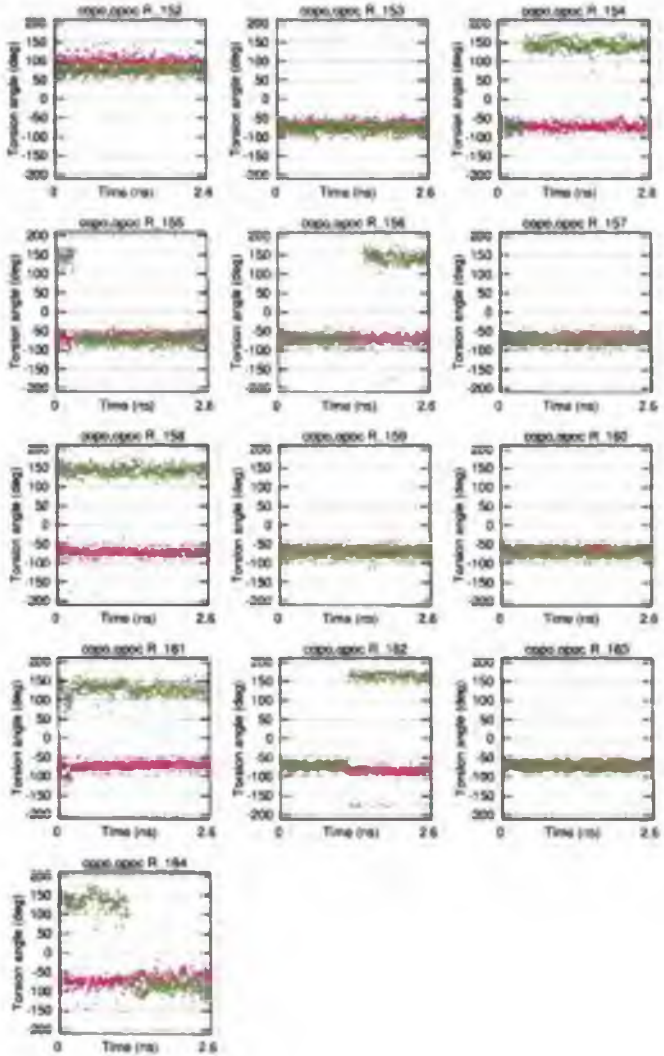


Figure B.5: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 7

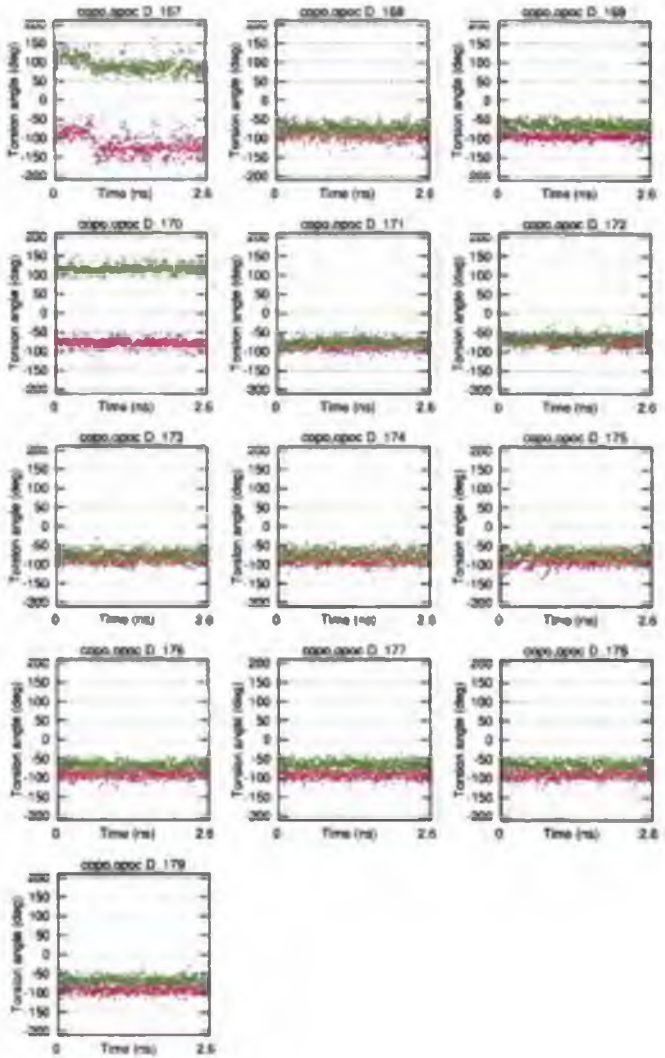


Figure B.6: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 7

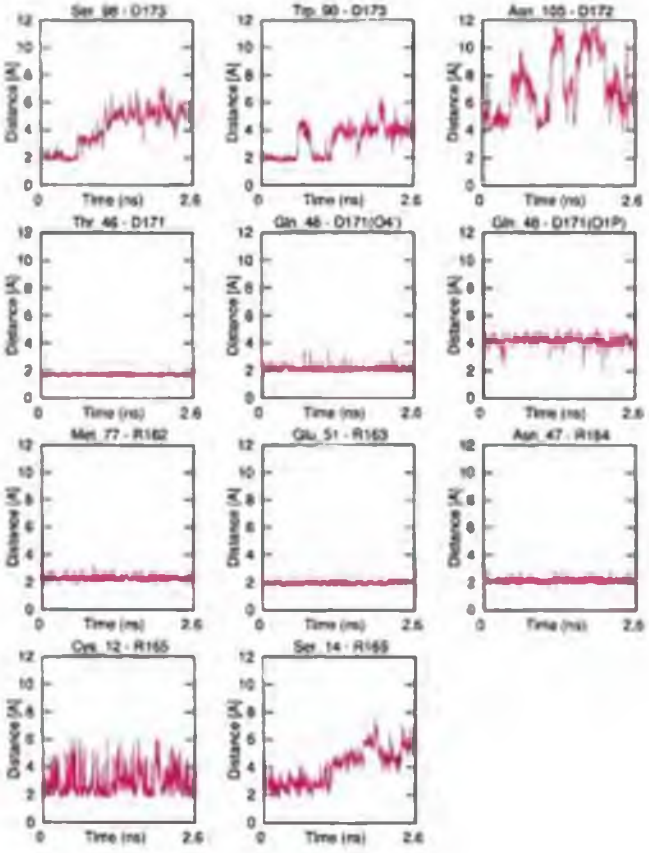


Figure B.7: The enzyme-substrate hydrogen bonds. Model 7

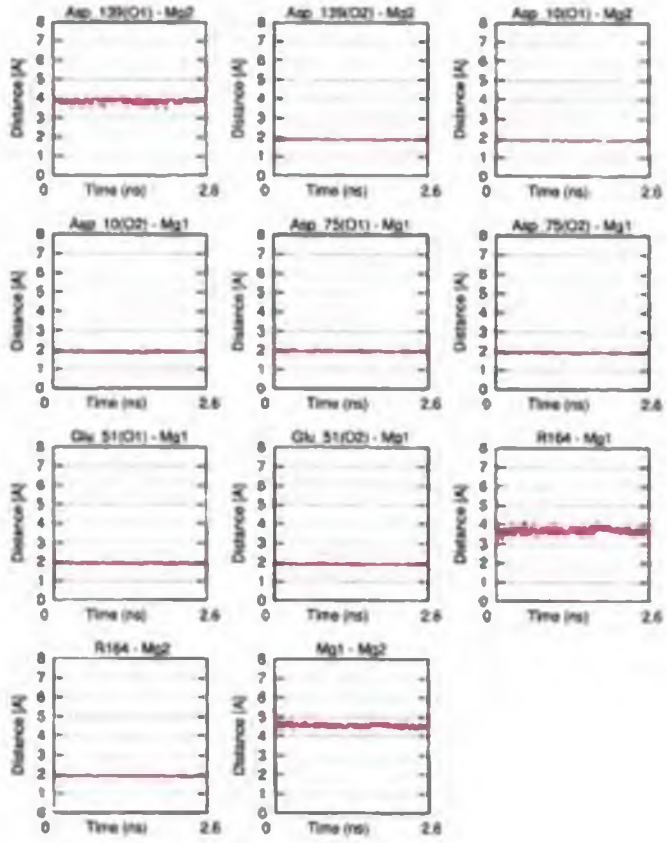


Figure B.8: Contacts between the Mg-ions and the DDE motif in the active site. Model 7

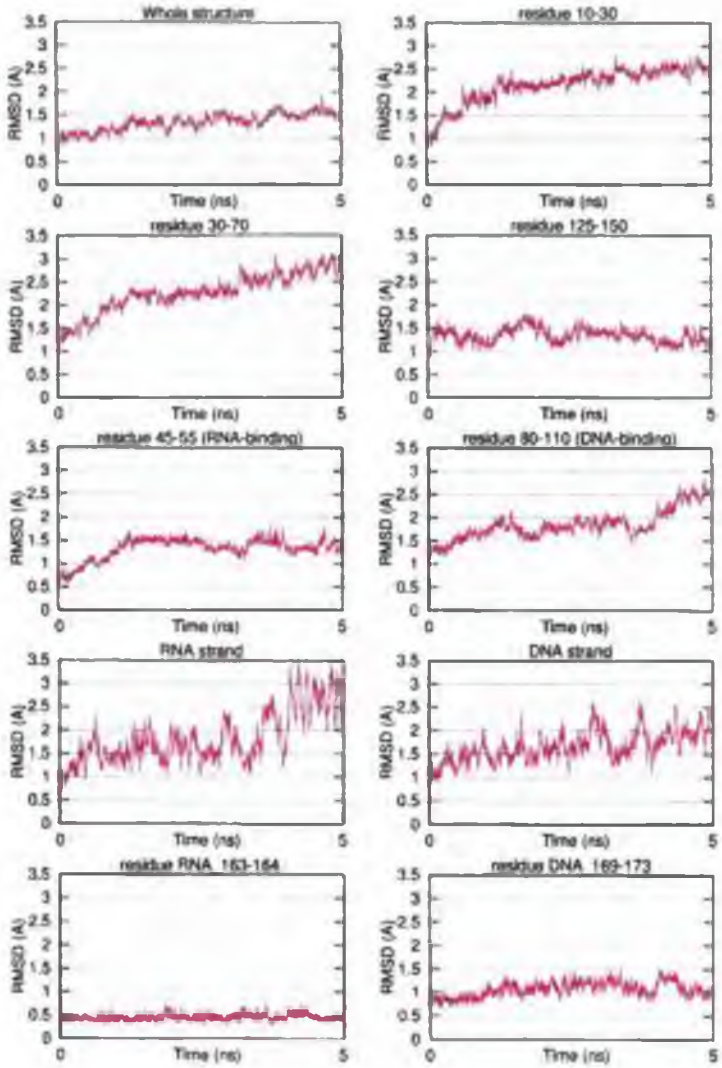


Figure B.9: RMSD of the whole structure and selected regions. Model 8

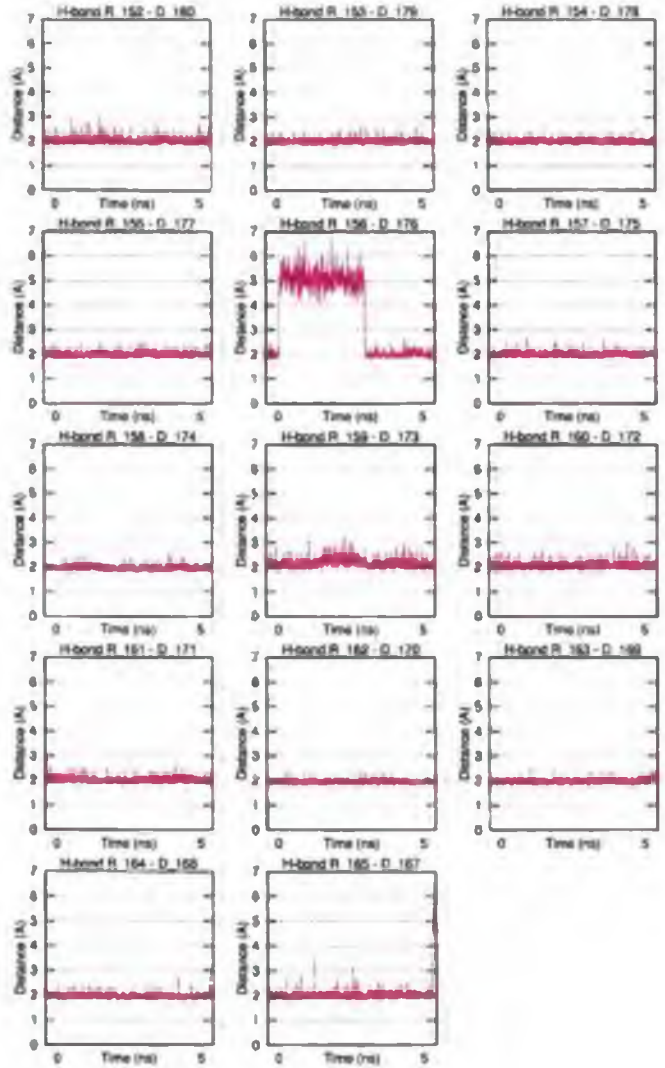


Figure B.10: The Watson-Crick hydrogen bonds. Model 8

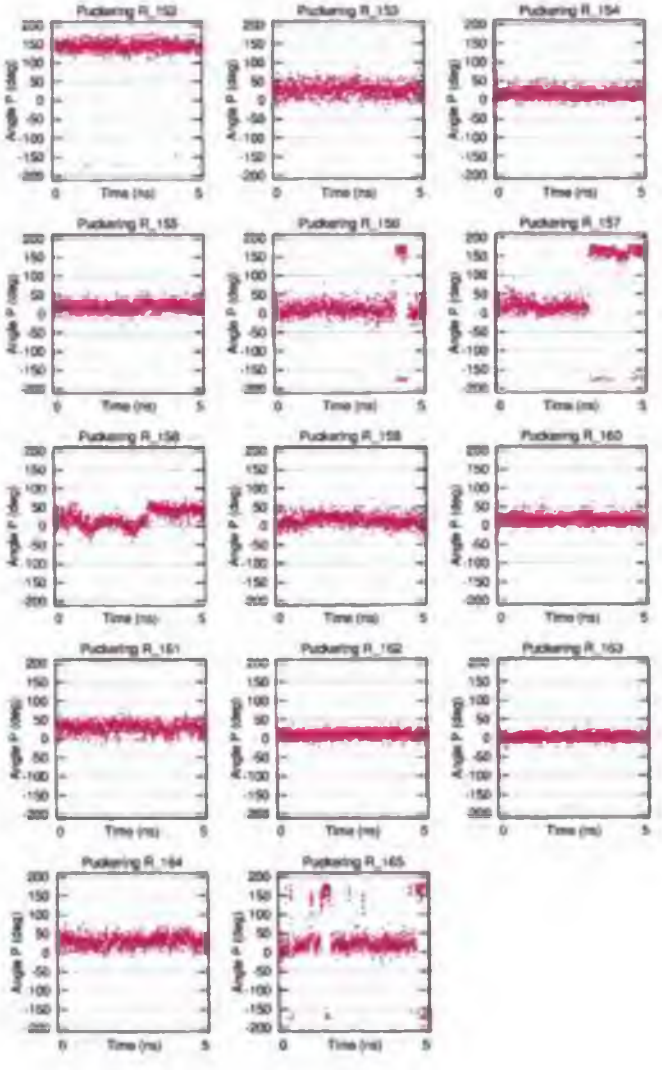


Figure B.11: Puckering of the RNA-strand. Model 8

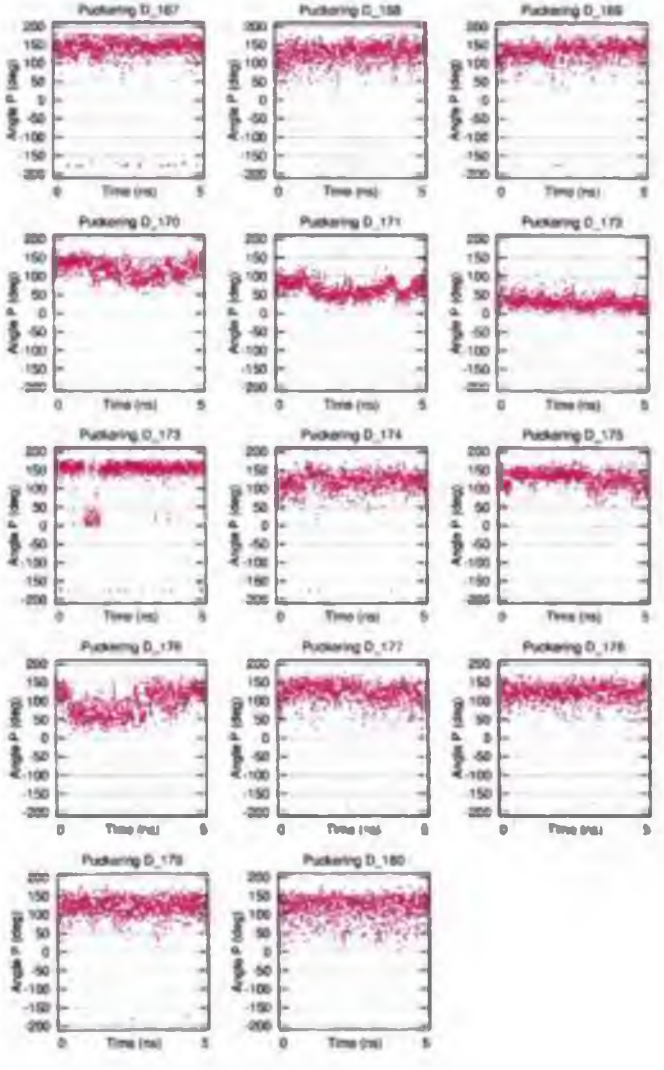


Figure B.12: Puckering of the DNA-strand. Model 8

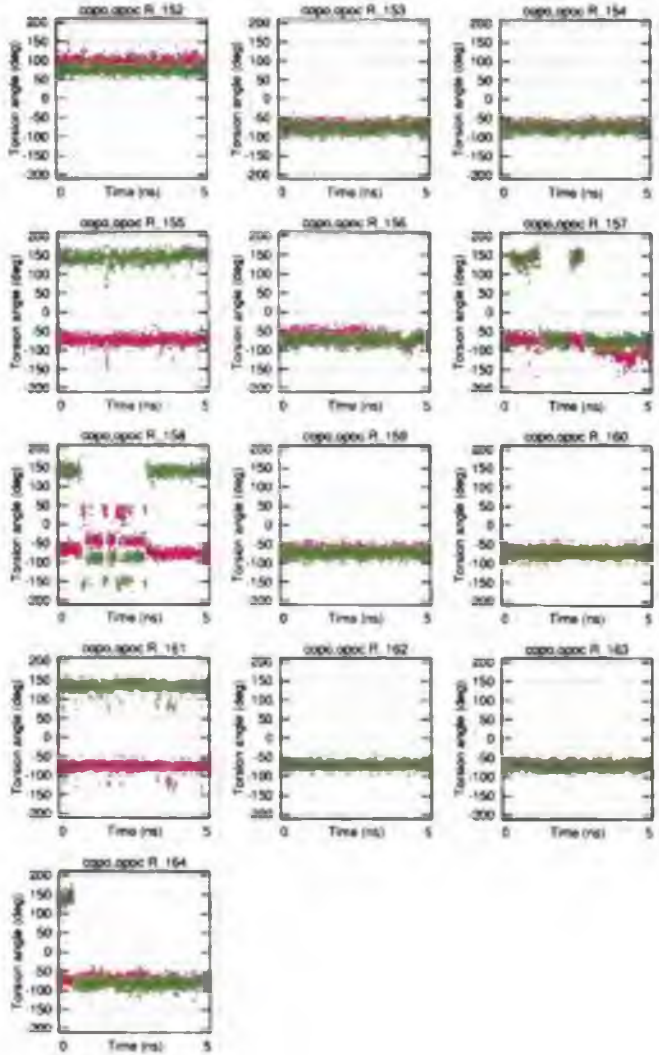


Figure B.13: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 8

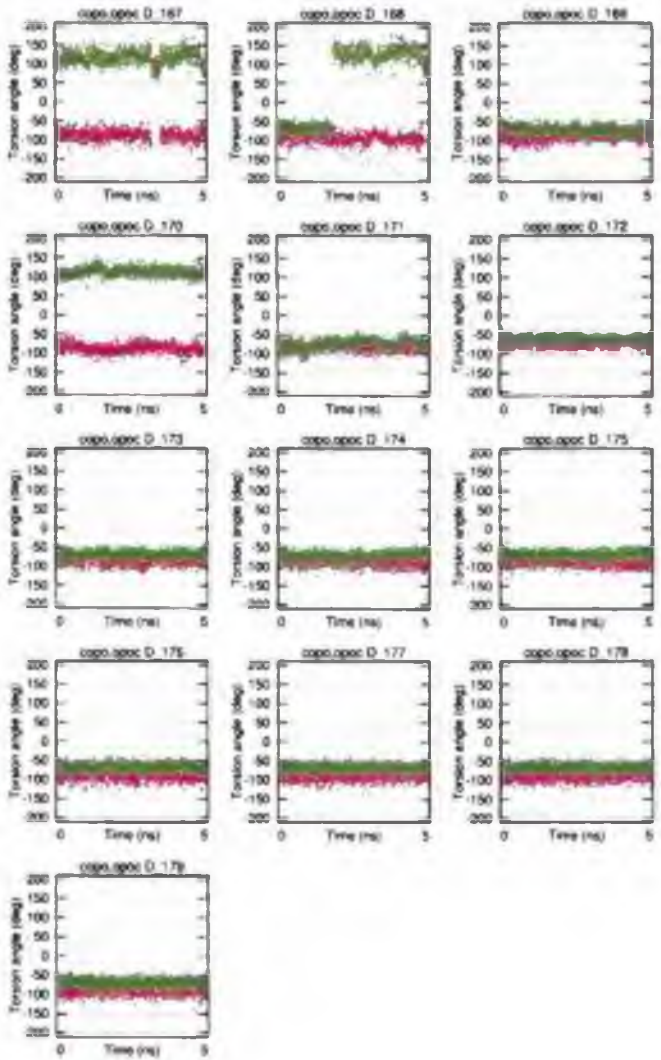


Figure B.14: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 8

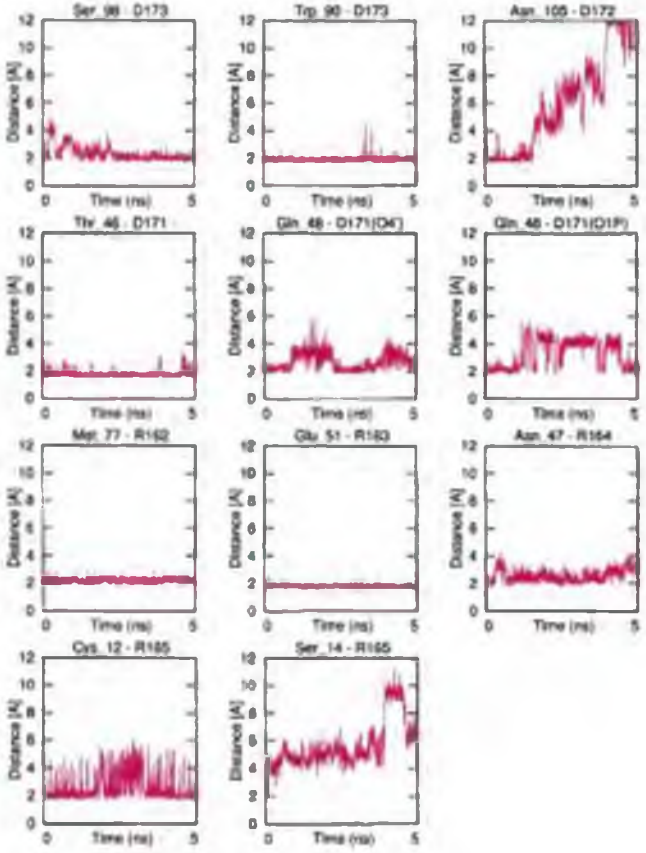


Figure B.15: The enzyme-substrate hydrogen bonds. Model 8

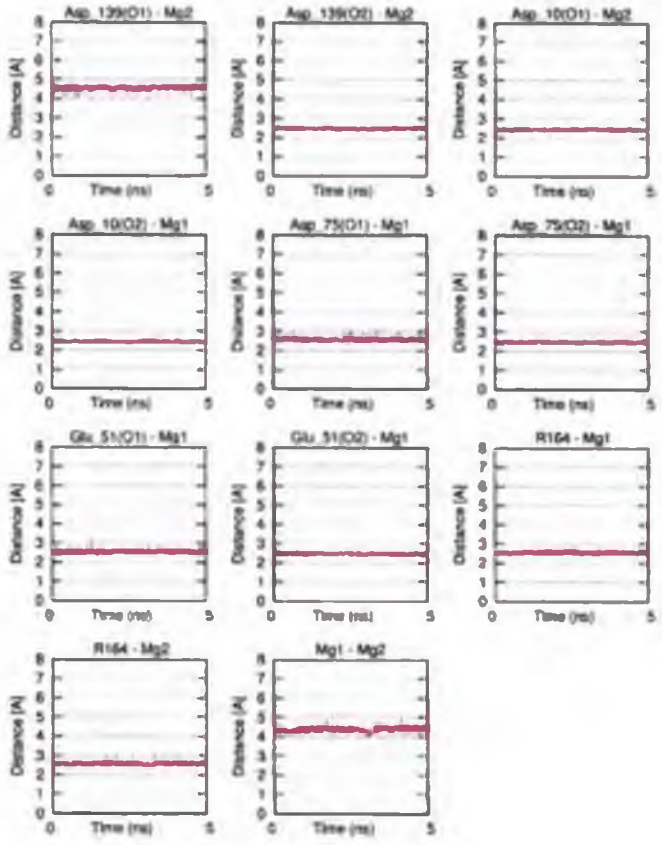


Figure B.16: Contacts between the Mg-ions and the DDE motif in the active site. Model 8

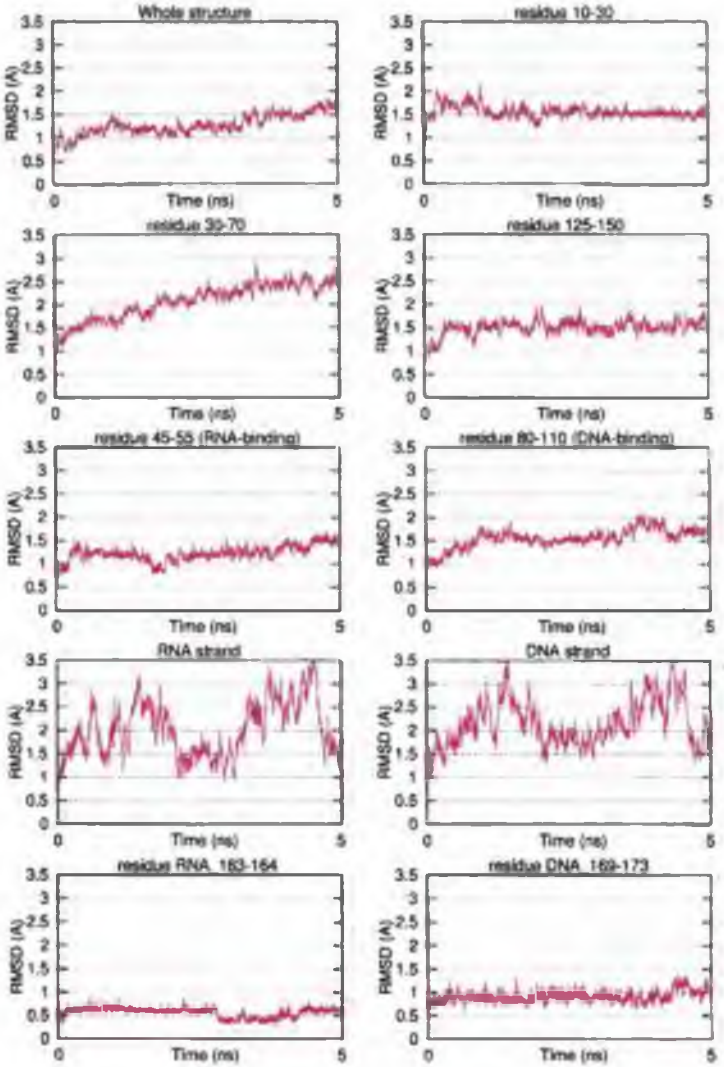


Figure B.17: RMSD of the whole structure and selected regions. Model 9

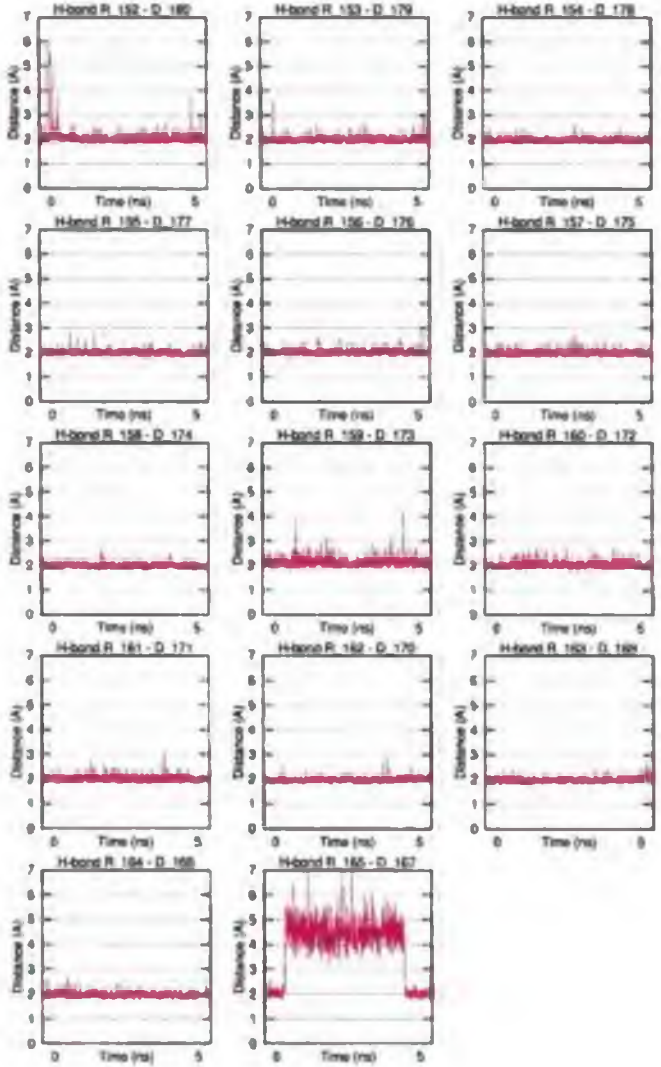


Figure B.18: The Watson-Crick hydrogen bonds. Model 9

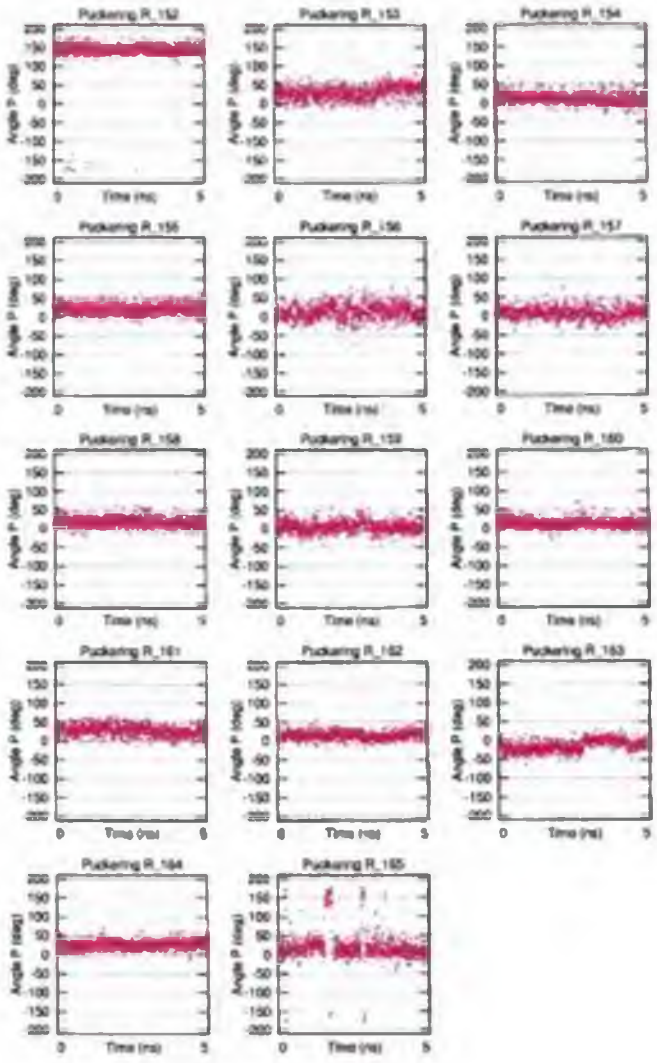


Figure B.19: Puckering of the RNA-strand. Model 9

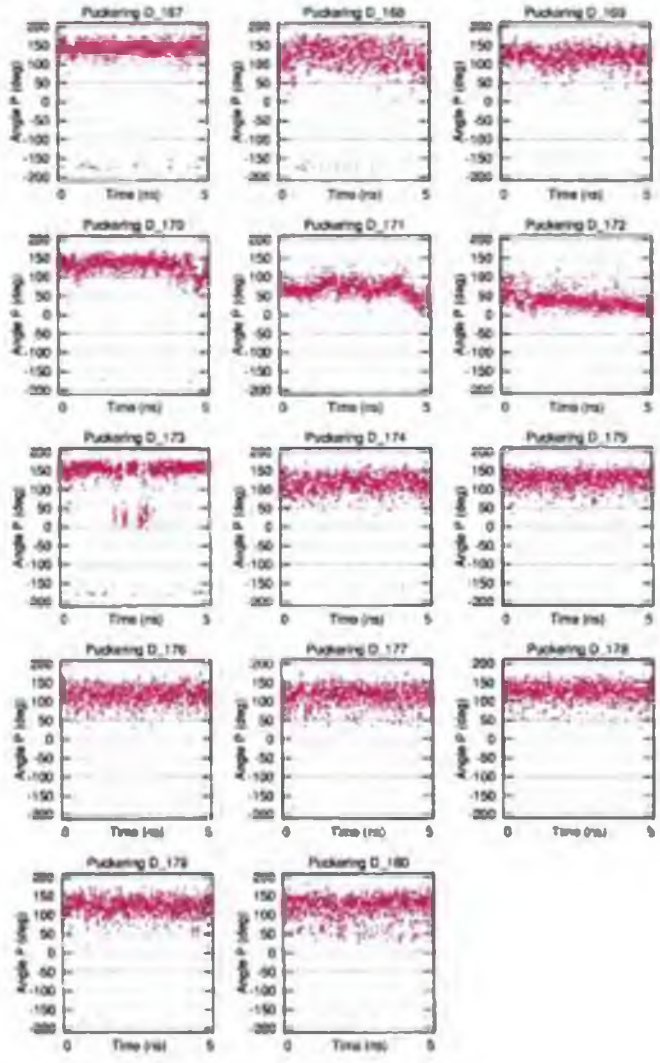


Figure B.20: Puckering of the DNA-strand. Model 9

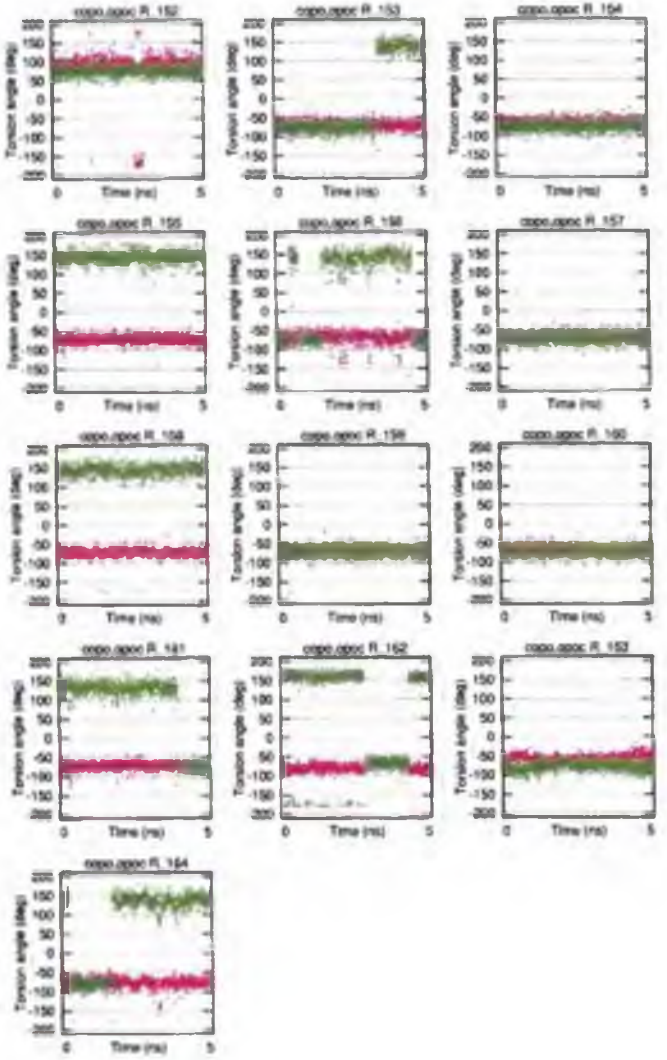


Figure B.21: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model 9

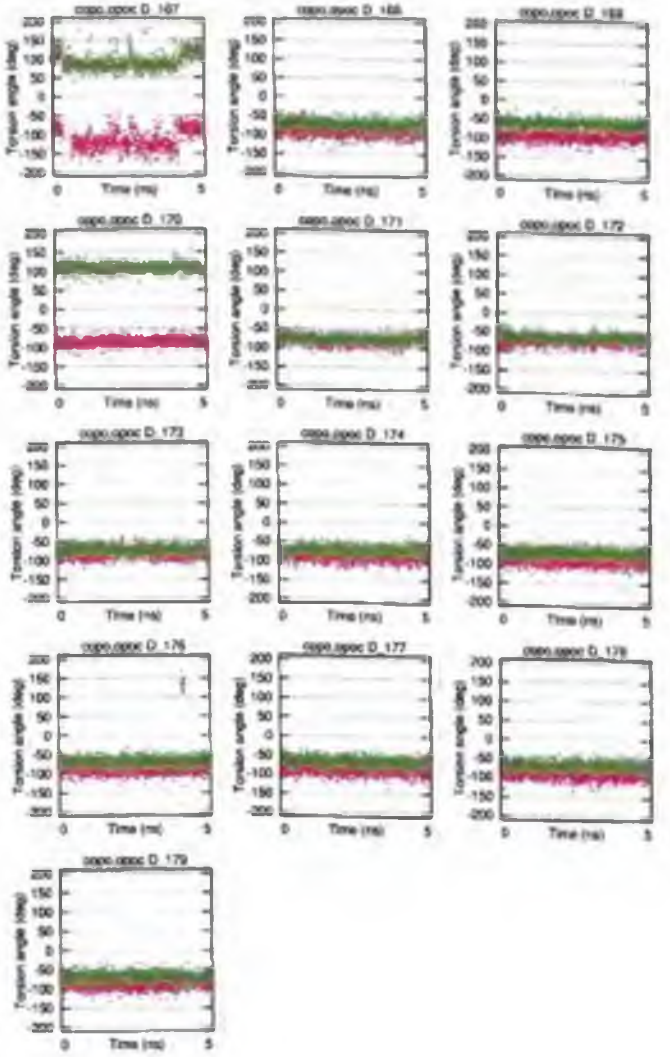


Figure B.22: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand. Model 9

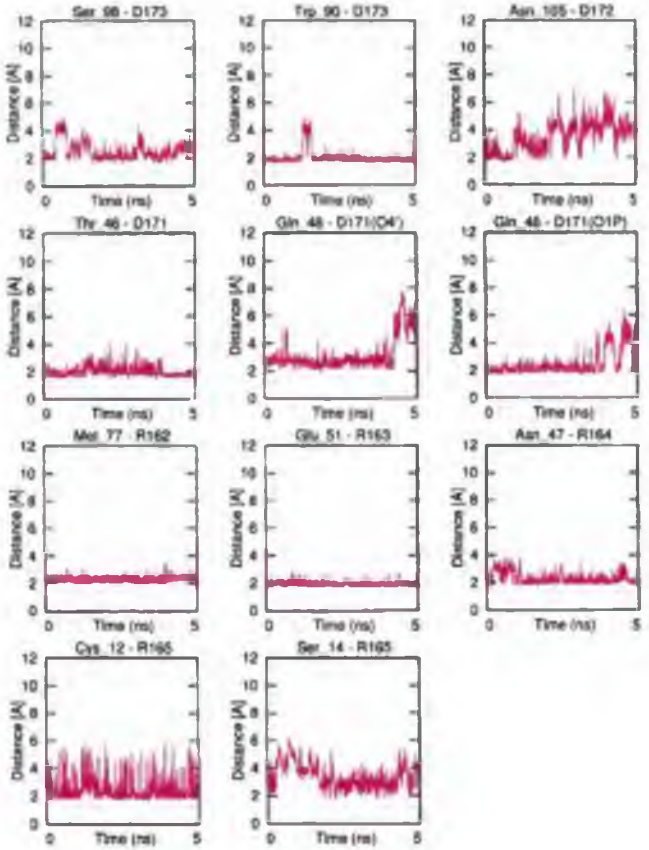


Figure B.23: The enzyme–substrate hydrogen bonds. Model 9

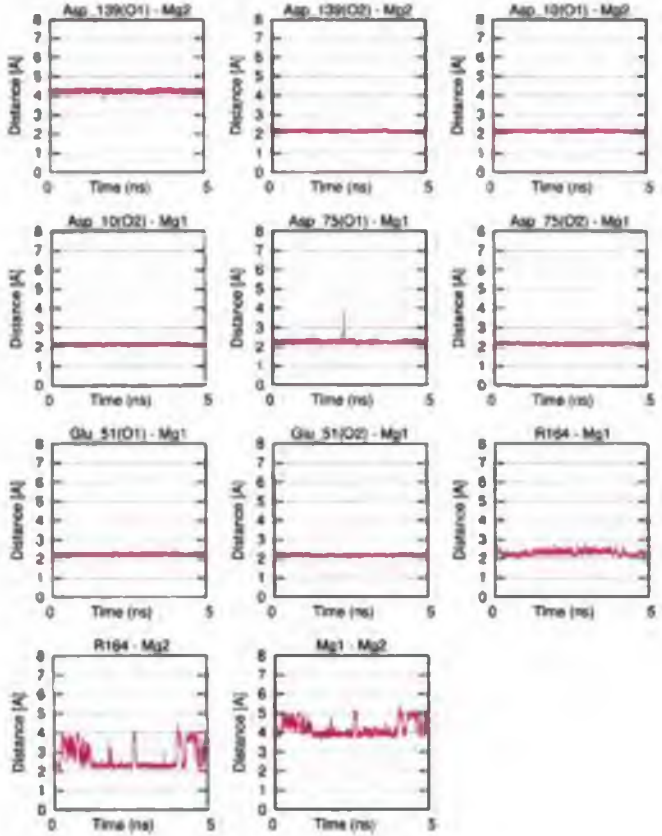
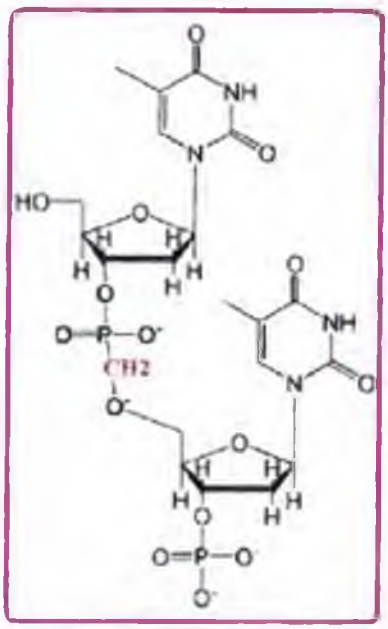


Figure B.24: Contacts between the Mg-ions and the DDE motif in the active site. Model 9

Appendix C

Graphs - models with a modified substrate

The C3'-O-P-*C*-O-C5' modification



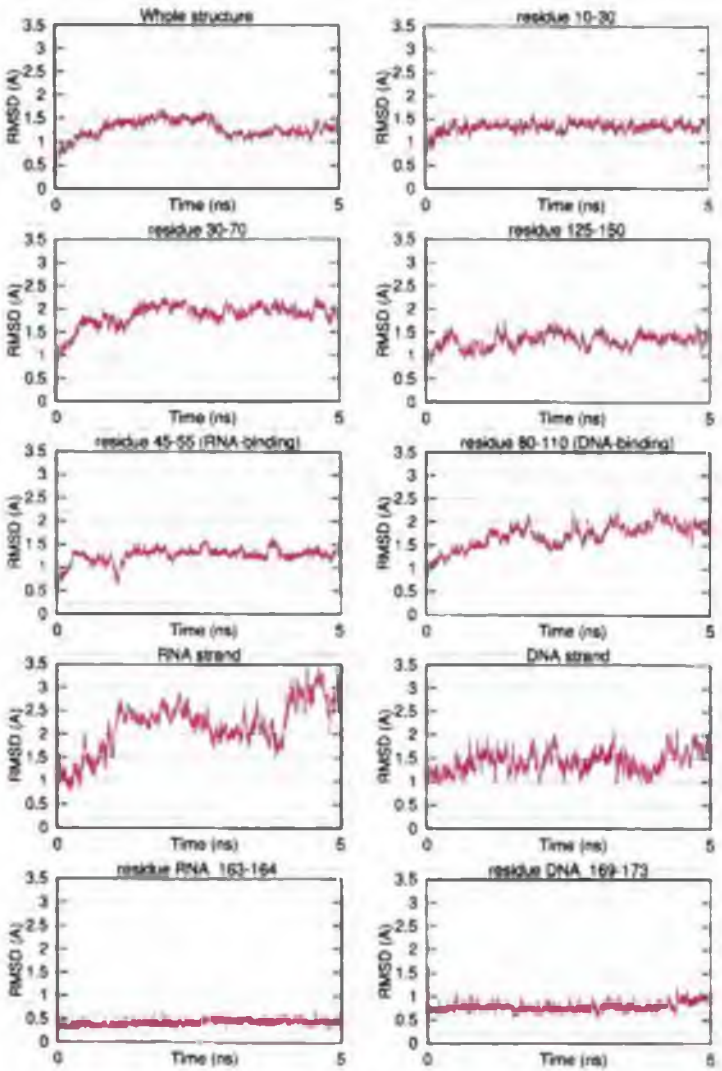


Figure C.1: RMSD of the whole structure and selected regions. Modification 1

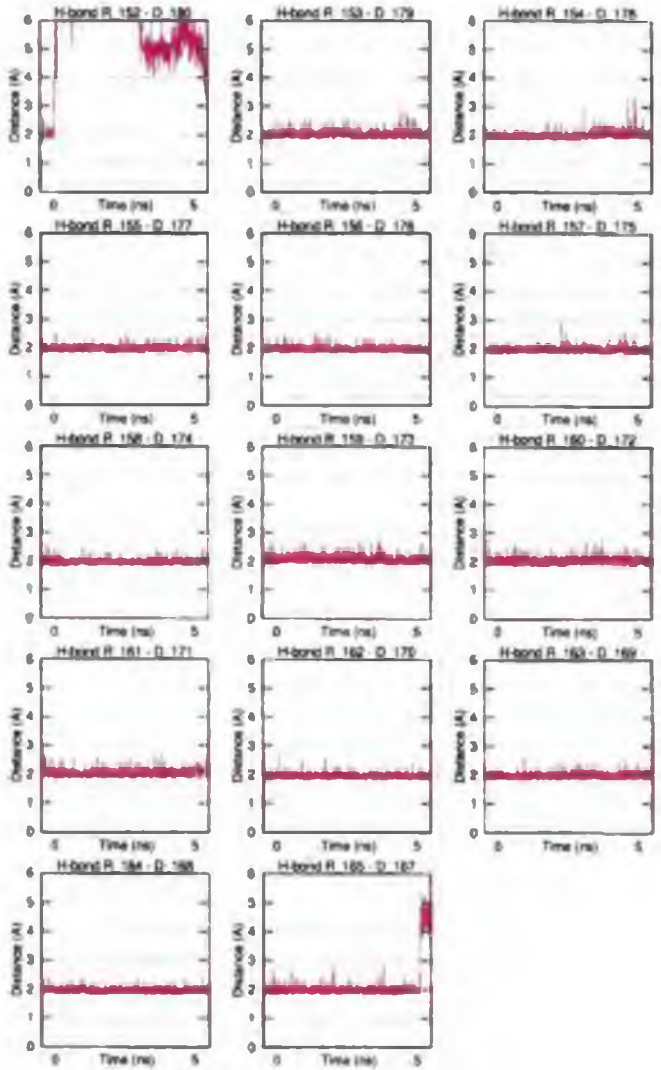


Figure C.2: The Watson-Crick hydrogen bonds. Modification 1

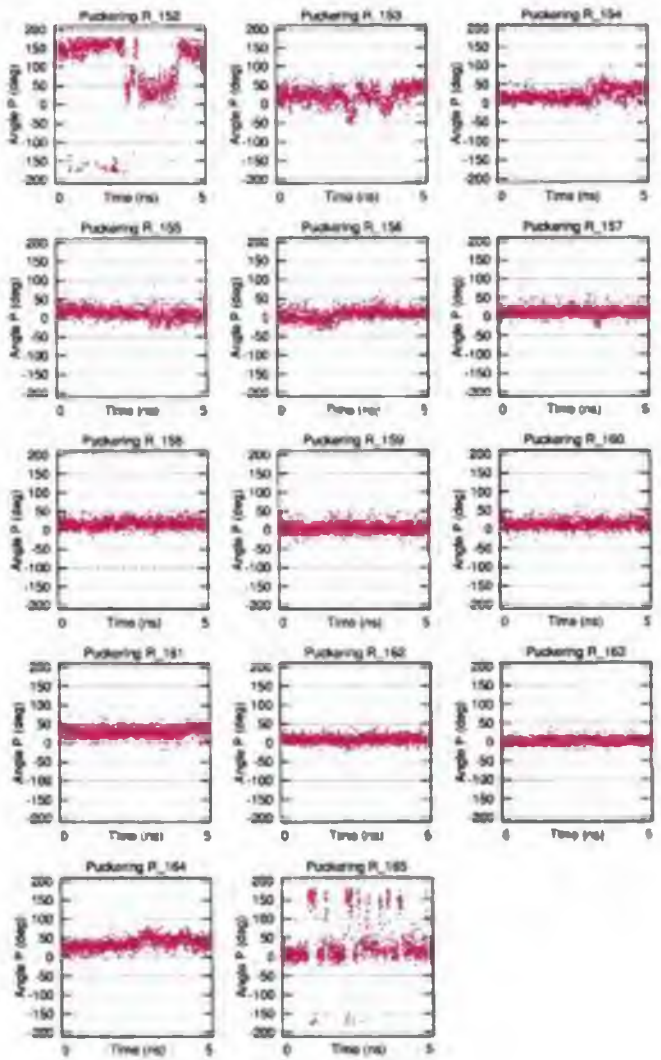


Figure C.3: Puckering of the RNA-strand. Modification 1

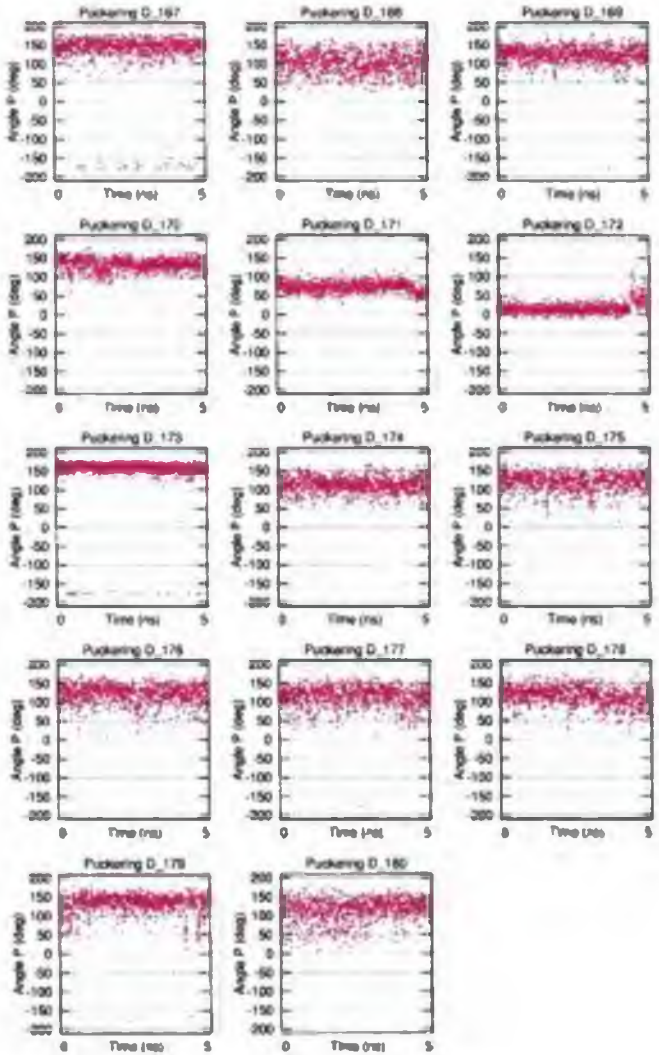


Figure C.4: Puckering of the DNA-strand. Modification 1

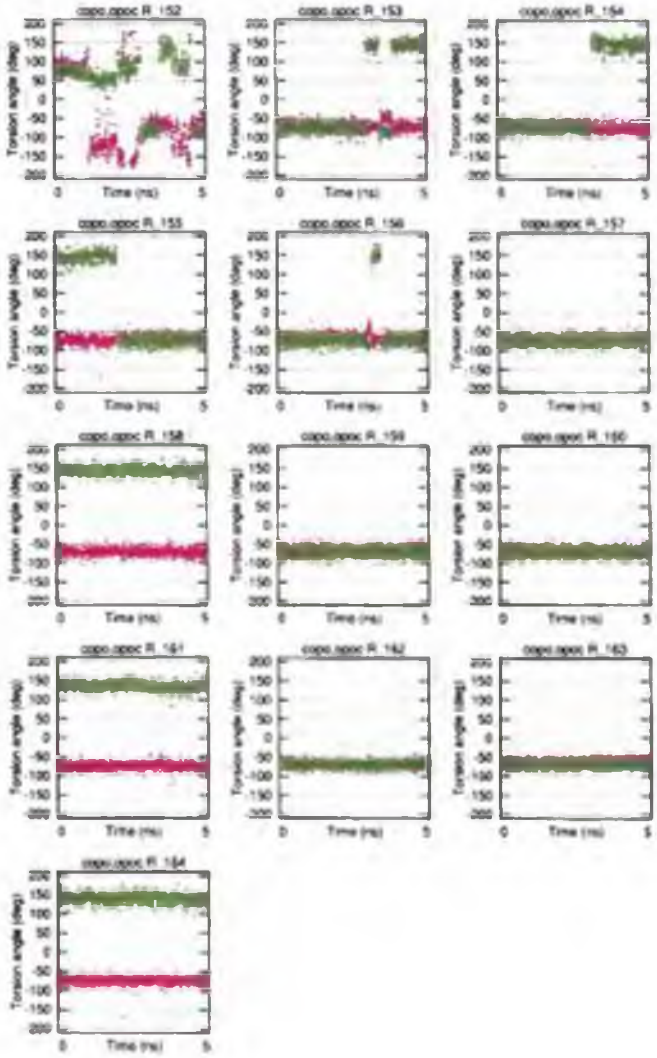


Figure C.5: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 1

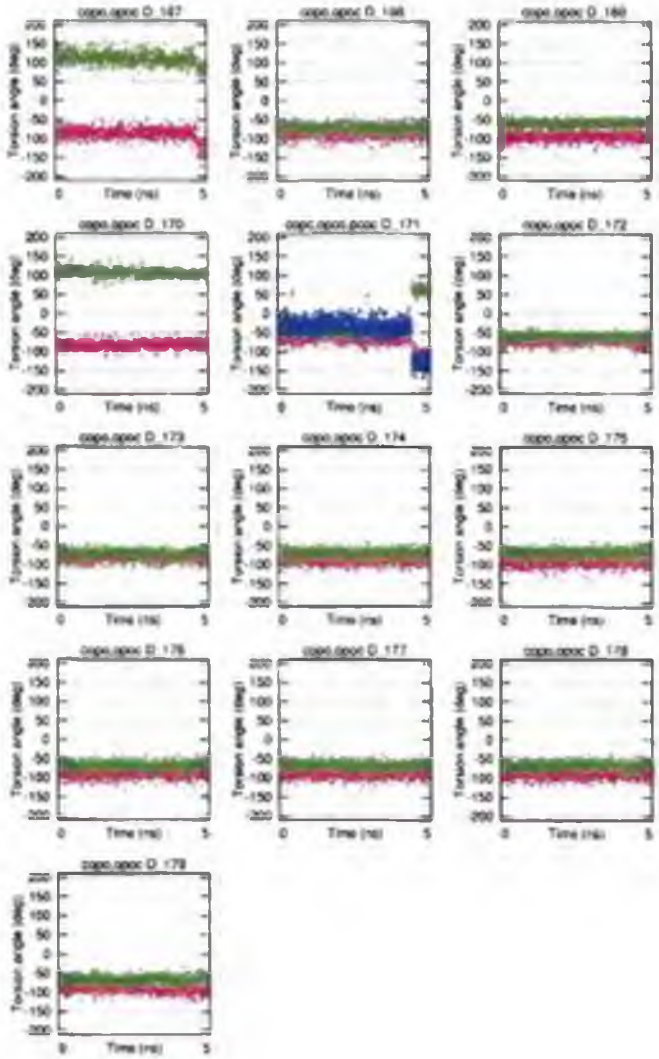


Figure C.6: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 171). Modification 1

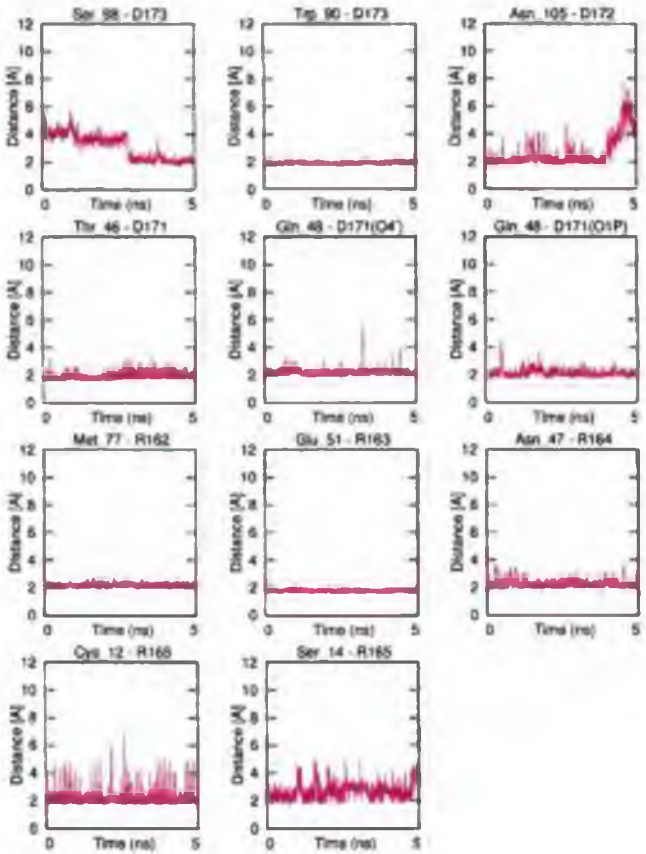


Figure C.7: The enzyme-substrate hydrogen bonds. Modification 1

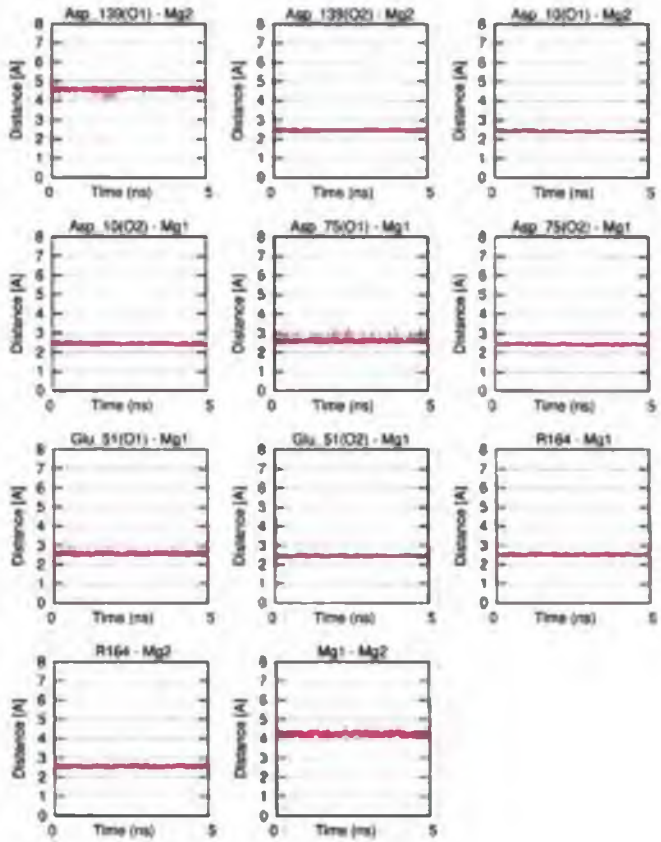


Figure C.8: Contacts between the Mg-ions and the DDE motif in the active site. Modification 1

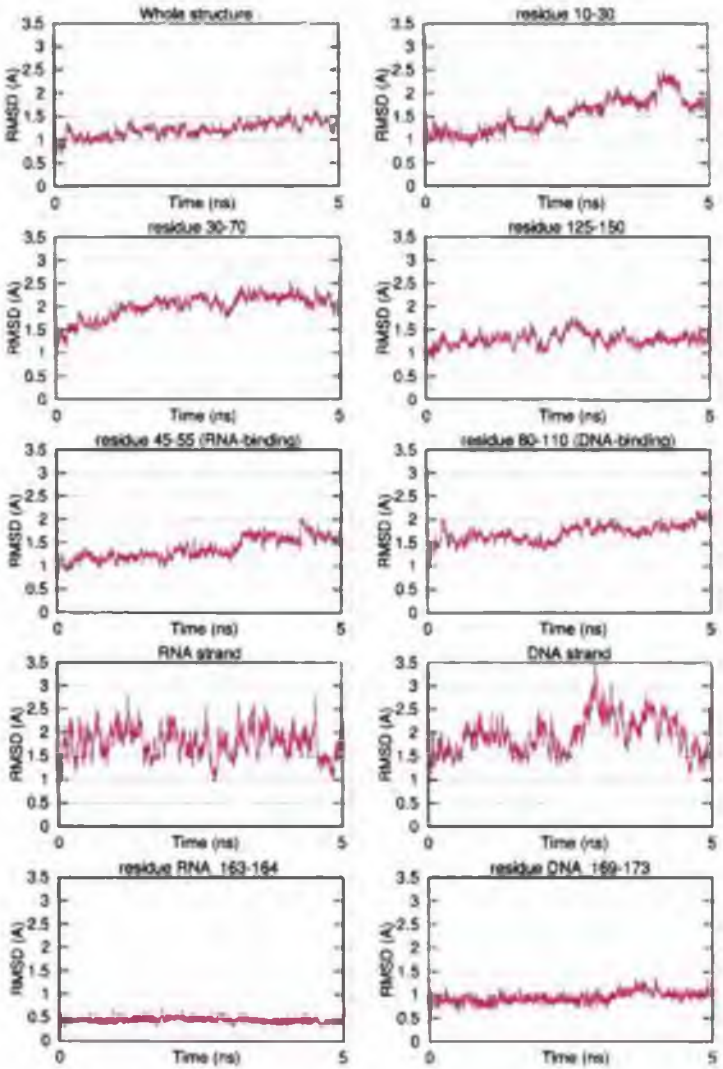


Figure C.9: RMSD of the whole structure and selected regions. Modification 2

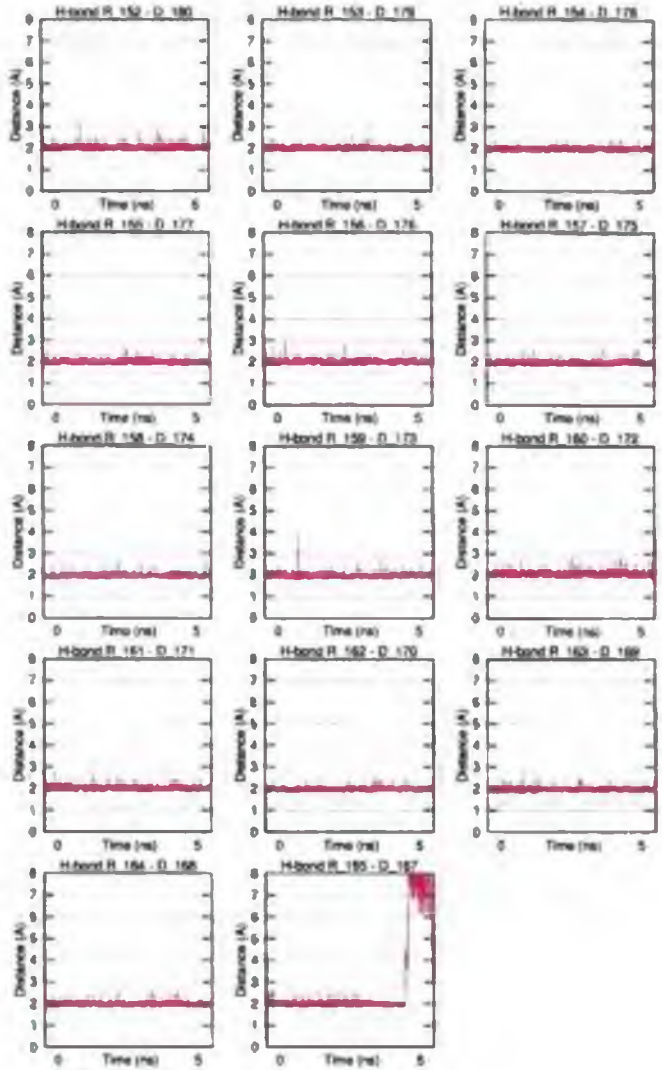


Figure C.10: The Watson-Crick hydrogen bonds. Modification 2

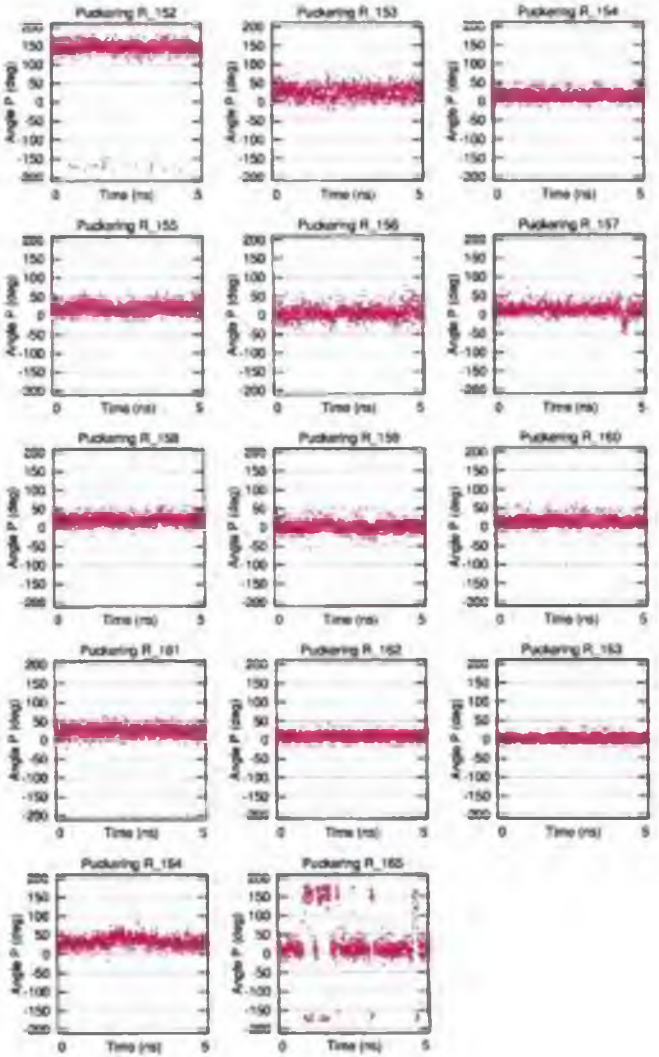


Figure C.11: Puckering of the RNA-strand. Modification 2

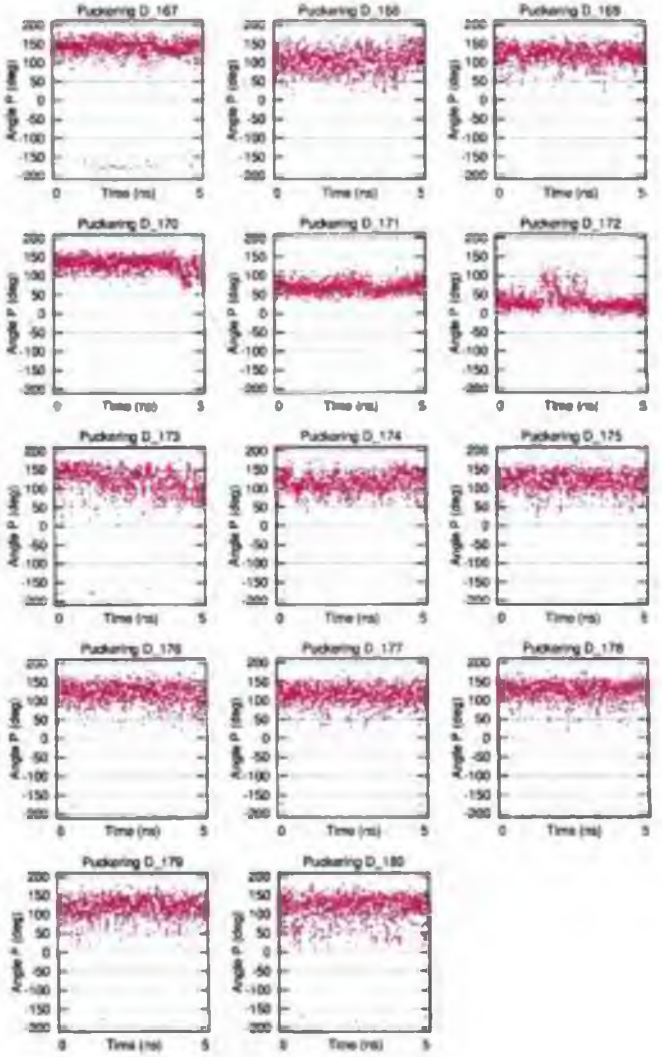


Figure C.12: Puckering of the DNA-strand. Modification 2

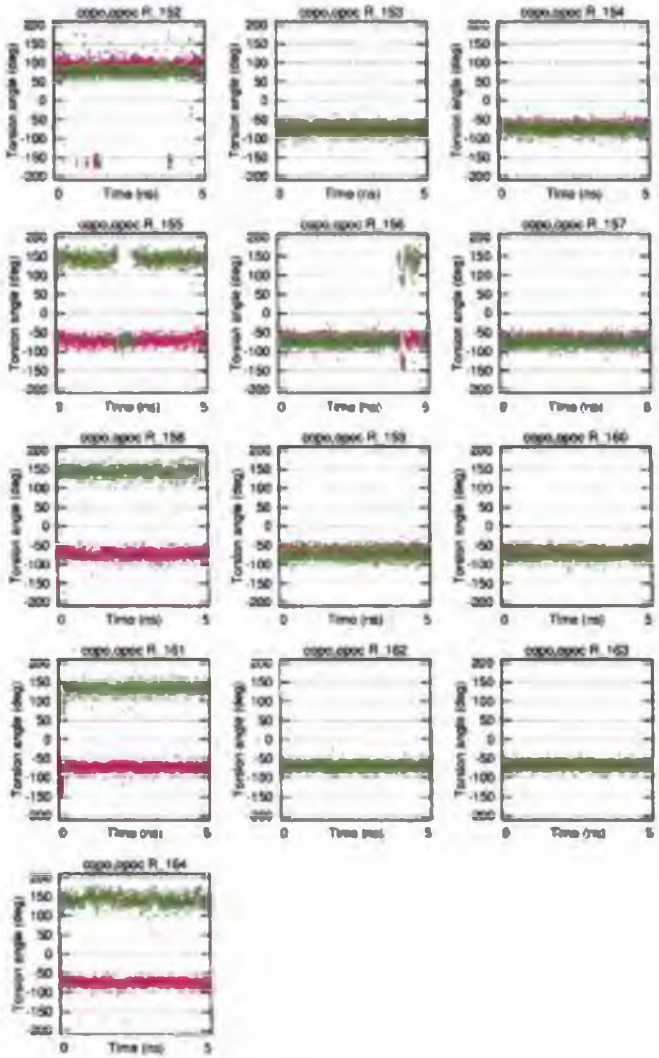


Figure C.13: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 2

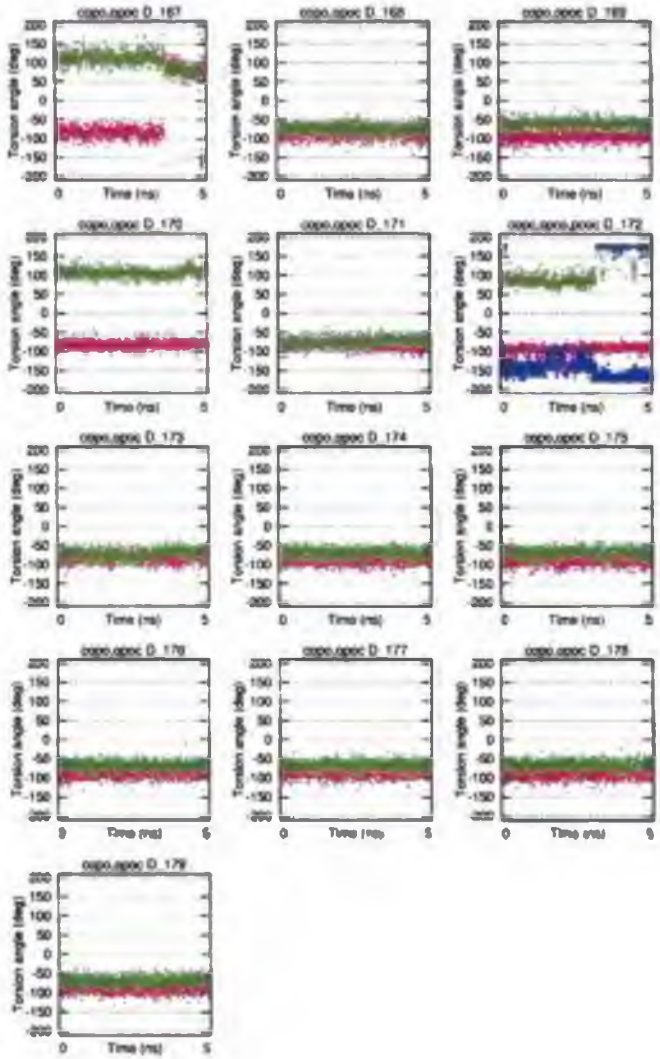


Figure C.14: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 172). Modification 2

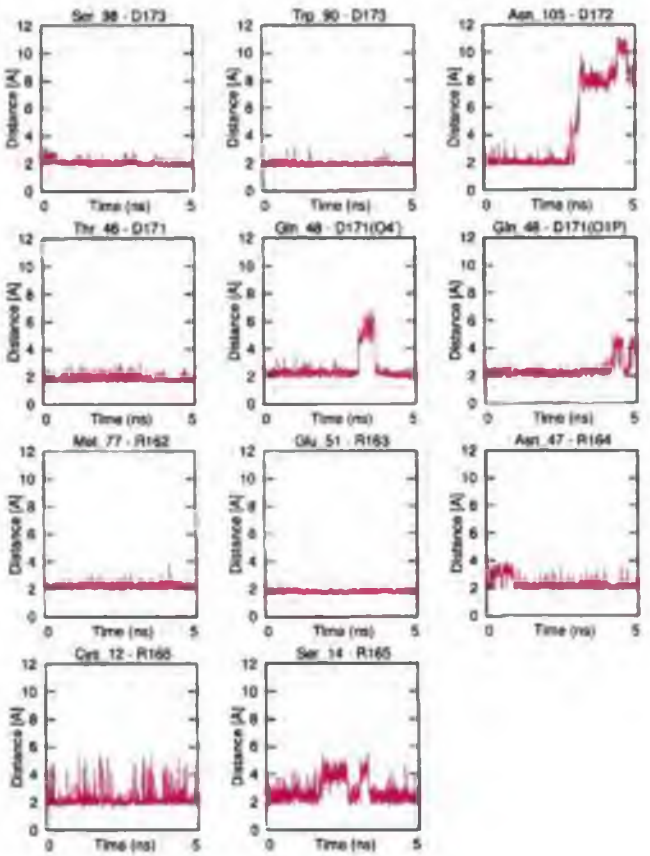


Figure C.15: The enzyme-substrate hydrogen bonds. Modification 2

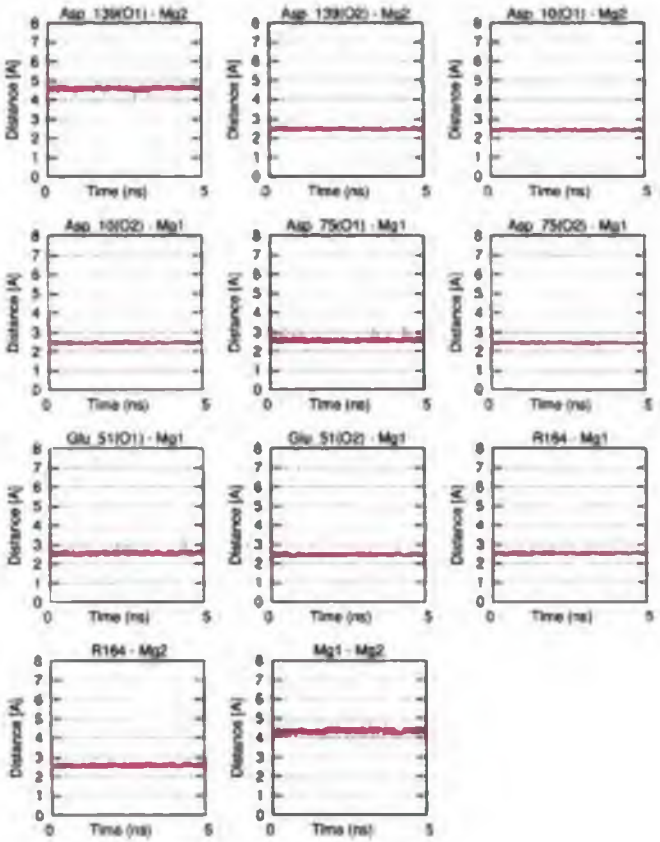


Figure C.16: Contacts between the Mg-ions and the DDE motif in the active site. Modification 2

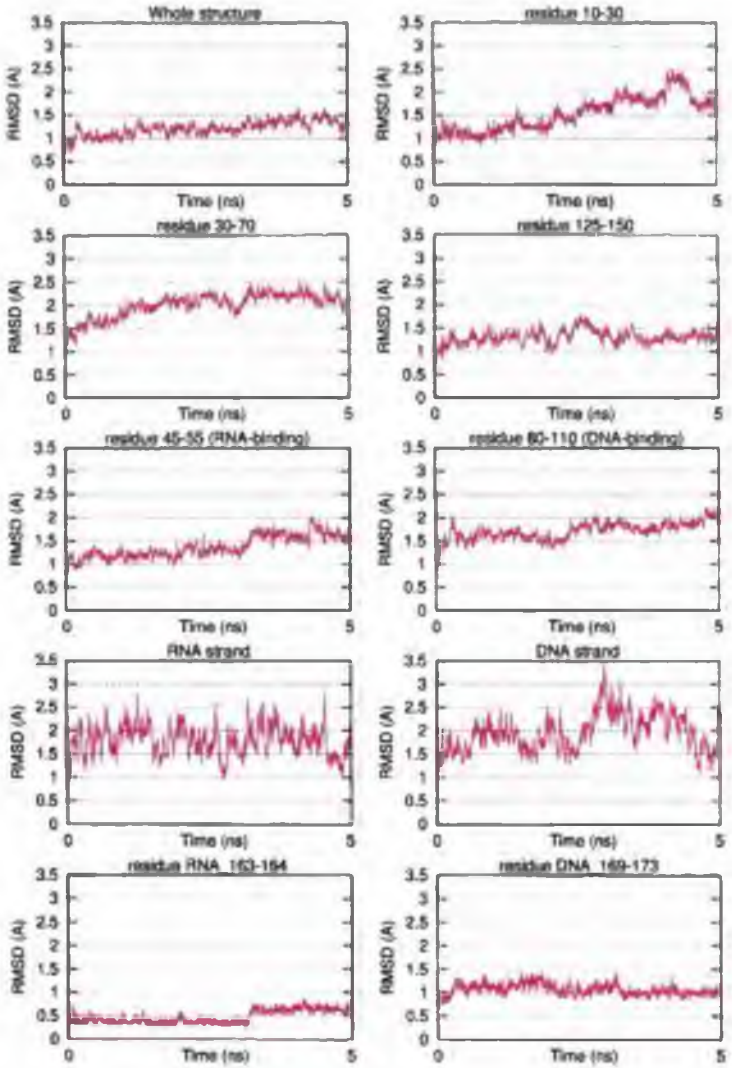


Figure C.17: RMSD of the whole structure and selected regions. Modification 3

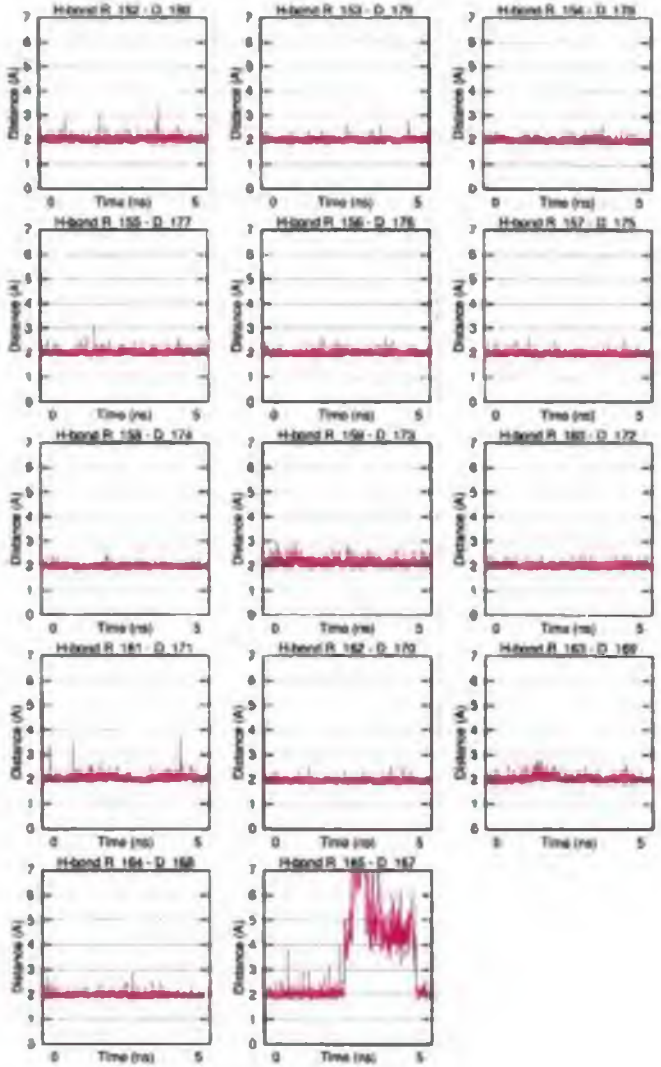


Figure C.18: The Watson-Crick hydrogen bonds. Modification 3

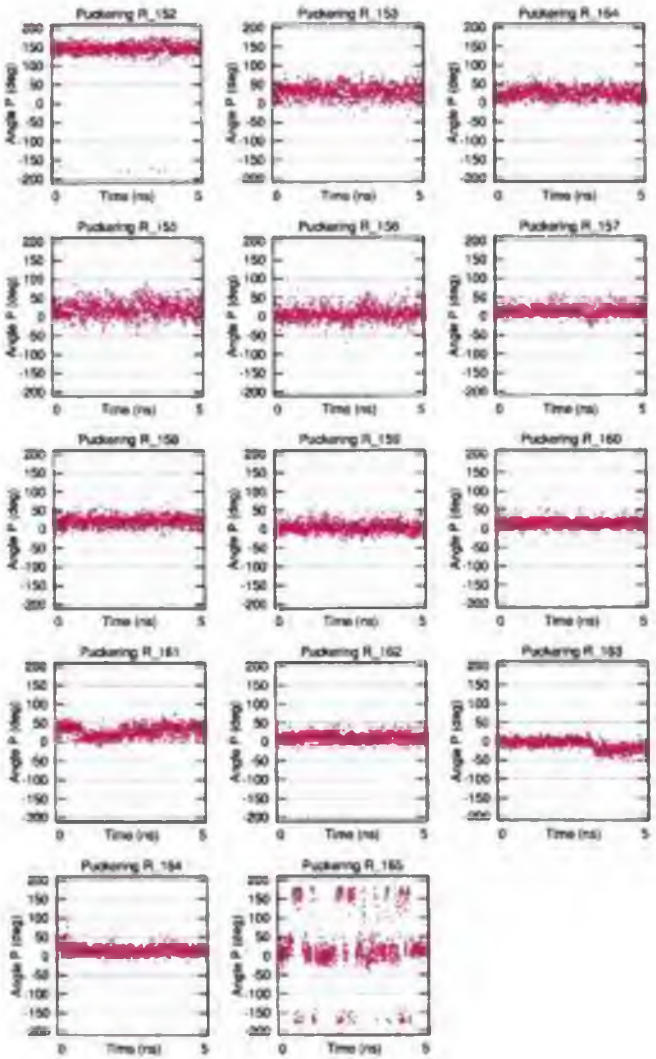


Figure C.19: Puckering of the RNA-strand. Modification 3

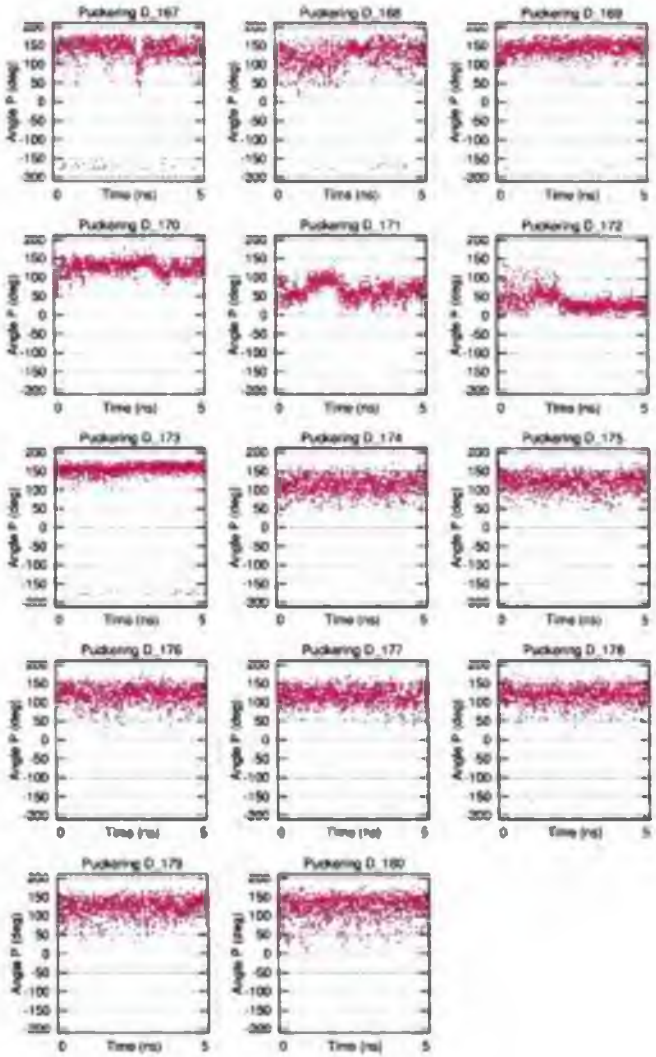


Figure C.20: Puckering of the DNA-strand. Modification 3

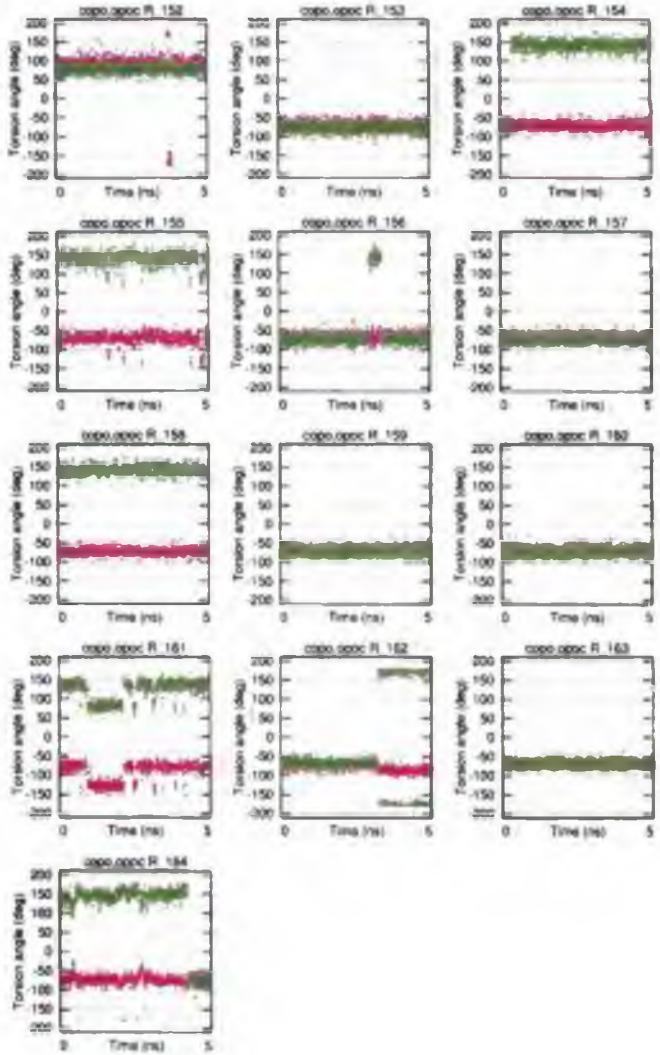


Figure C.21: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 3

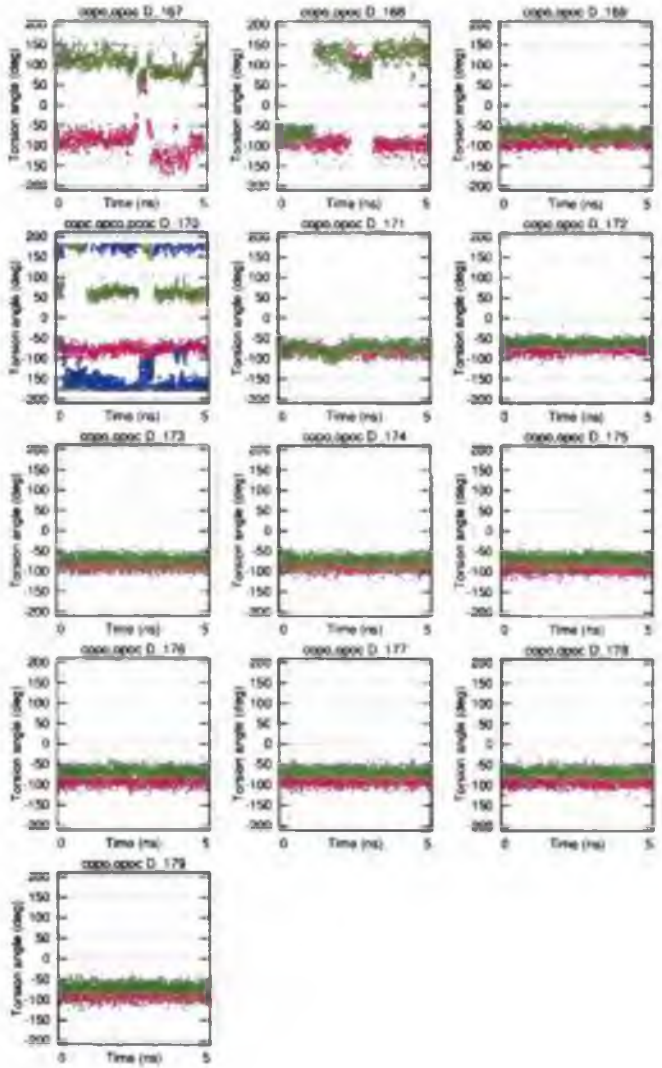


Figure C.22: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 3

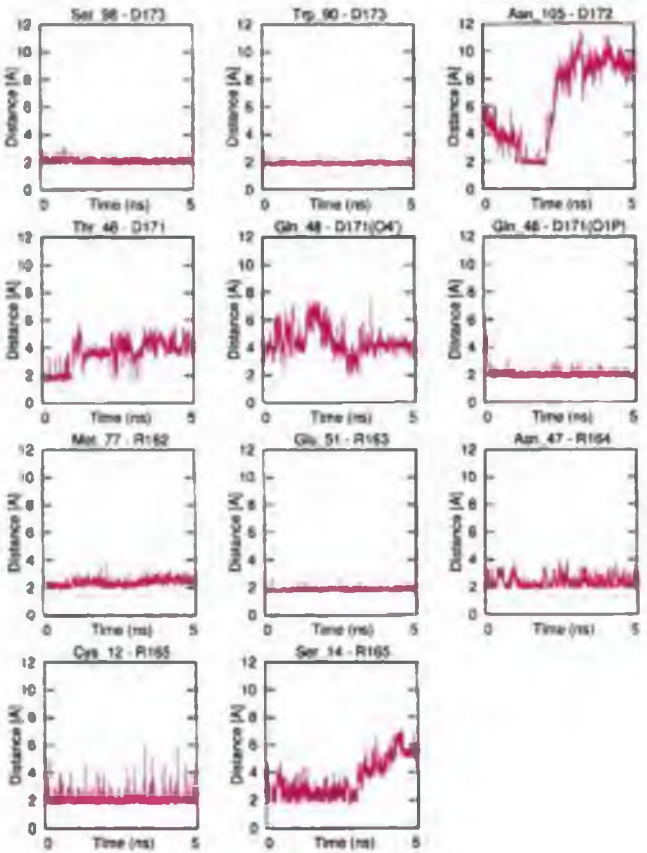


Figure C.23. The enzyme-substrate hydrogen bonds. Modification 3

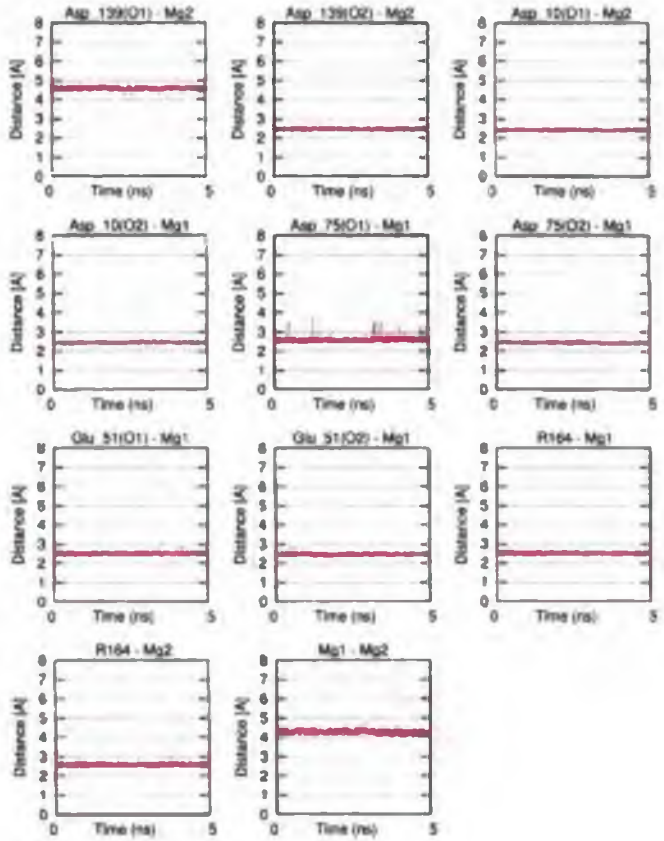


Figure C.24: Contacts between the Mg-ions and the DDE motif in the active site. Modification 3

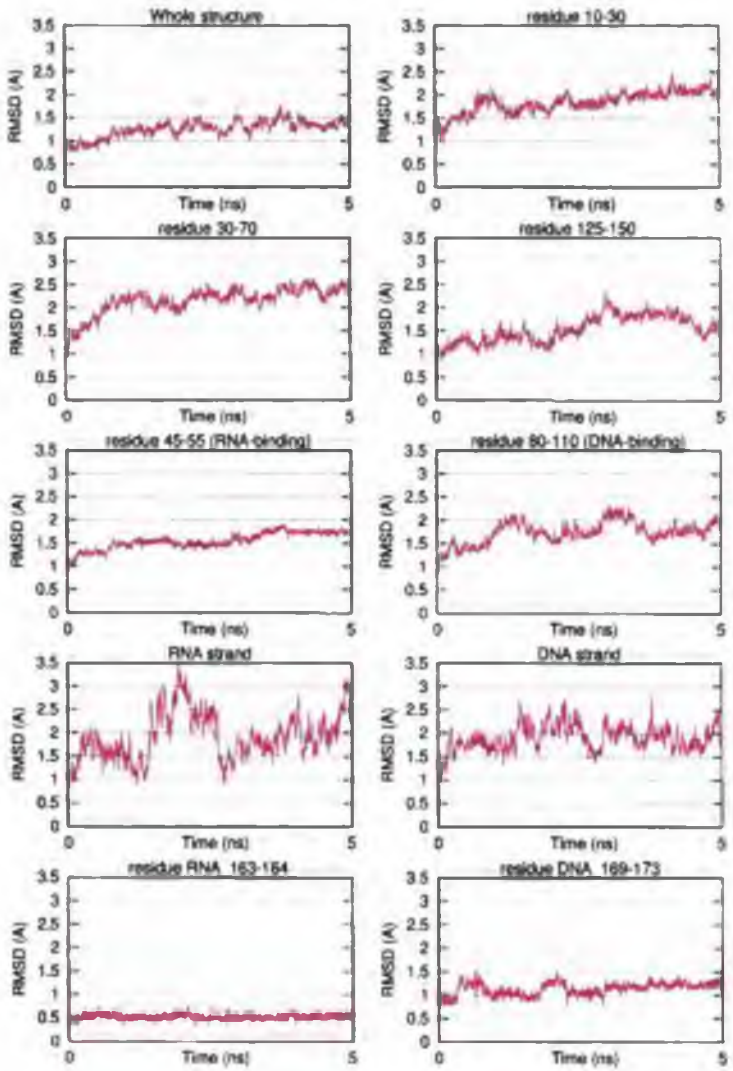


Figure C.25: RMSD of the whole structure and selected regions. Modification 4

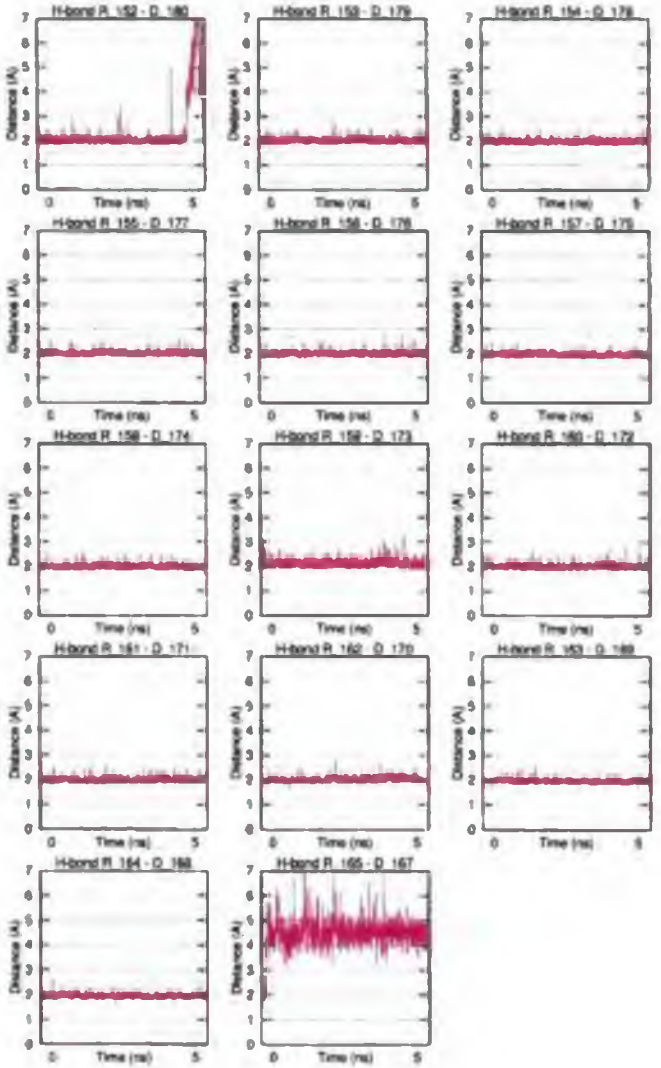


Figure C.26: The Watson-Crick hydrogen bonds. Modification 4

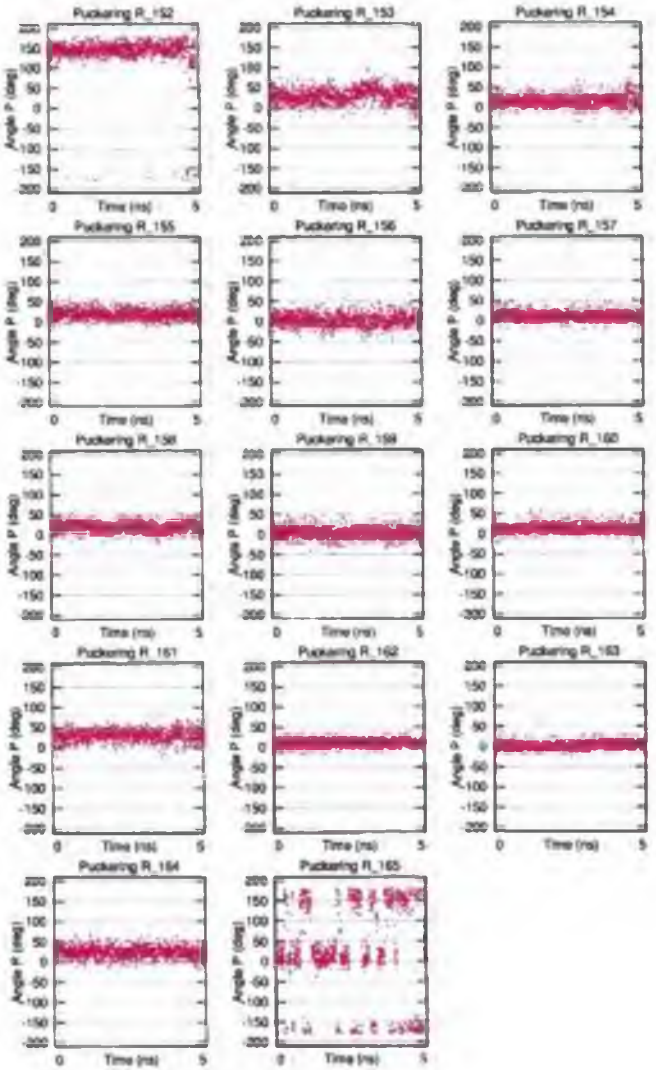


Figure C.27: Puckering of the RNA-strand. Modification 4

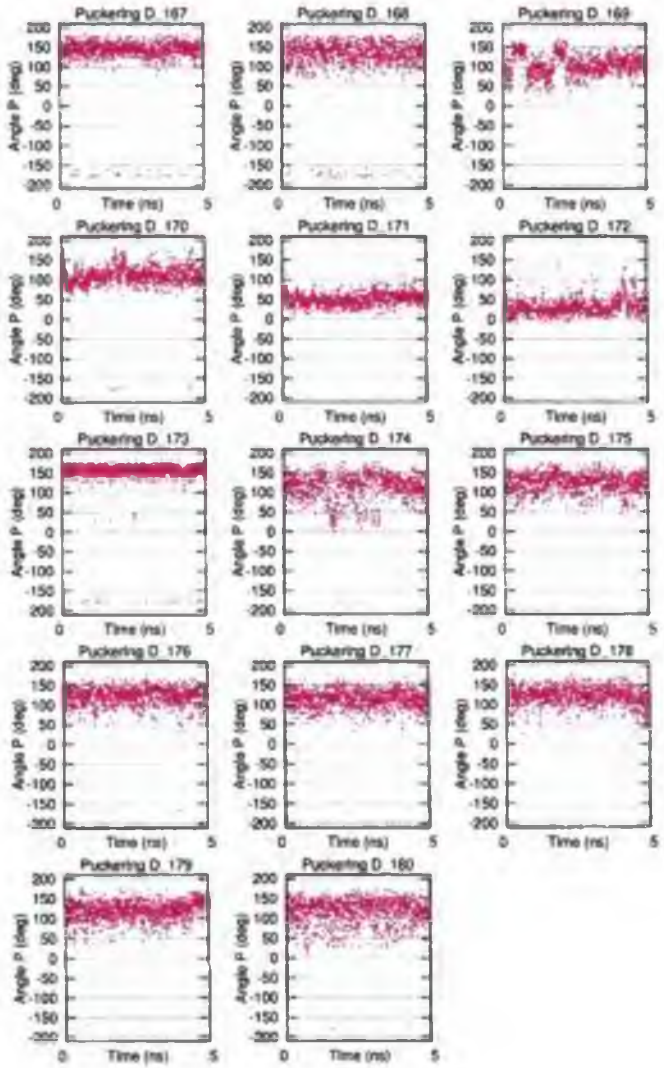


Figure C.28: Puckering of the DNA-strand. Modification 4

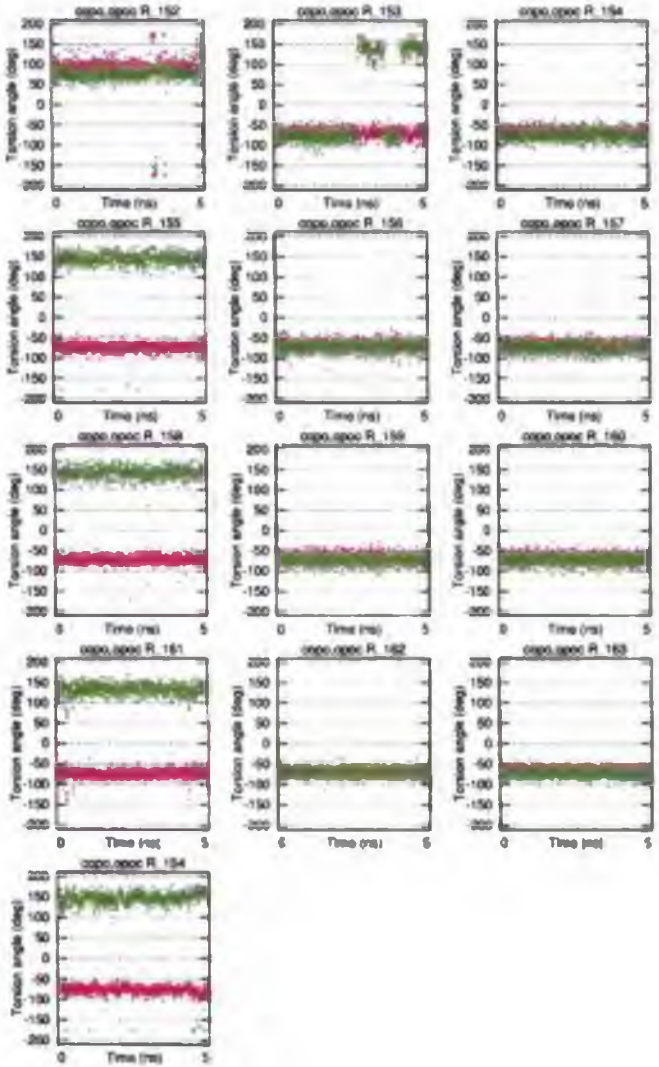


Figure C.29: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 4

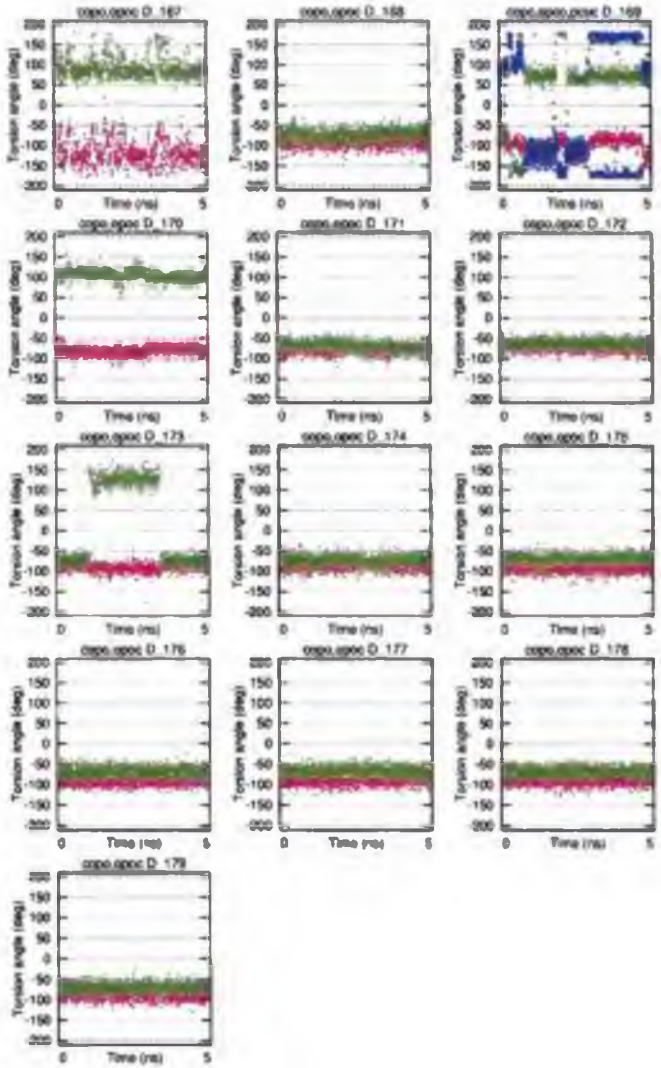


Figure C.30: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 4

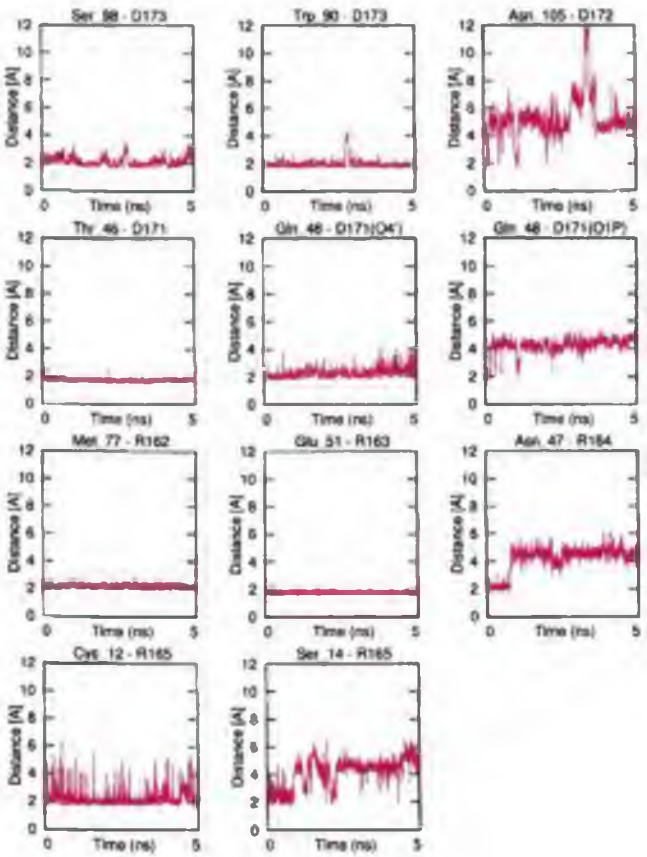


Figure C.31: The enzyme-substrate hydrogen bonds. Modification 4

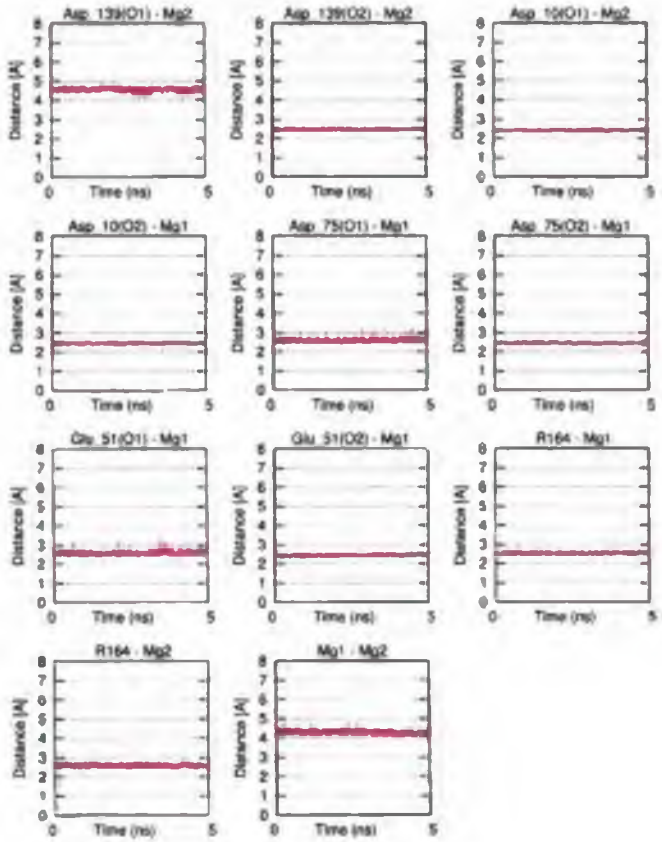


Figure C.32: Contacts between the Mg-ions and the DDE motif in the active site. Modification 4

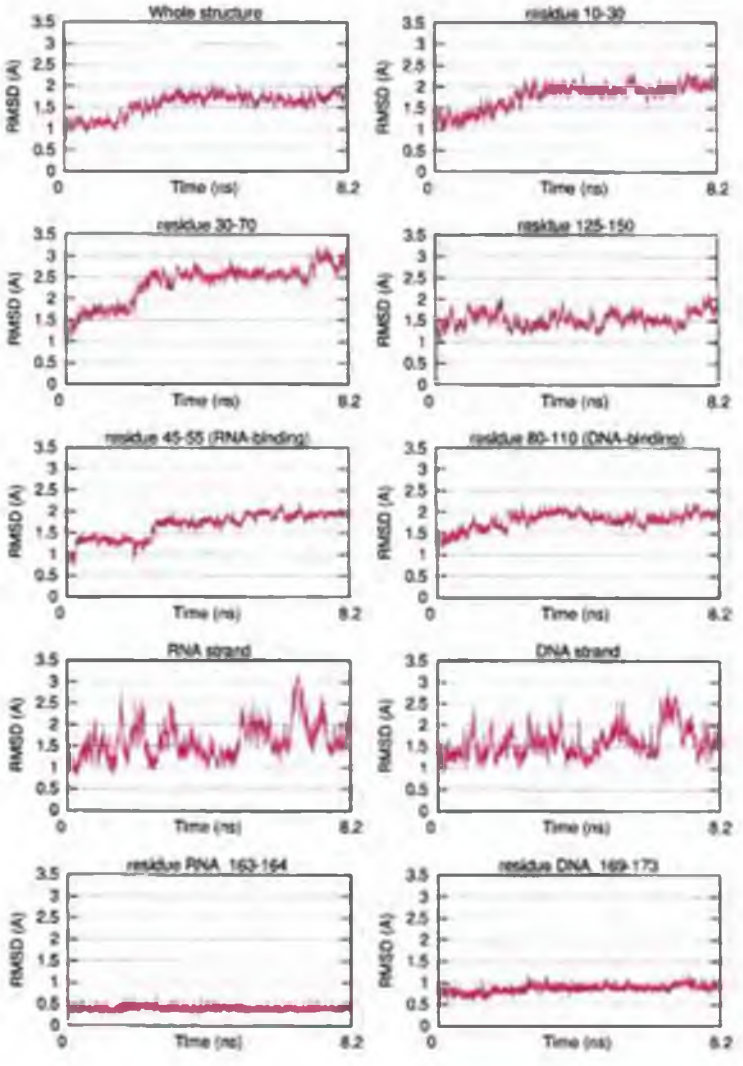


Figure C.33: RMSD of the whole structure and selected regions. Modification 5

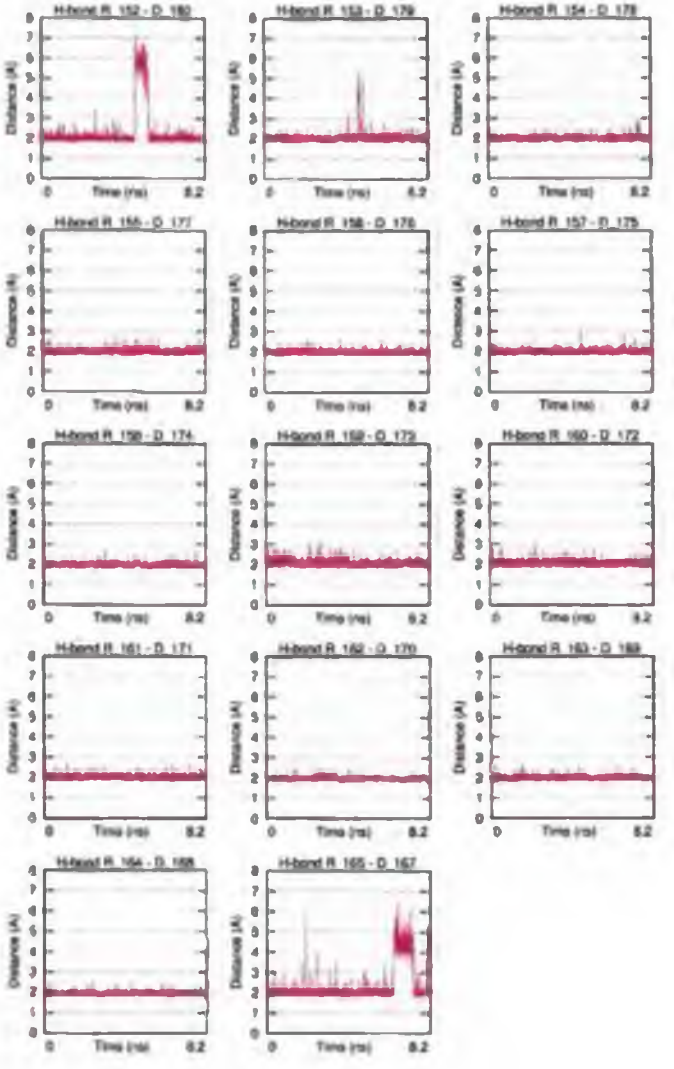


Figure C.34: The Watson-Crick hydrogen bonds. Modification 5

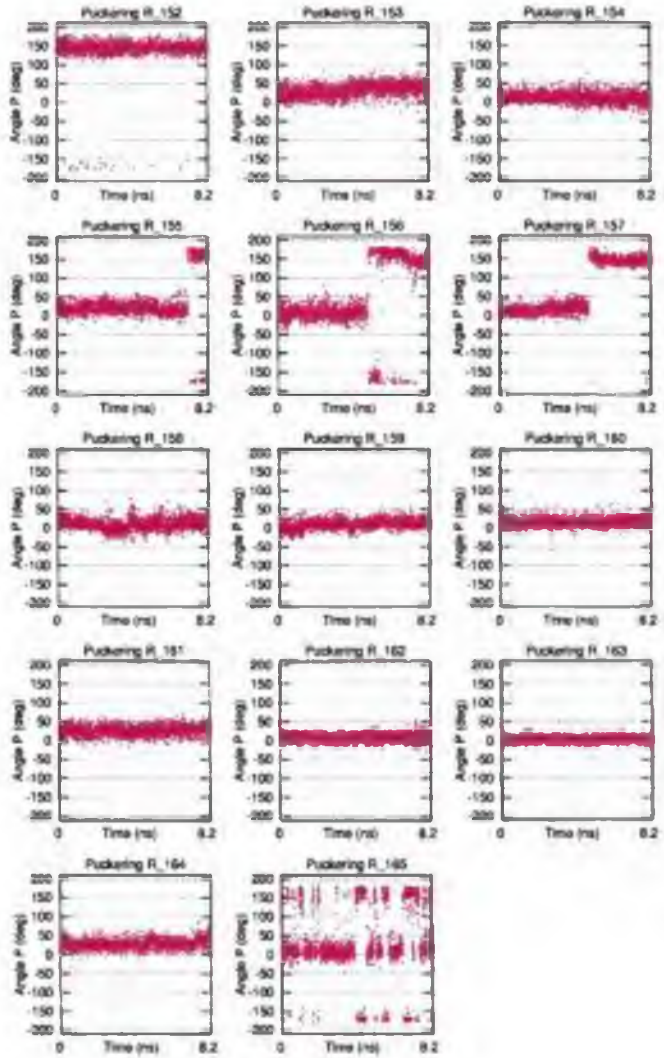


Figure C.35: Puckering of the RNA-strand. Modification 5

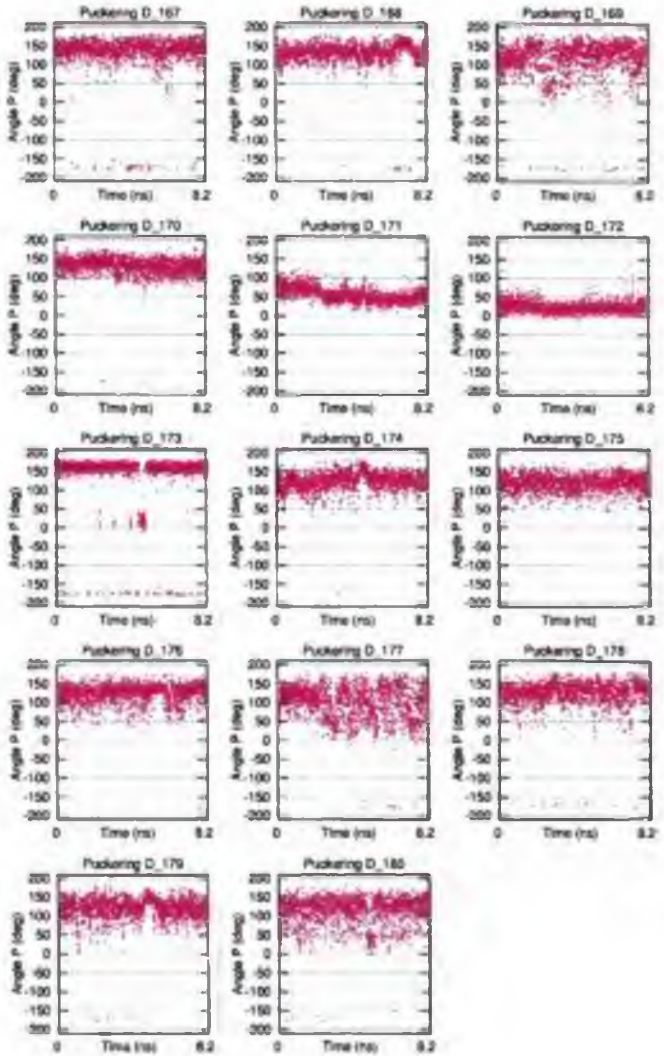


Figure C.36: Puckering of the DNA-strand. Modification 5

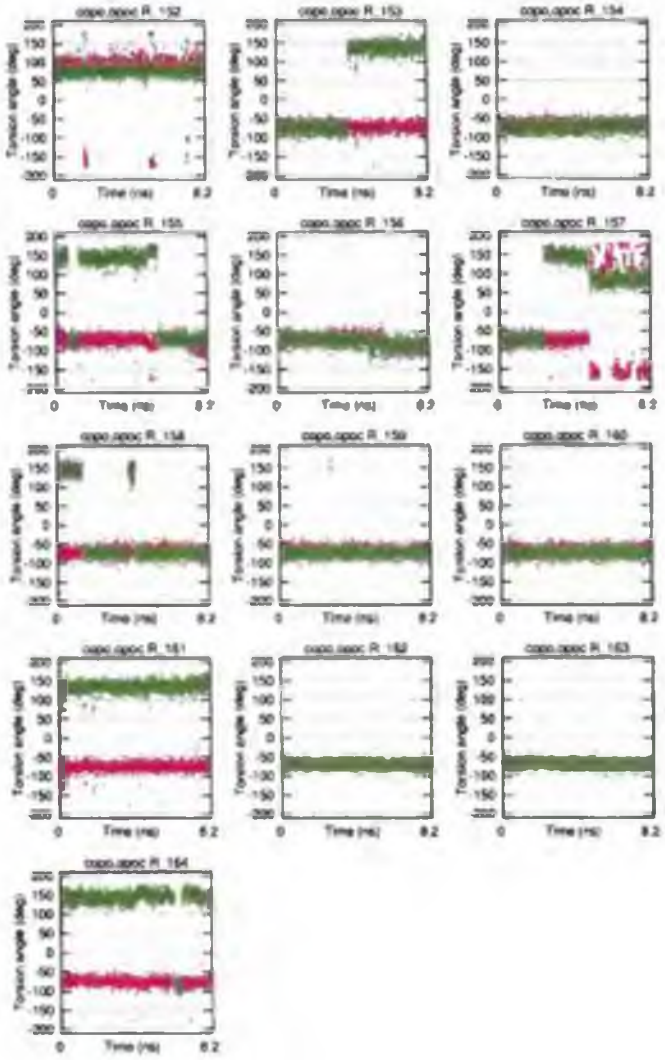


Figure C.37: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles: the RNA-strand. Modification 5

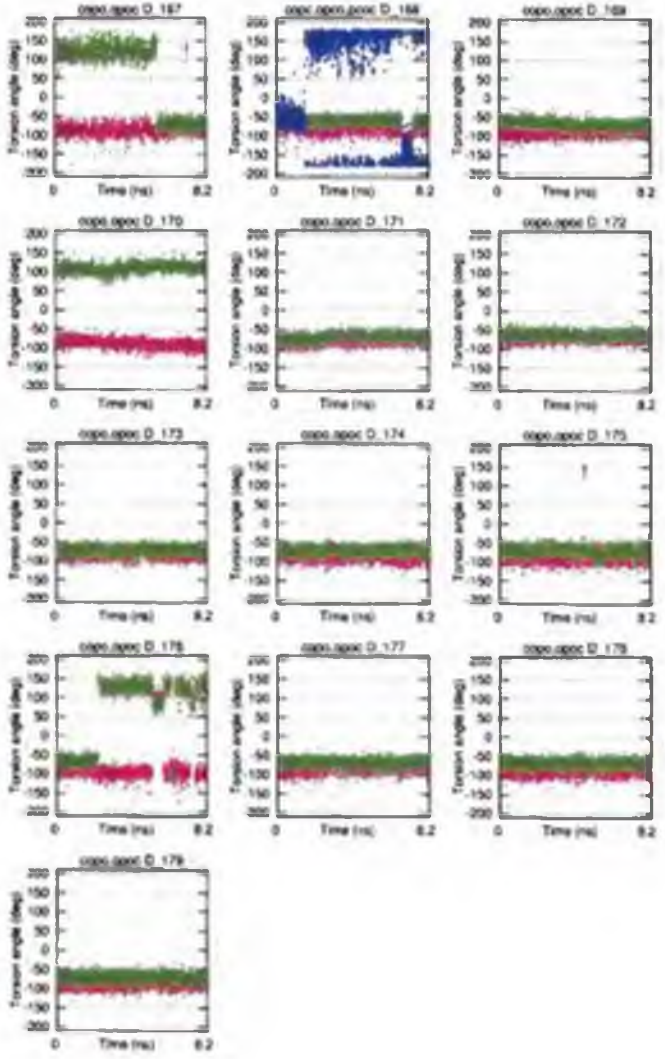


Figure C.38: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 5

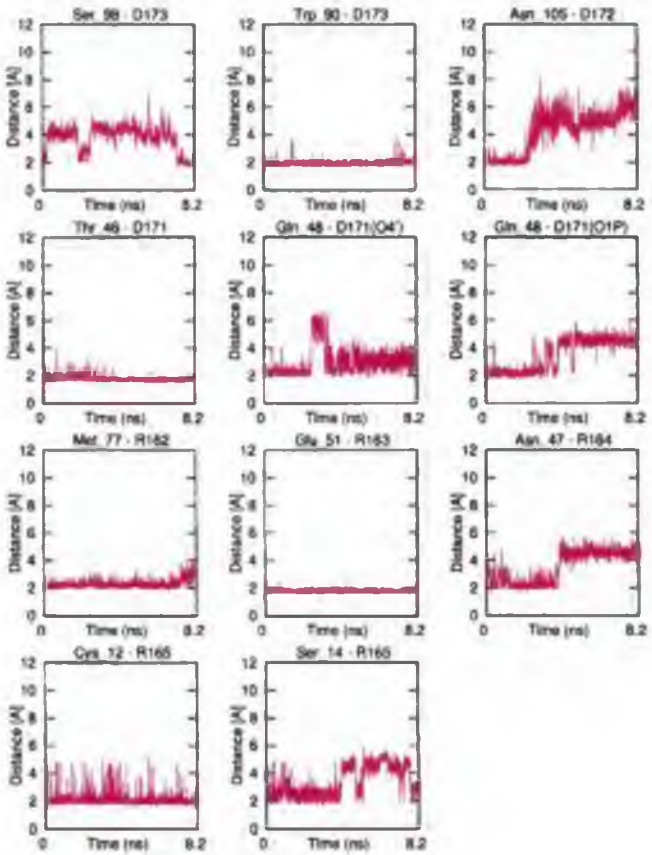


Figure C.39: The enzyme-substrate hydrogen bonds. Modification 5

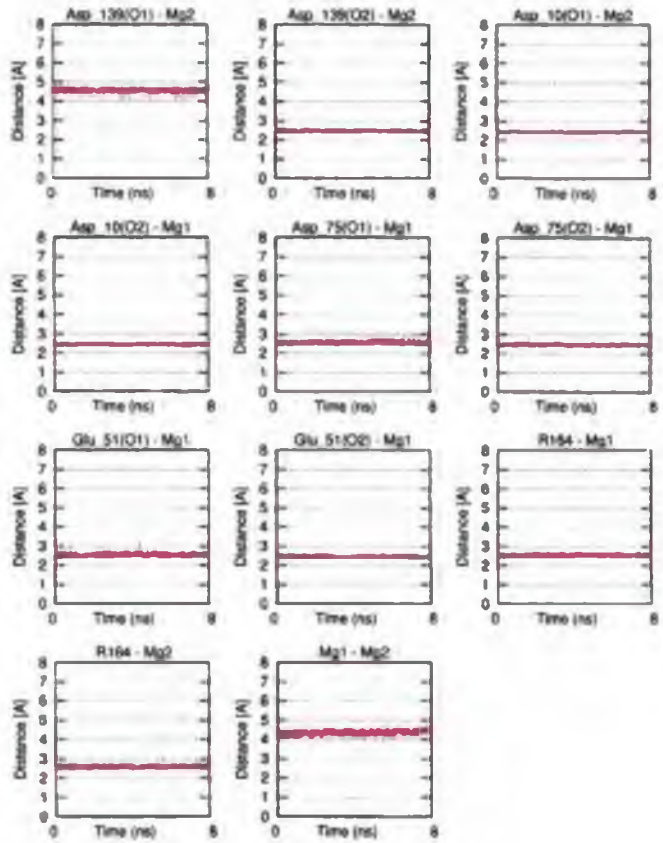
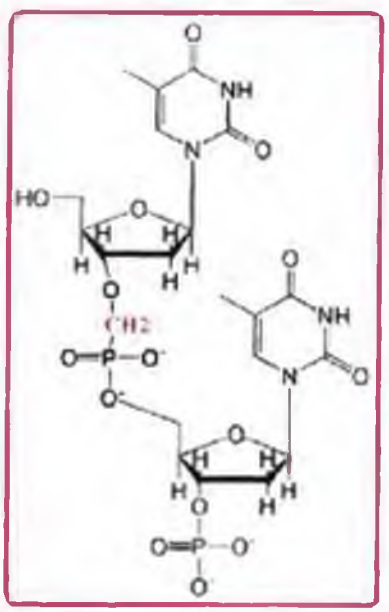


Figure C.40: Contacts between the Mg-ions and the DDE motif in the active site. Modification 5

Appendix D

Graphs - models with a modified substrate

The C3'-O-*C*-P-O-C5' modification



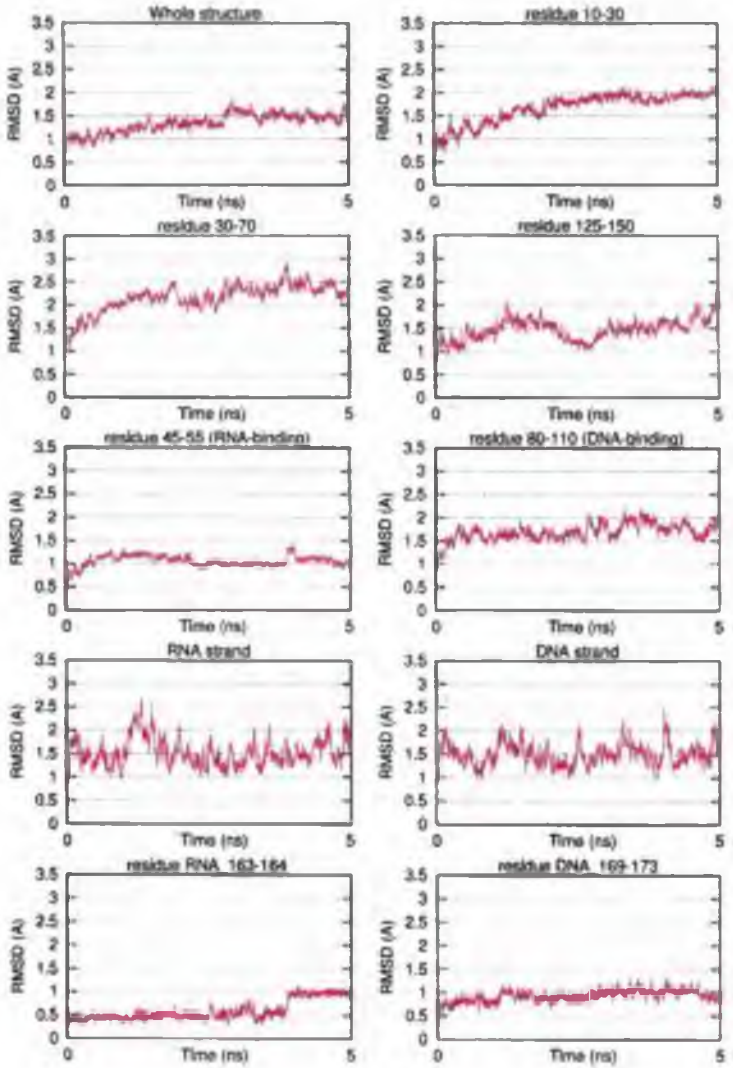


Figure D.1: RMSD of the whole structure and selected regions. Modification 6

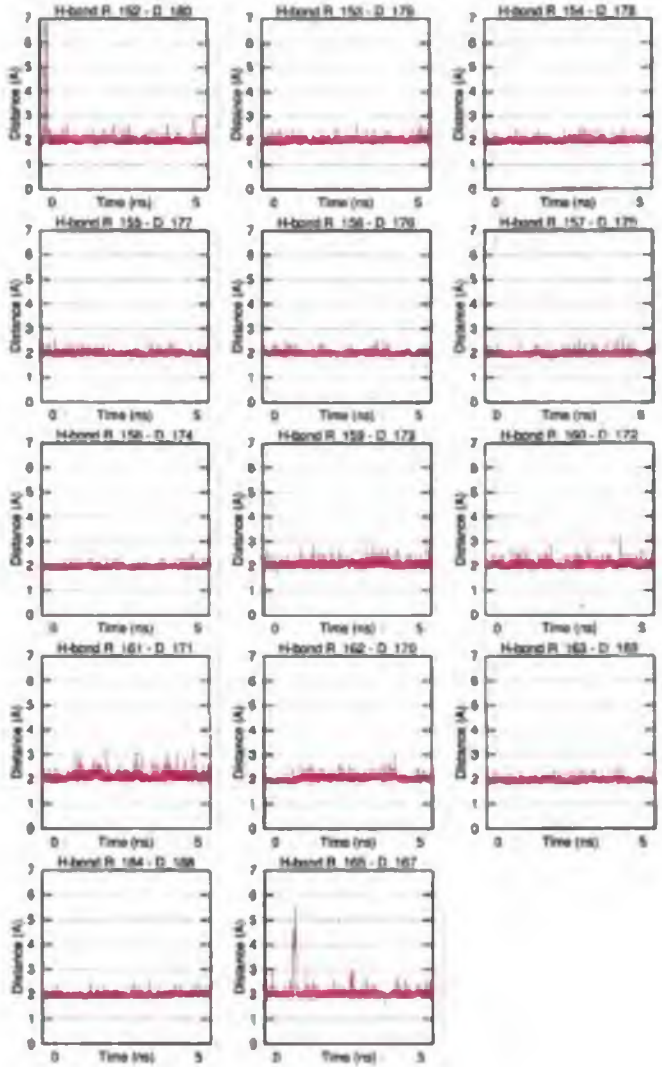


Figure D.2: The Watson-Crick hydrogen bonds. Modification 6

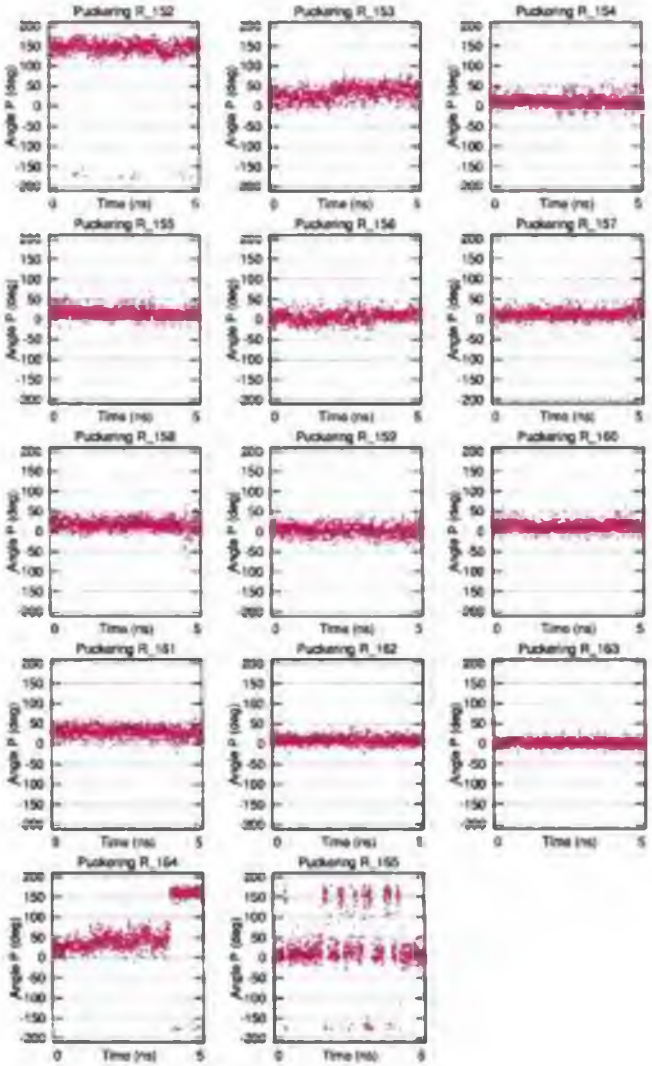


Figure D.3: Puckering of the RNA-strand. Modification 6

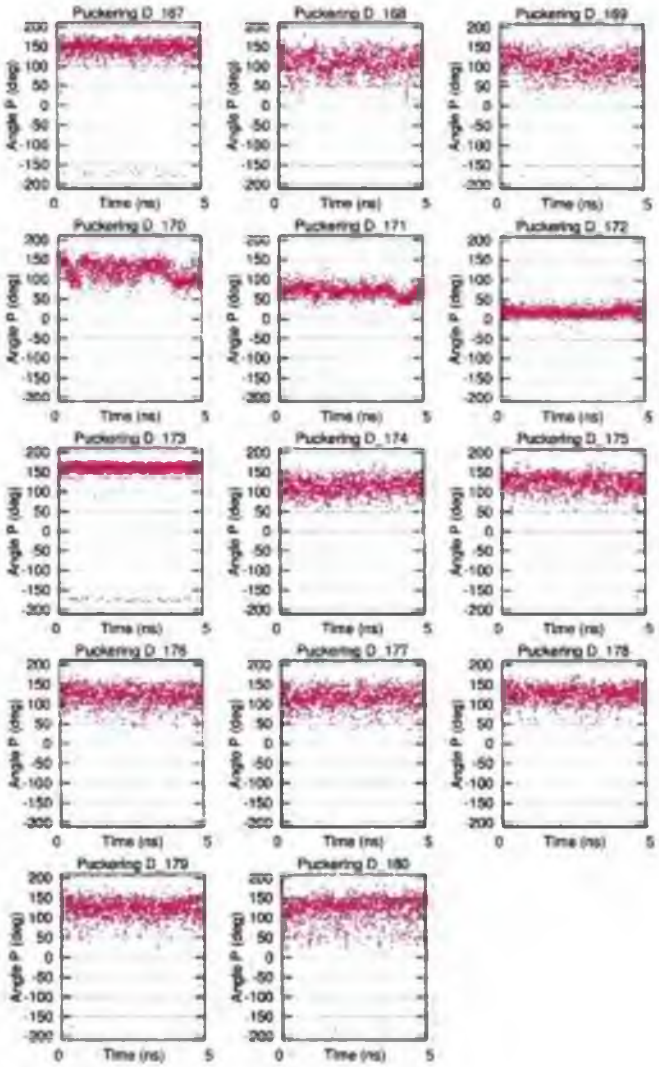


Figure D.4: Puckering of the DNA-strand. Modification 6

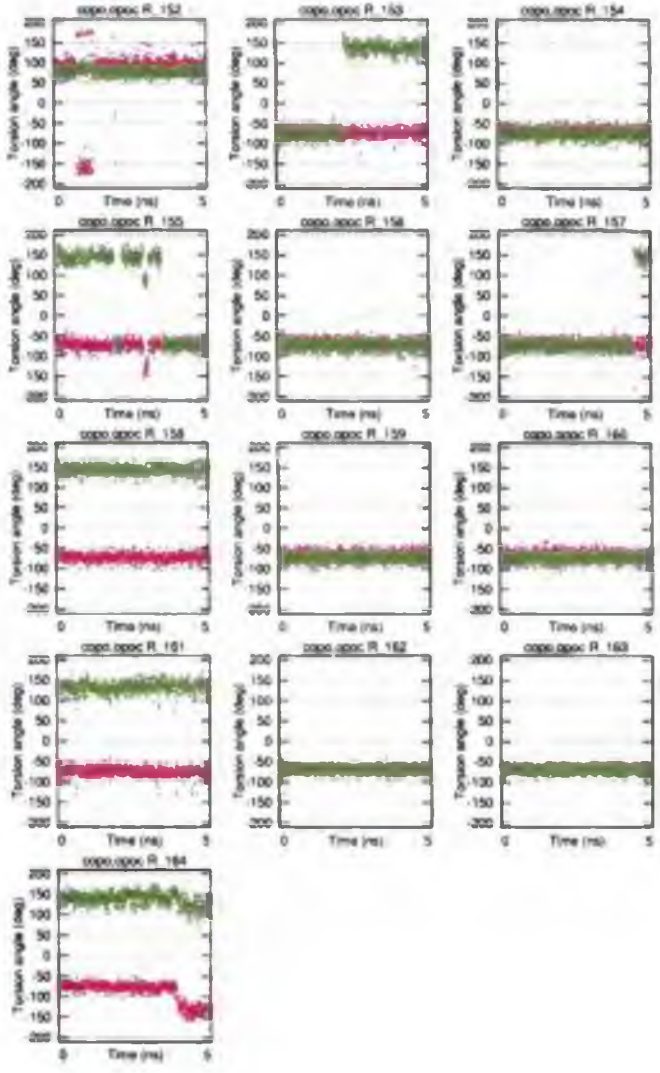


Figure D.5: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 6

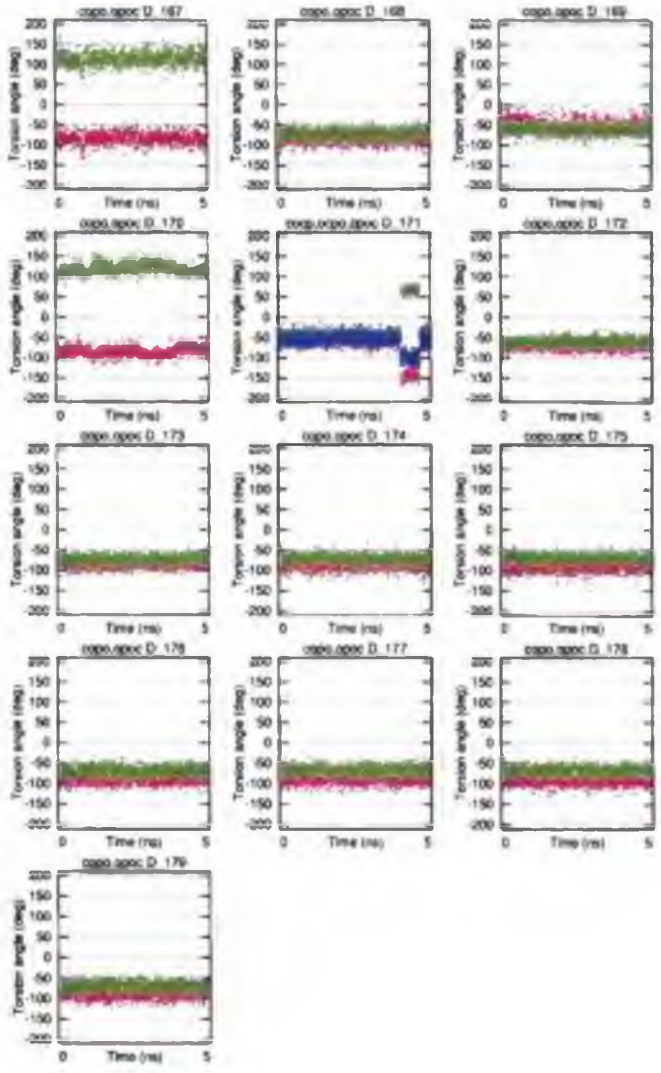


Figure D.6: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 6

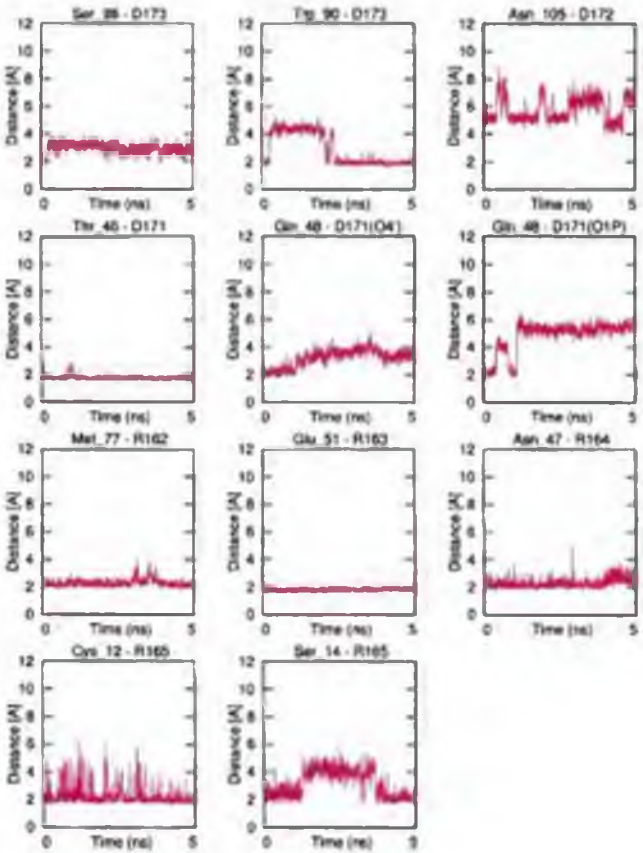


Figure D.7: The enzyme-substrate hydrogen bonds. Modification 6

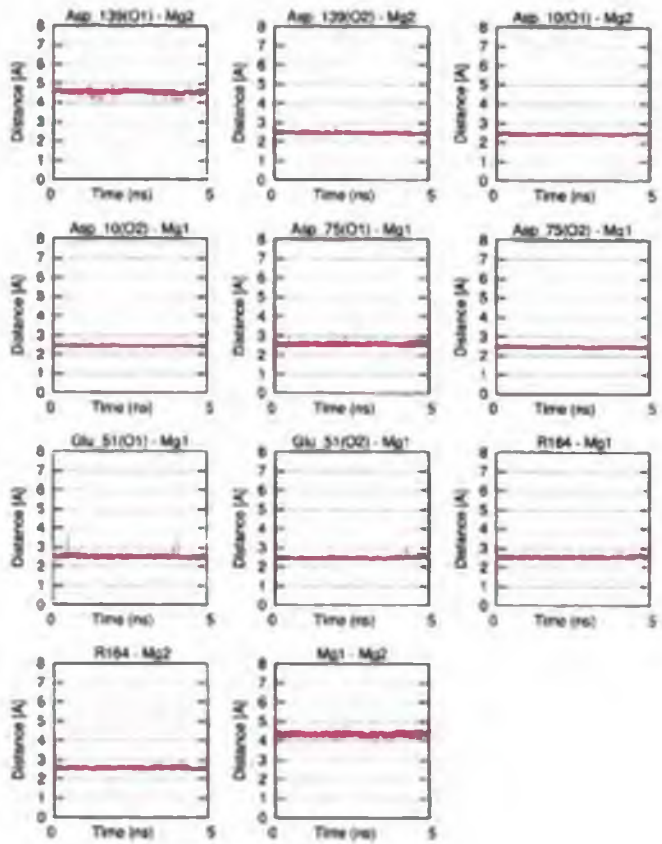


Figure D.8: Contacts between the Mg-ions and the DDE motif in the active site. Modification 6

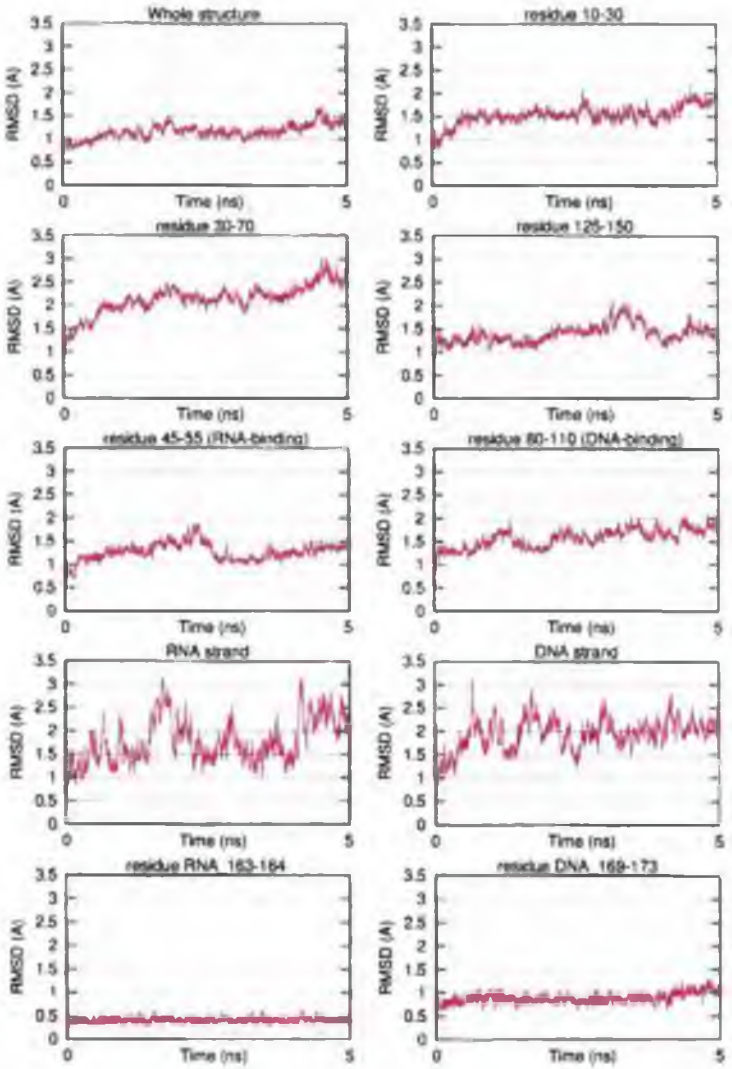


Figure D.9: RMSD of the whole structure and selected regions. Modification 7

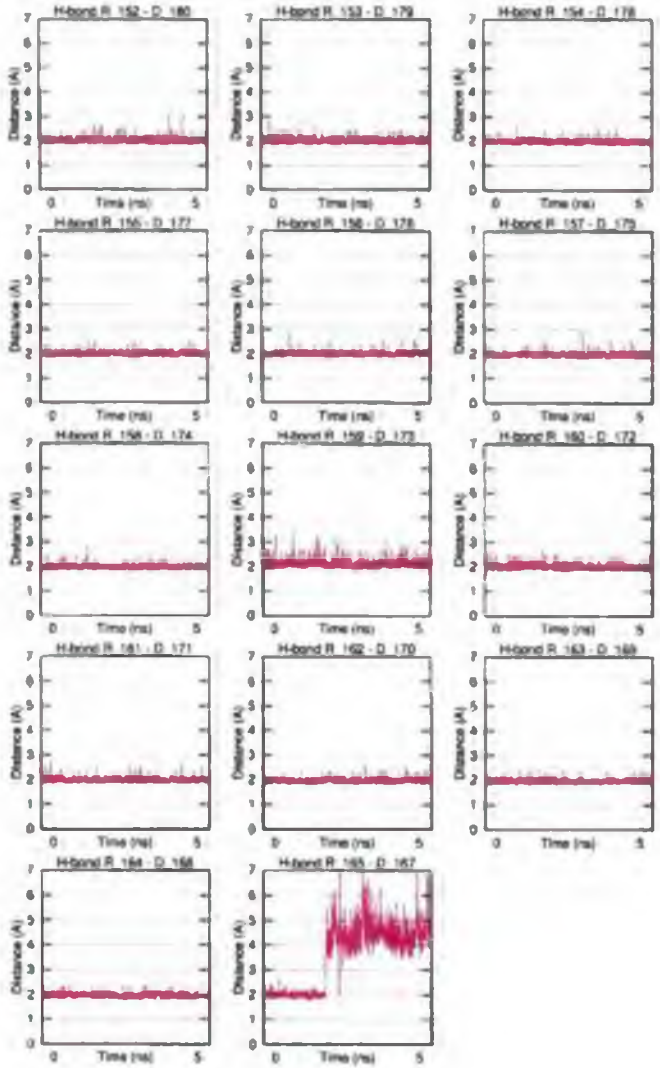


Figure D.10: The Watson-Crick hydrogen bonds. Modification 7

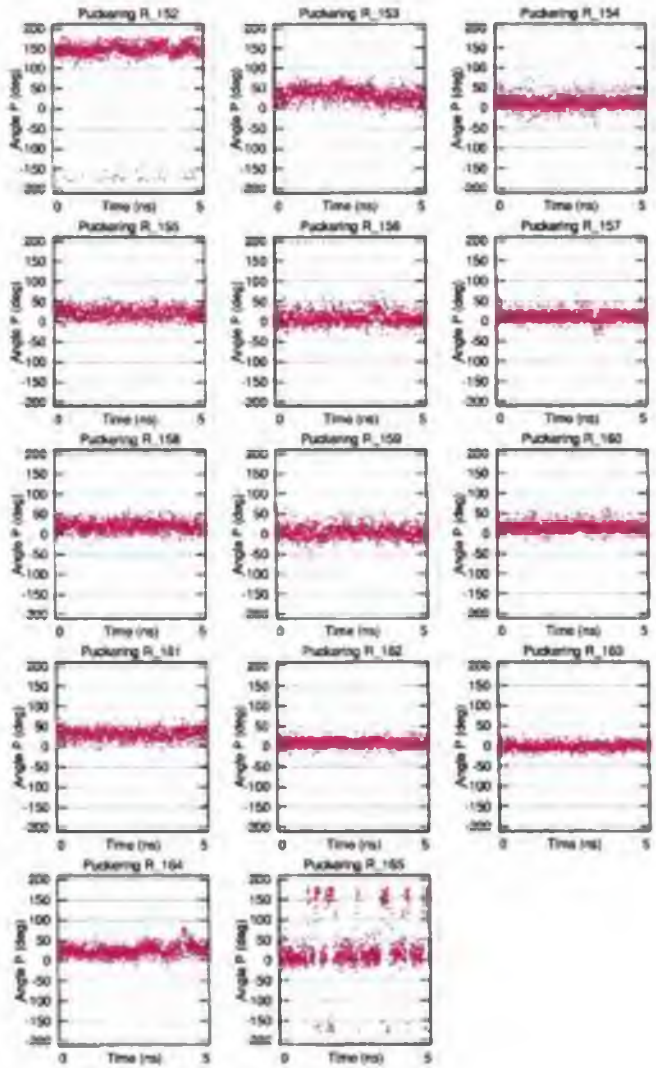


Figure D.11: Puckering of the RNA-strand. Modification 7

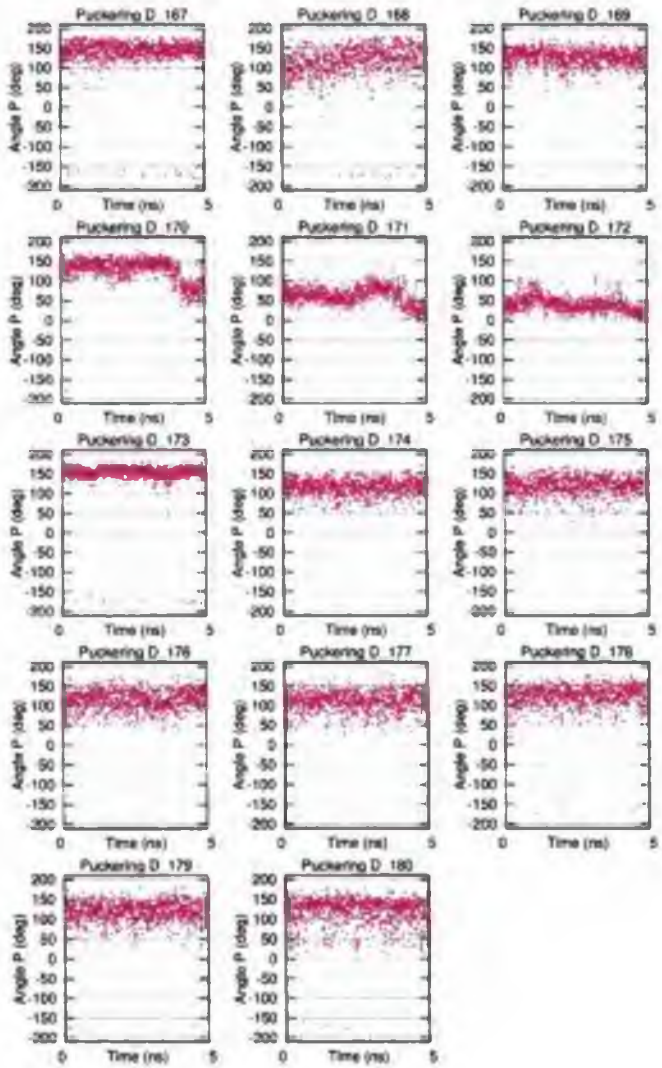


Figure D.12: Puckering of the DNA-strand. Modification 7

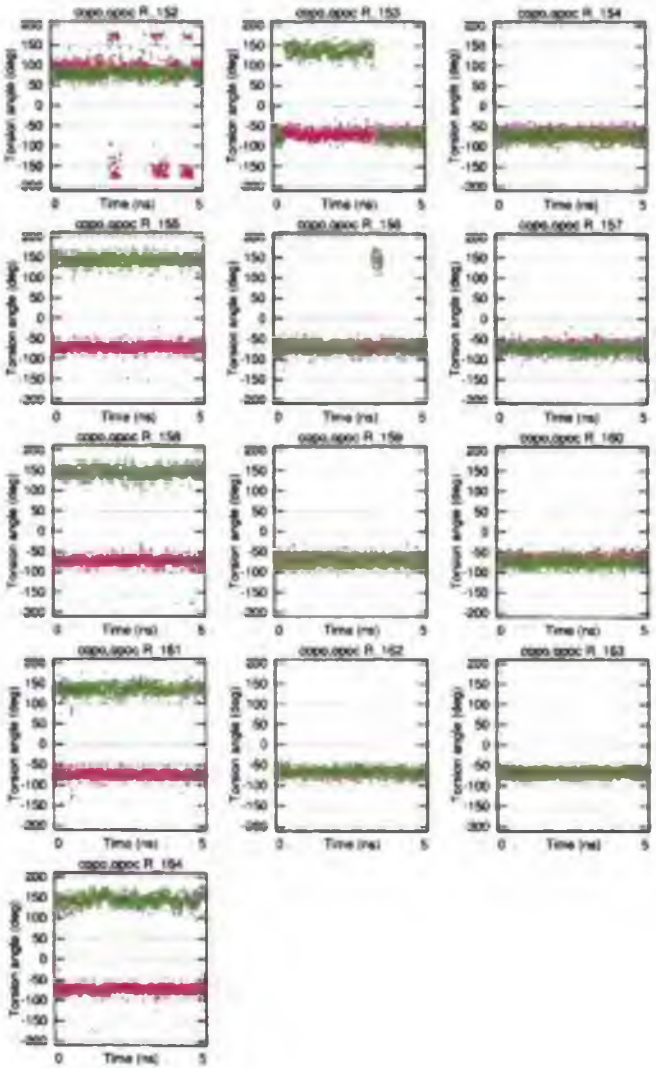


Figure D.13: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand Modification 7

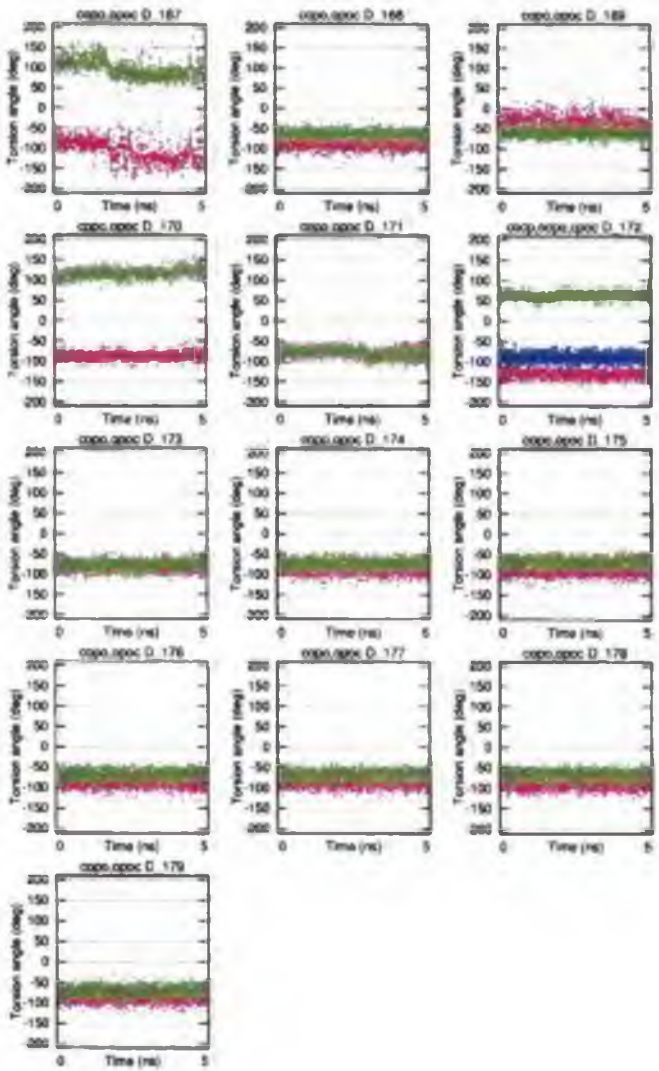


Figure D.14: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 7

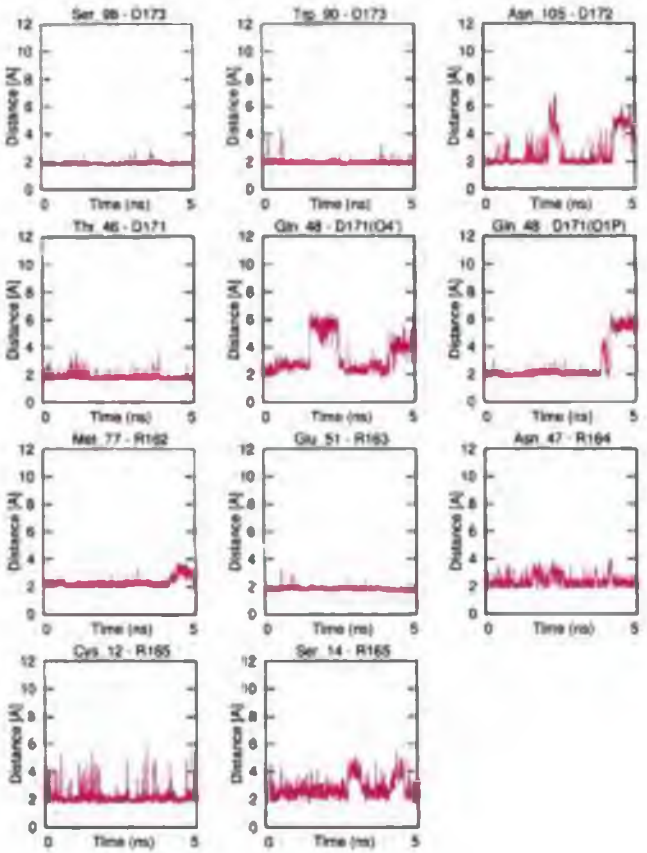


Figure D.15: The enzyme-substrate hydrogen bonds. Modification 7

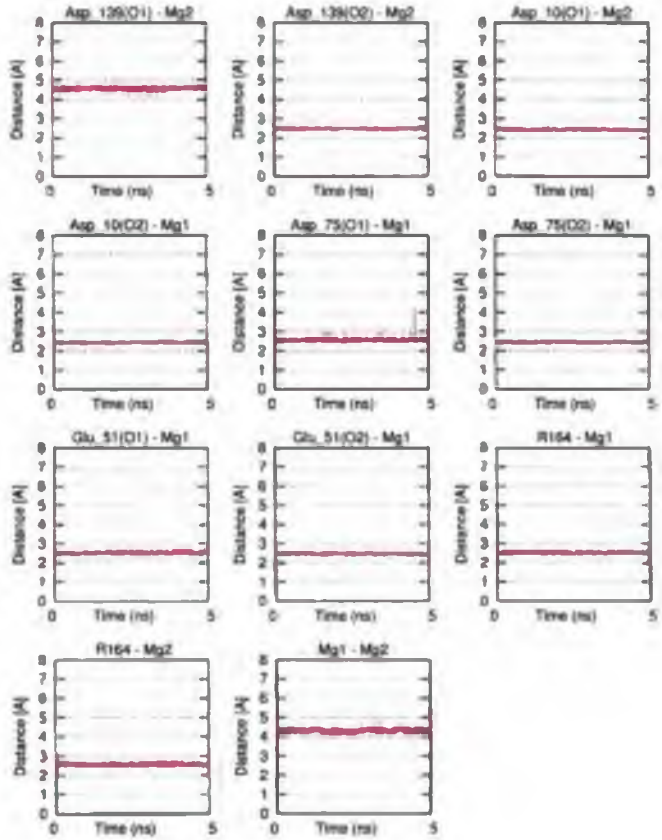


Figure D.16: Contacts between the Mg-ions and the DDE motif in the active site. Modification 7

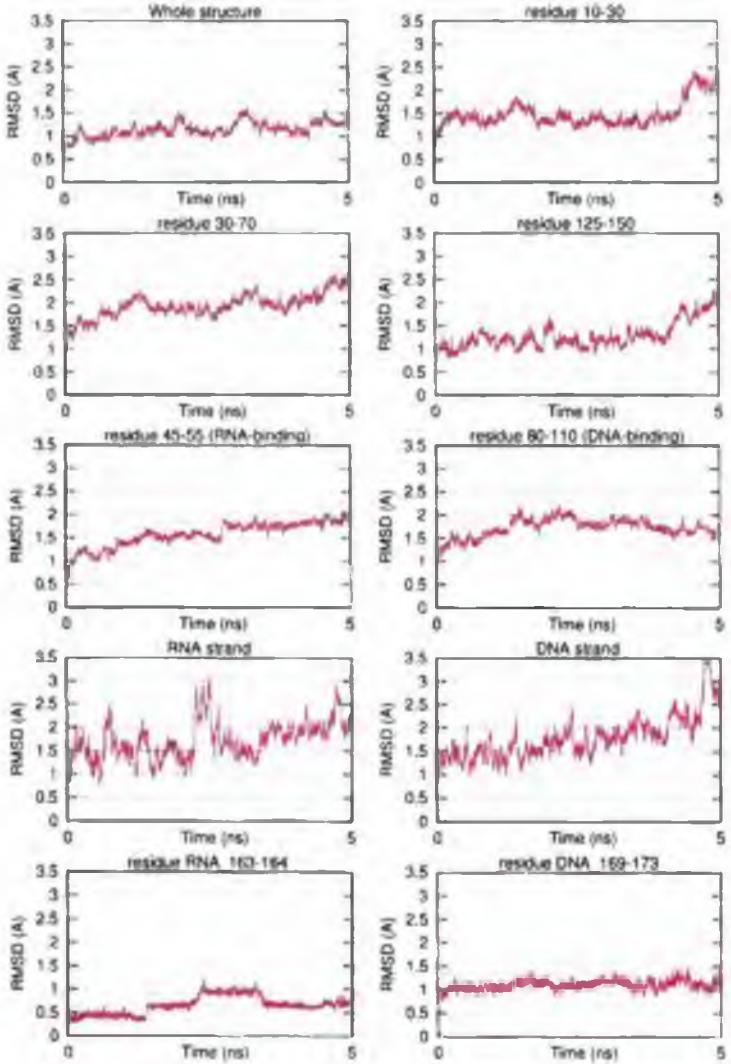


Figure D.17: RMSD of the whole structure and selected regions. Modification 8

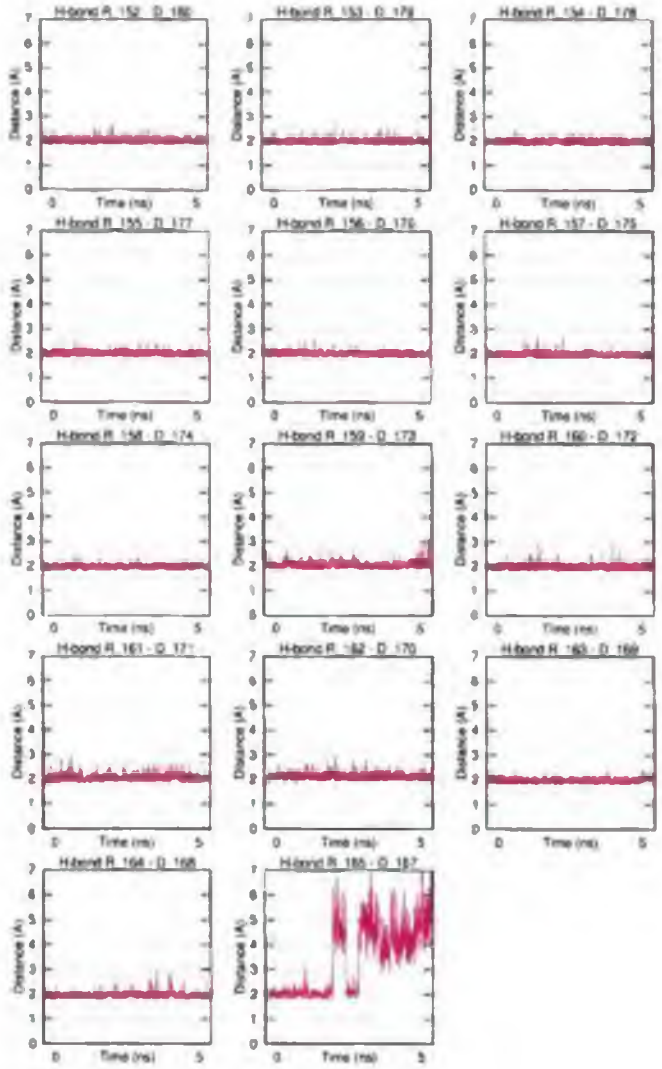


Figure D.18: The Watson-Crick hydrogen bonds. Modification 8

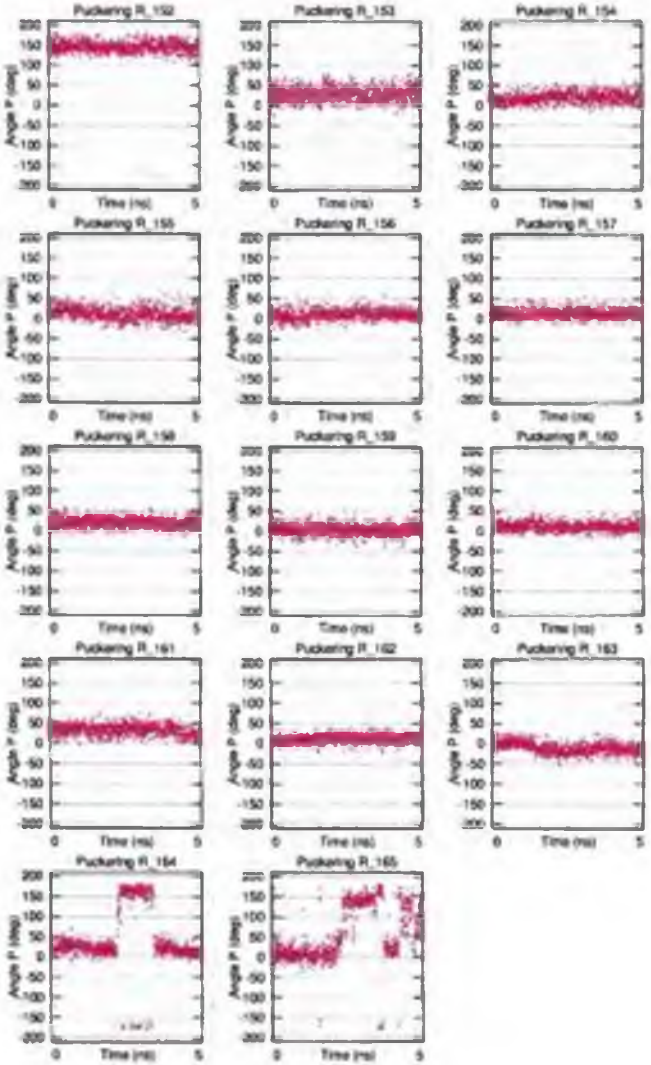


Figure D.19: Puckering of the RNA-strand. Modification 8

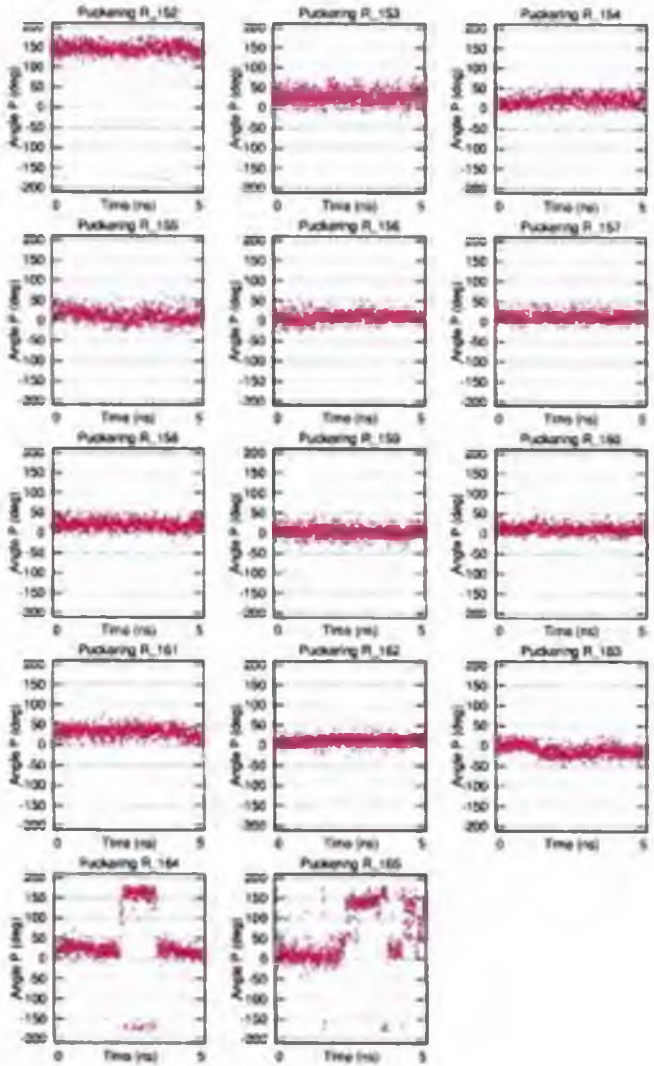


Figure D.20: Puckering of the DNA-strand. Modification 8

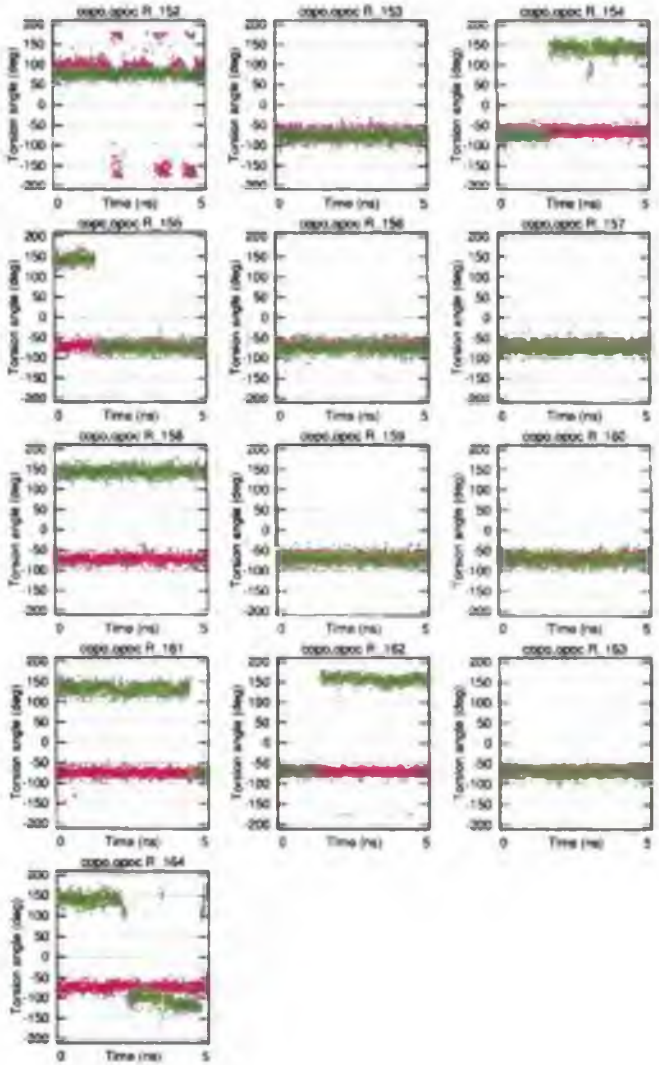


Figure D.21: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand Modification 8

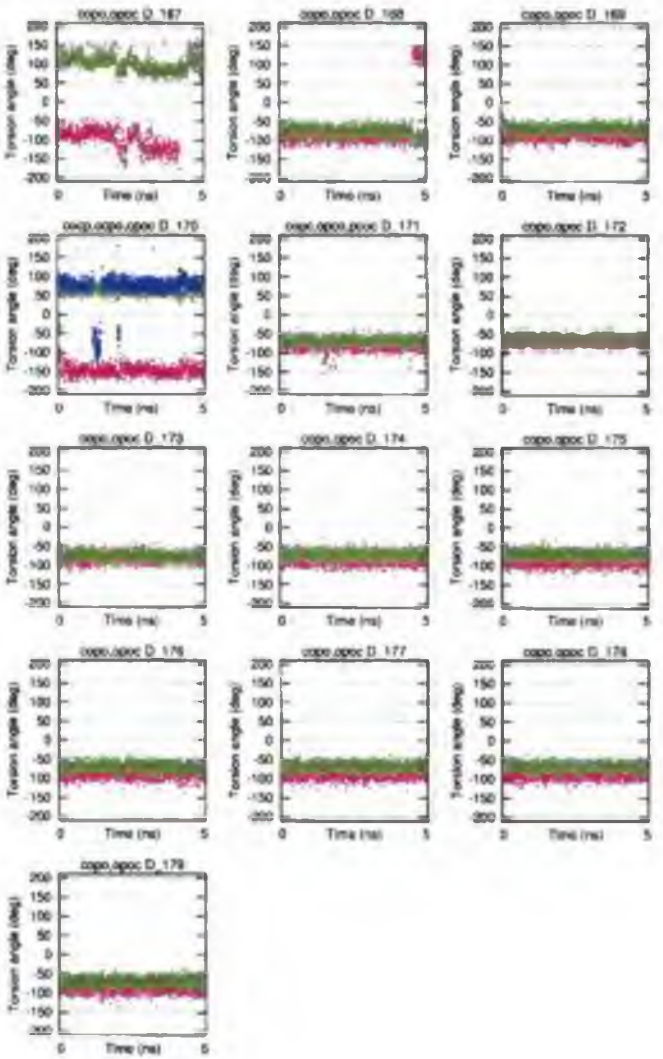


Figure D.22: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 8

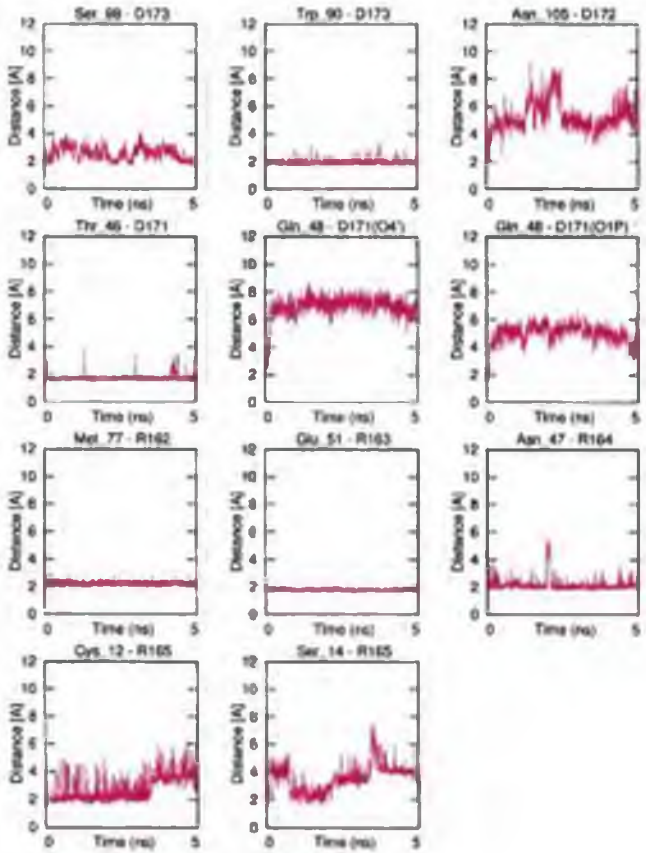


Figure D.23: The enzyme-substrate hydrogen bonds. Modification 8

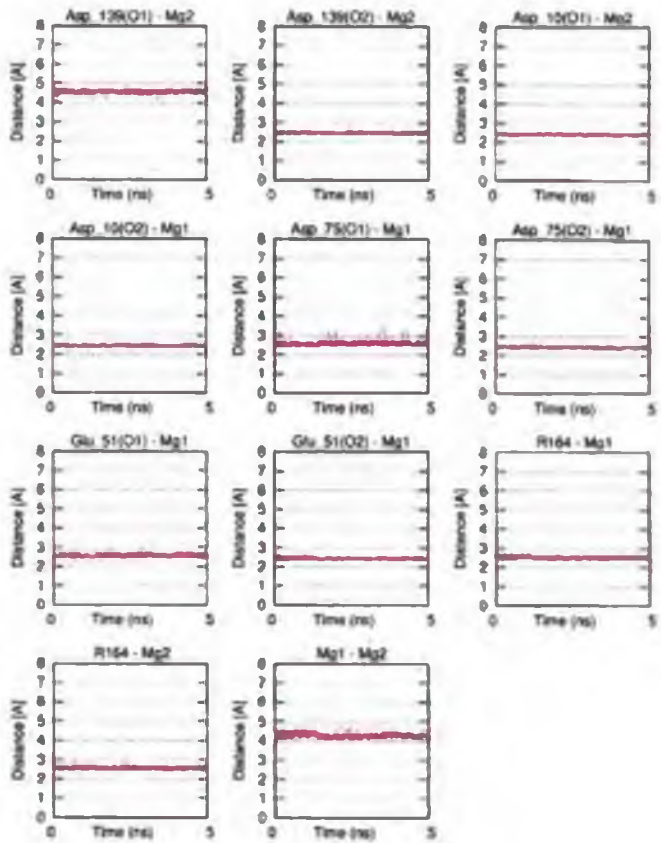


Figure D.24: Contacts between the Mg-ions and the DDE motif in the active site. Modification 8

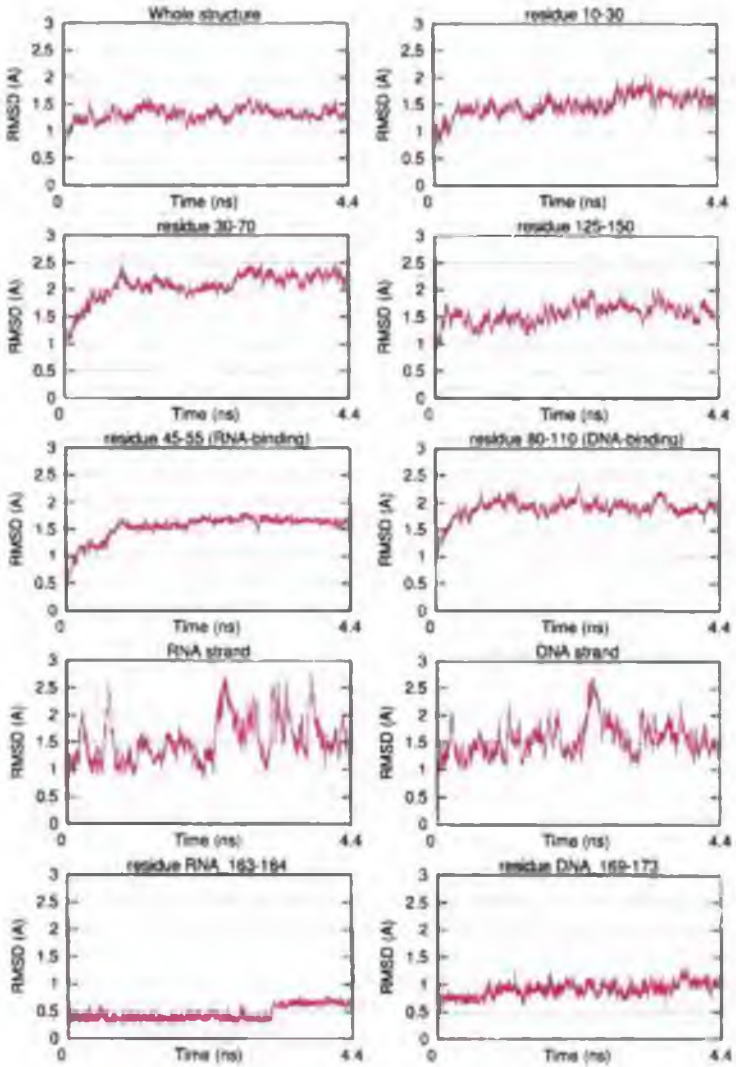


Figure D.25: RMSD of the whole structure and selected regions. Modification 9

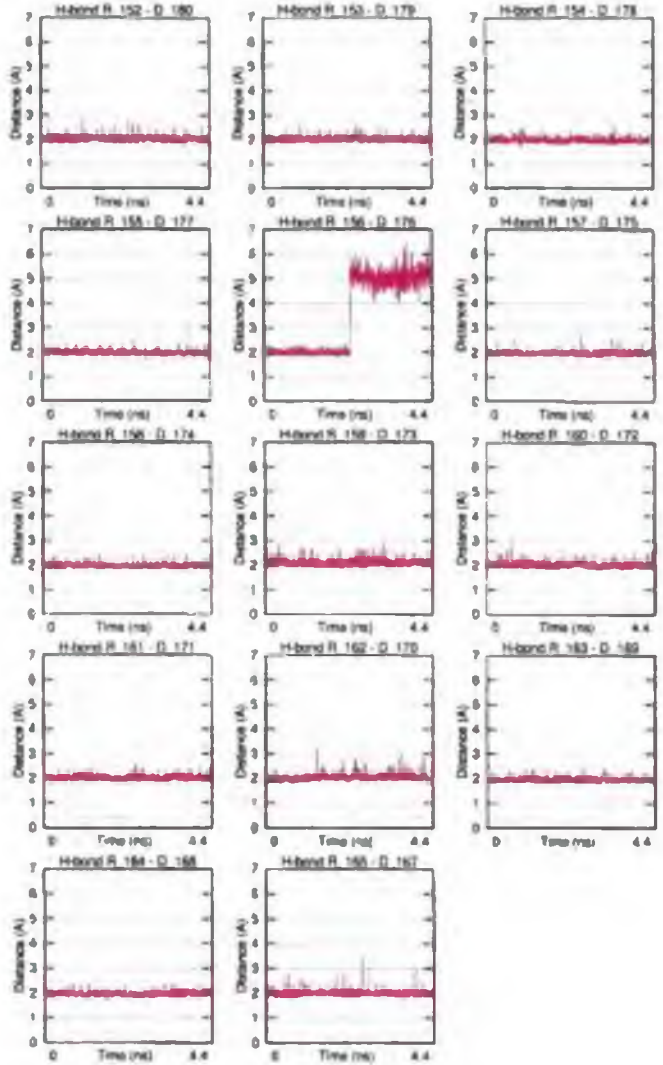


Figure D.26: The Watson-Crick hydrogen bonds. Modification 9

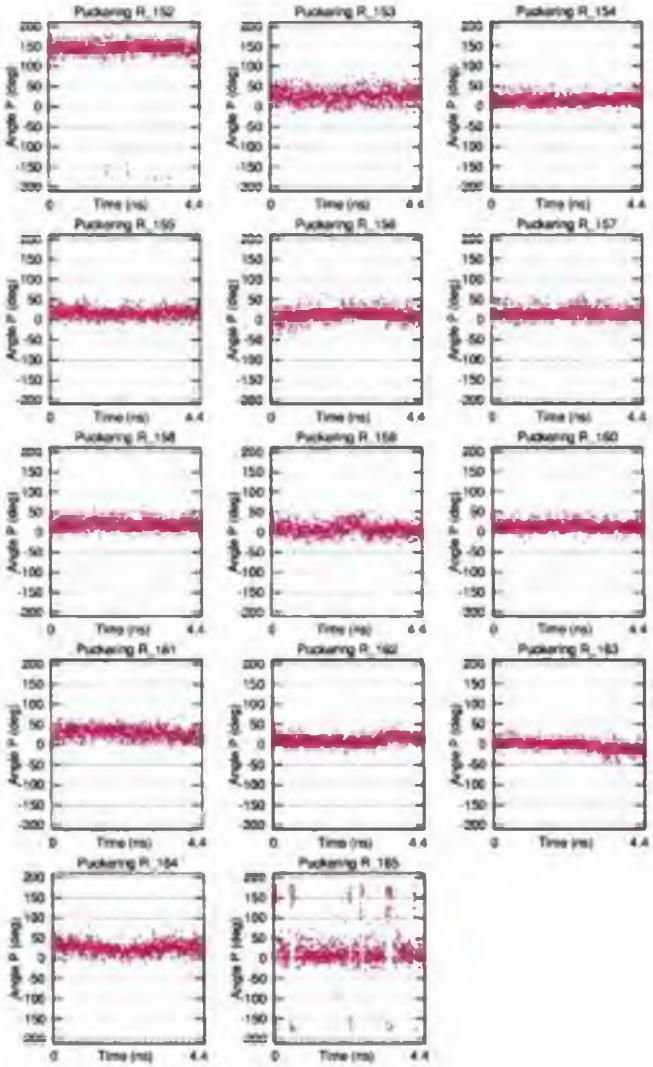


Figure D 27: Puckering of the RNA-strand. Modification 9

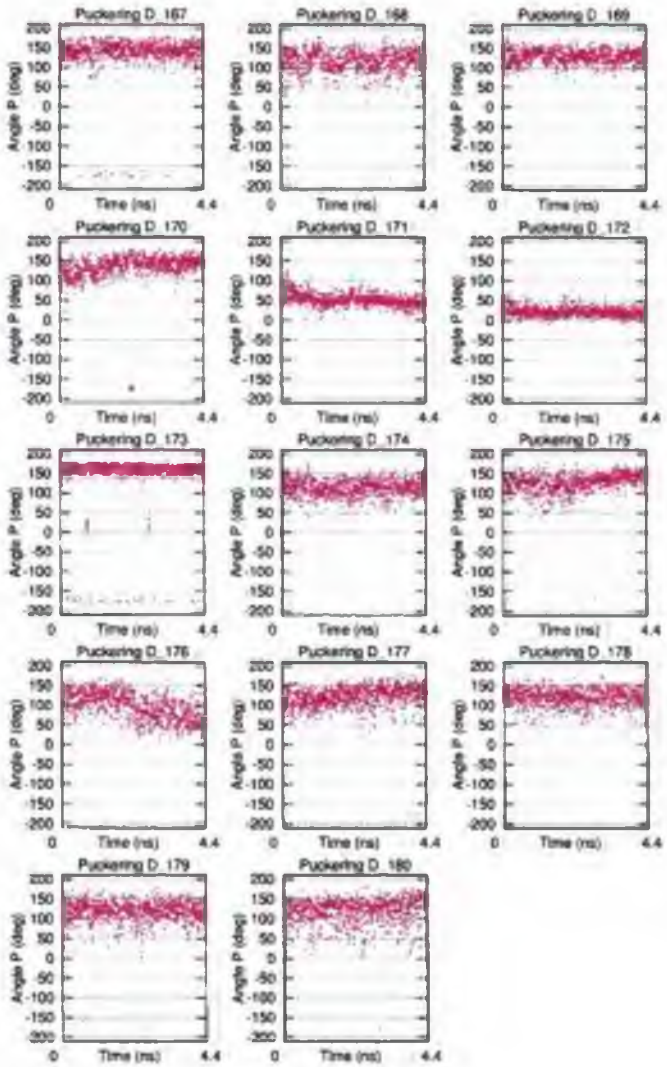


Figure D 28: Puckering of the DNA-strand. Modification 9

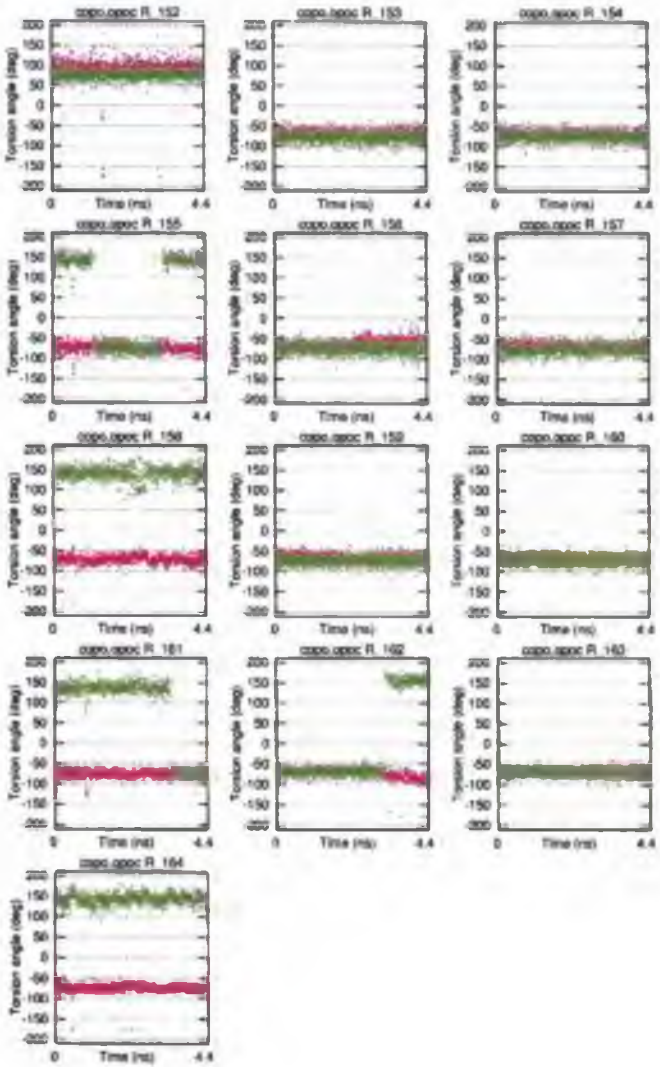


Figure D.29: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 9

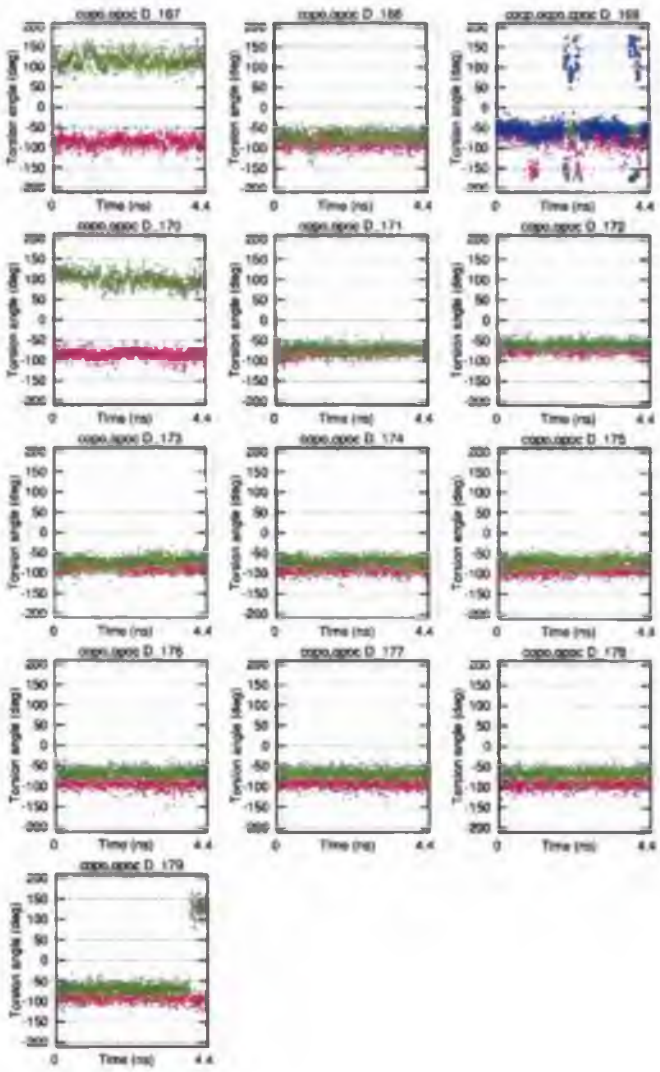


Figure D.30: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 9

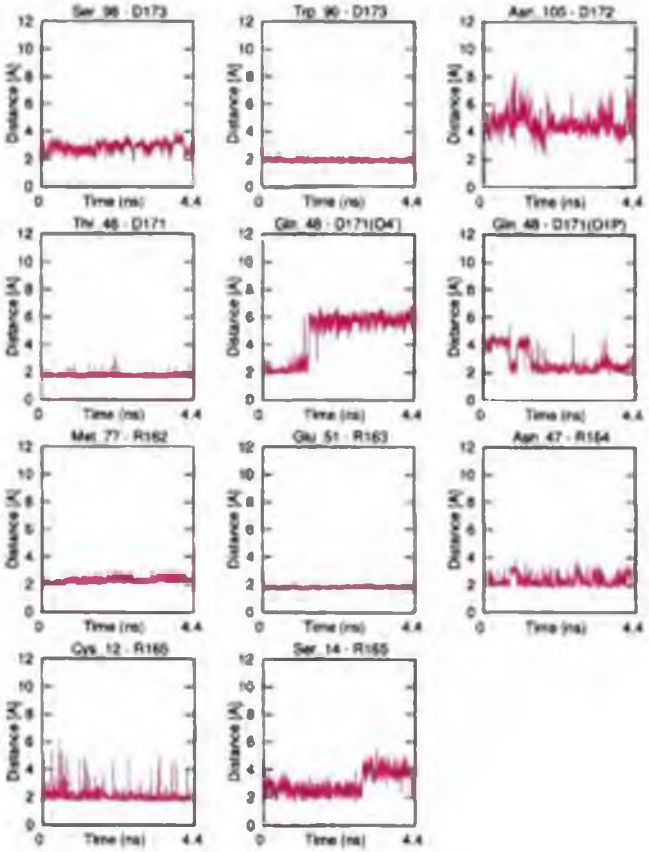


Figure D.31: The enzyme-substrate hydrogen bonds. Modification 9

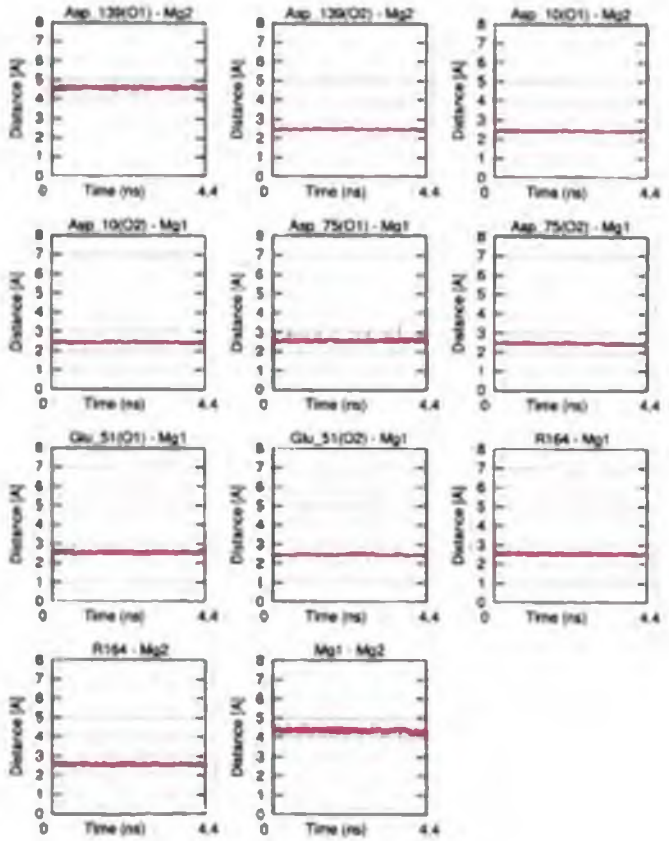


Figure D.32: Contacts between the Mg-ions and the DDE motif in the active site. Modification 9

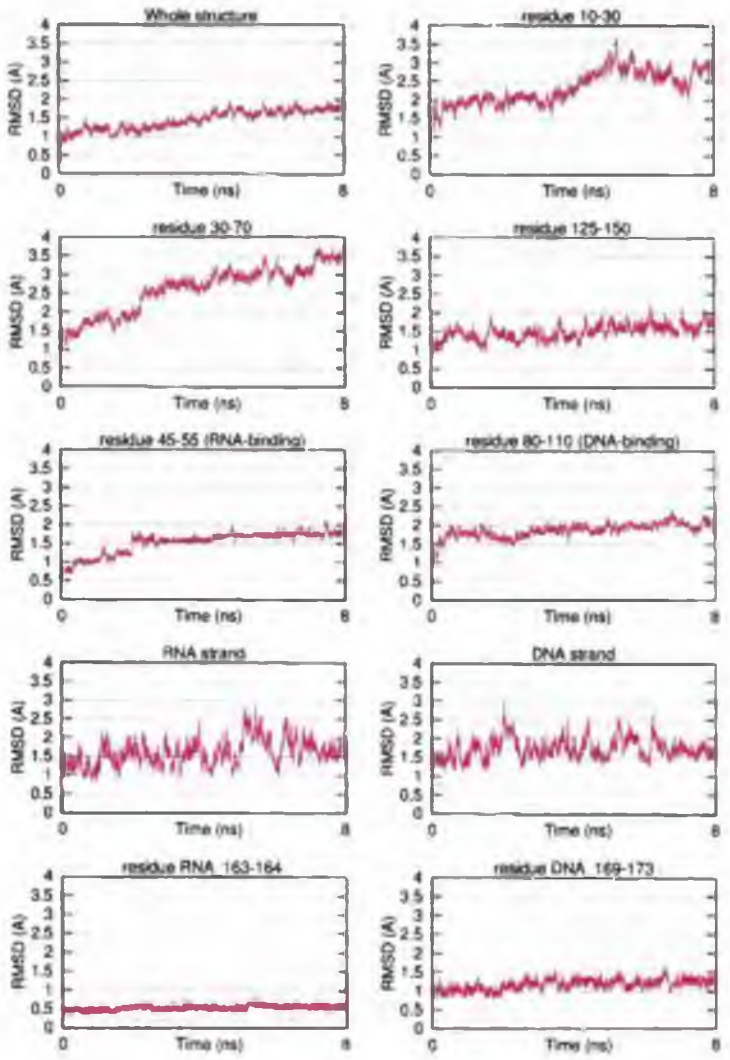


Figure D.33: RMSD of the whole structure and selected regions. Modification 10

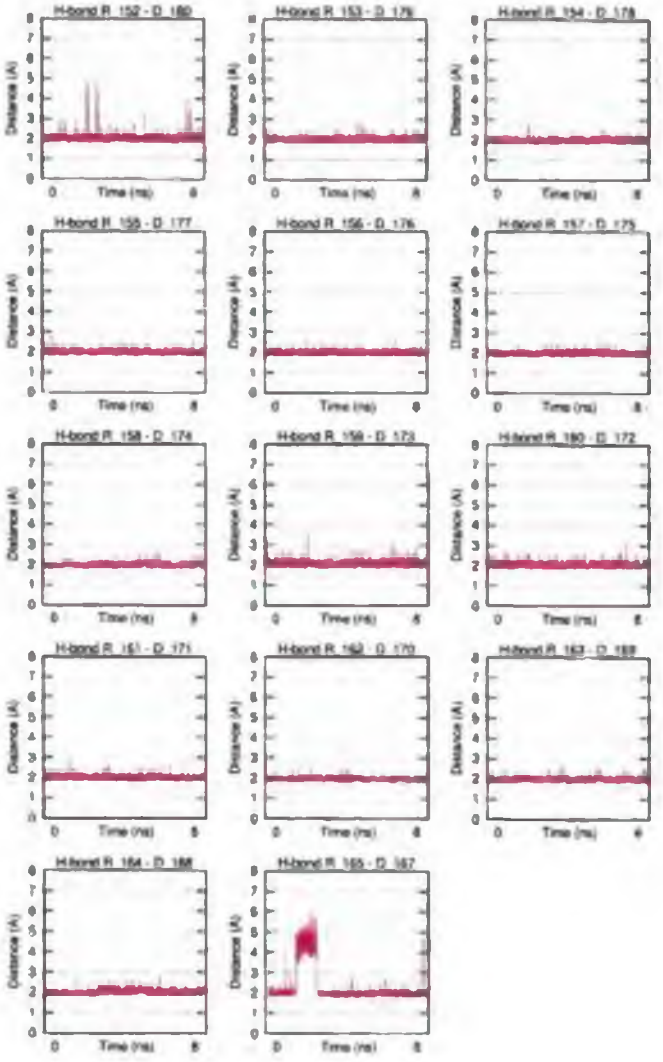


Figure D.34: The Watson-Crick hydrogen bonds. Modification 10

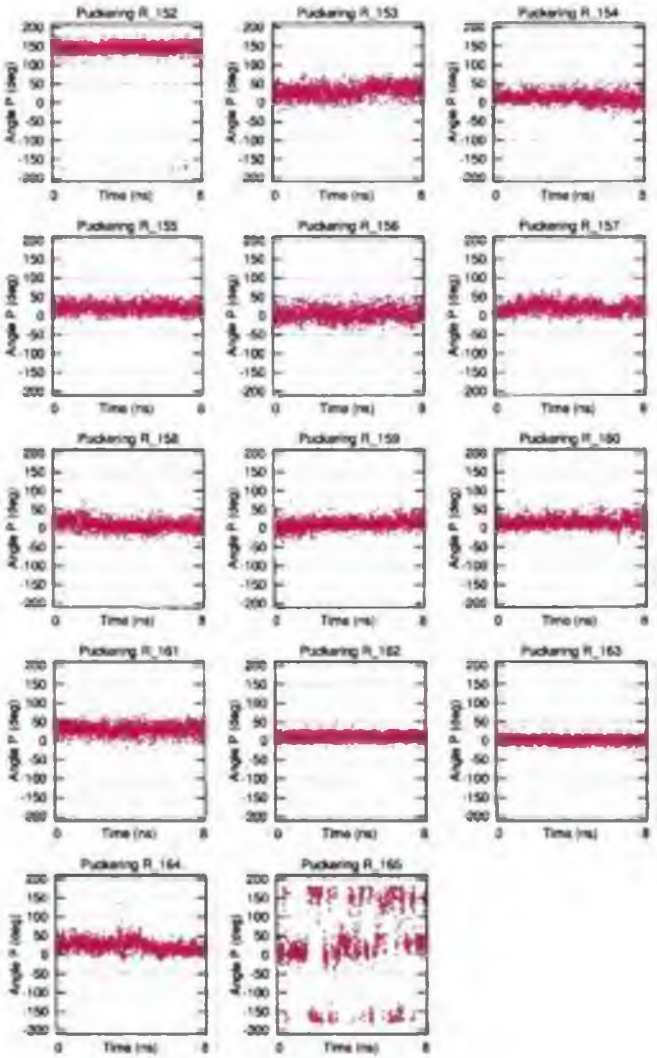


Figure D.35: Puckering of the RNA-strand. Modification 10

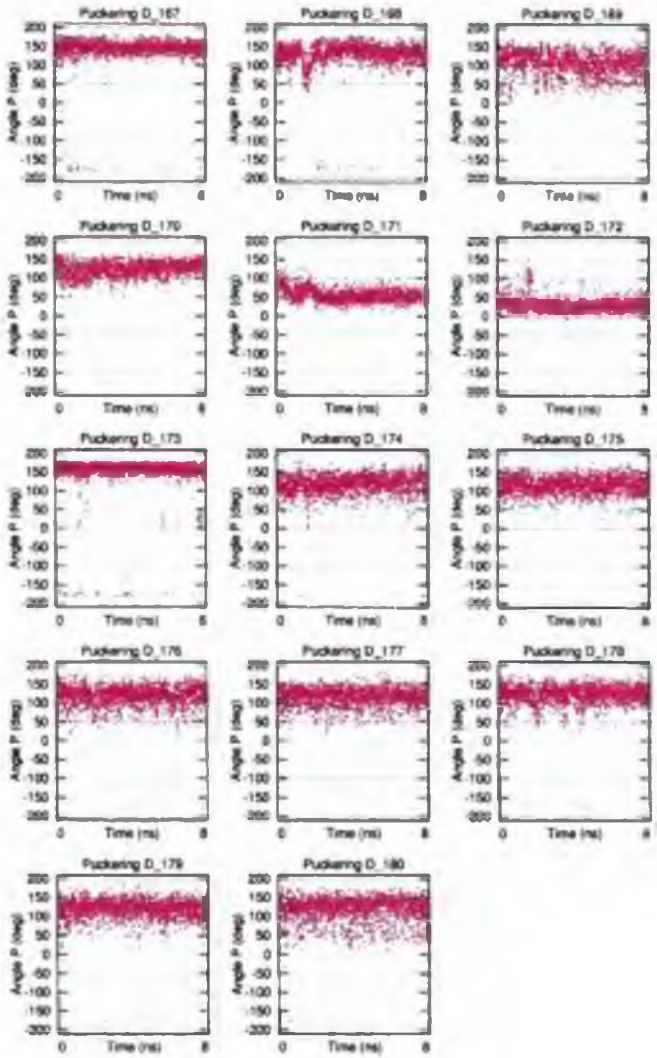


Figure D.36: Puckering of the DNA-strand. Modification 10

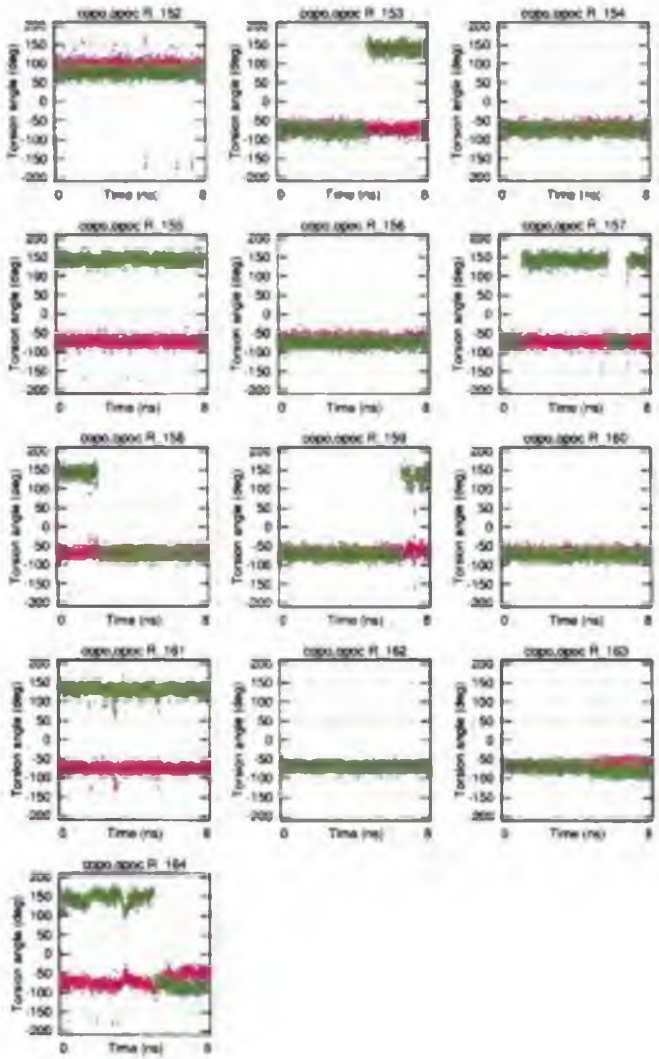


Figure D.37: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Modification 10

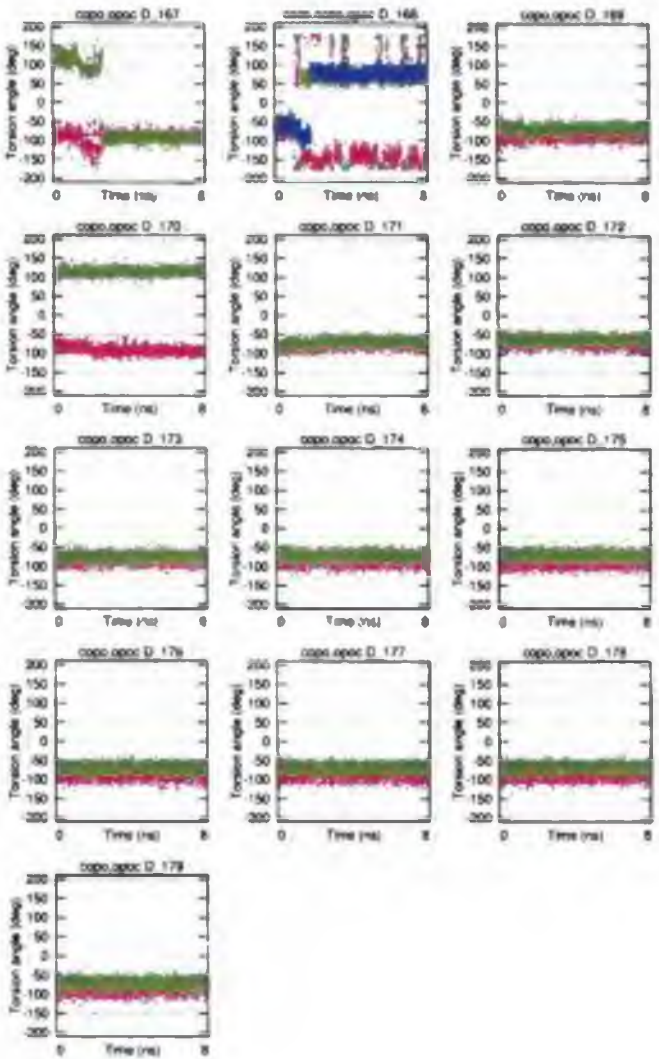


Figure D.38: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the DNA-strand (three torsion angles by the modified nucleotide 170). Modification 10

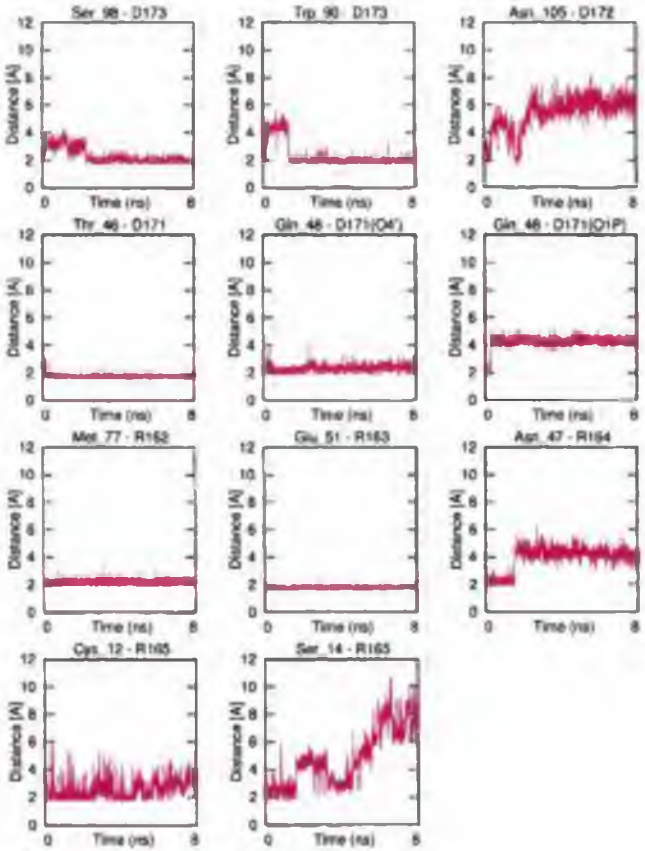


Figure D.39: The enzyme-substrate hydrogen bonds. Modification 10

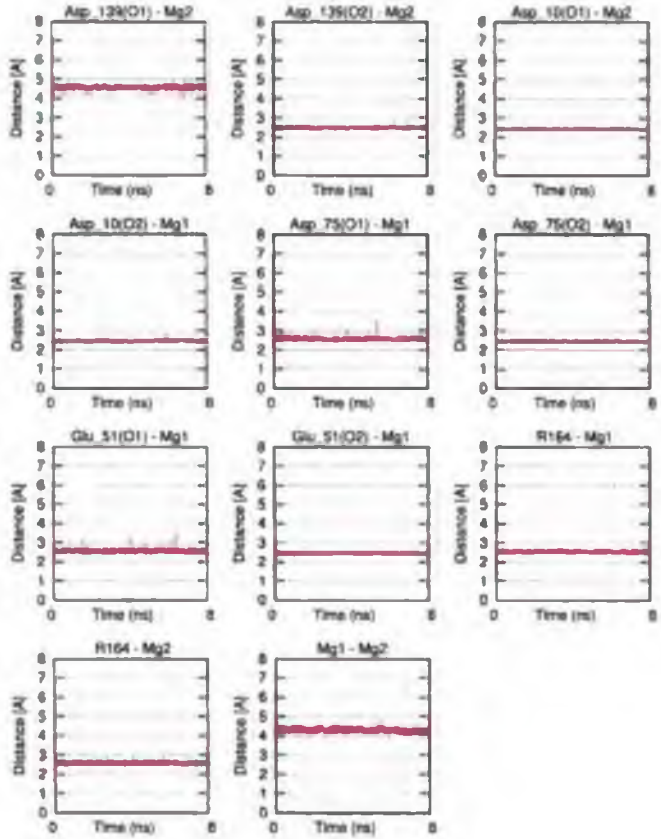
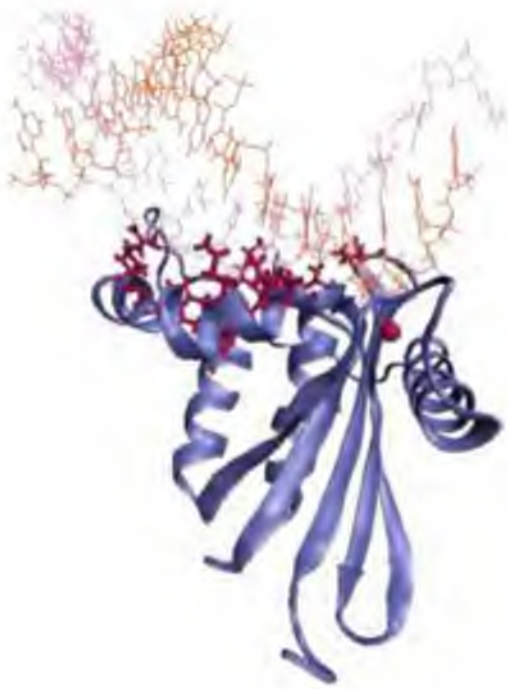


Figure D.40: Contacts between the Mg-ions and the DDE motif in the active site. Modification 10

Appendix E

Graphs - model of the E. coli RNase H with a substrate



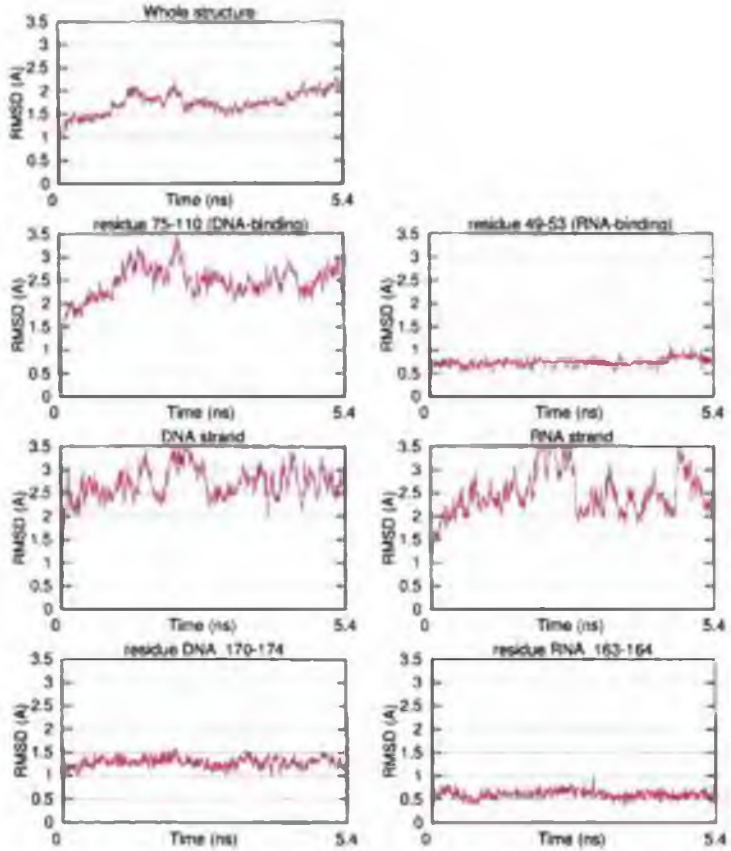


Figure E.1: RMSD of the whole structure and selected regions. Model E. coli

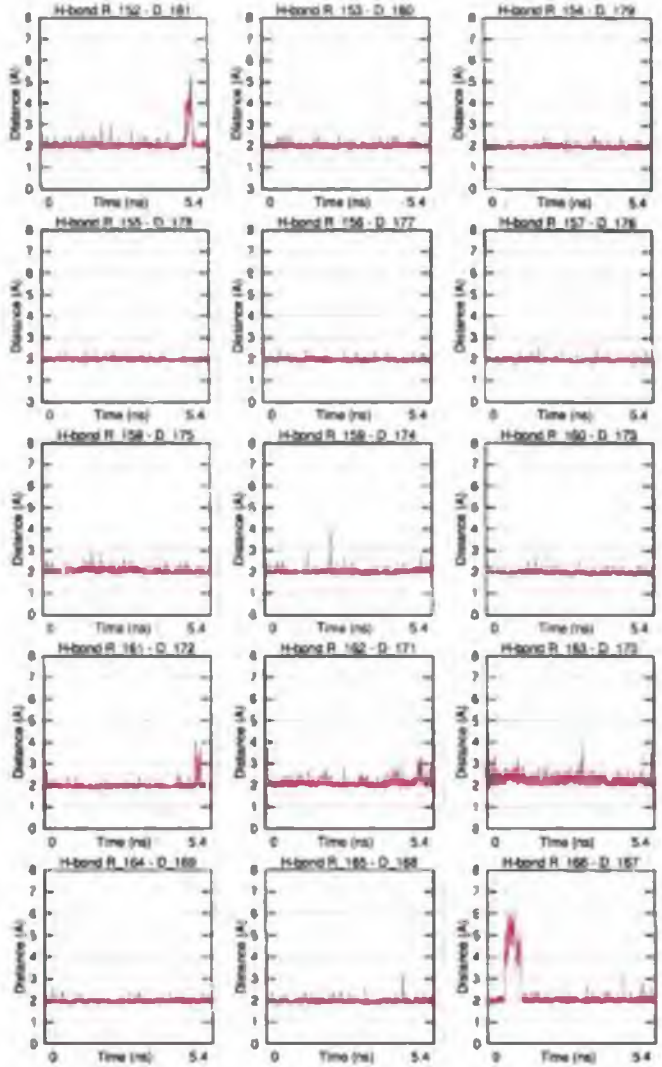


Figure E.2: The Watson-Crick hydrogen bonds. Model E. coli

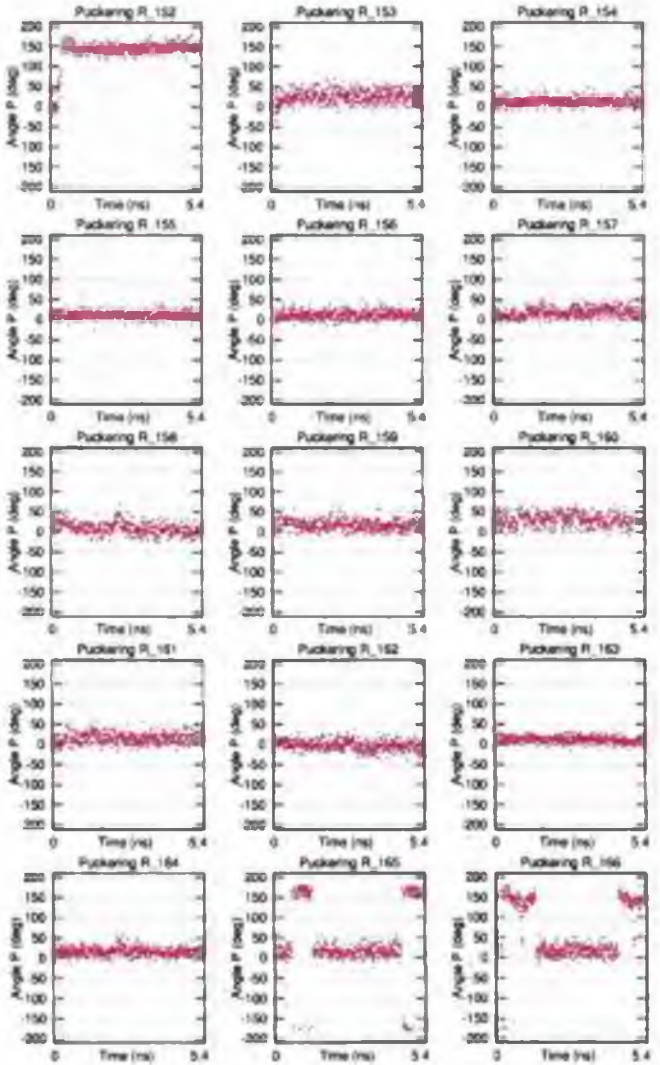


Figure E.3: Puckering of the RNA-strand. Model E. coli

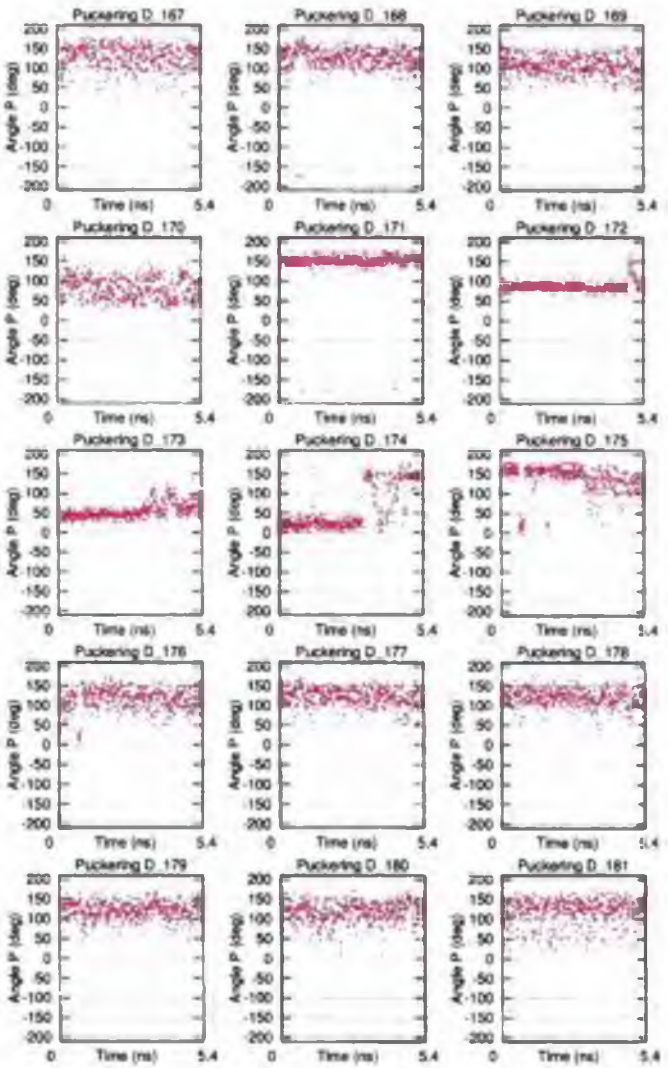


Figure E.4: Puckering of the DNA-strand. Model E. coli

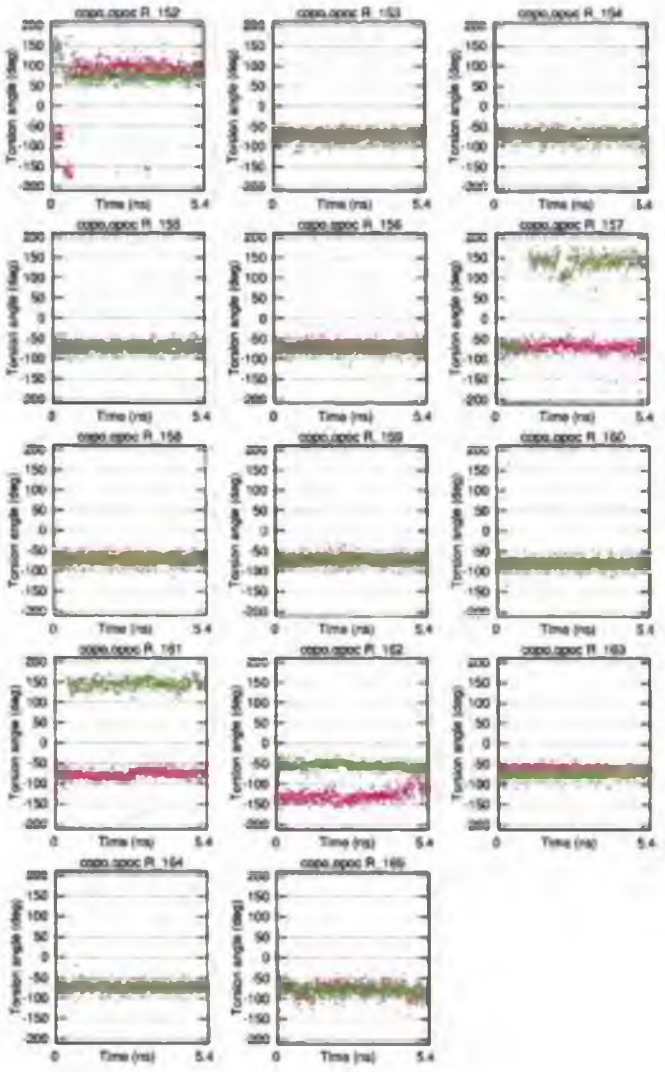


Figure E.5: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model E. coli

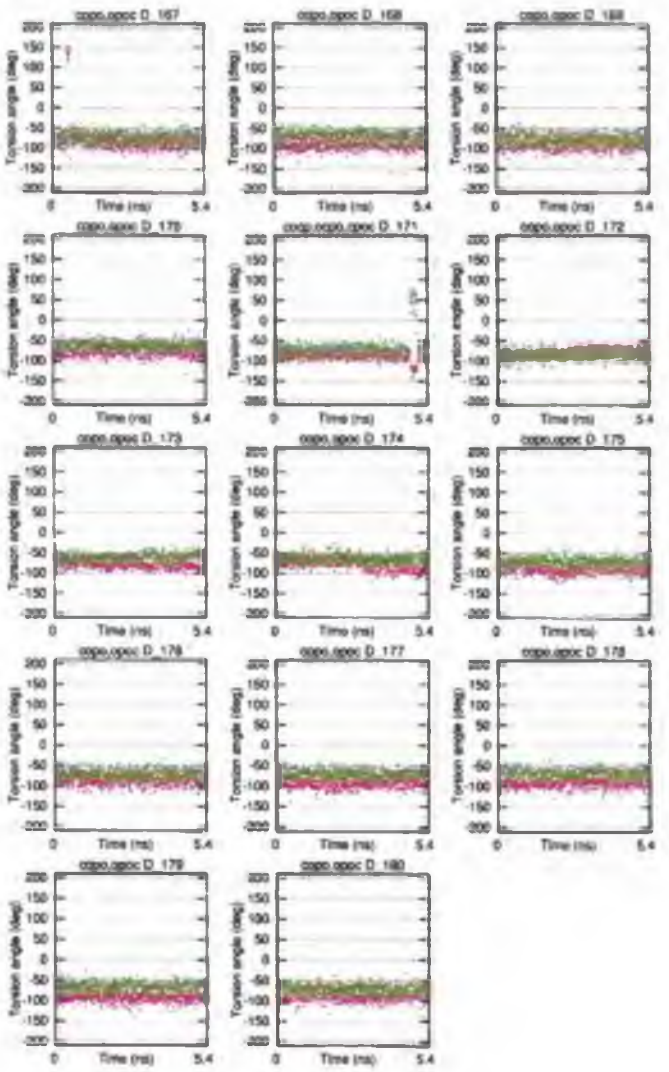


Figure E.6: The C3'-O3'-P-O5' (red) and O3'-P-O5'-C5' (green) torsion angles, the RNA-strand. Model E. coli

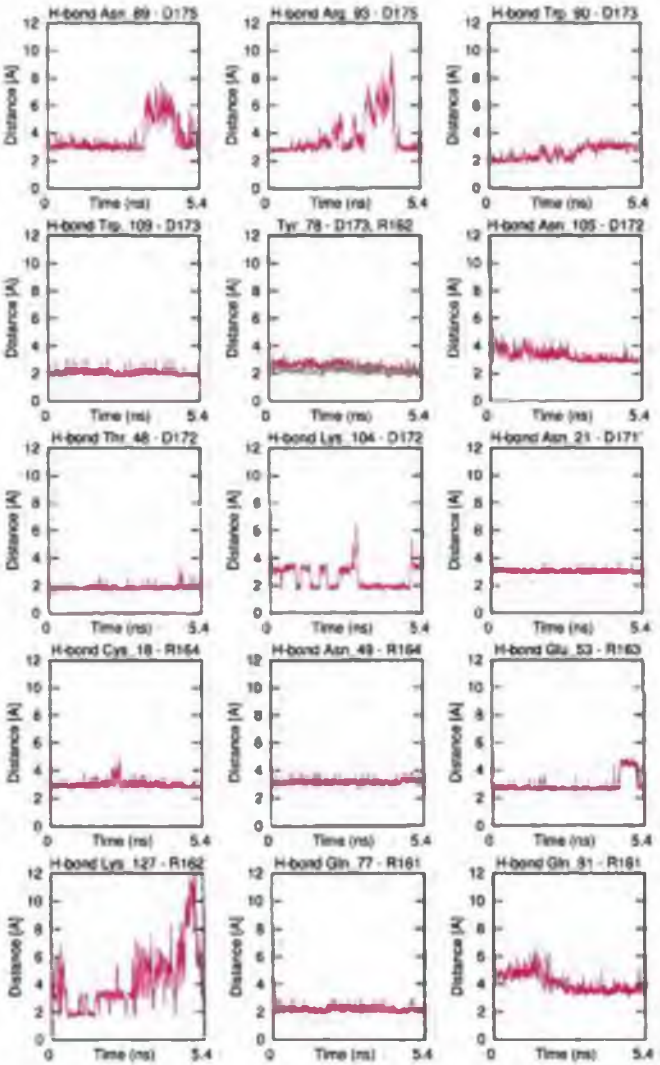


Figure E.7: The enzyme-substrate hydrogen bonds. Model E. coli

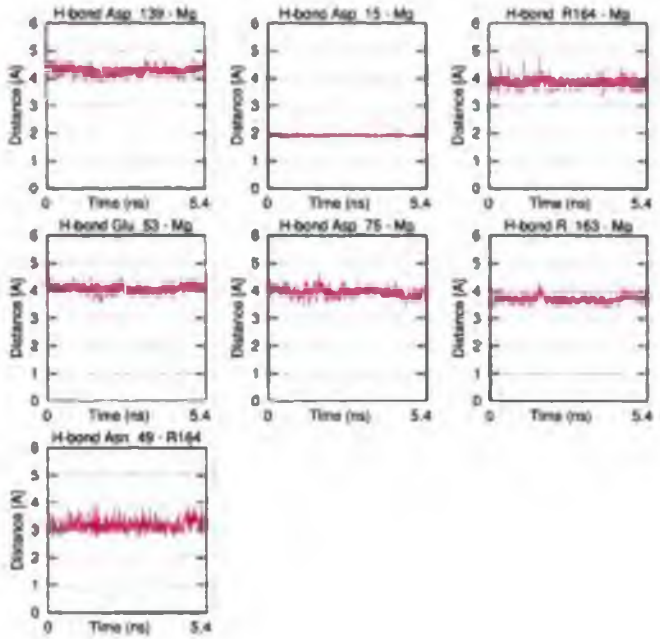


Figure E.8: Contacts between the active site residues and the Mg-ion. Model E. coli