ROMANIAN ACADEMY

INSTITUTE OF BIOCHEMISTRY

Ph.D. THESIS SUMMARY

Analysis methods applied in biomolecular modelling and simulation

Scientific Coordinator

Dr. Andrei-José Petrescu

Ph.D. Candidate

Marius Surleac

BUCHAREST

2017

TABLE OF CONTENTS

TABLE OF CONTENTS i

LIST OF FIGURES iii

LIST OF TABLES v

Acknowledgements vi

List of abbreviations vii

Aims of the thesis x

I. INTRODUCTION 1

I.1 Protein structure 2

I.1.1 Domains 3

I.1.2 Secondary structure and polypeptidic chain torsion 4

I.1.3 Structural motifs 5

I.2 Nucleic acids structure - DNA structure 5

I.2.1 Torsion angles on DNA backbone 9

I.2.2 Base pairs geometry 10

I.2.3 DNA bending and topology 11

I.3 Macromolecular structure determination methods 12

II. MOLECULAR MODELLING AND SIMULATION METHODS 14

II.1 Protein structure homology modelling 15

II.1.1 Protein templates identification 18

II.1.2 Target sequence based predictions 19

II.1.2.1 Secondary structure predictions 19

II.1.2.2 Disorder predictions 20

II.1.2.3 Motifs and domains predictions 20

II.1.2.4 PTM predictions 21

II.1.2.5 Hydrophobicity and folding predictions 21

II.1.2.6 Charge and variability predictions 22

II.1.3 Identification of the similarity between target and template sequences 22

II.1.4 3D model generation of target sequences for the studied proteins 23

II.1.5 Homology model validation 24

II.2 Proteins and nucleic acids MD simulation methods 25

II.3 Innovative method - *Coarse-grain* models' generation for DNA bending - case study: 12RSS and 23RSS DNA fragments in V(D)J recombination 27

II.4 Molecular docking simulations 28

III. V(D)J RECOMBINATION. RAG1/2 PROTEIC COMPLEX 30

III.1 Introduction into the V(D)J-RAG1/2 action mechanism 31

III.2 Results 33

III.2.1 Modelling the 12RSS & 23RSS DNA fragments 34

III.2.1.1 Structural model of the bent 23RSS DNA in PC 40

III.2.1.2 Structural model of the bent 12RSS DNA in PC and SC 45

III.2.2 Modelling the β-propeller RAG2 structure 49
III.2.3 Modelling the RNase-H catalytic domain of RAG1 63
III.3 Discussion 69
III.3.1 RAG1/2 - 3D structure comparison between the models and the crystals 69
III.3.2 Assembling the RAG1/2-12/23RSS complex 73
III.3.3 Comparison between the modelled and the crystallized RAG1/2-12/23RSS complex 78
III.4 Future perspectives 79
IV. TOPOISOMERASE IIα AND TOPOISOMERASE IIβ 80
IV.1 Introduction into the topoisomerase IIα and IIβ action mechanism 81
IV.2 Results 84
IV.2.1 Modelling the topoisomerase IIβ DNAbd 84
IV.2.2 Modelling the topoisomerase IIα DNAbd 87
IV.3 Discussion 90
IV.3.1 Modelling the dimers of topo IIα and IIβ isoforms 91
IV.3.2 Analysis of the topo IIα and IIβ DNAbd differences 92
IV.3.3 Analysis of the predicted DNAbd in both topo II isoforms 94
IV.3.4 Analysis of the regions around $Y_{\beta}656/Y_{\alpha}640$ 96
IV.3.5 Analysis of the CTD domains in both isoforms 99
IV.3.6 The action mode of the two isoforms in (+)/(-) supercoilings 101
IV.4 Future perspectives 103
V. "DECAPPING SCAVENGER" (DcpS) enzyme 104
V.1 Introduction into the mARN degradation mechanisms 105
V.2 Results 106
V.2.1 Modelling the dimer structures of DcpS_Ce enzyme 107
V.2.2 Docking the compounds into the DcpS_Ce active site 109
V.3 Discussion 114
V.3.1 Analysis of the m^7GpppG DcpS_Ce binding site 114
V.4 Future perspectives 116
Final conclusions 117
List of publications and conference participations 119
References 124

Acknowledgements

This thesis was done with the financial support of the next research projects:
UEFISCDI: PN-II-ID-PCE-2011-3-0342;
UEFISCDI: PN-II-PT-PCCA-2013-4-1407;
H2020: HIVERA-INinRAGI;

I would like to express my gratitude to:

- My family, for all the support during these years
- Dr. Andrei-José Petrescu and Dr. Ștefana Petrescu for the guidance and the help throughout the thesis.
- Dr. Laurențiu Spiridon
- Dr. Adina Milac
- Dr. Marius Micluță
- Dr. Mihai Ciubotaru, Immunobiology Department of Yale University, New Haven, Connecticut, US and IFIN-HH, Bucharest, Romania
- Professor David G. Schatz, Head of Immunobiology Department of Yale University, New Haven, Connecticut, US
- Dr. Mahrukh Ganapathi and Dr. Ram Ganapathi, Levine Cancer Institute, Carolinas Healthcare System, North Carolina, US
- Dr. Anna Wypijewska del Nogal and Dr. Elżbieta Bojarska, Faculty of Physics, University of Warsaw, Poland
- My colleagues from the Institute of Biochemistry

Aims of the thesis

The purpose of the studies included in this Ph.D. thesis was primarily to develop and implement special biomolecular modeling and simulation techniques, adjusted to study interactions and protein-induced effects on nucleic acids in vital biological processes such as somatic recombination, decatenation and degradation of mRNA; processes which are among

the most important topics and, also, intensively studied for decades by countless top research groups from prestigious universities.

- The somatic recombination or V(D)J - Variable, Diverse, Joining - is an extremely important process for the body's immune response to external factors (e.g. antigens) that can affect its integrity and plays a role in coding the hypervariable regions of immunoglobulins and in T-cell receptors. Our investigations focused here especially on the structural modeling of the main proteins and gene fragments that are involved in this mechanism, namely the RAG1 and RAG2 proteins and the 12RSS and 23RSS DNA fragments, and aimed to understand the mechanism of action of these types of proteins in interaction with DNA structures. Mutations, or other factors that affect the good behavior of these proteins, cause a number of autoimmune diseases that are often lethal.
- The decatenation and relaxation of the supercoiled chromosomal DNA - is an essential molecular mechanism encountered in a number of processes such as replication, transcription, DNA repair, cell cycle. The study focused on modeling the structures, and analysis of the mechanism of action, of the two human topoisomerase type II isoforms (topo IIα and topo IIβ) in controlling of the decatenation checkpoint, but also to understand their involvement in DNA supercoiling. These two proteins are vital for the normal cell functioning and mutations or factors that affect their function lead to their involvement in a range of diseases such as cancer, thus being extensively studied for drug development.
- The regulation of gene expression through the mRNA degradation process. This study focused on the modeling and analysis of the mechanism of action of the DcpS protein, having an important role in fragmentation of the mRNA ends, following the poly(A) tail removal in the deadenylation process. At the same time, the binding ability of various compounds (possibly therapeutics), analogous to the structure of the methylated nucleotide at the 5' end of the mRNA molecule, relative to the active site of the DcpS protein, was studied.

In addition, investigating these important systems has been a good opportunity to develop or improve a number of more general computational methods, which would also be helpful in future protein and nucleic acid modeling. Thus, I have developed, together with my colleagues, a method of bending DNA fragments through molecular dynamics simulations. I also improved the protein modeling methods, used in our department, by including multiple information related to predictions (secondary structure, disorder, PTMs, etc.), by setting up

charge profiles and variability profiles of sequences from multiple species. I have improved the analysis of the models and of the molecular dynamics simulations by developing programs / scripts written in programming languages such as AWK, Tcl, to calculate standard deviations between protein structures, distances between types of atoms, to create protein torsion angles values distributions, to analyze sequence databases, etc. In essence, the development of all these programs has the role of simplifying data analysis which, in their absence, could take hours or even countless working days; and they also cover research segments for which such tools have not been developed. All these are included in the scientific papers and book chapters I published.

INTRODUCTION AND METHODS

The cellular metabolism depends on thousands of interactions and reactions coordinated in space and time, dependent on both the gene instructions and operating environment. The effectors of this whole complex system are the proteins, on whose intricate structure and dynamics, the entire functionality of the biological system depends.

The brief introduction below summarizes the main concepts and structural techniques underlying the investigation of the protein-DNA systems that are the subject of this thesis.

Protein structure

The proteins are linear polymeric chains made up of 20 types of amino acids (aa) in sequences of tens, hundreds or thousands of units. Sequence information is encoded in the genome; and, according to needs, it is transcribed and translated into the protein chain on ribosomes, during the biosynthesis process (Alberts et al., 2008; Berg et al., 2011). Simultaneously with the biosynthesis, right during the elongation, the protein chain begins to spatially organize itself under a strictly cellular evaluation mediated by assistance proteins, the so-called chaperones (Hartl et al., 2011). The process of 3D packaging of the protein chain and bringing it to the native, functional form is called folding, and depending on the complexity of the chain it may take even hours after the end of the biosynthesis (Dill and MacCallum, 2012; Rognoni et al., 2014). The final, functional form of a protein may contain ordered regions with well-defined 3D structure as well as disordered intrinsic regions in which the polypeptide chain adopts multiple configurations. Experimentally, the structure of the foldable domains with well-defined 3D organization can be determined by X-ray crystallography (Van Benschoten et al., 2016), nuclear magnetic resonance (Ma et al., 2015) or more recently, electron

cryomicroscopy (Fernandez-Leiro and Scheres, 2016). These types of structural experiments are extremely complex, costly and time-consuming, which is why the number of structures determined so far is less than 150000, with about three orders of magnitude lower than those of the known sequences (Finn et al., 2017; Dawson et al., 2017). Due to the crucial importance that the knowledge of the 3D structure has in understanding the molecular basis of life in the last 20-30 years, sustained efforts have been dedicated to the analysis and organization of established structural data through the emergence of structural bioinformatics (Samish et al., 2015). If the polypeptide sequence is commonly known as protein's *primary structure*, protein packing analysis further identifies three more structural levels, specific to the spatial organization: the *secondary*, *tertiary* and *quaternary* structure. The *secondary structure* refers to the local conformation of the polypeptide chain - which can be repetitive, stabilized by hydrogen bonds (HB), or nonrepetitive, whether or not stabilized by HB (Nelson and Cox, 2004). The way in which secondary structure elements are grouped into space on longer sequence segments forms the so-called structural *motifs* or *super-secondary structure*; and the overall arrangement of all the atoms in the protein, including contacts between amino acids that are far apart from each other in the sequence, represents the third level of organization, the *tertiary structure*. The fourth level of organization is given by the *quaternary structure*. The quaternary structure consists in the arranging of several polypeptide chains / subunits in a single complex, units that may be similar (homo-oligomers) or different (hetero-oligomers). The subunits interact with each other and can form, for example, active sites in the protein, may be involved in interacting with other proteins. Taking into account all these levels of organization, the proteins can be classified into two main groups: - fibrous proteins (the polypeptide chains are arranged as long strips); - globular proteins (the polypeptide chains are folded in globular or spherical shape) (Nelson and Cox, 2004).

Nucleic acids structure - The structure of DNA

In a cell, the information-bearing molecules are of two types: DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), built based on essential units called *nucleotides* (Lodish et al., 2003). The DNA is the elementary molecule of each cell, it is found in the nucleus of the cell at eukaryotes, and in order to carry out its information-bearing function, it has to do more than copy itself, thus being involved in guiding the synthesis of the other molecules in the cell, primarily RNA and proteins (Lukacs et al., 2000). The DNA damage (under X-ray, UV or certain chemical agents such as reactive oxygen species) or the mutations (which may occur

and lead to chromosome aberrations during chromosomal separation) can cause various types of cancer, cell death or aging (Hoeijmakers, 2001). The three-dimensional structure of the DNA consists of two strands, twisted together to form the so-called *double helix* structure (Alberts et al., 2008).

The DNA backbone is formed by the phosphate-sugar groups (located on the exterior of the double helix) of adjacent nucleotides linked together, and the strands are held together by H-bonds that are formed between complementary bases on the inside of the double helix (Berg et al., 2011).

The stability of the double helix DNA structure is mainly due to two factors: the association between the bases in the complementary strands (through the H-bonds and the hydrophobic interactions between the bases) and the stacking interaction between the adjacent bases; which in turn are influenced by the melting temperature and the salt concentration (Yakovchuk et al., 2006), especially due to cations, that can neutralize the negative charges of the phosphate groups (Zhang et al., 2015).

The C3' carbon in the sugar unit is connected through a phosphate group to the C5' carbon of the next sugar unit. The bond is called 3'-5' phosphodiester bond (Pratt and Cornely, 2013). All the DNA strands are read from the 5' end to the 3' end, where the 5' end ends with a phosphate group and the 3' end ends with a sugar unit (Koolman and Roehm, 2005). Each unit is covalently bonded via the C1' atom to one of the 4 possible bases (at N1 atom of the pyrimidines or N9 atom of the purines) through an N-β-glycosidic bond (Nelson and Cox, 2004). The bases are perpendicularly oriented to the axis of the helix; and are hydrophobic in the perpendicular direction to the base plane (thus they can’t form H-bonds with the water molecules) (Kuriyan et al., 2013), while the exterior of the DNA is negatively charged. Because of its double helix shape, two grooves called the *minor groove* and the *major groove* (Pratt and Cornely, 2013) are formed along the length of the DNA molecule. These grooves are involved in the interaction with a range of proteins including transcription factors (Privalov et al., 2007). The 5'-3' strand is called *sense*, while the 3'-5' chain is called *anti-sense* (Koolman and Roehm, 2005).

There are three forms of DNA (A-, B- and Z-) of which the most common structure is B-DNA - which, depending on certain factors, can adopt the A-DNA (under low hydration) or Z-DNA (when the number of GC regions in the sequence is increased) forms (Koolman and Roehm, 2005). The DNA molecule is not a rigid structure - in fact, each of these types of structures is in a continuous thermal fluctuation, in interaction with other molecules (e.g. proteins, water

molecules), ionic interactions - all these lead to local twists, stretches, bends or strand disentanglements, etc. (Westman, 2006).

Innovative method - *Coarse-grain* models' generation for DNA bending - case study: 12RSS and 23RSS DNA fragments in V(D)J recombination

The distances measured through FRET (between fluorophores attached to various DNA bases), by our collaborators at Yale School of Medicine, suggest that when the PC complex is formed (23+12RSS) in the V(D)J recombination, the 12/23RSS DNA fragments are strongly bent by the RAG1/2+HMGB1/2 protein complex. To model the DNA curvature according to the experimentally observed FRET constraints, it is necessary to impose unequal constraints on the two sides of the DNA. In linear B-form DNA, the periodicity is 10.5bp per turn and the average distances between the equivalent atoms of the two antiparallel strands are 20Å and 14Å along the major and minor grooves (Ciubotaru et al., 2013). The *coarse-grain* model I have developed in this thesis, together with my colleagues, involves the easy bending of the DNA using a series of harmonic distance constraints imposed gradually and unequally on the major and minor grooves, in order to affect the *tilt* and *roll* parameters, discussed in the Introduction chapter. In order for the DNA structure not to deviate from B-form towards an irregular form, we put distance constraints on the σ, ω and κ parameters (corresponding to the opening of the base pairs, the twist propeller parameter and the bending of the base pairs, respectively). Thus, on one side of the double-stranded helix, we have imposed a series of longer distance constraints on both grooves (about 10% between the n, n+10 residues, etc.) to force a slight spacing between the bases on the convex side of the bend, and on the other side of the DNA we imposed distance constraints of about 10% less on both grooves (between the n+5, n+15 residues, etc.). I will discuss this method in more detail in the V(D)J recombination results chapter.

RESULTS

THE V(D)J RECOMBINATION. THE RAG1/2 PROTEIN COMPLEX

Until recently, little was known about the structure of 12RSS and 23RSS recombinant DNA fragments in complex with the RAG proteins, and about the detailed mechanism of the way they work. To address this issue, the first step in solving this "puzzle" was to generate theoretical structural models for the two RSSs based on FRET experimental data received

from our Yale University collaborators, as well as models for the catalytic site of RAG1 and the β-propeller domain of RAG2.

The B-form 23RSS DNA, which we used for modeling, contains a 65bp nucleotide sequence consisting of a 16bp region coding for hypervariable regions of the antigen receptors, a 7bp heptamer, a "spacer" of 23bp, a 9bp nonamer and a short fragment of 10bp not involved in the coding. The B-form 12RSS DNA, which we used, contains a sequence of 59bp, with the mention that the only differences to the 23RSS DNA are found in the spacer area (it has a length of 12bp) and in the non-coding region (this region has 15bp), while the other three areas are identical to those in the 23RSS.

The first step in shaping the 23RSS and 12RSS DNA structures was to generate 3D extended structures based on the two RSS sequences, using the NAB (Macke and Case, 1998) program that is part of the AMBER software suite (Salomon-Ferrer et al., 2013).

Based on the extended RSS structures, the next step was to prepare the input files required for energy minimization and Generalized Born Simulated Annealing(GBSA) molecular dynamics, as described in Methods. To begin with, for the preparation of input files we used the xLEaP program from Amber9 - I generated the topology and coordinate files, added the H atoms and the Mg^{2+} ions to the DNA structure, using the ff99SB force field parameters for proteins and nucleic acids (Hornak et al., 2006). Then I used the ambpdb program from Amber9 (Case et al., 2006) to convert the Amber files to the .pdb format.

In the next step, I prepared the constraint files. To preserve the DNA in its B-form, I used: a) dihedral constraints - two types of angular constraints applied to the base pairs to maintain their planarity; b) Watson-Crick distance constraints on the H-bonds between the DNA bases - to prevent segregation of the strands; c) distance constraints on the Mg^{2+} ions - in which we included the distances between the ions and the P-atoms in the backbone of the DNA, so that the Mg^{2+} ions would remain at an optimal distance from the DNA and won't disturb its local structure, but also not to be too far away - so can still exert some influence on the DNA.

In addition to these constraints, I have applied unequal and periodic distance constraints (Figure III.3) on the minor and major grooves, constraints which are also very important for generating the coarse-grain DNA bending model. These constraints are part of an innovative method developed within our department.

The simulations were performed on a high performance (HPC) Bull NovaScale R422/R423 HPC-cluster, 32- core CPU. Last but not least, we performed simulations of the final models in explicit solvent using the NAMD program (Phillips et al., 2005).

The resulting pattern for the 23RSS DNA in the PC shows that the DNA molecule is strongly bent in the form of the letter "U", with the "spacer" occupying the base of the "U" and the other DNA regions (e.g. heptamer, nonamer) residing on the "U" arms (Figure III.5.A), starting from an approx. 220Å distance between the ends of the DNA (in its extended form) and reaching approx. 77Å between its ends (in the bent form).

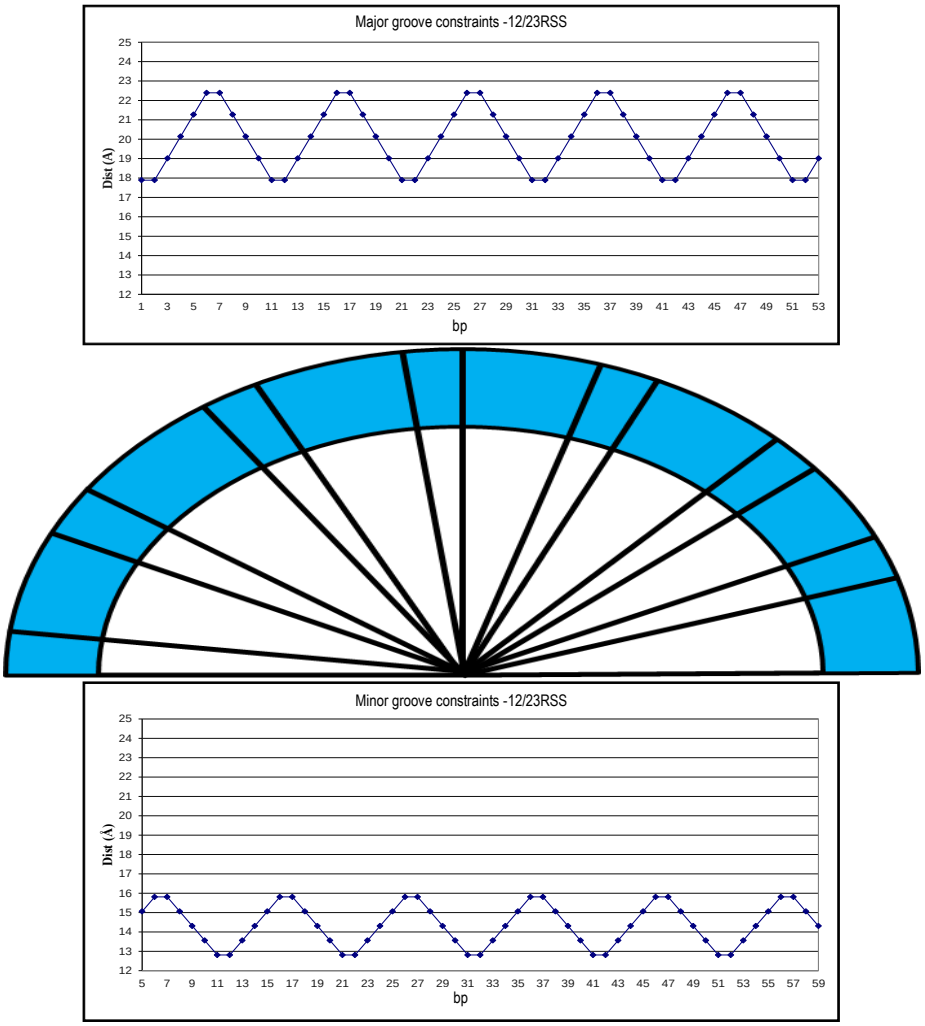

Fig. III.3 The distance constraints on major and minor grooves in 12/23RSS DNAs

Within the model, there are several distinct bends, of which the four most obvious are found at: the interface between the coding region and the heptamer (51°, external angle); the interface between the heptamer and the spacer (44º); the center of the spacer (49º) and the interface between the spacer and the nonamer (55º) (Fig III.5.B) (Ciubotaru et al., 2013).

Moreover, it is interesting that when interacting sites between RAG and DNA in the SC complex (experimentally observed by using the *ethylation interference* method) are represented on the model, most of the contacts are found on the inner surface (concave) of the

DNA. In contrast, sites that are sensitive to the Deoxyribonuclease I in the SEC complex, are predominantly found on the convex (outward) surface of DNA. It is to be taken into account that the latter sites are in the immediate vicinity of the most obvious bends in the spacer and of the bends from the spacer-nonamer interface (Figure III.5.C).

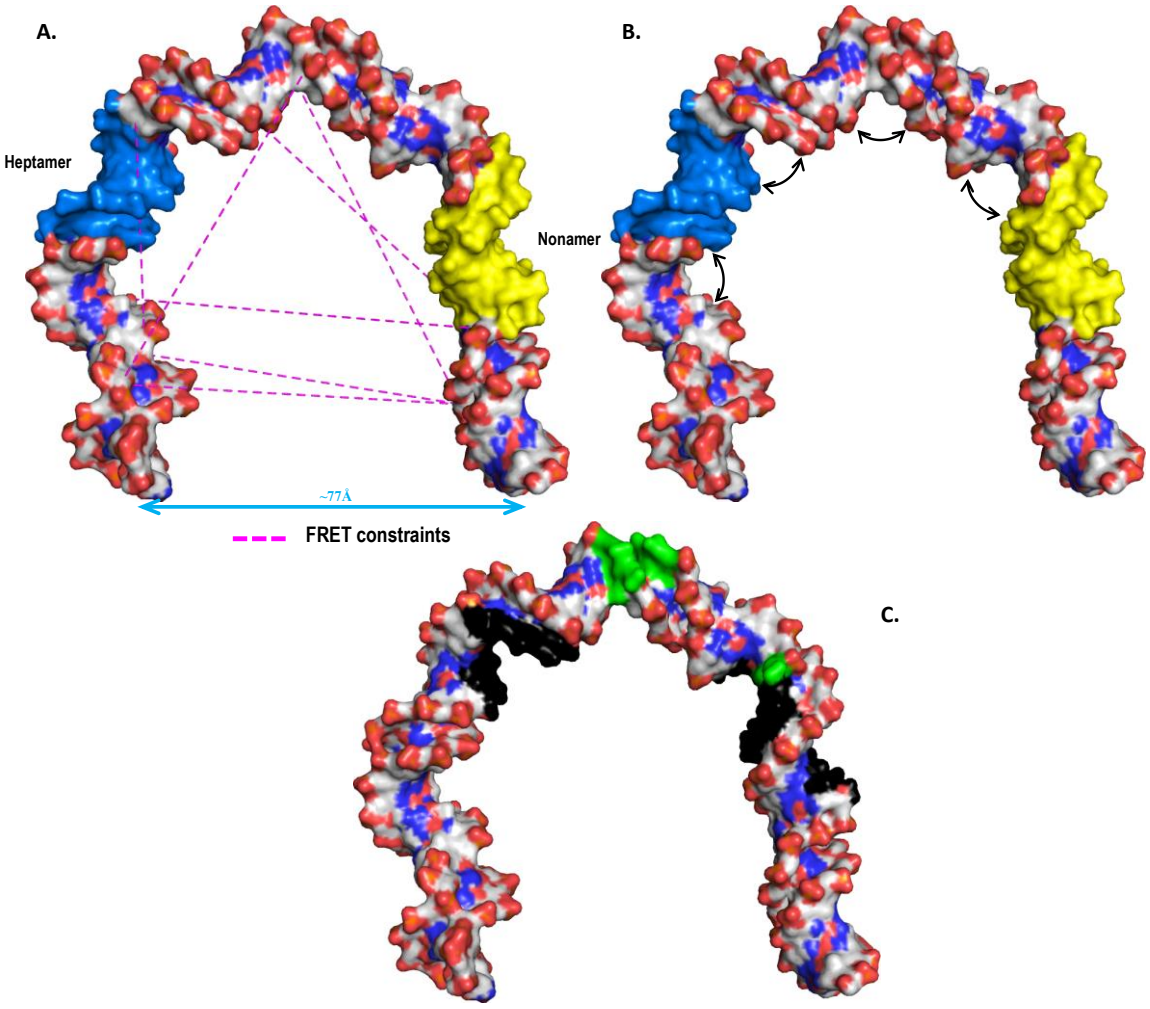

Fig. III.5 Final bending model of the 23RSS DNA
A. FRET constraints; **B.** Major bends in 23RSS; **C.** Ethylation (black) and hypersensitivity (green) sites

The analysis of the bent 23RSS DNA model shows that most of the distances in the model are consistent with those measured in FRET. In addition, the model was subsequently validated by new FRET experiments.
The bending method of the 12RSS DNA was similar to that of the 23RSS DNA bending.

The RAG2 protein, along with RAG1, is a component of the RAG protein complex that mediates the DNA cutting during the somatic recombination (or V(D)J). Within the complex,

RAG2 is not the catalytic component, this role being played by RAG1, but RAG2 is required for the catalytic activities mediated by RAG1.

RAG2 has 527 amino acids and is divided into two large regions: an N-terminal central domain (1-383aa) which is required for the recombination reaction, and a second C-terminal domain (384-527) which is not essential (Elkin et al., 2005). It is known that the central domain adopts a β-propeller structure, which is a common structural motif in many proteins.

Initially, our Yale collaborators provided us with the human RAG2 protein sequence for which we generated the first homology model and then, based on this model, we generated a model for the RAG2 protein from mouse. At the time of generating the RAG2 model, there was no information related to the three-dimensional structure (e.g. X-ray crystallography, NMR) of this domain.

To generate the RAG2 model, the first 351aa of the sequence were considered. Based on this sequence, secondary structure predictions, disorder predictions, charge profiles and similar template searches have been made, whose structure is known, using the methodology detailed in the Methods chapter.

RAG1 is the other protein of the RAG complex that mediates the V(D)J recombination, which assembles the variable regions of the immunoglobulin and T cell receptors, during the development of B and T lymphocytes (Zhang et al., 2015). RAG1 is the main component for DNA binding and cutting. In this study, the RAG1 sequence from mouse was used, it is 1040 amino acids long. This protein is comprised of a series of zinc finger domains (ZF) in the N-terminal region connected by a series of linkers, followed by a dimerization domain, a nonamer binding domain (NBD) which binds to the nonamers in 12RSS and 23RSS, a dimerization and DNA binding domain (DDBD), followed by an RNase-H domain towards the C-terminal end. The RNase-H domain is divided into two regions separated by an insertion domain (ID) of about 230aa. The central region, the minimum area of RAG1 that is required for activity, comprises aa384-1008, starting with the NBD domain. In addition to the ability of RAG1 to bind directly to the nonamer of the RSS, RAG1 is also able to interact with the RAG2 protein but also to recognize the heptamer from each RSS (Shinkai et al., 1992; Zhang et al., 2015).

The RNase-H domain represents the active part of the RAG1 protein and contains the DDE motif that forms the catalytic site of RAG1 and coordinates two Mg^{2+} ions. This DDE motif contains two aspartic acids and a glutamic acid, although in some proteins the glutamic acid is replaced by an aspartic acid. Due to the presence of this domain, RAG1 has a number of

similarities to other proteins that contain the same domain, including the *Tn5*, *Hermes* transposases or even the HIV-1, HIV-2, PFV integrases, that function by similar DNA-cutting mechanisms. Similar to RAG2 (at the time of this study), too little was known about the three-dimensional structure of RAG1, apart from the crystal of the NBD domain (the only region of known structure).

In the present study, a homology model for the RNase-H domain of RAG1 was generated, using the same methodology as for RAG2.

Unfortunately, the programs used to identify the template proteins couldn't find a global template protein for the entire RAG1 protein. Thus, by using remote homology modeling techniques, it was possible to construct the model for the RNase-H domain, and for two additional domains (extensions) at the ends of this domain, corresponding to the aa538-593 and aa997-1010 intervals (Zhang et al., 2015). For the remote homology modeling method, the information from the alignment of sequences of the RNase-H domains from multiple proteins was taken into account; thus, a multiple alignment of sequences was performed first which was later refined based on the alignment of the secondary structure elements by which this kind of domain is recognized, but in the same time the DDE (D600, D708 and E962) motif was "fixed" in the alignment. The RNase-H domain consists of a series of 9 secondary structure elements in the following order: β1-β2-β3-α1-β4-α2-β5-α3-α4 and corresponds to the aa594-732 and aa960-996 intervals of the RAG1 sequence.

To build the model, the following proteins containing the RNase-H domain were used: the *Hermes* transposase (4d1q.pdb) (Hickman et al., 2014), a metnase (3k9k.pdb) (Goodwin et al. 2010), the HIV2 integrase (3f9k.pdb) (Hare et al., 2009-B), the HIV1 integrase (4ovl.pdb) (Peat et al., 2014), the *Visna* virus integrase (3hpg.pdb) (Hare et al., 2009-A), the PFV (Prototype Foamy Virus) integrase (3oya.pdb) (Hare et al., 2010).

The 12/23 rule, which includes the 12RSS and 23RSS DNAs in PC, suggests RAG1/2's preference for a substrate asymmetry, but the exact reason for why this preference exists is not fully understood (Ciubotaru et al., 2015). The rationale behind this research, guided both on the RSS DNA fragments and on RAG proteins, was to understand as much as possible the mechanism of action of the V(D)J somatic recombination and if this substrate asymmetry could be reflected by the differences in the structures of 12RSS and 23RSS DNAs. From the results presented in this chapter it appears that the bending shape of both DNA structures is quite similar and the strong bending of both substrates is due to the RAG and HMGB proteins, the critical role being played by the RAG1 protein.

TOPOISOMERASE IIα AND TOPOISOMERASE IIβ

In the present study, homology models were generated for the DNA binding domains of human topoisomerases II; the dimer models of the two isoforms were generated, we studied the implications that two mutants, $Y_{\alpha}640F$ and $Y_{\beta}656F$, have on the function and we analyzed the sequences of each isoform's C-terminal domain, together with the implications that they may have. The two isoforms have an identity of about 68% and a similarity of 77% over the entire sequence. Both are similarly expressed in dividing cells, including tumors, but topo IIβ is more abundant in cells that do not depend on the cell cycle (Shapiro and Austin, 2014). Topo IIα is essential for cell survival *in vitro* (Akimitsu et al., 2003), whereas topo IIβ is not essential for *in vitro* cells but is necessary, for example, for neuronal differentiation (Lyu and Wang, 2003).

At the time of this study, there was little information about the three-dimensional structure of the two isoforms, especially X-ray crystallography data for each of them. Of the three domains, only the X-ray structure of the ATPbd domain was known, but not for the other two domains. The topoisomerase IIα (1531aa) and IIβ (1621aa) sequences, required to generate the DNAbd domain model of each of the two isoforms, were provided by our collaborators Ram Ganapathi and Mahrukh Ganapathi, Department of Cancer Pharmacology, from the Levine Cancer Institute, Carolinas Healthcare System, North Carolina, US.

The DNAbd domain is also divided into several other subdomains: TOPRIM (aa445-731), WHD (aa731-906), Tower and C-gate (aa906-1201); of which TOPRIM contains the acid triad that chelates the Mg^{2+}, while WHD contains the catalytic tyrosine (Wu et al., 2011). Taking into account the organization of the topo II from yeast, the DNAbd can also be divided into two large sub-domains A' and B', linked together by a '*hinge*' that allows an extra degree of freedom to the movement of a domain relative to the other and relative to the DNA (Classen et al., 2003).

To begin with, a homology model for the topo IIβ DNAbd domain using the aa449-1202 sequence interval was generated (Grozav et al., 2011). The same procedure as for RAG1/2 proteins was used to generate the model. Using the Phyre program (Kelley et al., 2009), 3 yeast topoisomerase II template sequences were found, with the following PDB codes: 3l4j.pdb (Schmidt et al., 2010), 1bgw.pdb (Berger et al., 1996), 1bjt.pdb (Fass et al., 1999). Aligning the sequences and studying the three-dimensional structures of these three structures showed that the sequence of the 3l4j.pdb structure is the most similar to the topo IIβ sequence

(47% identity and 75% similarity), it has the best resolution (2.48Å) and, distinct from the other two structures, does not present as many missing coordinates in its 3D structure.

The IIα and IIβ topoisomerases are part of a series of essential factors in the treatment of various cancers, by developing inhibitors (or "poisons"), but can also be used as antibacterial drug targets (Nitiss, 2009-B). Human topoisomerase II inhibitors are usually those that interfere with the DNA G-segment in tumors, where the expression of topoisomerases is increased (Durbecq et al., 2004, Skotheim et al., 2003). Due to the fact that the two isoforms are also involved in different processes relative to each other, they are intensively studied to understand the different mechanism of action.

Our collaborators discovered the following, studying the decatenation checkpoint: i) two mutants: $Y_{\alpha}640F$, $Y_{\beta}656F$ in the DNAbd which negatively affects the efficiency of the decatenation process; ii) they observed a different mode of action in the decatenation process - when the CTD domain is missing from the topo IIα, this isoform decatenates the DNA more efficiently, whereas the efficiency of the topo IIβ decatenation slightly decreases in the absence of the CTD domain compared to the wild-types of the two isoforms (Kozuki et al., 2017); iii) compared to topo IIα, topo IIβ wild type is more efficient in decatenation, but the decatenation is the most effective when both isoforms are present. In the missing CTD constructs, only the nuclear localization sequence sections of the CTD of each isoform were retained - the K1201-Q1453 & F1498-F1531 regions of the topo IIα and the K1219-S1521 & Q1574-N1621 regions of the topo IIβ (Kozuki et al., 2017).

Through this study, we tried to find answers to explain these differences that were experimentally observed. To start with, having the models of the two isoforms' monomers in the DNAbd domain, we have tried to address the mutant problem and for this it is necessary to know their positions in space within the topoisomerase dimer.

The dimer models for the two isoforms were generated by superimposing the monomer models on each monomer in the crystallized dimer structures from the bacteria, while the primary dimerization interface was remodeled using the 2wl2.pdb crystal as template (the only structure with coordinates in that region, compared to the missing coordinates in the 3l4j.pdb crystal).

Shortly after we sent to our collaborators the dimer models for the two isoforms, the crystallized structures of the dimeric DNAbd domain were published for both human isoforms - the topo IIβ structure first (PDB code - 3qx3.pdb) (Wu et al., 2011) and then the topo IIα structure (PDB code - 4fm9.pdb) (Wendorff et al., 2012).

From the comparison of the crystals (colors with open tones in Fig. IV.4) with the homology models (colors with dark shades in Fig. IV.4) for the dimers of both isoforms results a very good overlap on the approximately 1500aa (the total aminoacids in a DNAbd dimer), so the RMSD between the crystal and the model in the case of topo IIβ is 3.5Å, and in the case of topo IIα the RMSD is 2.8Å. These results indicate that the generated dimer models are of very good quality and, therefore, increase the confidence in the modeling methodology developed within our laboratory.

Since the regulation of the decatenation checkpoint in the G2 phase of the cell cycle is influenced by the mutation of the $Y_{\beta}656/Y_{\alpha}640$ tyrosines to phenylalanines and to try to explain their different response, we analyzed in more detail the subdomain B' of all published DNAbd crystal structures so far.

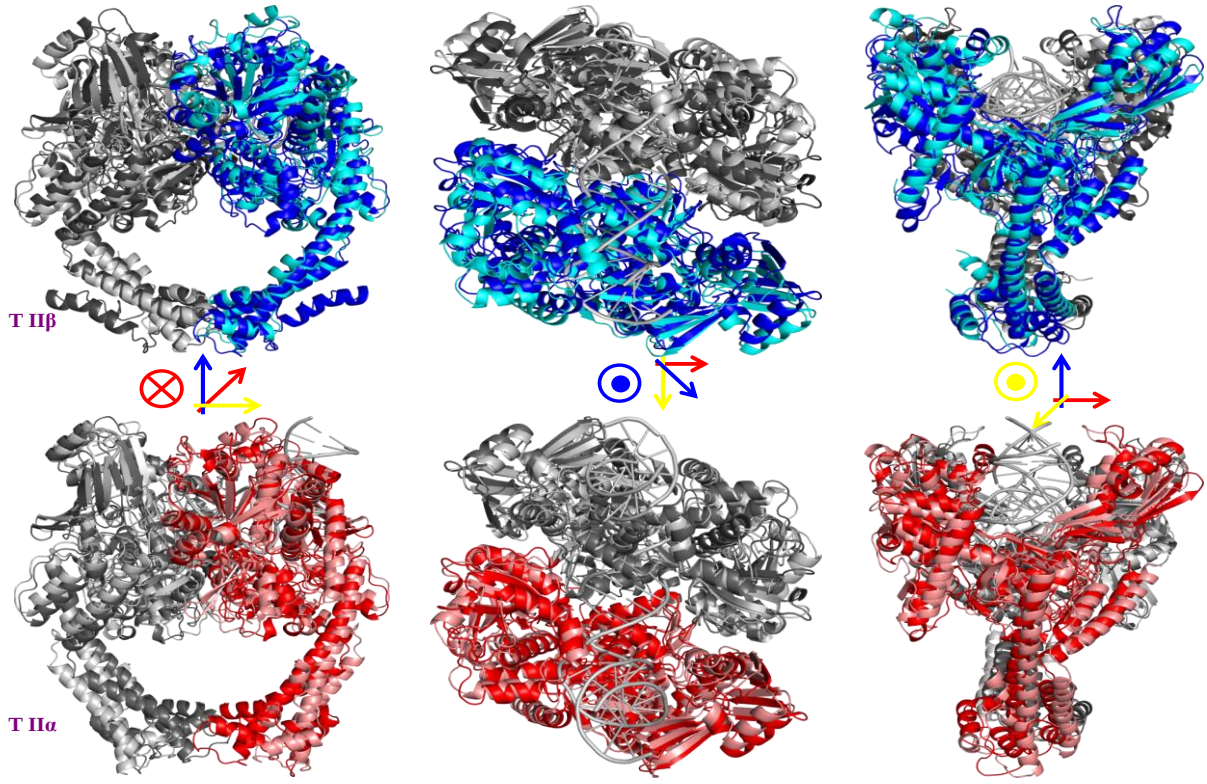

Fig. IV.4 The superposition of the topo IIα and topo IIβ crystals onto the homology models of the dimers

The analyzed structures are as follows: - in the case of topo IIα there is only one published crystal (4fm9.pdb) (Wendorff et al., 2012) whereas for topo IIβ there are 5 crystallized structures (3qx3.pdb, 4g0u.pdb, 4gov.pdb, 4g0w.pdb, 4j3n.pdb) (Wu et al., 2011; Wu et al., 2013). The crystal analysis reveals, at first glance, that in the top IIβ crystals there are two regions in the TOPRIM domain that have no coordinates (they are missing in the electronic density maps): - "*R-1*" (in the 593-IVKA...GTST-636 range, which corresponds to a "Greek

key"-like motif in topo IIα) and "*R-2*" (in the 696-LPEQ...YGTA-705 range, corresponding to an unstructured loop in topo IIα, in the proximity of $Y_{\alpha}640$ tyrosine).

The absence of these regions in the electronic density map of topo IIβ crystals suggests that they can be found in multiple configurations in this isoform (Kozuki et al., 2017). This situation may occur due to the lack of critical "*anchor*" contacts with respect to the rest of the protein and/or perhaps due to configuration changes induced by a stronger interaction with the T-segment DNA, in the case of topo IIβ.

A clear evidence of the lack of such “anchors” in topo IIβ is that in the case of topo IIα, the R-1 and R-2 regions are stabilized by a H-bond between $E_{\alpha}597$-$Y_{\alpha}684$. This connection is lost in the case of topo IIβ, since $Y_{\alpha}684$ is replaced by $F_{\beta}700$. Another possible anchor in topo IIα is the H-bond of $S_{\alpha}621$ (from the C-terminus of the R-1 region) to $D_{\alpha}963$ (from the DNAbd domain on the other monomer), a link that is lost in the case of topo IIβ, because the serine is replaced with the $A_{\beta}637$ alanine.

Using the crystallized topo IIα structure I measured all the distances ≤ 4.5Å between all atoms in the R-1 and R-2 regions with all the atoms in the rest of the protein (including those on the neighboring monomer) using a program that I wrote in the AWK programming language and which I developed in the present study (Kozuki et al., 2017).

To perform the same measurements in the case of topo IIβ, I modeled by homology modeling the two R-1 and R-2 regions. Based on the distance measurements for the two isoforms, a loss of favorable contacts was observed in the case of topo IIβ compared to topo IIα. Thus, in topo IIα there are more interactions which contribute to the general stability of the R-1 and R-2 regions (Kozuki et al., 2017).

It is interesting to note that around these "*flexible*" regions, the number of basic amino acids in topo IIβ is increased compared to topo IIα (Fig. IV.8).

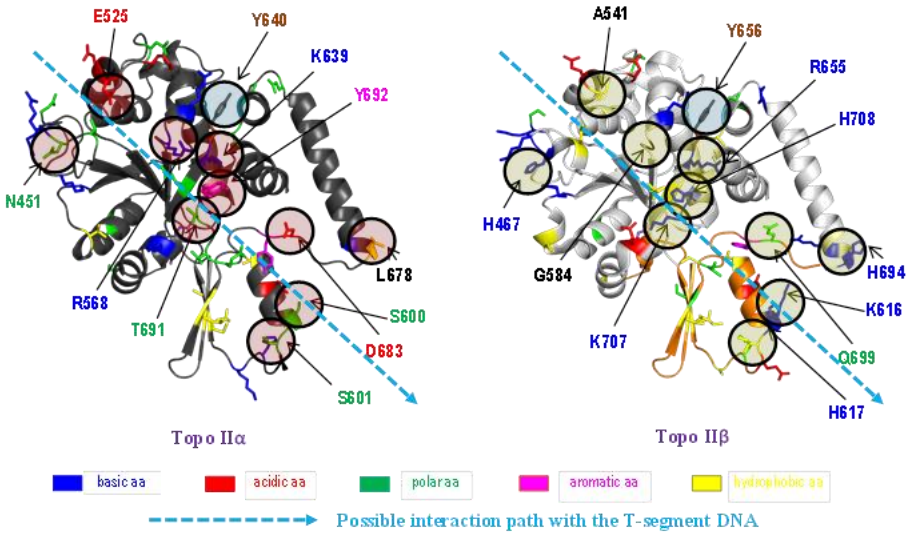

Fig. IV.8 The differences between topo IIα and topo IIβ on the TOPRIM
(adapted after (Kozuki et al., 2017))

Including the two R-1 and R-2 regions (orange-colored in Fig. IV.8), there are 5 additional positive charges compared to topo IIα in the TOPRIM domain of topo IIβ, while $D_{\alpha}683$ of topo IIα is replaced with $Q_{\beta}699$ in topo IIβ, as can be seen in Fig. IV.8 (labels were added only to the most different amino acids between the two isoforms). A possible interpretation of these observations is that the T-segment DNA could interact and could be better aligned with the TOPRIM domain of topo IIβ compared to topo IIα, where this interaction would not be as significant (Figure IV .8) (Kozuki et al., 2017). All of these observations resulting from the analysis of the region neighboring the $Y_{\beta}656/Y_{\alpha}640$ tyrosines would indicate that the interaction between the T-segment DNA with the TOPRIM domain, as well as the decatenation, could be affected by the $Y_{\alpha}640F$ and $Y_{\beta}656F$ mutations, which is consistent with the experimental results (Kozuki et al., 2017).

From the analysis of the CTD domains, TOPRIM subdomain analysis and the recent studies on the affinity of topo IIβ binding to DNA (Gilroy and Austin, 2011), and the fact that topo IIα decatenates negative supercoilings much faster than positive ones in the absence of the CTD (Seol Et al., 2013) (possibly involving T-segment arrest in positive supercoiling when CTD is present), it can be concluded that the T-segment could interact with the TOPRIM region in the close proximity of $Y_{\beta}656/Y_{\alpha}640$ tyrosines and their mutations to phenylalanine and/or specific differences in CTD could affect the process of decatenation.

The motivation behind this study was to better understand the mechanism of action of the topoisomerase II isoforms in humans, and we also tried to identify the structural properties that determine their different action under different conditions (e.g. mutations, lack of CTD domains) through sequence analysis, DNAbd domain modeling and predictions.

THE "DECAPPING SCAVENGER" (DcpS) ENZYME

Last but not least, in this thesis I also studied the DcpS enzyme from *Caenorhabditis elegans* (DcpS_Ce). The DcpS enzymes are part of the HIT pyrophosphatases family containing the HIT motif, crucial for the hydrolysis of triphosphate-5'-5' bond at the mRNA end.

The reason for choosing the *C. elegans* protein (despite the fact that the human counterpart structure is known) to study is to develop a series of DcpS inhibitors that will be used to treat diseases where a decrease of DcpS activity leads to a reduction of the pathology (e.g. spinal muscular atrophy) (Singh et al., 2008) but also as therapeutic agents in a series of cancers. In addition, *C. elegans* is one of the most studied, and easier to test, organisms but too few were

known about the structure of DcpS in this organism, but also about its affinity for native substrates or other derived compounds.

The DcpS_Ce sequence was made available by our collaborators at the Department of Biophysics of the Institute of Experimental Physics, Faculty of Physics, Warsaw University, Poland. They tested the substrate specificity of a series of analogous compounds to the dinucleotide ends, with modifications in the first m^7Guo (N-methylguanosine) or in the second nucleoside (Wypijewska del Nogal et al., 2013). They have observed experimentally that these changes do not affect the degradation of the structures on the ends and moreover they have noticed the impact that various functional groups have on the substrate specificity and on the hydrolysis rate. However, the structural bases and mechanisms that determine the selectivity of these enzymes for various compounds were unknown, which is the motivation of the present study. The studied analogues are based on the m^7GpppG structure which underwent a series of modifications, especially in the phosphate bridge region; and thus, differ in size, the electronegativity of the substituent, etc. For each analogue, equilibrium constants (K_{AS}) and Gibbs free energy of binding (ΔG^0) for the wild-type (WT) enzyme were measured by our collaborators, without affecting their native binding affinity (Wypijewska et al., 2010).

In order to better understand the binding of the analogues to the active site, but also to highlight the mechanism by which some compounds interact with the amino acids in the active site, we generated a homology model of the enzyme DcpS_Ce and then we performed docking experiments using a series of compounds. The DcpS enzyme is active as a homodimer.

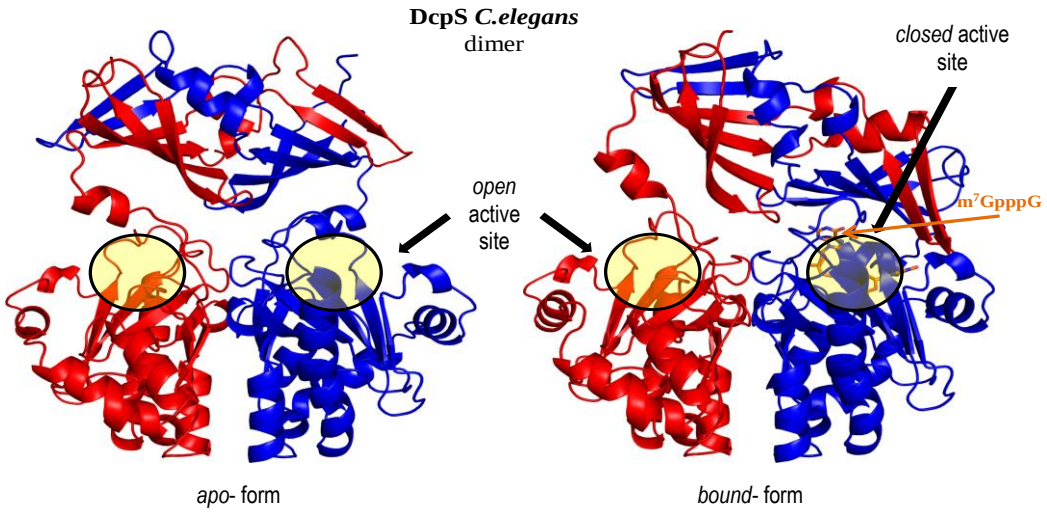

Fig. V.2 The DcpS_Ce models

We have generated three-dimensional models of the DcpS_Ce enzyme, in *bound* to mRNA end structure form (asymmetric conformation of the monomers, having an occupied site and a free one, on opposite sides of the dimer) and in the *apo* form (inactive or unbound, with symmetry between monomers), as there is a change in domain conformation.

In order to better understand the ligand binding pattern and the conformations in the active site of *C. elegans* DcpS, we performed DcpS (Ce) docking studies with several classes of ligands derived from m^7GpppG: - mononucleotides with varying number of phosphate groups (m^7GMP, m^7GDP and m^7GTP); - ARCA-type compounds ($m_2^{7,2'-O}GpppG$ and $m_2^{7,3'-O}GpppG$); - analogues containing bridge modifications in the phosphate cutting site (m^7GpCH_2ppG and $m^7GpNHppG$) (Fig. V.3) (Wypijewska del Nogal et al., 2013).
The docking experiments between DcpS and analogue compounds were performed using AUTODOCK 4.2 from AutoDock Tools 1.5.4 (Morris et al., 2009). For simplicity, we only considered the active site in closed conformation of the *C. elegans* DcpS enzyme.

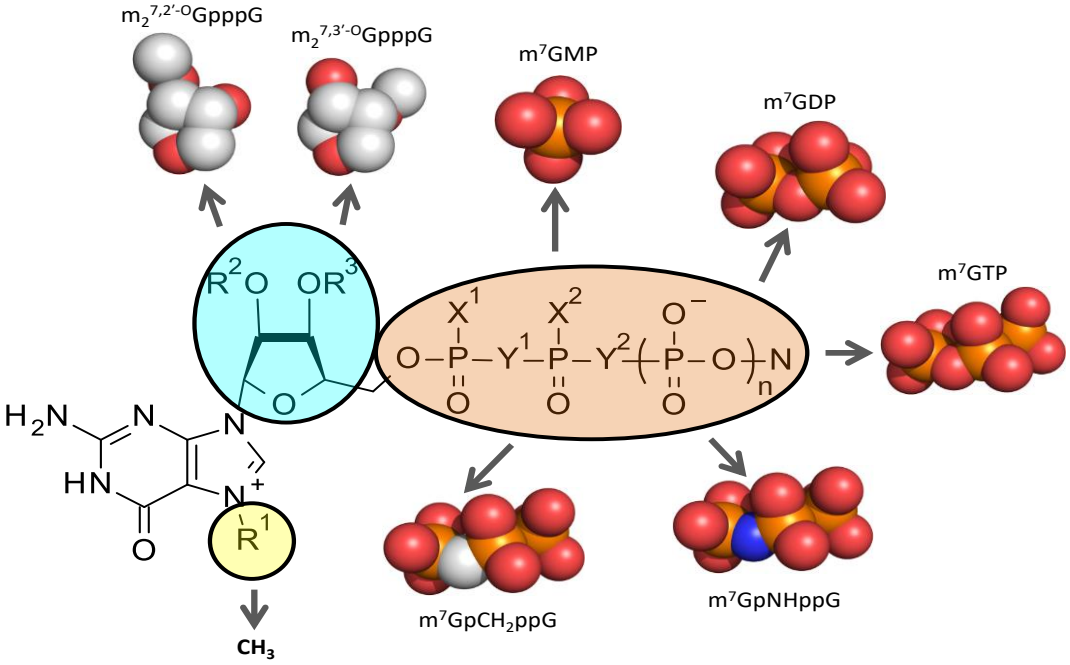

Fig. V.3 The structures of the m^7GpppG analogue ligands used in docking experiments, with modification in the ribose and phosphates regions
(adapted after (Wypijewska del Nogal et al., 2013))

We generated the analogous structures, starting from the m^7GpppG structure in the closed-site of the H. sapiens crystal 1st0.pdb, using the PyMol *Builder* module and the geometry of each ligand was optimized using the Avogadro 1.0.0 (Hanwell et al., 2012). From the classification of conformations obtained by docking for each ligand, based on the lowest free energy values,

the hierarchy of the docking experiments was similar to that of the experimental measurements.

Final Conclusions

On V(D)J Recombination topic, studies conducted by DBSB in collaboration with the Department of Immunobiology of the University of Yale, the following conclusions can be drawn:

- By modeling the RAG1/2 proteins and 12/23RSS DNA fragments we have explored into more detail and better understood the structural mechanisms of somatic recombination (V(D)J).
- Based on the 12/23RSS DNA fragments modeling, we have succeeded in developing a *coarse-grain* method for DNA bending, a method that can be used in our future modeling projects.
- The DNA bending procedure is a trusted one, this resulting also from the experimental validation (e.g. new FRET experiments based on the predicted structure).
- In RAG2 protein modeling, I developed a program (written in the AWK programming language) that helps to analyze more easily a large amount of molecular dynamics simulation trajectory, performing a detailed analysis of the Phi and Psi angles, and obtaining distributions of these torsion angles.
- The docking procedures underlying the formation of the RAG1/2-12/23RSS complex, as well as the information obtained from the analysis and modeling of all these structures, led to the assembly of the DNA (relative to the RAG1/2 heterotetramer) very similar to that observed by X-ray crystallography.

On the topic of the role of IIα/IIβ topoisomerases in decatenation, studies conducted by DBSB in collaboration with the Cancer Institute of the Carolina Healthcare System, the following conclusions can be drawn:

- The modeling of the two isoforms of human topoisomerase II helped to better understand their mechanism of action in DNA decatenation.
- The detailed analysis of the sequences of the two isoforms, the post-translational modifications predictions and the intense literature analysis resulted in very similar models to the structures obtained experimentally.

- Based on distance calculation programs (written in the AWK programming language), we have been able to analyze in detail the topoisomerase structures and the key interactions that have a role in the stability of some DNAbd regions.
- The suggested action models support the experimental results obtained by our collaborators.

On the topic of *C. elegans* DcpS enzyme inhibitor interactions, studies conducted by DBSB in collaboration with the Faculty of Physics of the University of Warsaw, the following conclusions can be drawn:

- Based on the *C. elegans* DcpS model we succeeded in developing a compound analysis using docking methods.
- The results, supported by the data obtained experimentally by our collaborators, help to understand the functional mechanisms of mRNA degradation, as well as the development of new therapeutic agents.

As a result of all these collaborations in which we have succeeded in developing some methods or improving others, in which we developed IT tools, the following conclusions can be underlined:

- The homology and remote homology modeling procedures described in this Ph.D. thesis have resulted in very good quality models which were subsequently validated by three-dimensional structures obtained by experimental methods such as X-ray crystallography.
- To facilitate the analysis of the three-dimensional structures of biomolecules, of contacts between them or contacts between amino acids in a particular region of a protein, of molecular dynamics trajectories, of selection of the best frames in a simulation, throughout this Ph.D. thesis I developed a series of small programs/scripts that will also be useful in the future projects within the laboratory
- Throughout this thesis I have developed programs for the analysis of sequence databases (from different species); in this case, I have carried out various statistical analyses of possible N-glycosylation sites.
- This PhD thesis, in addition to the results observed in each direction studied, has led to a better understanding of interactions between proteins and nucleic acids, and their

implications. Moreover, all this accumulated information will help me in future research projects.

List of publications and conference participations

Published papers:

Published articles and book chapters as first author:

1. Kozuki T*, Chikamori K*, ***Surleac MD****, Micluta MA, Petrescu AJ, Norris EJ, Elson P, Hoeltge GA, Grabowski DR, Porter ACG, Ganapathi RN, Ganapathi MK. "*Roles of the C-terminal domains of topoisomerase IIα and topoisomerase IIβ in regulation of the decatenation checkpoint*." ***Nucleic Acids Res***. **45(10)**, 5995-6010 (2017) [PMID: 28472494]

   IF: 10.162; **AI**: 3.5

   Kozuki T*, Chikamori K*, **Surleac MD***: authors with equal contributions

2. **Surleac MD**, Spiridon LN, Tacutu R, Milac AL, Petrescu SM, Petrescu AJ, *"The Structural Assessment of Glycosylation Sites Database – SAGS – An Overall View on N-Glycosylation."*, ***Glycosylation,*** Chap. 1, 1-22 (2012). [DOI: 10.5772/51690]

Published articles as a co-author:

3. Ciubotaru M, **Surleac MD**, Metskas LA, Koo P, Rhoades E, Petrescu AJ, Schatz DG, *"The architecture of the 12RSS in V(D)J recombination signal and synaptic complexes"*, ***Nucleic Acids Research***; 43(2):917-931, 2015. [PMID: 25550426]

   IF: 9.202**; AI**: 3.5

4. Ciubotaru M, Trexler AJ, Spiridon LN, **Surleac MD**, Rhoades E, Petrescu AJ and Schatz DG, *"RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complex"*, ***Nucleic Acids Research***; 41(4):2437-2454, 2013. [PMID: 23293004]

   IF: 8.808; **AI**: 3.4

5. Zhang YH, Shetty K, **Surleac MD**, Petrescu AJ, Schatz DG, *"Mapping and Quantitation of the Interaction Between the Recombination Activating Gene Proteins RAG1 and RAG2"*, ***J Biol Chem*****;** 290(19):11802-17, 2015. [PMID: 25745109]

IF: 4.258; **AI**: 1.7

6. Anna Wypijewska del Nogal, **Marius D. Surleac**, Joanna Kowalska, Maciej Lukaszewicz, Jacek Jemielity, Martin Bisaillon, Edward Darzynkiewicz, Adina L. Milac, Elzbieta Bojarska, *"Analysis of Decapping Scavenger (DcpS)-cap complex using modified cap analogs reveals molecular determinants for efficient cap binding"*, ***FEBS Journal***; Volume 280, Issue 24, pages 6508–6527, 2013. [PMID: 24119043]

IF – 3.986; AI - 1.3

7. Mihai Ciubotaru, **Marius Surleac**, Mihaela G. Mușat, Andreea M. Rusu, Elena Ioniță, Paul C. C. Albu, *„DNA bending in the synaptic complex in V(D)J recombination: turning an ancestral transpososome upside down"*, ***DISCOVERIES*, Vol. 2**, No. 1, P.1-15, 2014. [DOI: 10.15190/d.2014.5]

Conference presentations:

Oral presentations and prizes

- "*Mass Spectrometry interactomics of topoisomerase IIα and IIβ involved in human carcinoma cell lines*", IXth Symposium "Acad. Nicolae Cajal", 13-14 May 2014, Bucharest, Romania - (**Marius D. Surleac**, Cristian Munteanu, Ram Ganapathi, Mahrukh Ganapathi, Andrei J. Petrescu). I have been awarded by "Academician Nicolae Cajal" foundation the *Herbert Berler-Barbu prize for the best young researcher presentation, 2014*.
- "*Modeling protein-DNA interactions with experimental constraints. A study case on RAG proteins and the paired complex formation*", XIth Symposium "Acad. Nicolae Cajal", 17-19 March 2016, Bucharest, Romania - (Petrescu A-J, Surleac MD, Spiridon LN, Ciubotaru M, Schatz D).
- "*Computationally guided research in molecular life science at IBAR*", 25 Years of Promoting Molecular Life Sciences in Romania - International Conference of the Romanian Society of Biochemistry and Molecular Biology, 17-18 September 2015, Bucharest, Romania - (Andrei J. Petrescu, Adina Milac, Marius Micluţa, Laurențiu Spiridon, Marius Surleac, O. Căldăraru, Cristian V. Munteanu).
- "*Modelling Structures in the Twilight Zone and Beyond. Lessons from Resistance and Effector Gene Families*", COST FA 1208 Workshop: Structure-guided Investigation of Effector Function, Action and Recognition, 10-12 September 2014, Bucharest, Romania - (Spiridon LN, Milac AL, Surleac MD, Căldăraru O, Petrescu AJ).

- "*Identifying interactors of human topoisomerase IIα and IIβ through combined bioinformatics and Mass Spectrometry*", The Annual International Conference of the RSBMB, 5-6 June 2014, Băile Felix, Romania - (Marius D. Surleac, Cristian Munteanu, Ram Ganapathi, Mahrukh Ganapathi, Andrei J. Petrescu).

Poster presentations and prizes

- "*Structural insights into the functional divergence of human Topoisomerase IIα and IIβ on the decatenation checkpoint*", The Annual International Conference of the Romanian Society for Biochemistry & Molecular Biology, 8-9 June 2017, Timișoara, Romania - (**Surleac D. Marius**, Ganapathi Mahrukh, Ganapathi Ram, Petrescu J. Andrei). I have been awarded *The first prize*, *Correlation Structure–Properties in Biological Models – Drug Design section*, by the Romanian Society of Biochemistry and Molecular Biology (RSBMB).

- "*Modeling the bent PC-23RSS DNA based on FRET data and molecular dynamics simulations*" și "*Fluorescent quantification of Gibbs free energy of cap binding to DcpS reveals DcpS-cap interactions*", 9th EBSA European Biophysics Congress, 13-17 July 2013, Lisbon, Portugal - (**M. D. Surleac**, L. N. Spiridon, M. Ciubotaru, D. G. Schatz, A.-J. Petrescu) și respectiv (A. Wypijewska, M. D. Surleac, J. Kowalska, M. Lukaszewicz, J. Jemielity, M. Bisaillon, R. E. Davis, E. Darzynkiewicz, A. L. Milac, E. Bojarska). Participating to this conference I won a student scholarship, awarded by the European Biophysical Society (EBSA).

- "*Modelling of the human Topoisomerase II dimer*", 2011 International Conference of RSBMB, 28-30 Septembrie 2011, Craiova, Romania - (**Marius D. Surleac**, Laurențiu N. Spiridon, Adrian G. Grozav, Ram Ganapathi, Andrei-Jose Petrescu). I have been awarded *The best poster prize,* by the Romanian Society of Biochemistry and Molecular Biology (RSBMB).

- "*De novo Peptide Design for Enhanced Heavy Metal Accumulation*", The Annual International Conference of the Romanian Society for Biochemistry & Molecular Biology, 8-9 June 2017, Timișoara, Romania - (Martin Eliza. C., Caldararu Octav, Ruta L. Lavinia, Ghenea Simona, Surleac D. Marius, Spiridon Laurentiu, Milac Adina, Farcasanu C. Ileana, Petrescu J. Andrei).

- "*Modelling protein-DNA complexes using FRET constraints*", COST FA 1208 Workshop: Structure-guided Investigation of Effector Function, Action and Recognition, 10-12 September 2014, Bucharest, Romania - (Surleac MD, Spiridon LN, Ciubotaru M, Schatz DG, Petrescu AJ).

- "*The new model of mRNA degradation based on the recent advances in the DcpS enzyme specificity towards m7GDP*", 1st Congress of the Polish Biochemistry, Cell Biology, Biophysics and Bioinformatics BIO 2014, 9-12 September 2014, Warsaw, Poland - (Anna Wypijewska del Nogal, Marius D. Surleac, Joanna Kowalska, Maciej Lukaszewicz, Janusz Stepinski, Richard E. Davis, Martin Bisaillon, Jacek Jemielity, Edward Darzynkiewicz, Adina L. Milac and Elzbieta Bojarska).

- "*Structure-affinity relationship (SAFIR) for the mRNA cap analogs binding to C. elegans DcpS enzyme*", XIVth Congress of the Spanish Biophysical Society (SBE 2014), 11-13 June 2014, Alcalá de Henares, Spain - (Anna Wypijewska del Nogal, Marius D. Surleac, Joanna Kowalska, Maciej Lukaszewicz, Jacek Jemielity, Martin Bisaillon, Edward Darzynkiewicz, Adina L. Milac and Elzbieta Bojarska).

- "*Role of EDEM3 in ERAD. one step at a time*", The Annual International Conference of the RSBMB, 5-6 June 2014, Băile Felix, Romania - (Cristian Marian Butnaru, Mărioara Chirițoiu, Marius Surleac, Andrei J. Petrescu, Ștefana M. Petrescu).

- "*Molecular modelling and docking offer structural insights into mRNA cap binding by C. elegans DcpS scavenger decapping enzyme*" și "*Structural Analysis of Wheat Pm3 Disease Resistance Protein*", 2012 International Conference of RSBMB, 13-14 September 2012, Bucharest, Romania - (Marius Surleac, Anna Wypijewska, Jacek Jemielity, Elzbieta Bojarska, Edward Darzynkiewicz, Andrei-Jose Petrescu, Adina-Luminița Milac) și respectiv (Laurențiu N. Spiridon, Marius Surleac, Andrei J. Pestrescu).

- "*An update on SAGS, the Structural Assessment of Glycosylation Sites database*", 21st International Symposium on Glycoconjugates, 21-26 August 2011, Viena, Austria - (Petrescu A-J., Spiridon L., Tăcutu R., Milac A., Surleac M.).

- "*Modelling the bend of DNA 23- & 12-RSS with experimental constraints*", "*Modelling, MD simulation and in-silico mutation of cecropin P for optimizing the interaction with tumor cell membranes*" și "*Modelling and MD simulation of CC domains of some R proteins*", Annual International Conference of the RSBMB, 23-24 September 2010, Bucharest, Romania - (Marius D. Surleac, Laurențiu N. Spiridon, Mihai Ciubotaru, David Schatz, A-J Petrescu), (Marius A. Micluță, Cătălina A. Nenu, Marius D. Surleac, Laurențiu N. Spiridon, A-J Petrescu) și respectiv (Laurențiu N. Spiridon, Marius A. Micluță, Marius D. Surleac, A-J Petrescu).

Selected references:

CHEN, N., WALSH, M.A., LIU, Y., PARKER, R., SONG, H. 2005. Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. *J Mol Biol*, 347(4):707-18.

CIUBOTARU, M., KRIATCHKO, A.N., SWANSON, P.C., BRIGHT, F.V., SCHATZ, D.G. 2007. Fluorescence resonance energy transfer analysis of recombination signal sequence configuration in the RAG1/2 synaptic complex. *Mol Cell Biol*, 27(13):4745-58.

CORNEO, B., MOSHOUS, D., CALLEBAUT, I., DE CHASSEVAL, R., FISCHER, A., DE VILLARTAY, J.P. 2000. Three-dimensional clustering of human RAG2 gene mutations in severe combined immune deficiency. *J Biol Chem*, 275(17):12672-5.

GILROY, K.L., AUSTIN, C.A. 2011. The impact of the C-terminal domain on the interaction of human DNA topoisomerase II α and β with DNA. *PLoS One*, 6(2):e14693.

GROZAV, A.G., WILLARD, B.B., KOZUKI, T., CHIKAMORI, K., MICLUTA, M.A., PETRESCU, A.J., KINTER, M., GANAPATHI, R., GANAPATHI, M.K. 2011. Tyrosine 656 in topoisomerase IIβ is important for the catalytic activity of the enzyme: Identification based on artifactual +80-Da modification at this site. *Proteomics*, 11(5):829-42.

GRUNDY, G.J., RAMÓN-MAIQUES, S., DIMITRIADIS, E.K., KOTOVA, S., BIERTÜMPFEL, C., HEYMANN, J.B., STEVEN, A.C., GELLERT, M., YANG, W. 2009. Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. *Mol Cell*, 35(2):217-27.

HUANG, S., TAO, X., YUAN, S., ZHANG, Y., LI, P., BEILINSON, H.A., ZHANG, Y., YU, W., PONTAROTTI, P., ESCRIVA, H., LE PETILLON, Y., LIU, X., CHEN, S., SCHATZ, D.G., XU, A. 2016. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell*, 166(1):102-14.

KAWANO, S., KATO, Y., OKADA, N., SANO, K., TSUTSUI, K., TSUTSUI, K.M., IKEDA, S. 2016. DNA-binding activity of rat DNA topoisomerase II α C-terminal domain contributes to efficient DNA catenation in vitro. *J Biochem*, 159(3):363-9.

KIM, M.S., LAPKOUSKI, M., YANG, W., GELLERT, M. 2015. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*, 518(7540):507-11.

MCCLENDON, A.K., GENTRY, A.C., DICKEY, J.S., BRINCH, M., BENDSEN, S., ANDERSEN, A.H. & OSHEROFF, N. 2008. Bimodal recognition of DNA geometry by human topoisomerase II alpha: preferential relaxation of positively supercoiled DNA requires elements in the C-terminal domain. *Biochemistry*, 47, 13169-13178.

MILAC, A.L., PETRESCU, A.J. 2001. Protein Structure Prediction. (I) Homology Modelling. *Rom. J. Biochem.*, 38 (2), 249 -275.

NITISS, J.L. 2009-A. DNA topoisomerase II and its growing repertoire of biological functions. *Nat Rev Cancer*, 9(5):327-37.

PAPILLON, J., MENETRET, J.F., BATISSE, C., HELYE, R., SCHULTZ, P., POTIER, N. AND LAMOUR, V. 2013. Structural insight into negative DNA supercoiling by DNA gyrase, a bacterial type 2A DNA topoisomerase. *Nucleic acids research*, 41, 7815-7827.

PAWLOWSKI, M., BOGDANOWICZ, A. & BUJNICKI, J.M. 2013. QA-RecombineIT: a server for quality assessment and recombination of protein models. *Nucleic Acids Res*, 41, W389–W397.

PETRESCU, A.J. 1999. Computer simulation in molecular biology. (I). basic concepts and algorithms. *Rev. Roum. Biochim.*, 36, 1-2, ll5-142.

RU, H., CHAMBERS, M.G., FU, T.M., TONG, A.B., LIAO, M., WU, H. 2015. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*, 163(5):1138-52.

SATHYAPRIYA, R., VIJAYABASKAR, M.S. & VISHVESHWARA, S. 2008. Insights into protein-DNA interactions through structure network analysis. *PLoS computational biology*, 4, e1000170.

SCHMIDT, B.H., BURGIN, A.B., DEWEESE, J.E., OSHEROFF, N., BERGER, J.M. 2010. A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. *Nature*, 465(7298):641-4.

SINGH, J., SALCIUS, M., LIU, S.W., STAKER, B.L., MISHRA, R., THURMOND, J., MICHAUD, G., MATTOON, D.R., PRINTEN, J., CHRISTENSEN, J., BJORNSSON, J.M., POLLOK, B.A., KILEDJIAN, M., STEWART, L., JARECKI, J., GURNEY, M.E. 2008. DcpS as a therapeutic target for spinal muscular atrophy. *ACS Chem Biol*, 3(11):711-22.

SWANSON, P.C., KUMAR, S., RAVAL, P. 2009. Early steps of V(D)J rearrangement: insights from biochemical studies of RAG-RSS complexes. *Adv Exp Med Biol*, 650:1-15.

VAN DIJK, E., LE HIR, H., SÉRAPHIN, B. 2003. DcpS can act in the 5'-3' mRNA decay pathway in addition to the 3'-5' pathway. *Proc Natl Acad Sci*, 100(21):12081-6.

WENDORFF, T.J., SCHMIDT, B.H., HESLOP, P., AUSTIN, C.A., BERGER, J.M. 2012. The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. *J Mol Biol*, 424(3-4):109-24.

WYPIJEWSKA, A., BOJARSKA, E., STEPINSKI, J., JANKOWSKA-ANYSZKA, M., JEMIELITY, J., DAVIS, R.E., DARZYNKIEWICZ, E. 2010. Structural requirements for Caenorhabditis elegans DcpS substrates based on fluorescence and HPLC enzyme kinetic studies. *FEBS J*, 277(14):3003-13.

YIN, F.F., BAILEY, S., INNIS, C.A., CIUBOTARU, M., KAMTEKAR, S., STEITZ, T.A., SCHATZ, D.G. 2009. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat Struct Mol Biol*, 16(5):499-508.