THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Unveiling Mechanistic Details of Macromolecular Interactions: Structural Design and Molecular Modelling of DNA-Protein Systems in Their Active State

ANNA REYMER

Department of Chemical and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2012

Unveiling mechanistic details of macromolecular interactions:
Structural design and molecular modelling of DNA-protein systems in their active state

ANNA REYMER

© ANNA REYMER, 2012.
ISBN 978-91-7385-657-7

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3338
ISSN 0346-718X

Department of Chemical and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0) 31-772 1000

Cover Art: © Anna Reymer.
A jigsaw puzzle of a cartoon representation of a human Rad51 recombinase model structure comprising of four protein monomers (orange and plum colours) assembled on an ssDNA fragment (green) presumably by means of two putative DNA binding loops (red). ATP molecules inbound at the interfaces between adjacent protein subunits coloured green, Ca^{2+} cations accompanying ATP coloured cornflower blue.

Backside photo: © Alex Voronov.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2012

Dedicated to the memory of my grandfather Viktor

UNVEILING MECHANISTIC DETAILS OF MACROMOLECULAR INTERACTIONS: STRUCTURAL DESIGN AND MOLECULAR MODELLING OF DNA-PROTEIN SYSTEMS IN THEIR ACTIVE STATE

ANNA REYMER

Department of Chemical and Biological Engineering
Chalmers University of Technology

ABSTRACT: Molecular structure is fundamental for understanding mechanisms of molecular interactions. This applies not least to understanding biological function: every biological cell, whether bacterial or human, is an immensely complex system comprising of thousands of molecules that exist in constant motion and interaction with each other. Structural properties of the two main classes of biological macromolecules, nucleic acids and proteins, have been studied in this Thesis focusing on functional interactions of both DNA and the DNA-binding enzyme human recombinase Rad51.

DNA is a highly polymorphic molecule and its plasticity is important for its function. Conformational mechanics of the DNA helix was addressed to understand interactions with dumbbell-shaped ruthenium(II) polypyridyl compounds, known for their remarkable ability to slowly thread one of their bulky centres through a tightly packed DNA stack, probably invoking large transient conformational rearrangements of the helix. Thread-intercalation rate is accelerated by several orders of magnitude if the DNA target sequence is a stretch of at least ten base pairs of AT, as well as by the hydrophobicity of the auxiliary "dumbbell" ligands: counter-intuitively, a spatially smaller and less hydrophobic compound takes longer times to thread. It is hypothesized that thread-intercalation might be facilitated by an A-like DNA conformation, induced by the outside binding of the Ru(II) compounds. An NMR study, aiming to solve a thread-intercalated structure of the binuclear Ru(II)-DNA complex, resulted in a groove-binding geometry, probably representing an initial binding mode preceding intercalation, a result emphasizing the elusiveness and immense complexity of the threading process. Turning back to simpler monomeric propeller-shaped Ru(II) compounds it was deduced that, despite acting as classic intercalators, they can "read out" the chirality of the DNA helix by enantiospecifically kinking it, in a fashion analogous to several families of operatory DNA binding proteins.

Another operatory protein, human recombinase Rad51, that facilitates homologous recombination, the process of exchanging near identical-sequence DNA strands, is also mechanically acting on DNA, by stretching it. A 3-D high-resolution model structure of human Rad51 filament was solved by a combination of polarized-light spectroscopic data and molecular modelling. Highlighted by the model some interesting structural features could be addressed: strategic locations of two putative DNA binding loops inside the protein filament; as well as location of a putative ATP binding site at the interface between two protein subunits and in direct proximity to a supposed location of DNA – could hint about DNA docking mechanism and potential role of ATP in the protein function.

The Rad51-DNA model has proven itself useful also in a follow-up study on the stimulatory role of Ca^{2+} in the strand exchange reaction by human Rad51. A mechanism is proposed involving a high affinity DNA binding state of the Rad51 filament induced by Ca^{2+}, regulatorily crucial for the search of homology and subsequent DNA strands exchange.

KEYWORDS: DNA, ruthenium(II) polypyridyl compounds, intercalation, homologous recombination, human Rad51, molecular modelling.

LIST OF PUBLICATIONS AND CONTRIBUTION REPORT

This Thesis is based on the work presented in the following papers:

Paper I

DNA Polymorphism as an Origin of Adenine-Thymine Tract Length-Dependent Threading Intercalation Rate

Pär Nordell, Fredrik Westerlund, Anna Reymer, Afaf H. El-Sagheer, Tom Brown, Bengt Nordén and Per Lincoln, *Journal of American Chemical Society*, **2008**, 130 (44), pp 14651–14658

A.R. proposed a molecular mechanism of threading interaction facilitated by a local conformational transition of the DNA double-helix from B-form to A-like form as a result of outside binding of binuclear Ru(II)-compounds; contributed to the interpretation of the experimental data and to writing the paper. A.R. did not perform the experiments and did not participate in planning of the experimental part of the study.

Paper II

Catching Precursory States of DNA Threading: NMR Structure Reveals Λ,Λ-[μ-bidppz(bipy)$_4$Ru$_2$]$^{4+}$ in the Minor Groove of [d(CGCGAATTCGCG)]2 with Long-Range Exclusion

Lisha Wu, Anna Reymer, Cecilia Persson, Krzysztof Kazimierczuk, Tom Brown, Per Lincoln, Bengt Nordén and Martin Billeter, *manuscript*

A.R. planned and performed model refinement based on NMR data with molecular dynamics simulations, participated in the interpretation of the experimental data and in writing the paper. A.R. did not perform the experiments and did not participate in planning of the experimental part of the study.

Paper III

Enantiospecific kinking of DNA by partially intercalating metal complex

Anna Reymer and Bengt Nordén, *Chemical Communications 2012, in press*

A.R. designed, planned and performed the computational work, analysed the results. A.R. and B.N. wrote the paper.

Paper IV

Structure of Human Rad51 Protein Filament from Molecular Modeling and Site-Specific Linear Dichroism Spectroscopy

Anna Reymer, Karolin Frykholm, Katsumi Morimatsu, Masayuki Takahashi and Bengt Nordén, *Proceedings of the National Academy of Sciences of the United States of America*, **2009**, 106 (32), pp 13248-13253

A.R. planned and performed the structural design and refinement of the model structure of the human Rad51 protein based on the data obtained from Site-Specific Linear Dichroism spectroscopy, participated in the interpretation of the experimental data and wrote the paper. A.R. did not perform the experiments and did not participate in planning of the experimental part of the study.

Paper V

Ca^{2+} Improves Organization of Single Stranded DNA Bases in Human Rad51 Filament – Explaining Stimulatory Effect on Gene Recombination

Louise Fornander, Karolin Frykholm, Anna Reymer, Axelle Renodon-Corniere, Masayuki Takahashi and Bengt Nordén, *Nucleic Acids Research*, **2012**, doi:10.1093/nar/gks140

A.R. modelled the structures of human Rad51 in its active state in complex with ssDNA, ATP/ADP and Ca^{2+}/Mg^{2+}. Based on the experimental results and the results from molecular modelling, A.R. proposed a molecular mechanism of stimulatory effect of Ca^{2+} on the strand exchange activity of the protein. A.R wrote the parts about molecular modelling and contributed to writing the other parts of the paper. A.R. did not perform the experiments and did not participate in planning of the experimental part of the study.

CONTENTS

1. MOLECULAR STRUCTURE AND FUNCTION ARE TWO INTERDEPENDENT ENTITIES

A biological cell – a highly organized and an extremely complicated system, where each component has a certain function directly coupled to its structure. The major components of the cell are DNA, RNA and proteins.

Deoxyribonucleic acid (DNA) – a biopolymer – the ultimate miniature data bank in the world, with the main role to preserve and pass on the genetic instructions used for development and existence of all living cells, as was for the first time demonstrated by Avery and co-workers in 1944 (Avery et al., 1944). However, only with the discovery of the structure of DNA duplex by Watson and Crick in 1953 (Watson & Crick, 1953), the mechanism by which the passage of genetic information is facilitated and copied has become evident. Watson and Crick in their model for the first time showed and explained the principle of complementarity of DNA strands – the key element of the duplex stability and function. Without this knowledge about DNA structure the subsequent articulation of the central dogma of molecular biology (Crick, 1958), that the genetic information is transferred over to proteins via RNA, would probably never have happened.

Ribonucleic acid (RNA), whose backbone differs from DNA's by just one oxygen atom and for this reason can not be quite as regularly folded as DNA, adopts a wide variety of shapes which in turn reflects its much more active role in catalysing biological reactions, controlling gene expression, or sensing and communicating responses to cellular signals (Fire, 1999; Lippman & Martienssen, 2004). However, RNA similarly to DNA is also used as storage for genetic information for primitive organisms like some viruses. It has been suggested that RNA is the archetype macromolecule, which gave rise not only to DNA but also to all proteins (Hillis & Dixon, 1991; Dolja et al., 1994). The living and functioning evidence in support for this hypothesis is ribosome (which is 90% RNA) – a crucial component of cells that synthesizes and folds protein chains (Ban, 2000; Cech, 2000).

And finally, *proteins* – the third and the largest in quantity class of macromolecules, the often called "working-horses", as they participate in virtually every process within cells. In contrast to DNA and RNA, proteins have much broader set of elementary units to be composed from, 20 amino acids in proteins versus only 4 nucleotides in DNA and RNA. And thereby this extended "vocabulary" is the reason for such an incredible diversity of protein folds – spanning up globular 3-D shapes. Some proteins fold into a rigid structure with small fluctuations, while others have much higher flexibility and incessantly undergo large rearrangements from one conformation to another. Most often, the structure of a protein defines and regulates its function, or its activity like in

the case of an enzyme. Not all proteins require a pre-folding process in order to function, as some may function also in an initially unfolded state. In either case, a mistake in folding can result in a complete or partial protein malfunctioning (Dobson, 2003; Mu et al., 2006).

However, none of the macromolecules exist and, most importantly, function in isolation. They constantly appear to be in contact performing thousands of chemical reactions, most of which are of vital importance for every living cell. A biological reaction is a complicated chain of events where several macromolecules can be involved simultaneously, each performing its characteristic role. While they are interacting, DNA, RNA and proteins alter each other's structure in a specific way, which defines the course, mechanism and outcome of the reaction. In essence, the knowledge of the structure of macromolecules and their associates will help in revealing and correcting, if necessary, their function. This knowledge is of strategic importance and can be further used, for example, for a development of new medicines against cancer or genetically passed-on diseases.

In the focus of this Thesis lies a set of scientific illustrations of how the knowledge of macromolecular structure can assist in understanding of mechanisms of molecular function. The two main story lines depicted in the appended papers are: DNA interactions with ruthenium(II)-coordinated polypyridyl compounds (Vos & Kelly, 2006); and structure and mechanisms of function of recombination proteins, RecA and Rad51 (Shinohara et al., 1992; Benson et al., 1994; Lusetti & Cox, 2002).

Despite all advances in comprehension of DNA structure and function accomplished since Watson's and Crick's discovery, there are still white spots in understanding of how plastic the DNA structure is, and what extreme conformational transitions it may undergo, as well as what potential biological significance they have. For example, Wilhelmsson et al. in 2002 (Wilhelmsson et al., 2002) showed that an immense ruthenium(II)-coordinated polypyridyl binuclear compound of a dumbbell shape could thread one of its bulky centres through a tightly-packed DNA stack. This surprising result has not only shown what a tremendous flexibility is hidden inside DNA but also raised the question whether this property is biologically relevant. It has been postulated, "any physical property intrinsic to components or interactions of a biological system, even when counteracting the most direct function, is often exploited by nature for a globally optimal function" (unpublished, B. Nordén).

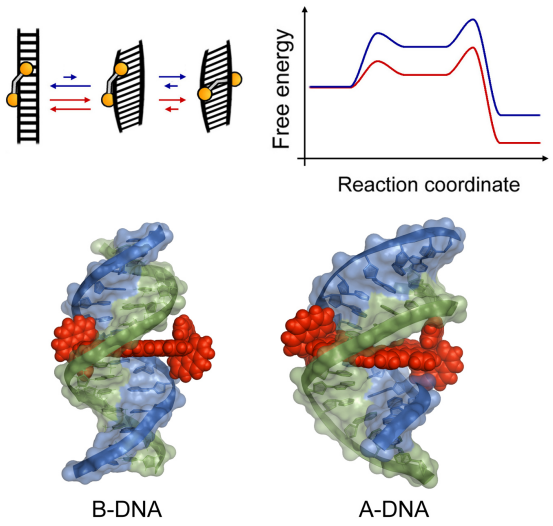
In *Papers I* and *II*, we have tried to address the question of function and possible relevance of thread-intercalation and to envisage, utilizing a variety of experimental techniques, assisted by molecular modelling, what will the DNA structure look like when subject to threading intercalation by this dumbbell-shaped ruthenium compound. In *Paper III*, we turned our attention to the simpler monomeric propeller-shaped ruthenium "mother molecule" of the dimeric derivative in order to get some mechanistic insight into interactions involved in the threading intercalation phenomenon of its more complex "offspring". There, addressing intercalative binding of Ru(II)tris(phenan-

throline) monomer to DNA we discovered why the two enantiomeric forms, the right- and left-handed propellers, give quite different hydrodynamic properties to the DNA duplex.

Recombination proteins promote the central steps of homologous recombination reaction (Shinohara et al., 1992; Kowalczykowski et al., 1994; Benson et al., 1994; Brendel et al., 1997) of exchanging DNA strands of similar sequences: they align and compare two DNAs and, depending on the degree of similarity, perform strand exchange between them. The homologous recombination reaction is of vital importance for maintenance of genetic diversity and DNA repair, and is ubiquitous in all living organisms from bacteria and fungi to mammalians. In spite of the evolutionarily conserved function and preserved key steps of the reaction, the structure of recombinases has evolved and the function appears to vary significantly between species such as bacteria and animals (Story & Steitz, 1992; Story et al., 1992; Brendel et al., 1997; Cox, 2003; Conway et al., 2004; Chi et al., 2006; Chen et al., 2008). To gain mechanistic insights into homologous recombination in humans we have designed a model of a filamentous structure of human recombination protein, HsRad51, *Paper IV*. In *Paper V,* through the structural analysis of the HsRad51 complexes formed on single-stranded DNA in the presence of ATP we suggested a mechanism of the stimulatory effect of Ca^{2+} on the strand exchange activity of human recombinase. Surprisingly, the bacterial recombinase, RecA, was never noticed to display a similar behaviour. Another peculiar discrepancy between Rad51 and RecA is the role and mechanism of ATP-hydrolysis, which I am currently investigating in detail for the case of RecA. Some of the results from that study form the basis of a separate chapter in the Thesis (see Chapter “On the Mechanism of ATP hydrolysis in Bacterial Recombinase RecA”).

To conclude, this Thesis aims to demonstrate the importance of knowledge of macromolecular structure in unveiling mechanistic details of their biological function; it provides examples of combinational approach coupling experimental data to molecular modelling to gain structural information and ways to utilize it.

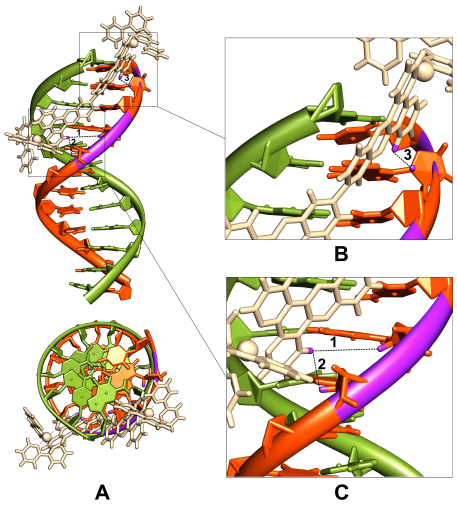
GRAPHICAL ABSTRACTS OF THE APPENDED PAPERS



Paper I

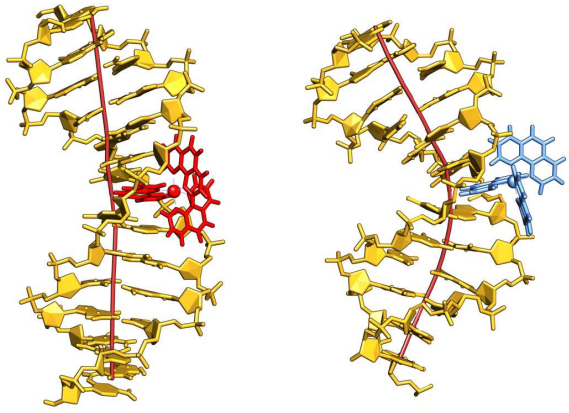
Top: Schematic representation of threading intercalation mechanism of a binuclear ruthenium(II) polypyridyl compound into DNA

Bottom: Two possible geometries of thread-intercalated state



Paper II

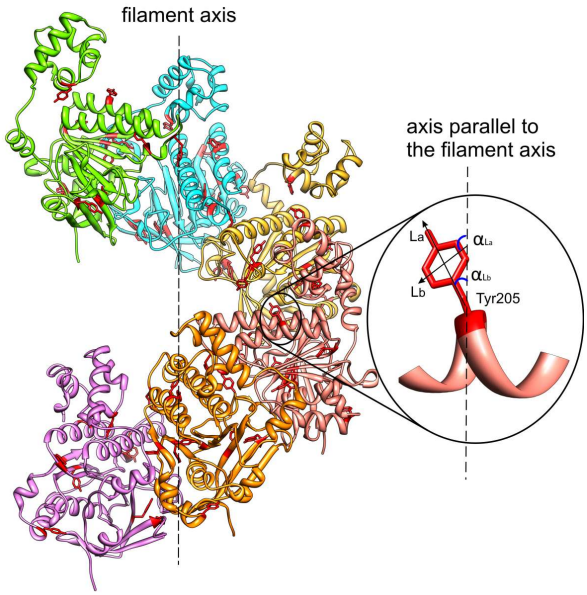
Catching precursory interactions of a binuclear ruthenium(II) polypyridyl compound and DNA – an NMR study



Paper III

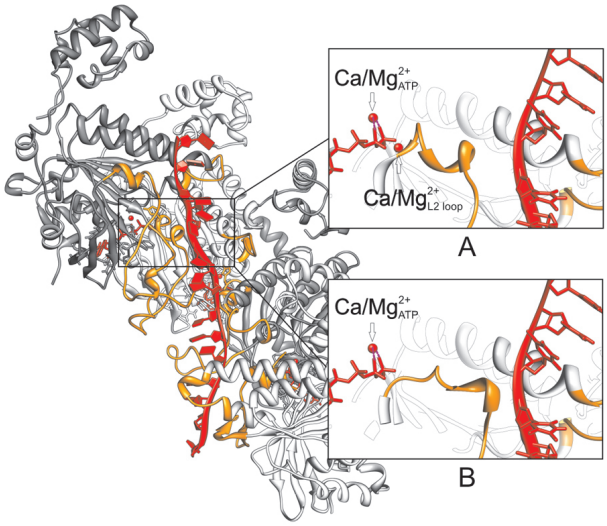
Binding of two enantiomeric forms of the very same mononuclear ruthenium(II) polypyridyl compound affects DNA differently

GRAPHICAL ABSTRACTS OF THE APPENDED PAPERS



Paper IV

Model structure of Human Rad51 protein – a modern analogue of a "Frankenstein monster", assembled from available structural information including a novel technique of Site-Specific Linear Dichroism spectroscopy.



Paper V

Model structure of the HsRad51-ssDNA-ATP filament. A cation, Ca^{2+} or Mg^{2+}, is found to stabilize an ordered conformation of a putative DNA binding loop L2, thus potentially regulating the binding affinity of ssDNA inside the protein filament.

2. BACKGROUND AND FUNDAMENTAL CONCEPTS

This chapter is intended to provide the basics of DNA and its structural polymorphism, describe possible ways of DNA-ligand interactions, introduce ruthenium(II)-coordinated polypyridyl compounds and their physico-chemical properties.

The reader will also get acquainted with the fundamental concepts of the homologous recombination reaction and its key players – recombination proteins RecA and Rad51, and learn about their structure and function.

The chapter concludes with a brief introduction to the most common structure-determination experimental techniques, their advantages and limitations.

NUCLEIC ACIDS AND RUTHENIUM COORDINATED POLYPYRIDIL COMPOUNDS

2.1. DNA Structure and Polymorphism

Deoxyribonucleic acid – DNA – the central molecule of life contains not only genetic information encrypting protein sequences but also instructions for the maintenance of the cell cycle. The molecule has an ingeniously simple structure, illustrated in **Figure 2.1**. DNA is a polymer built up of four different monomeric units, called nucleotides: adenosine- (A), guanosine- (G), thymidine- (T) and cytinidine- (C) monophosphates. Each nucleotide consists of a sugar, deoxyribose, attached via its 5' carbon atom to a phosphate group, and via its 1' carbon atom to an aromatic base: purine (A, G) or pyrimidine (T, C). Upon polymerization the phosphate group of one nucleotide connects to a 3' carbon atom of the sugar of another nucleotide, together they form a DNA's backbone of alternating phosphate-sugar groups. Two such chains can hybridize by the formation of hydrogen-bonded base-pairs to form a characteristic DNA double helix with anti-parallel negatively charged backbones on the outside and with nucleobases stacked like coins in a coin-pile in the centre due to hydrophobic and van der Waals interactions. Two DNA strands are held together by this base stacking and the specific pairing pattern between purine and pyrimidine bases, A always pairing with T by two hydrogen bonds and G always pairing with C by three hydrogen bonds.

DNA is a highly polymorphic molecule (**Figure 2.2**) and adopts a conformation that is the most favourable with respect to either surrounding physico-chemical environment, or local sequence motif, or upon interactions with other molecules. The most abundant DNA form in cells is the B-form – a double-stranded right-handed helix where the two strands are intertwined around each other with base pairs oriented with their planes nearly perpendicularly to the helix axis. Each helical turn comprises ten base pairs, the bases separated by 3.4 Å from each other and twisted by 36°. Two distinct deep grooves

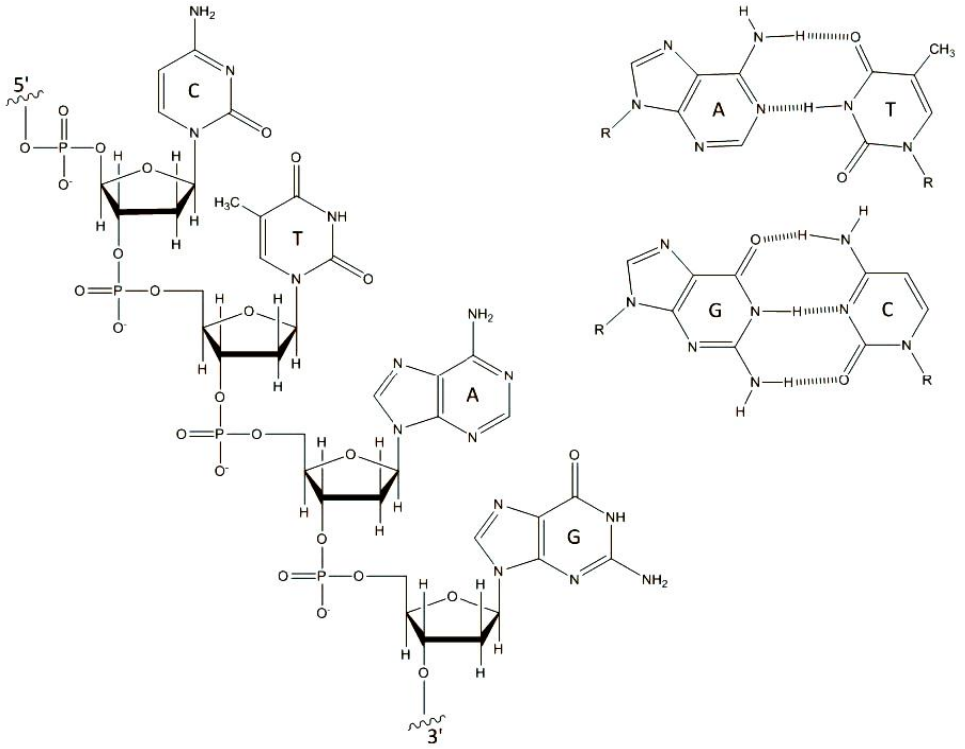


Figure 2.1. Left: A chemical structure of a DNA polynucleotide chain fragment showing alternating sugar-phosphate groups of the backbone. Right: Chemical structure and pairing pattern of purine-pyrimidine bases: adenine (A) = thymine (T) and guanine (G) ≡ cytosine (C) (R denotes connecting atom of backbone).

run along the duplex sides, one wide and one narrow, commonly known as major and minor grooves, respectively.

At low water activity a DNA helix adopts a different conformation, called A-form DNA, a wider and more compact duplex with highly inclined base pairs, positioned rather off from the helix centre, forming a hollow-barrel core. The helical rise, which is 3.4 Å in B-form, decreases to 2.9 Å, and the base pairs become slightly less twisted so that that the helical turn comprises 12 base pairs instead of 10. This results in a deepening of the major groove and a shallowing of the minor one. Besides that A-form occurs when DNA is in a surrounding of reduced water activity, double-stranded RNA also adopts the A-form conformation at normal solution conditions as a consequence of steric restrictions imposed by the presence of a hydroxyl group at the 2' C atom of the backbone sugar ring. Noteworthy is the fact that the A-form DNA, first found in stretched fibers and crystals with high concentrations of salts (Franklin & Gosling, 1953; Conner et al., 1984), was later rediscovered in a biological context, in a complex with TATA-box binding protein (Kim et al., 1993a; Kim et al., 1993b; Nikolov et al., 1996; Olson et al., 1998; Lu et al., 2000).

Another distinct conformation of DNA is Z-form. Being generally unfavourable, the conformation, though, has higher chances to occur in the regions with alternating purine-

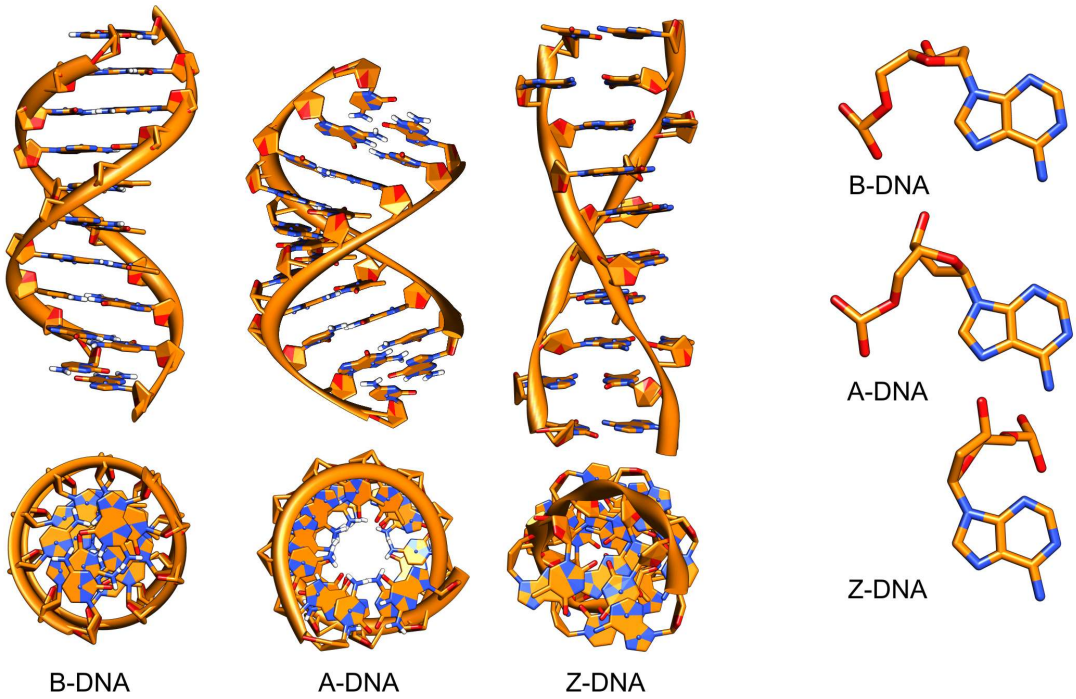


Figure 2.2. Left: B-, A-, Z- conformations of the DNA duplex. Right: the corresponding variations in the conformation of the backbone of adenosine monophosphate.

pyrimidine sequences, especially $[poly(dGdC)]_2$, at negative DNA supercoiling, or at high salt concentrations. In contrast to B- and A-forms, Z-DNA is a left-handed elongated double helix with the helical rise and twist of 3.8 Å and 30°, respectively. The major groove is flat or even convex, while the minor groove is deep and narrow. The potential biological role of Z-DNA, if any, has been debated. However, it is commonly believed that the form could provide torsional strain relief while DNA transcription occurs (Rich & Shuguang, 2003; Ha et al., 2005).

What makes DNA so extensively polymorphic is the flexibility of its backbone, and primarily the properties of the sugar (furanose) rings. The planar aromatic bases have themselves no internal conformational flexibility. The 5-membered furanose ring is acting like a mechanical joint between the base and a phosphate group. While changing its conformation at an energy cost of just a few kcal/mol (Olson & Sussman, 1982; Foloppe et al., 2002) it transforms the nucleotide unit, and the whole DNA conformation, from one form to another. The effect of a poly-conformational nature of the furanose ring is illustrated in **Figure 2.2** (to the right), by a comparison of structural variations for a nucleotide unit of B-, A- and Z-DNA forms.

Any DNA conformation can be described in terms of helical parameters (pitch and diameter of the helix), groove parameters (depth and width), a furanose ring conformation; rotational (inclination and tip) and translational (x- and y-displacement) base pair-axis parameters, together with six intra-base parameters: three rotational (buckle, propeller and opening) and three translational (shear, stretch and stagger), and six

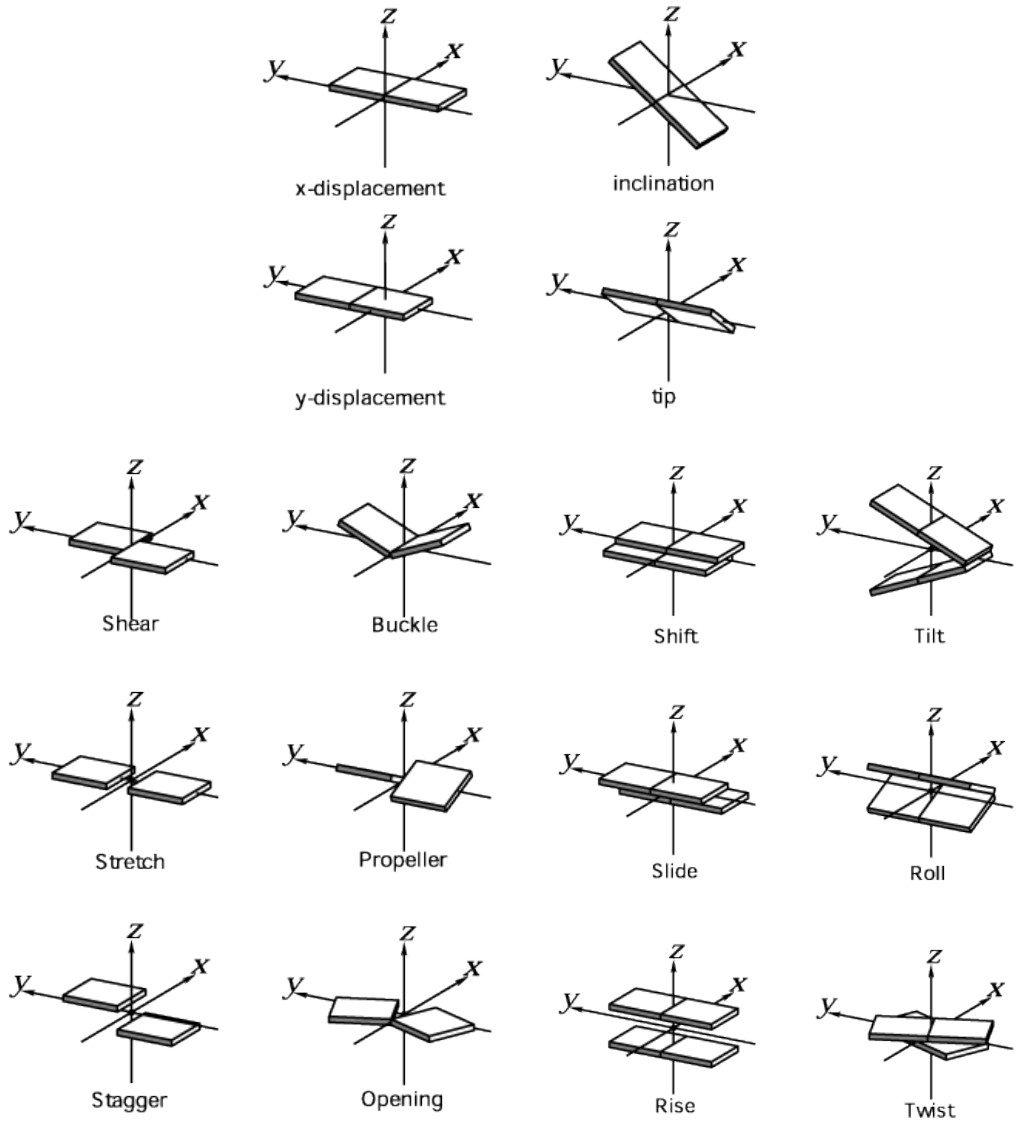


Figure 2.3. Translational and rotational base pair-axis parameters (top), inter-base pair parameters (lower left), and intra-base pair parameters (lower right). Fragments of the figure are adopted from http://3dna.rutgers.edu/x3dna/examples.

inter-base parameters: three rotational (tilt, roll and twist), and tree translational (shift, slide and rise). Base pair-axis, intra-and inter-base parameters are illustrated in **Figure 2.3.**

In addition to the mentioned B-, A- and Z-DNA forms there is a variety of transient or atypical conformations, which can be observed upon DNA interactions with proteins and ligands, and which will be briefly described in subsequent chapters.

2.2. DNA-Ligand Interactions

The primary DNA structure – the polynucleotide chain – is connected by strong covalent bonds, while, the overall 3-D DNA folding is determined by much weaker forces

including hydrogen bonds, van der Waals, electrostatic and hydrophobic interactions. These weaker forces, being comparable in order of magnitude to the average kinetic energy of molecules at physiologically relevant temperatures, account for most of the molecular interplay, structural equilibria and reorganizations inside a cell. While these interactions, balancing each other, define intramolecular interactions that determine secondary and tertiary structures, the intermolecular associations and binding-modes relative to DNA are predefined by structural features and can be divided into two distinct types: groove binding and intercalation.

DNA is a poly-anion, which defines the first binding-mode – groove binding. DNA's sugar-phosphate backbone with its negative charge, distributed along the walls of the grooves, makes DNA attractive to all positively charged molecules, from counterions, e.g. Na^+ or K^+, to any bigger cationic species: organic molecules, e.g. DAPI, or DNA-binding proteins. In the case of small and highly mobile monovalent counterions that bind in an unspecific entropically driven manner, their interaction with DNA is completely defined by electrostatic forces. A significant contribution to binding of a multivalent ligand is the entropy increase due to release of monovalent counterions. Bigger organic molecules, on the other hand, often associate with DNA in a more specific way. They can recognize DNA sequences or structure via a pattern of hydrogen bonds or a particular conformation. Noteworthy, smaller groove-binding agents, like DAPI or a family of Hoeschst 33258 microscopy stain dyes, bind preferentially into the minor groove where they can gain some hydrophobic protection for their aromatic ring-systems (Kubista et al., 1987; Pjura et al., 1987; Wilson et al., 1990; Clark et al., 1997). By way of contrast, DNA-binding proteins, like a homeodomain-binding protein or a zinc-finger protein, interact with DNA via the major groove. They recognize DNA bases via positively charged residues located on their externally positioned α-helix and thus need a wider docking spot (Billeter et al., 1996; Fraenkel et al., 1998; Tsui et al., 2000; Birrane et al., 2009; Schuetz et al., 2011; Xu et al., 2011). Some examples of groove-binders and their complexes with DNA are illustrated in **Figure 2.4**.

The hydrophobic stacking of aromatic residues, another prominent feature of the DNA duplex, determines the second essential binding mode – intercalation. It is defined as an insertion of a molecule or a part of it between adjacent DNA base pairs. This binding-mode is more abundant for fused aromatic ring-systems of hydrophobic nature, e.g. ethidium bromide (Reinhardt & Krugh, 1978; Laugaa et al., 1983) or 2,7-diazapyrene (Becker & Nordén, 1997). To accommodate an intercalator a sufficient spatial gap, comparable to the rise of one base pair (3.4 Å), has to be formed, which requires DNA to unwind. For example, in the case of ethidium bromide the unwinding was measured to be 26°. Since it was first discovered by Lerman in 1961 (Lerman, 1961), who studied DNA interactions with antiviral acridine proflavine, intercalation has for a long time been associated specifically with small planar aromatic ligands. However, with the increasing number of solved protein-DNA crystal structures it became evident that this binding mode is also relevant in the context of biomolecular interactions. Operative pro-

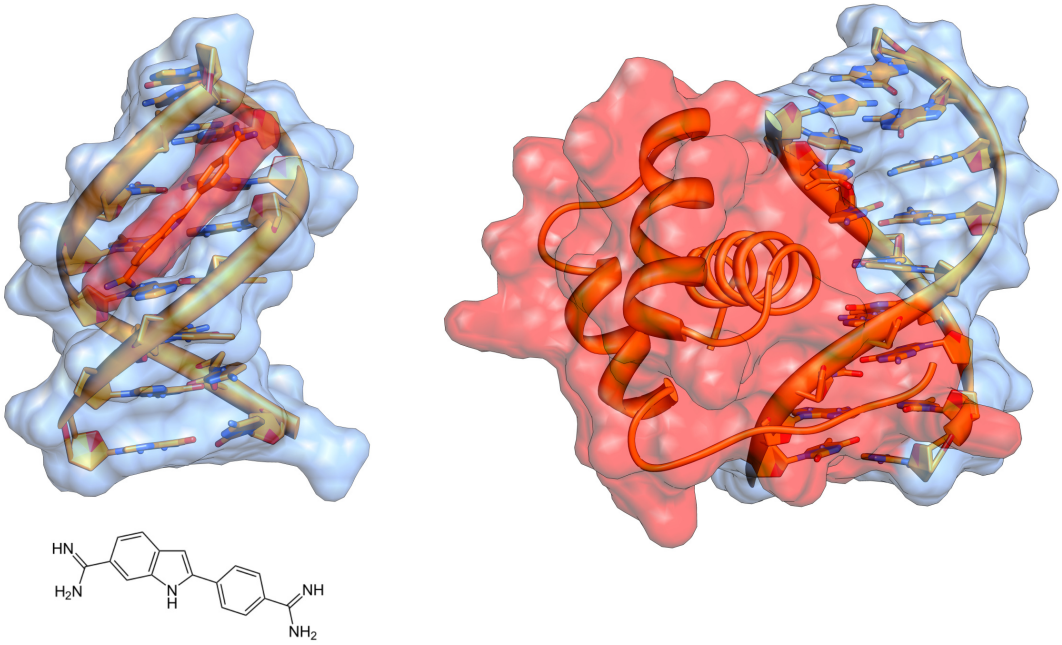


Figure 2.4. DNA groove binders: left – chemical structure of DAPI (below) and cartoon representation of its complexes with DNA (above); right – DNA-homeodomain-binding protein complex, cartoon representation.

teins, like recombination protein RecA or TATA-box binding protein TBP, use intercalative binding to manipulate the structure of DNA. RecA elongates and unwinds the DNA duplex preparing it for the strand exchange reaction (which will be further discussed in subsequent chapters), whereas TBP bends the DNA helix by as much as 80° (Kim et al., 1993a; Kim et al., 1993b; Nikolov et al., 1996) when initiating transcription reaction**.** Some examples of intercalating agents and their complexes with DNA are illustrated in **Figure2.5**.

Combining the best components from both types of interactions, where the groove binding provides the sequence specificity and the intercalation adds extra stability, the DNA structure and function could be controlled. Indeed, the formation of an intercalation pocket depends very little on the sequence of surrounding base pairs. Thus to gain sequence selectivity an intercalator must be complemented by some external moiety that could reside in either (or both) of the DNA grooves. In this way the groove binding part will selectively target a particular sequential pattern of DNA base pairs. These unique binding properties make such molecules potentially important for therapeutic applications, as exemplified by actinomycin D (**Figure 2.6**) (Chen, 1988; Robinson et al., 2001). Actinomycin is a potent antibiotic that binds to DNA and effectively inhibits RNA synthesis, primarily by interfering with the elongation phase of transcription (Sobell, 1985; Straney & Crothers, 1987). Another, less classical, representative of an intercalating agent is an anti-tumour drug ditercalinium (Gao et al., 1991; Crow et al., 2002). Ditercalinium (**Figure 2.6**) is a bisintercalator, which means that the molecule has two intercalating moieties connected by a flexible linker. Another type of

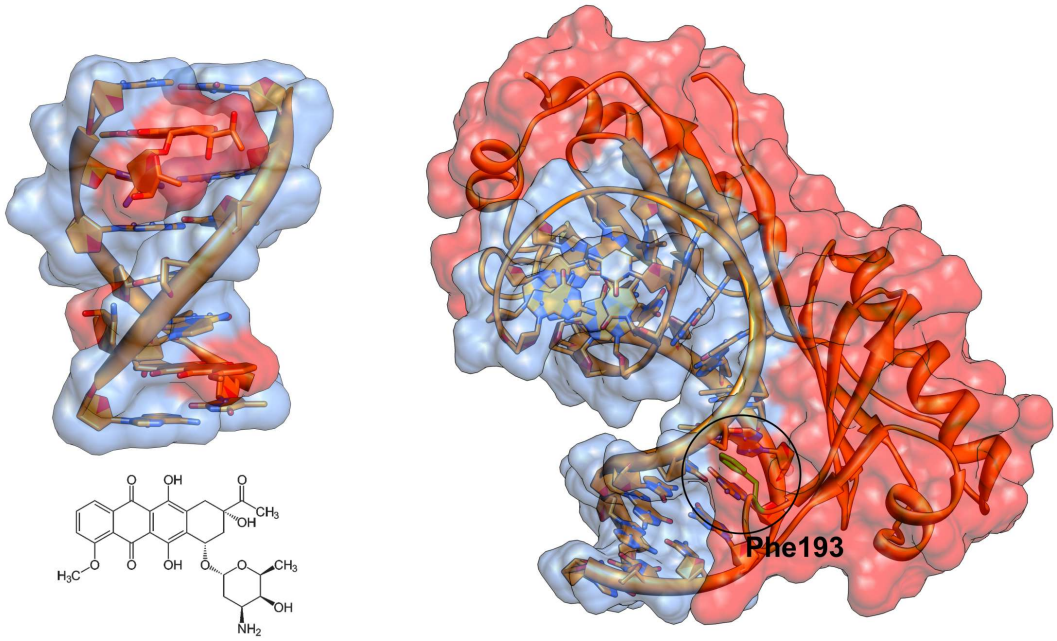
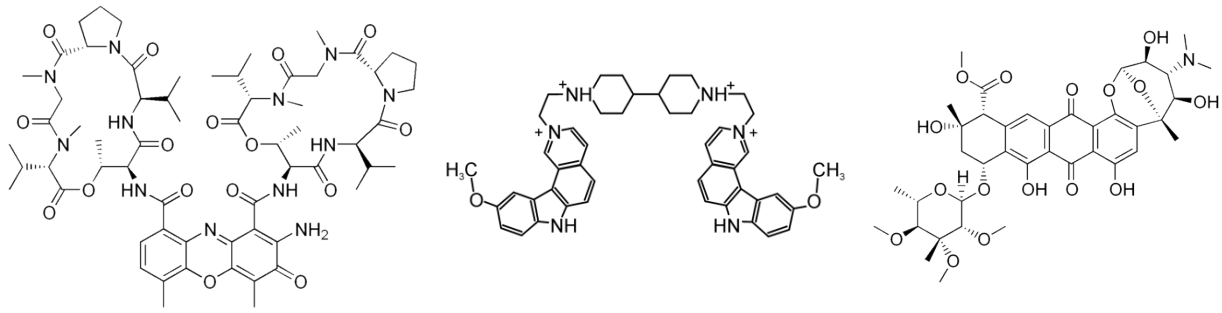


Figure 2.5. Intercalative binding to DNA: left – chemical structure of daunomycin (bottom) and cartoon representation of its complex with DNA (above); right – intercalative binding to DNA by TATA-box binding protein.

intercalation is threading intercalation for which the insertion between the adjacent DNA base pairs is sterically hindered by bulky substituents on both sides of the actual intercalating ring system. This, as it may seem at first sight, atypical form of intercalation is, however, biologically highly relevant. A natural antibiotic nogalamycin (**Figure 2.6**), produced by *Streptomyces nogalater*, is one example (Fox & Waring, 1984; Fox et al., 1985; Gao et al., 1990; Egli et al., 1991). To intercalate with its planar anthracycline moiety, either of nogalamycin's polar sugar groups has first to thread itself through the tightly packed, sterically rather inaccessible DNA base stack. To happen at all, threading intercalation would require some sufficiently large transient conformational reorganizations of the DNA double helix.

Despite extensive efforts, the details of DNA interactions with various ligands, small organic compounds as well as biological macromolecules, remain still poorly understood. As demonstrated in the previous paragraphs, the knowledge and ability of predicting how stable an intermolecular association is, and in what way it might influence the cell cycle is of strategic importance for the development of new efficient drugs. To gain mechanistic insights and to tackle the limits of DNA conformational flexibility from a new angle, a family of ruthenium(II)-coordinated compounds, which will be described in the following chapter, has been studied with respect to their interactions with DNA, their binding geometries and how they affect the DNA conformation.



a b c

Figure 2.6. Chemical structures of DNA intercalating agents found to exhibit some sequence-specific binding: a) actinomycin D, b) bisintercalator ditercalinium and c) threading intercalator nogalamycin.

2.3. Ruthenium(II)-Polypyridyl Compounds: Structure and Physico-Chemical Properties

As often happens in science, the ruthenium(II) coordinated polypyridyl compounds were originally developed with other intentions, namely to solve the energy crisis of the 1970s (Vos & Kelly, 2006). However, they were also successfully picked up as suitable DNA probes by Jacqueline Barton and co-workers almost thirty years ago (Barton et al., 1984) and are still used to study DNA-ligand interactions.

2.3.1. Mononuclear Ru(II)-polypyridyl Compounds

The pioneers among the ruthenium(II) coordination compounds (hereafter Ru(II)-compounds) to explore DNA double-helix were two three-bladed propeller-shaped polypyridil adducts of $[RuL_3]^{2+}$ type, where L denotes either bipyridine (bpy) or phenanthroline (phen) (**Figure 2.7**). These compounds gained a considerable popularity due to their remarkable chemical properties. Firstly, a covalent character of the bond between the divalent ruthenium ion and the ligands coordinated by it, combined with steric effects of the bulky ligands preventing new ligands to enter, makes the Ru(II)-compounds generally kinetically inert so that their association with DNA is reversible and does not involve binding to the ruthenium core. For the same reason the Ru(II) compounds are inversion-stable and exhibit an inherent chirality, thus we can differentiate between the right-handed Δ- and left-handed Λ-propellers (**Figure 2.7**). This property gives rise to handed discrimination (diastereomeric effect) upon association with a chiral hosting molecule, like DNA (the first time shown for DNA + $[FeL_3]^{2+}$ by Nordén and Tjerneld (Nordén & Tjerneld, 1976 b). And secondly, and most importantly, the Ru(II)-compounds as DNA-probes possess a rich repertoire of photo-physical properties, which are super-sensitive to variations in the micro-environment, allowing us to investigate DNA structure and function with the help of various spectroscopic techniques.

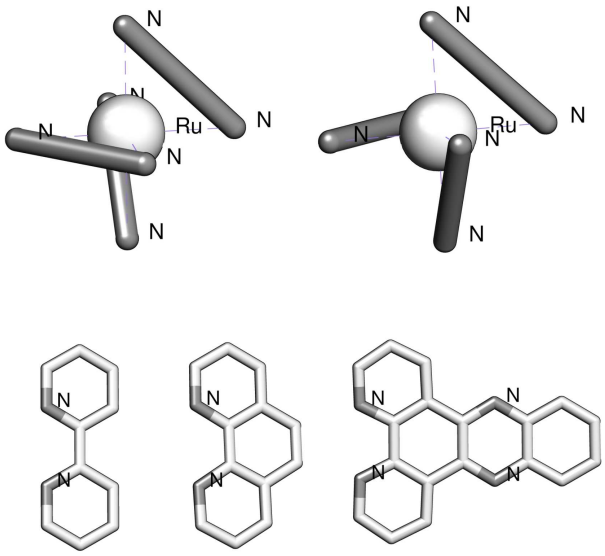


Figure 2.7. Top: Octahedral coordination of the ruthenium(II) centre gives rise to two stereo-isomeric forms: right-handed Δ- and left-handed Λ-propellers. Bottom: commonly used metal-coordination ligands: bpy = 2,2'-bipyridine, phen = 1,10-phenanthroline, dppz = dipyrido[3,2-a:2',3'-c] phenazine.

Photo-physical luminescence data for the $[RuBpy_3]^{2+}$ and $[RuPhen_3]^{2+}$ compounds does not reveal atom-level details about binding geometries of their complexes with DNA, which in 1990s led to a long debate regarding the particular binding mode (external groove binding or intercalation) and binding location (major groove or minor groove) (Kumar et al., 1985; Barton et al., 1986; Hiort et al., 1990; Satyanarayana et al., 1992, 1993; Eriksson et al., 1994). Further investigations, though, found indications that some of the compounds bind via intercalation and that the association takes place preferentially in the DNA minor groove (Lincoln & Nordén, 1998).

In 1985 Sauvage and co-workers synthesized a new ruthenium(II)-polypyridyl compound – $[RuL_2dppz]^{2+}$ (**Figure 2.7**) (Chambron et al., 1985) for which Barton et al. discovered that upon binding to DNA the compound exhibited a remarkable brilliant luminescence (Friedman et al., 1990; Jenkins et al., 1992). Their observation suggested intercalation: dppz, inserted between the base pairs of DNA, thus being shielded from quenching water. Hiort et al., using flow linear dichroism spectroscopy, showed that both right and left-handed enantiomers of $[RuL_2dppz]^{2+}$ bind to DNA and the elongated planar dppz moiety is oriented perpendicular to the DNA helix in consistency with intercalation (Hiort et al., 1993). And so, the prominent feature of $[RuL_2dppz]^{2+}$ to give rise to a striking increase in luminescence quantum yield once inserted into a hydrophobic, water-shielding, environment of the DNA stack, has lead to an extensive use of the dppz-based Ru(II) compounds as "Molecular-Light-Switch" probes in various biological contexts (Zhang & Lippard, 2003; Pierre et al., 2007; Antonarakis & Emadi, 2010;

Matson et al, 2011). Hence, $[RuL_2dppz]^{2+}$ proved to be less mysterious in regard to its binding-mode to DNA, however, it still posed the question about binding preferentiality: from the major or minor groove, a still on-going debate (Jenkins et al., 1992; Dupureur & Barton, 1994; Lincoln et al., 1996; Tuite et al., 1997; Holmlin et al., 1998). Recent high-resolution crystallographic data by Hall et al. on DNA complexes with $[RuTap_2dppz]^{2+}$ (almost identical to $[RuPheTap_2dppz]^{2+}$) showed minor groove binding (Hall et al., 2011).

Additional investigations reported that the dppz-containing compounds are high-affinity probes, with association constant $K > 10^6$ M^{-1} at 50 mM NaCl, but with moderate sequence preferentiality (K_{AT}/K_{GC}=5.9) (Holmlin et al., 1998). Being over 30 years in focus, the monomeric Ru(II)-coordinated compounds gave rise to a development of new generations of metal-coordinating ligands, where not only the peripheral ligands were systematically varied but also new metals have been tried, including iron(II), cobalt(II) and rhodium(II). The rhodium(II) compound, for example, relatively recently has shown to be of potential use for detecting sequence mismatches in B-DNA molecules (Pierre et al., 2007). The latter example confirms: the research is perpetually going on! In this Thesis, we have also addressed the long-lasting question of possibly quite different binding geometries of the enantiomeric forms Δ- and Λ-$[Ru(phen)_3]^{2+}$, the results presented in *Paper III*.

2.3.2. Binuclear Ru(II)-polypyridyl Compounds

In an attempt to tackle further horizons in probing DNA structure, but also to increase sequence selectivity of the Ru(II)-polypyridyl monomeric compounds, researchers took a swing into development of a series of binuclear ruthenium compounds and their photo-physical study. In the bisintercalating $[\mu\text{-}C4(cpdppz)_2Ru_2(phen)_4]^{4+}$ compound (**Figure 2.8**) the two dppz ligands are connected via a flexible linker (Önfelt et al., 1999; Önfelt et al., 2001). The compound has much higher affinity, in comparison to its mononuclear predecessors ($K \approx 10^9$ M^{-1} with calf thymus DNA at 100 mM NaCl). Upon association this bisintercalator has to thread its bulky ruthenium centres through the DNA stack and, respectively, unthread upon dissociation. This process, without any doubt, requires a complex multi-stage conformation-dependent pathway for both DNA and $[\mu\text{-}C4(cpdppz)_2Ru_2(phen)_4]^{4+}$. Thus, it is not surprising that the reaction kinetics is much slower than for a regular intercalator, for example the dissociation requires approximately 30 minutes at room temperature to be completed. In addition, the ΔΔ-enantiomer displays moderate selectivity for the alternating AT-DNA sequences, $poly(dAdT)_2$.

In order to facilitate the understanding of threading intercalation, some analogues with less flexible linkers have been synthesized: $[\mu\text{-}(bidppz)_2Ru_2L_4]^{4+}$, where L = bpy or phen, see **Figure 2.8**. In those the flexibility is limited to one single bond that connects the two monomeric subunits. Neither of these compounds has initially demonstrated any threading intercalation, though, virtually due to their 4+ charges, a strong binding was

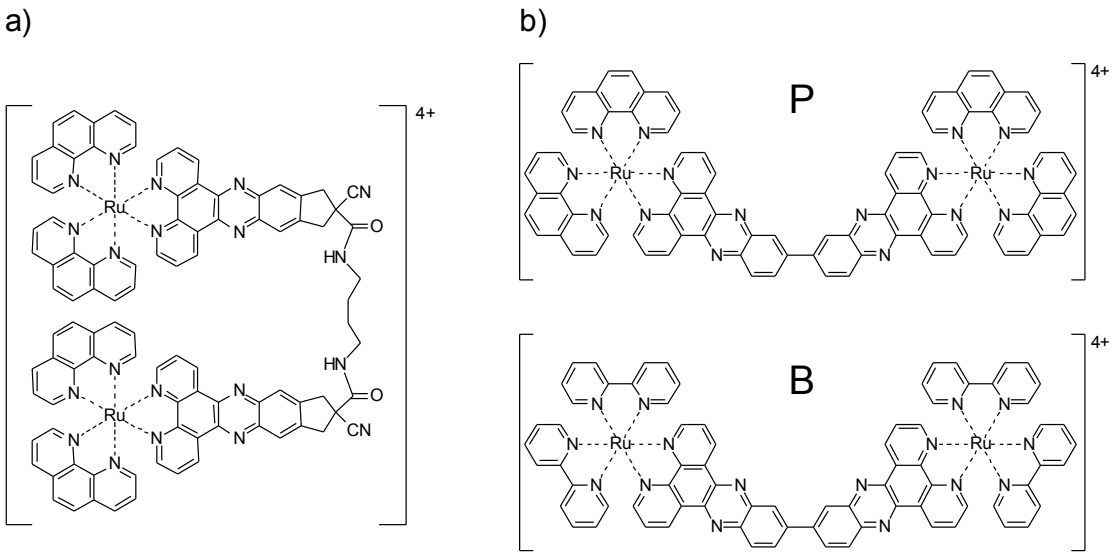


Figure 2.8. Chemical structure of binuclear ruthenium(II) compounds. a) bisintercalator [μ-C4(cpdppz)$_2$(phen)$_4$]$^{4+}$, b) thread-intercalators [μ-(bidppz)$_2$L$_4$]$^{4+}$, where L= 1,10-phenanthroline (top) or 2,2'-bipyridyl (bottom).

recorded (K ≈ 10^{12} M^{-1} at 10 mM NaCl) early concluded to be groove binding (Lincoln & Nordén, 1996).

However, later measurements on a sample of *Calf thymus* DNA and [μ-(bidppz)$_2$Ru$_2$(phen)$_4$]$^{4+}$, serendipitously left at room temperature for some two weeks' time, have revealed that threading intercalation is possible even in the case of such a conformationally-limited, sterically-unaffordable intercalating agent (Wilhelmsson et al., 2002). Further, by systematic variations of temperature and salt conditions, it has been found that the threading kinetics could be accelerated to be completed within only few hours at 50°C and in 100 mM NaCl with calf thymus DNA. Later both homo-chiral (ΛΛ) and meso (ΔΛ) configurations have been shown to intercalate DNA (Wilhelmsson et al., 2003). The follow-up studies with the help of linear dichroism spectroscopy detected no transient groove-bound state with alternating poly(dAdT)$_2$ DNA, while the threaded-state was detected to form already after a short equilibration time (Nordell et al., 2007). The length variation of alternating dAdT-tracts in AT-rich oligomers revealed the peculiar fact that at least one turn, 10 AT bases, is required for efficient threading. This AT-length threshold phenomenon was explained by a conformational transition of B-form DNA to A-form, in *Paper I* of this Thesis. In the first attempt to obtain a 3-D structure of the threaded-state of DNA in complex with [μ-(bidppz)$_2$Ru$_2$(bpy)$_4$]$^{4+}$, NMR studies were undertaken on oligonucleotides having a few AT base-pairs in the centre to which binding was expected to be directed. However, due to an apparently too short AT sequences, no thread intercalation was observed, but instead we managed to achieve data for a groove binding configuration of the ΛΛ-[μ-(bidppz)$_2$Ru$_2$(bpy)$_4$]$^{4+}$ compound, which forms the topic of *Paper II* of this Thesis.

STRUCTURE AND FUNCTION OF RECOMBINATION PROTEINS

2.4. Homologous Recombination Reaction

Homologous genetic recombination is an important biological process that involves the pairing and exchange of DNA between two homologous chromosomes or DNA molecules. The reaction is highly exploited in all living organisms, from bacteria to human, due to its vital importance for preservation of genomic integrity via error-free repair of DNA double-strand breaks but also for development of genetic diversity via creation of a unique offspring-DNA by shuffling maternal and paternal genomes (Kowalczykowski, 2000; Roca & Cox, 1997; San Filippo et al., 2008; Sung & Klein, 2006). Thus, being so essential for biological cells, the recombination reaction has attracted a lot of attention. And nowadays the level of knowledge already allows researchers to actively discuss and develop ways to utilize the recombination process for an artificial gene-repair targeting technology (Vispé et al., 1998; Ohnishi et al., 1998; Maacke et al., 2000; Liu et al., 2004; Nomme et al., 2008; Klug, 2010).

The recombination reaction is a very complex multi-step process. In a bacterium, *Escherichia coli,* for example, at least 25 different proteins are known to participate in entangled pathways of the reaction (Kowalczykowski et al., 1994), and even a larger number of proteins are involved in eukaryotes (Vispé & Defais, 1997; Baumann & West, 1998). The mechanism of recombination, which requires high degree of fidelity and precision and is achieved in the absence of any known DNA sequence selectivity of the involved proteins, has puzzled researches over the years. A number of models

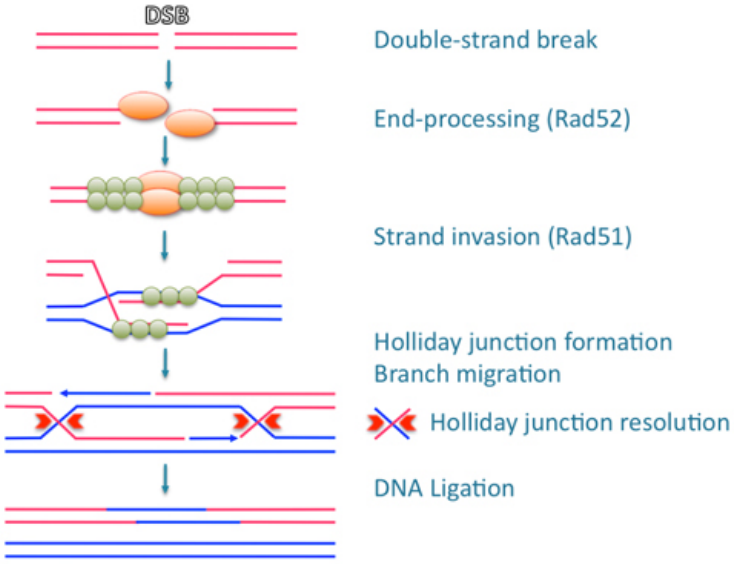


Figure 2.9. Schematic representation of the key-steps of homologous recombination reaction according to the double-strand break repair model, lustrated for eukaryotes (more exactly for yeast). The figure is adopted from http://www.biochem.mpg.de/en/rg/biertuempfel/research/index.html.

describing the reaction flow have been developed. **Figure 2.9** presents in a systematic way the key steps of the recombination reaction from the DNA double-strand breaks repair model (Kowalczykowski, 2000; Sung & Klein, 2006), which should be valid for all organisms. *Step 1*: initial appearance of a DNA double-strand break (or nick) leads to a formation of single-stranded DNA (hereafter ssDNA) overhangs. *Step 2*: a recombinase protein represented by e.g. prokaryotic RecA or eukaryotic Rad51, assisted by other supporting proteins (in the case of Rad51, while RecA rather easily binds to ssDNA), forms a presynaptic filament – a protein-ssDNA assembly. *Step 3*: in the presences of donor double-stranded DNA (hereafter dsDNA), a reaction intermediate – synaptic nucleoprotein filament – is formed. There, once the sequential homology is found, the repair synthesis and subsequent capture of the second end of the dsDNA break leads to a formation of a Holliday junction – a branched intermediate. *Step 4*: following the branch migration the completion of the DNA repair process is achieved by DNA synthesis, and finally ligation is made. *Step 5*: The recombined DNA molecules are released and the protein scaffold is disassembled.

2.5. Recombination Proteins RecA and Rad51: Structure and Mechanistic Insights

DNA strand exchange reaction and, preceding it, the search for homology are the two central steps of the homologous recombination reaction. Both steps are catalysed by the recombination proteins of the RecA family. The presence of RecA-like proteins, or their functional equivalents, in virtually all bacteria, archaea, and eukaryotes, confirms that the mechanism of homologous pairing and DNA strand exchange is evolutionarily conserved, essential and universal for all forms of life (Benson et al., 1994; Sehorn et al., 2004; Seitz et al., 1998; Shinohara et al., 1992).

2.5.1. RecA

Recombination protein A, RecA, from a bacterium *Escherichia coli* was first identified by Clark and Margulius in 1965 (Clark & Margulies, 1965). However, it was not until 1976 that the RecA gene was cloned for the first time by McEntee (McEntee, 1976). This made the protein available and allowed more systematic study of the strand exchange reaction, performed by RecA, *in vitro*. Since then, the RecA protein has become a model system for understanding the mechanism of homology search and strand exchange reaction (McEntee et al., 1979; Roca & Cox, 1997; Cox, 2007a). The monomer of RecA is a 352 amino acid long polypeptide with a molecular weight of 38 kDa. The first crystal structures of RecA in the presence and absence of a nucleotide co-factor ADP were resolved by Story and Steitz in 1992 (Story & Steitz, 1992; Story et al., 1992). Over the years, several other crystal structures of RecA have appeared (Datta et al., 2000; Datta et al., 2003a, 2003b; Xing & Bell, 2004a, 2004b; Krishna et al., 2006, 2007; Prabu et al., 2008). To promote search for sequential homology and catalyse DNA strand exchange the RecA protein, unlike many other proteins that participate in DNA metabolism, assembles around DNA in the presence of ATP (or ATP analogue) into a nucleoprotein

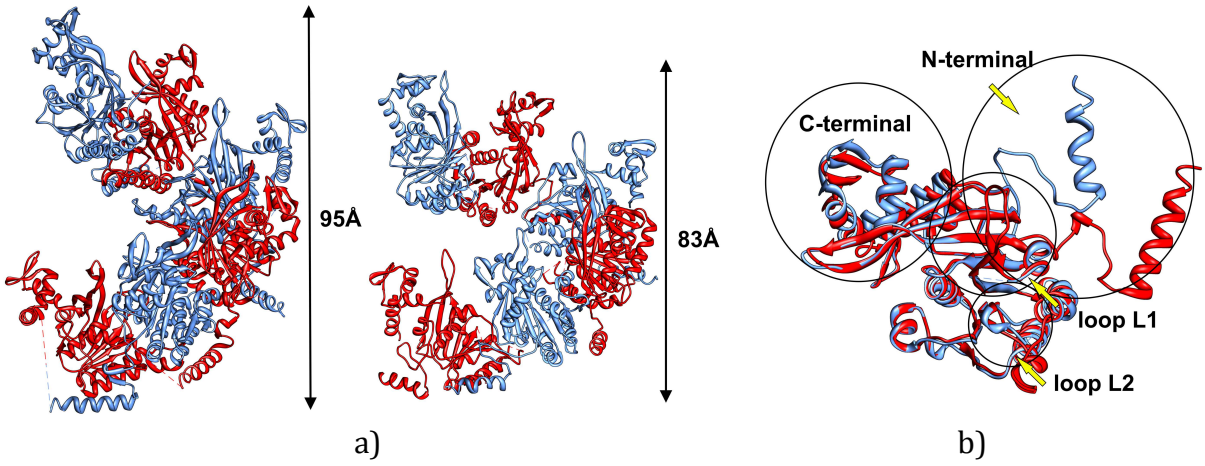


Figure 2.10. a) A cartoon representation of an "active" (left) and an "inactive" (right) RecA filament. The "inactive" form, (PDB ID: 3CMV) is formed in the absence of DNA: it is compressed and wider than the "active" filament (PDB ID: 3CMU), formed in the presence of ssDNA. b) A superposition of two RecA monomers of "active" (red) and "inactive" (blue) forms. DNA-binding loops, L1 and L2, as well as protein C-terminal and N-terminal domains also marked in the figure.

filament. The functional form of this assembly is a right-handed helix, where the DNA molecule is placed in the filamentous core. The filamentous helical pitch varies, but is usually reported to be from 90 to 100 Å in the case of an "active" filament, and from 65 to 85 Å in the case of an "inactive" filament (Egelman & Stasiak, 1993; Yu et al., 2001). The transition between the "inactive" and the "active" forms of filament is assured by a large motion of the N-terminal domain of the protein (**Figure 2.10**). This mechanistic detail and many other became available relatively recently with the appearance of high-resolution crystal structures of filamentous forms of RecA in complexes with ss- and dsDNA, as well as uncomplexed protein (Chen et al., 2008). These structures are of particular significance for the understanding of details of the strand exchange reaction. They represent the key intermediates of the reaction: free RecA, the presynaptic filament and finally the reaction product, as discussed in the previous chapter.

The DNA bound inside the filament is somehow forced to adopt an elongated conformation and is stretched by approximately 50% in comparison to canonical B-DNA. The stoichiometry of binding, which is one RecA monomer per three DNA bases, leads to an inhomogeneous stretching of DNA: a nucleotides triplet is flanked by a spatial gap, which is filled in by the intercalating residues of DNA-binding loops, L1 and L2 (**Figure 2.11** illustrates ssDNA binding, assured by the intercalative anchoring of Ile199 of L1). Interestingly, the loops are conformationally disordered when the DNA is absent, as confirmed by the crystal of the "inactive" RecA form. Whereas the primary DNA binding site, L1 and L2, is located in the interior of the filamentous construct, the second binding site, intended for the donor dsDNA, is proposed to be located at the C-terminal domain, which possibly acts as a gateway for the incoming DNA duplex (Aihara et al.,

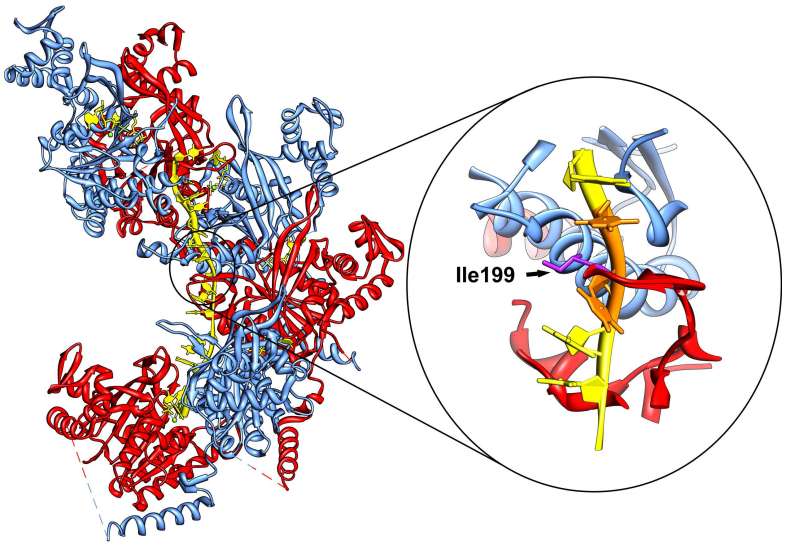


Figure 2.11. A tight binding of ssDNA (yellow) inside the "active" RecA filament is assured by the intercalation of Ile199 of L1 DNA-binding loop.

1997; Kurumizaka et al., 1996). The ability to interact with the incoming donor dsDNA molecule, the subsequent search for homology and, finally, the DNA strand exchange appear all to be highly dependent on the concentration of magnesium ion. High levels of Mg^{2+} have been reported to elevate the rates of DNA strand exchange, facilitated by a conformational transition in the C-terminal region (Lusetti et al., 2003).

2.5.2. Rad51

The eukaryotic homologue Rad51 is similar to RecA in size and also in structural organization. The human Rad51 (hereafter HsRad51), for example, has 338 amino acid residues and a molecular weight of 37 kDa. Rad51 analogously polymerizes on DNA in the presence of ATP (or ATP-analogue), forming a nucleoprotein filament as reported by electron microscopy (Ogawa et al., 1993; Yu et al., 2001). To date (January 2012) there are only a few examples of high-resolution structures of Rad51: two crystal structures of surprisingly high-pitched (130 Å versus normally observed 95 Å) *Saccharomyces cerevisiae* (Sc) filaments (Conway et al., 2004; Chen et al., 2010) an NMR structure of an N-terminal region (Aihara et al., 1999), and a crystal structure of an ATP-core domain of HsRad51 (Pellegrini et al., 2002). The latter, however, has a number of missing fragments, the most important of which are the putative DNA-binding loops, which are also absent in either of the crystals of ScRad51 filaments. The main structural difference between the promoters of Rad51 and RecA is the absence of C-terminal domain in Rad51 that has been obviously evolutionarily substituted by an extended N-terminal region (Benson et al., 1994; Ogawa et al., 1993; Yu et al., 2001). The ATP-binding domain is, on the other hand, surprisingly well conserved, both sequentially, with almost 50% identical residues, and structurally, with the RMSD value of C_α protein backbone trace of just 2.6 Å. Similarly to bacterial RecA, eukaryotic Rad51 has two putative DNA-binding

loops, L1 and L2. However, those have interchanged functionalities in comparison to the L1 and L2 loops of RecA, as reported by mutagenesis and tryptophan fluorescence analysis of the loop regions (Matsuo et al., 2006; Prasad et al., 2006). The N-terminal domain of Rad51 has been suggested, in analogy to C-terminal domain of RecA, to be the secondary binding site (Aihara et al., 1999). Unlike RecA, the strand exchange activity of HsRad51 is weaker in the presence of Mg^{2+}, though the activity rises when Mg^{2+} is substituted by Ca^{2+} (Bugreev & Mazin, 2004). Intuitively one could envision that Mg^{2+} may induce a fast ATP hydrolysis, which transforms the "active" filament into its "inactive" ADP-bound state, while Ca^{2+} acts as an inhibitor of ATPase activity.

To understand the details of structural organization of the HsRad51 filaments and gain some mechanistic insight into the DNA strand exchange reaction performed by this protein, a particular effort was made in the creation of a protein filament structural model (*Paper IV*).

This model of HsRad51 filament structure was further utilized for structural studies of the stimulatory effect of Ca^{2+} on the strand exchange activity of the protein: as a result an "ATP-hydrolysis free" mechanism was proposed (*Paper V*).

2.6. ATP hydrolysis by RecA

Both RecA and Rad51 exhibit DNA-dependent ATPase activity. Both recombinases require ATP (or ATP analogue) upon the assembly of active form filaments. ATP is acting as glue, as it is sandwiched in the interface between two protein monomers (Story & Steitz, 1992; Datta et al., 2003a, 2003b; Wu et al., 2004; Qian et al., 2005; Chen et al., 2008; Renodon-Cornière et al., 2008). However, the rate of ATP hydrolyses by RecA is much higher than by eukaryotic Rad51 (Bianco et al., 1998), indicating differences in mechanisms and potential roles of ATP hydrolysis between the two recombinases. Despite 40 years of studies the role of ATP hydrolysis in RecA is not yet fully understood, nevertheless several models have been proposed that couple ATP hydrolysis to DNA strand exchange. Due to the high level of homology to ATP synthase – a motor-like protein that controls levels of ATP inside cells – RecA has been suggested to possess an ATP-driven motor function, which could facilitate the rotation of two DNA substrates around each other when the search for homology is to be executed (Cox, 2003; Cox, 2007b). Other models have suggested that ATP hydrolysis has a recycling role: it initiates the disassembly of the whole construct once DNA strand exchange has occurred (Alberts & Miake-Lye, 1992; Kowalczykowski et al., 1994). Yet another model proposes that ATP hydrolysis is coupled to the redistribution of RecA monomers during homologous recombination reaction (Menetski & Kowalczykowski, 1989; Rehrauer & Kowalczykowski, 1993; Kowalczykowski & Krupp, 1995). While the role of ATP hydrolysis in RecA has reached some level of understanding, for Rad51, on the other hand, this question remains unclear. None of eukaryotic recombinases has ever exhibited a motor protein function, and findings to date suggest that the ATP hydrolysis in human Rad51

promotes the filament's disintegration (Bugreev & Mazin, 2004; Chi et al., 2006; Ristic et al., 2005).

The appearance of high-resolution crystal structures of the RecA-DNA filamentous complexes (Chen et al., 2008) with ADP+AlF_4 (an ATP transition state analogue) has brought new insights, but also new questions about the role and possible mechanism of ATP-hydrolysis. The hydrogen bond network, represented in the crystal structure, suggests an allosteric coupling of ATP to DNA via the protein environment. Interestingly, an analogous allosteric coupling has been suggested for human Rad51 too (Renodon-Corniere et al., 2008). Moving back to RecA, it has also been confirmed that the binding of ATP induces a different RecA-RecA interface, which leads to the conformational rearrangement of the filament into its active form. The crystal data further suggest that the initial binding of ssDNA in cooperation with ATP may initiate correct positioning of the donor dsDNA, but only when the second strand of a newly formed DNA duplex is correctly positioned in the active filament where ATP hydrolysis occurs. However, none of these crystal structures has revealed any solvent molecules close to ATP. The absence of water molecules in the ATP-binding site for either of ss- and dsDNA complexes was suggested to be crucial for stimulation of the strand exchange reaction (Chen et al., 2008), however it also posed a question about the possibility and actual mechanism of ATP hydrolysis, for which to occur the presence of solvent molecules around is crucial. Facilitated by the access of high-resolution structures of RecA-DNA complexes including presence of ATP analogues, with the help of quantum mechanics and molecular dynamics computations, the mechanism of ATP hydrolysis by RecA at atomic scale has been addressed in Chapter "On the mechanism of ATP hydrolysis in Bacterial Recombinase RecA" of this Thesis.

EXPERIMENTAL TECHNIQUES FOR MOLECULAR STRUCTURE DETERMINATION

The facts and conclusions about the details of molecular rearrangements and function would not be possible to deduce without the knowledge of molecular structure. The two major experimental techniques to gain high-resolution 3-D images of molecular systems, X-ray crystallography and nuclear magnetic resonance spectroscopy, have their own limitations, the most significant being related to the size and regularity of an investigated system. Alternative experimental techniques, that may overcome size limitations, however, can only provide low-resolution structural information e.g. electron microscopy, or very limited data like Förster resonance energy transfer spectroscopy (FRET) or linear and circular dichroism spectroscopy (LD and CD).

Although there are alternative experimental ways of obtaining structural information, only X-ray crystallography, NMR spectroscopy and linear dichroism spectroscopy will be introduced here due to their direct relations to the results presented in this Thesis.

2.7. X-ray Crystallography

The historical impact of X-ray diffraction for the development of molecular biology is, no doubt, enormous. It was an X-ray radiation study of DNA fibres that lead to *THE* discovery of the DNA double helix by Watson and Crick in 1953 (Watson & Crick, 1953). In 1958 John Kendrew and his co-workers resolved the first protein structure, myoglobin – a small protein responsible for the storage of oxygen in muscles (Kendrew et al., 1958). This structure of myoglobin was the first atomic resolution structure of a macromolecule. Only 20 years later, when it became relatively easy to synthesize DNA chemically in big enough quantities to grow crystals, the first atomic-resolution DNA structure was obtained. It was a mixed A-B-form DNA tetramer, (d-pApTpApT): deoxyribose was in a C3'-endo conformation when attached to a purine, and C2'-endo – when attached to a pyrimidine (Viswamitra et al., 1978). To date (January 2012) there are almost 70 000 crystal structures deposited in Protein Data Bank (Berman et al., 2000), they include DNA, RNA, protein, and their complexes.

To gain high-resolution structural information on 3-D arrangements of atoms in a molecule, its crystal is subjected to X-ray radiation. First, the diffraction pattern is collected

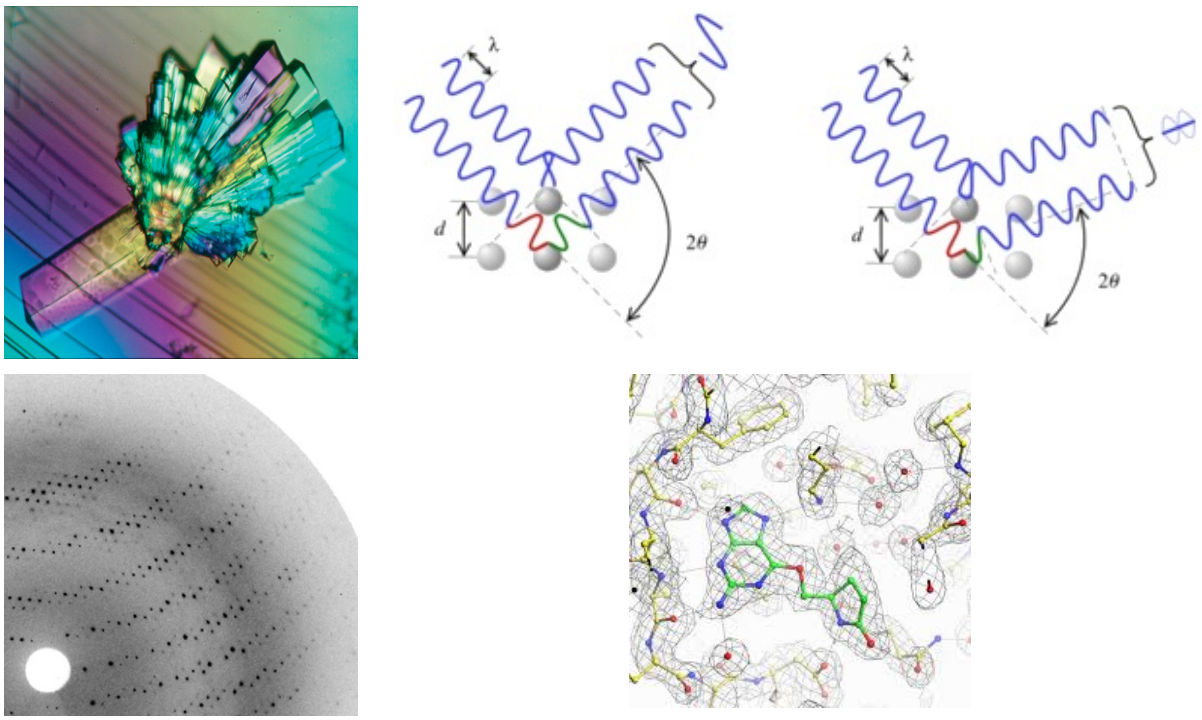


Figure 2.12. X-ray crystallography information flow: top left – an example of molecular crystals (source*:* Max Planck working groups, Hamburg); top right – the main principle of X-ray crystallography based on Bragg's diffraction law (source: http://en.wikipedia.org/wiki/Bragg_angle); bottom left – a typical diffraction pattern collected after X-ray beam scattered on the molecular lattice (source: http://www.flickr.com/photos/sc63/2852072432/); bottom right – a fragment of an electron density map with its atoms fitted into it (source: http://biop.ox.ac.uk/www/lab_journal_1998/endicott.html).

after the X-ray beam has passed through and been scattered by the lattice of atoms and molecules. Then, according to Bragg's law, the diffracting pattern can be resolved and inversely translated into an electron density map that will be later filled in with molecular fragments, and finally assembled into a complete 3-D structural model of the molecule (**Figure 2.12**).

As the molecule of interest becomes larger and more complex, the atomic level 3-D picture, provided by X-ray crystallography, loses its high precision and becomes less well resolved. Moreover, it is difficult to grow the crystals for larger molecules, as well as to achieve the regularity of the crystal packing. And most importantly, there could always remain a question open for a debate: are structures obtained via X-ray crystallography really biologically relevant? In a crystal, at first, a molecular system is subjected to extreme conditions compared to a live system: dense packing of the bio-macromolecules and high local concentrations of ions or other agents, added to stabilize the structure. These may clearly differ from the conditions in a native environment of a cell. Secondly, being immobilized in a crystal, the bio-system may represent a state close to equilibrium, but maybe not the one of significance to the live cell: instead, the structure might be only a representative of, so-called, dead matter...

2.8. Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance was first described and measured in molecular beams by Isidor Rabi in 1938 (Rabi et al., 1938). In 1946, Felix Bloch and Edward Purcell expanded the technique to liquids and solids. The first NMR solution structure of a small protein (BPTI) appeared in 1980, resolved by Richard Ernst and Kurt Wüthrich (Kumar et al., 1980). The latter fact means that NMR spectroscopy is 27 years younger than X-ray crystallography. To date in the Protein Data Bank there are only approximately 10 000 NMR structures to be compared to almost 70 000 crystal structures.

Nuclear magnetic resonance spectroscopy is based upon a phenomenon, which occurs when atomic nuclei, immersed in a static magnetic field, are exposed to a second oscillating magnetic field. These nuclei first absorb the electromagnetic radiation and then re-emit it. The energy of this emission has a specific resonance frequency, which depends on the strength of the magnetic field and magnetic properties of nuclei. Not all atoms can exhibit NMR but only those that have an odd number of nucleons (e.g. ^{1}H, ^{13}C, ^{14}N, ^{31}P, etc.) in their nuclei have an intrinsic magnetic moment and angular momentum, in other words a nonzero nuclear spin, which can be detected. The intensity of a nuclear spin signal and the shift of the resonance frequency due to variation of local magnetic field, also known as a chemical shift, depend on the shielding effect of neighbouring spins and also on a surrounding environment. The spin polarization from one spin can be transferred to another spin either via chemical bonds or via space, which gave rise to the development of different 2-D NMR techniques: COSY (Aue et al., 1976) and NOESY (Jeener et al., 1979) respectively, see (Cavanagh et al., 2007) for an overview. The COSY technique, where spin-spin coupling is observed if two nuclei are separated by 2-3

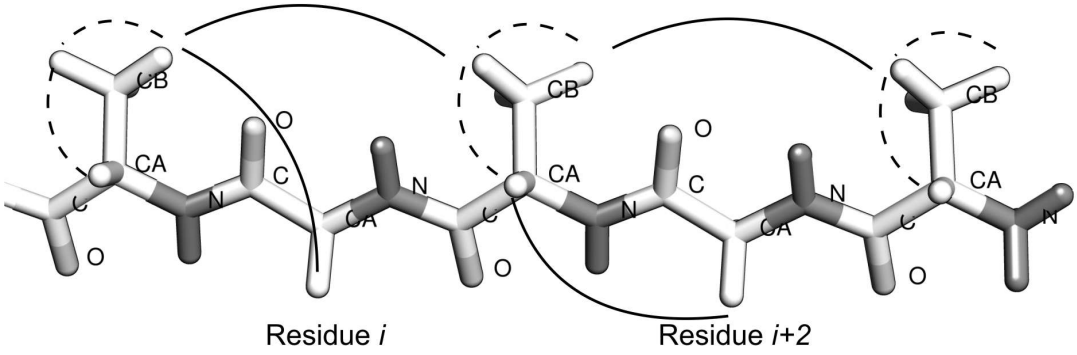


Figure 2.13. Example of spin polarization transfer in case of COSY (dashed line) and NOESY (solid line) NMR spectroscopy, not all possible transfers are shown to avoid image overcrowding. Examples of COSY and NOESY spectra could be found in *Paper II* of this Thesis.

chemical bonds, is used to determine local intermolecular atomic relations. NOESY, where, on the other hand, spin-spin coupling occurs via space if two spins are closely located, is used to define a molecular conformation or intramolecular contacts (the principles of COSY and NOESY are illustrated in **Figure 2.13**). Both NOESY and COSY are used simultaneously to resolve a 3-D molecular structure. TOCSY (Braunschweiler & Ernst, 1983; Weber et al., 1987) is another 2-D NMR technique, where spin-spin coupling is observed not only if two nuclei are separated by 2-3 chemical bonds, but also even further away if there is a non-zero spin nucleus.

The main advantage of NMR spectroscopy to X-ray crystallography is that experimental conditions, namely aqueous solutions of salts, resemble to a greater extent an intracellular environment. Another, not less important, advantage of NMR is that it can provide information about dynamics of a molecular system. The major disadvantage, on the other hand, is the size of the investigated systems possible to study, the limit lies somewhere around 70 kD. In addition, the structural information deduced from the NMR spectra provides structural information on the approximate distances between adjacent atoms or groups of atoms, and thus the model structure, collected from these bits into a 3-D model, requires further refinement.

2.9. Linear Dichroism Spectroscopy

Linear Dichroism is a long known phenomenon, dichroism means ”two colours”, which are seen when looking through dichroic crystals and rotating the plane of polarisation by 90°. The ultimate LD device is, thus, a polarizer for which the Malus law: $I(\alpha) = I(0)cos^2(\alpha)$, predicts that the transmitted light intensity varies as a cos^2 function of an angle, α, of a plane of light polarization relative to the maximum absorption direction.

The LD as a technique was pioneered by Eggers, Thulstrup and Michl in 1970, for determining directions of transition moments of planar aromatic molecules by forming linear combinations of the two spectra $A_{\parallel}$and $A_{\perp}$ (Michl et al., 1970). A device, which nowadays is the most commonly associated, at least for me, with LD – a ”Couette cell” – was pre-

sented by Wada and Kozawa in 1964 (Wada & Kozawa, 1964). With the help of their Couette cell, made of two transparent concentric quartz cylinders, between which, when spinning either of the cylinders at a speed of typically 1000 rotation per minute (rpm), a DNA orientation could be studied in the ultraviolet region. Unfortunately, the sensitivity of this approach was rather poor, which made more quantitative studies in dilute solutions difficult. A breakthrough occurred in 1972 when ”high-sensitivity LD” was introduced using phase modulation, the same technique that allowed circular dichroism to be measured with sufficient sensitivity. It was also shown rigorously that a commercial CD instrument could be converted to accurately measure LD and how non-linear effects could be easily taken care of (Davidsson & Nordén, 1976).

With this technique very small molecules were shown possible to study for the first time, in a polymer matrix (Nordén & Davidsson, 1972; Nordén, 1973). Also, using Wada’s Couette cell, DNA binding by intercalators (Nordén & Tjerneld, 1976a) as well as groove binders (Nordén & Tjerneld, 1977) could now be studied with unprecedented sensitivity. The DNA binding geometries of these small molecules could be analysed by using the orientation of DNA bases as internal standard for determining the flow orientation factor S. In this way, by testing various DNA binding compounds, it was discovered that substitution-inert metal complexes, e.g. Ru(II)-polypyridyl compounds, could bind to DNA (Nordén & Tjerneld, 1976b). Since then, various system specific orientation techniques have appeared, such as lamellar membrane systems (Nordén et al., 1977) and flow-deformed liposome membrane systems (Ardhammar et al., 1998).

Linear dichroism (LD) is defined as the differential absorption of linearly polarized light with polarization parallel and perpendicular to some macroscopic orientation axis of a system ($LD = A_{\parallel} - A_{\perp}$). Thereby, LD can be measured on a system that is either spontaneously oriented (for example a membrane protein in a crystal or in a lipid bilayer at a surface), or the orientation can be induced, for example, by applying an electric or magnetic field, or a mechanical force (stretching a polymer or introducing a shear flow gradient, etc.). The LD signal depends on the local as well as macroscopic orientation of molecular chromophores within a system. In a case of a protein-DNA system, these chromophores are the aromatic amino acid residues: tyrosine, tryptophan and phenylalanine, and all four DNA bases. Each of the chromophores exhibits a number of electronic transitions with characteristic so-called electric dipole transition moments (preferential oscillation directions of electrons upon resonance interaction with photon) that absorb the electromagnetic radiation at different wavelengths and thus can be differentiated.

If the average polarization of the transition dipole moments of all chromophores in a molecule, absorbing at a certain wavelength, is more parallel than perpendicular to the macroscopic reference axis, a positive LD will be recorded, and vice versa, if it is perpendicular, the LD signal will be negative. Hence, the LD technique can be used to estimate the orientation of a chromophore relative to a macroscopic orientation (**Figure 2.14**). A more quantitative characterization of a chromophore orientation can

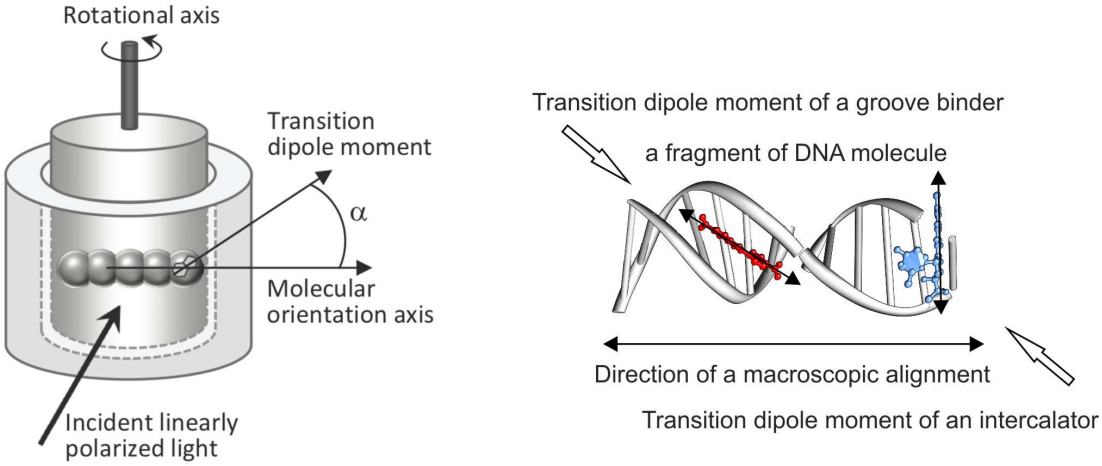


Figure 2.14. An illustration of LD principles. Left: a Couette cell with inner and outer rotating cylinders that can be used to produce macroscopic orientation of a liquid sample of elongated particles by creating a shear flow gradient (source: Frykholm, 2010). Right: a schematic representation of possible ways of DNA-ligand interaction. From the sign and strength of the LD signal it can be immediately detected whether the ligand is intercalated (signal is negative) or groove-bound (signal is positive).

be gained when the LD is normalized by dividing it by the isotropic absorption of the sample. This quantity is referred to as the Reduced Linear Dichroism (LD^r) and under conditions of a local uniaxial orientation, for an absorption band consisting of a single electronic transition, *i*, the LD^r can be written in the following way: $LD^r = LD/A_{iso} = 3 \cdot S \cdot (3\cos^2 \alpha_i - 1)/2$. Here, α_i is an angle between a particular transition moment of a chromophore and the molecular orientation axis, e.g. DNA helix axis. S is the orientation factor, which is equal to unity for a perfectly oriented system and zero for an isotropic sample. In case when the sample is also macroscopically uniaxial, the orientation factor can be determined by: $S = (3 < cos^2\Theta > -1)/2$, where Θ is the angle between the molecular orientation axis (e.g. DNA helix axis) and the macroscopic orientation axis (e.g. stretch direction of a polymer film) ("<...>" represents ensemble average over orientation distribution).

One of the limitations of this technique is a difficulty in determination of the orientation factor S. This hurdle can be overcame by using an internal orientation standard in the system that gives an opportunity to calibrate S by the LD^r measured for a chromophore with known orientation, e.g. the perpendicularly oriented DNA base pairs in a protein-DNA system. Another limitation is that the technique only observes an ensemble-averaged angular orientation of chromophores' transition moments and, in addition, is complicated by the effects of overlapping absorption bands, for example, in a system with several chromophores absorbing at the same wavelength. This limitation can also be resolved by a systematic mutation of a chromophore of interest to another chromophore with non-overlapping absorption bands. This approach, named Site-Specific

Linear Dichroism (SSLD), was successfully applied to the structural studies of the human recombination protein, Rad51 (see *Paper IV* and Results section of this Thesis).

3. METHODS

In this chapter a brief introduction to the computational chemistry techniques that were used in the papers presented in this Thesis will be given.

Computational chemistry is a branch of chemistry where the chemical experiments are conducted in a computer, or in other words *in silico*. The computational chemistry field can be divided into several sub-branches, depending on the way atoms and molecules are treated. In quantum mechanics (QM) or, when applied to molecules, "quantum chemistry", molecular orbitals and electrons occupying them are described for a molecular system, giving an opportunity to calculate chemical and physical properties which directly depend upon the electron distribution. Quantum chemical methods can be subsequently clustered into groups such as *ab initio*, or "from first principles", *semi-empirical* and *density functional theory*. Each of these methods has certain advantages, application areas, as well as limitations among which the size of molecular system or, in other words, the number of atoms is the most prominent. To overcome the size limitation, semi-empirical and later empirical methods have been developed. In comparison to QM, where even today the feasible size of the investigated molecular system does not exceed a few hundred atoms, the empirical methods, such as *Molecular Mechanics* (MM) or *Molecular Dynamics* (MD), can easily handle several hundred thousand atoms including salts and explicit solvent. The presence of the latter is highly important for the conduction of trustworthy *in silico* experiments and is especially relevant for bio-macromolecular systems, which exist and function in aqueous solutions. This giant leap of systems' size was possible due to relative simplicity of the MM and MD potential energy functional form, which led to that the accuracy of the empirical methods is highly dependent on the set of empirical parameters describing atoms in the system. Besides the mentioned QM, MM and MD methods a number of bioinformatics tools have also appeared over the last years, relevant in the context of this Thesis. By combining computer science expertise with knowledge of molecular biology and methodology of computational chemistry it is now possible, for example, based upon amino acids sequence, to predict with a reasonable accuracy a protein's secondary structure or even to create a 3-D atom-resolution model of a protein. On the subsequent pages some of these computational chemistry and bioinformatics methods that were used to achieve results presented in this Thesis will be introduced in greater detail.

QUANTUM MECHANICAL CALCULATIONS

Despite that the beginning of quantum mechanics could be associated with the discovery of cathode rays by Michael Faraday in 1839 and with the statement of the black body radiation problem by Gustav Kirchhoff in 1859, or even with the suggestion made by

Ludwig Boltzmann in 1877 that the energy states of a physical system could be discrete, I personally think that quantum mechanics started with the Schrödinger equation, formulated by Erwin Schrödinger in 1925. The equation tells how the quantum state of a system described by a *wave function* may evolve with time. The wave function, $\Psi(\mathbf{r}, t)$, is an abstract entity that has no physical analogy but has two properties. Firstly, it is an amplitude function whose phase is important upon interactions with other wave functions of matter or radiation, secondly, it is a probability function and its square $\Psi^*\Psi$ may be interpreted as the probability density for finding the system at time *t* and point ***r*** of the configuration space. The wave function could be also thought of as a statistical "weighting operator" with the help of which, very conveniently, information about what is the expected value of any dynamic quantity, corresponding to an operator $\hat{A}$, can be calculated as:

$$\overline{A} = \frac{\int_{-\infty}^{\infty} \Psi^* \hat{A}\, \Psi d\boldsymbol{r}}{\int_{-\infty}^{\infty} \Psi^* \Psi d\boldsymbol{r}} = \frac{\langle \Psi^* \left| \hat{A} \right| \Psi \rangle}{\langle \Psi^* | \Psi \rangle} \qquad (1)$$

where $\langle | \rangle$ is the "bra-ket" notation, introduced by Dirac, and Ψ^* is the conjugate wave function. When $\hat{A}$ is the operator of energy (i.e. the Hamiltonian), the expression provides the total energy of the system that the wave function represents.

While the mathematical basis of quantum mechanics had been developed by Schrödinger in 1925, the first quantum chemical calculation on the hydrogen molecule was performed in 1927 by Walter Heitler and Fritz London, in their method (known as Valence Bond method) the interatomic interactions were treated in a pairwise fashion. An alternative approach was developed by Friedrich Hund and Robert Mulliken in 1929. They suggested the electrons to be described by mathematical functions delocalized over the entire molecule. The Hund-Mulliken method is generally known as molecular orbital (MO) method and formed the basis of the Hartree-Fock and post-Hartree-Fock methods (which will be introduced later). Yet another really alternative approach developed by Llewellyn Thomas and Enrico Fermi in 1927. The Thomas-Fermi model was the first attempt to use the electron density instead of the wave function to describe a multi-electron system. Though their approach was not very successful for the description of the entire molecule, it laid the ground for the development of the Density Functional Theory.

Since quantum mechanical calculations have just been a secondary tool for the results presented in this Thesis, and since the background of the computational quantum mechanics is exhaustively dealt with in numerous textbooks (Hehre et al., 1986; Pilar, 1990; Griffiths, 1995; Leach, 2001; Cramer, 2004), only the principal concepts and main approximations will be briefly introduced in the following sections, omitting heavy mathematical derivations.

3.1. The Schrödinger Equation

Quantum Mechanics methods originate from the Schrödinger Equation. As deceptively simple as it might look, this equation is so complicated that it can be analytically solved for only very few simple systems, e.g. the hydrogen atom. However, using various approximations and iterative schemes, the equation can be numerically solved even when applied to much more complex systems. The full, time-dependent form of this equation is:

$$\widehat{H}\Psi(\mathbf{r},\mathrm{t}) = E\Psi(\mathbf{r},\mathrm{t})$$

$$\left\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+\mathcal{V}(\boldsymbol{r},t)\right\}\Psi(\boldsymbol{r},t) = i\hbar\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} \qquad (2)$$

This equation refers to a single particle of mass m, which is moving through space, its position given by the vector $\mathbf{r}$(x, y, z) and time (t), under the influence of an external field $\mathcal{V}$, which could, for example, be the electrostatic potential due to the nuclei of a molecule. E is the energy of the particle, Ψ is the wave function that characterizes the particle, $\widehat{H}$ is the *Hamiltonian* operator. The Schrödinger equation falls into the category of partial differential eigenvalue equations in which an operator acts on a function (eigenfunction) and returns the function multiplied by a scalar (eigenvalue).

3.2. Hierarchy of Approximations Made to Approach the Schrödinger Equation

As mentioned, the Schrödinger equation can be analytically solved for the exclusive case of the hydrogen atom. However, even for this trivial system some approximations have to be made. For more complex systems like multi-electron atoms or molecules further approximations are required in order to solve the equation. To provide a better overview on how the ideal theory has evolved to the state of current modern practical QM methods the major approximations are mentioned here very briefly.

For the majority of nowadays calculations the time-independent Schrödinger equation is employed, which describes the stationary states of the system. The time-independent or stationary Schrödinger equation does not describe the relativistic effects, thus, the latter are commonly neglected. While the non-relativistic treatment is appropriate for the lighter elements of the periodic table, up to chlorine, for heavier atoms, where relativistic effects become more and more significant, a number of approximate approaches have been developed (Moss, 1973; Pyykkö, 1988).

The second approximation, Born-Oppenheimer approximation, is no doubt the most important for the development of computational quantum chemistry (Born & Oppenheimer, 1927). Based on the fact that nuclei are approximately 2000 times heavier than electrons, Max Born proposed to separate the total wave function into the electronic wave function, $\Psi_e(\boldsymbol{r},\boldsymbol{R})$, and the one of the nuclei, $\Psi_N(\boldsymbol{R})$, so that the motion of latter can be considered as stationary with respect to that of the electrons.

$$\Psi(\boldsymbol{r},\boldsymbol{R}) \approx \Psi_e(\boldsymbol{r},\boldsymbol{R})\Psi_N(\boldsymbol{R}) \tag{3}$$

where **r** denotes a position vector of an electron and **R** – of a nucleus. This approximation is satisfactory to most systems in the ground state. Thereby, the time independent Schrödinger equation looks somewhat simpler than the one introduced above, it can be decoupled into electronic (4):

$$\begin{gathered}\widehat{H}_e\Psi_e(\boldsymbol{r},\boldsymbol{R}) = E(\boldsymbol{R})\Psi_e(\boldsymbol{r},\boldsymbol{R}) \\ \left\{-\frac{\hbar^2}{2m}{\nabla_r}^2 + \mathcal{V}(\boldsymbol{R},\boldsymbol{r})\right\}\Psi_e(\boldsymbol{r},\boldsymbol{R}) = E(\boldsymbol{R})\Psi_e(\boldsymbol{r},\boldsymbol{R})\end{gathered} \tag{4}$$

and nuclear equation (5):

$$\begin{gathered}\widehat{H}_N\Psi_N(\boldsymbol{R}) = E(\boldsymbol{R})\Psi_N(\boldsymbol{R}) \\ \left\{-\frac{\hbar^2}{2m}{\nabla_R}^2 + \mathcal{V}(\boldsymbol{R})\right\}\Psi_N(\boldsymbol{R}) = E(\boldsymbol{R})\Psi_N(\boldsymbol{R})\end{gathered} \tag{5}$$

In virtue of the Born-Oppenheimer (B.-O.) approximation, it is only necessary to solve the electronic Schrödinger equation and then add a nuclear repulsion term to get the total internal energy. Though the B.-O. approximation has drastically simplified the wave function representation, for a system with multiple electrons to solve the Schrödinger equation further approximations have to be considered which circumvent the analytic solution of the electron-electron repulsion. Douglas Hartree developed a method in which a plausible approximate atom wave function was written as the product of one-electron wave functions, known as the Hartree product:

$$\Psi = \psi(1)\psi(2)\psi(3)\ldots\psi(n) \tag{6}$$

which, however, did not account for the Pauli principle, namely, if the coordinates of two indistinguishable electrons are exchanged, the wave function should change its sign.

The problem was corrected by Slater and Fock (Slater, 1930; Fock, 1930). According to their method a wave function is composed of orthogonal spin orbitals. A spin orbital, ϕ_i, is the product of a spatial orbital, ψ_i and a spin function, α or β. To ensure antisymmetry, Slater functions, which are written in the form of determinants, provide the desired mathematical formalism:

$$\Phi(1\ldots n) \stackrel{\text{def}}{=} \frac{1}{\sqrt{n!}}\begin{vmatrix}\phi_1(1) & \cdots & \phi_1(n) \\ \vdots & \ddots & \vdots \\ \phi_n(n) & \cdots & \phi_n(n)\end{vmatrix} \tag{7}$$

where $1/\sqrt{n!}$ is the normalization factor. Further on, as mentioned earlier, in order to obtain the Hartree product the electron repulsion term has to be neglected. All these approximations lead to a complete exclusion of the electron repulsion term from the Schrödinger equation. To get meaningful results, however, the electron repulsion has to be somehow considered. As a consequence an effective potential term was introduced,

which assumes that an electron is affected by the average potential of the other electrons in the system.

$$V^{eff}(i) = \sum_{j=1}^{n_e} \langle \phi_j(j) | \frac{1}{r_{ij}} \left(1 - \hat{P}_{ij}\right) | \phi_j(j) \rangle \qquad (8)$$

where $V^{eff}(i)$ depends on every electron except i^{th}, $\hat{P}_{ij}$ is the exchange operator, which exchanges the electrons of two one-electron orbitals. With all the approximations and necessary adjustments in hand it will be most logical to proceed now to the introduction of the Hartree-Fock method (HF), the simplest and the fastest of *ab initio* methods. To avoid heavy mathematic derivation, the reader is referred elsewhere (Hehre et al., 1986; Pilar, 1990; Griffiths, 1995; Leach, 2001; Cramer, 2004) for details. However, it is still worth mentioning here that the HF method is searching for a solution of the Schrödinger equation as a variational problem by minimizing the internal energy with respect to ϕ_i, with the constraint of the orthogonality of the spin orbitals. This variational problem could be solved with the help of the Lagrange method, followed by an iteration procedure where the initial guess of one-electron functions, $\phi_i^0(j)$, and the subsequent solution of the HF equations lead to an update of the initial $\phi_i^0(j)$ to $\phi_i^1(j)$, and the cycle starts again. This process will be repeated until some convergence criterion is satisfied; this is called the self-consistent-field (SCF) procedure.

The Hartree-Fock-Roothaan method – an alternative to the Hartree-Fock method is applied for the molecular orbitals (MO). The molecular wave function in the Hartree-Fock-Roothaan (HFR) method is represented as a linear combination of the atomic orbitals (LCAO).

$$\psi_i = \sum_{\mu=1}^{m} c_{i\mu} \chi_\mu \qquad (9)$$

where χ_μ is a *basis function* (corresponding to an atomic orbital), chosen from an appropriate basis set, *m* is a finite number referring to the number of the basis functions, $c_{i\mu}$ are the linear coefficients of the i^{th} orbital. The Roothaan-Hall equations and the way of solving them are analogous to those of Hartree-Fock but correspond to the molecular orbitals, where the coefficients $c_{i\mu}$ are subjected to variation instead of the spinorbitals. The Hartree-Fock and Hartree-Fock-Roothaan methods give a fairly good description of the electronic structures. However, all the approximations made on the way may lead to significant deviations from the exact solution. As a consequence, numerous methods have been developed, known as the post-HF methods, e.g. the perturbation treatment of atoms and molecules proposed in 1934 by Møller and Plesset (MP2, MP4, etc) (Møller & Plesset, 1934), the coupled-cluster (CC) methods (Cizek, 1966; Szalay & Bartlett, 1995; Hrusak et al., 1997) and configuration interaction methods (CI) (Pople et al., 1987; Siegbahn, 1992).

3.3. Density Functional Theory

When the original Thomas-Fermi model resulted in fiasco, the density functional theory (DFT) was not used in quantum chemistry until 1990s. However todays DFT has been a popular approach since 1980s for calculations in the field of solid-state physics. The popularity came along with an original publication of the Kohn-Sham method (Kohn, 1965). The modern DFT calculations are based on the two Hohenberg-Kohn theorems (Hohenberg & Kohn, 1964), which say that all the properties of a molecule in a ground state could be defined by the ground state electron density function $\rho_0(\boldsymbol{r})$, the energy of the ground state can, thus, be considered as a function of the electron density $E_0(\rho_0(\boldsymbol{r}))$; the second theorem claims that any trial energy is either higher or equal to true ground state energy. If the energy E_V obtained from calculations with a trial electron density of electrons moving under influence of some external potential *V(R)*, then the second DFT theorem can be written as:

$$E_V(\rho_{trial}(\boldsymbol{r})) \geq E_0(\rho_0(\boldsymbol{r})) \qquad (10)$$

Subsequently, analogously to the previously described variational problem of the Hartree-Fock and the Hartree-Fock-Roothaan *ab initio* methods, DFT calculations are also solved with a variational approach. The Kohn-Sham (KS) equations are the analogues of the HF equations. The KS method is the most common implementation of density functional theory, where the many-body problem of interacting electrons, moving in a static external potential, is reduced to a problem of non-interacting electrons moving in an effective potential. The latter is the sum of the potential induced by the Coulomb interactions between the electrons plus the external potential. The most popular DFT method, which was also used in a supportive manner for the calculations presented in the Thesis, is the Becke3-Lee-Yang-Parr (B3LYP) hybrid functional (Lee et al., 1988; Becke, 1993; Stephens et al., 1994). Although most DFT methods (e.g. the latter hybrid method) are not real *ab initio* approaches (they contain some empirical parameters), their great advantage is that, while being a bit more computationally expensive than HF calculations, DFT methods provide quality closer to post-HF methods and thus allow relatively accurate molecular calculations of significantly bigger systems than those to be considered by the post-HF methods.

3.4. Basis Sets

The accuracy of the quantum mechanical calculations is highly dependent on the chosen basis set. The basis sets most commonly used are composed of atomic functions. The best choice would obviously be Slater type orbitals (STO) obtained from the solution of the Schrödinger equation for one-electron atoms. STOs are of $f(\boldsymbol{r})e^{-\xi \boldsymbol{r}}$ functional form where $f(\boldsymbol{r})$ is a polynomial function of $\boldsymbol{r}$. Unfortunately, Slater type functions are not particularly amenable to implementation in calculations of either multi-electron atoms or molecular orbitals. Thus it is common in QM calculations to use as a basis set the functions based upon Gaussians, which have the functional form of $e^{-\alpha \boldsymbol{r}^2}$, and are

substantially easier in use. The idea of using Gaussian functions in quantum mechanical calculations is often ascribed to Boys (Boys, 1950). The Gaussian type orbitals (GTO) have the functional form of:

$$g(l,m,n;\boldsymbol{r}(x,y,z);\zeta) = Nx_l y_m z_n e^{-\zeta \boldsymbol{r}^2} \tag{11}$$

where N is a normalization constant, ζ is called exponent, x, y, z are Cartesian coordinates, and l, m, n are integral exponents. The exponential term is responsible for the radial dependence, where $\boldsymbol{r}$ is the distance from the centre of the function, which is usually an atom's position in a molecule. The $x_l y_m z_n$ term characterizes the angular dependence in the function. Practically, the $l + m + n$ expression can be correlated to s, p, d, f ... type atomic functions. Namely, for a Cartesian s-function $l + m + n = 0$, for a p-function $l + m + n = 1$, and so on. However, it should be emphasized that l, m, n do not represent the quantum numbers but simply integral components. Modern computer codes employ Gaussian functions. Compared to STOs, GTOs do not have a cusp at the position of the nuclei and they have much sharper fall-off with increasing r, thus, they are of lesser quality functions. To improve the quality, the shape of the STO function is approximated by summing up a number of GTOs with different exponents and coefficients. STO-type basis functions are usually contracted by linear combination of primitive Gaussians, g_i:

$$\chi_{STO} = \sum_{i=1}^{L} c_i g_i \tag{12}$$

Due to extensive research and subsequent developments (Davidson & Feller, 1986; Clementi et al., 1990; Almlöf & Taylor, 1991), a diverse selection of atom-centred standard basis sets is available nowadays. A minimum basis set includes one contracted function for each of the occupied core and valence orbitals. For example, for an oxygen atom, the minimum basis set is composed of two s-functions (1s and 2s), and three p-functions ($2p_x$, $2p_y$, and $2p_z$). Another type of basis set is the split-valence basis set, which is based on the argument that core electrons are less influenced by chemical bonding and, thus can be treated with lower accuracy. In practice this is done by having a single-zeta basis set for the core shell electrons, and double-zeta or even triple-zeta basis sets for the valence shell electrons. Polarization and diffuse functions are often used to enhance the description of the valence electron distribution, they permit the electron distribution to be displaced along a particular direction, and thus a more anisotropic electron distribution can be established. These functions are symbolized with an asterisk, "*" and "+", respectively, if they are applied only to heavy atoms of a molecule, and with two asterisks, "**" or "++", if they are applied to hydrogen atoms too.

With all these notations in mind it is now possible to decipher what does the notation of a particular basis set 6-31++G, one of the basis sets used for the work presented in this Thesis, stands for. Number 6 before the hyphen represents how many primitive

Gaussians were used for obtaining the single core basis functions; numbers 3 and 1 after the hyphen represent a double zeta basis set, with three and one contracted primitive Gaussians, respectively per basis function; and finally the diffuse functions are applied to all elements. Often, to complete the notation of the level of theory, an acronym stands before the basis set, defining the method to be used. In my case, since I worked with sufficiently large molecules, DFT B3LYP method was used, hence the complete notation was: B3LYP/6-31++G.

3.5. Potential Energy Surfaces and Geometry Optimization

The concept of potential energy surface (PES) arises from the Born-Oppenheimer approximation. The approximation, that atomic nuclei of a molecule are considered stationary with respect to electrons, allows us to assign a shape of a surface of PES of a molecule as function of the nuclear coordinates. Characteristic shapes, or in other words structures of the molecule, can be explored by searching the PES that might result in finding of stationary points with minimal energy on the surface corresponding to stable structures of the molecule, i.e. molecular conformers. Thus, the concept of PES is central to geometry optimization and conformational studies (Schlegel, 1995; Schlegel, 1998). The potential energy surface or hypersurface describes the molecular energy as a function of structural parameters: bonds lengths (l_i), bond angles (θ_i) and torsional angles (ω_i):

$$E = E(l_1, .., l_i;\ \theta_1, .., \theta_{i-1};\ \omega_1, .., \omega_{i-2}) \tag{13}$$

In such a way the potential energy surface describes the conformational space of the molecule. Thus, the conformational analysis of the molecule can be identified with the location of points of interest on the potential energy surface, resulting in for example energy minima points and also *saddle points,* such as *Transition States* (TS). The process of locating the energy minima or TSs is often referred to as geometry optimization. Geometry optimization methods employ gradients, or first order derivatives, of the PES with respect to molecular structural parameters. The negative of the first derivatives of the potential energy function are the forces that act on the atoms of the molecule. When these forces are equal to zero, the corresponding point on the PES is called a *stationary point*, or a *critical point* (the latter nomenclature corresponds to topological analysis of hypersurfaces). In order to identify the character of such a critical point, i.e. whether it is energy minimum or a transition state, the second derivatives of the potential energy function by the structural parameters have to be considered.

$$H_{ij} = \frac{\partial^2 E}{\partial r_i \partial r_j} \tag{14}$$

where r_i and r_j are any of the internal coordinates of the molecule. The second derivatives represent the Hessian matrix. If the eigenvalues of the Hessian matrix are positive, then the critical point is a minimum. If the eigenvalues are negative, the critical

point is a maximum. If there are both negative and positive eigenvalues, the critical point is a saddle point. In the case of only one negative eigenvalue, the point in question is a transition state.

Note, that besides conformational analysis, the concepts of PES, minima and transition states are also valid when a mechanism of a chemical reaction is explored, such as the ATP hydrolysis reaction presented in the last chapter of the Results section. When characterizing a chemical reaction, the focus is on the reaction coordinate, which may involve changes of several bond lengths, angles, and dihedral angles. Location of minima and transition states are done in a similar fashion. The minima represent: the initial reactant state, the final product state, and the stable intermediates in between. For investigating enzymatic reaction mechanisms, both QM active site models and combined quantum mechanics/molecular mechanics (QM/MM) methods are commonly employed (Dittrich et al., 2003; Beke-Somfai et al., 2011). As the preliminary insight gained so far on the mechanism of ATP hydrolysis in RecA have only a supporting role in this Thesis, the methods employed are presented briefly.

The majority of the geometry optimization methods find equilibrium structures, i.e. structures at energy minima, and TS structures without constructing the full PES. For instance, quasi-Newton and Newton-Raphson algorithms employ a quadratic approximation of the PES around the critical point, and take a series of steps on this surface until they converge to a minimum. However, these optimization algorithms are effective in finding only the local stationary points that might correspond to the local energy minima structures, if the input geometry of the molecule is far from the global minimum. To resolve this problem, the complete PES could be estimated at a lower level of theory, and then all points of interest will be re-evaluated at a higher level of theory, sufficient for a proper description of the molecular system of interest. The level of theory used for the geometry optimization of mononuclear $[Ru(phen)_2dppz]^{2+}$, $[Ru(phen)_3]^{2+}$ and binuclear $[Ru_2(phen)_4bidppz]^{4+}$ compounds was DFT B3LYP method with hybrid basis set LANL2DZ (Hay & Wadt, 1985), to account for the transition metal.

3.6. Atomic Charge Calculations

The concept of an atomic charge does not exist in quantum chemistry, it cannot be unambiguously calculated from the wave function. However, the atomic charge is an important physical object that is widely used in various empirical-type molecular modelling algorithms, for instance, molecular mechanics or molecular dynamics. This explains the variety of models for determination of the atomic charges, e.g. Mulliken (Mulliken, 1955), Löwdin (Löwdin, 1950), Bader (Bader, 1985), CHELPG (Breneman & Wiberg, 1990), RESP (Bayly et al., 1993), etc. charge models, as well as a never-ending debate on which charge derivation scheme performs better. The conventional wisdom is that one should use whatever charge derivation scheme that the force field one is using was itself parameterized with. In the case of the work presented in this Thesis the Amber family of force fields (Cornell et al., 1995; Duan et al., 2003; Pérez et al., 2007)

were employed, thus all atomic charges of the molecular systems, including proteins and nucleic acids, were derived with the RESP model.

The derivation of the RESP charges of a molecule of interest starts with geometry optimization calculations performed with, most often, *ab initio* QM calculations. The geometries of all compound conformers have to be optimized, in case of a multi-conformational nature of the molecule of interest. Luckily, in the case of ruthenium(II) polypyridyl compounds, $[Ru(phen)_3]^{2+}$ and $[Ru(phen)_2dppz]^{2+}$ the conformational space is limited to two enantiomeric forms, Λ and Δ, that are mirror images of each other and thus the atomic charges set should also support this symmetry. In the case of a binuclear $[Ru_2(phen)_4bidppz]^{4+}$ compound the RESP charges derivation accounted for an extra degree of freedom, namely the rotation around a single bond connecting the two bridging dppz ligands. Once the geometries of all conformers are optimized, the molecular electrostatic potential (MEP) has to be calculated. These calculations are also performed with QM calculations. According to the definition the electrostatic potential at a point $\boldsymbol{r}$, $\phi(\boldsymbol{r})$ is defined as the work it takes to bring a unit positive charge from infinity to this point. The electrostatic potential has contributions from both the nuclei and the electrons of the molecule. The contributions of the nuclei *M* (15), and the electrons (16) to the electrostatic potential are:

$$\phi_{nucl}(\boldsymbol{r}) = \sum_{A=1}^{M} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \tag{15}$$

$$\phi_{elec}(\boldsymbol{r}) = \int \frac{d\mathbf{r}'\rho(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} \tag{16}$$

where $\rho(\mathbf{r})$ is the electron density and can be calculated from the Born interpretation of the wavefunction as a sum of squares of the spinorbitals at the point $\boldsymbol{r}$ for all occupied molecular orbitals. Thus, for a system of *N* electrons occupying *N/2* orbitals it can be written as:

$$\rho(\boldsymbol{r}) = 2\sum_{i=1}^{N/2} |\psi_i(\boldsymbol{r})|^2 \tag{17}$$

Then the total electrostatic potential equals to the sum of the nuclear and the electronic contributions:

$$\phi(\boldsymbol{r}) = \phi_{nucl}(\boldsymbol{r}) + \phi_{elec}(\boldsymbol{r}) \tag{18}$$

The electrostatic potential varies through space, however, since the non-covalent interactions between molecules often occur at their molecular surfaces, or in other words at separations where the van der Waals radii of the interacting atoms are just touching, it is most common to examine the electrostatic potential in these regions.

When the molecular electrostatic potential is derived, the next step is to proceed to the "mapping" procedure, namely the calculation of a set of atomic charges that would reproduce in the best way the quantum mechanically obtained electrostatic potential at a series of points on the molecular surface. The fitting procedure of the RESP charges originates from the model called ESP charges (Singh & Kollman, 1984), where the electrostatic potential was fitted onto the series of molecular surfaces, constructed using gradually increasing van der Waals radii for the atoms. The points at which the electrostatic potential was fitted were located on these atomic shells. The RESP model (Restrained Electrostatic Potential fit) introduced by Bayly et al. (Bayly et al., 1993) uses hyperbolic restraints on the non-hydrogen atoms of the molecule, so that the atomic charges of the buried atoms are effectively reduced. This was implemented to avoid high charges on the buried elements produced by the standard ESP method, which might result is an overestimation of the impact of electrostatics onto intramolecular interactions. RESP charges are also less sensitive to the molecular conformations, so that the charges of different conformers of the same molecule are comparable.

MOLECULAR DYNAMICS SIMULATIONS

Nowadays, empirical methods, such as MD, as was mentioned before, are able to model time evolution of a molecular system of size comparable with a real biological subsystem, e.g. a protein-DNA complex, with the time scale, e. g. several milliseconds, required for understanding of a biological process. Several program packages have been developed over the years for simulating atomistic behaviour of biomolecules, the most common are AMBER (Case et al., 2005), CHARMM (Brooks et al., 2009), GROMOS (Christen et al., 2005), GROMACS (van der Spoel et al., 2005) and NAMD (Phillips et al., 2005). The first simulation of a bio-macromolecule was performed in 1977 by Martin Karplus and co-workers (McCammon et al., 1977). It was a study of a bovine pancreatic trypsin inhibitor, a small 58 a.a.r. protein with a simulation time of just 8.8 ps. As the starting point the crystal structure with four water molecules, bound in the interior of the protein, was used and no additional solvent. Much has evolved since then. One of the most impressive examples of a large MD simulation is that of the Satellite Tobacco Mosaic Virus, which contained one million atoms and was simulated for 50 ns (Freddolino et al., 2006). The computer time needed to achieve a couple of ns for such a big system could be many days. While simulations of protein systems continued to improve in capacity of size and simulation times, computations of nucleic acids has lagged behind. The first simulations of nucleic acids were those of two dsDNA molecules of 12 and 24 base pairs long conducted by Michael Levitt in 1983 (Levitt, 1983). At about the same time the group of Martin Karplus investigated three dsDNA hexamers (Tidor et al., 1983). Due to the specific importance of hydrophobic interactions and the highly charged nature of DNA chains, it was crucial to treat the solvent and electrostatics more carefully than was needed for the proteins (see (Beveridge & Ravishanker, 1994; Beveridge & McConnell, 2000) for the reviews of MD simulations of nucleic acids). Since

hydrophobic as well as electrostatic effects both involve the solvent water, explicit solvent treatment became an essential attribute of a nucleic acid MD simulation. However, when the explicit solvent is included, up to 90% of the total simulation time is spent on computing solvent-to-solvent interactions. Another reason for the difficulties in simulation of nucleic acids is the structural polymorphism of these macromolecules. When the first DNA simulations started to appear, there were not so many high-resolution experimentally solved structures to be used to validate the simulation results. Thus the development of accurate force field parameters for nucleic acids was much delayed. With the increase in number of high-resolution DNA structures of various conformations, solved by X-ray crystallography or NMR, new sets of force fields began to emerge and the latest one ff-99bsc0 force field was contributed in 2007 (Pérez et al., 2007). Despite all the efforts, however, current force fields are still rather blunt empirical tools and much remains to be done in order to be able to capture the polymorphic nature of DNA and interactions with proteins and other DNA binding molecules with reasonable confidence.

3.7. Force Fields

The term *force field* has a dual meaning, it can refer to the form of a potential energy functional that describes a molecular system, or to the parameter sets used in this energy functional that explicitly characterize the atoms of the system. These atomic parameters were derived with the help of either experimental data or high-level quantum chemistry calculations on small model systems. The potential energy is calculated by adding up the energy terms that describe interactions between bonded atoms (bonds, angles and torsion) and terms that describe the non-bonded interactions, such as van der Waals and electrostatic interactions:

$$\begin{aligned} V(r^N) = & \sum_{all\ bonds} k_l(l-l_0)^2 + \sum_{all\ angles} k_\theta(\theta-\theta_0)^2 \\ & + \sum_{all\ torsions} \frac{1}{2}V_n[1+\cos(n\omega-\gamma)] \\ & + \sum_{j=1}^{N}\sum_{i=j+1}^{N}\left\{\epsilon_{i,j}\left[\left(r_{0ij}/r_{ij}\right)^{12} - 2\left(r_{0ij}/r_{ij}\right)^{6}\right] + {}^{q_i q_j}/_{4\pi\varepsilon_0\varepsilon r_{ij}}\right\} \end{aligned} \tag{19}$$

The bonded terms represent the stretching of bonds (l), bending of valence angles (θ) and rotation of dihedral angles (ω). Three force constants: k_l, k_θ and V_n characterize how difficult it is, or how much energy should be paid, to increase the equilibrium values of a bond length (l_0) or an angle (θ_0) and rotate around a torsion angle, respectively. The torsion term represents a periodic rotation of a dihedral angle with periodicity n and phase γ. The non-bonded energy is the sum of repulsion, attraction and electrostatics between non-bonded atoms. The parameter $\epsilon_{i,j}$ is related to the well-depth of the Lennard-Jones (L.-J.) potential, r_{0ij} is the distance at which the L.-J. potential has its

minimum, q_i is the partial atomic charge, ε_0 is a vacuum permittivity, ε is a relative permittivity, and r_{ij} is the distance between atom *i* and atom *j*. The L.-J. potential and the Columbic term describe the short-range non-bonded interactions. The evaluations of the long-range electrostatic interactions can be difficult and were often ignored beyond a specific cut-off distance resulting in approximations in a calculation. With the introduction of Ewald summation and particle mesh Ewald (PME) method the long-range electrostatic calculations have become more accurate (Darden et al., 1993; Sagui & Darden, 1999).

There are currently a number of empirical force fields families, which are used for macromolecules: AMBER (Cornell et al., 1995; Duan et al., 2003; Pérez et al., 2007), CHARMM (MacKerell, Jr. et al., 1998; Foloppe & MacKerell, Jr., 2000), OPLS (Jorgensen et al., 1996) and GROMOS (Oostenbrink et al., 2004). These vary slightly as to the functional form of the potential energy, mainly in the non-bonded terms, as well as values of specific atomic parameters. For a review on the development of empirical force fields for biomolecules, see (Mackerell, 2004). In the papers appended to this Thesis the AMBER family of force fields, including ff-03 and ff-10, was used.

3.8. Derivation of Force Field Parameters for Ruthenium(II) Coordination Sphere

As was discussed before, in order to perform an MD simulation the molecular system of interest has to be fully parameterized. This means that every atom of the system has to be assigned a set of parameters, including chemical bond parameters: equilibrium bond length (l_0) and bond force constant (k_l); bond angle parameters: equilibrium bond angle (θ_0) and bond angle constant (k_θ); torsion angle parameters: phase factor (γ), multiplicity (n), and barrier height (V_n); van der Waals parameters: ($\epsilon_{i,j}$) and (r_{0ij}); and point electrostatic charges (q_i).

These parameters sets are usually well defined for biological moieties, e. g., peptides, proteins, nucleic acids and polysaccharides (Cornell et al., 1995; Duan et al., 2003; Pérez et al., 2007; Kirschner et al., 2008); or can be relatively straight-forwardly derived for organic molecules, consisting of H, C, N, O, S and P, using for example Antechamber utility of AmberTools (Case et al., 2010) in association with generalized AMBER force fields, GAFF (Wang et al., 2004). However, for metal-organic compounds, such as used in this Thesis mononuclear $[Ru(phen)_2dppz]^{2+}$, $[Ru(phen)_3]^{2+}$ and binuclear $[Ru_2(phen)_4–bidppz]^{4+}$, GAFF had to be complemented by the set of MD parameters describing the Ru(II)-coordination sphere. There are three options as to how to describe the metal-ligand bonding (Hay, 1993; Comba, 1994; Landis et al., 1995; Zimmer, 1995): to treat the metal as a free atom/ion, using solely non-bonded interactions; to introduce a metal-ligand bonds but to describe all ligand-ligand interactions by non-bonded interactions; or to treat the entire compound as a covalently bound entity, using standard stretching, bending, and torsional potentials to describe intramolecular interactions. The latter option was chosen to describe metal-ligand associations. Derivation of a new set of

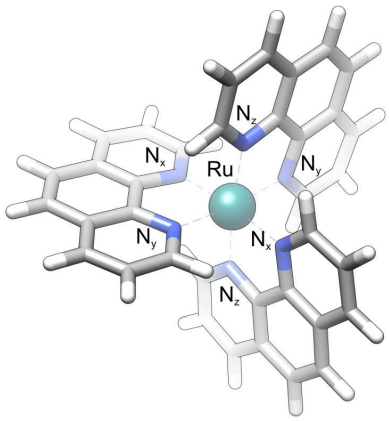


Figure 3.1. A 3-D structure of Δ-[Ru(phen)$_3$]^2, opposite, "cross-ruthenium", aza-nitrogen atoms of the phenanthroline ligands were assigned to N_x, N_y and N_z atom types.

parameters from scratch is a time consuming and tedious procedure. Luckily, for the Ru(II)-coordination sphere in Ru(II) polypyridyl compounds, the parameters set was already calculated for another empirical force field, molecular mechanics MM3* (Brandt et al., 1998).

The functional form of molecular mechanics force field is a little more sophisticated than that of MD, it includes cross terms, e.g. a bond stretch-angle bend term. Follow the parameters derivation "recipe" of Norrby and Liljefors (Norrby & Liljefors, 1998) and Brand et al. (Brand et al., 1998), ruthenium(II) force field parameters, including chemical bond, bond angle and torsion angle parameters, were validated and tuned to perform the best with Amber force fields. The remaining, non-bonded, parameters such as van der Waals parameters were adopted without any further refinement from Allinger et al. (Allinger et al., 1994), point electrostatic charges were calculated for each of the ruthenium(II) compounds separately according to the RESP charge derivation procedure (Bayly et al., 1993), preceding the actual refinement of the bonded parameters. As a parameters "training set", I have used QM-optimized structures of both enantiomeric forms of [Ru(phen)$_2$L]$^{2+}$, where L is either a phen- or a dppz-ligand, together with two crystal structures of [Ru(phen)$_3$]$^{2+}$ (Breu & Stoll, 1996) and [Ru(bpy)$_3$]$^{2+}$ (Tamura et al., 1996). Although the GAFF force filed parameters set contained parameters for an N_{sp2} type nitrogen atom, for the matter of convenience, three new N_{sp2} types for phenanthroline aza-nitrogens, "N_x", "N_y" and "N_z", were introduced (**Figure 3.1**). The full set of MD parameters compatible to Amber GAFF is provided in Supporting Information to *Paper III*. All QM calculations were performed with Gaussian 03 software package (Frisch et al., 2003).

3.9. Molecular Dynamics Calculations

In an MD simulation the energy potential is used to calculate the forces between the atoms. To initiate such a simulation all atoms of a system should have a defined starting position in space, which can for example be obtained from high-resolution X-ray and NMR structures or from a model building. Apart from the starting coordinates, atoms should also have starting velocities, which are generated from a random number series. The Newton equations of motion are then solved to calculate a time-dependent space evolution of each atom in a molecular system and, when collected together, they form a so-called MD trajectory. The force $\boldsymbol{F}_i$, which acts on an atom *i* by interactions from the rest of the system, is calculated as a negative gradient of the potential energy function with respect to the position, indicated by a 3-D vector $\mathbf{r}_i$ to give the atom *i* acceleration $\boldsymbol{a}_i$ (equations (20) and (21) respectively):

$$\boldsymbol{F}_i = -\frac{dV}{d\boldsymbol{r}_i} \tag{20}$$

$$\boldsymbol{a}_i = \frac{d\boldsymbol{r}_i^2}{dt^2} = \frac{\boldsymbol{F}_{r_i}}{m_i} \tag{21}$$

where m_i and $\boldsymbol{r}_i$ are the mass and position of the atom *i*, respectively. The change of the atom's position after a short time interval (Δt) can then be calculated using Taylor series. There is a number of algorithms for the integrations of the motion equations (reviewed in (Leach, 2001)), the integration is broken down into small steps, separated by usually 1-2 fs time steps. The Verlet algorithm (Verlet, 1967) is a widely common method, it uses atomic positions $\mathbf{r}_i(t)$ and accelerations $\boldsymbol{a}_i(t)$ at a time t together with the positions from the prior step, $\mathbf{r}_i(t-\Delta t)$ to determine a new position , $\mathbf{r}_i(t+\Delta t)$:

$$\boldsymbol{r}_i(t+\Delta t) = 2\boldsymbol{r}_i(t) - \boldsymbol{r}_i(t-\Delta t) + \Delta t^2 \boldsymbol{a}_i(t) \tag{22}$$

The velocities, however, do not explicitly appear in the Verlet integration algorithm but can be calculated using the following central difference approximation:

$$\boldsymbol{v}_i(t) = \frac{\boldsymbol{r}_i(t+\Delta t) - \boldsymbol{r}_i(t-\Delta t)}{2\Delta t} \tag{23}$$

At each time step for each atom equations (20) – (23) have to be calculated. And in such a way an MD trajectory is obtained. It contains the positions of all atoms in the system as a function of time.

The very first MD simulations were carried out in vacuum, but since biological molecules are usually surrounded by solvent (excluding membrane systems) most of MD simulations today are recorded in explicit solvent. A molecular system of interest is solvated by a finite water shell, which on one side shares an interface with the solvent but on the other with vacuum. A well-used solution for such a problem is to multiply the system

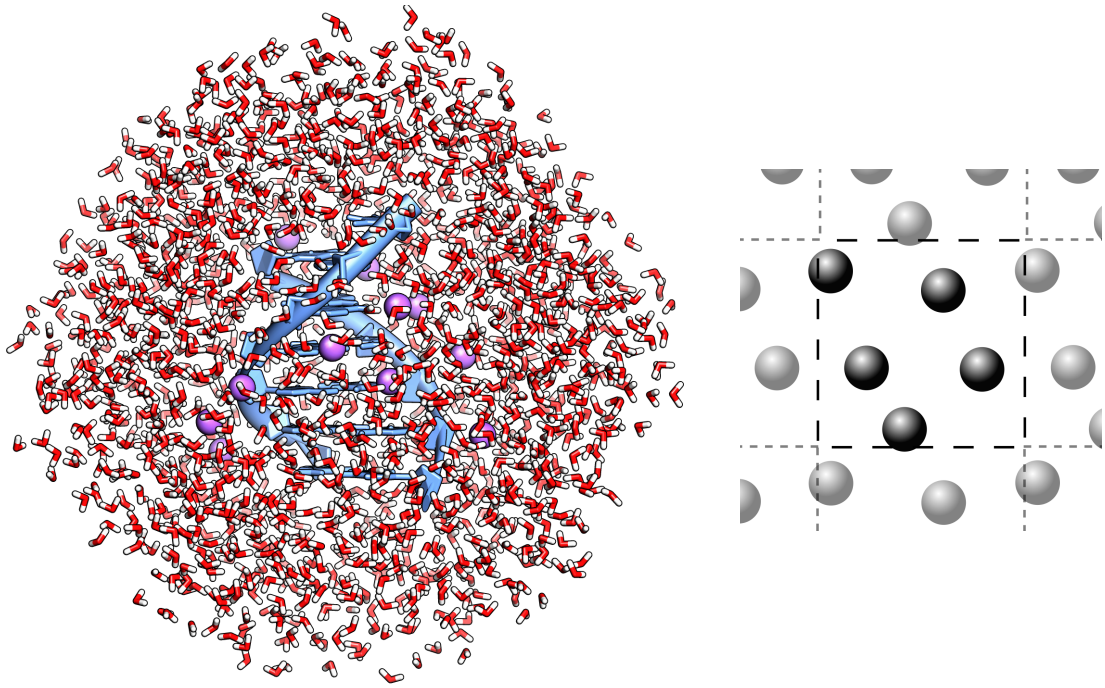


Figure 3.2. Left: a DNA molecule solvated in a truncated octahedron of explicit water. Right: a schematic representation of periodic boundary condition.

periodically in all 3 directions to represent an infinite system (**Figure 3.2**). When the so-called periodic boundary conditions are applied, all simulated atoms are surrounded by neighbours. Periodic boundary conditions can be used with various geometries, a 3-D cubic box, a rhombic dodecahedron, or a truncated octahedron. The latter is often used to reduce the number of solvent molecules in the system to speed up the computation. Another way of minimizing the number of atoms in a simulation is to apply a spherical (non-periodical) boundary condition. This setup is useful when studying a particular part of a large system, where the region of interest is enclosed within a shell, where the atoms are subjected to MD, whereas the outside atoms are kept fixed to their initial positions. Solvent can also be represented via continuum models, usually referred to as implicit solvent.

Molecular dynamics experiments are usually performed in conditions as close as possible to experimental, so physical values like constant pressure or volume and temperature are carefully chosen. In this way, the most common setups of MD experiments can be divided into: NVE or microcanonical ensemble, where the thermodynamic state is characterized by a constant number of particles (N), volume (V) and energy (E); NVT or the canonical ensemble where all N, V and temperature (T) are constant; or NPT – the isobaric-isothermal ensemble where N, pressure (P) and T are constant. In order to keep the constant temperature, the velocities can be rescaled (Woodcock, 1971) during the course of simulation, or the system can be coupled to a temperature bath (Hoover, 1985). The Langevin thermostat temperature control

method (Izaguirre et al., 2001) was used for the calculations represented in this Thesis. Constant pressure can be maintained by allowing the volume to fluctuate or to use a "pressure bath" in analogy to the temperature bath. In the work included in this Thesis the simulations were carried out in either NVT of NPT ensembles.

One of the main limitations of MD simulations is still a feasible time frame. With the development of supercomputers, as well as graphic processing units, and ways to parallelize the calculations, it is now possible to record longer and longer trajectories. However, the time scale is not yet sufficiently long to illustrate, for example, a biological reaction. The length of standard simulations, performed in the studies included into this Thesis, is fluctuating between 10-50 ns for systems from 25 000 atoms to 125 000 atoms. Longer (> 100 ns) trajectories, despite becoming more and more common, are still very computationally intense. To speed up the calculation rate all the fastest modes of bonds vibrations (time scale of approximately femto-seconds) are kept frozen using SHAKE algorithm (Ryckaert et al., 1977). The hydrogen bonds are constrained to fixed length and by imposing this restraint larger integrating time steps can be used without a significant loss of accuracy of the calculations.

3.10. Unrestrained Molecular Dynamics (MD) versus Steered Molecular Dynamics (SMD)

The major difference between unrestrained molecular dynamics or, simply MD, from a steered MD approach is the ability of the latter to bias the system's evolution along some predefined coordinate by applying an external force, which drives a change in coordinates within a certain time. Several applications of SMD have been developed in Klaus Schulten's group (Izrailev et al., 1998; Jensen et al., 2002), one of which is when the coordinate of interest is moving with a constant velocity. This implementation, done by the group of Dario Estrin (Crespo et al., 2005), has been included into AMBER11 software package and used in the studies included in this Thesis. This method is based upon that the centre of the restraint is time-dependent as in:

$$V_{restraint}(t) = \frac{1}{2}k(x - x_0(t))^2 \qquad (24)$$

where x could be a distance, an angle, or a torsion angle between atoms or groups of atoms. This approach can be used to drive various physical processes, for example, ion diffusion or ligand binding (as illustrated in *Paper III*), conformational changes, etc. By integrating the force over time or distance, a generalized work can be calculated. Subsequently, this work can be used to compute free energy differences using the Jarzynski relationship (Jarzynski, 1997; Hummer & Szabo, 2001). According to the method the free energy difference between two states in question, A and B, with different coordinates (x) is equal to:

$$e^{-\Delta G/k_B T} = \langle e^{-W/k_B T} \rangle_A \qquad (25)$$

And so, by calculating the work between the two states, A and B, and averaging over the initial state, A, equilibrium free energies can be extracted from non-equilibrium calculations. The SMD approach proved to be useful in various studies of protein mechanics (Israelewitz et al., 2001) and folding and unfolding pathways (Lu et al., 1998; Liu et al., 2010; Pradhan et al., 2011).

3.11. Replica-Exchange Molecular Dynamics (REMD)

A "standard" MD simulation does not usually provide an accurate canonical distribution of molecular conformations at room temperatures. Usually in simulations at temperatures around 300 K, a molecular structure tends to get trapped in one of a huge number of local minimum-energy states. To overcome this multiple-minima problem, a Replica-Exchange Molecular Dynamics approach (Sugita & Okamoto, 1999) was developed, a simulation based on non-Boltzmann probability weight factors so that a random walk in energy space may be achieved. Random walks allow the simulation to overcome any energy barrier and to sample a much wider conformational space than by "standard" methods. While a molecular system performs a free 1-D random walk in energy space, its *simulated tempering* does a free random walk in a temperature space. The latter, in turn, induces a random walk in energy space and allows the simulation to enrich the conformational sampling in high-energy regions, by escaping from local minima-energy states. In the replica exchange MD method, non-interacting copies of a system (replicas) are simulated concurrently at different values of temperature or some other independent variable. With some periodicity the replicas are subjected to a Monte Carlo type evaluation, thus effecting exchange between the values of, for example, temperature. And so, the REMD method enables simulation in a generalized ensemble, in which states may be weighted by non-Boltzmann probabilities (Mitsutake et al., 2001). The Replica-Exchange MD method in temperature space was successfully applied in a number of folding studies of proteins and RNA (García & Onuchic, 2003; Andrec et al., 2005; Mu et al., 2006; Garcia & Paschek, 2008).

3.12. MM-PB(GB)SA approach for analysis of binding free energies

Free energy is one of the most important thermodynamics quantities, for a molecular system it envisages the system's tendency to associate and react. Free energy is usually expressed as the Helmholtz function (A), suitable for a system with a constant number of particles, temperature and volume (NVT ensemble), or the Gibbs function (G), appropriate for a system with a constant number of particles, temperature and pressure (NPT ensemble). Since most of experiments are conducted under conditions of the NPT ensemble, the G function is normally used. In molecular biology binding free energies are commonly associated with a relative affinity, for example, of a protein to its cognate DNA compared to a mutated DNA molecule. Calculations of the free energies of biological macromolecules, due to their extensive conformational space with many local minimum energy configurations separated by low energy barriers, are immensely difficult and computationally intense. "Standard" molecular dynamics techniques, due to

inadequate sampling of the conformational space, thus, cannot accurately determine the free energy. A number of free energy calculation techniques (see for reviews (Beveridge & DiCapua, 1989; Straatsma & McCammon, 1992; Kollman, 1993) on free energy calculations) to circumvent these problems have been developed over the last 30 years since first free energy calculations were done by Berendsen (Postma et al., 1982) and McCammon (Tembre & McCammon, 1984). Each of the methods is particularly useful when applied to a certain class of problems. For example, free energy perturbation (FEP) is usually referred to as a computational alchemy method since it involves transformations of atoms from one molecular unit to another. Unlike FEP, potential of mean force and umbrella sampling technique represents a physically achievable process, the free energy is related to the probability distribution of states as a function of a reaction coordinate along which the free energy is sampled. The accurate MD free energy calculations are notoriously time consuming to perform, thereby alternative approximate "rapid" methods have been developed, one of which is MM-PB(GB)SA (Srinivasan et al., 1998; Kollman et al., 2000).

The Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) and the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods represent the post-processing analysis for the evaluation of binding free energies of two molecules or absolute free energies of molecules in solution. Sets of molecular structures are collected with molecular dynamics or Monte Carlo methods. The MM-PB(GB)SA approaches combine the molecular mechanical energies, which represent the internal energy, van der Waals and electrostatic interactions, with the continuum solvent approaches. The electrostatic contribution to the solvation free energy is calculated with the Poisson-Boltzmann equation, where the solvent is modelled as a continuum of high dielectric, which estimates the electrostatic contribution to the solvation free energy. The hydrophobic contribution to the solvation free energy is determined with solvent-accessible-surface-area-dependent term. The binding free energies are evaluated according to the equations (26) – (28). The contribution of the change to the conformational entropy during complex formation between two molecules can be estimated with the help of normal mode analysis on a set of conformational snapshots taken from MD simulations.

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{solvation} - T\Delta S \quad (26)$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{VDW} \quad (27)$$

$$\Delta G_{solvation} = \Delta G_{PB/GB} + \Delta G_{SA} \quad (28)$$

where ΔE_{MM}, $\Delta G_{solvation}$ and $-T\Delta S$ are the changes of the gas phase molecular mechanics energy, the solvation free energy, and the conformational entropy upon binding, respectively. ΔE_{MM} includes $\Delta E_{internal}$ (bond, angle, and dihedral angle energies), $\Delta E_{electrostatic}$ (electrostatic), and ΔE_{VDW} (van der Waals) energies. $\Delta G_{solvation}$ is the sum of

electrostatic solvation energy (polar contribution), $\Delta G_{PB/GB}$, and the non-electrostatic solvation component (nonpolar contribution), ΔG_{SA}.

The MM-PB(GB)SA approach has been successfully applied to various protein-ligand systems (Kuhn & Kollman, 2000; Wang & Kollman, 2001; Hou et al., 2003) or protein-protein/peptide complexes (Gohlke & Case, 2004; Wang & Kollman, 2000), but their performance is system-dependent (Kuhn et al., 2005; Pearlman, 2005). In addition, MM-PB(GB)SA is sensitive to simulation protocols, such as the sampling strategy of generating snapshots and entropy calculation methods as well as other parameters, e.g., charge models, force fields, the solute dielectric constant, and radius parameters in continuum solvent models.

BIOINFORMATICS TOOLS

According to the definition, *bioinformatics* is the application of computer science and information technology to the fields of molecular biology, biochemistry and medicine, which deals with data mining, structural and sequential search algorithms, data bases, molecular modelling and simulations, etc. Historically bioinformatics was first applied in the creation and maintenance of nucleotide and amino acid sequences databases to store biological information at the beginning of the "genomic revolution". Interestingly, the term "bioinformatics" has appeared long before the actual genomic revolution, it was coined by Paulen Hogeweg in 1978 (Hogeweg, 1978). Since then, the field of bioinformatics has evolved, nowadays various applications have been developed to analyze, interpret and compare DNA/RNA and protein sequences, molecular domains and conformations, and protein structures and protein structure prediction, an overview of bioinformatics tools can be found elsewhere (Leach, 2001).

3.13. Homology modelling

There is a striking likeness between the 3-D structures of some proteins, most of these proteins are often homologues, namely they are performing identical functions but in different organisms, but it is also possible for biologically unrelated proteins to show significant structural similarity. Homology modelling (or comparative modelling) exploits these similarities in protein folding by constructing a 3-D structural model based upon known high-resolution structure(s) of the best-related proteins. The process of creation of a new protein model has three main stages: finding of the template protein structure, sequential alignment between the target and the template proteins, and finally, structure refinement (usually performed with MD methods). In cases when the protein's function is not known, sequential alignment provides invaluable insight by screening amino acids databases in search for the best alignment score, to relate the protein to a particular protein family.

The resulting alignment of the target protein and its template can have both sequentially converged regions and sequentially variable regions. The former correspond to the

maximum sequentially identical stretches, where it is expected that the conformation of the unknown protein is very alike to the template protein(s). The sequentially and structurally conserved regions are often found in the core of the protein or at the protein active site. The structurally variable regions usually, but not necessarily, correspond to polypeptide loops. These loops might show some variation as to their length and are usually disordered.

There are three major classes of homology modelling methods (Baker & Sali, 2001; Zhang, 2008). The first class is based on piecing together as rigid bodies the bits of the 3-D structure(s) of the template protein(s) (the method is also known as Fragment Assembly). The second class of methods (Segment Assembly) assemble the target protein by joining together the small segments of the template protein(s) or by reconstructing the set of coordinates by dividing the target protein into a series of short segments, for each of which will be tried to find a perfect sequential match with its own template from the, for example, Protein Data Bank. Thus, sequence alignment is done over segments rather than over the entire protein like in the first class of methods. The third approach generates a set of spatial constraints from the template protein(s). The constraints set is used in conjunction with an optimization procedure, very analogous to the NMR structure refinement, to derive the new structure.

A number of software tools are available for performing the homology modelling, the most widely used are SWISS-MODEL web-server and SWISS-PDB-viewer (Arnold et al., 2006), MODELLER (Sali & Blundell, 1995; Fiser et al., 2000) and WHATIF (Vriend, 1990). All homology-modelling methods and tools are put to test every two years through the Critical Assessment of Techniques for Protein Structure Prediction (CASP). CASP is a community-wide prediction experiment that challenges protein prediction teams to submit structural models for a number of sequences whose structures have recently been solved experimentally but have not yet been published.

3.14. Multiple sequence alignment

In 1977 the first DNA sequence of Phage Φ-X174 was identified (Sanger et al., 1977), since then thousands of organisms have been decoded and their genome sequences stored in databases. These sequences are analysed to determine genes, regulatory sequences, structural motifs, etc. By sequential comparison of genes within a species or between different species it is possible to identify similarities between protein function and structure, as well as evolutionary relations between species. The relationship between proteins' sequence, structure and function was first examined by Chothia and Lesk in 1986 (Chothia & Lesk, 1986), who have showed that proteins with similar sequences tend to have similar 3-D folding. Since then the sequential alignment gained wide popularity, which gave rise to the development and further implementations of various alignment algorithms. The objective of a sequence-alignment algorithm is to position the amino acid sequences (or nucleotide sequences in case of nucleic acids

comparison) so that the matched stretches correspond to common structural or functional features.

Three main types of sequence-alignment algorithms can be identified: first type of algorithms is designed to attempt to match along the entire length of the sequences; second algorithms search for local alignments involving sections from the sequences. The best-known and commonly used examples of these two approaches are Needleman and Wunsch (Needleman & Wunsch, 1970) and Smith and Waterman (Smith & Waterman, 1981) algorithms, respectively (see (Leach, 2001) for details). Both types of these algorithms are computationally intense. Therefore, a third type of algorithms, more heuristic and approximate, was developed, e.g. “word methods”, BLAST (Altschul et al., 1990) and FASTA (Lipman & Pearson, 1985). “Word methods” are especially useful in large-scale database searches where a large proportion of the candidate sequences will have essentially no significant match with the query sequence. “Word methods” identify a series of short, non-overlapping subsequences ("words") in the query sequence, which are then matched to candidate database sequences. The relative positions of the “word” in the two compared sequences are subtracted to obtain an offset; this will indicate a region of alignment if multiple distinct words produce the same offset. Multiple sequence-alignment is a comparison of three and more sequences, performed in attempt to identify conserved regions across a group of sequences. Simultaneous alignment of all the sequences in the query is an NP-complete problem (Cook, 1971; Wang & Jiang, 1994; Elias, 2006) (a solution could be found by e.g. testing all possible combinations, which takes time exponential in the size of the sequences, and no polynomial algorithm exists to date).

To shorten the time, progressive alignment methods were developed; they first align the most similar sequences and then add successively less related sequences or groups to the alignment until the entire query set has been incorporated into the solution. The initial alignment is done by pairwise comparison methods, like earlier named FASTA. Progressive alignment results depend on the choice of "most related" sequences and thus are sensitive to inaccuracies in the initial pairwise alignments. Most progressive multiple sequence alignment methods additionally weight the sequences in the query set according to their evolutional relatedness to reduce the likelihood of making a poor choice of initial sequences and thus improve the alignment accuracy. The most widely used implementation of progressive multiple sequence-alignment is Clustal and various modifications like ClustalW or ClustalX (Higgins & Sharp, 1988; Larkin et al., 2007; Thompson et al., 1994). ClustalW was used in *Papers IV* and *V* of this Thesis.

3.15. Molecular Docking

Molecular Docking is a widely used method for prediction of either a ligand’s orientation in a receptor’s binding site, or of a binding site on receptor’s surface, or of binding arrangements of interacting macromolecules (Blaney & Dixon, 1993; Lengauer & Rarey, 1996). This knowledge can be then used to estimate binding affinities between the

molecules. Molecular docking can be considered as a "key-lock" problem, especially when applied to a macromolecule-ligand system, where one is interested in finding a correct relative orientation of the "key" that will fit the "lock". Molecular docking may be also identified with the optimization problem to achieve the "best-fit" orientation of a ligand or the "best-fit" ligand when applied to a computer-aided drug design problem inside the binding site of a target protein. Since both ligand and target binding-sites are often flexible, the rigid "key-lock" analogy was exchanged for a more appropriate "hand-in-glove" one (Jorgensen, 1991). And thus, the objective of molecular docking is to achieve an optimized global conformation of both a protein and a ligand, and the relative orientation between them, so that the energy of the overall system is minimized. The "docking problem" can be divided into two: generation and evaluation of plausible structures of intramolecular complexes.

Docking involves many degrees of freedom: the 6 degrees of translational and rotational freedom of one molecule relative to another, as well as the conformational degrees of freedom (DOF) of each molecule. Various algorithms have been developed to tackle the problem; they can be characterized according to the number of DOF that they ignore. Thus the simplest algorithm treats molecules as rigid bodies and explores only the 6 DOF. In this case a manual "hands-on" approach using interactive computer graphics is usually useful, especially in the case when the binding site is well known and both of the interacting molecules have countless DOF. However, for small molecules with less DOF, to decrease the bias, automated docking approaches that usually consider many more possible binding arrangements were developed. A well-known example of such an algorithm is the DOCK program (used in *Paper V*) of Kuntz and co-workers (Kuntz, 1992; Meng et al., 1992; Shoichet et al., 1992). DOCK is designed to find the ligand's orientation inside the receptor's binding site to be as highly shape-complementing as possible. The program first derives a "negative" image of the receptor's binding site from the molecular surface or water assessable molecular surface (designed by rolling a probe of a predefined radius, often chosen as 1.4Å, the radius of a water molecule). The negative image is then translated to a collection of overlapping spheres of varying radii, each of which has two points in contact with the receptor's binding site. Then ligand atoms or groups of atoms matched to the spheres centres to find matching sets in which all the distances between the ligand atoms in the set (or clique) are equal to the corresponding sphere centre-sphere centre distances within some predefined tolerance. The ligand is then oriented so that the least square fit of the ligand atoms to the sphere centres becomes minimum. The final orientation is checked to ensure that there are no unacceptable steric clashes. The successful pre-screened candidates are then subjected to energy minimization with the help of molecular mechanics with continuous solvent models, and finally scored to release the top best-fit orientations.

To perform conformationally flexible docking conformational DOFs have to be taken into account. Due to immense number of DOF of the receptor, which would lead to

computationally intense calculations, most of the flexible docking methods consider only the conformational space of the ligand, while the receptor is still treated as a rigid body.

3.16. Analysis of Electrostatic Potential Surface

In *Paper V* of this Thesis an electrostatic potential surface analysis has been performed to facilitate the "rigid-body" type docking of ssDNA fragment inside the human Rad51 (HsRad51) protein filament. To generate the electrostatic potential surface a 3-monomers long fragment of the HsRad51 protein filament was subjected to Poisson–Boltzmann electrostatics calculations with APBS Poisson–Boltzmann solver software (Baker et al., 2001; Baker, 2004). In Poisson-Boltzmann (PB) approach the solute is treated as a body of constant low dielectric permittivity (permittivity is set to be usually between 2 and 4), the solvent, on the other hand, is modelled as a continuum of high dielectric permittivity. The Poisson equation relates the variation in the electrostatic potential ϕ within a medium of uniform dielectric constant ε to the charge density ρ:

$$\nabla^2\phi(\boldsymbol{r}) = -\frac{\rho(\boldsymbol{r})}{\varepsilon\varepsilon_0} \tag{29}$$

For a set of point charges in a constant dielectric the Poisson equation reduces to the Coulomb's law. However, if the dielectric is not constant but varies with position, the Coulomb's law is no longer applicable and the Poison equation becomes more complex:

$$\nabla(\varepsilon(\mathbf{r})\nabla\phi(\boldsymbol{r})) = -\frac{\rho(\boldsymbol{r})}{\varepsilon_0} \tag{30}$$

In the presence of mobile ions the Poisson equation has to be further modified, to account for their redistribution in the solution in response to the electric potential. Thus the ion distribution is described by a Boltzmann distribution of the following form:

$$n(\boldsymbol{r}) = Ne^{-\mathcal{V}(\boldsymbol{r})/k_BT} \tag{31}$$

where $n(\boldsymbol{r})$ is the number density of ions at a particular location $\boldsymbol{r}$, N is the bulk number density and $\mathcal{V}(\boldsymbol{r})$ is the energy change to bring the ions from infinity to the position $\boldsymbol{r}$. When incorporated together the Poisson equation becomes the Poisson-Boltzmann equation, which represents the non-linear differential equation and to be solved it can be expanded as a Taylor series of hyperbolic sine functions. The linearized Poisson-Boltzmann equation looks like this:

$$\nabla\big(\varepsilon(\mathbf{r})\nabla\phi(\boldsymbol{r})\big) - \kappa^2\phi(\boldsymbol{r}) = -\frac{\rho(\boldsymbol{r})}{\varepsilon_0} \tag{32}$$

where κ is related to the Debye-Hückel inverse length. Even the linearized equation can be analytically solved for simple geometric forms. Before the computers' era, the proteins were treated as spheres or ellipsoids; DNA as a uniformly charged cylinder, and

membranes as planes. With computers a variety of numerical approaches can be used, including finite element and boundary element methods. The finite element approach formed the bases of APBS Poisson–Boltzmann solver software, which was used in this Thesis. In the method the cubic lattice is superimposed onto the solute(s) and the surrounding solvent. Values of the electrostatic potential, charge density and ionic strength are assigned to each of the grid points. The atomic charges are allocated to the eight surrounding grid points in a such a way that the closer the charge to the grid point the greater the proportion of its total charge that is allocated. The derivatives of the Poisson-Boltzmann equation are then determined by the finite difference formula. To perform a PB calculation it is necessary to assign a value of the dielectric constant to each of the grid points. This is done by the definition of the boundary between solute and solvent to be either molecular surface or solvent accessible surface. All the points outside the surface are identified with high dielectric constant (e.g. 80 for water) and the ionic strength value. The inner-surface points are identified with the dielectric constant of a macromolecule.

4. RESULTS

In this chapter the main results and conclusions achieved in Papers I-V are summarized and put into context, as well as some preliminary unpublished results of relevance are presented.

DNA, the central molecule of life, has always attracted a lot of attention of researchers. Despite the popularity there are still countless questions on the mechanisms of molecular function, interactions with small ligands and proteins, ways of utilizing existing information for creating effective medicines, etc. Being a highly polymorphic molecule, DNA dynamically adjusts itself to changes in surrounding environments, and to a great extent, this conformational flexibility defines the character of intramolecular associations. Some extremes of DNA's plasticity could be tackled in experimental models, such as by studying the interactions with a dumbbell-shaped ruthenium binuclear compound, which has been reported to intercalate into DNA containing a sufficient number of adjacent AT base pairs. Some two weeks at room temperature was the time the compound, being conformationally rigid, took before it managed to thread one of its bulky ruthenium centres through a tightly packed DNA stack. Such an exceptionally slow kinetics indicates large structural reorganizations of the DNA duplex. In *Paper I* an attempt is made to highlight the mechanism of threading intercalation: several different DNA constructs were tested which resulted in the hypothesis that a transition from B- to A-DNA might be induced by the presence of the initially outside binding of the Ru(II)-compound that in turn might facilitate the threading event.

In *Paper II*, further effort was made to resolve, at atomic scale, a structure of a DNA-ruthenium compound complex, using NMR spectroscopy supported by molecular modelling. Due to solubility problems with the more hydrophobic phenanthroline ruthenium compound, that had shown the highest threading intercalation rates, we had instead to use its much slower but less hydrophobic bipyridyl analogue. Although the end product is not an intercalated state, but one in which the compound remains bound on the outside of the duplex, the obtained structure is interesting, as it may represent the precursory state before threading, i.e. the binding geometry that Wilhelmsson et al. observed initially, characterized by a positive LD (Wilhelmsson et al., 2002).

In *paper III* a computational study was conducted aiming to gain some mechanistic insight into molecular interactions, preceding threading intercalation, exhibited by binuclear ruthenium compounds, as well as to resolve an old debate about details of intercalative binding to DNA by ruthenium-coordinated mononuclear compounds.

The structural studies of recombinases: human Rad51 protein, conducted in *Papers IV-V*, and bacterial RecA protein, preliminary results are described in the last chapter of this section, similarly to high-resolution structural techniques, X-ray and NMR spectroscopy,

helped us gain insight into mechanisms of the proteins' function. In the case of human Rad51, where no filament structure was available and for the monomeric protein only fragments of structure were resolved, the initial step was to assemble an atomistic filament model (*Paper IV*). This model can be described as a 3-D jigsaw puzzle, collected bit by bit from all accessible structural information.

In *Paper V* this model was further refined and successfully exploited to address a particular mechanism of the stimulatory effect of Ca^{2+} on the strand exchange activity of the protein.

In the case of prokaryotic RecA, high-resolution crystal structures of the free filament as well as in the complex with ss- and dsDNA are available (Chen et al., 2008). These represent the key-step structures of homologous recombination reaction: free recombinase – presynaptic filament – postsynaptic product. And thus, the next logical step, to push the horizons of understanding of homologous recombination reaction performed by RecA, is the detailed atomistic studies targeting a specific mechanism, e.g. ATP hydrolysis, as illustrated with some preliminary results in a chapter "On the Mechanism of ATP-Hydrolysis in Bacterial Recombinase RecA".

PAPER I. MECHANISM OF DNA THREADING INTERCALATION

Threading intercalation of the binuclear ruthenium(II)-coordinated polypyridyl compounds into $[poly(dAdT)]_2$ in comparison to mixed sequences of calf thymus DNA (ctDNA), at room temperature, may occur within a few minutes up to several days respectively, as detected by the luminescence intensity traces (Nordell et al., 2007). The appearance of a bright luminescence indicates a hydrophobic shielding of the nitrogen atoms of the bidppz bridging ligand from solvent water – or in other words intercalation. When phenanthroline (P) ancillary peripheral ligands of the bidppz bridge are substituted to less hydrophobic, though spatially smaller ligands, bipyridines (B), the rate of threading drops about 1.3-fold in case of $[poly(dAdT)]_2$ and 20x for ctDNA at 37°C, see **Figure 4.1** (Nordell & Lincoln, 2005; Nordell et al., 2007). Such a dramatic sequence dependency, as well as the fact that a smaller object requires longer time to penetrate through a hole than a bigger one, which is clearly counterintuitive, led to the conclusion that the mechanism of threading intercalation into DNA lies beyond a regular equilibrium binding. Further studies of threading intercalation phenomena revealed yet another interesting detail: that a sufficiently long stretch of alternating AT base pairs is required for the efficient intercalation (Nordell et al., 2007).

A detailed analysis of threading rates dependency on the length of AT-tract was performed in *Paper I* with a help of a series of HEG-(hexaethylene glycol)-linked hairpin oligonucleotides containing a central AT-step repeating from 3 to 22 times. The result was enigmatic: a threshold when the threading intercalation becomes efficient was found, lying between 10 and 14 base pairs of alternating AT. Any further increase of the central AT length showed only a moderate increase in the rates. Interestingly, a pure A-

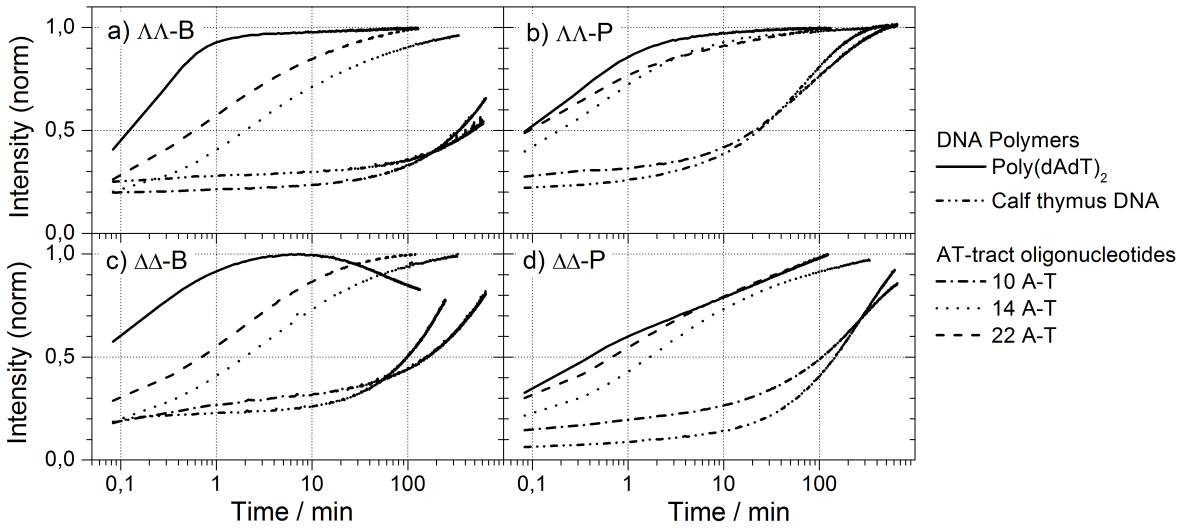


Figure 4.1. AT-tract length dependence of threading intercalation efficiency. Luminescence after mixing the Ru(II) compounds with DNA polymers and HEG-linked oligonucleotide hairpins at 37˚C (luminescence intensity normalized to final/maximum intensity). B = bipyridine compound, P = phenanthroline compound. The figure is prepared by Pär Nordell.

stretch, 14 successive adenine bases, displayed a much slower and weaker luminescence increase, just behaving like alternating GC sequences which had earlier been observed to completely prevent threading (Nordell, 2009).

The hypothetic threading intercalation mechanism, suggested in *Paper I* to explain the observed threshold phenomenon, can be described as a jigsaw puzzle. Two experimental facts that a hydrophobic compound, despite being spatially bigger, takes shorter time to introduce one of its bulky ruthenium centres through the tight DNA stack; and that a stretch of 10 to 14 alternating AT base pairs is required for the efficient threading intercalation, were the two starting puzzle bits. Alternating AT sequences are highly biologically relevant (Lifton et al, 1978; Smale & Kadonaga, 2003). This is a signature composition of promoter regions of genes in archaea and eukaryotes: approximately 24% of human genome contains a so-called TATA-box within the core promoter regions (Yang et al., 2007). The TATA sequence is the binding target of transcription factor proteins as well as histones; the binding of either prevents the binding of the opposite. When investigating high-resolution crystallographic data of DNA complexes with TATA-box binding proteins (TBP) (one of which is of 1.9 Å resolution, PDB ID: 1CDW (Nikolov et al., 1996)) (**Figure 2.5**, right, page 13), I noticed structural similarities between the TBP and the ruthenium(II) phenanthroline dimer, namely the spatial span of the binding interface as well as the hydrophobic character of the interaction. In the crystal structure the DNA helix is bent as much as 80° and, most excitingly, resembles no longer a B-duplex but has rather an A-like DNA structure. The fact that the outside binding of a hydrophobic species could induce a conformational rearrangement of the DNA helix was, for me, yet another bit in the jigsaw puzzle. However, to thread through DNA it is

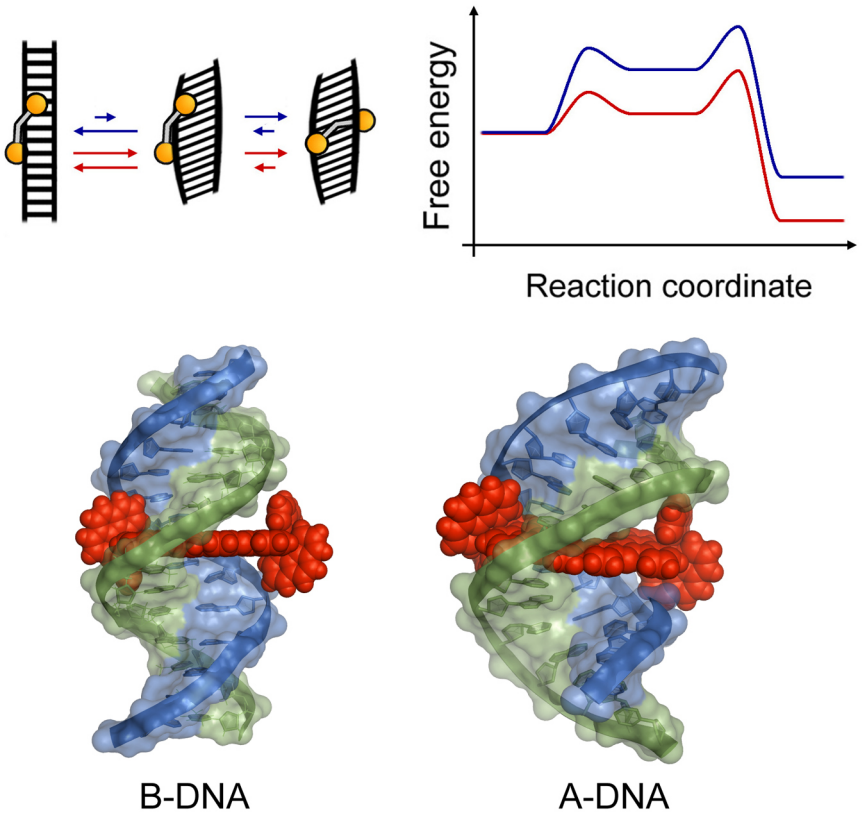


Figure 4.2. Top: Schematic representation of proposed mechanism of threading intercalation of binuclear Ru(II) compounds into AT-tracts oligonucleotides. Bottom: Two models of the final intercalated binding geometry of ΔΔ-P in a 14 AT base pairs duplex.

not enough to modify DNA structure, at least one base pair has also to be opened to allow the threading of the bulky ruthenium centre, as was concluded from DNA stretching experiments with optical tweezers in the presence of ΛΛ-P (Paramanathan et al., 2008). Further data mining revealed theoretical evidence in support for that the suspected conformational rearrangement of the DNA helix into an A-like form could indeed assist threading. From free energy calculations of an AT base pair opening it was suggested that the C3´-endo conformation of a DNA backbone furanose (A-form DNA furanose) facilitates a base pair opening (Giudice et al., 2001). All pieces of the puzzle made a nice story: the groove binding of a hydrophobic phenanthroline ruthenium dimer, probably by lowering the water activity, initiates a local conformational rearrangement of DNA from B- to A-like form. Thereafter, the A-DNA sugars would facilitate the base pair opening to finalize the event of threading intercalation. In the case of the less bulky, but also less hydrophobic bipyridyl compound it was proposed that the B-A transition is less effective and the conformational rearrangement of DNA, hence, requires longer times.

To support the proposed threading mechanism, outlined in **Figure 4.2**, a set of molecular dynamics simulations was initiated. Treating ruthenium(II) ion as a dummy

ion, with a frozen coordination-sphere parameters (since no Ru(II) MD parameters were available at the time of the study), allowed calculations of groove bound and intercalated states to be performed. Analogously to the binding geometry of the TBP-DNA complex in the crystal structure, the ΔΔ-P compound was placed in the minor groove, while the intercalation spot was chosen to be in the centre. A 10 ns-long MD run of the groove-bound state indeed resulted in a conformation resembling the one of the TBP-case (**Figure 4.3**): with a helical bending of 45°, and most excitingly a C3'-endo furanose conformation of the two central bases of the 3'-5' strand. The simulation of the intercalated state, when starting with ΔΔ-P positioned in-between base pairs of a B-form duplex, not surprisingly though, did not result in any A-form-like threaded product. The MD calculations suggested that the intercalation pocket formed in B-DNA could only protect one of the two dppz-moieties, while the second one was fully exposed to solvent (**Figure 4.2** bottom left). On the other hand, when placed between the adjacent base pairs of an A-form duplex the bidppz bridging ligand of the ΔΔ-P compound was much better shielded from the surrounding waters (**Figure 4.2** bottom right). The latter geometry, thus, seems to be more relevant to explain the observed massive increase in fluorescence intensity. Additional attempts were made to model the actual threading with the help of the steered MD approach, although I considered this exercise less meaningful due to the absence of any high-resolution data.

Although, the simulations did not result in any illustration of the complete threading intercalation, starting from an outside binding geometry and finishing at a thread-intercalated state of the ΔΔ-P compound, they still provided some interesting mechanistic insight into the process. The outside binding of the ΔΔ-P compound has

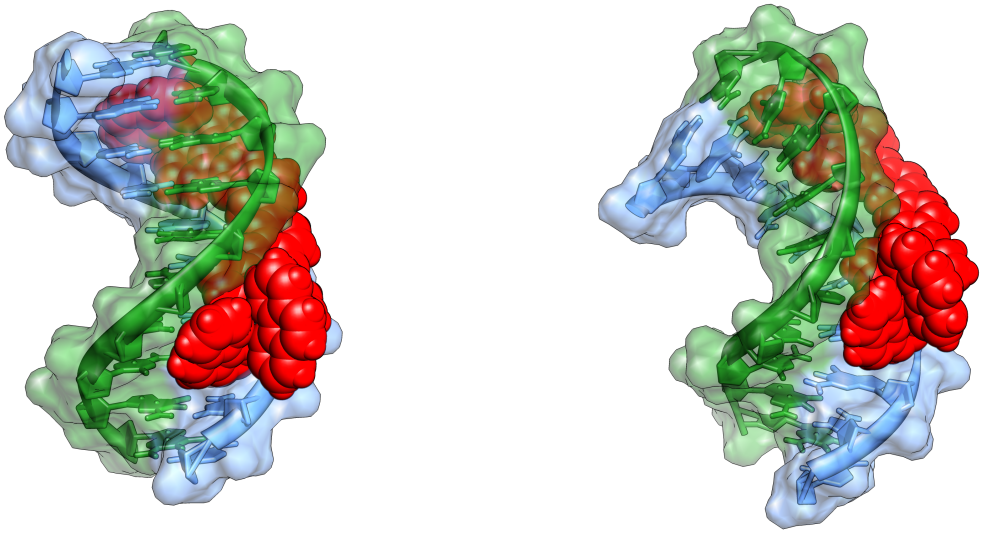


Figure 4.3. Left: starting conformation of binuclear Ru(II)-ΔΔ-P-DNA ($[GCATATATCG]_2$) complex. Right: the conformation of the complex after 10 ns MD simulation

obviously made the DNA duplex diverge away from its starting B-conformation (**Figure 4.3**). Although the mechanistic jigsaw puzzle pieces seem to fall nicely in position together, with some support from the MD simulations too, there is still a high degree of elusiveness as to how the mechanism of threading intercalation by the binuclear ruthenium complexes indeed happens. Despite considerable effort, not just of *Paper I*, the only firm facts are that the efficient threading intercalation requires at least one turn of the DNA helix and that the more hydrophobic but also spatially bigger Ru(II)-phenanthroline dimer threads 20 times faster into mixed sequences of calf thymus DNA than its less hydrophobic smaller bipyridine analogue. The research is going on…

PAPER II. PRECURSORY STATES BEFORE THREADING INTERCALATION – AN NMR STUDY

The value of a 3-D structure for the understanding of mechanistic details of any biological process is no doubt enormous. And so, stimulated by the hypothesis about the mechanism of threading intercalation (described in the previous chapter), an NMR study of DNA-Ru(II)-compound complex structure was initiated. The initial composition of the system included a DNA oligonucleotide $[d(CGCGAATTCGCG)]_2$ and a ΛΛ-P binuclear compound in proportions either 1:1 or 1:0.5, respectively. The central AT-step was chosen to be only four base pairs, positioned in an asymmetrical order. Despite the reported earlier results on the threshold length of the central AT-tract, which has to be at least ten base pairs long for treading intercalation to be effective, four A-T pairs seemed to be a good starting point. Such a short AT-stretch was chosen from the considerations of NMR's incapacity of differentiating between subsequent adenines or thymines in a longer AT-pattern and also because too long AT stretch would give a heterogeneous distribution of intercalation sites thereby obviating distinct NMR signals. However, as was figured out later, the shortness of the central AT-tract was only the first trade-off, further "concessions" were necessary to be made. The hydrophobic character of the phenanthroline-type binuclear compound at concentrations required for the NMR experiments (1 mM) resulted in a formation of aggregates and a subsequent precipitation of the sample. And thus the threading intercalator was substituted to the smaller and less hydrophobic ΛΛ-bipyridine (ΛΛ-B) compound (**Figure 2.8** (**b**) on page 17). A subsequent fluorescence and circular dichroism analysis of the DNA-ΛΛ-B mixtures unfortunately indicated no intercalation with $[d(CGCGAATTCGCG)]_2$ or with mixed calf thymus DNA, while the $poly(dAdT)_2$ sequences still exhibited threading.

The NMR experiments, as well as the optical spectroscopy, showed no signs of intercalation: high intensity chemical shifts of imino base-pair hydrogen bonds protons, the presence of which indicates the integrity of a base-pair, were recorded. If intercalation had occurred, this intensity would be much smaller due to the dramatic changes in the environment in the presence of ΛΛ-B. The connectivity between subsequent base pairs would be broken had any intercalation occurred.

Table 4.1. Observed and assigned NOEs used for structural calculations, DNA signals are all from strand β, H1′ denotes a hydrogen atom of a C1' carbon atom of a furanose ring; a and b denote opposite monomers of the ruthenium ΛΛ-B compound. Average distances from steered MD simulations shown for comparison.

NOE	Proton1	Proton2	Upper limit (Å)	Average distance (Å)
1	H1' β–G4	H4Ca	5.0	4.6
2	H1'β–A5	H4Ca	3.5	2.8
3	H1'β–G2	H4Cb	3.5	2.9

Despite the failure of producing an intercalative binding mode, NMR was able to diagnose two different outside-binding configurations based on the assignments of the DNA strands and interactions with protons on the Ru(II) dimer. Upon interaction between ΛΛ-B and DNA, the original symmetry in both molecules broke. Both of the monomer units of ΛΛ-B and two strongly asymmetrically perturbed DNA strands (labelled for convenience α and β) become individually visible. In addition a much less perturbed symmetrical third DNA strand (γ), assigned to be very similar or identical to one strand of a free DNA duplex, was identified. The most prominent feature of the latter DNA assignment was that the chemical shifts of the terminal base pairs were changing continuously (indicating fast exchange) upon adding increasing amounts of the ruthenium compound. This indicates a DNA-terminal binding which we denoted as the top-binding mode. The first mentioned assignment of DNA interaction with Ru(II) compound – two asymmetrical strands α and β – on the other hand showed no variations upon gradual addition of ΛΛ-B, indicating stronger binding and slow exchange. Due to a very crowded spectrum with much overlap, it was only possible to identify three intramolecular NOE resonances signifying geometric information about the interaction of the ruthenium compound with the DNA. The arrangement of these NOEs: NOE1 and NOE2 unambiguously place one of the ruthenium monomers close to the DNA centre while the other monomer has close contact with the DNA's termini

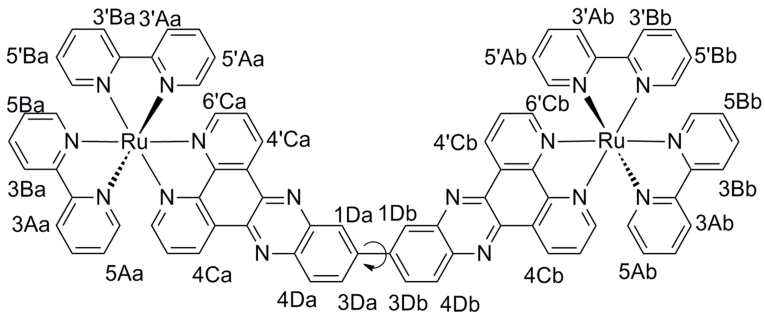


Figure 4.4. Atom nomenclature of the ruthenium ΛΛ-B compound, the compound is depicted in its *syn*-conformation.

(NOE3), also excludes intercalation. These NOEs are given in **Table 4.1**, atom nomenclature of the ΛΛ-B ruthenium compound is given in **Figure 4.4**.

Despite the extensive overlap of resonances from both the DNA and the ruthenium compound, allowing only few NOEs to be identified, their structural assessment seemed to be a great success in the virtue of restricted conformational flexibility of ΛΛ-B, which is limited to the rotation around a chemical bond connecting two ruthenium centres (**Figure 4.4**). However, the structural response of DNA, a highly polymorphic molecule, could not be assigned based upon these three NOEs. To assist NMR in the refinement of the DNA-ΛΛ-B complex structure, molecular dynamics (MD) was employed, using as a starting configuration a canonical B-DNA with ΛΛ-B in its *anti*-conformation fitted into the minor groove to satisfy NOE constraints. In a first attempt, a 20 ns unrestrained MD simulation revealed that the immediate response of the DNA was to expel the ruthenium

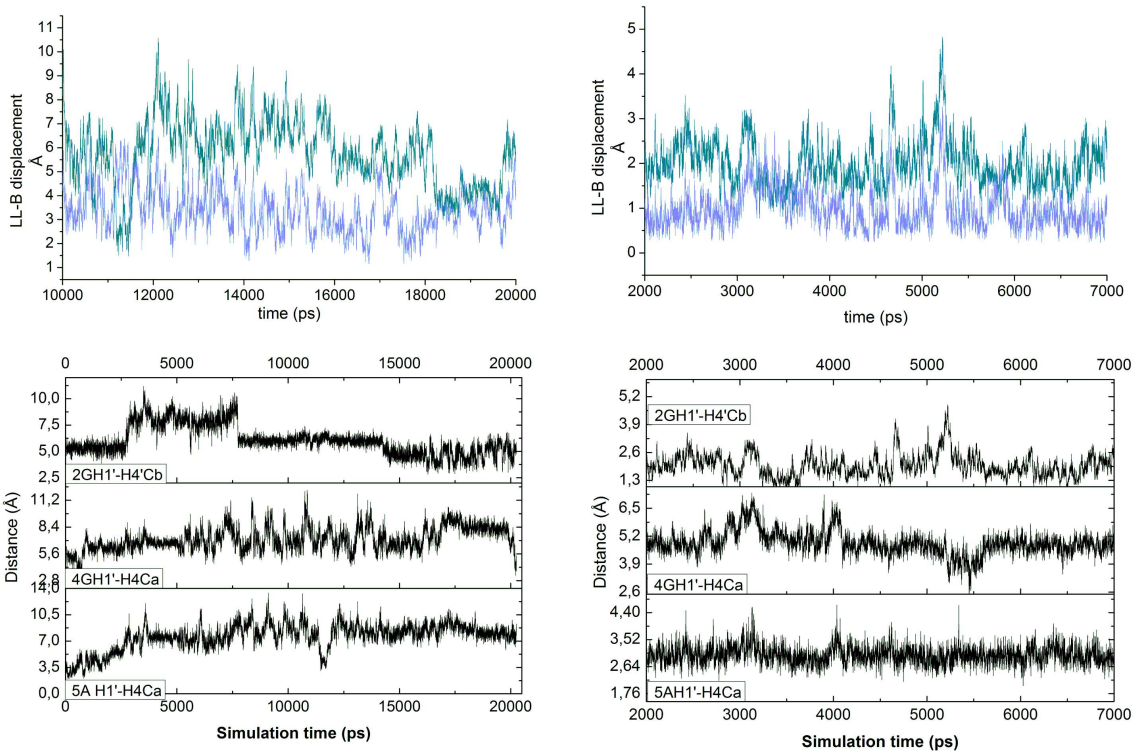


Figure 4.5. Averaged structural variations vs. simulation time for unrestrained MD (left) and steered MD (right) simulations. For the steered MD only the last unrestrained 5 ns are shown and considered. The bottom three panels show the time evolution for proton–proton distances corresponding to the measured NOEs: (b), (c), and (d) correspond to NOEs denoted 3, 1, and 2, respectively, in **Table 4.1**. Top panel shows the average displacement of heavy atoms of ΛΛ-B between the snapshot structures and the starting structure (green curve), and between the snapshot structures and the mean structure of the corresponding trajectory interval (purple curve). For these displacements, the DNA backbone of the ten inner base pairs has been superimposed, followed by calculation of differences of positions for the ΛΛ-B heavy atoms.

compound from the minor groove towards the DNA's termini (**Figure 4.5**), this observation seems to confirm the fast-exchange top-binding mode. However, the sliding of the ΛΛ-B complex off the DNA groove signifies a high interfacial tension between the two species, which has to be overcome in order to confirm also the second, groove-bound structure. With the suspicion that the starting conformation of ΛΛ-B was wrongly assigned, a quantum chemistry (QM) calculation aiming to scan the torsion angle of the intramonomer chemical bond was performed. The results from the QM run, however, confirmed the preferential *anti*-conformation with a torsion angle of 33° as well as reported the energy barrier for a transition from *syn*- to *anti*-conformation to be about 2.6 kcal/mol in vacuum.

And thus, it became clear that the DNA structure has to be forced to adjust itself to the presence of the binuclear ruthenium outside binder. To find stable configurations of DNA-ΛΛ-B complex that correspond to the minor-groove binding mode, and at the same time satisfy the intramolecular NOEs, further MD simulations were made (details are given in *Paper II*). With the help of the steered MD approach coupled to replica-exchange MD (REMD), for better sampling of conformational space, the validity and existence of the groove-bound mode structure was found, the structure is depicted in **Figure 4.6**. The details of the simulation protocols can be found in the supplementary information to *Paper II* and will not be presented here. In contrast to the dummy-ion approach used

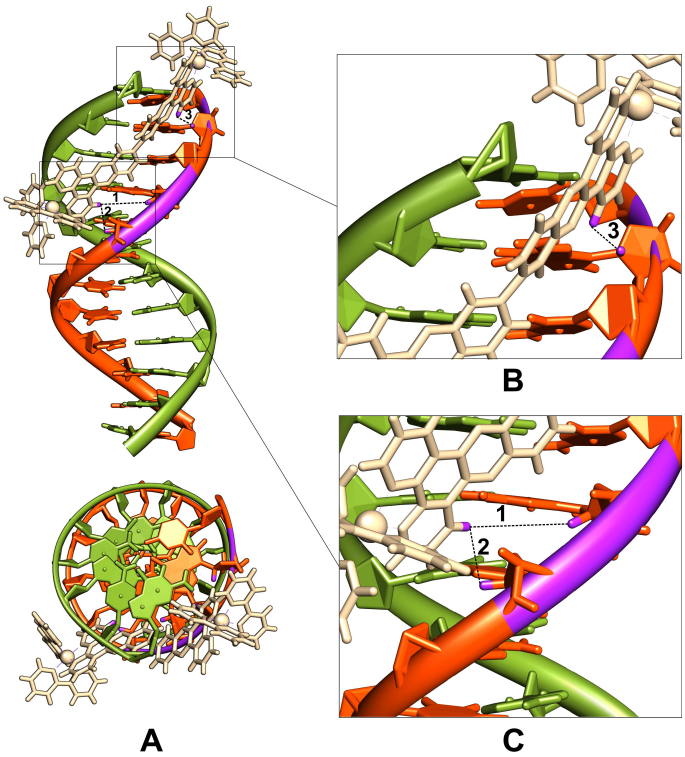


Figure 4.6. DNA–ΛΛ-B structure as determined from steered MD simulation. Panel A: Interactions between DNA and ruthenium complex ΛΛ-B, side view in upper picture, top-to-bottom view in lower panel. Panels B and C show distances 1, 2 and 3 corresponding to **Table 4.1**.

in *Paper I* for modelling the ruthenium(II) ion, in *Paper II* the metal and surrounding coordination sphere were properly parameterized.

To conclude, two outside binding modes (detected by NMR and refined by MD) have been identified: one fast exchange top-binding mode (which is probably irrelevant in the biological context of long DNA molecules) and one more stable groove-bound mode, of which the latter may represent the transient binding state prior to threading intercalation in such DNA structures where an array of alternating ATs will allow that binding mode. But, still, we are hunting for the threaded state and the research is to be continued…

PAPER III. STRUCTURAL VARIATIONS OF A DNA DUPLEX – EFFECTS OF INTERCALATIVE BINDING

In an attempt to discover some potentially-useful details helping to understand the mechanism of threading intercalation by binuclear $[Ru_2(phen)_4bidppz]^{4+}$, a structural study of binding associations of Ru(II) monomers, $[Ru(phen)_3]^{2+}$ and $[Ru(phen)_2dppz]^{2+}$, to DNA was initiated. In addition, the study aimed to resolve the long lasting discussion as to the details of binding geometries and preferentiality of Ru(II) monomers to reside in either of the DNA grooves.

The extensive in-house research of DNA interactions with Ru(II) polypyridyl compounds has provided a variety of experimental data as well as a number of new questions. One is how to understand the enantiospecific character of binding of the three–bladed propeller complex $[Ru(phen)_3]^{2+}$. A significant drop in the orientation factor, S=0.68, for right-handed Δ-$[Ru(phen)_3]^{2+}$, but not for left-handed Λ, with S=0.91, deduced from flow linear dichroism spectroscopy (Lincoln & Nordén, 1998), together with a reduced relative viscosity of DNA (Satyanarayana et al., 1992) and strongly different electrophoretic migration properties between the enantiomeric complexes with DNA (Gisselfält et al., 2000), were all indicative of a DNA kinking occurring specifically upon interaction with the Δ-enantiomer.

This enantiospecificity, however, became much less prominent when one of the three phenanthroline-propeller blades was replaced by the more elongated dipyrido-phenazine ligand, dppz. The $[Ru(phen)_2dppz]^{2+}$ compound binds to DNA by intercalating the large planar dppz moiety. This binding mode was confirmed by strong luminescence as a result of protection of the dppz aza-nitrogens from surrounding water when in DNA, as well as by LD spectroscopy that showed the long axis of dppz to be oriented roughly perpendicular to DNA helix axis (Jenkins et al., 1992; Holmlin et al., 1998; Lincoln & Nordén, 1998). Due to intercalative binding $[Ru(phen)_2dppz]^{2+}$ gains sufficient affinity ($K>10^6$ M^{-1} at 50 mM NaCl) in comparison to the "mother" compound $[Ru(phen)_3]^{2+}$ ($K\approx10^4$ M^{-1} at 50 mM NaCl), which, presumably, might obscure the delicacy of enantio-recognition of the dppz-based compound by chiral DNA.

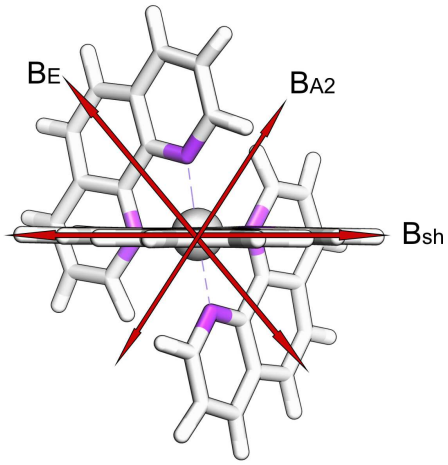


Figure 4.7. Three major electric dipole transition moment directions of $[Ru(phen)_2dppz]^{2+}$, deduced in (Lincoln et al., 1996).

Nonetheless, from the flow LD spectra of DNA complexes with $[Ru(phen)_2dppz]^{2+}$ it has been deduced yet another interesting detail, namely an asymmetry in the way the two enantiomers bind to the duplex. One of the electric dipole transition moments of the Ru(II) chromophore, called B_E (**Figure 4.7**), absorbing at 450 nm, was detected to give rise to a negative LD signal for Δ but positive for Λ. Here it is worth recalling that a zero-LD signal of an aligned sample corresponds to exactly 55° angle between the orientation of an electric dipole transition moment of a chromophore and the macroscopic orientation axis of the sample. When bound to DNA the B_E electronic transition moment of Δ-$[Ru(phen)_2dppz]^{2+}$ is oriented at an angle bigger than 55° to the DNA helix axis, while Λ-$[Ru(phen)_2dppz]^{2+}$ – with an angle smaller than 55° (for the LD spectra the reader is referred to (Lincoln & Nordén, 1998)). Simple structural estimates suggest that if the plane (short axis) of dppz had been perfectly perpendicular to the DNA helix axis, the B_E transition moment had been accidentally at exactly 55° and, thus, given zero LD at 450 nm. Deviations of the short axis of dppz from perpendicularity were calculated to be +11° for Δ (a clockwise "roll" of the dppz plane, or inclination of the dppz plane in terms of base pair nomenclature, see **Figure 2.3** on page 10 for reference) and +15° for Λ (also clockwise), these values obtained from the experimental LD spectra when flow-oriented in aqueous solution (Lincoln & Nordén, 1998). The conclusion was: the $[Ru(phen)_2dppz]^{2+}$ is bound in such a way that it is rotated around the dppz long axis at an angle of around 10-15° in the clockwise direction when viewed into the minor groove where the complexes were assumed to reside.

This remarkable deviation from the perpendicularity of the intercalating moiety orientation is interesting from two perspectives: first, it is in the same direction (clockwise) for both enantiomeric forms of the compound, and second, it could indicate a conformational transition in DNA. The latter was suspected due to the fact that the plane of an intercalating moiety is, in general, co-planar with the orientation of the DNA bases, and if the bases are inclined by as much as 10-15° this could indicate a deviation

from a characteristic B-DNA form (for which the bases are perpendicular) towards an A-form DNA (base pairs inclination of 19° in clock-wise direction when viewed from the minor groove!).

These two questions, namely the enantiospecific DNA kinking by $[Ru(phen)_3]^{2+}$ and the possible DNA conformational transition induced by the binding of $[Ru(phen)_2dppz]^{2+}$, were both addressed by means of molecular dynamics (MD) simulations. MD indeed confirmed that Δ-$[Ru(phen)_3]^{2+}$, but not Λ, induces DNA kinking upon interaction by introducing one of its phenanthrolines into the DNA stack in a wedge-like fashion, while DNA-Λ association resulted in a geometry closely resembling the one of classical intercalative binding with insignificant DNA bending. In the case of $[Ru(phen)_2dppz]^{2+}$ MD did not detect any conformational transition for either of the enantiomer-DNA complexes. Instead I suggest that the observed experimental inclination of the dppz plane be, mostly, an effect of a preferred DNA bending in a right-handed way when viewed into the minor groove, induced by the insertion of the dppz ligand into the base pairs stack, more prominent in the case of Λ-$Ru(phen)_2dppz]^{2+}$.

To perform an MD simulation, aiming for more subtle structural details, the system has to be fully parameterized (the reader is referred to the MD methods section of this Thesis for details). While such parameterization existed for DNA and proteins, or could be easily derived (with the help of available tools, e.g. Antechamber utility of AmberTools (Case et al., 2010) in association with generalized AMBER force fields, GAFF (Wang et al., 2004)) for an organic compound (consisting of H, C, N, O, P and S), a metal-organic compound, in particular Ru(II)-coordination centre, was lacking MD parameterization data. Luckily, after some data-mining it was found that the Ru(II)-coordination sphere in Ru(II)-polypyridyl compounds has been parameterized for molecular mechanics (MM3*) (Brandt et al., 1998). Due to some variations of potential energy functional form between Amber MD force field and MM3* force field (FF), the Ru(II)-coordination sphere parameters derived by Brandt et al. were tuned to be used with Amber FF according to the procedure described in (Norrby & Liljefors, 1998). Particular attention was paid to the calculations of RESP point charges (Bayly et al., 1993; Wang et al, 2000) of $[Ru(phen)_3]^{2+}$ and $[Ru(phen)_2dppz]^{2+}$. The calculations of RESP charges as well as the parameters derivation procedure required high level quantum mechanical (QM) calculations, details of which are briefly specified in the method section of this Thesis. All QM calculations were performed with Gaussian03 software package (Frisch et al., 2003). The next step was to use the developed force fields and to proceed to actual MD calculations.

As no experimental structures were available, the first step was to obtain a reasonable set of binding geometries of DNA complexes with $[Ru(phen)_3]^{2+}$ and $[Ru(phen)_2dppz]^{2+}$ for both enantiomeric forms of either of the compounds. While the binding mode of $[Ru(phen)_3]^{2+}$ was not completely resolved, groove binding vs. "quasi-intercalation" (Lincoln & Nordén, 1998; Gisselfält et al., 2000), in the case of $[Ru(phen)_2dppz]^{2+}$ the intercalative binding was experimentally confirmed (Friedman et al., 1990; Jenkins et al.,

1992; Hiort et al., 1993; Holmlin et al., 1998; Lincoln & Nordén, 1998). So I decided to start with the latter compound, probing both minor and major grooves, as well as varying the DNA sequence content and length, testing several 12 base pairs and 16 base pairs long DNA fragments in search for consistency.

Trying several simulation setups, including unrestrained and steered MD coupled to replica-exchange MD for better sampling of conformation space, starting from either a pre-intercalated position of the Ru(II) compound as well as fully separated states. The resulting geometries for Δ- and Λ-$[Ru(phen)_2dppz]^{2+}$ when placed in the major groove were concluded to be inconsistent with the experimental data, as the dppz ligand when protruding through the DNA stack brought the aza-nitrogens too much forward, which should expose them for a direct contact with water on the other side of the base stack, in conflict with the experimental strong luminescence enhancement. By contrast, when placed in the minor groove, both Δ- and Λ-$[Ru(phen)_2dppz]^{2+}$ had the dppz aza-nitrogen atoms positioned in the shielding hydrophobic environment of the DNA intercalation pocket, in this way confirming the preferential binding from the minor groove. The resulting geometries have also revealed an asymmetry as to the position of the Δ-enantiomer in the intercalation pocket: one of the phenanthroline wings was found to hydrophobically stack with a DNA backbone sugar, resulting in somewhat inclined orientation of the short axis of dppz relative to the planes of the surrounding DNA base pairs. This arrangement resulted in a somewhat skewed orientation of the dppz ligand relative to the dyad axis of DNA (**Figure 4.8**), as was earlier hypothesized by Barton and co-workers (Friedman et al., 1990; Jenkins et al., 1992).

However, none of the simulated structures of the DNA-$[Ru(phen)_2dppz]^{2+}$ complexes have shown any significant deviation from B-form DNA. Yet, the intercalative binding of both Δ- and Λ-$[Ru(phen)_2dppz]^{2+}$ induced a helical axis bending by about 20° for Δ and by about 30° for Λ towards the major groove (conformational analysis performed with Curves+ program (Lavery et al., 2009)). An analogous DNA bending was also computationally predicted for the case of intercalative binding of daunomycin (Mukherjee et al., 2008). The DNA bending had also an impact on the overall distribution of base pairs' inclination angle relative to the DNA helical axis. Due to chirality of the DNA structure, namely the major and minor grooves turning in a right-handed way, there is also an asymmetry in the way the helix is bending over towards the major groove. The bending, initiated by the presence of an intercalator, thus, results in a deviation from the perpendicularity of the base pairs. The bigger values of the helical bending induced by Λ-$[Ru(phen)_2dppz]^{2+}$ is due to that the compound tries to gain as much as possible hydrophobic stacking interactions with the DNA sugar rings of the intercalation pocket, towards the walls of the minor groove, thus protruding deeper into the DNA stack.

The computational experiments, when re-performed with longer (16-mer) DNA, revealed that bending spread over in total eight base pairs (four above and four below the intercalation site) and indeed had impact on the average base pair inclination measured

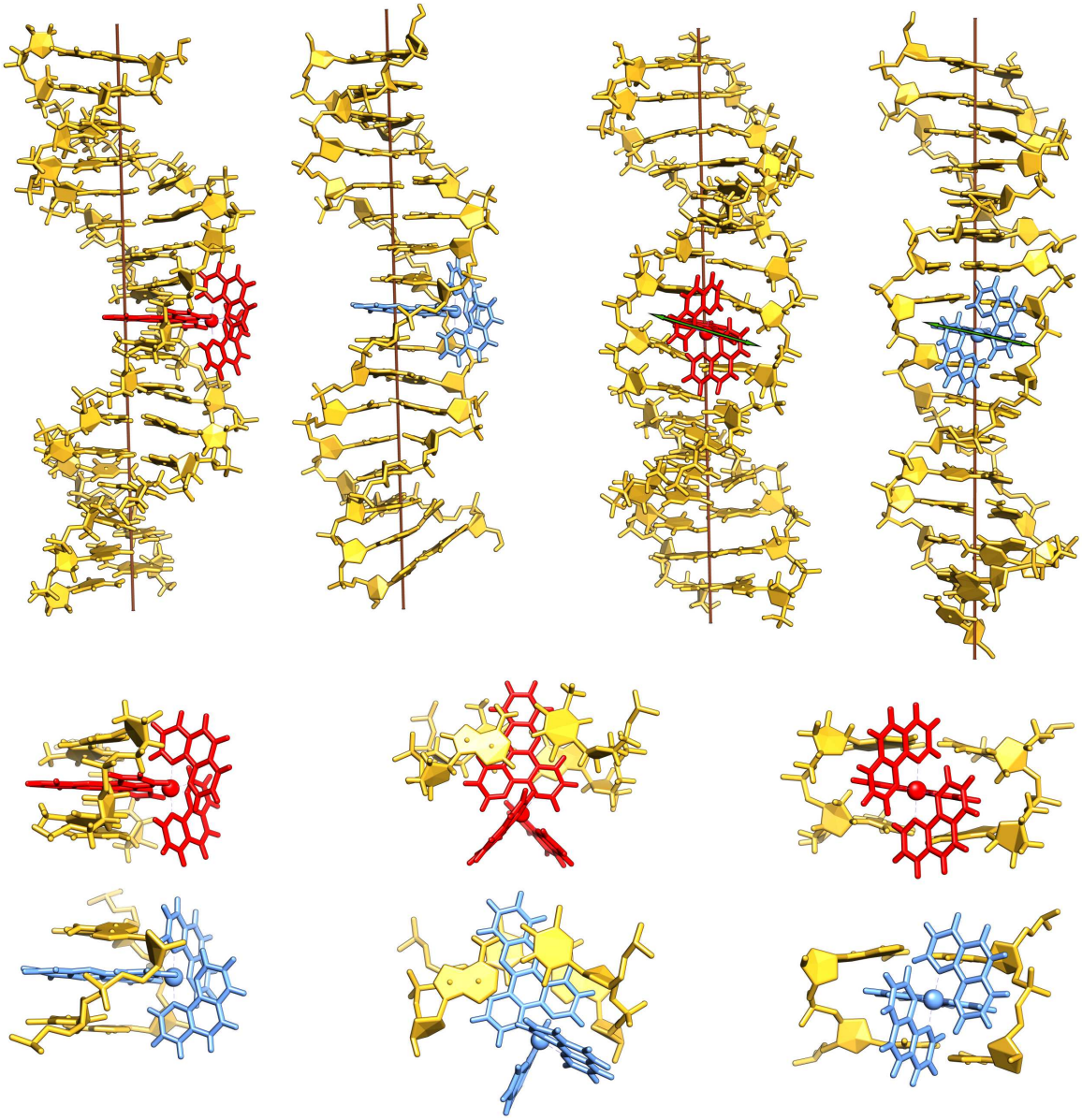


Figure 4.8. Top: binding geometries of the Δ- and Λ-enantiomers of $[Ru(phen)_2dppz]^{2+}$ with 16mer DNA. Left: Δ (blue) and Λ (red) enantiomers induce bending of the DNA helix, of about 20° for Δ and of about 30° Λ. The brown-coloured axis represents the macroscopic alignment axis of the sample in analogy with LD experiments. Green-coloured arrows show the orientation of the transition moment B_{sh} deviating from perpendicularity. Bottom: side, top and front zoom-in of the DNA intercalation site illustrating symmetric binding of the Λ-enantiomer and asymmetric binding of Δ.

relative to the DNA helix (which is the macroscopic axis in the LD experiments). And thus, the experimentally detected clock-wise base pair inclination, which we first thought could be due to a local conformational transition to an A-like form of DNA, is possibly instead the result of a chiral DNA bending induced by intercalation: making the DNA bend over towards the major groove and at the same time turning to the right when viewed into the minor groove. To test this hypothesis we studied the LD spectra of

oligonucleotides confined and aligned in stretched matrices of humid PVA, a medium in which DNA has been confirmed to exist in a B-form like conformation at 100% relative humidity. In this case the LD spectra of opposite enantiomers were found indeed to be very similar and coalesce in the region of the B_E transition, thus, indicating that the transition moment is now effectively at 55° and, thus, the dppz short axis perpendicular to the helix axis: no significant base pair inclination! (Reymer, Hanczyc, Lincoln and Nordén, to be published). The difference between flow orientation and stretched PVA matrix orientation LD spectra could thus be that while in the former the dynamics of DNA when free in solution could enhance bending (and base inclination), while when stretched and confined in PVA matrix this effect is suppressed.

Analogous computational experiments have been performed for the $[Ru(phen)_3]^{2+}$ compound. Testing the compound's binding from the major groove resulted in a complete fiasco. Even when the compound was placed in a "pre-made" intercalation pocket, neither enantiomeric forms stayed there, dissociation was instant. By contrast, the test of the minor groove binding was more successful. Both Δ- and Λ-$[Ru(phen)_3]^{2+}$ could form stable complexes with DNA. The MD experiments demonstrated that the Δ-form of the compound indeed induces a substantial kink of the DNA helix (of 53°), while the Λ-form rather resembles a typical intercalator initiating only an inconspicuous bending of 16° (conformational analysis performed with Curves+ program (Lavery et al., 2009)). The resulting structures (**Figure 4.9**) suggest that the kink produced by the Δ-enantiomer is partially due to a wedge effect, separating the base pairs, but also involves, analogously to the above described Δ-$[Ru(phen)_2dppz]^{2+}$, an asymmetric hydrophobic stacking with one of the sugars along the DNA minor groove. The Λ-enantiomer, by contrast, due to hydrophobic stabilization stacking of the two non-intercalated phenanthroline wings with the DNA sugars above and below the intercalation spot, is positioned rather symmetrically with respect to the groove opening and is inserted deeper into the DNA stack. The described scenario of the enantiomeric binding selectivity of Δ- and Λ-$[Ru(phen)_3]^{2+}$ not only explains the observed experimental variations, but is also interesting in the context of evolutionary development of nucleic acid interaction with operative proteins, such as the eukaryotic transcription factor TBP (TATA-box binding protein) (Kim et al., 1993a; Kim et al., 1993b; Nikolov et al., 1996) or the high mobility group proteins (HMGP) (Klass et al., 2003; Palasingam et al., 2009).

A recent crystallographic study by Hall et al. (Hall et al., 2011) on DNA complex with Λ-$[Ru(tetraazaphenanthrene)_2dppz]^{2+}$, a compound almost identical to $[Ru(phen)_2dppz]^{2+}$, has revealed two intercalative binding modes: a typical intercalation of the elongated dppz ring system and a "semi-intercalation" of a shorter tap ligand, both modes represent binding from the minor groove (!), the latter giving rise to a DNA helix kink of 51°. The first binding mode, typical intercalation, shows a skewed orientation of the Λ-$[Ru(tap)_2dppz]^{2+}$ compound in the intercalation site, in contrast to our simulation results which show that it is Δ, but not Λ, that is asymmetrically placed in the intercalation

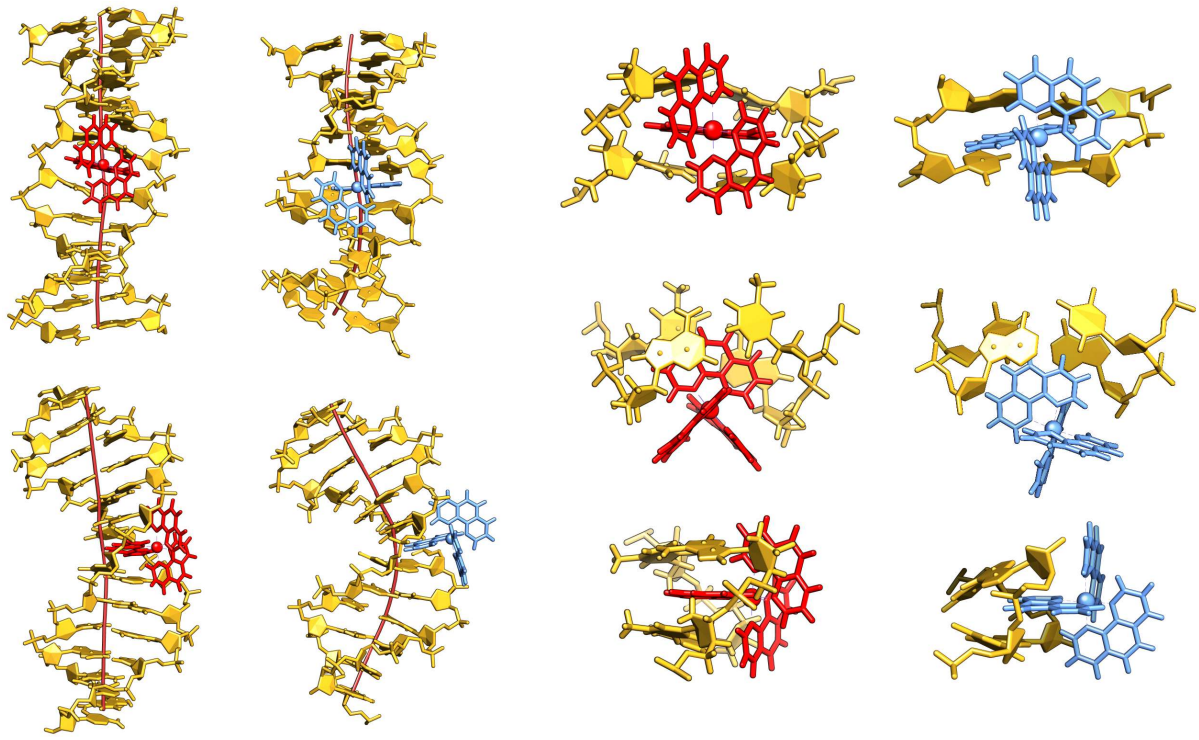


Figure 4.9. Binding geometries of the Δ- and Λ-enantiomers of $[Ru(phen)_3]^{2+}$ with 12mer DNA. Left: the Δ-enantiomer (blue) induces a 53° kink indicated by the helical axis (brown); by contrast, the Λ-enantiomer (red) bends DNA only inconspicuously (16°). The brown-coloured axis represents the helical axis of DNA. Right: front, top and side zoom-in of DNA intercalation spot illustrating the symmetric, resembling typical intercalation, binding of Λ and the asymmetric, wedge like, insertion of Δ.

pocket. The kinking of DNA observed in the crystal is produced by Λ, which contradicts both earlier experimental and computational results, which clearly indicate that it is only the Δ enantiomer that gives rise to such a perturbation of a DNA structure. This discrepancy may be due to an artefact of crystal packing, as the binding mode is detected at the interface between two DNA molecules.

The experimental binding free energies of the Δ- and Λ-enantiomers of $[Ru(phen)_3]^{2+}$ to DNA have been reported to be almost identical: -4.4 kcal/mol and -4.5 kcal/mol, respectively (Satyanarayana et al., 1992) with small variations with DNA base content and ionic strength (Hiort et al., 1990). Whereas the resulting computed structures vary significantly it was also interesting to identify which type of intramolecular interaction dominates in either case. The association free energies of Δ- and Λ-$[Ru(phen)_3]^2$ to DNA were evaluated with the MM-PBSA approach (Kollman et al., 2000; Kuhn & Kollman, 2000), which breaks down the energies into terms that implicitly describe electrostatic interactions; hydrophobic, solvation and dispersive interactions; and entropic contribution. However, the association energies produced by this method have been subject to debate and criticism for inaccuracy and high-dependence on the input parameters (Gohlke & Case, 2004; Kuhn et al., 2005; Singh & Warshel, 2010), yet some

interesting variations as to the distribution of the energetic terms between the enantiomers were noted. Smaller values of electrostatic and hydrophobic terms for the Δ-enantiomer correlate with its asymmetric orientation in the binding pocket, since Δ has only one of the non-intercalating phenanthrolines electrostatically interacting with the DNA backbone. Hydrophobic stacking of this phen-ligand with the furanose ring above the intercalation pocket, together with the partial wedge-intercalation of the other phenanthroline blade, obviously contribute less to the hydrophobic energy term than the three phen-ligands of Λ. There, one phen-ligand is intercalating and the other two are stacking with the sugars towards the walls of the minor groove, the latter arrangements are also giving rise to a bigger, in absolute value, electrostatic term. The entropic contribution to the association free energies was estimated with the normal-mode analysis and resulted in identical values for the enantiomers, which initially seemed to be strange. However, this is maybe not surprising, since the normal mode analysis, based upon vibrational modes of the system's component, will result in identical values for identical systems. And indeed these systems, comprising a DNA fragment and either of the enantiomers different only with respect to chirality of the ligands, are identical.

The total computed association free energies for Δ and Λ are different (!), with the lower value for Λ. These values, contradicting the experimental findings, might reflect that the entropic term was indeed much underestimated, suggesting that these systems might be an example where the entropy-enthalpy compensation principle is in action (Lee & Graziano, 1996; Cooper et al., 2001). This consideration is in agreement with the general finding of extensive entropy-enthalpy compensation in systems involving hydrophobic interaction.

To conclude, unlike the associative interactions of binuclear ruthenium compounds with DNA duplexes, where everything seems to be entangled and quite confusing, mono-nuclear ruthenium-DNA interactions are much easier to comprehend. The particular finding that the Δ-tris(phenanthroline)Ru(II) bends DNA by interfering with the duplex in a wedge-like fashion nicely correlates with the earlier experimental findings. Hydrophobic stacking interactions, which are obviously the reason of the asymmetric orientation of Δ and symmetric orientation of Λ, might be of particular importance in the initiation of the base pair opening in the case of threading intercalation by binuclear ruthenium compounds.

PAPER IV. MODEL STRUCTURE OF HUMAN RECOMBINASE RAD51

Guided by a successful study of structural variations of RecA filamentous complexes formed with ss- as well as dsDNA (Frykholm et al., 2006), an analogous project about human recombinase Rad51 was initiated. This new project, however, appeared to be a real challenge in many aspects…

In the preceding RecA study the structural information, namely the angular orientations of seven tyrosine residues collected by Site-Specific Linear Dichroism spectroscopy (hereafter SSLD), was utilized. Here it is worth reminding that LD only provides averaged angles, at which an electric dipole transition moment (absorbing at a particular wavelength) of all chromophores in a system is oscillating relative to a macroscopic orientation axis. Or, to be fully correct, average cosine square values of angles. This means generally 1) that angles with opposite signs are two possible solutions, and 2) the average will depend on the orientation distribution (which is unknown except for cases of perfect orientation such as when $\langle\cos^2(\alpha)\rangle = 0$ or 1).

Very conveniently, in the case of filamentous aggregates of recombination proteins, the macroscopic orientation axis of the sample coincides with the filament's axis. The main advantage of the SSLD method, used in the study, over classic LD is that the angular orientation is no longer averaged over all chromophores of the system but instead represent the orientation of a specific chromophore, namely a particular preselected tyrosine residue. Using SSDL data and facilitated by the presence of a high-resolution crystal structure of a RecA monomer, assuming that the monomer's conformation does not change along the course of the homologous recombination reaction, filamentous models for ss- and dsDNA-RecA complexes were assembled. Conclusions, drawn from these models, were later confirmed by the crystal structures of the DNA-RecA filaments (Chen et al., 2008), demonstrating the usefulness of the approach.

Moving over to the human Rad51 (HsRad51) project. In contrast to the RecA case, where the filamentous models were obtained by a straightforward rotation of monomers so that the angular orientation of the tyrosine residues would coincide with the SSLD data, the HsRad51 study was hampered by the lack of a full-length (339 a.a.r.) 3-D structure of the monomer. The HsRad51 monomer structure was only fragmentally available: including a crystal structure of the fusion complex of the central ATP-binding domain with the BRCA2 protein (PDB ID: 1N0W) (Pellegrini et al., 2002) and an NMR structure of the N-terminal domain (PDB ID: 1B22) (Aihara et al., 1999). These two structures, unfortunately, do not overlap, 1N0W, residues 99-339, and 1B22, residues 16-85. The missing residues, 86-98, correspond to the crucial part of the structure – the polymerization motif – a flexible joint connecting two domain regions, which is also responsible for the filament assembly. In addition, in the ATP-binding domain structure two putative DNA-binding loops and some other protein fragments were missing. So the first challenge was to construct a full-length model of the HsRad51 protein monomer.

The sequence alignment, performed with ClustalW (Larkin et al., 2007), of those HsRa51 homologues for which high-resolution structures were available in Protein Data Bank, including human DMC1 (Kinebuchi et al., 2004; Hikiba et al., 2008), yeast Rad51 (Conway et al., 2004), and archaea RadA (Shin et al., 2003; Wu et al., 2004; Qian et al., 2005), indicated the highest level of sequential identity between yeast and human Rad51. Therefore, yeast Rad51 (ScRad51, PDB ID: 1SZP) was chosen for homology modelling, acting as a structural scaffold in which the domain structures of HsRad51

were merged together. Particular attention was paid to the modelling of the two putative DNA binding loops, which were also missing in the ScRad51 structure. The longer L2 loop (residues, 268-292) was modelled according to the protein folding "recipe" provided by Simmerling and co-workers (Simmerling et al., 2002) while the configuration for the shorter L1 loop was assigned with the help of MODELLER (Fiser et al., 2000). The resulted monomer's model was then subjected to structural refinement with MD (details of simulating protocols are given in the methods section of *Paper IV*), and afterwards was exploited for the filament construction.

To get the SSLD data, to be used for the HsRad51 filament construction in analogy to the RecA study, was another challenge. The preparatory work, including site-directed mutagenesis for single amino acids substitution, protein expression and purification had to be made from scratch. Out of 10 tyrosine residues that were initially chosen as targets for SSLD, the two mutants Y191F and Y301F were excluded from the study due to insufficient purification levels. The other eight mutants together with the wild-type protein in complex with double-stranded calf thymus DNA and ATP, as a co-factor, were subjected to flow LD measurements, resulting angular orientations data listed in **Table 1** of *Paper IV*. Unfortunately the corresponding angular data for the HsRad51-ssDNA complex was not obtained. The protein-DNA interaction affinity, probably, was too weak, which resulted in a low intensity of the LD signal and inability to perform a

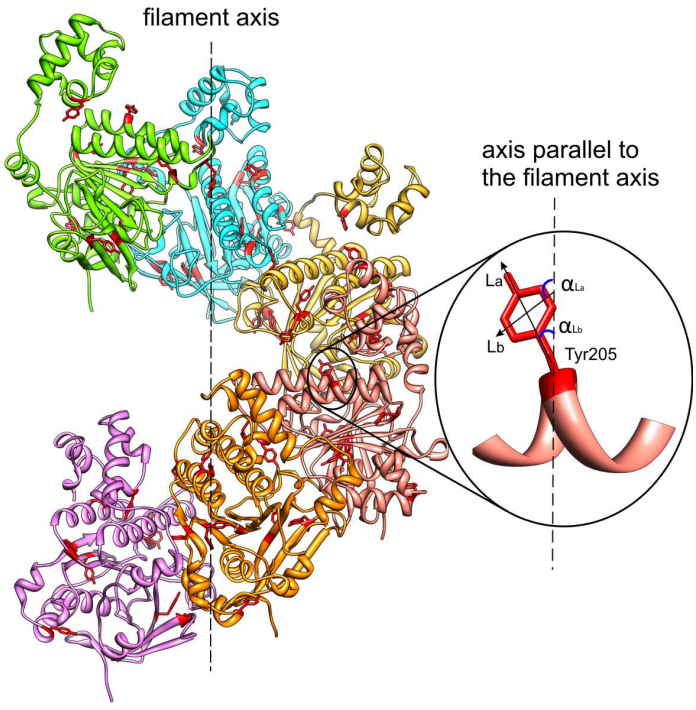


Figure 4.10. A model of the human Rad51 filament, constructed from a monomer structure based on known ATP-binding and N-terminal domain structures of HsRad51 and using ScRad51 as a homologues scaffold. Zoomed: angular orientations of the tyrosine electric dipole transition moment L_a and L_b relative to the filament axis.

qualitative analysis.

Therefore, at that stage of the study both the SSLD data and a "Frankenstein monster" monomer model, assembled from various pieces, were ready to be processed for the filament structure assembly. However, the two orthogonal angular orientations of each of the tyrosine residues are not enough to define the orientation of the monomer body in a 3-D Euclidean space: to describe such an orientation three parameters are required. So, the third challenge was to data-mine a parameter or a set of parameters that unambiguously would define the 3rd dimension. The data found included the helical parameters of the filament, namely a helical pitch of 99 Å that comprises 6.39 monomers (Yu et al., 2001; Galkin et al., 2005). They also included information from the tryptophan-scanning mutagenesis about the strategically positioned residues, for example: Phe126 and Phe129 of one monomer and His294 of an adjacent monomer were reported to be located close to the ATP-binding site on the interface between two monomers (Selmane et al., 2004; Renodon-Cornière et al., 2008), and that Tyr232 of the L1 loop is essential for ssDNA binding and thus has to be at the interior of the filament construct (Matsuo et al., 2006; Prasad et al., 2006).

The proceeding filament assembly consisted of three steps. Step 1, in which the SSLD data was utilized to define pitch and roll angles (Euler's angels nomenclature) of the monomer body with the help of a functional minimization procedure (details are given in *Paper IV*). Step 2, where the collected mutation analysis data was used as a landmark to define the 3rd angle. Step 3, in which, using the helical parameters, the model of the HsRad51 filament was finally accomplished. The assembled filament model was also subjected to structural refinement with MD. The final structure of the HsRad51 filament model is shown in **Figure 4.10**, where also the angular orientations of the tyrosine

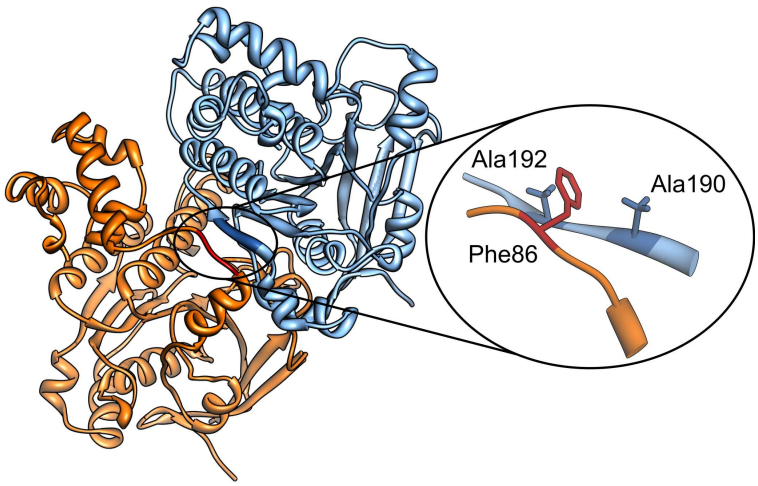


Figure 4.11. Two monomers of Rad51. Oligomerization of HsRad51 is coordinated by the β-strand region of inter-domain linker of one monomer together with the β-strand region of the adjacent monomer, and follows a "key-lock" principle: Phe86 is sandwiched between Ala190 and Ala192.

electric dipole transition moments are illustrated.

The filament model highlighted several interesting structural features. To start with, the polymerization of a filament is assured by a typical "key-lock" binding principle, illustrated in **Figure 4.11**, the Phe86 residue of one monomer is sandwiched between two alanine residues, Ala190 and Ala192, of an adjacent monomer. To continue, at the monomer-monomer interface earlier mentioned, Phe126, Phe129 and His294 are projected into a cavity, which is well suited for an ATP-binding pocket, though a co-factor was not explicitly included in the modelling. The putative ATP-binding site is in close proximity to the L2 loop and the overall configuration of this region is very similar to the one observed in the crystal structures of the RecA-DNA complexes (Chen et al., 2008). And finally, the structure of a double-stranded DNA molecule, adopted from the 3CMW structure of the RecA-dsDNA complex, was fitted inside the filament construct: encouragingly enough, the putative DNA-binding loops, L1 and L2, are both found to be strategically located enabling direct contacts with DNA. The model suggested that the L1 loop interferes with the DNA stack via insertion of the Tyr232 residue between two DNA bases, possibly acting as an anchor during the filament assembly, and electrostatic interaction of positively charged Arg235 with the negatively charged DNA backbone. Proposed protein-DNA interactions are depicted in **Figure 4.12**.

The model of the HsRad51 protein filament, together with the dsDNA structure fitted into it, was co-aligned with a 3-dimensional reconstruction of an electron micrograph of HsRad51-DNA filament (Galkin et al., 2005). The good agreement between the two,

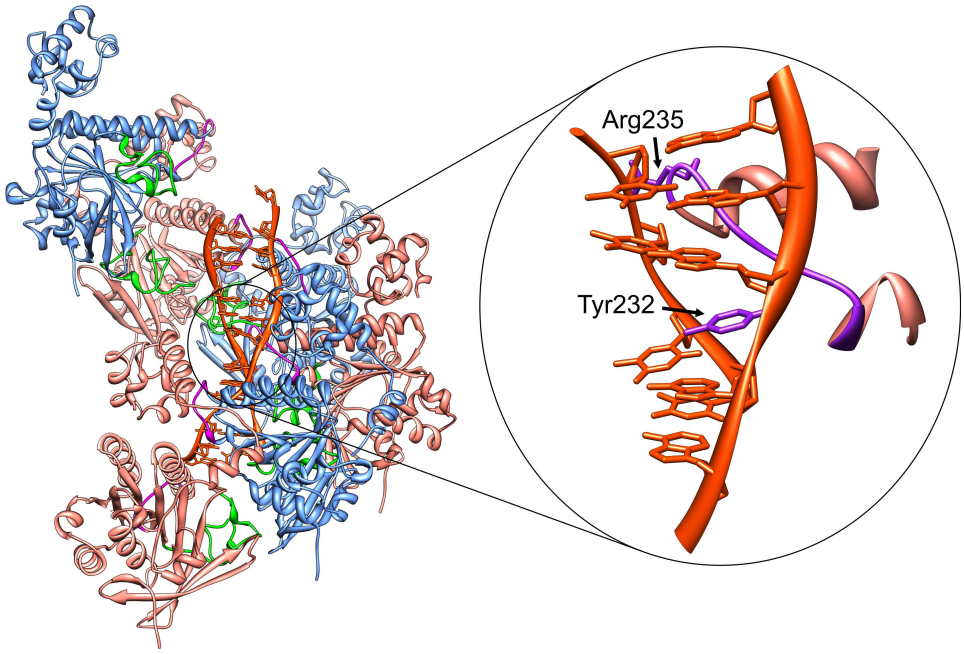


Figure 4.12. Rad51-DNA binding. Both loops, L1 (magenta) and L2 (green), are facing the interior of the filament, which allows the housing of a dsDNA molecule. The zoomed-in part illustrates the interactions of the L1 loop with DNA, suggested by the model.

illustrated in **Figure 5** of *Paper IV*, verifies the quality of the model and justifies the assumptions made.

To conclude, despite all challenges and missing structure data, a self-consistent model of the HsRa51 filament was successfully constructed. In addition to interesting highlights of the overall structure of the fibrous complex, the model also opens opportunities for further investigations of mechanisms of the homologous recombination reaction performed by human recombinase Rad51, for example, as pursued in Paper V.

PAPER V. Ca^{2+} STIMULATES DNA STRANDS EXCHANGE BY HUMAN RAD51 – MOLECULAR MECHANISM

The human recombinase, HsRad51, apart from its major role of catalysing DNA strand exchange, also possesses a moderate ATPase activity, which was reported to amplify in the presence of Mg^{2+} and decay in the presence of Ca^{2+} (Bugreev & Mazin, 2004). However, the LD experiments, performed in *Paper IV*, studying the filamentous complexes of HsRad51-dsDNA-ATP-Mg^{2+} have not detected any structural variations on the time frame of several hours. This observation initiated a question: if ATP was hydrolysed, why no variations were detected in the LD spectra, or otherwise, why does HsRad51 in the complex with double-stranded DNA not exhibit any ATPase activity despite that the ATP was accompanied by Mg^{2+}? The subsequent LD experiments of HsRad51-dsDNA recorded with ATP+Ca^{2+} as well as with ADP+Ca^{2+} resulted in variations of the spectral intensity but not in the shape of the LD signal. With ATP+Ca^{2+} the signal was the strongest, almost twice the value than that with ATP+Mg^{2+}, while with ADP+Ca^{2+} the signal was the smallest. These intensity variations are presumably coupled to the stiffness of the nucleoprotein aggregates: the stronger the signal is – the more regular and more rigid the filament is. The identity of the spectral shapes, on the other hand, signified that the structures of HsRad51-dsDNA were very similar regardless of the type of a nucleotide co-factor and a counterion.

Analogous LD experiments for the complexes of HsRad51-ssDNA, with ATP+Mg^{2+}, ATP+Ca^{2+} and ADP+Ca^{2+}, showed much greater variations both in spectral intensities and signal profiles. First of all, in comparison to the LD spectra of HsRad51-dsDNA, the LD measurements of the HsRad51-ssDNA complexes confirmed the overall much more flexible character of the system (indeed an ssDNA molecule is much more flexible than dsDNA). And secondly, and even more interestingly, the two spectral sets: the ones with dsDNA and the others with ssDNA showed considerable differences in their spectral profiles as response to variation of nucleotide co-factor and divalent cation. While the ds- and ssDNA-HsRad51-ATP-Ca^{2+} gave rise to reasonably similar shapes of LD spectra, though the shoulder around 230nm (indicator of a well-aligned protein) is much more prominent with dsDNA, the other two combinations, namely those with Mg^{2+}, are barely comparable. In either of the spectra of HsRad51-ssDNA-ATP-Mg^{2+} or -ADP-Mg^{2+} the 230nm-shoulder is hardly noticeable. Otherwise, the spectral shapes are similar, though

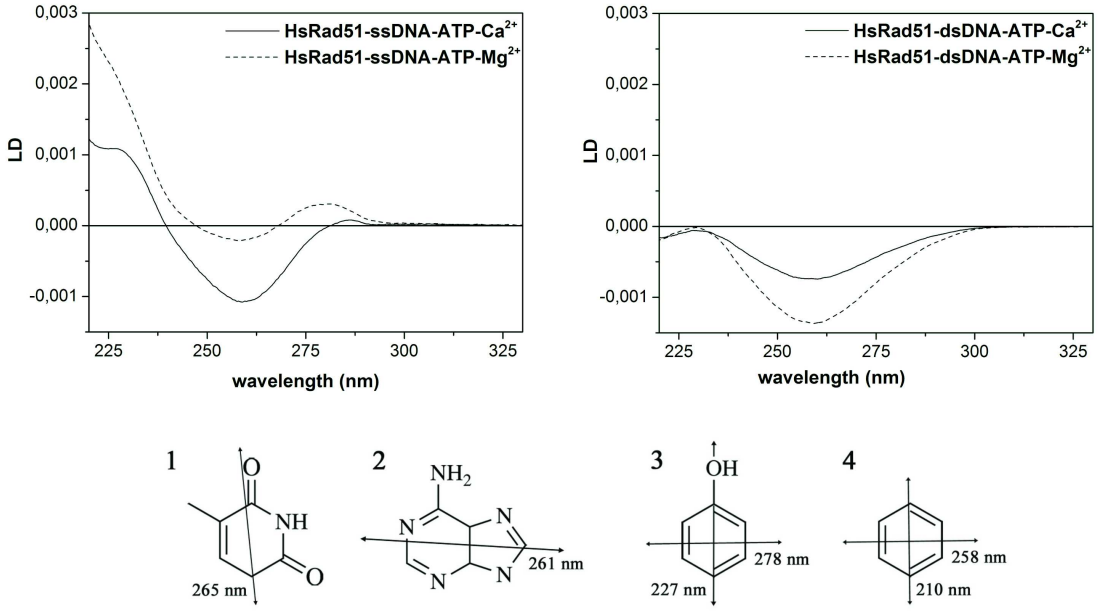


Figure 4.13. Top: flow LD spectra of HsRad51-DNA-ATP complexes: comparison between Ca^{2+} and Mg^{2+} conditions. LD spectra were measured after two hours incubation at room temperature. Bottom: transition moments responsible for LD signals at indicated wavelengths: thymine (1), adenine (2), tyrosine (3) and phenylalanine (4), reported in (Nordén et al., 1992; Holmén et al., 1997) for the DNA bases, and in (Hooker & Schellman, 1970) for the protein residues.

the one with ATP showed 2-fold higher intensity at 280 nm. The LD spectra of the HsRad51-ssDNA/dsDNA-ATP-Mg^{2+}/Ca^{2+} are collected in **Figure 4.13**.

In all spectra recorded for HsRad51-dsDNA the LD signal around 260 nm (a wavelength at which the DNA bases absorb) is distinct and negative, which is indicative of well-ordered DNA bases, perpendicularly oriented relative to the filament axis. However, in the case of HsRad51-ssDNA a corresponding conclusion was not possible – the dip at 260 nm was just a miniature of the normal DNA spectrum. This could signify either a very low level of orientation of the protein-DNA sample or that the DNA bases have no preferential orientation inside the filament. The LD spectrum showed relatively strong amplitudes at 230-280 nm, wavelengths at which both protein and DNA chromophores absorb, indicating aligned filament and excluding the former reason. Neither ssDNA nor monomeric HsRad51 alone would give rise to any LD signal. In addition, the ATP chromophore, which also absorbs at 260 nm could also contribute to the observed "260 nm"-dip. Due to spectral overlaps of DNA and ATP it was decided to repeat the HsRad51-ssDNA experiments with a DNA analogue, etheno-DNA (chemical structure is given in **Figure 4.14**) that has different spectral properties. The etheno-DNA (poly(dεA)), used in the experiments, has an extra electronic transition absorbing outside the common absorption area of DNA and protein, namely at 310 nm, which makes it possible to identify in the LD spectrum. In **Figure 4.14** the two spectra

corresponding to HsRad51-poly(dεA)-ATP-Mg^{2+}/Ca^{2+}complexes clearly show that with Ca^{2+} the DNA bases are rather perpendicularly oriented, while with Mg^{2+} they are totally disordered. The HsRad51 strand exchange activity also varies with two ionic conditions, Mg^{2+} vs. Ca^{2+}: Ca^{2+} was shown to exhibit a stimulatory effect on the strand exchange activity of the human recombinase but not Mg^{2+} (Bugreev & Mazin, 2004). In addition, Ca^{2+} was reported to stiffen the filaments of an archaea recombinase, RadA, while Mg^{2+} was deficient there too (Qian et al., 2006). To exclude that ATP-hydrolysis had any significant impact on the results, the LD experiments were re-performed on samples in presence of an ATP-regeneration system, resulting in almost identical spectra and, thus, the conclusion that ATP hydrolysis is not a reason for the observed structural variations.

To assist mechanistic understanding of the stimulatory effect of Ca^{2+} on the strand exchange activity of HsRad51, as well as to explain the observed variations of the LD spectra, the model structure of human recombinase filament, obtained in *Paper IV*, was recalled. As neither DNA- nor ATP-binding was explicitly modelled therein, the structure needed further refinement. Carefully checking the crystals of homologue proteins, human DMC1 (Kinebuchi et al., 2004; Hikiba et al., 2008), yeast Rad51 (ScRad51) (Conway et al., 2004; Chen et al., 2010), *Methanococcus voltae* RadA (MvRadA) (Shin et al., 2003; Wu et al., 2004; Wu et al., 2005; Qian et al., 2005; Qian 2006; Li et al., 2009) and RecA (Chen et al., 2008), I noticed that a C-terminus of the L2 loop has an α-helical conformation in almost all of the mentioned structures, but MvRadA also formed with Mg^{2+} where the L2 loop region was absent completely. This L2 loop, as was earlier discussed, is involved in DNA binding but also "hosts" ATP. Quite surprisingly, neither DMC1 nor ScRad51 has any co-factor in their crystals. On the other hand, the MvRad structures, both containing Mg^{2+} and Ca^{2+}, include AMP-PNP (a non-hydrolysable ATP analogue). In the Ca^{2+} case of the MvRadA structures there are two counterions present:

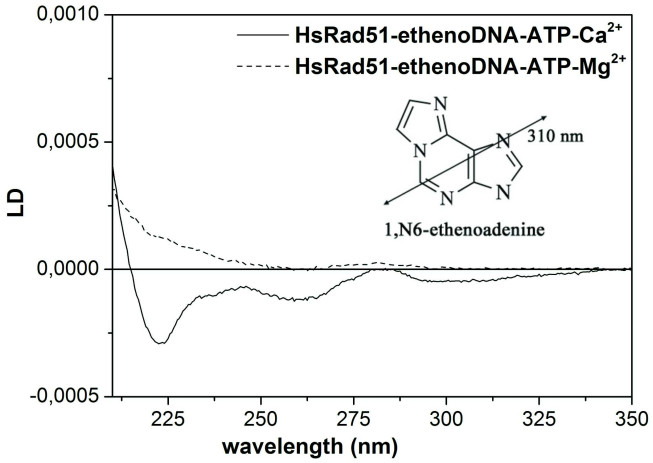


Figure 4.14. LD spectra of HsRad51-poly(dεA)-ATP-Mg^{2+}/Ca^{2+}complexes. Indicated in the figure transition moment of etheno-DNA responsible for LD signal at 310 nm is reported in (Holmén et al., 1994).

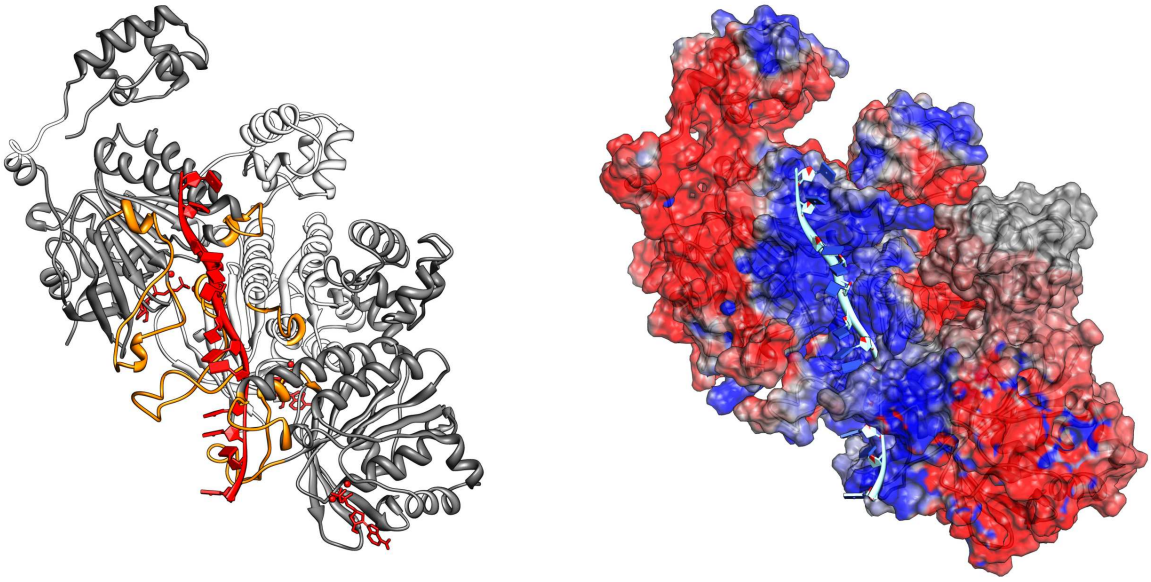


Figure 4.15. Two putative DNA binding loops, L1 and L2 (orange), form a positively charged (blue, right) channel on the inner surface of the protein filament, providing a tight binding pocket for DNA. Left: The overall geometry of the HsRad51-ssDNA-ATP-Mg^{2+}/Ca^{2} complex. Right: The surface electrostatic potential map of HsRad51, positive electrostatic potential (blue), red represents negative and white is neutral electrostatic potential.

one accompanying AMP-PNP and one at the C-terminus of the L2 loop, coordinating carboxyl oxygens of the protein backbone. Moving further, in the case of RecA the α-helical configuration of the L2 loop is observed only when DNA is in-bound, the ATP analogue – ADP+AlF_4 (artificial transition state analogue), accompanied by a single Mg^{2+} cation is also there. Further sequential comparison revealed why in either of DMC1, RecA and ScRad51 there is no need for a cation at the L2 loop C-terminus – the α-helical configuration is supported by the positively charged lysine residues projecting exactly towards the point where MvRadA has its 2^{nd} Ca^{2+} cation.

To understand why Ca^{2+}, but not Mg^{2+}, initiates this specific L2 loop configuration and how this echoes on the overall stiffness of the filament, the L2 loop region of the HsRad51 model was reconstructed analogously to the one observed in MvRadA (the L2 loop sequences of the two are nearly identical). Subsequent search for the DNA binding pocket on the interior of the filament with the help of electrostatic potential surface analysis revealed a well-defined positively charged tunnel located between the two putative DNA-binding loops (**Figure 4.15**), there the ssDNA, adopted from the 3CMU crystal structure of RecA (Chen et al., 2008) was positioned. When the new refined filament model, including ssDNA and ATP, was finalized, a set of MD simulations varying the cationic conditions, and co-factors, as well as presence of DNA inside the filament was initiated (the details of simulating protocols as well as system setup configurations are given in *Paper IV,* methods section).

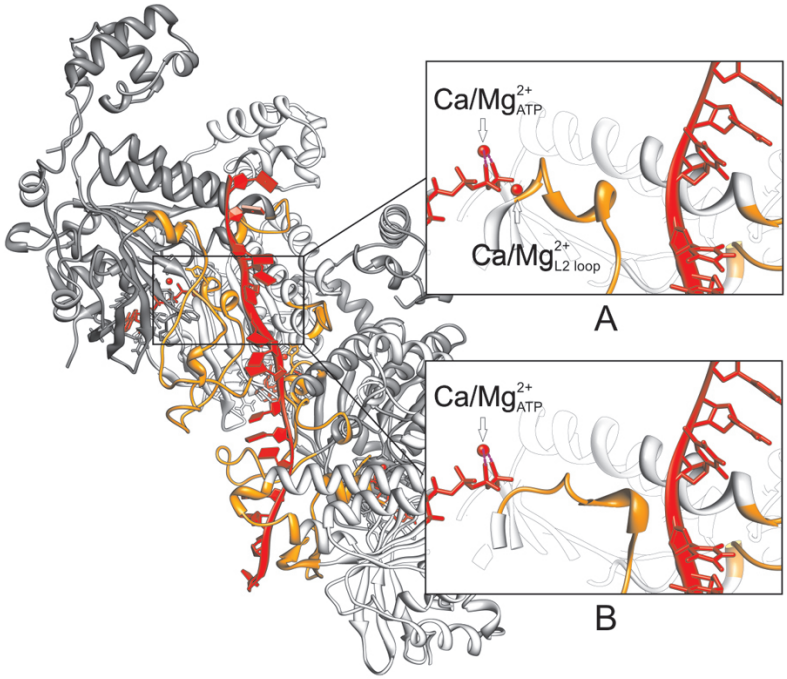


Figure 4.16. Model structure of the HsRad51-ssDNA-ATP filament with and without a divalent cation, (A) and (B) respectively, at the C-terminus of the L2 loop (orange). One cation (magnesium or calcium) coordinates the oxygen atoms of P_γ of ATP (cation labelled "ATP"), whereas another cation coordinates the α-helical region of loop L2 (cation labelled "L2 loop"). The α-helical part of the L2 loop, which secures the DNA binding inside the nucleoprotein filament, is stable in presence of the divalent cation (Ca^{2+} or Mg^{2+}). Without the divalent ion at the C-terminus of the loop, the α-helical region becomes elongated, significantly disturbing the DNA-binding.

The initial results indicated no structural differences between the complexes of HsRad51-ssDNA-ATP-Mg^{2+}/Ca^{2+} between the runs with two Mg^{2+} molecules and two Ca^{2+} molecules. However, a conducted literature survey suggested that polyvalent cations are still poorly described by the molecular dynamics force fields, which results in a tendency of such ions to stay around at the initial position. To overcome this limitation, MD runs including only one, ATP-coordinating cation (Ca^{2+} or Mg^{2+}), were performed. The results were quite inspiring: the L2 loop in the absence of the coordinating its C-terminal metal ion becomes elongated, as depicted in **Figure 4.16**, which ruins the integrity of the positively charged tunnel needed for stable accommodation of ssDNA inside the filament. The L2 loop in MD simulations of the dsDNA case, along with the experimental results, did not exhibit similar trend, probably, due to that dsDNA sits much tighter inside the filament, thus confining the ordered conformation of the L2 loop. And therefore we suggested that the Mg^{2+} ion being smaller and therefore much more mobile could easily leave the spot, resembling the MD run with only one cation at the ATP site. In conclusion, the α-helical conformational of C-terminus of the L2 loop, electrostatically stabilized by the bigger Ca^{2+} cation, secures an

ordered binding of ssDNA with bases oriented perpendicularly relative to the filamentous axis. This arrangement of the DNA bases might be of potential significance for the search of homology and subsequent strands exchange facilitated by the protein. This interplay between types of the cation and stability of the ordered conformation of the L2 loop most probably has an evolutionary meaning and is the subject of further research.

ON THE MECHANISM OF ATP-HYDROLYSIS IN BACTERIAL RECOMBINASE RECA (WITH PRELIMINARY RESULTS)

As said before, RecA is a prokaryotic recombinase. One of the protein's roles is to promote a template-directed repair of DNA double strand breaks. Although homologous recombination of DNA for creating genomic diversity by combining paternal and maternal genomes is not applicable to sex-less bacterial species, there is, however, an uptake and incorporation of exogenous DNA (transformation, conjugation) by homologous recombination to enrich the genomic information. This is certainly the origin of horizontal gene transfer and may be involved in the quick appearance of antibiotic resistance among various pathogenic bacteria, in this process RecA plays an essential role. RecA is also involved in the segregation of DNA for cell division (Roca & Cox, 1997; Cox, 2007a; Caro-Quintero et al., 2011). RecA aligns and compares two DNA molecules, and then depending on the degree of sequence similarity facilitates the strand exchange reaction. According to the scenario of the homologous recombination reaction, schematically described in chapter 2.4 of this Thesis, after a resection of broken DNA to produce a single-stranded DNA (ssDNA) overhang, RecA assembles on this ssDNA into an active form filament (presynaptic filament). RecA does not simply bind ssDNA upon the assembly; the protein inhomogeneously stretches DNA by 50% and unwinds it to 18 bp/turn of filament helix (Chen et al., 2008) so that each RecA subunit in the filament covers three nucleotides, preparing the strand for searching for its pairing partner. In search for homology the nucleoprotein filament invades a donor – double-stranded DNA (dsDNA) – to form the reaction intermediate (synaptic filament). Once homology is encountered, the ensuing strand exchange reaction results in a postsynaptic filament in which the complementary strand of the secondary DNA duplex is paired with the original ssDNA.

As was also discussed earlier, RecA is a DNA dependent ATPase. However, ATP hydrolysis is not required for the strand exchange reaction (Menetski & Kowalczykowski, 1985; Rosselli & Stasiak, 1990): the reaction also proceeds in the presence of a slowly hydrolysable ATP-γS cofactor. From the strand exchange reaction studies in the presence of non-covalently complexed cofactor ADP+AlF_4, it was suggested that the functionally active high DNA-binding affinity RecA filament is necessary and sufficient for the reaction to occur, and there is no requirement for any energy-supplying cofactor (!) (Kowalczykowski & Krupp, 1995). Kowalczykowski & Krupp, developing further the hypothesis of the possible role of ATP hydrolysis in RecA, speculated that the hydrolysis

plays a role of a molecular switch, permitting the cycling of RecA between two allosteric states: inactive, low DNA-binding affinity state, and active, high DNA-binding affinity state. The free energy needed to activate the DNA substrate is derived from the energy of binding of ATP, or an ATP analogue, at the interface between the two neighboring protein subunits, and is transferred to DNA via a cofactor-binding induced conformational transition of the protein's N-terminal, as deduced by comparing structures of inactive and active protein monomers (Chen et al., 2008).

Indeed, from the crystal structures of ssDNA- and dsDNA-RecA and uncomplexed protein filaments (PDB ID: 3CMU, 3CMX and 3CMV, correspondingly (Chen et al., 2008)), two DNA binding loops, L1 and L2, of the DNA-containing filaments form a well-defined binding pocket for DNA, while in the DNA-free filament, both of the loops are quite disordered (**Figure 4.17**). Interestingly, in the structure of the uncomplexed filament formed with AMP-PNP, a non-hydrolysable adenine triphosphate, despite being structurally analogous to ATP, the cofactor did not convert the filament into its active state, indicating the necessity of both a DNA substrate, acting as a filamentous "skeleton", as well as a cofactor, triggering a conformational change in the N-terminal region of the protein. In addition, the conformational transition from inactive to active state also induces a rotation of the cofactor's binding site from the interior of the filaments to the protein's subunit-subunit interface, as deduced from the analysis of crystal structures of the inactive form RecA (Story & Steitz, 1992; Story et al., 1992; Datta et al., 2000; Datta et al., 2003a, 2003b; Xing & Bell, 2004a, 2004b; Krishna et al., 2006, 2007; Prabu et al., 2008; Chen et al., 2008).

Another interesting observation is that the ATP-binding core – the central domain of a

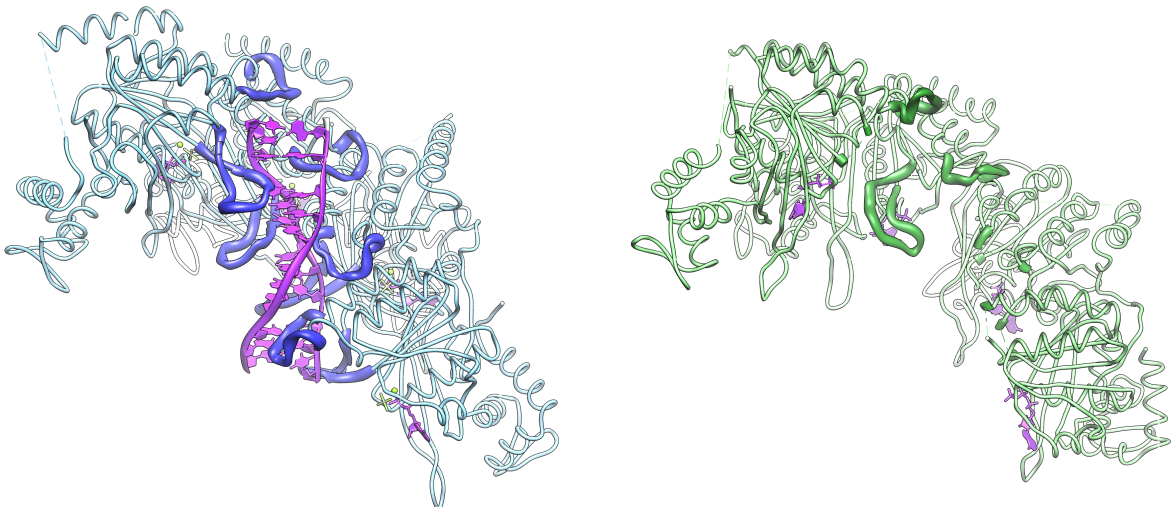


Figure 4.17 Crystal structures of the dsDNA-RecA (left) and DNA-free RecA (right) filamentous complexes with cofactor: ADP+AlF_4 for RecA-DNA, and AMP-PNP for uncomplexed RecA. DNA and cofactors colored magenta. DNA binding loops in active RecA colored purple, in inactive – green. DNA binding loops create a binding pocket for DNA in the interior of the filament.

RecA monomeric body – and in particular the amino acids composition in the ATP binding site, is found to be similar to various members of the RecA-like helicase family, including F_0F_1 ATP-synthase (Abrahams et al., 1994; Amano et al., 1994; Bowler et al., 2007), multiple helicases and DNA-transport proteins (Bianchet et al., 1997; Yu & Egelman, 1997; Bird et al., 1998; Egelman, 2000). These structural similarities might be indicative of some direct role of ATP hydrolysis in invoking a rotational degree of freedom of a RecA monomer or a part of it. Such a motion might be used for some specific protein functionality, in addition to the recycling function. Indeed, the RecA-mediated ATP hydrolysis is not accidental (Cox, 2007): it facilitates DNA manipulation when RecA is functioning at replication forks. For example, in the case of three-strand homologous recombination, the strand exchange proceeds in either direction, but becomes unidirectional, 5′→3′, relative to initially bound ssDNA, upon ATP hydrolysis (Cox & Lehrman, 1981; Konforti & Davis, 1992; Jian et al., 1994).

While formed on an ssDNA substrate, RecA does not stop at the boundary of this ssDNA, it moves on into the dsDNA region, resulting in a four DNA strands filament. This transition is ATP hydrolysis dependent and is always unidirectional (!), 5′→3′ relative to initially bound ssDNA (Kim et al., 1992; Shan et al., 1996). In addition, in the case of four-strand homologous recombination, ATP hydrolysis allows the bypass of a heterologous DNA substrate thus resolving significant structural barriers coupled to RecA-mediated unwinding of the entire heterologous segment of DNA. The reaction was reported to be completely blocked in the absence of ATP hydrolysis (Rosselli & Stasiak, 1991; Kim et al., 1992; Shan et al., 1996). The protein's ability to unwind DNA in an ATP hydrolysis dependent manner indeed suggests rotational swing motions of RecA monomers or parts of them, which might be indicative of a functional homology between RecA and a motor protein F_0F_1 ATP synthase.

Intrigued by the recombinase machinery and still great unclearness regarding the potential roles of ATP hydrolysis in homologous recombination by RecA, as well as molecular mechanisms utilizing this energy together with the fact of compositional similarities of ATP binding sites in RecA and F_0F_1 ATP synthase, I initiated a joint quantum mechanics-steered molecular dynamics study (QM/SMD) (in collaboration with Tamás Beke-Somfai) aiming to resolve some of these "white spots", using *in silico* tools as a "molecular microscope" monitoring the ATP hydrolysis reaction in DNA-RecA filaments. The original idea of the study is to determine the transition state structure(s) (TS) in the molecular mechanism of ATP hydrolysis at the binding site with the help of QM methods. This would be achieved by starting calculations from the ATP-state and proceeding towards the final ADP+P_i-state (ADP plus an inorganic phosphate ion, $H_2PO_4^-$). Then, using SMD, the ATP binding site will be reconfigured to satisfy the TS geometries obtained with QM, resulting in a set of SMD simulations. Monitoring the structural response of the DNA-RecA system onto the introduced changes in the ATP binding site, hopefully, will provide insight into the role and molecular mechanism of the

ATP hydrolysis in RecA function. Some very preliminary results and interesting observations will be briefly described below.

To start with, the ATP hydrolysis reaction requires water around the reaction site (by definition of hydrolysis). However, neither of the crystals of ssDNA-RecA (PDB ID 3CMU) and dsDNA-RecA (PDB ID 3CMX) have sufficiently high resolution to resolve solvent water molecules around ATP. Based on previous studies and structures of ATPase proteins, the target of the nucleophilic water in ATP hydrolysis is the γ-phosphate (Dittrich et al., 2003; Beke-Somfai et al., 2011). It was speculated, however, that the absence of water in the ATP-binding site for either of the complexes could be crucial for the stimulation of the strand exchange reaction (Chen et al., 2008). On the other hand, the lack of solvent molecules poses a question regarding the actual mechanism of ATP hydrolysis. It is important to understand in what way water molecules, necessary for the actual hydrolysis step, may penetrate into the site, and what structural changes their appearance and subsequent ATP hydrolysis will impose on the global DNA-RecA filamentous geometry.

To exclude the possibility that solvent was omitted in the crystals due to insufficient resolution of 2.8-4.2 Å (e.g. water oxygens become visible at 2-2.5 Å resolution), the DNA-RecA structures were processed with DOWSER (Zhang & Hermans, 1996) – a program that searches for possible locations of buried water molecules. However, the program did not locate any potentially important water molecules, within < 4.5 Å proximity to the ATP's P_Y atom, indeed indicating the structural tightness of the ATP binding site.

To investigate the above intriguing issue and to introduce solvent molecules into the catalytic site, I built a model system of the dsDNA-RecA complex based on PDB structure 3CMX. The model consisted of three RecA monomers, a DNA fragment of 10 base pairs, two ATP molecules and two Mg^{2+}. The cofactor is originally represented in the 3CMX structure by ADP+AlF_4, it thus was substituted by ATP by superimposing the adenine base. The model system was subjected to an MD equilibration in explicit water environment and 300 K temperature aiming to introduce solvent molecules into the catalytic site, as well as to relax possible steric tension induced by e.g. crystal packing. The system was prepared according to standard protocols; prior to the simulation the model was also processed with PROPKA analysis (Li et a., 2005) for correct assignment of the protein protonation state and then subjected to 10 ns unrestrained MD simulation. Throughout the simulation, however, the catalytic site did not open up enough to let any water molecules in. Moreover, the overall structure of the nucleoprotein fragment was quite stable, RMSD values of the C_α atoms of the protein's backbone fluctuated below 3 Å. These observations suggest that the initiation step of ATP hydrolysis, namely the appearance of water molecules in the site, probably, is not triggered by any thermal fluctuations of the system but involves somewhat bigger conformational rearrangements of the nucleoprotein filament structure!

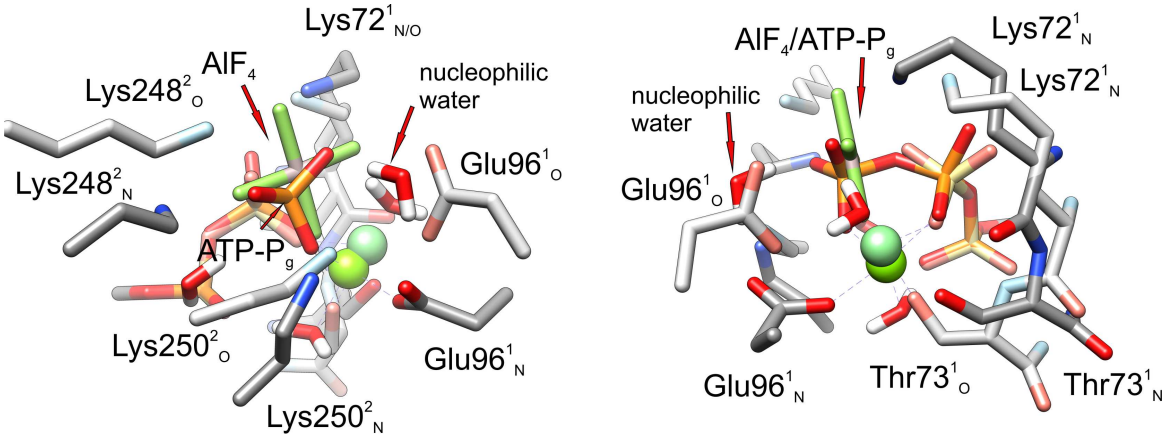


Figure 4.18. The ATP binding site in the dsDNA-RecA complex. Only the moieties of the most conspicuously deviating from the crystal structure residues are depicted. To avoid overcrowding, only water hydrogens are shown. Pale colors plus a subscript "O" (from "old") denote the crystal structure; brighter colors plus a subscript "N" ("new") represent the calculated ATP-state configuration of the active site. Superscript "1" and "2" denote different RecA monomers. Color coding: oxygen coral -> red; nitrogen sky blue -> blue; carbon light grey -> grey; phosphorus wheat yellow -> orange red; magnesium light green -> green.

To envisage the character and amplitude of these conformational rearrangements, a new unrestrained MD equilibration was initiated, but in contrast to the system setup in the first MD run, four water molecules were added manually into the active site, positioned at < 4 Å proximity to the ATP's γP atom, analogous to the ATP binding site of F_oF1 ATP synthase (from PDB ID 2JDI, Bowler et al., 2007). Interestingly, the structure did not remain as stable as in the first simulation, some of the residues comprising the ATP binding site fluctuated markedly. Among those the most fluctuating were: Glu96^1, Lys248^2 and Lys250^2, and somewhat less fluctuating – Lys72^1 and Thr73^1 (superscripts 1 and 2 denote the two neighboring RecA monomers). The latter two residues are in the so-called P-loop, the loop that accommodates the adenine triphosphate. Not less interesting observation is that the rest of the P-loop residues did not fluctuate significantly The averaged structure of the nucleoprotein filament fragment from the second MD equilibration run, cooled down to 0 K, was used to create the ATP binding site for the subsequent QM active site calculations (Himo, 2006), starting with the ATP-state configuration of the active site. By having a compromise between including all presumably important fragments of the nearby amino acid residues for the hydrolysis reaction and between keeping the size of the system feasible for QM calculations, the QM model of the active site was constructed having a total of 164 atoms. The site included the tri-phosphate moiety of ATP up to the C5' atom of the ATP ribose ring, a Mg^{2+} ion, the backbone and side chains of the residues Glu68^1-Thr73^1 of the P-loop, as well as side chains of additional residues: Glu96^1, Gln194^1, Phe217^2-Ala219^2, Lys248^2, Lys250^2. This QM model of the catalytic site was employed to locate the most crucial critical points of

the ATP hydrolysis reaction with geometry optimization performed at the B3LYP/6-31G* level of theory.

The superposition of the optimized ATP site and the original active site from the 3CMX structure (fragmentally illustrated in **Figure 4.18**) may provide some interesting mechanistic insight. The putative catalytic Glu96[1] (Skiba et al., 1994: Skiba et al., 1999; Chen et al., 2008) residue has shifted in order to accommodate a nucleophilic water molecule. In the crystal structure this glutamate residue is connecting via a hydrogen bond network, spanning across the subunit-subunit interface, to dsDNA. This network is disrupted, presumably (my speculation) imposing structural fluctuation of the DNA binding site. Two lysine residues, 248 and 250, which are believed to stabilize (Chen et al., 2008) the cofactor's transition state, have also shifted as result of the site expansion due to the presence of water molecules. Position of these water molecules is within 3.5 Å of the P_γ atom. Presumably the hydrolysis step progresses via a TS with planar configuration of the $P_\gamma O_3^-$ moiety. The most suitable water molecule to perform such a nucleophilic attack initiating the bond breaking between P_γ and P_β is termed nucleophilic water in (**Figure 4.18**).

Encouragingly, preliminary results from QM calculations of the ATP hydrolysis energetics with the help of the obtained active site model provide a 22.8 kcal/mol barrier height for the hydrolysis step, which from a computational perspective is rather close to the free energy value derived back from the experimental rates (18 kcal/mol (Brenner et al., 1987)). While this validates our model to some extent, one may still argue that the designed ATP binding site deviates significantly from that of the crystal structure and thus might not be justified as a structure to be employed for QM calculations of the transition states and the subsequent SMD simulations. Still though, to provide any progress as to accurate mechanistic understanding of the hydrolysis reaction, in absence of any high-resolution structure that includes water at the active site, various attempts with trial structures are clearly motivated. It should also be taken into account that significant fluctuations are likely to occur in the active sites of a functional and dynamic enzyme, such as RecA, which could result in significant structural deviations compared to the static crystal structures, as exemplified by the results of the second MD simulation. To investigate this further as well as compare models with and without water molecules in the active site, we have ongoing investigations using a similar QM procedure on the model built without water molecules in the active site. In addition, the QM results will be supplemented with combined quantum mechanics/molecular mechanics calculations (QM/MM). And finally we plan to perform the SMD simulations described above using information from both sets of models. More results to come!

5. CONCLUDING REMARKS

In the collection of scientific illustrations presented on the pages of this Thesis I have tried to emphasize and put forward the necessity and importance of a molecular structure for unveiling of mechanisms of molecular interactions, as well as to capture the elusiveness of the dynamic nature of biomolecular existence and interplay. The two main "characters" of my "stories", namely the DNA molecule and the family of recombinase proteins, exemplify the two main classes of biological macromolecules. The polymorphic character of the former and highly dynamic nature of the latter, as well as, certainly, their vital importance for every living cell, put these two biomacromolecules into focus of many scientific studies, both experimental and theoretical, providing versatile knowledge as well as new (virtually endless) questions as to the mechanisms of their function.

In this Thesis the question of the DNA polymorphism was addressed via structural studies of intercalative binding of transition metal compounds: mononuclear $[Ru(phen)_3]^{2+}$ and $[Ru(phen)_2dppz]^{2+}$; and binuclear, $[Ru_2(phen)_4bidppz]^{4+}$ and $[Ru_2(bpy)_4bidppz]^{4+}$. The relatively small number of atoms, in comparison with, for instance, DNA-binding proteins, allows studies of these systems *in silico*, where the number of atoms comprising the system of interest, despite rapid increase in computational resources, still matters. Being sensitive photophysical probes, these compounds permit direct monitoring of their interactions with DNA, providing invaluable experimental data that in combination with molecular modelling could result in atomic-level mechanisms of molecular interactions, as exemplified in *Papers I* and *III*.

The binuclear ruthenium compounds, being thread intercalators, in order to pass through the tightly-packed DNA stack one of their bulky centres, have to invoke large transient conformational rearrangements of the helix, as indicated by the extremely slow, two weeks time at room temperature, kinetics. The hydrophobic nature of the auxiliary ligands seems to be the "key-stone" for the understanding of the multistage tread-intercalation process, as illustrated by the counterintuitive experimental observation that the spatially smaller but also less hydrophobic $[Ru_2(bpy)_4bidppz]^{4+}$ takes longer times to thread than a fatter "colleague". In *Paper I*, discussing thread-intercalation mechanism – based on a number of experimental and theoretical observations, it is proposed that threading could be facilitated by a preceding B- to A-like-form DNA conformational transition – a hypothesis that remains to be verified. One of the plausible theoretical approaches suitable for this task would be comparative free energy calculations coupled to replica-exchange MD approach for better exploration of conformational space performed for both $[Ru_2(phen)_4bidppz]^{4+}$ and $[Ru_2(bpy)_4bidppz]^{4+}$.

An experimental attempt made, with the help of nuclear magnetic resonance spectroscopy, to capture the thread-intercalated or any of its possible transition states on the way to intercalation by $[Ru_2(bpy)_4bidppz]^{4+}$, so far resulted in fiasco as to adding to understanding the threading mechanism, emphasizing once more the immense complexity of the process. On the other hand, the study, illustrated in *Paper II*, produced a highly interesting structure of the outside binding configuration of a $[Ru_2(bpy)_4bidppz]^{4+}$-DNA complex. This binding mode may represent the precursory state before threading, i.e. the binding geometry that Wilhelmsson et al. observed initially, characterized by a positive LD (Wilhelmsson et al., 2002).

Monomeric ruthenium compounds, on the other hand, while acting as classic intercalators induce, maybe not as dramatic as in the case of binuclear threading intercalators, but still noticeable conformational rearrangements of the DNA double helix. One of the particularly exciting observations, reported in *Paper III*, is the spectacular enantiospecific character of binding of the smaller $[Ru(phen)_3]^{2+}$: right-handed Δ induces DNA helix kinking by as much as 53°(!), while its mirror image, left-handed, Λ produces only inconspicuous helical bending. The chiral recognition of DNA is interesting in the context of evolutionary selection, and particularly exciting since analogous binding arrangements seem to be widely exploited by operatory protein, for instance the eukaryotic transcription factor TBP (Kim et al., 1993; Nikolov et al., 1996), the high mobility group proteins (Klass et al., 2003; Palasingam et al., 2009), and thermophile chromosomal proteins (Robinson et al., 1998).

Another outstanding example of DNA's polymorphic nature is the molecule's conformation inside protein filaments formed by recombinases, e.g. RecA and Rad51. Upon the nucleoprotein filament assembly the recombinase monomers do not simply bind DNA, they inhomogeneously stretch the molecule by as much as 50% and unwind it to 18 base pairs per turn of protein filament helix, as seen from the recently solved crystal structures of RecA-DNA complexes (Chen et al., 2008). The human recombinase, HsRad51, arranges analogously to its bacterial homologue RecA, in long filamentous fibers, as was deduced from the electron microscopy studies (Yu et al., 2001). However, in the absence of any high-resolution structure of the HsRad51 filament researchers could only speculate about how two evolutionarily separated recombinases, RecA and Rad51, affect the DNA structure in similar or different manners.

To answer this question, as well as to gain a mechanistic insight into the function of human recombinase, a 3-D model structure of the protein was assembled using a combination of Site-Specific Linear Dichroism spectroscopy and molecular modelling (*Paper IV*). The obtained model indeed highlighted a number of important structural features. In particular, the two putative DNA binding loops, L1 and L2, are found to be positioned on the interior of the filament enabling direct contact with a DNA molecule accommodated in the centre of the filamentous "shaft". The non-uniformly stretched DNA structure, adopted from the RecA-dsDNA crystal (PDB ID: 3CMX), fits well: the binding stoichiometry of one protein monomer per three nucleotides is analogous to the

RecA structure. The most exciting observation is the binding arrangements of the L1 and L2 loops. Taken for what it may be worth, the model suggests a mechanism in which the human Rad51 filament assembly employs a Tyr residue of the L1 loop as a binding anchor for the protein monomeric subunits on the DNA scaffold. As exciting as it might sound, this protein-DNA interaction is yet a speculation and subject of further investigations (work in progress).

Another interesting structural feature deduced from the model is the location of a putative ATP-binding pocket at the interface between two protein subunits and in close proximity to DNA, yet another analogy with RecA. The binding of ATP is known to be essential for the filaments assembly although its role is not known in detail (Cox, 2007a; Cox, 2007b). Sandwiched between the two protein subunits, ATP could act as a "glue" during the filaments assembly but it is also believed to initiate a high affinity DNA binding state of the protein filament (Kowalczykowski & Krupp, 1995). Proposed in *Paper V* is a molecular mechanism explaining the stimulatory effect of Ca^{2+} on the strand exchange reaction performed by HsRad51, in a way, "tempered" by the hypothesis of Kowalczykowski & Krupp, originally proposed for bacterial RecA. In the case of human Rad51, it seems, that the high affinity DNA binding state could not be solely induced by ATP, the presence of a second co-factor, namely a Ca^{2+} cation, is required. A conformational stability of the L2 loop, electrostatically supported by the cation, secures an ordered binding of ssDNA with bases oriented perpendicularly relative to the filamentous axis. The latter arrangement might be of potential significance for the search of homology and subsequent strands exchange facilitated by the protein.

In conclusion, the results presented in this Thesis are in my opinion remarkably incomplete: for every question answered – new, not less interesting questions tend to appear – which I personally consider to be the greatest achievement of my work. The mechanisms of DNA dynamic response to changes of physico-chemical environment around the molecule induced by DNA binding proteins or small ligands, as well as the structure and mechanisms of function of recombination proteins RecA and Rad51 have been addressed in this Thesis. The Thesis also provides some examples of a combinatory way of approaching a scientific problem, using methods or molecules from widely different contexts.

6. POPULAR SCIENCE SUMMARY

Every biological cell, whether bacterial or human, is an immensely complex system comprising of millions of microscopic objects – molecules – that exist in constant motion and incessantly interact with each other. One of the central molecules in this biological "society" is DNA – the most compressed miniature data bank in the world. By the sequence of nucleotides, the elementary units or code letter of DNA, the molecule preserves and passes on the genetic information needed for the development and reproduction of all living organisms. The structure of DNA, a double helix, over the years has become a popular symbol of molecular science and much more. It is associated in minds with something *new*, some breakthrough scientific achievements.

However, very few know that a DNA molecule can change its form, it can become wider, narrower, it can be bent, stretched and compressed – a biomolecular chameleon! This so-called structural polymorphism, utilized over millions of years of natural evolution, is of vital importance: it is needed for DNA to be able to specifically interact with various molecules during its life cycle. The molecule dynamically responds to any change in the surrounding environment by adopting the most suitable form to continue function under new conditions. This structural plasticity of DNA is also used by scientists, for example, for the development of medicines against genetic diseases or cancer.

To utilize DNA polymorphism it is crucial to understand in greater details the structural mechanics and dynamics of the DNA molecule in different environments and under interactions with other molecules, the topic which was addressed in this Thesis.

One direction, investigated in this Thesis, is DNA interaction with relatively small dumbbell-shaped synthetic molecules, which first bind from the outside of DNA and then in about two weeks time thread one of their bulky "heads" through a densely packed DNA stack, requiring large transient rearrangements of the double helix structure.

Another direction taken in this Thesis regards DNA interaction with a human enzyme, named Rad51. The main role of the enzyme is to exchange near identical-sequences of DNA against each other, a process used for DNA repair but also for creating genetic diversity by combining paternal and maternal genomes upon cell division in sexual reproduction. In order to function Rad51 must first assemble itself on DNA by mechanically stretching and unwinding the DNA molecule to form long filamentous constructs.

As one may guess, the rest of the world of molecules is not less dynamics then of those described above, and probably equally, if not more, complex. The mechanisms of molecular interplay have been barely touched upon in this Thesis, and could be subjects of many further investigations. The research to be continued!

7. AUTHOR'S ACKNOWLEDGEMENTS

A number of people have contributed in one way or another, directly or indirectly, to this Thesis. I am deeply grateful for all: support, help, advices, ideas, etc., during this period of my life. I feel most lucky to have met so many talented, enthusiastic, smart, humorous, and simply very nice people during this exciting in many ways journey, which has come to an end… So,

First of all I would like to thank to my supervisor, ***Bengt Nordén***, for conducting this "experiment" with me ☺. These PhD years have been an extremely challenging experience with a lot of ups and downs. Thank you for never loosing belief in me, or at least never showing that you lost it ☺. Thank you for being more a friend than a supervisor, because if the opposite was true, then I would probably have quit this job long ago. Thank you for letting me choose my projects and guiding me to become as independent as I am now. Thank you for all your help, small and big! Thank you for a friendly kick now and then (e.g. Sunday morning 7 a.m. ☺) Thank you for introducing me to sailing – it has indeed become a true love of mine. I have much more to thank you for, but I shall stop now to leave some space for others, not least important to be thanked for.

Second, I would like to thank to my co-supervisor, ***Per Lincoln***, for always having time for my "5-minutes"questions ☺. Thank you for not pressing me too much with the ruthenium compounds. I have been always impressed by the deepness of your thoughts and questions, which always kept my brain in a good shape.

I would also like to thank to other seniors with whom I was lucky to collaborate. ***Masayuki Takahashi*** for a fruitful collaboration, for teaching me basics of molecular biology, introducing me to the field of recombination proteins and for all our funny conversations. You have always been my 3rd supervisor. ***Martin Billeter*** for introducing me to the field of computational chemistry at a time when you were my master project supervisor, and for interesting collaboration later.

Now switching to youngsters ☺. I would like to thank to all my co-authors, to start with ***Pär Nordell*** – for introducing me to the mysterious field of ruthenium compounds and for being a great roommate during my first 6 months; ***Karolin Frykholm*** – for letting me take over the Rad51 project, for teaching me Swedish and, most importantly, you were the only one (others, please, no offence) who tried to get into all computational details of the project, which was actually very helpful; ***Lisha Wu*** – for providing me with an opportunity to curb my temper ☺, sorry about being harsh with you sometimes; ***Louise Fornander*** – for showing me that sometimes things seem much better than they are, but everything works out fine in the very end; ***Tamás Beke-Somfai*** – for teaching me quantum chemistry and for your friendship; ***Piotr Hanczyc*** – for the "PVA-LD"

experiments that have helped me to move on with the Ru-DNA project, and for the "diplomacy lessons" ☺; ***Bobo Feng*** – despite our "insulin inversion" paper never came out, I am grateful to you for teaching me how to do real experiments; ***Catherine Kitts*** – for an interesting collaboration to be continued, for delicious cakes and for introducing me to a different point of view on various things, looking forward to see you in the Netherlands ☺.

Apart from my co-authors, whose contribution to the accomplishment of this Thesis is more direct, I would also like to thank people that were creating my daily atmosphere - my roommates: ***Kristina Fant***, ***Moheb Nayeri*** and ***Janette Ulama***. Thank you guys for all entertaining conversations and for being a real support when I was facing both difficulties with work and personal problems over the years.

I would like to thank ***Bo Albinsson*** and ***Per Lincoln*** for running the division in a good way. ***Anna Molander***, ***Carina Pettersson*** and ***Gunilla Saethe*** for all help with the administrative and practical life issues.

I am very grateful to ***all people at Physical Chemistry***, present and former, for creating such a nice and friendly atmosphere at the department. It indeed matters if one feels nice and welcome at work.

Special thanks to ***Alex Voronov***, ***Tamás Beke-Somfai***, ***Masayuki Takahashi*** and ***Bengt Nordén*** for proofreading of this Thesis!

I also would like to acknowledge the ***two gangs*** of the Greek "crusades". It was a great time, thank you all! Gang number 1: thank you for pushing me off my comfort zone when talking to me in Swedish – otherwise I would have probably never started. Gang number 2: 7 days in such a heterogeneous cultural atmosphere made a deep impression on me.

King Abdullah University of Science and Technology is kindly acknowledged for sponsoring my PhD projects, as a part of award (KUK-11-008-23) to Bengt Nordén.

Special thanks go to all ***my friends*** from the department and outside it for their moral support, understanding, for letting me be who I am and accepting me with all my good and bad. Special thanks, guys, for all the chocolate, which I have received during the last few months! ☺ I am sorry that I have abandoned you during my last PhD year.

My deepest acknowledgement I would like to send over to my very first supervisor ***Sergey N. Volkov***. Thank you for introducing me to the field of structural studies of biological macromolecules, for bringing into me this sparkle of interest in Science.

And finally I would like to acknowledge the most important for me people – my ***family***. This Thesis is dedicated to the memory of my ***grandfather Viktor***, who always believed in me and supported me, from the first day of my life till the last day of his. You made me believe that everything is possible! I miss you very much...

My other ***grandfather***, ***Jacob***, I would like to thank for providing me with an opportunity to see Europe and come to Sweden and see Chalmers, back in 2004. Danke, Opa, dass du immer erinnerst mich, dass ich Sorge um mich muss zu nehmen haben.

My ***grandmother Alla,*** I would like to thank for her constant reminding me of not stop working and keep moving forward every day even if little.

I would like to thank my ***mum*** who always tries very enthusiastically to help me with all my problems including scientific research. My greatest acknowledgment, in this regard, goes to my ***dad*** for actually slowing down my mum's enthusiasm ☺ but to be serious, you two are the best parents I could have ever imagined for myself! Despite being 2000 km apart I always feel – you are close.

And last but not least, I would like to thank ***Alex***. Over the years you have been a great support for me both workwise and in real life. Thank you for being my best friend and more. You mean so much to me! And sorry that I made you learn so "common" in everyday life words like *HsRad51* and $[Ru(phen)_2dppz]^{2+}$ ☺

8. BIBLIOGRAPHY

Abrahams, J. P., Leslie, A. G., Lutter, R., & Walker, J. E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*, *370*(6491), 621-8.

Aihara, H., Ito, Y., Kurumizaka, H., Terada, T., Yokoyama, S., & Shibata, T. (1997). An interaction between a specified surface of the C-terminal domain of RecA protein and double-stranded DNA for homologous pairing. *J. Mol. Biol.*, *274*, 213-221.

Aihara, H., Ito, Y., Kurumizaka, H., Yokoyama, S., & Shibata, T. (1999). The N-terminal domain of the human Rad51 protein binds DNA: structure and a DNA binding surface as revealed by NMR. *J. Mol. Biol.*, *290*, 495-504.

Alberts, B., & Miake-Lye, R. (1992). Unscrambling the puzzle of biological machines: the importance of the details. *Cell*, *68*, 415-420.

Allinger, N. L., Zhou, X., & Bergsma, J. (1994). Molecular mechanics parameters. *J. Mol. Struc.: TheoChem*, *312*, 69-83.

Almlöf, J., & Taylor, P. R. (1991). Atomic Natural Orbital (ANO) basis sets for quantum chemical calculations. *Adv. Quan. Chem.*, *22*, 301-373.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, *215*, 403-410.

Amano, T., Yoshida, M., Matsuo, Y., & Nishikawa, K. (1994). Structural model of the ATP-binding domain of the F1-β subunit based on analogy to the RecA protein. *FEBS Letters*, *351*(1), 1-5. doi:10.1016/0014-5793(94)00796-9

Andrec, M., Felts, A. K., Gallicchio, E., & Levy, R. M. (2005). Protein folding pathways from replica exchange simulations and a kinetic network model. *Proc. Natl. Acad. Sci. U.S.A.*, *102*, 6801-6806.

Antonarakis, E. S., & Emadi, A. (2010). Ruthenium-based chemotherapeutics: are they ready for prime time? *Cancer Chemother. Pharm.*, *66*, 1-9.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-Model workspace: a web-based environment for protein structure homology modelling. *Bioinform.*, *22*, 195-201.

Ardhammar, M., Mikati, N. and Nordén, B. (1998) Chromophore orientation in liposome membranes probed with flow linear dichroism. *J. Am. Chem. Soc., 120*, 9957-9958.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.*, *64*, 2229.

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.*, *79*, 137-158.

Bader, R. F. W. (1985). Atoms in molecules. *Acc. Chem. Res.*, *18*, 9-15.

Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. *Science*, *294*, 93-96.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, *98*, 10037-10041.

Ban, N. (2000). The complete atomic structure of the large ribosomal subunit at 2.4Å resolution. *Science*, *289*, 905-920.

Barton, J. K., Danishefsky, A., & Goldberg, J. (1984). Tris(phenanthroline)ruthenium(II): stereoselectivity in binding to DNA. *J. Am. Chem. Soc.*, *106*, 2172-2176.

Barton, J. K., Goldberg, J. M., Kumar, C. V., & Turro, N. J. (1986). Binding modes and base specificity of tris(phenanthroline)ruthenium(II) enantiomers with nucleic acids: tuning the stereoselectivity. *J. Am. Chem. Soc.*, *108*, 2081-2088.

Baumann, P., & West, S. C. (1998). Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends in Biochemical Sciences*, *23*(7), 247-251.

Bayly, C. I., Cieplak, P., Cornell, W., & Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, *97*, 10269-10280.

Becke, A. D. (1993). A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.*, *98*, 1372.

Beke-Somfai, T., Lincoln, P., & Nordén, B. (2011). Double-lock ratchet mechanism revealing the role of alphaSER-344 in FoF1 ATP synthase. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(12), 4828-33. doi:10.1073/pnas.1010453108

Becker, H. C., & Nordén, B. (1997). DNA binding properties of 2,7-diazapyrene and its N-methylated cations studied by linear and circular dichroism spectroscopy and calorimetry. *J. Am. Chem. Soc.*, *119*, 5798-5803.

Benson, F. E., Stasiak, A., & West, S. C. (1994). Purification and characterization of the human rad51 protein, an analogue of E. coli RecA. *EMBO J.*, *13*, 5764-5771.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., et al. (2000). The Protein Data Bank. *Nucl. Ac. Res.*, *28*, 235-242.

Beveridge, D. L., & Dicapua, F. M. (1989). Free energy via molecular simulation: applications to chemical and biomolecular systems. *Ann. Rev. Biophys. Chem.*, *18*, 431-492.

Beveridge, D. L., & McConnell, K. J. (2000). Nucleic acids: theory and computer simulation, Y2K. *Curr. Opin. Struc. Biol.*, *10*, 182–196.

Beveridge, D. L., & Ravishanker, G. (1994). Molecular dynamics studies of DNA. *Curr. Opin. Struc. Biol.*, *4*, 246–255.

Bianchet, M. A., Ko, Y. H., Amzel, L. M., & Pedersen, P. L. (1997). Modeling of nucleotide binding domains of ABC transporter proteins based on a F1-ATPase/recA topology: structural model of the nucleotide binding domains of the cystic fibrosis transmembrane conductance regulator (CFTR). *Journal of bioenergetics and biomembranes*, *29*(5), 503-24. doi:10.1023/A:1022443209010

Bianco, P. R., Tracy, R. B., & Kowalczykowski, S. C. (1998). DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci. J. Virt. Lib.*, *3*, 570-603.

Billeter, M., Güntert, P., Luginbühl, P., & Wüthrich, K. (1996). Hydration and DNA recognition by Homeodomains. *Cell*, *85*, 1057-1065.

Bird, L. E., Subramanya, H. S., & Wigley, D. B. (1998). Helicases: a unifying structural theme? *Current Opinion in Structural Biology*, *8*(1), 14-18. doi:10.1016/S0959-440X(98)80004-3

Birrane, G., Soni, A., & Ladias, J. A. A. (2009). Structural basis for DNA recognition by the human PAX3 Homeodomain. *Biochem.*, *48*, 1148-1155.

Blaney, J. M., & Dixon, J. S. (1993). A good ligand is hard to find: automated docking methods. *Persp. Drug Diss. Des.*, *1*, 301-319.

Born, M., & Oppenheimer, R. (1927). Zur Quantentheorie der Molekeln. *Ann. Der Phys.*, *389*, 457-484.

Bowler, M. W., Montgomery, M. G., Leslie, A. G. W., & Walker, J. E. (2007). Ground state structure of F1-ATPase from bovine heart mitochondria at 1.9 A resolution. *The Journal of biological chemistry*, *282*(19), 14238-42. doi:10.1074/jbc.M700203200

Boys, S. F. (1950). Electronic wave functions. I. A general method of calculation for the stationary states of any molecular system. *Proc. Royal Soc. A: Math. Phys. Engin. Sci.*, *200*, 542-554.

Brandt, P., Norrby, T., Åkermark, B., & Norrby, P. O. (1998). Molecular mechanics (MM3*) parameters for ruthenium(II)-polypyridyl complexes. *Inorg. Chem.*, *37*, 4120-4127.

Braunschweiler, L., & Ernst, R. R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.*, *53*, 521-528.

Brendel, V., Brocchieri, L., Sandler, S. J., Clark, A. J., & Karlin, S. (1997). Evolutionary comparisons of RecA-like proteins across all major kingdoms of living organisms. *J. Mol. Evol.*, *44*, 528-541.

Breneman, C. M., Wiberg, K. B., Breneman, & Wiberg. (1990). Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comp. Chem.*, *11*, 361-373.

Brenner, S. L., Mitchell, R. S., Morrical, S. W., Neuendorf, S. K., Schutte, B. C., & Cox, M. M. (1987). recA protein-promoted ATP hydrolysis occurs throughout recA nucleoprotein filaments. *J. Biol. Chem.*, *262*(9), 4011-4016.

Breu, J., & Stoll, A. J. (1996). Tris(1,10-phenanthroline)ruthenium(II) Bis(hexafluoro-phosphate). *Acta Crystallographica C Crys. Struc. Comm.*, *52*, 1174-1177.

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., et al. (2009). CHARMM: the biomolecular simulation program. *J. Comp. Chem.*, *30*, 1545-1614.

Bugreev, D. V., & Mazin, A. V. (2004). Ca^{2+} activates human homologous recombination protein Rad51 by modulating its ATPase activity. *Proc. Natl. Acad. Sci. U.S.A.*, *101*, 9988-9993.

Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Höfle, M. G., Klappenbach, J., & Konstantinidis, K. T. (2011). Unprecedented levels of horizontal gene transfer among spatially co-occurring Shewanella bacteria from the Baltic Sea. *The ISME journal*, *5*(1), 131-40.

Case, D.A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., et al. (2010). *AMBER 11*. San Francisco.

Case, David A, Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., et al. (2005). The Amber biomolecular simulation programs. *J .Comp. Chem.*, *12255*, 1668-1689.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., & Skelton, N. J. (2007). *Protein NMR Spectroscopy: Principles and Practice* (p. 885). Elsevier Inc.

Cech, T. R. (2000). Structural biology. The ribosome is a ribozyme. *Science*, *289*, 878-879.

Chambron, J. C., Sauvage, J. P., Amouyal, E., & Koffi, P. (1985). $Ru(Bipy)_2$(dipyrido-phenazine)$^{2+}$ a complex with long-range directed charge transfer exited state. *Nouveau journal de chimie*, *9*, 527-529.

Chen, F. M. (1988). Kinetic and equilibrium binding studies of actinomycin D with some d(TGCA)-containing dodecamers. *Biochem.*, *27*, 1843-1848.

Chen, J., Villanueva, N., Rould, M. A, & Morrical, S. W. (2010). Insights into the mechanism of Rad51 recombinase from the structure and properties of a filament interface mutant. *Nucl. Ac. Res.*, *38*, 4889-4906.

Chen, Z., Yang, H., & Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, *453*, 489-494.

Chi, P., Van Komen, S., Sehorn, M. G., Sigurdsson, S., & Sung, P. (2006). Roles of ATP Binding and ATP hydrolysis in human Rad51 recombinase function. *DNA Repair*, *5*, 381-391.

Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.*, *5*, 823-826.

Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., Heinz, T. N., et al. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *J. Comp. Chem.*, *26*, 1719-1751.

Clark, A.J., & Margulies, A. D. (1965). Isolation and characterization of recombination-deficient mutants of Escherichia coli K12. *Proc. Natl. Acad. Sci. U. S. A.*, *53*, 451-459.

Clark, G. R., Boykin, D. W., Czarny, A., & Neidle, S. (1997). Structure of a bis-amidinium derivative of Hoechst 33258 complexed to dodecanucleotide d(CGCGAATTCGCG)$_2$: the role of hydrogen bonding in minor groove drug-DNA recognition. *Nucl. Ac. Res.*, *25*, 1510-1515.

Clementi, E., Chakravorty, S. J., Corongiu, G., & Sonnad, V. (1990). Independent electron models: Hartree-Fock for many-electronatoms. In E. Clementi (Ed.), *Mod. Tech. Comp. Chem.: MOTECC 1990* (pp. 47-140). Springer.

Comba, P. (1994). Solution structures of coordination compounds. *Comm. Inorg. Chem.*, *16*, 133-151.

Conner, B. N., Yoon, C., Dickerson, J. L., & Dickerson, R. E. (1984). Helix geometry and hydration in an A-DNA tetramer: IC-C-G-G. *J. Mol. Biol.*, *174,* 663-695.

Conway, A. B., Lynch, T. W., Zhang, Y., Fortin, G. S., Fung, C. W., Symington, L. S., & Rice, P. A. (2004). Crystal structure of a Rad51 filament. *Nat. Struc. Mol. Biol.*, *11*, 791-797.

Cook, S. A. (1971). *The complexity of theorem-proving procedures*. *Proceedings of the third annual ACM symposium on Theory of computing - STOC '71* (pp. 151-158). New York, USA: ACM Press.

Cooper, A., Johnson, C. M., Lakey, J. H., & Nöllmann, M. (2001). Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.*, *93*(2-3), 215-230.

Cornell, W., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, *117*, 5179-5197.

Cox, M. M., & Lehman, I. R. (1981). Directionality and polarity in recA protein-promoted branch migration. *Proc. Natl. Acad. Sci. U.S.A.*, *78*(10), 6018-6022.

Cox, M. M. (2003). The bacterial RecA protein as a motor protein. *Ann. Rev. Microbiol.*, *57*, 551-577.

Cox, M. M. (2007a). Motoring along with the bacterial RecA protein. *Nat. Rev. Mol. Cell Biol.*, *8*, 127-138.

Cox, M. M. (2007b). The bacterial RecA protein: structure, function, and regulation. *Molecular Genetics of Recombination* (pp. 53-94). Springer.

Cramer, C. J. (2004). *Essentials of Computational Chemistry: Theories and Models* (2nd ed., p. 618). Wiley.

Crespo, A., Martí, M. A., Estrin, D. A., & Roitberg, A. E. (2005). Multiple-steering QM-MM calculation of the free energy profile in chorismate mutase. *J. Am. Chem. Soc.*, *127*, 6940-6941.

Crick, F. H. C. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.*, *12*, 138-163.

Crow, S. D. G., Bailly, C., Garbay-Jaureguiberry, C., Roques, B., Shaw, B. R., & Waring, M. J. (2002). DNA sequence recognition by the antitumor drug ditercalinium. *Biochem.*, *41*, 8672-8682.

Čížek, J. (1966). On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.*, *45*, 4256.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, *98*, 10089-10092.

Datta, S., Ganesh, N., Chandra, N. R., Muniyappa, K., & Vijayan, M. (2003). Structural studies on MtRecA-nucleotide complexes: insights into DNA and nucleotide binding and the structural signature of NTP recognition. *Proteins*, *50*, 474-485.

Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K., & Vijayan, M. (2003). Crystal structures of Mycobacterium smegmatis RecA and its nucleotide complexes. *J. Bacteriolg.*, *185*, 4280-4284.

Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K., & Vijayan, M. (2000). Crystal structures of Mycobacterium tuberculosis RecA and its complex with ADP-AlF_4: implications for decreased ATPase activity and molecular aggregation. *Nucl. Ac. Res.*, *28*, 4964-4973.

Davidson, E. R., & Feller, D. (1986). Basis set selection for molecular calculations. *Chem. Rev.*, *86*, 681-696.

Davidsson, Å., & Nordén, B. (1976). Conversion of Legrand-Grosjean circular dichroism spectrometers to linear dichroism detection. *Chemica Scripta*, *9*, 49-53.

Dittrich, M., Hayashi, S., & Schulten, K. (2003). On the mechanism of ATP hydrolysis in F1-ATPase. *Biophysical journal*, *85*(4), 2253-66.

Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, *426*, 884-890.

Dolja, V. V., Karasev, A. V., & Koonin, E. V. (1994). Molecular biology and evolution of closteroviruses: sophisticated build-up of large RNA genomes. *Ann. Rev. Phytopath.*, *32*, 261-285.

Duan et al., Y. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comp. Chem.*, *24*, 1999-2012.

Dupureur, C. M., & Barton, J. K. (1994). Use of selective deuteration and 1H NMR in demonstrating major groove binding of Δ-$[Ru(phen)_2dppz]^{2+}$ to $d(GTCGAC)_2$. *J. Am. Chem. Soc.*, *116*, 10286-10287.

Egelman, E. H., & Stasiak, A. (1993). Electron microscopy of RecA-DNA complexes: two different states, their functional significance and relation to the solved crystal structure. *Micron*, *24*, 309-324.

Egelman, E. (2000). A common structural core in proteins active in DNA recombination and replication. *Trends in biochemical sciences*, *25*(4), 179-82.

Egli, M., Williams, L. D., Frederick, C. A., & Rich, A. (1991). DNA-nogalamycin interactions. *Biochem.*, *30*, 1364-1372.

Elias, I. (2006). Settling the intractability of multiple alignment. *J. Comp. Biol.: J. Comp. Mol. Cell Biol.*, *13*, 1323-1339.

Eriksson, M., Leijon, M., Hiort, C., Nordén, B., & Graslunds, A. (1994). Binding of Δ- and Λ-$[Ru(phen)_3]^{2+}$ to $[d(CGCGATCGCG)]_2$ studied by NMR. *Biochem.*, *33*, 5031-5040.

Fire, A. (1999). RNA-triggered gene silencing. *Tren. Genet.*, *15*, 358-363.

Fiser, A., Do, R. K., & Sali, A. (2000). Modeling of loops in protein structures. *Prot. Sci.: Publ. Prot. Soc.*, *9*, 1753-1773.

Fock, V. (1930). Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik A: hadrons and nuclei*, *61*, 126-148.

Foloppe, N., & Mackerell, Jr., A. D. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, *21*, 86-104.

Foloppe, N., Hartmann, B., Nilsson, L., & Mackerell, A. D. (2002). Intrinsic conformational energetics associated with the glycosyl torsion in DNA: a quantum mechanical study. *Biophys. J.*, *82*, 1554-1569.

Fox, K. R., & Waring, M. J. (1984). Evidence of different binding sites for nogalamycin in DNA revealed by association kinetics. *Biochim. Biophys. Acta*, *802*, 162-168.

Fox, K. R., Brassett, C., & Waring, M. J. (1985). Kinetics of dissociation of nogalamycin from DNA: comparison with other anthracycline antibiotics. *Biochim. Biophys. Acta*, *840*, 383-392.

Fraenkel, E., Rould, M. A., Chambers, K. A., & Pabo, C. O. (1998). Engrailed Homeo-domain-DNA complex at 2.2Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.*, *284*, 351-361.

Franklin, R. E., & Gosling, R. G. (1953). Molecular configuration in sodium thymo-nucleate. *Nature*, *171*, 740-741.

Freddolino, P. L., Arkhipov, A. S., Larson, S. B., mcpherson, A., & Schulten, K. (2006). Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, *14*, 437–449.

Friedman, A.E., Chambron, J. C., Sauvage, J., Turro, N. J., & Barton, J. K. (1990). Molecular "light-switch" for DNA: $Ru(bpy)_2(dppz)^{2+}$. *J. Am. Chem. Soc.*, *112*, 4960-4962.

Frisch M. J. et al., (2003). Gaussian03. *Gaussian, Inc., Wallingford CT*.

Frykholm, K., Morimatsu, K., & Nordén, B. (2006). Conserved conformation of RecA protein after executing the DNA strand-exchange reaction. A Site-Specific linear dichroism structure study. *Biochem.*, *45*, 11172-11178.

Frykholm, K. (2010). *Aspects of DNA Strand Exchange: Recombination Proteins and Model System Studies*. PhD thesis. Chalmers University of Technology, Göteborg.

Galkin, V. E., Esashi, F., Yu, X., Yang, S., West, S. C., & Egelman, E. H. (2005). BraA2 Brc motifs bind Rad51-DNA filaments. *Proc. Natl. Acad. Sci. U.S.A.*, *102*, 8537-8542.

Gao, Q., Williams, L. D., Egli, M., Rabinovich, D., Chen, S. L., Quigley, G. J., & Rich, A. (1991). Drug-induced DNA repair: X-ray structure of a DNA-ditercalinium complex. *Proc. Natl. Acad. Sci. U.S.A.*, *88*, 2422-2426.

Gao, Y. G., Liaw, Y. C., Robinson, H., & Wang, A. H. (1990). Binding of the antitumor drug nogalamycin and its derivatives to DNA: structural comparison. *Biochem.*, *29*, 10307-10316.

Garcia, A. E., & Paschek, D. (2008). Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. *J. Am. Chem. Soc.*, *130*, 815-817.

García, A. E., & Onuchic, J. N. (2003). Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. U.S.A.*, *100*, 13898-13903.

Gisselfält, K., Lincoln, P., Nordén, B., & Jonsson, M. (2000). Interactions of tris(phenanthroline)ruthenium(II) enantiomers with DNA: effects on helix flexibility studied by the electrophoretic behaviour of reptating DNA in agarose gel. *J. Phys. Chem. B*, *104*, 3651-3659.

Giudice, E., Várnai, P., & Lavery, R. (2001). Energetic and conformational aspects of A:T base-pair opening within the DNA double helix. *ChemPhysChem*, *2*, 673-677.

Gohlke, H., & Case, D. A. (2004). Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comp. Chem.*, *25*, 238-250.

Griffiths, D. J. (1995). *Introduction to Quantum Mechanics* (p. 394). Prentice Hall.

Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y. G., & Kim, K. K. (2005). Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature*, *437*, 1183-1186.

Hall, J. P., Sullivan, K. O., Naseer, A., Smith, J. A., Kelly, J. M., & Cardin, C. J. (2011). Structure determination of an intercalating ruthenium dipyridophenazine complex which kinks DNA by semiintercalation of a tetraazaphenanthrene ligand. *Proc. Natl. Acad. Sci. U.S.A.*, *108*, 17610-17614.

Hay, P. J., & Wadt, W. R. (1985). Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. *J. Chem. Phys.*, *82*, 270.

Hay, B. P. (1993). Methods for molecular mechanics modelling of coordination compounds. *Coord. Chem. Rev.*, *126*, 177-236.

Hehre, W. J., Radom, L., Schleyer, P. V., & Pople, J. (1986). *Ab initio molecular orbital theory* (p. 576). John Wiley & Sons.

Higgins, D. G., & Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, *73*, 237-244.

Hikiba, J., Hirota, K., Kagawa, W., Ikawa, S., Kinebuchi, T., Sakane, I., Takizawa, Y., et al. (2008). Structural and functional analyses of the Dmc1-M200V polymorphism found in the human population. *Nucl. Ac. Res.*, *36*, 4181-4190.

Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Quart. Rev. Biol.*, *66*, 411-453.

Himo, F. (2005). Quantum chemical modeling of enzyme active sites and reaction mechanisms. *Theor. Chem. Acc.*, *116*(1-3), 232-240.

Hiort, C., Lincoln, P., & Nordén, B. (1993). DNA binding of Δ- and Λ-$[Ru(phen)_2dppz]^{2+}$. *J. Am. Chem. Soc.*, *115*, 3448-3454.

Hiort, C., Nordén, B., & Rodger, A. (1990). Enantiopreferential DNA binding of [Ru(II)(1,10-phenanthroline)$_3$]$^{2+}$ studied with linear and circular dichroism. *J. Am. Chem. Soc.*, *112*, 1971-1982.

Hogeweg, P. (1978). Simulating the growth of cellular forms. *Simulation*, *31*, 90-96.

Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Phys. Rev.*, *136*, B864-B871.

Holmén, A., Albinsson, B., & Nordén, B. (1994). Electronic Transition Dipole Moments of the 1,N6-Ethenoadenine Chromophore. *The Journal of Physical Chemistry*, *98*(51), 13460-13469. doi:10.1021/j100102a006

Holmén, A., Broo, A., Albinsson, B., & Nordén, B. (1997). Assignment of Electronic Transition Moment Directions of Adenine from Linear Dichroism Measurements. *Journal of the American Chemical Society*, *119*(50), 12240-12250. doi:10.1021/ja9710262

Holmlin, R. E., Stemp, E. D. A., & Barton, J. K. (1998). $Ru(phen)_2dppz^{2+}$ luminescence: dependence on DNA sequences and groove-binding agents. *Inorg. Chem.*, *37*, 29-34.

Hooker, T. M., & Schellman, J. A. (1970). Optical activity of aromatic chromophores. I. O, m, and p-tyrosine. *Biopolymers*, *9*(11), 1319-48. doi:10.1002/bip.1970.360091104

Hoover, W. (1985). Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, *31*, 1695-1697.

Hou, T., Zhu, L., Chen, L., & Xu, X. (2003). Mapping the binding site of a large set of quinazoline type EGF-R inhibitors using molecular field analyses and molecular docking studies. *J. Chem. Inform. Comp. Sci.*, *43*, 273-287.

Hrušák, J., Ten-no, S., & Iwata, S. (1997). Quadratic configuration interaction versus coupled-cluster theory: Importance of orbital relaxation phenomena in Cuh and Cuf. *J. Chem. Phys.*, *106*, 7185.

Hummer, G., & Szabo, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U.S.A.*, *98*, 3658-3661.

Israelewitz B., Gao, M., & Schulten, K. (2001). Steered molecular dynamics and mechanical functions of proteins. *Curr. Op. Struc. Biol.*, *11*, 224-230.

Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., & Skeel, R. D., (2001). Langevin stabilization of molecular dynamics. *J. Chem. Phys., 114*, 2090.

Izrailev, S., Stepaniants, S., Isralewitz, B., Kosztin, D., Lu, H., Molnar, F., Wriggers, W., et al. (1998). Steered molecular dynamics. *Science*, *4*, 39-65.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, *78*, 2690-2693.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, *71*, 4546-4554.

Jenkins, Y., Friedman, A. E., Turro, N. J., & Barton, J. K. (1992). Characterization of dipyridophenazine complexes of ruthenium(II): the light switch effect as a function of nucleic acid sequence and conformation. *Biochem.*, *31*, 10809-10816.

Jensen, M. Ø., Park, S., Tajkhorshid, E., & Schulten, K. (2002). Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U.S.A.*, *99*, 6731-6736.

Jain, S. K., Cox, M. M., & Inman, R. B. (1994). On the role of ATP hydrolysis in RecA protein-mediated DNA strand exchange. III. Unidirectional branch migration and extensive hybrid DNA formation. *J. Biol. Chem.*, *269*(32), 20653-61.

Jorgensen, W. L. (1991). Rusting of the lock and key model for protein-ligand binding. *Science*, *254*, 954-955.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, *118*, 11225-11236.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, *181*, 662-666.

Kim, J., Cox, M., & Inman, R. B. (1992). On the Role of ATP Hydrolysis in RecA Protein-mediated DNA Strand Exchange. *Biochemistry*, 16438-16443.

Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, *365*, 520-527.

Kim, Y., Geiger, J. H., Hahn, S., & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, *365*, 512-520.

Kinebuchi, T., Kagawa, W., Enomoto, R., Tanaka, K., Miyagawa, K., Shibata, T., Kurumizaka, H., et al. (2004). Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein Dmc1. *Mol. Cell*, *14*, 363-374.

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comp. Chem.*, *29*, 622-655.

Klass, J., Iv, F. V. M., Fouts, S., Serenil, M., Changela, A., Siple, J., & Churchill, M. E. A. (2003). The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity. *Nucl. Ac. Res.*, *31*, 2852-2864.

Klug, A. (2010). The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Quarterly reviews of biophysics*, *43*(1), 1-21.

Kohn, W., & Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Physic. Rev.*, *140*, A1133-A1138.

Kollman, P. A. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.*, *93*, 2395-2417.

Kollman, P. A., Massova, I., Reyes, C. M., Kuhn, B., Huo, S., Chong, L., Lee, M., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, *33*, 889-897.

Konforti, B. B., & Davis, R. W. (1992). ATP hydrolysis and the displaced strand are two factors that determine the polarity of RecA-promoted DNA strand exchange. *J. Mol. Biol.*, *227*(1), 38-53.

Kowalczykowski, S. C. (2000). Initiation of genetic recombination and recombination-dependent replication. *Tren. Biochem. Sci.*, *25*, 156-165.

Kowalczykowski, S. C., & Krupp, R. A. (1995). DNA-strand exchange promoted by RecA protein in the absence of ATP: implications for the mechanism of energy transduction in protein-promoted nucleic acid transactions. *Proc. Natl. Acad. Sci. U.S.A.*, *92*, 3478-3482.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., & Rehrauer, W. M. (1994). Biochemistry of homologous recombination in Escherichia coli. *Microbiol. Rev.*, *58*, 401-465.

Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K., & Vijayan, M. (2006). Crystallographic identification of an ordered C-terminal domain and a second nucleotide-binding site in RecA: new insights into allostery. *Nucl. Ac. Res.*, *34*, 2186-2195.

Krishna, R., Prabu, J. R., Manjunath, G. P., Datta, S., Chandra, N. R., Muniyappa, K., & Vijayan, M. (2007). Snapshots of RecA protein involving movement of the C-domain and different conformations of the DNA-binding loops: crystallographic and comparative analysis of 11 structures of Mycobacterium smegmatis RecA. *J. Mol. Biol.*, *367*, 1130-1144.

Kubista, M., Åkerman, B., & Nordén, B. (1987). Characterization of interaction between DNA and 4',6-diamidino-2-phenylindole by optical spectroscopy. *Biochem.*, *26*, 4545-4553.

Kuhn, B., & Kollman, P. A. (2000). Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.*, *43*, 3786-3791.

Kuhn, B., Gerber, P., Schulz-Gasch, T., & Stahl, M. (2005). Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.*, *48*, 4040-4048.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Research Comm.*, *95*, 1-6.

Kumar, C. V., Barton, J. K., & Turro, N. J. (1985). Photophysics of ruthenium complexes bound to double helical DNA. *J. Am. Chem. Soc.*, *107*, 5518-5523.

Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. *Science*, *257*, 1078-1082.

Kurumizaka, H, Aihara, H., Ikawa, S., Kashima, T., Bazemore, L. R., Kawasaki, K., Sarai, A., et al. (1996). A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. *J. Biol. Chem.*, *271*, 33515-33524.

Landis, C. R., Root, D. M., & Cleveland, T. (1995). Molecular mechanics force fields for modeling inorganic and organometallic compounds. In K. B. Lipkowitz & D. B. Boyd (Eds.), *Rev. Comp. Chem.,* (Vol. 6). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., mcgettigan, P. A., mcwilliam, H., Valentin, F., et al. (2007). ClustalW and ClustalX version 2.0. *Bioinform.*, *23*, 2947-2948.

Laugaa, P., Delbarre, A., Le Pecq, J. B., & Roques, B. P. (1983). Comparative binding of ethidium and three azido analogs to dinucleotides: affinity and intercalation geometry. A 1H NMR and visible spectroscopy study. *Europ. J. Biochem. FEBS*, *134*, 163-173.

Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., & Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucl. Ac. Res.*, *37*, 5917-5927.

Leach, A. R. (2001). *Molecular Modeling: Principles and Applications* (2nd ed., 744 p.). Harlow, England: Pearson Education Limited.

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, *37*, 785-789.

Lee, B., & Graziano, G. (1996). A Two-State Model of Hydrophobic Hydration That Produces Compensating Enthalpy and Entropy Changes. *Journal of the American Chemical Society*, *118*(22), 5163-5168.

Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. *Curr. Op. Struc. Biol.*, *6*, 402-406.

Lerman, L. S. (1961). Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.*, *3*, 18-30.

Levitt, M. (1983). Computer simulation of DNA double-helix dynamics. *Cold Spring Harbor Symposia on Quantitative Biology*, *47*, 251-262.

Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, *61*(4), 704-21.

Li, Y., He, Y., & Luo, Y. (2009). Crystal structure of an archaeal Rad51 homologue in complex with a metatungstate inhibitor. *Biochem.*, *48*, 6805-6810.

Lifton, R. P., Goldberg, M. L., Karp, R. W., & Hogness, D. S. (1978). The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. *Cold Spring Harbor Symposia on Quantitative Biology*, *42*, 1047-1051.

Lincoln, P., & Nordén, B. (1996). Binuclear ruthenium(II)phenanthroline compounds with extreme binding affinity for DNA. *Chem. Comm.*, 2145.

Lincoln, P., & Nordén, B. (1998). DNA binding geometries of ruthenium(II) complexes with 1,10-Phenanthroline and 2,2′-Bipyridine ligands studied with linear dichroism spectroscopy. Borderline cases of intercalation. *J. Am. Chem. Soc.*, *102*, 9583-9594.

Lincoln, P., Broo, A., & Nordén, B. (1996). Diastereomeric DNA-binding geometries of intercalated ruthenium(II)trischelates probed by linear dichroism: $Ru(phen)_2dppz$ and $Ru(phen)_2bdppz$. *J. Am. Chem. Soc.*, *118*, 2644-2653.

Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, *227*, 1435-1441.

Lippman, Z., & Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. *Nature*, *431*, 364-370.

Liu, L., Maguire, K. K., & Kmiec, E. B. (2004). Genetic re-engineering of Saccharomyces cerevisiae RAD51 leads to a significant increase in the frequency of gene repair in vivo. *Nucleic acids research*, *32*(7), 2093-2101.

Liu, Y., Hsin, J., Kim, H., Houdusse, A., Sweeney, H. L., Selvin, P. R., & Schulten, K. (2010). Steered molecular dynamics simulation of unfolding of myosin VI proximal tail domain. *Biophys. J. 98*, 724a.

Lu, H., Isralewitz, B., Krammer, A., Vogel, V., & Schulten, K. (1998). Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, *75*, 662-671.

Lu, X. J., Shakked, Z., & Olson, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.*, *300*, 819–840.

Lusetti, S. L., & Cox, M. M. (2002). The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Ann. Rev. Biochem.*, *71*, 71-100.

Lusetti, S. L., Shaw, J. J., & Cox, M. M. (2003). Magnesium ion-dependent activation of the RecA protein involves the C-terminus. *J. Biol. Chem.*, *278*, 16381-16388.

Löwdin, P.-O. (1950). On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J. Chem. Phys.*, *18*, 365.

Maacke, H., Opitz, S., Jost, K., Hamdorf, W., Henning, W., Krüger, S., Feller, a C., et al. (2000). Over-expression of wild-type Rad51 correlates with histological grading of invasive ductal breast cancer. *Internatl. J. Cancer. J. Internl. du Cancer*, *88*, 907-913.

Mackerell, Jr., A. D., Bashford, D., Dunbrack,, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, *102*, 3586-3616.

Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: overview and issues. *J. Comp. Chem.*, *25*, 1584-1604.

Matson, M., Svensson, F. R., Nordén, B., & Lincoln, P. (2011). Correlation between cellular localization and binding preference to RNA, DNA, and phospholipid membrane for luminescent ruthenium(II) complexes. *J. Phys. Chem. B*, *115*, 1706-1711.

Matsuo, Y., Sakane, I., Takizawa, Y., Takahashi, M., & Kurumizaka, H. (2006). Roles of the human Rad51 L1 and L2 loops in DNA binding. *FEBS J.*, *273*, 3148-3159.

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, *267*, 585-590.

McEntee, K. (1976). Specialized transduction of RecA by bacteriophage lambda. *Virology*, *70*, 221-222.

McEntee, K., Weinstock, G. M., & Lehman, I. R. (1979). Initiation of general recombination catalyzed in vitro by the RecA protein of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, *76*, 2615-2619.

Menetski, J. P., & Kowalczykowski, S. C. (1985). Interaction of recA protein with single-stranded DNA: Quantitative aspects of binding affinity modulation by nucleotide cofactors. *J. Mol. Biol.*, *181*(2), 281-295.

Menetski, J. P., & Kowalczykowski, S. C. (1989). Enhancement of Escherichia coli RecA protein enzymatic function by dATP. *Biochem.*, *28*, 5871-5881.

Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (1992). Automated docking with grid-based energy evaluation. *J. Comp. Chem.*, *13*, 505-524.

Michl, J., Thulstrup, E. W., & Eggers, J. H. (1970). Polarization spectra in stretched polymer sheets. Physical significance of the orientation factors and determination of π-π* transition moment directions in molecules of low symmetry. *J. Phys. Chem.*, *74*, 3878-3884.

Mitsutake, A., Sugita, Y., & Okamoto, Y. (2001). Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopol.*, *60*, 96-123.

Moss, R. E. (1973). *Advanced Molecular Quantum Mechanics: Introduction to Relativistic Quantum Mechanics and the Quantum Theory of Radiation (Study in Chemical Physics)* (p. 280). Chapman and Hall.

Mu, Y., Nordenskiöld, L., & Tam, J. P. (2006). Folding, misfolding, and amyloid protofibril formation of WW domain FBP28. *Biophys. J.*, *90*, 3983-3992.

Mukherjee, A., Lavery, R., Bagchi, B., & Hynes, J. T. (2008). On the molecular mechanism of drug intercalation into DNA: a simulation study of the intercalation pathway, free energy, and DNA structural changes. *J. Am. Chem. Soc.*, *130*, 9747-9755.

Mulliken, R. S. (1955). Electronic Population Analysis on LCAO(Single Bond)MO molecular wave functions. I. *J. Chem. Phys.*, *23*, 1833.

Møller, C., & Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Phys. Rev.*, *46*, 618-622.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, *48*, 443-453.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., & Burley, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. U.S.A.*, *93*, 4862-4867.

Nomme, J., Takizawa, Y., Martinez, S. F., Renodon-Cornière, A., Fleury, F., Weigel, P., Yamamoto, K., et al. (2008). Inhibition of filament formation of human Rad51 protein by a small peptide derived from the Brc-motif of the Brca2 protein. *Genes to cells : devoted to molecular & cellular mechanisms*, *13*, 471-481.

Nordell, P. (2009). *Kinetic Recognition of Nucleic Acids: Studies on the DNA Binding Selectivity of Threading Ruthenium Complexes*. Chalmers University of Technology.

Nordell, P., & Lincoln, P. (2005). Mechanism of DNA threading intercalation of binuclear Ru complexes: uni- or bimolecular pathways depending on ligand structure and binding density. *J. Am. Chem. Soc.*, *127*, 9670-9671.

Nordell, P., Westerlund, F., Wilhelmsson, L. M. M., Nordén, B., & Lincoln, P. (2007). Kinetic recognition of AT-rich DNA by ruthenium complexes. *Ang. Chem.*, *119*, 2253-2256.

Nordén, B. (1973). Detection of π-π* transitions in pyridine and pyrazine in polyethylene solution by linear dichroism. *Chem. Phys. Lett.*, *23*, 200-202.

Nordén, B., & Davidsson, Å. (1972). Polarized absorption spectra of benzene, naphthalene and anthracene obtained with a high-sensitive linear dichroism technique. *Tetrahedron Lett.*, *30*, 3093.

Nordén, B., & Tjerneld, F. (1976). High-sensitivity linear dichroism as a tool for equilibrium analysis in biochemistry- stability constant of DNA-ethidiumbromide complex. *Biophys. Chem.*, *4*, 191-198.

Nordén, B., & Tjerneld, F. (1976). Binding of inert metal complexes to deoxyribonucleic acid detected by linear dichroism. *FEBS lett.*, *67*, 368-370.

Nordén, B., & Tjerneld, F. (1977). Binding of methyl green to deoxyribonucleic acid analyzed by linear dichroism. *Chem. Phys. Lett.*, *50*, 508-512.

Nordén, B., Kubista, M., & Kurucsev, T. (1992). Linear dichroism spectroscopy of nucleic acids. *Quart. Rev. Biophys.*, *25*(1), 51-170.

Nordén, B., Lindblom, G. & Jonas, I. (1977) Linear dichroism spectroscopy as a tool for studying molecular orientation in model membrane systems. *J. Phys. Chem.*, 81, 2086-2093.

Norrby, P. O., & Liljefors, T. (1998). Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data. *J. Comp. Chem.*, *19*, 1146-1166.

Ogawa, T., Yu, X., Shinohara, A., & Egelman, E. H. (1993). Similarity of the yeast Rad51 filament to the bacterial RecA filament. *Science*, *259*, 1896-1899.

Ohnishi, T., Taki, T., Hiraga, S., Arita, N., & Morita, T. (1998). In vitro and in vivo potentiation of radiosensitivity of malignant gliomas by antisense inhibition of the Rad51 gene. *Biochem. Biophys. Res. Comm.*, *245*, 319-324.

Olson, W. K., & Sussman, J. L. (1982). How flexible is the furanose ring? A comparison of experimental and theoretical studies. *J. Am. Chem. Soc.*, *104*, 270-278.

Olson, W. K., Gorin, A. A., Lu, X.J., Hock, L. M., & Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.*, *95*, 11163-11168.

Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comp. Chem.*, *25*, 1656-1676.

Palasingam, P., Jauch, R., Keow, C., Ng, L., & Kolatkar, P. R. (2009). The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *J. Mol. Biol.*, *388*, 619-630.

Paramanathan, T., Westerlund, F., mccauley, M. J., Rouzina, I., Lincoln, P., & Williams, M. C. (2008). Mechanically manipulating the DNA threading intercalation rate. *J. Am. Chem. Soc.*, *130*, 3752-3753.

Pearlman, D. A. (2005). Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a congeneric series of ligands to P38 Map kinase. *J. Med. Chem.*, *48*, 7796-7807.

Pellegrini, L., Yu, D. S., Lo, T., Anand, S., Lee, M., Blundell, T. L., & Venkitaraman, A. R. (2002). Insights into DNA recombination from the structure of a Rad51-Brca2 complex. *Nature*, *420*, 287-293.

Phillips J. C. et al. (2005). Scalable molecular dynamics with NAMD. *J. Comp. Chem.*, *26*, 1781-1802.

Pierre, V. C., Kaiser, J. T., & Barton, J. K. (2007). Insights into finding a mismatch through the structure of a mispaired DNA bound by a rhodium intercalator. *Proc. Natl. Acad. Sci. U.S.A.*, *104*, 429-434.

Pilar, F. L. (1990). *Elementary Quantum Chemistry*. Mcgraw-Hill.

Pjura, P. E., Grzeskowiak, K., & Dickerson, R. E. (1987). Binding of Hoechst 33258 to the minor groove of B-DNA. *J. Mol. Biol.*, *197*, 257-271.

Pople, J. A., Head-Gordon, M., & Raghavachari, K. (1987). Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.*, *87*, 5968.

Postma, J. P. M., Berendsen, H. J. C., & Haak, J. R. (1982). Thermodynamics of cavity formation in water. A molecular dynamics study. *Faraday Symp. Chem. Soc.*, *17*, 55-67.

Prabu, J. R., Manjunath, G. P., Chandra, N. R., Muniyappa, K., & Vijayan, M. (2008). Functionally important movements in RecA molecules and filaments: studies involving mutation and environmental changes. *Acta Crystal. D, Biol. Crystal.*, *64*, 1146-1157.

Pradhan, S. M., Katti, D. R., & Katti, K. S. (2011). Steered molecular dynamics study of mechanical response of full length and short collagen molecules. *J. Nanomech. Micromech.*, *1*, 104.

Prasad, T. K., Yeykal, C. C., & Greene, E. C. (2006). Visualizing the assembly of human Rad51 filaments on double-stranded DNA. *J. Mol. Biol.*, *363*, 713-728.

Pyykkö, P. (1988). Relativistic effects in structural chemistry. *Chem. Rev.*, *88*, 563-594.

Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A, & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, *92*, 3817-3829.

Qian, X., He, Y., Ma, X., Fodje, M. N., Grochulski, P., & Luo, Y. (2006). Calcium stiffens archaeal Rad51 recombinase from Methanococcus voltae for homologous recombination. *J. Biol. Chem.*, *281*, 39380-39387.

Qian, X., Wu, Y., He, Y., & Luo, Y. (2005). Crystal structure of Methanococcus voltae RadA in complex with ADP: hydrolysis-induced conformational change. *Biochem.*, *44*, 13753-13761.

Qin, S., & Zhou, H.-xiang. (2007). Do electrostatic interactions destabilize protein–nucleic acid binding? *Biopol.*, *86*, 112-118.

Rabi, I., Zacharias, J., Millman, S., & Kusch, P. (1938). A new method of measuring nuclear magnetic moment. *Phys. Rev.*, *53*, 318-318.

Rehrauer, W. M., & Kowalczykowski, S. C. (1993). Alteration of the nucleoside triphosphate (NTP) catalytic domain within Escherichia coli RecA protein attenuates NTP hydrolysis but not joint molecule formation. *J. Biol. Chem.*, *268*, 1292-1297.

Reinhardt, C. G., & Krugh, T. R. (1978). A comparative study of ethidium bromide complexes with dinucleotides and DNA: direct evidence for intercalation and nucleic acid sequence preferences. *Biochem.*, *17*, 4845-4854.

Renodon-Cornière, A., Takizawa, Y., Conilleau, S., Tran, V., Iwai, S., Kurumizaka, H., & Takahashi, M. (2008). Structural analysis of the human Rad51 protein-DNA complex filament by tryptophan fluorescence scanning analysis: transmission of allosteric effects between ATP binding and DNA binding. *J. Mol. Biol.*, *38*, 575-587.

Rich, A., & Shuguang, Z. (2003). Z-DNA : the long road to biological function. *Nature Rev. Genetics*, *4*, 566-573.

Ristic, D., Modesti, M., van der Heijden, T., van Noort, J., Dekker, C., Kanaar, R., & Wyman, C. (2005). Human Rad51 filaments on double- and single-stranded DNA: correlating regular and irregular forms with recombination function. *Nucl. Ac. Res.*, *33*, 3292-3302.

Robinson, H.; Gao, Y. G.; mccrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. J. (1998). The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Lett. Nature*, *392*, 202-205.

Robinson, H., Gao, Y. G., Yang, X., Sanishvili, R., Joachimiak, A., & Wang, A. H. (2001). Crystallographic analysis of a novel complex of actinomycin D bound to the DNA decamer CGATCGATCG. *Biochem.*, *40*, 5587-5592.

Roca, A. I., & Cox, M. M. (1997). RecA protein: structure, function, and role in recombinational DNA repair. *Prog. Nucl. Ac. Res. Mol. Biol.*, *56*, 129-223.

Rosselli, W., & Stasiak, A. (1990). Energetics of RecA-mediated recombination reactions. Without ATP hydrolysis RecA can mediate polar strand exchange but is unable to recycle. *J. Mol. Biol.*, *216*(2), 335-52.

Rosselli, W., & Stasiak, A. (1991). The ATPase activity of RecA is needed to push the DNA strand exchange through heterologous regions. *EMBO Journal*, *10*(13), 4391-4396.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, *23*, 327-341.

Sagui, C., & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Ann. Rev. Biophys. Biomol. Struc.*, *28*, 155-179.

Sali, A., & Blundell, T. L. (1995). Comparative protein modelling by satisfaction of spatial restraints. *Mol. Med. Today*, *1*, 270-277.

San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Ann. Rev. Biochem.*, *77*, 229-257.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., et al. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, *265*, 687-695.

Satyanarayana, S., Dabrowiak, J. C., & Chaires, J. B. (1992). Neither Λ- nor Δ-tris(phenanthroline)ruthenium(II) binds to DNA by classical intercalation. *Biochem.*, *31*, 9319-9324.

Satyanarayana, S., Dabrowiak, J. C., & Chaires, J. B. (1993). Tris(phenanthroline)-ruthenium(II) enantiomer interactions with DNA: mode and specificity of binding. *Biochem.*, *32*, 2573-2584.

Schlegel, H. B. (1995). Geometry optimization on potential energy surfaces. In D. Yarkony (Ed.), *Modern electronic structure theory, Part 1* (pp. 459-500). World Scientific Publishing.

Schlegel, H. B. (1998). Geometry Optimization: 1. *Encyclopaedia of Computational Chemistry*. Wiley.

Schuetz, A., Nana, D., Rose, C., Zocher, G., Milanovic, M., Koenigsmann, J., Blasig, R., et al. (2011). The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cell. Mol. Life Sci.: CMLS*, *68*, 3121-3131.

Sehorn, M. G., Sigurdsson, S., Bussen, W., Unger, V. M., & Sung, P. (2004). Human meiotic recombinase Dmc1 promotes ATP-dependent homologous dna strand exchange. *Nature*, *429*, 433-437.

Seitz, E. M., Brockman, J. P., Sandler, S. J., Clark, A. J., & Kowalczykowski, S. C. (1998). RadA protein is an archaeal RecA protein homolog that catalyses DNA strand exchange. *Genes & Devel.*, *12*, 1248-1253.

Selmane, T., Camadro, J.M., Conilleau, S., Fleury, F., Tran, V., Prévost, C., & Takahashi, M. (2004). Identification of the subunit-subunit interface of xenopus Rad51 protein: similarity to RecA. *J. Mol. Biol.*, *335*, 895-904.

Shan, Q., Cox, M. M., & Inman, R. B. (1996). DNA strand exchange promoted by RecA K72R. Two reaction phases with different Mg2+ requirements. *J. Biol. Chem.*, *271*(10), 5712-24.

Shin, D. S., Pellegrini, L., Daniels, D. S., Yelent, B., Craig, L., Bates, D., Yu, D. S., et al. (2003). Full-length archaeal Rad51 structure and mutants: mechanisms for Rad51 assembly and control by Brca2. *EMBO J.*, *22*, 4566-4576.

Shinohara, A., Ogawa, H., & Ogawa, T. (1992). Rad51 protein involved in repair and recombination in S. cerevisiae is a Rec-like protein. *Cell*, *69*, 457-470.

Shoichet, B. K., Kuntz, I. D., & Bodian, D. L. (1992). Molecular docking using shape descriptors. *J. Comp. Chem.*, *13*, 380-397.

Siegbahn, P. E. M. (1992). The configuration interaction method. In B. O. Roos (Ed.), *Lecture Notes in Quantum Chemistry*. Springer.

Simmerling, C., Strockbine, B., & Roitberg, A. E. (2002). All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, *124*, 11258-11259.

Singh, U. C., & Kollman, P. A. (1984). An approach to computing electrostatic charges for molecules. *J. Comp . Chem.*, *5*, 129-145.

Singh, N., & Warshel, A. (2010). A comprehensive examination of the contributions to the binding entropy of protein-ligand complexes. *Proteins*, *78*, 1724-1735.

Slater, J. C. (1930). Note on Hartree's Method. *Phys. Rev.*, *35*, 210-211.

Smale, S. T., & Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Ann. Rev. Biochem.*, *72*, 449-479.

Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular sub-sequences. *J. Mol. Biol.*, *147*, 195-197.

Sobell, H. M. (1985). Actinomycin and DNA transcription. *Proc. Natl. Acad. Sci. U.S.A.*, *82*, 5328-5331.

Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., & Case, D. A. (1998). Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc.*, *120*, 9401-9409.

Stephens, P. J., Devlin, F. J., Chabalowski, C. F., & Frisch, M. J. (1994). Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, *98*, 11623-11627.

Story, R. M. M., & Steitz, T. A. (1992). Structure of the RecA protein-ADP Complex. *Nature*, *355*, 374-376.

Story, R. M., Weber, I. T., & Steitz, T. A. (1992). The Structure of the E. Coli RecA protein Monomer and Polymer. *Nature*, *355*, 318-325.

Straatsma, T. P., & McCammon, J. A. (1992). Computational alchemy. *Ann. Rev. Phys. Chem.*, *43*, 407-435.

Straney, D. C., & Crothers, D. M. (1987). Effect of drug-DNA interactions upon transcription initiation at the LAC promoter. *Biochem.*, *26*, 1987-1995.

Sugita, Yuji, & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Let.*, *350*, 141-151.

Sung, P., & Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature Rev. Mol. Cell Biol.*, *7*, 739-750.

Szalay, P. G., & Bartlett, R. J. (1995). Approximately extensive modifications of the multireference configuration interaction method: a theoretical and practical analysis. *J. Chem. Phys.*, *103*, 3600.

Tamura, H., Ikeda, N., Iguro, T., Ohno, T., & Matsubayashi, G. E. (1996). The pseudo-racemic complex bis[tris(2,2'-bipyridine)ruthenium(II)] hexacyanocobaltate(III) chloride octahydrate, $[Ru(bpy)_3]^{2+}[Co(CN)_6]Cl_8H_2O$. *Acta Crystal. C, Crystal Struc. Comm.*, *52*, 1394-1399.

Tembre, B. L., & McCammon, J. A. (1984). Ligand-receptor interactions. *Comp. Chem.*, *8*, 281-283.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Ac. Res.*, *22*, 4673-4680.

Tidor, B., Irikura, K. K., Brooks, B. R., & Karplus, M. (1983). Dynamics of DNA oligomers. *J. Biomol. Struc. & Dynam.*, *1*, 231-252.

Tsui, V., Radhakrishnan, I., Wright, P. E., & Case, D. A. (2000). NMR and molecular dynamics studies of the hydration of a zinc finger-DNA complex. *J. Mol. Biol.*, *302*, 1101-1117.

Tuite, E., Lincoln, P., & Nordén, B. (1997). Photophysical evidence that Δ- and Λ-$[Ru(phen)_2dppz]^{2+}$ intercalate DNA from the minor groove. *Biochem.*, *2*, 10286- 10287.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *J. Comp. Chem.*, *26*, 1701-1718.

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, *159*, 98-103.

Vispé, S., & Defais, M. (1997). Mammalian Rad51 protein: a RecA homologue with pleiotropic functions. *Biochimie*, *79*(9-10), 587-92.

Vispé, S., Cazaux, C., Lesca, C., & Defais, M. (1998). Overexpression of Rad51 protein stimulates homologous recombination and increases resistance of mammalian cells to ionizing radiation. *Nucl. Ac. Res.*, *26*, 2859-64.

Viswamitra, M. A., Kennard, O., Jones, P. G., Sheldrick, G. M., Salisbury, S., Falvello, L., & Shakked, Z. (1978). DNA double helical fragment at atomic resolution. *Nature*, *273*, 687-688.

Vos, J. G., & Kelly, J. M. (2006). Ruthenium polypyridyl chemistry; from basic research to applications and back again. *Dalton transactions (Cambridge, England : 2003)*, *41*, 4869-4883.

Vriend, G. (1990). WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.*, *8*, 52-56.

Wada, A., & Kozawa, S. (1964). Instrument for the studies of differential flow dichroism of polymer solutions. *J. Pol. Sci. Part A: General Papers*, *2*, 853-864.

Wang, L., & Jiang, T. (1994). On the complexity of multiple sequence alignment. *J. Comp. Biol.*, *1*, 337-348.

Wang, J., Cieplak, P., & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.*, *21*, 1049.

Wang, W., & Kollman, P. A. (2000). Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.*, *303*, 567-582.

Wang, W., & Kollman, P. A. (2001). Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. U.S.A.*, *98*, 14937-14942.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *J. Comp. Chem.*, *25*, 1157-74.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, *171*, 737-738.

Weber, P. L., Sieker, L. C., Samy, T. S. A., Reid, B. R., & Drobny, G. P. (1987). Two-dimensional coherence transfer NMR spectroscopy by isotropic mixing: application to protein NMR assignments. *J. Am. Chem. Soc.*, *109*, 5842-5844.

Wilhelmsson, L. M., Westerlund, F., Lincoln, P., & Nordén, B. (2002). DNA-binding of semirigid binuclear ruthenium complex Δ,Δ-[μ-(11,11′bidppz)$(phen)_4Ru_2]^{4+}$: extremely slow intercalation kinetics. *J. Am. Chem. Soc.*, *124*, 12092-12093.

Wilhelmsson, L. M., Esbjörner, E. K., Westerlund, F., Nordén, B., & Lincoln, P. (2003). Meso stereoisomer as a probe of enantioselective threading intercalation of semirigid ruthenium complex [μ-(11,11′-bidppz)$(phen)_4Ru_2]^{4+}$. *J. Phys. Chem. B*, *107*, 11784-11793.

Wilson, W. D., Tanious, F. A., Barton, H. J., Jones, R. L., Fox, K., Wydra, R. L., & Strekowski, L. (1990). DNA sequence dependent binding modes of 4′,6-diamidino-2-phenylindole (DAPI). *Biochem.*, *29*, 8452-8461.

Woodcock, L. V. (1971). Isothermal molecular dynamics calculations for liquid salts. *Chem. Phys. Let.*, *10*, 257-261.

Wu, Y., He, Y., Moya, I. A., Qian, X., & Luo, Y. (2004). Crystal structure of archaeal recombinase RadA: a snapshot of its extended conformation. *Mol. Cell*, *15*, 423-435.

Wu, Y., Qian, X., He, Y., Moya, I. A., & Luo, Y. (2005). Crystal structure of an ATPase-active form of Rad51 homolog from Methanococcus voltae. Insights into potassium dependence. *J. Biol. Chem.*, *280*, 722-778.

Xing, X., & Bell, C. E. (2004). Crystal structures of Escherichia coli RecA in complex with Mg-ADP and Mn-AMP-PNP. *Biochem.*, *43*, 16142-16152.

Xing, X., & Bell, C. E. (2004). Crystal structures of Escherichia coli Reca in a compressed helical filament. *J. Mol. Biol.*, *342*, 1471-1485.

Xu, C., Bian, C., Lam, R., Dong, A., & Min, J. (2011). The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nature Comm.*, *2*, 227.

Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., & Martinez, E. (2007). Prevalence of the initiator over the TATA-box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, *389*, 52-65.

Yu, X., & Egelman, E. H. (1997). The RecA hexamer is a structural homologue of ring helicases. *Nat. Struct. Biol.*, *4*(2), 101-104.

Yu, X., Jacobs, S. A., West, S. C., Ogawa, T., & Egelman, E. H. (2001). Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proc. Natl. Acad. Sci. U.S.A.*, *98*, 8419-8424.

Zhang, L., & Hermans, J. (1996). Hydrophilicity of cavities in proteins. *Proteins*, *24*(4), 433-8.

Zhang, C. X., & Lippard, S. J. (2003). New metal complexes as potential therapeutics. *Curr. Op. Chem. Biol.*, *7*, 481-489.

Zhang, Y. (2008). Progress and challenges in protein structure prediction. *Curr. Op. Struc. Biol.*, *18*, 342-348.

Zimmer, M. (1995). Bioinorganic molecular mechanics. *Chem. Rev.*, *95*, 2629-2649.

Önfelt, B., Lincoln, P., & Nordén, B. (1999). A molecular staple for DNA: threading bis-intercalating $[Ru(phen)_2dppz]^{2+}$ dimer. *J. Am. Chem. Soc.*, *121*, 10846-10847.

Önfelt, B., Lincoln, P., & Nordén, B. (2001). Enantioselective DNA threading dynamics by phenazine-linked $[Ru(phen)_2dppz]^{2+}$ dimers. *J. Am. Chem. Soc.*, *123*, 3630-3637.