DNA conformational polymorphism

Habilitation thesis by Lukas Trantirek, 2015

Acknowledgements

I would like to thank all of my past supervisors and mentors who contributed to my scientific education. In particular, I thank Prof. Jaroslav Koca and Prof. Vladimir Sklenar, who supervised me during my undergraduate and graduate studies. I wish to thank my postdoc supervisors, Prof. Juli Feigon and Prof. Norbert Muller, for their support and all of the stimulating discussions we had. Additionally, I would like to thank all of the colleagues and collaborators I had the pleasure to work with, particularly Prof. Vacha, Prof. Sponer, Prof. Vorlickova, Prof. Fajkus, Dr. Sychrovsky, Dr. Hansel, Prof. Dotsch, and Prof. Plavec. Special thanks belong to my wife for all her help, support and inspiration in both my personal as well as scientific life.

Abbreviations

CD	Circular Dichroism
CP	Conformational Polymorphism
DNA	Deoxyribonucleic Acid
EPR	Electron Paramagnetic Resonance
FRET	Förster Resonance Energy Transfer
G4	G-quadruplex
GBA	Glycosidic Bond Angle
NA	Nucleic Acid
NMR	Nuclear Magnetic Resonance
PEG	PolyEthylene Glycol
PELDOR	Paramagnetic Electronic Double Resonance
RDC	Residual Dipolar Coupling
XRD	X-Ray Diffraction

Table of Contents

Appendix

Preface

This habilitation work is a compilation of selected scientific publications to which I have contributed as the corresponding author or a co-author in the course of my independent scientific career. These articles were published between 2005 and 2015. A list of these publications is given on page 45. All of these publications have a common theme related to DNA structural polymorphisms. In particular, they are focused on the mechanistic understanding of the phenomenon of environmentally promoted structural polymorphisms of DNA, context-dependent DNA polymorphisms, as well as the development of novel methods to study DNA polymorphisms. The accompanying text highlights my contribution to the field of DNA structural biology and also contains a brief introduction to the topic. Comprehensive information on the individual topics can be found in the enclosed original publications. The enclosed publications also include three review articles.

1. Introduction

Deoxyribonucleic acid (DNA) is an abundant biopolymer in all living entities, where it functions to encode, transmit, and express genetic information. The physiological functions of DNA are predefined by its conformational plasticity. Disturbances of the native DNA structure correlate with dozens of pathological human conditions, including cancer (1-5).

DNA is a fundamentally attractive drug target. The essence of the "antigene" strategy is that it is advantageous to attack disease targets at their source, at the level of gene expression (6,7). A protein drug target is the product of a particular gene. At each stage of progression through the central dogma (DNA transcription to RNA and the subsequent translation to protein), the absolute number of target molecules to be hit by a drug inhibitor dramatically increases. A single gene makes multiple mRNA copies, each of which is translated to make multiple copies of the target protein. The number of target molecules is amplified at each stage in the process. By targeting the DNA, a single gene, rather than the numerous resulting protein molecules, should promote more selective and efficient drug action. In past two decades, various DNA structural motifs have become valid targets for new anticancer drugs and many leading compounds that target these motifs have entered pre-clinical or clinical trials (8).

In addition to the biological significance of DNA, its unique nano-scale geometry, biocompatibility, biodegradability, and molecular recognition capacity have made it a promising candidate for the construction of novel functional nano-materials and nano-devices (9-15). In addition, site-specific surface modification of NAs enables the presentation of bioactive compounds at defined distances and stochiometries, which has promoted their use in a variety of novel biomedical applications, such as tailored cell targeting and substance delivery on demand (10,14)

DNA structure

From a chemical perspective, DNA can be regarded as a hetero-polymer composed of four basic units (monomers), namely adenosine mono-phosphate (A), cytosine mono-phosphate (C), guanosine mono-phosphate (G), and thymidine mono-phosphate (T). The sequence of monomeric unit defines the DNA primary structure. In terms of its biological function, DNA can be considered a code. While some of the DNA functions are coded by the DNA primary structure alone, such as coding for RNA

transcription, other DNA functions, such as those related to the regulation of gene expression or maintenance of genome integrity, are encoded in its secondary, tertiary, and/or quaternary structure, collectively referred to as higher-order DNA structure. While the primary DNA code is rather simple and comprises only four "letters", the code of the higher order DNA structure is enormously complex. The complexity of this code is due to the inherent structural plasticity of the DNA. In addition to the hetero-dimeric form composed of two DNA strands (in an antiparallel orientation) that are held together by hydrogen bonds, known as B-DNA and the predominant structural arrangement of the genomic DNA, DNA can adopt a broad spectrum of structural motifs. Figure 1 displays some of the basic classes of DNA motifs.

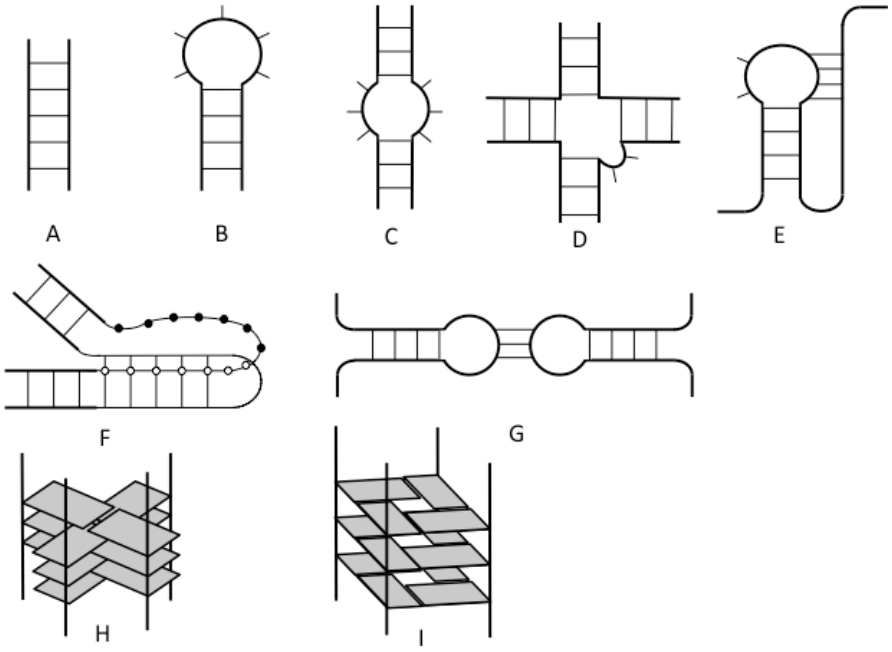

Figure 1. Schematic representations of basic DNA structural motifs: A) parallel and antiparallel duplex, B) hairpin, C) bulge, D) junction, E) pseudo-knot, F) H-DNA, G) kissing-hairpins, H) i-motif, and I) G-quadruplex

Importantly, a number of structurally distinct patterns exist within each of the major classes of DNA motifs. For example, a single DNA sequence forming antiparallel double helix can adopt at least six distinct conformations, referred to as A-, A/B-, hybrid-, composite, B-, and Z-DNA (16).

The complexity of the DNA conformational space can be well illustrated by the G-quadruplex motif. The G-quadruplex (also known as G4-DNA) is a four-stranded structure formed from sequences that are rich in guanine. In G-quadruplex,

four guanine bases align in a pseudo-plane through hydrogen-bond alignments involving the Watson–Crick edge of a guanine and the Hoogsteen edge of its partner, resulting in a (G:G:G:G) tetrad, as shown in Figure 2A. The quadruplex stem is composed of stacked tetrads, with phosphodiester backbones delimiting the cavities, termed grooves. The tetrads are held together by cations and interactions of the π orbitals of stacked aromatic bases. Even very simple G-quadruplex-forming sequences, such as $d(G_2N_3)_4$, where N stands for an arbitrary nucleotide, can, in principle, adopt hundereds of distinct conformations that differ in their strand orientations, loop topologies, and/or disposition of the glycosidic bond angle of the intervening bases, which can assume either an *anti* or a *syn* orientation (Figure 2B & 2C).

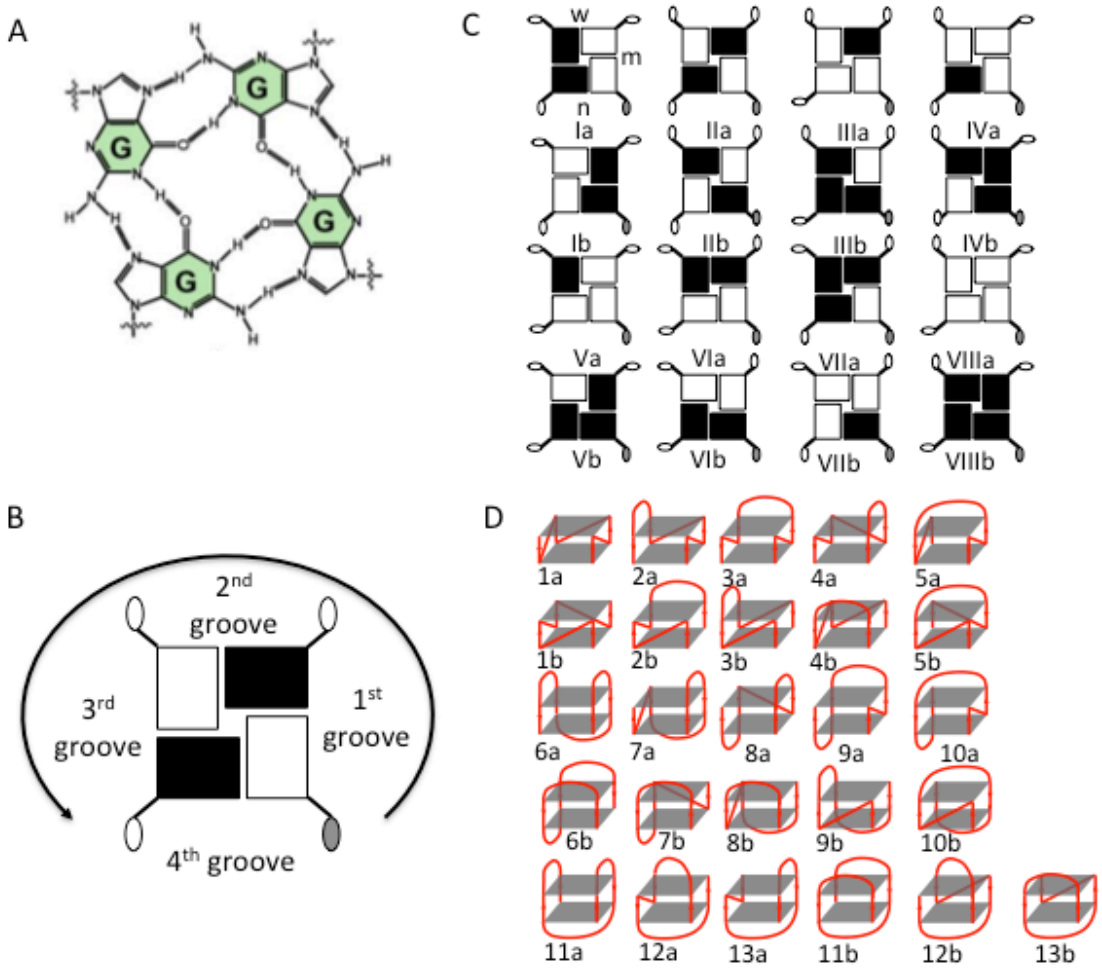

Figure 2. A) Schematic representation of the G-tetrad. B) The frame of reference for describing quadruplex topologies. The origin, the 5'-end, sits on a strand of the quadruplex stem progressing towards the viewer with polarity indicated by grey ellipse on the sugar-pucker. Looping can be anticlockwise, as denoted by the arrow,

or clockwise; (-) and (+), respectively. The description of the grooves follows anticlockwise from first to fourth. Black and white rectangles mark G residues with *syn*- and *anti*- glycosidic bond angle (GBA), respectively. C) All possible combinations of GBA for (G:G:G:G) tetrads. In Ia, the definitions of medium (m), wide (w), and narrow grooves (n) according to the adjacent disposition of GBA are shown. D) Representations of all looping topologies possible for three loop uni-molecular quadruplex topologies. The topologies denoted by "a" start with anticlockwise progressing loops, and conversely the topologies denoted by "b" start with clockwise progressing loops. The sub-figures B-D were adopted from reference (17).

2. DNA conformational polymorphism

Conformational polymorphism is an inherent property of DNA. DNA conformational polymorphism (CP) refers to the ability of a specific DNA sequence to exist in more than one conformation. In structural biochemistry and biophysics, we distinguish three main classes of conformational polymorphisms: i) sequence context-promoted CP, ii) environmentally promoted CP, and iii) CP arising from kinetic partitioning in the course of DNA folding. The borderlines between the individual classes are not strictly defined and a number of DNA sequences display behavior that falls into all three classes. It needs to be mentioned that the classes are primarily defined by the limitations imposed by the experimental procedures used to explore the DNA structure. The individual classes of CP and their implications for fields of structural biology, molecular engineering and DNA drug design are discussed below.

2.1 Conformational polymorphisms due to sequence context

In eukaryotic cells, the DNA is localized in the cell nucleus and mitochondria. At present, there is no available experimental technique that provides high-resolution information on the DNA structure and dynamics in the context of unperturbed genomic/mitochondrial DNA. Essentially all information on structure of biologically interesting regions from the genomic/mitochondrial DNA comes either from X-ray diffraction studies or solution NMR investigations conducted on model oligonucleotides (18). The use of short model oligonucleotides presumes the "excision" of the sequence to be studied out of the genomic context. The concept of excision implicitly assumes that structure and dynamics of the model oligonucleotide

are identical to that of corresponding sequences in the genome; in other words, the structure and dynamics of the model oligonucleotide are independent of the flanking sequence/structural context. However, a number of examples exist, where two oligonucleotides with marginal difference in sequence (differing typically by virtue of one or two extra flanking nucleotides) show notably different structural properties. In this respect, we define sequence context-promoted CP as the situation when two or more distinct "excisions" of an identical structural segment from the genomic/mitochondrial DNA display distinct structural properties.

One of the most studied cases of sequence context-promoted CP is provided by a DNA sequence corresponding to the 3'-G-rich single-stranded overhang (G-overhang) from human telomeric DNA. The 50- to 200-nucleotide-long G-overhang consists of repeating d(GGGTTA) elements (19). The G-overhang was observed to adopt G-quadruplex (G4) structures *in vitro* and *in vivo* (20-22). A number of studies suggested that the G-telomeric overhang consists of multiple G4 units (each unit comprises four telomeric repeats) that are arranged in a "bead-on-the-string" fashion (reviewed in, e.g., (22)). Folding of the G-overhang into a four-stranded G-quadruplex has been demonstrated to inhibit telomerase, an enzyme that is activated in more than 80% of cancers (reviewed in, e.g., (23)). As several small molecular weight ligands that stabilize telomeric G-quadruplex structures have displayed promising anticancer activity in tumor xenograft models, it has been proposed that stabilization of telomeric G-quadruplexes might be applicable to the treatment of a wide range of human cancers (reviewed in, e.g., (24,25)). However, *in vivo*, all of these small molecules failed to selectively target telomeric G4-DNA relative to other G4-forming regions in the genome (26,27). Significant interest in improving the selectivity of small molecular weight ligands towards telomeric G4 prompted structural studies to characterize the physiologically relevant conformation of the G4 unit within the G-overhang.

However, numerous studies using short telomeric constructs demonstrated that the G4 conformation strongly depends on the 5'- and 3'-flanking residues immediately adjacent to the core G4 forming sequence, namely d(GGG(TTAGGG)$_3$) (28). Thus far, the short oligonucleotides based on four telomeric repeat segments were found to be capable of adopting at least four folding topologies in presence of potassium, namely hybrid 1, hybrid 2, parallel and an anti-parallel two tetrad G4 (Figure 3), depending on the nucleotides flanking the basic G4 core (22,28-31). These

observations raised concerns about whether the structural behavior of the G4 units within the telomeric G-overhang might be realistically assessed in studies of short, model telomeric sequences forming single G4 units.

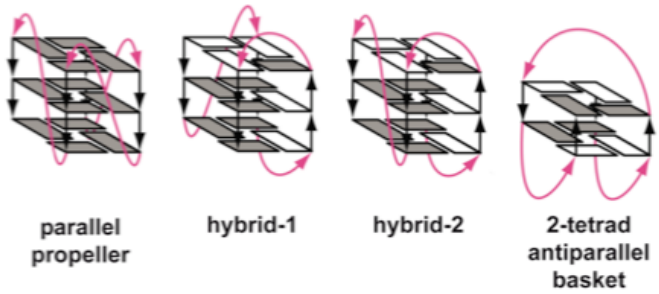

Figure 3. Schematic representation of G4-DNA folding topologies observed for short human telomeric DNA model constructs in presence of potassium, namely parallel (32), hybrid-1 (33), hybrid-2 (34), and an anti-parallel two tetrad basket (29).

To gain insight into the conformational properties of individual G4 units within the context of the telomeric G-overhang, we investigated site-specifically labeled extended constructs spanning 8 and 12 telomeric repeats using high-resolution NMR spectroscopy (35). While site-specific labeling of the construct based on 8 telomeric repeats, hereafter referred to as the double core construct, was tailored to identify the conformation of the 3'-terminal G4 unit in the G-overhang, the specific labeling of the construct based on 12 telomeric repeats, hereafter referred to as the triple core construct, was adopted to provide structural information about the conformation of the internal G4 unit(s) in the G-overhang. With this experimental setup, we were able to show that the addition of 3'-nucleotides and the whole 5'-G4 unit to the double core construct had no significant influence on the folding topology of the internal G4 unit. In other words, our data suggest that the 5'- and 3'-flanking-dependent conformational polymorphisms observed for single-core G4 sequences are diminished in long telomeric sequences that form multiple units. In addition, detailed NMR investigations revealed that the G4 units in the G-overhang sequences coexist in 2-tetrad antiparallel basket and hybrid-2 conformations. Our investigations identified the 2-tetrad antiparallel basket and hybrid-2 topologies as structural targets for the development of telomere-specific ligands, while marking parallel and hybrid-1 conformations as "artifacts of the experimental design"(35). The example given above illustrates a situation when marginal sequence variations give rise to dramatic structural rearrangements and bias the assessment of the biological relevance of

individual conformations within the studied DNA. In this particular case, the observed flanking nucleotide-dependent CP for telomeric sequences forming a single G4 unit can essentially be regarded as an artificial problem of the "inappropriate" design of the model oligonucleotide.

Although, as demonstrated above, the sensitivity of the DNA structure to sequence variation might impose a problem in assessing the biological relevance of the structural data, we recently showed that it could be positively exploited to interrogate the relationships between the DNA structure and its function, with a number of advantages. While the precise knowledge of the G4 topology in the telomeric G-overhang is of crucial importance for rational drug design applications, the question is whether the physiological function of the G4 units within the G-overhang requires a particular G4-topology. We set to address this issue by "tracing" the phylogenetically conserved structural properties of the telomeric G-rich DNA in the course of the evolution of eukaryotic species (36). While the function of telomeric DNA is conserved among eukaryotes, the sequences of the telomeric DNA differ between species. We based our analysis on the assumption that only functionally relevant structural features are subject to evolutionary selection pressure (and are evolutionarily conserved), while other functionally silent properties manifest themselves as phylogenetic variables. We found that the telomeric DNA from *Caernohabditis elegans* is based on a fold-back-based structure, while the telomeric DNA from other species is based on a G4 structure. These data indicate that only the formation of the secondary structure is required for function of telomeric DNA, but not a particular structural type. From the point of view of biomedical applications, our data suggest that the function of (human) telomeric DNA is not connected with any specific G4 topology (36).

From the stereo-chemical point of view, the telomeric DNA from *C. elegans* itself represents an important example of context-dependent CP. The G-rich telomeric DNA from *C. elegans*, emulated by the $d(GGCTTA)_3GG$ oligonucleotide, was previously shown to preferentially fold into a G4-quadruplex structure stabilized by Hoogsteen base-pairs (37). However, we demonstrated that the extension of the construct by a single 3' nucleotide leads to the complete remodeling of the conformation space towards the formation of the fold-back structure stabilized by Watson-Crick base pairs (36). We could show that this fold-back structure remains in the preferred folded topology in the extended telomeric constructs. This example also

well illustrates problems with design of model DNA constructs: the addition/deletion of the single nucleotide to the model construct might promote a transition between the different structural classes.

The specific category within this class is context-dependent CP promoted by (epigenetic) modifications of the DNA-constituting moieties, such as nucleic acid bases or the sugar-phosphate backbone. Typical examples of such modifications are 5-methylation of cytosine residues, deamination of adenosines, or formation of abasic sites in the polynucleotide chain (38-40). Depending on the sequence context, the individual modifications may display a notably distinct impact on the conformational behavior of DNA. For example, we recently showed that the generation of abasic sites within the G4-forming sequences from human telomeric DNA are connected with the promotion of the G4 polymorphism (41). In contrast, 5-methylation of the cytosine residues in i-motif-forming sequences were not found to promote the polymorphisms, but instead modulate the thermodynamic stability of the basic i-motif scaffold (42)

Together, the provided examples indicate that the basic assumption of the excision concept in structural biology is not universally valid and that transmitting the information on the oligonucleotide structure to its behavior in the context of the genomic DNA needs to be done with caution. Arguably, the disregard for the context-dependent CP is currently one of most frequent artifacts in structural biological analysis of DNA. On a positive note, context-dependent CP can be exploited for not only molecular engineering for the rational design of various DNA nano-machines and devices, but also for identification of the relationships between the DNA structure and its function.

The following articles of the applicant are related to the above topic:
(***K*** *marks article corresponding author*)

1: Hänsel R, Löhr F, Trantirek L, Dötsch V. High-resolution insight into G-overhang architecture. *J Am Chem Soc*. **2013** *135*(7):2816-24. **IF=12.1**

2**K**: Školáková P, Foldynová-Trantírková S, Bednářová K, Fiala R, Vorlíčková M, Trantírek L. Unique *C. elegans* telomeric overhang structures reveal the evolutionarily conserved properties of telomeric DNA. *Nucleic Acids Res*. **2015** 43(9):4733-45. **IF=9.1**

3: Konvalinová H, Dvořáková Z, Renčiuk D, Bednářová K, Kejnovská I, Trantírek L, Vorlíčková M, Sagi J. Diverse effects of naturally occurring base lesions on the structure and stability of the human telomere DNA quadruplex. *Biochimie*. **2015** 118:15-25. **IF = 3.0**

2.2 Environmentally promoted conformational polymorphisms

Due to the importance of DNA in living systems and materials science, the structures of DNA have been the subjects of intense biophysical investigations for the last 60 years, including high-resolution methods such as x-ray diffraction (XRD) and nuclear magnetic resonance (NMR) spectroscopy. However, instead of revealing the unique 3D structural motifs for specific DNA sequences, in many cases, these studies provided a puzzling picture of the DNA conformational polymorphisms that appeared to be promoted by the DNA sequence and the experimental conditions used in the structural investigations (28,43-48).

A common property among the broad spectrum of DNA motifs that emerged from structural investigations was the inherent sensitivity of the DNA structure and dynamics to non-specific physical/chemical environmental factors. A number of studies demonstrated that the stability, local DNA dynamics and the folding topology of DNA might strongly depend on the water activity, nucleic acid concentration, molecular crowding, pH, temperature, viscosity, or nature and concentration of the counter ion(s) (21,28,43,45,46,48-61). The folding of the DNA G-quadruplex motifs represents a textbook example of this type of behavior. The DNA G-quadruplexes formed by human telomeric repeats adopt at least three distinct folding topologies and differ in their strand polarities, loop orientations, and the orientation of the guanine bases with respect to the sugar moiety, depending on the environmental conditions and experimental conditions employed for their investigations (28). Hydration-dependent A- to B-DNA helix transitions (43,55) are among the other well-known examples of chemical/physical control over the DNA structure. Typical examples of environmentally controlled DNA structural polymorphisms are shown in Figure 4.

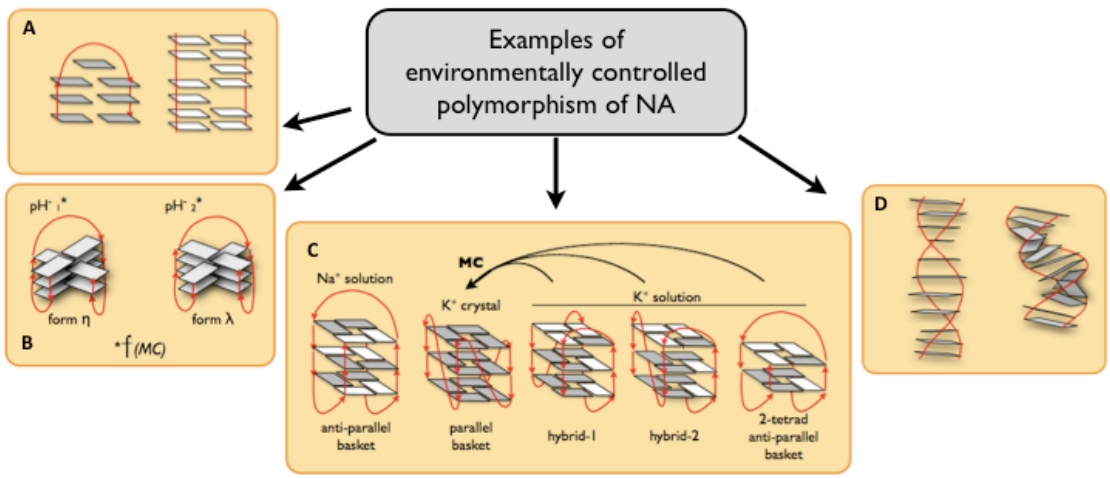

Figure 4. Example of environmentally controlled polymorphism of NA structures. A) concentration dependent hairpin to helix transitions (61), B) pH dependent rearrangements of i-motifs in human centromeric DNA (45,62), C) ion type, molecular crowding, and water activity dependent polymorphism of G-quadruplex structure (reviewed in (28)), D) Hydration-dependent A to B DNA helix transitions (55).

From a practical point of view, we define the environmentally promoted structural polymorphisms of DNA as DNA structural rearrangements (shifts in conformational equilibriums) that are induced by perturbation of one or more non-specific physical-chemical parameters in environment.

As already mentioned above and as evidenced from the structural database statistics, essentially all of the available high-resolution information on the DNA structure comes either from XRD or solution NMR investigations (http://www.rcsb.org/pdb/statistics/holdings.do). It needs to be stressed that the vast majority of these studies have been motivated by the biological relevance of the investigated DNA motifs. However, the observation that the structural behavior of DNA might strongly depend on the subjective choice of experimental conditions and procedures indicates that the conventional XRD and/or NMR techniques (in fact any in vitro spectroscopic technique) must be used with caution when addressing the question of the physiologically relevant structure of DNA molecules. Both XRD and NMR examine the structural properties of the isolated DNA under unnatural and rather simplistic conditions. The environmental conditions in XRD studies are constrained to conditions that support mono-crystal growth. In the majority of cases, these conditions are rather artificial and involve crystallization at non-physiological pH or in the presence of additives that facilitate the DNA crystallization process. In addition, the crystallization process itself exposes the DNA to notable degree of dehydration.

Considering the specifics of the XRD experimental procedure, it is of no surprise that the DNA structures derived using XRD are often found to be distinct from structures detected by spectroscopic methods in diluted solutions (48,52,55,61). Nonetheless, it needs to be stressed that analyzing the DNA structure/dynamics in solution does not necessarily ensure the physiological relevancy of the acquired structural/dynamical information, as the choice of solution conditions is limited by the

fact that not all of the relevant factors that can modulate the DNA structure in vivo are known.

Influence of counter ions

A statistical analysis of the conditions used in solution NMR spectroscopy over last 30 years to determine DNA structures revealed that more than 85% of the available structural information on DNA in solution currently comes from analyses in sodium-based buffers (63) (Note: Sodium is a prevalent ion in the extracellular space, where no DNA is present, while potassium is the prevalent ion in intracellular space (64)). This analysis highlights one of the most commonly applied assumptions in solution NMR studies that presume that the nature of the counter ion has no influence on DNA structure and dynamics. However, shortcomings stemming from this assumption are only now beginning to be recognized. For example, the DNA G-quadruplexes were shown to adopt different folding topologies in the presence of sodium and potassium (28). Ion type-dependent bending of double-stranded DNA polyA/polyT sequences represents another known example (65). Although in a number of cases the overall DNA structure, particularly double-stranded DNA, might be essentially identical in both Na^+ and K^+-based solutions, we recently demonstrated that the local DNA dynamics in the presence of Na^+ are notably distinct from those in the presence of K^+ (66).

Molecular crowding

There have been ongoing attempts to assess the physiologically relevant DNA structure in vitro by emulating the composition of the intracellular space using adjusted buffer compositions. In addition to salts and small molecular weight compounds, the milieu that surrounds the chromosomes, the inter-chromatine compartment (67), is crowded by the presence of (macro)-molecules up to ~ 100mg/ml (68). A simulation of the so-called molecular crowding effect with organic additives showed that the DNA structure is notably sensitive to crowding. The simulated crowding effect, mostly using polyethylene glycol (PEG) 200-400, was shown to affect the equilibrium between the DNA duplex and G-quadruplex/i-motif, promote the inter-conversion between various G-quadruplex topologies, strengthen the interaction between DNA and small molecular weight ligands, or promote stabilization of DNA triplex (59,69). However, we have recently demonstrated that the molecular crowding DNA experiences inside cells and/or in crude cellular

homogenates is notably distinct from the conditions DNA experiences in PEG-supplemented solutions (21,70,71). It was only recently revealed that the majority of the observed structural phenotypes in the simulated crowded environment were due to a direct interaction of the PEG molecules with the DNA rather than due to the influence of molecular crowding itself (51).

Together, the experimental observations demonstrating that the folding of NAs depends on environmental factors strongly suggest that the *quantitative characterization of physiologically relevant NA structures and dynamics should ideally be performed either under native conditions in vivo or under in vitro conditions that realistically emulate the complex environment of living cells.* This consideration has recently sparked interest in the development of novel tools that allow the structural characterization of NAs in the complex cellular context.

Inception of structural analysis of NAs in living cells

The introduction of the concept of in-cell spectroscopic analysis, namely in-cell NMR spectroscopy, in-cell EPR spectroscopy, and in-cell single particle FRET, was the first important step towards high-resolution structural analysis of NAs in their native environment (72-74). All of the mentioned methods are based on a similar underlying principle, which is outlined in Figure 5. In all of the techniques, the exogenous NA fragments are introduced into living cells, followed by the respective spectroscopic investigations. Despite the illusion that the in vivo measurements are directly analogous to the conventional in vitro NMR/EPR/FRET experiments, in-cell spectroscopic analysis has developed into a separate field. The established techniques for in vitro NMR/EPR/FRET structural analysis of NAs generally fail to provide an expected readout under in vivo conditions due to the limited resolution and reduced sensitivity of in-cell spectra, as well as due to the limitations imposed by biological factors, such as cell death, accidental NA leakage from the cells, the NA cellular localization, or degradation - for details, see (70-72). As a result, in-cell spectroscopic analysis requires the application of unconventional approaches not only for sample preparation and in-cell spectra acquisition but also for the interpretation of the in-cell spectroscopic data. While in-cell spectroscopic analysis of NA structure and/or dynamics in the native, complex cellular environment is feasible and provides structural insights that are impossible to envision using the conventional isolated in vitro methods, its general application remains limited due to various technical

reasons. The limitations and application potential of the individual methods are discussed in detail below.

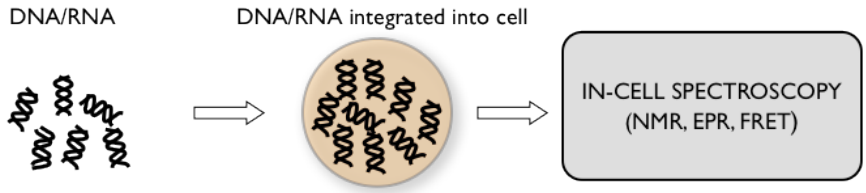

Figure 5. Schematic representation of general in-cell spectroscopic setup.

2.2.1 In-cell NMR spectroscopy

Initially, in-cell NMR spectroscopy was devised to selectively observe overexpressed $^{13}C/^{15}N$-labeled proteins in bacterial cells (75,76). Labeling with $^{13}C/^{15}N$ isotopes was shown to be indispensable to resolve the NMR signals from the protein of interest from the signals originating from the cellular background and to enhance the sensitivity of the in-cell NMR detection (71,75-79). However, this approach is limited by the requirement for high levels of the overexpressed proteins. In 2006, this concept was extended to allow observations of proteins in eukaryotic cells, namely in *Xenopus laevis* oocytes (79,80). The original and new concepts differ in the way in which the investigated protein is deposited in the living cells. In the original approach, the isotopically labeled protein is directly produced within the bacteria, which grows in isotopically labeled medium. In the latter approach, the isotopically labeled protein is not produced in oocytes, but is delivered into oocytes via microinjection. The microinjection delivers several advantages over endogenous overexpression of the protein. First, the deposition of the isotopically labeled protein into the isotopic label-free cellular background allows us to monitor the introduced protein without interference from the cellular background. Second, in contrast to the original method based on protein overexpression, this approach can be applied to biomolecules other than proteins.

In 2009, our group adapted the method of in-cell NMR spectroscopy to investigate polymorphisms of nucleic acids (NAs) in the complex environment of injected *X. laevis* oocytes (73). In an in-cell NMR experiment, the unlabeled or isotopically labeled DNA fragment, which is prepared either by enzymatic or chemical synthesis, is mechanically introduced into *X. laevis* oocytes via microinjection. The intracellular concentration of the introduced DNA typically needs

to be in the 50-250 μM range to observe and characterize the DNA structure inside a cell. Approximately 200 oocytes need to be injected for one in-cell NMR sample. The injected oocytes are then transferred to an NMR tube, submerged in a buffer mimicking the composition of the extra-cellular environment and subjected to NMR investigations (73).

In contrast to its application to proteins, the in-cell NMR of NA has identified specific problems, such as the degradation of nucleic acids inside the living cell (70,71,73). The DNA is exposed to nucleases when it is introduced into the cells. The nucleolytic activity results in the relatively fast depletion of the studied NA fragment inside the cell. The "life-span" of the DNA in vivo strongly depends on its primary sequence and folding topology. In addition to DNA degradation, there are also other factors that might impose additional "time" restrictions, such as leakage of introduced DNA from cells or cell death. In general, due to the limitations arising from DNA degradation, leakage of the DNA from cells, and/or cell death, the time-window accessible for routine DNA exploration under in vivo conditions is typically less than 6 hours (70,71,73). This time restriction dictates the type of NMR experiments than can be used for *in vivo* NA structure investigations.

Another source of restrictions in applications of in-cell NMR spectroscopy of nucleic acids stems from the heterogeneous nature and viscosity of the intracellular environment, as well as the heterogeneity of the NMR sample composed of stacked oocytes. All of these factors are responsible for the generally low resolution and S/N ratio of *in-cell* NMR spectra. Typically, the signals observed in the *in-cell* NMR spectra are significantly broader than the signals from purified samples (70,71,73).

In general, the quantitative structure determinations using NMR spectroscopy presume the use of isotopically labeled samples coupled with the exploitation of an elaborate suite of hetero-nuclear NMR experiments. While there was one reported successful protein structure determination *in vivo* using *in-cell* NMR spectroscopy (81,82), for nucleic acids, de novo structure determination *in vivo* appears impractical, although, in principle, it is feasible. The main obstacle is low *in vivo* stability of the nucleic acids coupled with a prohibitively high cost for isotopically labeled samples. An additional obstacle might be toxicity. We have observed that some NA fragments were toxic to the injected cells at the standard concentrations required for *in-cell* NMR investigations (70,71,73). Although reducing the effective concentration of the injected NA usually diminishes toxicity, it is inherently connected with a decreased

sensitivity of the *in-cell* NMR experiment. As a result, the interpretation of the *in-cell* NMR data adopted thus far has been primarily based on comparisons of the *in vitro* and *in vivo* spectral fingerprints, rather than *ab initio* structure determination (70,71). Comparing the spectral fingerprints of NA sequences under various *in vitro* conditions that promote miscellaneous structural arrangements within the studied NA fragments with *in/ex vivo* spectral patterns serves two primary purposes: i) to identify defined *in vitro* conditions, mostly a specific buffer composition, to be employed for a detailed characterization of the physiologically relevant structures using conventional *in vitro* NMR spectroscopy; and ii) to identify the property of the intracellular environment that is responsible for defining the effective conformation of the nucleic acid *in vivo*. While the *in vivo* and *ex vivo* spectral patterns are recorded in living *X. laevis* oocytes and/or *X. laevis* egg extracts, respectively, the reference *in vitro* data are individually recorded at conditions simulating various degrees of molecular crowding, viscosity, miscellaneous pH and temperature, or different compositions of a buffer in terms of low-molecular weight compounds (21,70). In general, this approach benefits from the high sensitivity of NMR chemical shifts to both local and global structural changes of the nucleic acids. In principle, each NA folding topology is marked by a unique pattern of NMR signals. As a result, an agreement between the *in/ex vivo* and *in vitro* spectral patterns is a reliable indication of the representative *in vitro* conditions under which the nucleic acid fragment adopts the same folding arrangement as in a cell. A schematic representation of the interpretation process based on the comparison of NMR spectral fingerprints is outlined in Figure 6.

In contrast to conventional in vitro NMR, the in-cell NMR experiment is extremely technically challenging. The approach used to interpret the in-cell NMR data based on comparisons between the spectral fingerprints of the NA sequences acquired in vivo and those acquired in vitro completely fails when none of the considered in vitro conditions provide a spectrum identical to that recorded under native conditions in vivo. Unfortunately, these cases are biologically the most interesting. This approach also becomes quite problematic when the in vivo and in vitro spectra share some degree of similarity but are not entirely identical. In this case, a philosophical problem arises in terms of the meaning of the in vitro structural data. While the degree of spectral similarity will somehow reflect the accuracy of the derived structural information, the precision of in the vitro-derived structure will become meaningless in this case (please note that in XRD and in vitro NMR analyses,

the structural precision is the only measure of the structural quality). For a detailed discussion of the application potential and limitations of in-cell NMR spectroscopy, please refer to our recent review (70,71). Undoubtedly, the development of techniques allowing *ab initio* NA structural determination in vivo is an essential prerequisite for the general applicability of this approach and future development of the field.

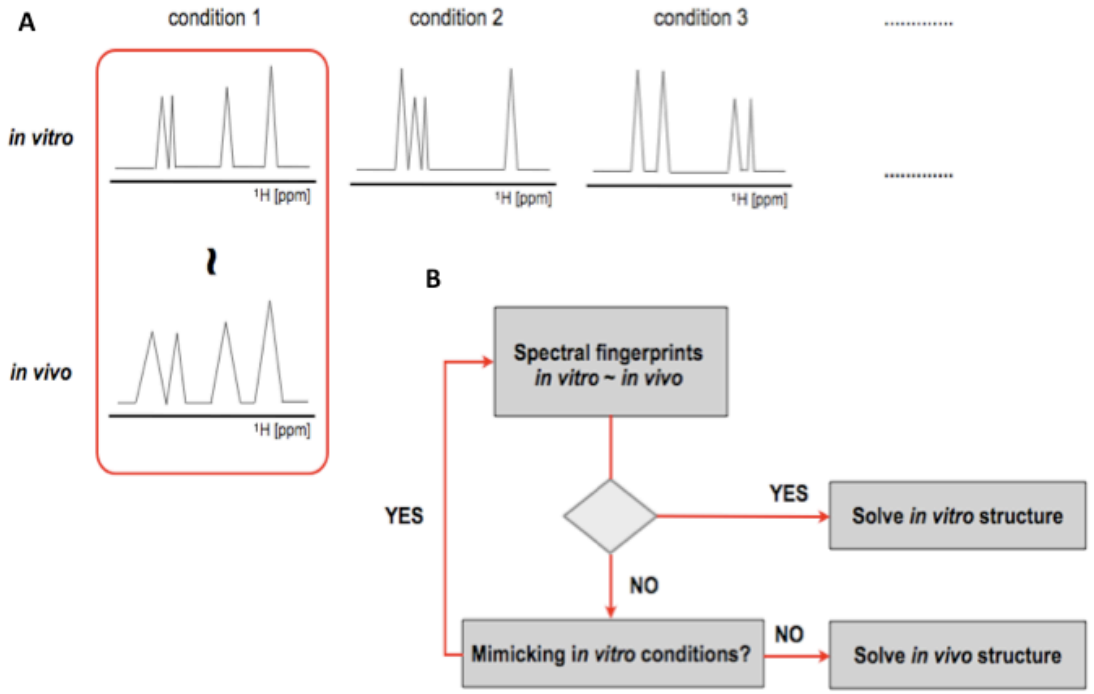

Figure 6. A) and B) Outline of the interpretational process of in-cell NMR spectra based on comparison of spectral fingerprints between in vivo and various in vitro conditions. The in vitro conditions are employed to generate different structural arrangements of NA fragment under the study. The figure was reproduced from our recent review (70).

2.2.2 In-cell EPR spectroscopy

In-cell EPR spectroscopy of DNA was introduced in 2011 by Prof. Prisner and colleagues (74). In terms of sample preparation, the in-cell EPR procedure is directly analogous to the in-cell NMR experiment. Whereas NMR spectroscopy makes use of the magnetic properties of the nuclear spin, EPR spectroscopy is based on the magnetic moment of unpaired electrons. Of particular interest is a technique called PELDOR (or DEER), which measures the dipolar coupling between unpaired electrons that are separated by distances of 1.5–10 nm. The primary components of nucleic acids are devoid of paramagnetic electrons; therefore, the target of interest

must be labeled with a stable paramagnetic probe. Site-directed spin labeling (SDSL) is routinely used to attach small nitroxides, such as 3-maleimido-2,2,5,5-tetramethyl-1-pyrrolidinyloxy (3-maleimido-PROXYL, 5-MSL) or 2,2,5,5-tetramethyl-1-oxyl-3-methyl methanethiosulfonate (MTSL), to the NA base, sugar, or phosphate-backbone moiety during or after chemical synthesis (50,71,74,83,84). In contrast to liquid-state in-cell NMR spectroscopy, which is carried out at physiological temperatures, the fast relaxation of the nitroxide spin labels at room temperature requires that the PELDOR experiments are conducted at cryogenic temperatures (50 K), and the cells are shock-frozen in liquid nitrogen prior to the experiment (71,85).

EPR spectroscopy cannot provide the same amount of detailed information about the conformational and dynamic state of a macromolecule as NMR spectroscopy. However, the advantages of in-cell EPR spectroscopy are the lack of cellular background signals, and the possibility of measuring distances by the PELDOR EPR method independent of the tumbling rate, and, therefore, the molecular weight of the macromolecule (85). Instead of 200 oocytes, as required for in-cell NMR experiments, only 30–50 injected oocytes are needed to occupy the active volume of the EPR resonator. Thus, injection and loading can be performed within 10–15 min, while the injection of the 200 oocytes required for in-cell NMR analysis usually takes hours.

Undoubtedly, the biggest challenge for in-cell EPR spectroscopy is the rapid reduction of the currently available spin labels in the cellular environment, which limits the measurement time. Igarashi et al. observed that the nitroxide spin label attached to the target under study was quickly reduced once present in the cellular environment, with an estimated half-life of 1 h. (86). More sterically protected spin labels are currently being developed and will have a major impact on the application of EPR spectroscopy in cellular systems (87) .

2.2.3 Is the DNA inside the cells automatically in the native conformation?

The “Catch-22” of all of the *in vivo* techniques for nucleic acid structural characterization, including in-cell NMR and in-cell EPR, is that while these techniques were devised for biomolecular structure analysis under native conditions, their implementations are always connected with various degrees of disturbance of the native environment (71,78). Without any doubt, the introduction of high

concentrations of exogenous DNA into a living cell is inherently connected with alterations of the physico-chemical properties of the intracellular environment, which might bias the structural readout.

While the deposition of large quantities of DNA into cells is an inherent attribute of in-cell NMR and in-cell EPR, the cells appear to be able to diminish the adverse effects and restore the homeostasis of their intracellular environment. There are two options to confirm the integrity of the manipulated cells. The first is based on a comparison of the cell viability injected with a buffer versus cells injected with a buffer containing the DNA (70,71). The second uses the addition of progesterone to injected oocytes and determines whether the oocytes can be driven from a G2/M arrested oocyte to metaphase of meiosis II, i.e., to the stage of a mature egg (70,71). As the enzymatic activities required for the transition from an oocyte to a matured egg are sensitive to disturbances in the environment, this simple test provides a good indication that the injected cells were able to restore the homeostasis of the intracellular environment. A dramatic disturbance of the intracellular environment is typically recognized within 30 minutes of the injection and is marked by increased cell mortality.

2.2.4 *In-cell single particle FRET*

For both the in-cell NMR and EPR investigations, there are three principal issues that stem from the requirement to deposit ultra-high concentrations of exogenous DNA into the cells (73,74,84). First, as mentioned above, the introduction of high concentrations of exogenous DNA into a living cell is inherently connected with alterations of the physico-chemical properties of the intracellular environment, which might bias the structural/physiologically relevant readout (78). Second, it is not only likely that there is possible disturbance of the intracellular environment, but also the inability to ensure that the entire introduced DNA is localized to its native, namely the cell nucleus. Although, it was shown that small DNA fragments accumulate in the nucleus after injection into the cytoplasm of eukaryotic cells, the localization is not quantitative, and ~10% of the introduced DNA remains in the cell cytosol (70). Third, our inability to deliver these high concentrations of exogenous material into bacterial and mammalian cells limits (at the present) the in-cell NMR/EPR analysis to a single cell type, namely *X. laevis* oocytes.

To resolve all of these issues, we have devised the concept of in-cell single particle FRET of nucleic acids to allow us to characterize the DNA structure in the complex environment of living cells at the single molecule level (72). The method is based on the common principle of all in-cell methods: exogenous DNA is introduced into the living cells, and, in this case, it is labeled with fluorophore tags. Quantitative FRET analysis is then employed to convey the required structural information about the nucleic acid fragment. The principle difference between the in-cell EPR/NMR method and in-cell spFRET is that in in-cell spFRET, the required information on the DNA structure is derived from measurements from a single molecule (a single emitted photon). As the single molecule can be delivered into almost any type of mammalian and/or bacterial cell, this method is not limited to large cells that can be mechanically injected, such as *X laevis* oocytes, in contrast to in-cell NMR/EPR. The delivery of a single molecule (or only a few molecules) into the cells has no toxic effect (which is often observed with in-cell NMR/EPR applications), and the disturbance of the native environment is diminished. Last, but not least, the fluorescence from the studied molecule provides a means of verifying the DNA localization inside the cell, and only the fluorescence from the DNA localized in the cell nucleus can be used to quantitatively assess the DNA structure (72).

However, in terms of the data interpretation procedures, the in-cell spFRET and in-cell EPR are essentially identical (72). The current approaches for the interpretation of the in-cell FRET/EPR data require the existence of reference 3D structural models. This model-based interpretation presumes that all of the possible structural arrangements of the NAs can either be predicted or determined prior to in-cell data interpretation using conventional approaches, such as XRD and/or NMR. The interpretation of the EPR/FRET data is then essentially reduced to a statistical evaluation of the closest match between the distances estimated from the set of reference structures and the in vivo distances between either paramagnetic or fluorescent tags acquired by in-cell EPR or in-cell FRET, respectively. This approach has two principle weaknesses. i) It presumes the existence of accurate and precise structural representations for the studied NA. When this requirement, which cannot be guaranteed a priori, is not fulfilled, the procedure, which is based on an evaluation of the closest match, will provide a “false positive” identification of the “physiological” structure. ii) Most importantly, due to the notable uncertainty in the paramagnetic/fluorescent tag location with respect to the NA and the altered spectral

properties of the tag in the intracellular environment (72), this approach might fail to unambiguously discriminate even quite distinct structural motifs.

For a detail comparison of the application potential and limitations of the individual in-cell spectroscopic techniques see (71,72).

Concluding remarks on in-cell spectroscopic methods: The development of in-cell spectroscopic methods allowed us to characterize the DNA structure within its native environment. The reported applications of these methods provided important insights into the role of the intracellular environment in modulating the DNA structure/function, and it is also important information for formulating the composition of the buffers that are being used to emulate the parameters of the intracellular space in in vitro spectroscopic studies. The importance of in-cell spectroscopic methods can also be illustrated in a recent in-cell NMR study by Selgado et al. (88), which showed that the DNA-drug interactions (drug binding mode) observed in vivo might be distinct from those observed under artificial in vitro conditions. This observation has profound implications for the field of drug development, particularly for rational drug design, which entirely relies on the knowledge of DNA structure and the DNA-drug binding modes.

However, at the same time, it needs to be stressed that the in-cell NMR/EPR/spFRET analysis is far from being routine. The applications of these methods are primarily hampered by technical difficulties in preparing the in-cell samples, the high costs of the labeled material, and the limitations stemming from problems with the acquisition and interpretation of the in-cell spectroscopic data. Currently, only a few laboratories around the globe (less than ten) are capable of producing and interpreting the in-cell data. It is clear than more methods should be developed to make these techniques available to the research community. The construction of advanced bioreactors to keep the cells alive under the harsh experimental conditions, the development of ultra-fast schemes for the acquisition of the in-cell NMR data, the development of interpretational schemes that allow the determination of the *ab initio* structure from the in-cell NMR/EPR/spFRET data, development of stable spin tags, development of novel approaches for geometry-defined attachment of spin/fluorophore tags to DNA, and the development of efficient and novel methods for the delivery of exogenous DNA into cells, among others, will be essential.

2.2.5 Molecular basis for the modulation of the DNA structure by environmental factors

The majority of studies dealing with environmentally promoted conformational polymorphisms are primarily descriptive and limited to the identification of the phenomena for particular DNA sequence. Only a small fraction of these studies provides insight into the molecular mechanism by which the environmental factors transmit their effects into the DNA structure and dynamics. Conceptually, the impact of the water exclusion effect due to molecular crowding/confinement is well understood: the DNA prefers to adopt conformation(s) generally marked by the smallest hydrodynamic radius (the smaller/smallest active area for interactions with water molecules) in environments with reduced water activity (50,53,54,56-58,89-95). In addition to the water exclusion effect, the entropic effect is cumulative, and the DNA disfavors extended conformations in a spatially confined environment to minimize the collisions of the DNA with the internal surface of elastic "nano-cavity" (60,95,96).

The DNA-ion interactions are the most problematic to study and as result the least understood. The major impediment to the characterization of the DNA-ion interactions, which are transient and weak in nature, is a general lack of unbiased techniques that allow their investigation. High-resolution information on DNA-ion interactions usually comes from either X-ray crystallography or solution NMR or EPR analysis (97-101). In XRD studies, a low degree of hydration in the crystalline state has been shown to change the nature of the DNA-ion interactions, favoring direct interactions between DNA and ions over the water-mediated DNA-ion interactions observed in solution (55). In contrast to XRD studies, NMR and EPR measurements can be performed in a physiologically relevant hydration state. Unfortunately, the NMR and EPR properties of the physiologically relevant counter ions (e.g., Na^+, K^+, Mg^{2+}) do not allow the application of standard techniques to identify their interaction sites. These limitations are usually overcome by the application of mimicking systems, referred to as "substitutionary ions" (e.g., NH_4^+ for monovalent ions or Mn^{2+} for Mg^{2+}) (99-101). However, the use of the substitutionary ions brings about possible source of bias due to the differences between properties of the native and mimicking ions. To avoid the use of the substitutionary ions, we developed a simple approach based on monitoring the NMR cross-correlated relaxation rates (Γ) between the aromatic carbon chemical shift anisotropy (CSA) and

the proton–carbon dipolar interaction to identify the binding sites between the physiologically relevant counter ions and the DNA (66). The technique proved to be extremely sensitive and can be used to detect very weak and transient interactions on a time scale much shorter than the DNA correlation time (τc). Such interactions are virtually invisible to conventional NMR techniques, such as chemical shift mapping and NOESY-based techniques, which are not likely to detect ions with occupancy less than 10% (102). Our study revealed that Na^+ and K^+ have distinct preferential binding sites on DNA and that the ion binding to DNA distinctly affects the local intra-molecular dynamics in the vicinity of the ion-coordination site(s) (66).

However, it is not just the transient and weak character of the water/ion-DNA interactions that makes them difficult to study. As demonstrated by our studies, the structural manifestations of the DNA water/ion interactions are often well below the resolution of conventional NMR methods and/or schemes used to interpret the NMR data (66,103-107). For many years, the hetero-aromatic bases, a constituent unit of DNA, were considered to be planar and conformationally rigid (108). In 2009, our group demonstrated that in reality, the nucleic acid bases are non-planar systems (104). In our subsequent study, we showed that the degree of the non-planarity is fine-tuned by the interaction of the DNA bases with the environment, represented by networks of water molecules and ions interacting with the individual bases (106). At the same time, our data indicated that the non-planarity is strongly correlated with the conformation of the glycosidic bond (104,106). These data suggested that the environment modulated the flexibility of the hetero-aromatic bases and may transmit information about the DNA environment into the arrangement of the DNA helix (106).

Our database study based on an inspection of the ultra-high resolution XRD structures of DNA that revealed the details of the localized water molecules and interacting ions indicated that the structural manifestation of the interactions between the DNA and its environment, particularly the non-planarity of nucleic acid bases, are, in fact, reflected in the number of observable NMR parameters, such as direct and indirect spin-spin interactions or the NMR relaxation parameters related to local DNA dynamics (103,105,107). However, a re-examination of the number of primary experimental NMR data points on DNA accompanied by a series of QM calculations revealed that the information on the environmentally dependent structural variations is “canceled” and “spoiled” by the approximations employed in the course of the

NMR data interpretation. For example, information about glycosidic nitrogen pyramidalization, a quantitative measure of the non-planarity of a nucleic acid base, is canceled by inappropriate parameterization of the Karplus equation used to interpret 3JH1'-C6/8 or 3JH1'-C2/4; the corresponding parameters implicitly presume an idealized sp^2 hybridization at the glycosidic nitrogen (107). This presumption results in over/underestimation of the glycosidic torsion angle. A similar situation holds true for other primary NMR parameters, such as the cross-correlation relaxation rates between glycosidic nitrogen CSA and C1'-H1' dipole-dipole or NMR relaxation data that depend on C6/8 and/or C1' CSA (103,105).

The assumption of planarity of the nucleic acid bases that is implicitly embedded in NMR data interpretational schemes not only "destroys" the information about the interactions of individual DNA bases with the local environment but also biases the entire structure determination process. The situation is further complicated by errors in the empirical parameterization of the current generation of the force fields employed in the (restrained)-MD simulations used to determine the DNA structure from the NMR data. While the current generations of the force fields correctly emulated the degree of nonplanarity of the nucleic acid bases, they fail to correctly assess the chirality at the individual atoms of DNA bases (104).

Thus far, our understanding of the DNA-ion/water interactions remains limited. The major obstacle to the characterization of the DNA-ion interactions is a general lack of unbiased experimental techniques that allow their investigation at a sufficient resolution.

The following articles of the applicant are related to the above topics:
(***K*** *marks article corresponding author*)

1**K**: Sychrovský V, Müller N, Schneider B, Smrecki V, Spirko V, Sponer J, Trantírek L. Sugar pucker modulates the cross-correlated relaxation rates across the glycosidic bond in DNA. *J Am Chem Soc*. **2005**;127(42):14663-7. **IF=12.1**

2**K**: Brumovská E, Sychrovský V, Vokácová Z, Sponer J, Schneider B, Trantírek L. Effect of local sugar and base geometry on 13C and 15N magnetic shielding anisotropy in DNA nucleosides. *J Biomol NMR*. **2008** 42(3):209-23. **IF=3.1**

3**K**: Sychrovsky V, Foldynova-Trantirkova S, Spackova N, Robeyns K, Van Meervelt L, Blankenfeldt W, Vokacova Z, Sponer J, Trantirek L. Revisiting the planarity of nucleic acid bases: Pyramidilization at glycosidic nitrogen in purine bases is modulated by orientation of glycosidic torsion. *Nucleic Acids Res.* **2009** 37(21):7321-31. **IF=9.1**

4**K**: Hänsel R, Foldynová-Trantírková S, Löhr F, Buck J, Bongartz E, Bamberg E, Schwalbe H, Dötsch V, Trantírek L. Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy. *J Am Chem Soc.* **2009** *131*(43):15761-8. **IF=12.1**

5**K**: Vokácová Z, Trantírek L, Sychrovský V. Evaluating the effects of the nonplanarity of nucleic acid bases on NMR, IR, and vibrational circular dichroism spectra: a density functional theory computational study. *J Phys Chem A*. **2010** 114(37):10202-8. **IF=2.7**

6**K**: Fiala R, Spacková N, Foldynová-Trantírková S, Sponer J, Sklenár V, Trantírek L. NMR cross-correlated relaxation rates reveal ion coordination sites in DNA. *J Am Chem Soc.* **2011** 133(35):13790-3. **IF=12.1**

7**K**: Hänsel R, Löhr F, Foldynová-Trantírková S, Bamberg E, Trantírek L, Dötsch V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **2011** *39*(13):5768-75. **IF=9.1**

8: Sychrovský V, Sochorová Vokáčová Z, Trantírek L. Guanine bases in DNA G-quadruplex adopt nonplanar geometries owing to solvation and base pairing. *J Phys Chem A*. **2012** 116(16):4144-51. **IF=2.7**

9**K**: Fessl T, Adamec F, Polívka T, Foldynová-Trantírková S, Vácha F, Trantírek L. Towards characterization of DNA structure under physiological conditions in vivo at the single-molecule level using single-pair FRET. *Nucleic Acids Res*. **2012** *40*(16):e121. **IF= 9.1**

10: Hänsel R, Löhr F, Trantirek L, Dötsch V. High-resolution insight into G-overhang architecture. *J Am Chem Soc*. **2013** *135*(7):2816-24. **IF=12.1**

11**K**: Hänsel R, Foldynová-Trantírková S, Dötsch V, Trantírek L. Investigation of quadruplex structure under physiological conditions using in-cell NMR. *Top Curr Chem*. **2013** *330*:47-65. **IF= 5.6**

6: Hänsel R, Luh LM, Corbeski I, Trantirek L, Dötsch V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angewandte Angew Chem Int Ed Engl*. **2014** 53(39):10300-14. **IF = 11.1**

2.3 Conformational polymorphisms related to the kinetic control of DNA folding

CP arising due to kinetic partitioning in the course of DNA folding account for a situation(s) when a single DNA sequence adopts two or more folding topologies, whose populations do change over time. While this definition is general enough to account for a simple transition from an unfolded to a folded state, the term is being employed to denote a DNA folding process that operates on very long time scales (hours, days, months) and along a very complex conformational surface that is rugged, with deep competing basins of attractions.

This type of CP is characteristic for the folding of G-quadruplex-forming sequences (44,47,109-111). In NMR, this type of CP is typically experimentally manifested by the i) time-dependence of the NMR spectra and ii) high sensitivity of spectral pattern to experimental (environmental) conditions including sample storage (112). Regarding the time-dependence of the NMR spectra, the spectrum recorded one day is notably different from the spectrum recorded on the same sample several hours (days) later. Even for samples that were "stabilized" by so called annealing, the procedure serves to shift the conformational equilibriums towards the most thermodynamically stable state, and the NMR spectra are often found to change over time. Regarding the sensitivity of the spectra to the environment, perturbations in buffer composition have a pronounced effect on the appearance of the NMR spectra in terms of the number of species present and their relative populations (44-46,49,109-111,113).

This kind of behavior seems to suggest that DNA G4 folding is fundamentally different from the funnel-like folding of small RNAs or proteins, which fold on the time-scale of μs to ms (114,115). The folding of small proteins (or RNAs) is directed by the formation of native, local secondary structure contacts (116). This contrasts with the G4 DNA, where the most critical part of the folding may involve the formation of the native, consecutive, ion-stabilized quartets. The quartets are based on non-local interactions and, in a given fold, must have a specific combination of *syn* and *anti* nucleotides (17,117). A decisive role of the local contacts (which would include formation of the loop topologies) for the complete G4 fold is also not consistent with the fact that many G4 structures fold to diverse topologies upon changes in the environment or flanking sequences (21,44,46,49,110,113,118). In fact, this indicates the presence of several (or even many, depending on the time

resolution) competing basins of attractions. It is likely that only a small fraction of the molecules directly attain the most thermodynamically stable ensemble. The remaining molecules would be first trapped in competing sub-states, i.e., misfolded states, resulting in a multiple pathway folding process. The native and most significant competing basins of attraction may be interchanged upon changing the experimental conditions, which may explain why the dominant folding topology in the final thermodynamic equilibrium is so sensitive to subtle changes in the folding conditions. Most molecules likely initially attempt different (competing) folds and may need to unfold to attempt another fold (112). This is entirely consistent with our simulation studies identifying numerous highly stable potential intermediates (119,120). This has already been visualized experimentally for the folding of the human telomeric hybrid-1 type GQ, for which the hybrid-2 structure has been shown as a major competing sub-state (basin of attraction) with respect to the global minimum (112). Of note, the study by Bessi et al. (112) showed that the transition between the two folds appears to be realized via the unfolded ensemble. It cannot be ruled out that the folding has also been affected by other sub-states, which are already below the resolution of the experiment. A very long time is definitely required for the most thermodynamically stable structure (under given conditions) to be observed in the experiment.

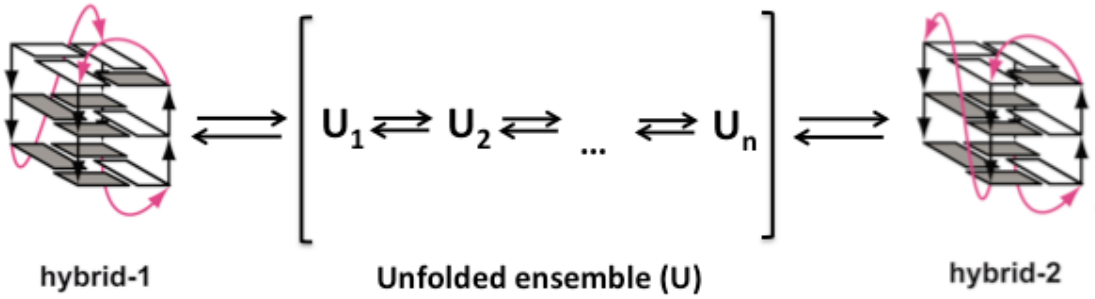

Figure 6. Proposed mechanism for interconversion between hybrid-1 and hybrid-2 G4-DNA model constructs based on human telomeric DNA (112).

In protein or RNA structural biology, it is implicitly presumed that the biological function is related to the lowest energy conformation. In this case, the folding times are comparable or even below the time scale of the related biological process. However, this assumption might not be valid for the G4 structures. Considering the extremely slow time scale of G4 folding, which requires hours (days) to reach the

most thermodynamically stable state, it is much more likely that kinetically favored states, which can be considered as folding intermediates, are responsible for the biological function. This consideration implies that the G4 folding intermediates might represent more efficient drug targets compared to the G4 thermodynamically controlled states (121).

To gain insight into the folding pathway of the telomeric G4 structures, we have used an MD simulation to assess the stabilities of the plausible folding intermediates, namely G-hairpins and G-triplexes (119,120). The investigation of G-hairpins is relevant not only to the earliest stages of the folding pathways (the hairpins were suggested to fold on the ms time scale), but also to various later phases of folding, as the G-hairpins may be important parts of the ensembles during the inter-conversions among different triplex and quadruplex arrangements. Our identified structural arrangements of G-hairpins and G-triplexes are stable enough to contribute to the G4 folding pathways (119,120).

To date, GQ folding remains elusive, despite the intense experimental efforts to understand the process. The basic limitation of the experimental studies is that most of them do not allow a confident determination of the structures that are populated in time; they are limited in terms of time, structural resolution and/or sensitivity to the low populated species. The most recent data suggest that it may be feasible to target the G4 folding intermediates, such as G-hairpins, G-triplexes, or kinetically controlled G-quadruplexes, for anticancer drug designs that target G-rich regions.

The following articles are related to the above topic:

1: Stadlbauer P, Trantírek L, Cheatham TE 3rd, Koča J, Sponer J. Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie* **2014** 105:22-35. **IF=3.1**

2: Stadlbauer P, Kuhrova P, Banas P, Koca J, Bussi G., Trantírek L, Otyepka M, Sponer J. Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res*. 2015, *in press*(gkv994). **IF=9.1**

3. Biological relevance of environmentally controlled DNA polymorphisms

The sensitivity of the DNA structure and dynamics to non-specific physical/chemical factors is an inherent property of DNA. While this phenomenon is being heavily exploited in molecular engineering to construct dynamic DNA assemblies, such as logical gates, conformational switches, nano-machines, and/or biosensors (122-124), it remains unclear whether this phenomenon is, in fact, biologically relevant. Although a number of DNA related processes, such as those connected to changes in gene/ncRNA expression, are accompanied by natively occurring fluctuations in the composition and properties of the intracellular space, such as pH, [Ca2+], [K+]/[Na+] fluctuations in the course of cell- cycle, stress responses like hypoxia or hyperthermia, apoptosis, or the degree of molecular crowding/confinement in the course of chromatin remodeling (68,125-129), the direct link between environmentally controlled DNA polymorphisms and these environmental changes/biological process has not yet been demonstrated. One of the plausible explanations might be that within the ranges of the environmental conditions tolerated by the mechanisms controlling cellular homeostasis, *the phenomenon of environmentally controlled DNA polymorphisms does not manifest itself and/or is simply functionally silent in vivo.*

Similar considerations do apply to the DNA structural polymorphisms that arise from kinetic control of DNA folding. As shown for the G-quadruplexes formed from human telomeric DNA repeats, certain G4 topologies are kinetically favored over the topology corresponding to the lowest energy conformation (111,112). However, in this case, the manifested polymorphism is "temporary". In the initial stages of the folding process, the polymorphic mixture consists of a small fraction of the slowly formed, most thermodynamically stable species and a predominant fraction of the (multiplex) kinetically favored ones. The ratio between the kinetically and thermodynamically favored species changes over time. In thermodynamic equilibrium, the polymorphism is essentially diminished and G4 adopts a topology corresponding to the most thermodynamically stable conformation. The kinetically promoted polymorphism needs to be considered because G4 folding operates on an extremely slow time scale, which, in certain cases, may take months or even years. However, what is the biological relevance of this kinetically induced polymorphism? Biological processes take place at millisecond-to-second time scales. As result, the thermodynamically favored structure is unlikely to participate in the related biological

process. The complexity of the folding pathway indicates that a number of conformationally distinct species co-exist, even in the initial stage of the folding process. It is still an open question whether any particular conformation from the kinetically controlled mixture bears the function, or whether an equilibrium among the co-existing species is important for the function.

4. Future prospects

Most of the current research interest about DNA relates to its conformational polymorphisms. In materials science, the DNA conformational polymorphisms are being exploited for the construction of a controllable assembly of static structures, such as one-dimensional nanowires and 3D nano-structures (DNA origami); macro-sized materials, such as DNA hydrogels; or the construction of dynamic assemblies, including molecular motors, biosensors, fabricated nano-containers for controlled release (smart drug delivery systems), molecular logic gates, unidirectional DNA-walkers, surfaces for reversible cell adhesion; and reversible systems for the aggregation and dispersion of carbon nano-materials or hydrophobic dendrimers. In addition to the nano-technology applications, DNA has also recently regained attention as an attractive drug target in a wide variety of human pathological conditions, particularly due to the increasing functional importance of G4 structures for the regulation of gene expression and/or maintenance of genome integrity. Development in these fields strongly relies on our understanding of the DNA structure/dynamics.

Despite all of these intense efforts, our understating of some of the fundamental DNA properties, particularly DNA folding and its response to non-specific physical-chemical factors, remains limited. Undoubtedly, the major impediment to our understanding is a general lack of unbiased techniques that allow their investigation at a sufficient temporal resolution and in the complex environments of living cells. However, recent advances and the development of novel tools and methods for studying the DNA structure, such as in-cell spectroscopic techniques that allow us to investigate DNA structure and its interaction with drug-like molecules in the physiological environment, holds promise for the future.

5. Summary

Deoxyribonucleic acid (DNA) is abundant biopolymer in all living entities. In terms of biological function, DNA can be considered a code. While some of the DNA functions are coded by the DNA primary structure alone, such as RNA transcription, other DNA functions, such as those related to the regulation of gene expression or the maintenance of genome integrity, are encoded in its secondary, tertiary, and/or quaternary structure. While the primary DNA code, is rather simple and comprises only four "letters", the code of the higher order DNA structure is enormously complex. The complexity of this code is due to the inherent structural plasticity of the DNA that is finely tuned by a variety of non-specific environmental factors and/or specific interactions, such as those between DNA and DNA-binding proteins. In this habilitation thesis, I outline a role for non-specific physical-chemical factors in remodeling the DNA conformational space. The habilitation thesis summarizes my contribution to the field of DNA structural biology. In particular, the thesis recapitulates my contribution to the mechanistic understanding of the phenomenon of environmentally promoted and context-dependent structural polymorphisms of DNA and also recaps my contribution to the development of methods to investigate the DNA structure in the complex environment of living cells.

References:

1. Bacolla, A. and Wells, R.D. (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog*, **48**, 273-285.
2. Chuzhanova, N., Chen, J.M., Bacolla, A., Patrinos, G.P., Ferec, C., Wells, R.D. and Cooper, D.N. (2009) Gene conversion causing human inherited disease: evidence for involvement of non-B-DNA-forming sequences and recombination-promoting motifs in DNA breakage and repair. *Hum Mutat*, **30**, 1189-1198.
3. Iyer, R.R., Pluciennik, A., Napierala, M. and Wells, R.D. (2015) DNA triplet repeat expansion and mismatch repair. *Annu Rev Biochem*, **84**, 199-226.
4. Kamat, M.A., Bacolla, A., Cooper, D.N. and Chuzhanova, N. (2015) A Role for Non-B DNA Forming Sequences in Mediating Micro-Lesions Causing Human Inherited Disease. *Hum Mutat*.
5. Wu, Y. and Brosh, R.M., Jr. (2010) G-quadruplex nucleic acids and human disease. *FEBS J*, **277**, 3470-3488.
6. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.L., Thuong, N.T., Lhomme, J. and Helene, C. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-

[alpha]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res*, **15**, 7749-7760.
7. Moser, H.E. and Dervan, P.B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science*, **238**, 645-650.
8. Ali, A. and Bhattacharya, S. (2014) DNA binders in clinical trials and chemotherapy. *Bioorg Med Chem*, **22**, 4506-4521.
9. Bhatia, D., Sharma, S. and Krishnan, Y. (2011) Synthetic, biofunctional nucleic acid-based molecular devices. *Curr Opin Biotechnol*, **22**, 475-484.
10. Collie, G.W. and Parkinson, G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem Soc Rev*, **40**, 5867-5892.
11. Feldkamp, U. and Niemeyer, C.M. (2006) Rational design of DNA nanoarchitectures. *Angew Chem Int Ed Engl*, **45**, 1856-1876.
12. Heckel, A. and Famulok, M. (2008) Building objects from nucleic acids for a nanometer world. *Biochimie*, **90**, 1096-1107.
13. Krishnan, Y. and Simmel, F.C. (2011) Nucleic acid based molecular devices. *Angew Chem Int Ed Engl*, **50**, 3124-3156.
14. Sedoris, K.C., Thomas, S.D., Clarkson, C.R., Muench, D., Islam, A., Singh, R. and Miller, D.M. (2012) Genomic c-Myc quadruplex DNA selectively kills leukemia. *Mol Cancer Ther*, **11**, 66-76.
15. Wengel, J. (2004) Nucleic acid nanotechnology-towards Angstrom-scale engineering. *Org Biomol Chem*, **2**, 277-280.
16. Dickerson, R.E. and Ng, H.L. (2001) DNA structure from A to B. *Proc Natl Acad Sci U S A*, **98**, 6986-6988.
17. M, d.S. (2007) Geometric Formalism for DNA Quadruplex Folding. *Chem. Eur. J*, **13**, 9738-9745.
18. Database, N.A. http://ndbserver.rutgers.edu/.
19. Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D. and Shay, J.W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev*, **11**, 2801-2809.
20. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, **5**, 182-186.
21. Hansel, R., Lohr, F., Foldynova-Trantirkova, S., Bamberg, E., Trantirek, L. and Dotsch, V. (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res*, **39**, 5768-5775.
22. Phan, A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J*, **277**, 1107-1117.
23. Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol*, **19**, 414-422.
24. Crees, Z., Girard, J., Rios, Z., Botting, G.M., Harrington, K., Shearrow, C., Wojdyla, L., Stone, A.L., Uppada, S.B., Devito, J.T. *et al.* (2014) Oligonucleotides and G-quadruplex stabilizers: targeting telomeres and telomerase in cancer therapy. *Curr Pharm Des*, **20**, 6422-6437.
25. Xu, Y. (2011) Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.*, **40**, 2719-2740.
26. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*, **10**, 261-275.

27. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*, **99**, 11593-11598.
28. Dai, J., Carver, M. and Yang, D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172-1183.
29. Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N. and Phan, A.T. (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J Am Chem Soc*, **131**, 4301-4309.
30. Xue, Y., Kan, Z.Y., Wang, Q., Yao, Y., Liu, J., Hao, Y.H. and Tan, Z. (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition. *J Am Chem Soc*, **129**, 11185-11191.
31. Zhang, Z., Dai, J., Veliath, E., Jones, R.A. and Yang, D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res*, **38**, 1009-1021.
32. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876-880.
33. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*, **34**, 2723-2735.
34. Dai, J., Punchihewa, C., Ambrus, A., Chen, D., Jones, R.A. and Yang, D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res*, **35**, 2440-2450.
35. Hansel, R., Lohr, F., Trantirek, L. and Dotsch, V. (2013) High-resolution insight into G-overhang architecture. *J Am Chem Soc*, **135**, 2816-2824.
36. Skolakova, P., Foldynova-Trantirkova, S., Bednarova, K., Fiala, R., Vorlickova, M. and Trantirek, L. (2015) Unique C. elegans telomeric overhang structures reveal the evolutionarily conserved properties of telomeric DNA. *Nucleic Acids Res*, **43**, 4733-4745.
37. Tran, P.L., Mergny, J.L. and Alberti, P. (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res*, **39**, 3282-3294.
38. Bauer, N.C., Corbett, A.H. and Doetsch, P.W. (2015) The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Res*.
39. Berquist, B.R. and Wilson, D.M., 3rd. (2012) Pathways for repairing and tolerating the spectrum of oxidative DNA lesions. *Cancer Lett*, **327**, 61-72.
40. Breiling, A. and Lyko, F. (2015) Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin*, **8**, 24.
41. Konvalinova, H., Dvorakova, Z., Renciuk, D., Bednarova, K., Kejnovska, I., Trantirek, L., Vorlickova, M. and Sagi, J. (2015) Diverse effects of naturally occurring base lesions on the structure and stability of the human telomere DNA quadruplex. *Biochimie*.

42. Xu, B., Devi, G. and Shao, F. (2015) Regulation of telomeric i-motif stability by 5-methylcytosine and 5-hydroxymethylcytosine modification. *Org Biomol Chem*, **13**, 5646-5651.
43. Kypr, J., Kejnovska, I., Renciuk, D. and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res*, **37**, 1713-1725.
44. Long, X. and Stone, M.D. (2013) Kinetic partitioning modulates human telomere DNA G-quadruplex structural polymorphism. *PLoS One*, **8**, e83420.
45. Miyoshi, D., Matsumura, S., Li, W. and Sugimoto, N. (2003) Structural polymorphism of telomeric DNA regulated by pH and divalent cation. *Nucleosides Nucleotides Nucleic Acids*, **22**, 203-221.
46. Palacky, J., Vorlickova, M., Kejnovska, I. and Mojzes, P. (2013) Polymorphism of human telomeric quadruplex structure controlled by DNA concentration: a Raman study. *Nucleic Acids Res*, **41**, 1005-1016.
47. Prislan, I., Lah, J. and Vesnaver, G. (2008) Diverse polymorphism of G-quadruplexes as a kinetic phenomenon. *J Am Chem Soc*, **130**, 14161-14169.
48. Li, J., Correia, J.J., Wang, L., Trent, J.O. and Chaires, J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res*, **33**, 4649-4659.
49. Abu-Ghazalah, R.M., Rutledge, S., Lau, L.W., Dubins, D.N., Macgregor, R.B., Jr. and Helmy, A.S. (2012) Concentration-dependent structural transitions of human telomeric DNA sequences. *Biochemistry*, **51**, 7357-7366.
50. Amiri, A.R. and Macgregor, R.B., Jr. (2011) The effect of hydrostatic pressure on the thermal stability of DNA hairpins. *Biophys Chem*, **156**, 88-95.
51. Buscaglia, R., Miller, M.C., Dean, W.L., Gray, R.D., Lane, A.N., Trent, J.O. and Chaires, J.B. (2013) Polyethylene glycol binding alters human telomere G-quadruplex structure by conformational selection. *Nucleic Acids Res*, **41**, 7934-7946.
52. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K. and Miura, K. (1994) Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nucleases and heat. *Nucleic Acids Res*, **22**, 576-582.
53. Hud, N.V., Smith, F.W., Anet, F.A. and Feigon, J. (1996) The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR. *Biochemistry*, **35**, 15383-15390.
54. Jana, B., Pal, S., Maiti, P.K., Lin, S.T., Hynes, J.T. and Bagchi, B. (2006) Entropy of water in the hydration layer of major and minor grooves of DNA. *J Phys Chem B*, **110**, 19611-19618.
55. Kypr, J., Chladkova, J., Zimulova, M. and Vorlickova, M. (1999) Aqueous trifluorethanol solutions simulate the environment of DNA in the crystalline state. *Nucleic Acids Res*, **27**, 3466-3473.
56. Miller, M.C., Buscaglia, R., Chaires, J.B., Lane, A.N. and Trent, J.O. (2010) Hydration is a major determinant of the G-quadruplex stability and

conformation of the human telomere 3' sequence of d(AG3(TTAG3)3). *J Am Chem Soc*, **132**, 17105-17107.
57. Miyoshi, D., Fujimoto, T. and Sugimoto, N. (2013) Molecular crowding and hydration regulating of G-quadruplex formation. *Top Curr Chem*, **330**, 87-110.
58. Miyoshi, D., Muhuri, S., Mimura, K. and Sugimoto, N. (2009) DNA junction structure stabilized by molecular crowding conditions. *Nucleic Acids Symp Ser (Oxf)*, 59-60.
59. Miyoshi, D. and Sugimoto, N. (2008) Molecular crowding effects on structure and stability of DNA. *Biochimie*, **90**, 1040-1051.
60. Pramanik, S., Nagatoishi, S., Saxena, S., Bhattacharyya, J. and Sugimoto, N. (2011) Conformational flexibility influences degree of hydration of nucleic acid hybrids. *J Phys Chem B*, **115**, 13862-13872.
61. Sunami, T., Kondo, J., Tsunoda, M., Sekiguchi, T., Hirao, I., Watanabe, K., Miura, K. and Takenaka, A. (2002) X-ray structure of d(GCGAAGC); switching of partner for G:A pair in duplex form. *Nucleic Acids Res Suppl*, 181-182.
62. Nonin-Lecomte, S. and Leroy, J.L. (2001) Structure of a C-rich strand fragment of the human centromeric satellite III: a pH-dependent intercalation topology. *J Mol Biol*, **309**, 491-506.
63. Stosek, J. (2013) Influence of buffer composition on structure of nucleic acids. *Bachelor Theisis, Faculty of Science, Masaryk University*.
64. al., J.e. (2008) Vnitřní prostředí. *Grada Publishing a.s.*
65. Stefl, R., Wu, H., Ravindranathan, S., Sklenar, V. and Feigon, J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. *Proc Natl Acad Sci U S A*, **101**, 1177-1182.
66. Fiala, R., Spackova, N., Foldynova-Trantirkova, S., Sponer, J., Sklenar, V. and Trantirek, L. (2011) NMR cross-correlated relaxation rates reveal ion coordination sites in DNA. *J Am Chem Soc*, **133**, 13790-13793.
67. Albiez, H., Cremer, M., Tiberi, C., Vecchio, L., Schermelleh, L., Dittrich, S., Kupper, K., Joffe, B., Thormeyer, T., von Hase, J. *et al.* (2006) Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Res*, **14**, 707-733.
68. Hancock, R. (2007) Packing of the polynucleosome chain in interphase chromosomes: evidence for a contribution of crowding and entropic forces. *Semin Cell Dev Biol*, **18**, 668-675.
69. Nakano, S., Miyoshi, D. and Sugimoto, N. (2014) Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chem Rev*, **114**, 2733-2758.
70. Hansel, R., Foldynova-Trantirkova, S., Dotsch, V. and Trantirek, L. (2013) Investigation of quadruplex structure under physiological conditions using in-cell NMR. *Top Curr Chem*, **330**, 47-65.
71. Hansel, R., Luh, L.M., Corbeski, I., Trantirek, L. and Dotsch, V. (2014) In-cell NMR and EPR spectroscopy of biomacromolecules. *Angew Chem Int Ed Engl*, **53**, 10300-10314.
72. Fessl, T., Adamec, F., Polivka, T., Foldynova-Trantirkova, S., Vacha, F. and Trantirek, L. (2012) Towards characterization of DNA structure under

physiological conditions in vivo at the single-molecule level using single-pair FRET. *Nucleic Acids Res*, **40**, e121.
73. Hansel, R., Foldynova-Trantirkova, S., Lohr, F., Buck, J., Bongartz, E., Bamberg, E., Schwalbe, H., Dotsch, V. and Trantirek, L. (2009) Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy. *J Am Chem Soc*, **131**, 15761-15768.
74. Krstic, I., Hansel, R., Romainczyk, O., Engels, J.W., Dotsch, V. and Prisner, T.F. (2011) Long-range distance measurements on nucleic acids in cells by pulsed EPR spectroscopy. *Angew Chem Int Ed Engl*, **50**, 5070-5074.
75. Serber, Z. and Dotsch, V. (2001) In-cell NMR spectroscopy. *Biochemistry*, **40**, 14317-14323.
76. Serber, Z., Ledwidge, R., Miller, S.M. and Dotsch, V. (2001) Evaluation of parameters critical to observing proteins inside living Escherichia coli by in-cell NMR spectroscopy. *J Am Chem Soc*, **123**, 8895-8901.
77. Dotsch, V. (2008) Investigation of proteins in living bacteria with in-cell NMR experiments. *Top Curr Chem*, **273**, 203-214.
78. Ito, Y. and Selenko, P. (2010) Cellular structural biology. *Curr Opin Struct Biol*, **20**, 640-648.
79. Serber, Z., Selenko, P., Hansel, R., Reckel, S., Lohr, F., Ferrell, J.E., Jr., Wagner, G. and Dotsch, V. (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc*, **1**, 2701-2709.
80. Selenko, P., Serber, Z., Gadea, B., Ruderman, J. and Wagner, G. (2006) Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proc Natl Acad Sci U S A*, **103**, 11904-11909.
81. Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H. *et al.* (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature*, **458**, 106-109.
82. Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Walchli, M. *et al.* (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature*, **458**, 102-105.
83. Azarkh, M., Okle, O., Eyring, P., Dietrich, D.R. and Drescher, M. (2011) Evaluation of spin labels for in-cell EPR by analysis of nitroxide reduction in cell extract of Xenopus laevis oocytes. *J Magn Reson*, **212**, 450-454.
84. Azarkh, M., Okle, O., Singh, V., Seemann, I.T., Hartig, J.S., Dietrich, D.R. and Drescher, M. (2011) Long-range distance determination in a DNA model system inside Xenopus laevis oocytes by in-cell spin-label EPR. *Chembiochem*, **12**, 1992-1995.
85. Krstic, I., Endeward, B., Margraf, D., Marko, A. and Prisner, T.F. (2012) Structure and dynamics of nucleic acids. *Top Curr Chem*, **321**, 159-198.
86. Igarashi, R., Sakai, T., Hara, H., Tenno, T., Tanaka, T., Tochio, H. and Shirakawa, M. (2010) Distance determination in proteins inside Xenopus laevis oocytes by double electron-electron resonance experiments. *J Am Chem Soc*, **132**, 8228-8229.

87. Jagtap, A.P., Krstic, I., Kunjir, N.C., Hansel, R., Prisner, T.F. and Sigurdsson, S.T. (2015) Sterically shielded spin labels for in-cell EPR spectroscopy: analysis of stability in reducing environment. *Free Radic Res*, **49**, 78-85.
88. Salgado, G., F., Cazenave, C., Kerkour, A. and Mergny, J.L. (2015) G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chemical Science*, **6**, 3314-3320.
89. Fujimoto, T., Miyoshi, D., Tateishi-Karimata, H. and Sugimoto, N. (2009) Thermal stability and hydration state of DNA G-quadruplex regulated by loop regions. *Nucleic Acids Symp Ser (Oxf)*, 237-238.
90. Miyoshi, D., Karimata, H. and Sugimoto, N. (2007) Hydration regulates the thermodynamic stability of DNA structures under molecular crowding conditions. *Nucleosides Nucleotides Nucleic Acids*, **26**, 589-595.
91. Miyoshi, D., Nakamura, K., Tateishi-Karimata, H., Ohmichi, T. and Sugimoto, N. (2009) Hydration of Watson-Crick base pairs and dehydration of Hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions. *J Am Chem Soc*, **131**, 3522-3531.
92. Muhuri, S., Mimura, K., Miyoshi, D. and Sugimoto, N. (2009) Stabilization of three-way junctions of DNA under molecular crowding conditions. *J Am Chem Soc*, **131**, 9268-9280.
93. Nakano, M., Tateishi-Karimata, H., Tanaka, S., Tama, F., Miyashita, O., Nakano, S.I. and Sugimoto, N. (2015) Thermodynamic properties of water molecules in the presence of cosolute depend on DNA structure: a study using grid inhomogeneous solvation theory. *Nucleic Acids Res*.
94. Son, I., Shek, Y.L., Dubins, D.N. and Chalikian, T.V. (2014) Hydration changes accompanying helix-to-coil DNA transitions. *J Am Chem Soc*, **136**, 4040-4047.
95. Workman, H. and Flynn, P.F. (2009) Stabilization of RNA oligomers through reverse micelle encapsulation. *J Am Chem Soc*, **131**, 3806-3807.
96. Pramanik, S., Nagatoishi, S. and Sugimoto, N. (2012) DNA tetraplex structure formation from human telomeric repeat motif (TTAGGG):(CCCTAA) in nanocavity water pools of reverse micelles. *Chem Commun (Camb)*, **48**, 4815-4817.
97. Egli, M. (2002) DNA-cation interactions: quo vadis? *Chem Biol*, **9**, 277-286.
98. Halle, B. and Denisov, V.P. (1998) Water and monovalent ions in the minor groove of B-DNA oligonucleotides as seen by NMR. *Biopolymers*, **48**, 210-233.
99. Hoogstraten, C.G., Grant, C.V., Horton, T.E., DeRose, V.J. and Britt, R.D. (2002) Structural analysis of metal ion ligation to nucleotides and nucleic acids using pulsed EPR spectroscopy. *J Am Chem Soc*, **124**, 834-842.
100. Hud, N.V. and Feigon, J. (2002) Characterization of divalent cation localization in the minor groove of the A(n)T(n) and T(n)A(n) DNA sequence elements by (1)H NMR spectroscopy and manganese(II). *Biochemistry*, **41**, 9900-9910.
101. Hud, N.V., Sklenar, V. and Feigon, J. (1999) Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J Mol Biol*, **286**, 651-660.
102. Denisov, V.P. and Halle, B. (2000) Sequence-specific binding of counterions to B-DNA. *Proc Natl Acad Sci U S A*, **97**, 629-633.

103. Brumovska, E., Sychrovsky, V., Vokacova, Z., Sponer, J., Schneider, B. and Trantirek, L. (2008) Effect of local sugar and base geometry on 13C and 15N magnetic shielding anisotropy in DNA nucleosides. *J Biomol NMR*, **42**, 209-223.
104. Sychrovsky, V., Foldynova-Trantirkova, S., Spackova, N., Robeyns, K., Van Meervelt, L., Blankenfeldt, W., Vokacova, Z., Sponer, J. and Trantirek, L. (2009) Revisiting the planarity of nucleic acid bases: Pyramidilization at glycosidic nitrogen in purine bases is modulated by orientation of glycosidic torsion. *Nucleic Acids Res*, **37**, 7321-7331.
105. Sychrovsky, V., Muller, N., Schneider, B., Smrecki, V., Spirko, V., Sponer, J. and Trantirek, L. (2005) Sugar pucker modulates the cross-correlated relaxation rates across the glycosidic bond in DNA. *J Am Chem Soc*, **127**, 14663-14667.
106. Sychrovsky, V., Sochorova Vokacova, Z. and Trantirek, L. (2012) Guanine bases in DNA G-quadruplex adopt nonplanar geometries owing to solvation and base pairing. *J Phys Chem A*, **116**, 4144-4151.
107. Vokacova, Z., Trantirek, L. and Sychrovsky, V. (2010) Evaluating the effects of the nonplanarity of nucleic acid bases on NMR, IR, and vibrational circular dichroism spectra: a density functional theory computational study. *J Phys Chem A*, **114**, 10202-10208.
108. Gelbin, A., Schneider, B., Clowney, L., Hsieh, S., H., Olson, W., K. and Berman, H., M. (1996) Geometric parameters in nucleic acids: nitrogeous bases. *J. Am. Chem. Soc.*, **118**, 509-5518.
109. Dailey, M.M., Miller, M.C., Bates, P.J., Lane, A.N. and Trent, J.O. (2010) Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Res*, **38**, 4877-4888.
110. Gray, R.D. and Chaires, J.B. (2008) Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res*, **36**, 4191-4203.
111. Xue, Y., Liu, J.Q., Zheng, K.W., Kan, Z.Y., Hao, Y.H. and Tan, Z. (2011) Kinetic and thermodynamic control of G-quadruplex folding. *Angew Chem Int Ed Engl*, **50**, 8046-8050.
112. Bessi, I., Jonker, H.R., Richter, C. and Schwalbe, H. (2015) Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex. *Angew Chem Int Ed Engl*, **54**, 8444-8448.
113. Lannan, F.M., Mamajanov, I. and Hud, N.V. (2012) Human telomere sequence DNA in water-free and high-viscosity solvents: G-quadruplex folding governed by Kramers rate theory. *J Am Chem Soc*, **134**, 15324-15330.
114. Lindorff-Larsen, K., Piana, S., Dror, R.O. and Shaw, D.E. (2011) How fast-folding proteins fold. *Science*, **334**, 517-520.
115. Thirumalai, D., O'Brien, E.P., Morrison, G. and Hyeon, C. (2010) Theoretical perspectives on protein folding. *Annu Rev Biophys*, **39**, 159-183.
116. Best, R.B., Hummer, G. and Eaton, W.A. (2013) Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci U S A*, **110**, 17874-17879.
117. Karsisiotis, A., I.,, O'Kane, C. and da Silva, M., W.,. (2013) A Tutorial on Quadruplex Topologies. *Methods*, **64**, 28-35.

118. Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res*, **36**, 5482-5515.
119. Stadlbauer, P., Kuhrova, P., Banas, P., Koca, J., Bussi, G., Trantirek, L., Otyepka, M. and Sponer, J. (2015) Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res*.
120. Stadlbauer, P., Trantirek, L., Cheatham, T.E., 3rd, Koca, J. and Sponer, J. (2014) Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie*, **105**, 22-35.
121. Panda, D., Debnath, M., Mandal, S., Bessi, I., Schwalbe, H. and Dash, J. (2015) A Nucleus-Imaging Probe That Selectively Stabilizes a Minor Conformation of c-MYC G-quadruplex and Down-regulates c-MYC Transcription in Human Cancer Cells. *Sci Rep*, **5**, 13183.
122. Chen, Y.J., Groves, B., Muscat, R.A. and Seelig, G. (2015) DNA nanotechnology from the test tube to the cell. *Nat Nanotechnol*, **10**, 748-760.
123. Dong, Y., Yang, Z. and Liu, D. (2014) DNA nanotechnology based on i-motif structures. *Acc Chem Res*, **47**, 1853-1860.
124. Linko, V., Ora, A. and Kostiainen, M.A. (2015) DNA Nanostructures as Smart Drug-Delivery Vehicles and Molecular Devices. *Trends Biotechnol*, **33**, 586-594.
125. Madshus, I.H. (1988) Regulation of intracellular pH in eukaryotic cells. *Biochem J*, **250**, 1-8.
126. Matsuyama, S., Llopis, J., Deveraux, Q.L., Tsien, R.Y. and Reed, J.C. (2000) Changes in intramitochondrial and cytosolic pH: early events that modulate caspase activation during apoptosis. *Nat Cell Biol*, **2**, 318-325.
127. Matsuyama, S. and Reed, J.C. (2000) Mitochondria-dependent apoptosis and cellular pH regulation. *Cell Death Differ*, **7**, 1155-1165.
128. Nixon, V.L., McDougall, A. and Jones, K.T. (2000) Ca2+ oscillations and the cell cycle at fertilisation of mammalian and ascidian eggs. *Biol Cell*, **92**, 187-196.
129. Wadsworth, W.G. and Riddle, D.L. (1988) Acidic intracellular pH shift during Caenorhabditis elegans larval development. *Proc Natl Acad Sci U S A*, **85**, 8435-8438.

APENDIX

List of publications of the applicant included in this habilitation work

K *marks article corresponding author*

1: Stadlbauer P, Kuhrova P, Banas P, Koca J, Bussi G., Trantírek L, Otyepka M, Sponer J. Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Research, accepted, pii: gkv994.* **IF=9.1**

2K: Školáková P, Foldynová-Trantírková S, Bednářová K, Fiala R, Vorlíčková M, Trantírek L. Unique *C. elegans* telomeric overhang structures reveal the evolutionarily conserved properties of telomeric DNA. *Nucleic Acids Res*. **2015** 43(9):4733-45. **IF=9.1**

3: Konvalinová H, Dvořáková Z, Renčiuk D, Bednářová K, Kejnovská I, Trantírek L, Vorlíčková M, Sagi J. Diverse effects of naturally occurring base lesions on the structure and stability of the human telomere DNA quadruplex. *Biochimie*. **2015** 118:15-25. **IF=3.0**

4: Stadlbauer P, Trantírek L, Cheatham TE 3rd, Koča J, Sponer J. Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie* **2014** 105:22-35. **IF=3.0**

5: Hänsel R, Luh LM, Corbeski I, Trantirek L, Dötsch V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angewandte Angew Chem Int Ed Engl*. **2014** 53(39):10300-14. **IF=11.1**

6: Hänsel R, Löhr F, Trantirek L, Dötsch V. High-resolution insight into G-overhang architecture. *J Am Chem Soc*. **2013** *135*(7):2816-24. **IF=12.1**

7K: Hänsel R, Foldynová-Trantírková S, Dötsch V, Trantírek L. Investigation of quadruplex structure under physiological conditions using in-cell NMR. *Top Curr Chem*. **2013** *330*:47-65. **IF= 5.6**

8K: Fessl T, Adamec F, Polívka T, Foldynová-Trantírková S, Vácha F, Trantírek L. Towards characterization of DNA structure under physiological conditions in vivo at the single-molecule level using single-pair FRET. *Nucleic Acids Res*. **2012** *40*(16):e121. **IF= 9.1**

9: Sychrovský V, Sochorová Vokáčová Z, Trantírek L. Guanine bases in DNA G-quadruplex adopt nonplanar geometries owing to solvation and base pairing. *J Phys Chem A*. **2012** 116(16):4144-51. **IF=2.7**

10K: Fiala R, Spacková N, Foldynová-Trantírková S, Sponer J, Sklenár V, Trantírek L. NMR cross-correlated relaxation rates reveal ion coordination sites in DNA. *J Am Chem Soc*. **2011** 133(35):13790-3. **IF=12.1**

11K: Hänsel R, Löhr F, Foldynová-Trantírková S, Bamberg E, Trantírek L, Dötsch V. The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res*. **2011** *39*(13):5768-75. **IF=9.1**

12K: Vokácová Z, Trantírek L, Sychrovský V. Evaluating the effects of the nonplanarity of nucleic acid bases on NMR, IR, and vibrational circular dichroism spectra: a density functional theory computational study. *J Phys Chem A*. **2010** 114(37):10202-8. **IF=2.7**

13K: Sychrovsky V, Foldynova-Trantirkova S, Spackova N, Robeyns K, Van Meervelt L, Blankenfeldt W, Vokacova Z, Sponer J, Trantirek L. Revisiting the planarity of nucleic acid bases: Pyramidilization at glycosidic nitrogen in purine bases is modulated by orientation of glycosidic torsion. *Nucleic Acids Res.* **2009** 37(21):7321-31. **IF=9.1**

14K: Hänsel R, Foldynová-Trantírková S, Löhr F, Buck J, Bongartz E, Bamberg E, Schwalbe H, Dötsch V, Trantírek L. Evaluation of parameters critical for observing nucleic acids inside living Xenopus laevis oocytes by in-cell NMR spectroscopy. *J Am Chem Soc*. **2009** *131*(43):15761-8. **IF=12.1**

15K: Brumovská E, Sychrovský V, Vokácová Z, Sponer J, Schneider B, Trantírek L. Effect of local sugar and base geometry on 13C and 15N magnetic shielding anisotropy in DNA nucleosides. *J Biomol NMR*. **2008** 42(3):209-23. **IF=3.1**

16K: Sychrovský V, Müller N, Schneider B, Smrecki V, Spirko V, Sponer J, Trantírek L. Sugar pucker modulates the cross-correlated relaxation rates across the glycosidic bond in DNA. *J Am Chem Soc*. **2005**;127(42):14663-7. **IF=12.1**