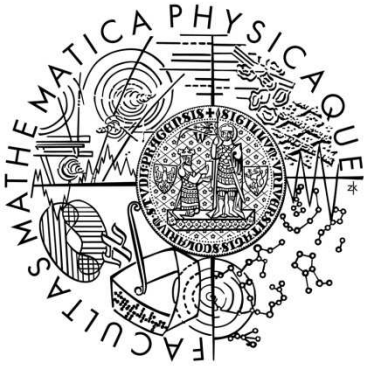
Charles University in Prague

Faculty of Mathematics and Physics

# DOCTORAL THESIS




Mgr. Jan Palacký

## Biologically important non-canonical structures of nucleic acids in complexes with cationic porphyrins

Institute of Physics, Division of Bimolecular Physics

Supervisor of the doctoral thesis: Assoc. Prof. Peter Mojzeš, CSc.

Study programme: Physics
Specialization: Biophysics, chemical and macromolecular physics

Prague 2012

# Acknowledgments

At this point, I would like to express my deepest appreciation to all the staff members as well as the students at the Division of Biomolecular Physics at Institute of Physics of Charles University in Prague for making my long stay there amazing experience that will stay with me for the rest of my life. First and foremost, my utmost gratitude goes to Associate professor Peter Mojzeš, for his mentorship, guidance and patience during my doctoral study. He always looked at the positive side of things and encouraged me to keep doing the research at the moments of my frequent scientific despair and depression. He was always at hand when I posed awkward questions to him or needed small experimental help. His attitude to science, in general, showed me that scientific life is the best, happiest and fullest way of living. The determining moment of my doctoral study was that my supervisor introduced me to Professor Michaela Vorlíčková and Dr. Iva Kejnovská from Institute of Biophysics of the Academy of Sciences of the Czech Republic in Brno. I would like to thank them for their cooperation. Without the joint experiments and intensive personal correspondence between our and their group, my thesis would never have been brought to completion. My great thanks also belong to Dr. Lucie Bednárová who made me possible to use a CD spectrometer in her laboratory at Institute of Organic Chemistry and Biochemistry of the Academy of Sciences of the Czech Republic in Prague and who assisted me with my experiments with porphyrins. I am also indebted to Dr. Ivan Barvík for introducing me to the basics of molecular dynamics simulations of biomolecules, Professor Josef Štěpánek for theoretical discussions concerning the extraction of thermodynamical data from calorimetric and spectroscopic experiments, Associate professor Jiří Bok from theoretical department for providing me with the software for pre-processing of Raman spectra and for some practical numerical advice, and to Dr. Jan Lang from Department of Low Temperature Physics of Charles University in Prague for performing some NMR experiments for us. Last but not least, I appreciate continuous support from my family and their understanding for my obsession with my work. I also highly acknowledge financial support from the Grant Agency of the Charles University in Prague (the project number 122208).

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In…Prague..... date..... signature

Název práce: Studium biologicky významných nekanonických struktur nukleových kyselin v komplexech s kationickými porfyriny

Autor: Jan Palacký
Katedra / Ústav: Fyzikální ústav, Oddělení fyziky biomolekul
Vedoucí doktorské práce: doc. RNDr Peter Mojzeš CSc., Fyzikální ústav, Oddělení fyziky biomolekul
Abstrakt:

Guaninové kvadruplexy tvoří třídu neobvyklých konformací nukleových kyselin založených na stohování guaninových tetrád stabilizovaných prostřednictvím Hoogsteenovského párování a koordinace kationtů. Podílejí se celé řadě buněčných procesů včetně replikace, rekombinace nebo transkripce. Guaninové kvadruplexy jsou velmi rozšířené v lidském genomu, ale nejvíce se vyskytují v jednovláknových oblastech telomer bohatých na guanin. O telomerní guaninové kvadruplexy roste zájem díky jejich schopnosti inhibovat aktivitu enzymu telomeráza, který je zodpovědný za proliferaci nádorových buněk. Konkrétně jsme zkoumali konformační polymorfismus základní lidské telomerní sekvence $G_3(TTAG_3)_3$ vyvolaný koncentrací DNA, kovových kationtů ($K^+$, $Na^+$) a/nebo teplotní renaturací. Ramanova spektroskopie byla použita jako primární metoda pro toto studium, neboť, na rozdíl od běžných spektroskopických metod, jsme mohli sledovat strukturu kvadruplexu za velmi vysokých koncentrací DNA napodobujících podmínky molekulárního shlukování v buňce. Ukazujeme, že kvadruplex $G_3(TTAG_3)_3$ přechází mezi antiparalelním a paralelním uspořádáním řetězců v závislosti na koncentraci DNA a iontů $K^+$. Kromě toho také ukazujeme, že kationické porfyriny mohou sloužit jako citlivé sondy konformace kvadruplexů. Závěrem diskutujeme možnou dimerizaci zajímavého guaninového kvadruplexu známého jako jako trombinový aptamer.

Klíčová slova: DNA, guanin, kvadruplex, porfyriny, Ramanova spektroskopie

Title: Biologically important non-canonical structures of nucleic acids in complexes with cationic porphyrins

Author: Jan Palacký
Department / Institute: Institute of Physics, Division Of Bimolecular Physics
Supervisor of the doctoral thesis: Assoc. Prof. Peter Mojzeš, CSc., Institute of Physics, Division Of Bimolecular Physics
Abstract:

Guanine quadruplexes are a class of unusual nucleic acids conformations based on stacked planar guanine tetrads stabilized via Hoogsteen pairing and cation coordination. They are implicated in numerous cellular processes including replication, recombination or transcription. Guanine quadruplexes are widespread within the human genome but their occurrence is highest in the single stranded guanine-rich regions at telomeres. Telomeric guanine quadruplexes are gaining growing interest due to their ability to inhibit the activity of the telomerase enzyme, which is responsible for the proliferation of tumor cells. Specifically, we investigated the conformational polymorphism of the human core telomeric sequence $G_3(TTAG_3)_3$ conditioned by the concentration of DNA, metal cations ($K^+$, $Na^+$) and/or annealing. Raman spectroscopy was employed as the primary method for this study because, unlike common spectroscopic methods, it allowed us to monitor the quadruplex structure at very high DNA concentrations mimicking molecular crowding conditions in the cell. We demonstrate that the $G_3(TTAG_3)_3$ quadruplex switch between the antiparallel and parallel strand alignment as a function of nucleoside and potassium concentration. In addition, we demonstrate that cationic porphyrins can be used as sensitive probes of the quadruplex conformation. Finally, we discuss possible dimerization of interesting guanine quadruplex known as thrombin binding aptamer.

Keywords: DNA, guanine, quadruplex, porphyrins, Raman spectroscopy

# Contents

# Preface

The presented thesis summarizes the research done during my PhD study at the Division of Biomolecular Physics at Institute of Physics of Charles University in Prague. The primary objective of the dissertation was to investigate the biophysical properties of non-canonical DNA structures. The whole work deals exclusively with an interesting group of the four-stranded nucleic acids structures known as guanine (G-) quadruplexes or tetraplexes. These structures are ubiquitous in the human genome. It is believed that they are involved in numerous cellular processes such as the regulation of gene expression or recombination.

G-quadruplexes occurring at the single-stranded G-rich regions of telomeres are most widely studied since their formation and stabilization may modulate the activity of telomerase enzyme that is responsible for immortality and proliferation of cancer cells. Human telomeric DNA sequences are highly polymorphic in nature. Despite intensive research, their conformations assumed under physiological conditions remain largely unknown. A myriad of experimental data available to date suggests that the structure of human telomere is a consequence of the complex interplay of a range of factors such as the exact oligonucleotide sequence and concentration, the type and concentration of stabilizing metal cations, the presence of molecular ligands, dehydration and molecular crowding. Specifically, it has been recently proposed that drastically different oligonucleotide concentrations appropriate for various experimental techniques such as CD and NMR spectroscopy may be a reason of why contradictory structures are revealed under otherwise similar conditions **[Renčiuk, 2009]**. We decided to clarify this matter by employing Raman spectroscopy as a bridge between the methods restricted to low and high DNA concentrations only.

Our work is pioneering in many respects. We followed by Raman spectroscopy the switch between the antiparallel and parallel quadruplex arrangement of the human core telomeric sequence $G_3(TTAG_3)_3$ induced by oligonucleotide concentration, potassium concentration or/and annealing. Raman spectroscopy has been rarely applied for studying quadruplex nucleic acids. Our study is one of the few systematic studies in a decade employing conventional non-resonant Raman spectroscopy to probe G-quadruplex structures in solution. Thermal melting measurements carried out by us at extremely high nucleoside concentrations are first

of its kind. The possibility of Raman spectroscopy to characterize high concentrated DNA samples mimicking the conditions in the cell demonstrated by our study may attract interest in the application of Raman spectroscopy among the scientific community interested in G-quadruplexes. The Raman data on human telomere DNA were confronted with CD spectroscopic measurements and native polyacrylamide gel electrophoresis conducted in cooperation with Dr. Iva Kejnovská and prof. Michaela Vorlíčková at Institute of Biophysics of the Academy of Sciences of the Czech Republic (Brno). Long-term cooperation between our and their group culminated in a joint publication concerning polymorphism of human telomeric quadruplex structure controlled by DNA concentration **[Palacký et al., 2012, submitted, in attachments, pages 195-220]**.

Apart from telomeric G-quadruplexes, this thesis also deals with other guanine quadruplex structures. We measured resonance Raman spectra of cationic porphyrins in order to characterize their interaction with various intramolecular and intermolecular G-quadruplexes. To our knowledge, the similar Raman study has not been published yet. The primary idea was to use cationic porphyrins as sensitive and specific probes for G-quadruplex DNA. Raman scattering experiments were correlated with the data provided by absorption and CD spectroscopy. CD spectra were recorded again in the group of prof. Michaela Vorlíčková but also in the group of Dr. Lucie Bednárová from Institute of Organic Chemistry and Biochemistry of the Academy of Sciences of the Czech Republic (Prague). The obtained results are convincing although they do not necessarily bring new insights into understanding of G-quadruplexes. We simply demonstrate that cationic porphyrins can be employed to follow indirectly the structural changes in the conformation of DNA guanine quadruplexes provoked by varying experimental conditions such as introducing specific metal cations or changing temperature.

Of all the studied nontelomeric G-rich sequences, we narrowed our detailed research to TBA (thrombin binding aptamer). TBA is a simple model system of a guanine quadruplex, having low thermal stability and only two G-tetrads. This quadruplex is most frequently studied DNA aptamer. It is biologically important because it binds with high affinity to thrombin enzyme, which is implicated in blood clotting. It is widely accepted that TBA is an intramolecular G-quadruplex. However, on the basis of bimolecular-like electrophoretic migration of TBA in native

polyacrylamide gels, it was previously proposed that TBA is in fact bimolecular structure **[Fialová et al., 2006]**.

In an effort to clarify the reason why distinct experimental techniques (NMR, gel electrophoresis) provide inconsistent results concerning the molecularity of TBA, we combined a host of experimental methods including optical spectroscopic methods, NMR, differential scanning calorimetry and gel electrophoresis to probe the solution structure of TBA and its interaction with cationic porphyrins. A crucial finding of our study is the observation of a strong induced CD signal of the bisignate shape for mixtures of TBA with the copper cationic porphyrin. Based on this and other results, we conclude that TBA aptamer forms some kind of dimeric arrangement rather than an intramolecular G-quadruplex, as it is generally believed.

The presented results on G-quadruplexes could not be obtained without the use of sophisticated multivariate techniques (based on SVD) for processing of very large quantities of spectroscopic data. During the course of doctoral study, a range of useful programs were written in Matlab® (version 7.4, Mathworks) in order to simplify data treatment and visualization. They include, for instance, the program for manual correction of background, the program for intensity normalization and solvent subtraction exploiting the properties of water stretching vibrations or the program for visualization of the results of 2D imaging experiments. It is worth noting that our semi-automated procedure for spectra normalization and solvent subtraction exploiting the properties of water stretching vibrations was recently published **[Palacký et al., 2011, in attachments, pages 227-250]**.

# 1. Introduction

## 1.1 Nucleic acid structure – basic concepts

The nucleic acids are biopolymers composed of repeating building blocks called nucleotides. The nucleotide can be divided into the three major components, that is, a pentose sugar, a heterocyclic nucleobase and one or more phosphate groups. The most frequently occurring nucleobases in nucleic acids are adenine, guanine (derived from purine) and thymine, cytosine and uracil (derived from pyrimidine). Some other less common nucleobases such as inosine are normally found in natural nucleic acids (especially in RNA), but in a rather small amount. The major difference between the structure of DNA and RNA is that the sugar in RNA has extra oxygen at the 2´ carbon. Furthermore, uracil is almost exclusively found in RNA, whereas thymine is much more common in DNA. The nucleotide that lacks the phosphate group(s) is a called nucleoside. Nucleosides can be phosphorylated at hydroxyls linked to the 2´ (in RNA), 3´ and 5´ carbons. As a convention, the DNA or RNA strand directionality is defined in terms of the orientation of phosphodiesther linkage (from the carbon C5´ to C3´). Two chains of natural nucleic acids assuming double helical structure run normally in the opposite directions. Short DNA or RNA fragments, denoted as oligonucleotides, can be nowadays routinely synthesized at high purity for required base sequences. It is quite common to synthesize oligonucleotides that do not occur in nature such as those with the phosphodiesther bond linking C2´ and C5´ carbons or those having some chemical modifications on the nucleobases or sugar rings. As opposed to aromatic nitrogen heterocycles of nucleobases, which are more or less planar, the geometry of (deoxy) – ribofuranose rings of DNA (RNA) strongly deviates from planarity **[Saenger, 1983]**. The conformation of the sugar moiety is defined in terms of 5 endocyclic torsion angles $\nu_0$, $\nu_1$, $\nu_2$, $\nu_3$, $\nu_4$ (see figure 1 and table 1), and it is theoretically best represented by the pseudorotation cycle **[Saenger, 1983]** that introduces two parameters, namely the pseudorotation ring torsion angle (pseudorotation amplitude) and the pseudorotation phase angle (see figure 2). The pseudorotation cycle was first employed for cyclopentane, where there is practically no difference between individual puckering states.

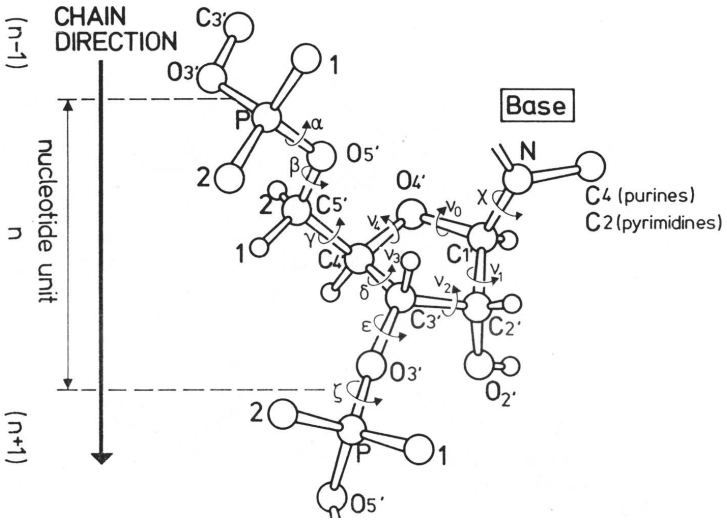



**Figure 1.** Atomic numbering scheme and definitions of torsions angles for polyribonucleotide chain. Taken from *[Saenger, 1983]*.

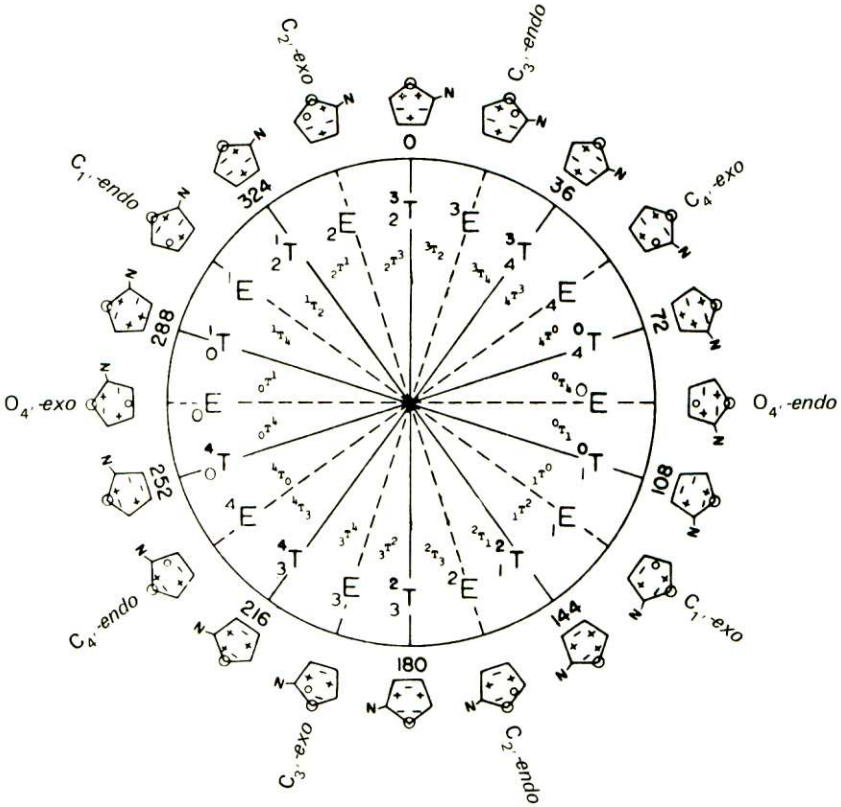



**Figure 2.** Pseudorotation cycle of the furanose ring in nucleosides. Values of phase angles are given in multiples of 36°. Envelope E and twist T forms alternate every 18°. Sugar puckers differing by 180° are mutual mirror images. The signs (+) and (-) indicate positive and negative torsion angles, respectively. Taken from *[Saenger, 1983]*.

| Torsion angle | Atoms involved |
|---|---|
| $\alpha$ | $_{(n-1)}O_{3'} - P - O_{5'} - C_{5'}$ |
| $\beta$ | $P - O_{5'} - C_{5'} - C_{4'}$ |
| $\gamma$ | $O_{5'} - C_{5'} - C_{4'} - C_{3'}$ |
| $\delta$ | $C_{5'} - C_{4'} - C_{3'} - O_{3'}$ |
| $\varepsilon$ | $C_{4'} - C_{3'} - O_{3'} - P$ |
| $\xi$ | $C_{3'} - O_{3'} - P - O_{5'(n-1)}$ |
| $\chi$ | $O_{4'} - C_{1'} - N_1 - C_2$ (pyrimidines)<br>$O_{4'} - C_{1'} - N_9 - C_4$ (purines) |
| $\nu_0$ | $C_{4'} - O_{4'} - C_{1'} - C_{2'}$ |
| $\nu_1$ | $O_{4'} - C_{1'} - C_{2'} - C_{3'}$ |
| $\nu_2$ | $C_{1'} - C_{2'} - C_{3'} - C_{4'}$ |
| $\nu_3$ | $C_{2'} - C_{3'} - C_{4'} - O_{4'}$ |
| $\nu_4$ | $C_{3'} - C_{4'} - O_{4'} - C_{1'}$ |

**Table 1.** Definition of torsion angles in nucleotides. Adapted from ***[Saenger, 1983]***.

The pseudorotation phase angle P can be obtained by means of the endocyclic torsion angles $\nu_0$, $\nu_1$, $\nu_2$, $\nu_3$, $\nu_4$ according to equation 1.1.

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

Torsions angles are related to the phase angle P through equation 1.2, where $j$ = goes from 0 to 4, $\varphi$ = 720°/5 = 144° and $\nu_{max}$ is sugar puckering amplitude **[Bloomfield et al., 2000]**:

$$\nu_j = \nu_{max} \cdot \cos(P + (j - 2) \cdot \varphi) \quad (1.2)$$

As can be expected, asymmetric attachment of substituent groups results in the restriction of conformational flexibility. This is the case of nucleic acids having the purine or pyrimidine bases connected to the $C_{1'}$ carbon of the sugar where one can predominately observe either $C_{2'}$-endo (144° ≤ P ≤ 180°) or $C_{3'}$- endo (0° ≤ P ≤ 36°) sugar puckers **[Saenger, 1983]**. The conformations $C_{2'}$- endo and $C_{3'}$- endo can be simply envisioned as such deviation of the carbons $C_{2'}$ and $C_{3'}$, respectively, that these lies in the same half-plane as the carbon $C_{5'}$ with respect to the plane formed by the remaining atoms.

Apart from the endocyclic torsion angles, the structure of nucleic acids can be characterized by the arrangement of the sugar phosphate backbone, which is

identified with the sequence of four atoms P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (see figure 1). The conformational flexibility of the backbone is expressed in terms of six torsion angles, that is, $\alpha$, $\beta$, $\gamma$, $\delta$, $\varepsilon$, $\zeta$ (see figure 1 and table 1). Each torsion angle generally varies only over a certain range of values due to steric clashes between surrounding groups of atoms. The common terminology for describing the range of torsions angles has been introduced by Klyne and Prelog (see figure 3).

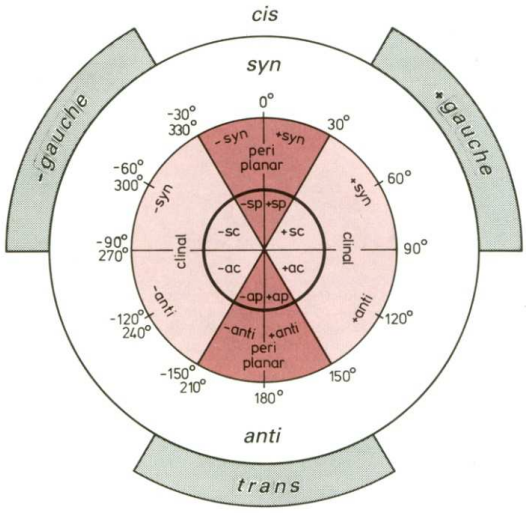



**Figure 3.** Schematic correlation of torsion angle ranges used by spectroscopists (*cis*, *trans*, *+gauche*, *-gauche*) with ranges defined by Klyne and Prelog (*syn* or *synperiplenar*, *anti* or *antiperiplenar*, *+synclinal*, *- synclinal*, *+anticlinal*, *-anticlinal*). Taken from ***[Saenger, 1983]***.

The orientation of nucleobase around N-glycosidic bond is defined by the torsion angle $\chi$ (see figure 1 and table 1), which is normally restricted to the two major orientations, that is *syn* or *anti* **[Bloomfield et al., 2000]**. Roughly speaking, the conformation *anti* is characterized by less overlap between the sugar and the base than the *syn* conformation (pyrimidine ring in purine and $O_2$ in pyrimidine bases is directed towards the sugar in the *syn* conformation). Helical double-stranded nucleic acid structures, mononucleotides, as well as pyrimidine nucleosides prefer the *anti* over *syn* sugar orientation **[Bloomfield et al., 2000; Saenger et al., 1983]**. The ratio of *syn* to *anti* conformers can be influenced by the nucleobase modification. The nucleobases with the bulky substituents can be found rather in the *syn* than in the *anti* conformation (see figure 4) **[Bloomfield et al., 2000; Saenger et al., 1983]**.

**Figure 4.** Two different orientations of nucleobases around the glycosidic bond. While uridine has the base in the *anti* conformation (on the left), its 6-methylated derivative prefers the *syn* geometry (on the right). Taken from ***[Saenger, 1983]***.

There is a notable correlation between the previously mentioned sugar puckering modes and the torsion angle $\chi$. The *syn* conformation of glycosidic bond is associated with the $C_2'$- endo puckering, whereas the *anti* conformation is more typically found with the $C_3'$- endo sugar puckers **[Bloomfield et al., 2000; Saenger, 1983]**. Interestingly, RNA helical nucleic acids or oligonucleotides are more prone to adopt the $C_3'$- endo puckering then their DNA counterparts, which prefer the $C_2'$- endo puckering. The summary of the prevailing sugar puckering modes and the corresponding base orientations is compiled for comparison purposes in table 2.

Nucleic acids are highly polymorphic molecules **[Bloomfield et al., 2000; Kypr et al., 2009; Saenger, 1983]**. These include various helical structures formed by one **[Bell et al., 1997]**, two (duplexes) **[Movileanu et al., 2002; Ottová et al., 2011]**, three (triplexes) **[Herrera & Štěpánek, 2010; Liquier et al., 1999]**, four (tetramolecular guanine quadruplexes **[Mergny et al., 2005a]** and cytosine quadruplexes **[Benevides et al., 1996]**) and five or more nucleic acid strands **[Borbone et al., 2011]**, as well as hairpins **[Baumruk et al., 2001; Vachoušek & Štěpánek, 2008]**, RNA-DNA heteroduplexes **[Kříž et al., 2008; Vondrušková et al., 2008]** and other less frequent conformations **[Bacolla & Wells, 2004]**.

Right-handed helical DNA and RNA conformations can be roughly divided into two distinct structural categories known as A and B family **[Saenger, 1983]**. The the two families are primarily distinguished according to the pucker of furanose rings. Specifically, whereas polynucleotides of the family A are found exclusively with $C_3'$- endo sugar pucker, the family B is predominately associated with $C_2'$- endo sugar puckering ($C_1'$- exo, $C_3'$- exo, $C_4'$- exo or $O_4'$- endo puckers are, however, also possible). **[Bloomfield et al., 2000; Saenger, 1983]**. $C_3'$- endo

puckering is reflected in a short interphosphate distance of 5.9 Å. $C_2$´- endo is associated with a significantly higher interphosphate distance of 7.0 Å. As a result, the polynucleotides belonging to the family B exhibits larger conformational variability than the polynucleotides of the family A (the average structural parameters for both nucleic acid families are given in table 3) **[Saenger, 1983]**. Accordingly, RNA duplexes can be found only in two closely related forms A and A´(A´ conformation prevails at higher salt concentrations), whereas DNA duplexes adopt not only A conformation (typical of RNA) but also a range of other conformations (B, B´, C, C´, C´´, D, E, T) belonging to the B family **[Saenger, 1983]**. The comparison of A, B and Z form (see below) of DNA is given in figure 5.

The particular DNA (RNA) conformation is dependent on many factors such as the sequence context, ionic strength, temperature, pH, presence of other interacting molecules or chemical modifications. Generally, right-handed B-form of DNA is the predominant conformation at physiological conditions (high relative humidity, and the presence of $K^+$ and/or $Na^+$ cations). The A-DNA conformation is limited to dehydrated conditions (and higher ionic strengths) typical of crystals. Most RNA-DNA hybrids usually adopt duplex or triplex conformations with the A geometry **[Saenger, 1983]**. Although RNA molecules can form duplex or triplex structures, they are more commonly found as single-stranded (with the exception of some RNA viruses). Single-stranded RNA structures are known for their high polymorphism, partly due to a high percentage of non-canonical bases or/and base pairs, as compared to DNA. The example of a very complex RNA secondary and tertiary structure is a nucleoprotein complex called ribosome.

Apart from the above-mentioned right-handed nucleic acid structures, DNA as well RNA can adopt left-handed structure known as Z-form (in fact, there are several types of Z-DNA, similarly to the B family of structures). The Z DNA helix is typically observed for alternating purine-pyrimidine sequences in the presence of high salt concentrations or alcohol **[Bloomfield et al., 2000; Saenger, 1983; Vorlíčková et al., 2012]**. A basic repeat unit of Z-structure forms two nucleotides: purine in *C3'-endo/syn* and pyrimidine in *C2'-endo/anti* conformation with sugar-phosphate bacbone running in zigzag manner (see figure 5). The selected structural parameters for Z-form of DNA are tabulated in table 3. Interestingly, the Z-form was also observed for non-alternating purine-pyrimidine sequences and with trinucleotide repeats **[Vorlíčková et al., 2012; Wang et al., 1985]**.

| | Base / sugar | Pyrimidine | | Purine | |
|---|---|---|---|---|---|
| | | Sugar puckering | Base orientation | Sugar puckering | Base orientation |
| **Nucleoside** | **Ribose** | *$C_2$-endo*<br>*$C_3$-endo* | ***anti***<br>*syn* | *$C_2$-endo*<br>*$C_3$-endo* | *anti*<br>*syn* |
| | **Deoxyribose** | ***$C_2$-endo***<br>*$C_3$-endo* | ***anti***<br>*syn* | ***$C_2$-endo***<br>*$C_3$-endo* | *anti*<br>*syn* |
| **Nucleotide** | **Ribose** | *$C_2$-endo*<br>*$C_3$-endo* | *anti* | *$C_2$-endo*<br>*$C_3$-endo* | *anti* |
| | **Deoxyribose** | *$C_2$-endo*<br>*$C_3$-endo* | *anti* | *$C_2$-endo*<br>*$C_3$-endo* | *anti* |
| **Helical polynucleotide** | **Ribose** | *$C_3$-endo* | *anti* | *$C_3$-endo* | *anti* |
| | **Deoxyribose** | *$C_2$-endo*<br>*$C_3$-endo* | *anti* | *$C_2$-endo*<br>*$C_3$-endo* | *anti* |

**Table 2.** Preferred sugar puckers and base orientations for nucleosides and poly(nucleotides). Dominating conformations are depicted in red bold font. Adapted from *[Saenger, 1983]*.

| **DNA form** | | **A** | **B** | **Z** |
|---|---|---|---|---|
| **Helix sense** | | right-handed | right-handed | left-handed |
| **Diameter** | | 2.6 | 2.0 | 1.8 |
| **Base pairs per one turn** | | 11 | 10 | 12 (6 CG dimers) |
| **Rotation per one base pair** | | 32.7° | 36° | 10° for pyrimidine-purine<br>50° for purine-pyrimidine |
| **Height per one turn** | | 3.20 | 3.40 | 4.50 |
| **Inclination of base pairs to helical axis** | | +12° | +2.4° | -6.2° |
| **Deviation of base pairs from helical axis** | | 0.41 | 0.08 | 0.3 |
| **Major groove** | **depth** | 1.35 | 0.85 | shallow |
| | **width** | 0.27 | 1.17 | |
| **Minor groove** | **depth** | 0.28 | 0.75 | 0.90 |
| | **width** | 1.10 | 0.57 | 0.40 |
| **The puckering modes of sugar** | | *$C_3$-endo* | *$C_2$-endo* | *$C_2$-endo* for pyrimidines<br>*$C_3$-endo* for purines |
| **The orientation of N-glycosidic bond (angle χ)** | | *anti* | *anti* | *anti* for pyrimidines<br>*syn* for purines |

**Table 3.** The selected values of average structural parameters for three basic DNA conformations. Numerical values are in [nm]. Adapted from *[Bloomfield et al., 2000]*.

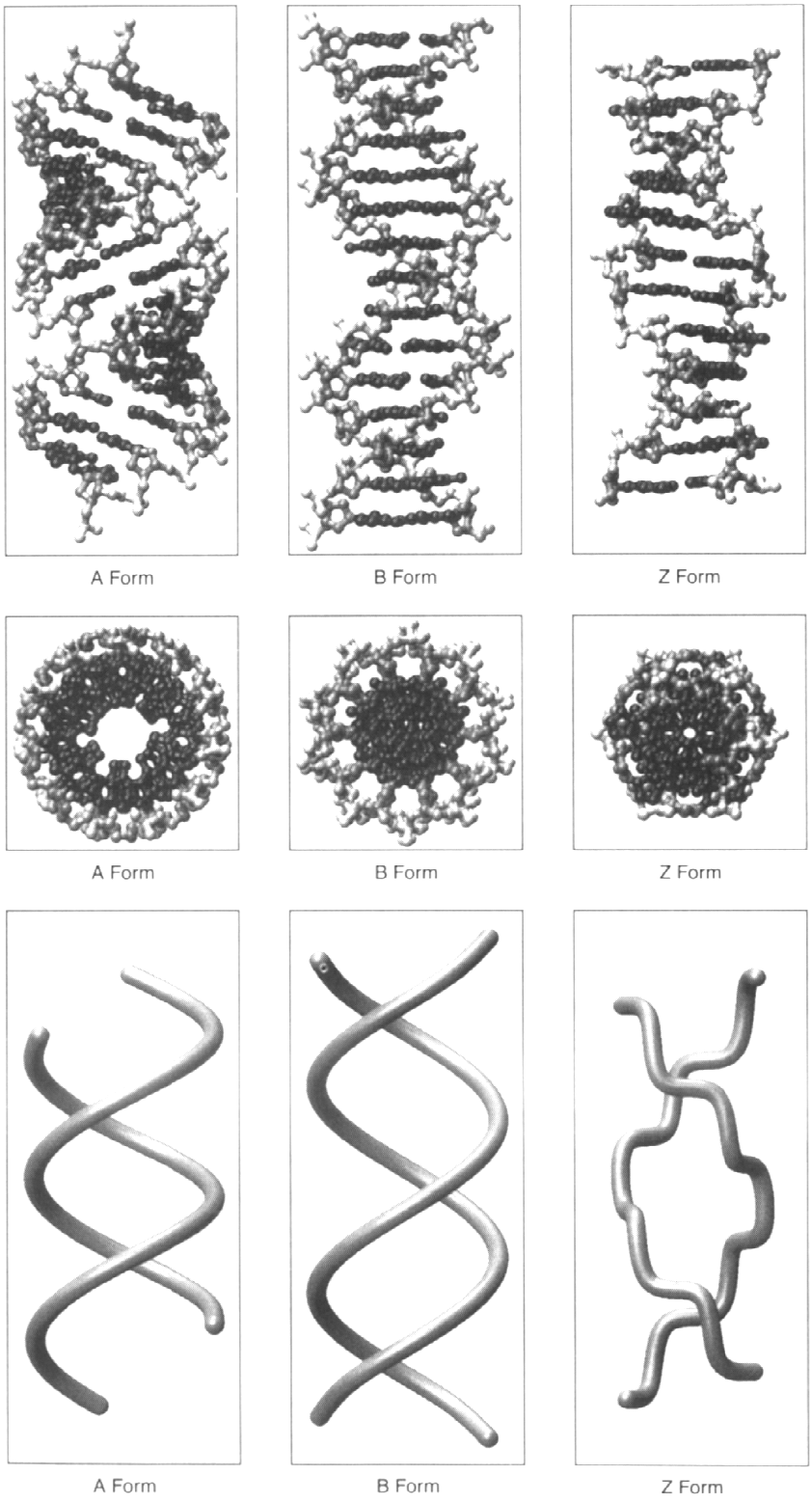



**Figure 5.** Three different representations of 3D structures of double helical DNA having 20 base pairs for the three most commonly occurring DNA structural forms: A, B and Z. Adapted from *[Bloomfield et al., 2000]*.

## 1.2 Non-canonical nucleic acid structures – a focus on guanine quadruplexes

Guanine (G-) quadruplexes (known also as guanine tetraplexes or simply quadruplexes) represent a diverse class of nucleic acid structures that are constituted by oligo and polynucleotides with a high content of consecutive guanine residues in their primary sequences. They differ from canonical nucleic acid structures (such as B-DNA) in many aspects. It has been known for a long time that guanosine and its derivatives form viscous gels in water at millimolar or higher concentrations. As early as 1962, it was proposed, based on X-ray diffraction analysis, that the formation of such gels can be explained by stacked planar tetrads of guanines linked together via Hoogsteen type of hydrogen bonding (see figure 6) **[Gellert, 1962]**. Stacked guanine tetrads constitute the core of a guanine quadruplex structure with a central channel allowing the incorporation of metal cations. This arrangement is notable for relative rigidity, which is reflected by rather high thermal stability of quadruplex nucleic acids. The bases that are not involved in the tetrad formation (mostly A and T) are very flexible and their conformation can be easily modulated by the interaction with other molecules. Quadruplexes are classified in the literature into diverse groups depending on various structural factors that are taken into account **[Simonsson, 2001]**. The classification of quadruplexes that is given below is not exhausting and may be even ambiguous since some structural parameters can be interrelated. A more theoretical approach to the topology of guanine quadruplexes founded on the assumption that there are just two predominant ranges of the glycosidic torsion angle $\chi$ (*syn* or *anti*) for guanine residues forming G-tetrads has been presented recently **[da Silva et al., 2007]**.

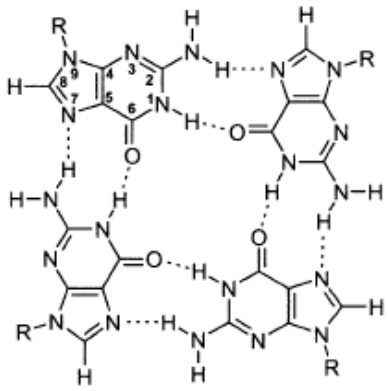


**Figure 6.** The structure of guanine tetrad.

The roughest classification of guanine quadruplexes can be done according to their stoichiometry, i.e. according to the number of interacting guanine-rich strands constituting one quadruplex structure. Although it is perhaps not well theoretically justified, all experimental findings are consistent with just the three basic stoichiometries (see figure 7). Quadruplexes are either intramolecular (unimolecular) or can be formed from 2 or 4 oligonucleotide strands. The higher aggregates (stoichiometry > 4) have been observed by gel electrophoresis, but these can be considered as a mixture of guanine quadruplexes of lower stoichiometries interacting one with another via non-covalent interactions **[Collie et al., 2010; Renčiuk, 2009]**.

The particular stoichiometry is primarily determined by the sequence, the oligonucleotide concentration and the presence and nature of stabilizing metal cations **[Hardin et al., 2001]**. As for the relation between the sequence and the stoichiometry, some restrictive rules can be deduced. First, intramolecular quadruplexes are formed only from the sequences having four or more guanine repeats (guanine repeats are the regions of two or more contiguous guanines within the sequence). Second, only the sequences with at least two guanine repeats have the propensity to form bimolecular structures. Last, all the guanine-rich sequences can, at least theoretically, form tetramolecular structures. As will be discussed later, tetramolecular guanine quadruplexes are formed extremely slowly if the oligonucleotide concentration is not sufficiently high **[Mergny et al., 2005a]**.

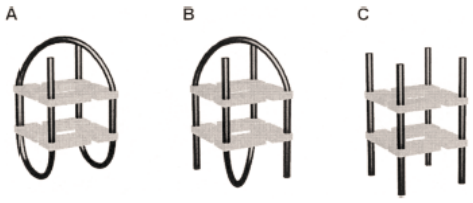



**Figure 7.** Quadruplex stoichiometry. A, B, C denotes a schematic example of unimolecular, bimolecular and tetramolecular quadruplex, respectively. Taken from ***[Simonsson, 2001]***.

Another distinguishing feature of guanine quadruplexes is the polarity of their strands. Strand backbone polarity is not only related to intermolecular structures (as it may seem), but it is associated more generally with the adjacent sugar phosphate backbone polarities. It is obvious that just four different polarities are possible (see figure 8).

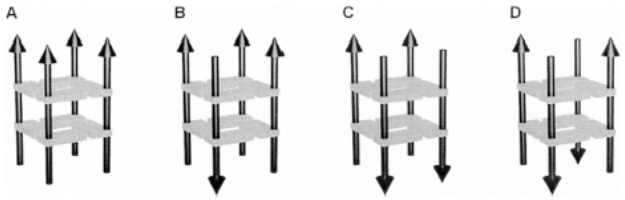



**Figure 8.** Various strand backbone polarities. The adjacent backbones can be all parallel (A), three parallel and one antiparallel (B), adjacent parallel (C) or alternating antiparallel (D). Taken from ***[Simonsson, 2001]***.

The variability of guanine quadruplexes is further increased by a diverse arrangement of the bases that are not a part of guanine tetrads. Some examples of these so-called looping sequences (or simply loops) are given in figures 9, 10 for the case of unimolecular and bimolecular quadruplexes, respectively. Neglecting the strand polarity, one can distinguish the three most frequent types of loop geometry.

First, the so called lateral or edgewise loops connect two guanine bases within one tetrad that are hydrogen bonded to each other. Examples of lateral loops are depicted in figure 9 for the case of unimolecular quadruplexes (ignore the middle loops in figure 9B and 9C) and in figure 10B-E for the case of bimolecular quadruplexes. An intramolecular quadruplex with three edgewise loops (see figure 9A) is known in the literature as a chair conformation.

Second, the loops that span a guanine tetrad diagonally (in a similar fashion as do the middle loop of an unimolecular quadruplex in figure 9B and the loops of a bimolecular quadruplex in figure 10 A) so that connect two guanine bases within the same tetrad that are not hydrogen bonded are called diagonal loops. An intramolecular quadruplex with two edgewise loops and one central diagonal loop (see figure 9B) is frequently denominated as a ‘basket’ conformation.

Third, the so-called double-chain-reversal or propeller loops are depicted in figure 9C (middle loop) or in figure 11. In this case, the loops link bases that are in different tetrads but within the same groove so that the neighbouring phosphodiester backbones run parallel **[Parkinson, 2002]**.

A type of loop that is actually formed is determined mainly by its sequence and length. Although one can arrive at 26 possible looping topologies for an intramolecular quadruplex with the three aforementioned loops, only six of them were experimentally confirmed as yet **[da Silva, 2007]**. At present, no general rules can be given relating the primary sequence and the structure of quadruplex.

However, some empirical rules can be postulated. While lateral or propeller loops can have one or more residues, at least three residues are required to form diagonal loops **[da Silva, 2007]**. Propeller loops are frequently found in parallel folds with single residue loops **[Parkinson, 2002]**. In fact, three single residue loops are consistent only with parallel-stranded intramolecular quadruplexes **[Hazel et al., 2004].** Too long loops may have a destabilizing effect and, in the presence of a G-quadruplex complementary strand, duplex formation may compete with potential G-quadruplex folding **[Kumar et al., 2008]**. The modifications in loop sequences influence the stability of quadruplex **[Smirnov & Shafer, 2000a]** and its interaction with other molecules **[Takahama et al., 2011]**.

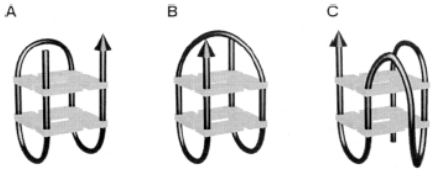



**Figure 9.** Loop variations for unimolecular quadruplexes ***[Simonsson, 2001]***.

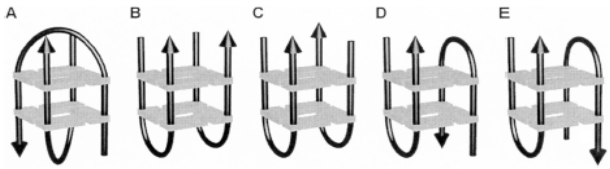



**Figure 10.** Loop variations for bimolecular quadruplexes ***[Simonsson, 2001]***.

The structure of guanine quadruplexes, similarly like canonical nucleic acids, can be characterized by a set of torsion angles (see figure 1 and table 1). The glycosidic torsion angle is particularly useful parameter for experimental discrimination between various quadruplex topologies. As contrasted with helical A or B-DNA structures, where the torsion angle $\chi$ is found predominantly in *anti* conformation, the guanines participating in the G-tetrad formation can naturally exist in both *syn* and *anti* conformations **[Gallagher, 2004; Hardin et al., 2001; Simonsson, 2001]**. If the values of glycosidic torsion angles are restricted to either *syn* or *anti* configuration, then there are 16 different combinations of how to build guanine tetrad (see figure 12).

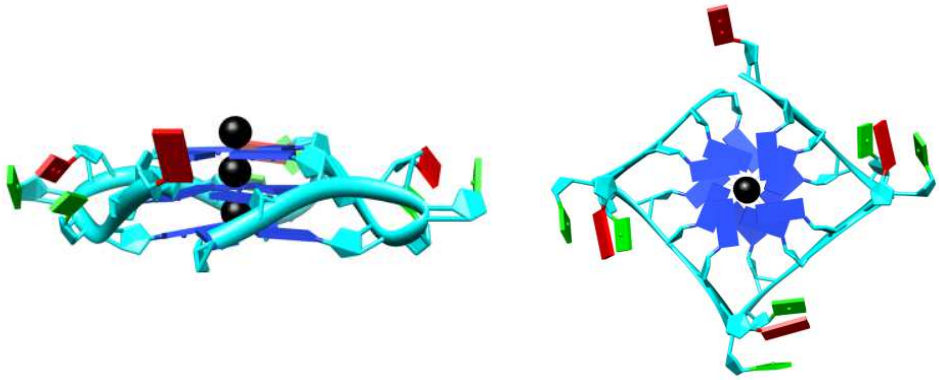


**Figure 11.** The X-ray structure of human telomeric DNA having the sequence 5'-$AG_3(TTA)_3G_3$. Sugar phosphate backbone is colored in cyan, adenine residues in red, thymine residues in green and guanine residues in blue. Stabilizing potassium ions are colored in black. On the left: side view. On the right: Top view. Three TTA loops protrude out of the quadruplex core reminding the blades of propeller. PDB code of structure is 1KF1.

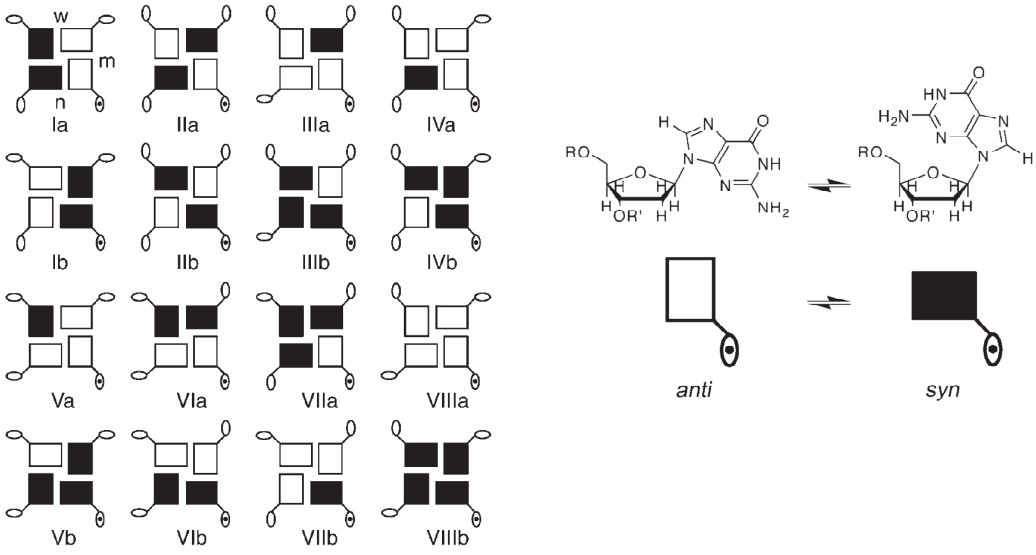



**Figure 12.** On the left: A schematic depiction of all possible combinations of glycosidic torsion angles for 4 guanine residues forming guanine tetrad. The definition of medium (m), wide (w) and narrow grooves (n) based on glycosidic angle conformations of adjacent guanine residues is shown in Ia. On the right: Chemical structures of *anti* and *syn* conformers. Adapted from ***[da Silva, 2007]***.

The glycosidic torsion angles of adjacent guanine residues can be used to define the type of a groove **[da Silva, 2007]**. As opposed to helical A and B-DNA, which show one minor and one major groove, a G-quadruplex can have generally up to three distinct grooves that differ in size (that is narrow, medium and wide groove) depending on the distribution of glycosidic torsion angles within guanine tetrads (see figure 12). Due to steric reasons, stable quadruplexes may be formed only if stacked guanine tetrads have the same groove width dimensions. As a result, there are just eight different groove width combinations (hence, one groove width combination corresponds to just two different configurations of a G-tetrad). Generally, guanine quadruplexes with mutually parallel strands have the same distribution of glycosidic torsion angles within guanine tetrads (see the schemes VIIIa and VIIIb in figure 12). It means that they have four grooves of medium size **[da Silva, 2007, Simonsson, 2001]**. On the basis of molecular mechanics calculations it was proposed that parallel four-stranded structures with guanine residues solely in *anti* conformation (see the scheme VIIIa in figure 12) are favored over structures with guanine residues that alternate in *syn* and *anti* conformation (see the schemes IIa and IIb in figure 12) **[Hardin et al., 2001; Simonsson, 2001]**. On the other hand, the alternation in glycosidic torsion angles of guanines is characteristic for intramolecular antiparallel folds having two wide and two narrow grooves **[da Silva, 2007, Simonsson, 2001]**. The purely *syn* structures (see the scheme VIIIb in figure 12) are possible, but they are rare and the least stable **[Hardin et al., 2001; Mohanty & Bansal, 1993]**. The grooves can serve as potential binding sites for drugs **[Georgiades et al., 2010]**. The interaction of guanine quadruplexes with drugs will be discussed later (see the section 1.5.4).

The classification of guanine quadruplexes given above illustrates only the basic concepts regarding their structure: the stoichiometry and the geometry of loops. The three-dimensional structures of guanine quadruplexes revealed mostly by NMR are more diverse than may appear at first glance. In addition to commonly known G-quadruplex motifs such as all-parallel tetramolecular quadruplex (see figure 8A), bimolecular hairpin-like structures (see figure 10) or intramolecular chair (see figure 9A) and basket quadruplex (see figure 9B), there many other less frequent, interesting and unusual G-quadruplex conformations. Most of these structures arise as a result of dimerization of structurally simpler quadruplex subunits. The dimerization is driven mainly by the interaction between the loops of the two

quadruplex partners. The bases within the loop of one quadruplex subunit may interact with one another as well as with the bases belonging to the loops of another quadruplex subunit by means of the formation of hydrogen bonds of either the regular Watson-Crick or mismatch type. Hydrogen bonding and stacking interactions between the two quadruplex subunits stabilize the overall quadruplex structure.

An example of unusual interlocked dimeric G-quadruplex is provided by Phan and co-workers for the sequence $G_4TG_3AG_2AG_3T$, a potent inhibitor of HIV-1 integrase **[Phan et al., 2005]**. This quadruplex dimerizes via the stacking of the two guanine tetrads, each of which is composed of three guanines that are part of one strand and one guanine that belongs to other strand (see figure 13). The interlocked quadruplex has all guanines but one in anti conformation, the adjacent sugar phosphate backbones all run in parallel and all loops are of the double-chain-reversal (propeller) type. This structure is also interesting in that each of its two monomer subunits contains one A•G•G•G•G pentad consisting of four guanines and one adenine.

In addition to the pentad observed by Phan and coworkers for the interlocked dimeric quadruplex **[Phan et al., 2005]**, unusual hydrogen-bonded planar network of nucleobases including triads **[Kuryavyi et al., 2000]**, tetrads (containing at least one non-guanine residue) **[Lim et al., 2009a]**, pentads **[Zhang et al., 2001]**, hexads **[da Silva, 2005]**, heptads **[Palumbo et al., 2008]** or octads **[Borbone et al., 2011]** was also observed for other G-quadruplexes. Further details on the general aspects regarding the structure of guanine quadruplexes can be found in an overview book **[Neidle & Balasubramanian, 2006]**.

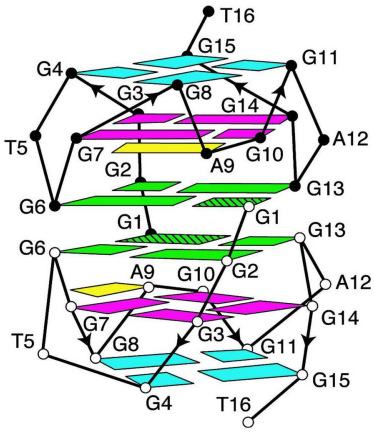



**Figure 13.** The structure of an interlocked dimeric quadruplex (the sequence: $G_4TG_3AG_2AG_3T$. Adapted from *[Phan et al., 2005]*.

## 1.3 Structural polymorphism of guanine quadruplexes

The conformation and stability of guanine quadruplexes depends on a range of factors, among which the ionic composition of solution is the most important **[Neidle & Balasubramanian, 2006]**. Metal cations can bind to guanine quadruplexes in two completely distinct ways. They can interact either with backbone phosphates or they coordinate to oxygen atoms within guanine tetrads. The interaction of cations with sugar phosphate backbone is non-specific and, similarly to the case of B-DNA, merely screens the negative charge on phosphates. Unlike other nucleic acid structures, dehydrated metal cations directly coordinate to carbonyl oxygens ($O_6$) of guanines participating in the tetrad formation. It explains why the conformations and thermal stabilities of G-quadruplexes differ significantly even for rather similar cations such as sodium or potassium **[Neidle & Balasubramanian, 2006]**. The propensity of sodium and potassium to form and stabilize guanine quadruplexes is superior to most other metal cations. As such and due to their biological relevance, these are a subject to the vast majority of studies. The precise mode of interaction of potassium and sodium with G-tetrads differs due to their slightly different ionic radii. The smaller ionic radius of sodium allows its facile coordination within the G-tetrad plane or anywhere in the central cavity of a G-quadruplex. By contrast, potassium cations are too large to fit within the G-tetrad plane and must be necessarily accommodated between the stacked G-tetrads (see figure 14). Although sodium ions better fit into the quadruplex cavity than potassium ions, their, in most cases, lower efficiency for stabilizing G-quadruplexes can be partly a consequence of the higher energetic cost associated with the dehydration process **[Hardin et al., 2001]**.

Apart from $Na^+$ and $K^+$ ions, the stabilization of G-quadruplexes by other monovalent ions has been extensively studied. The stabilizing potency of monovalent ions can be written as follows: $K^+ > Rb^+ > Na^+ > Cs^+ > Li^+$ **[Hardin et al., 2001]**. This order is a generalization of many experimental data rather than a universal law. According to another source, there is a little different order: $K^+ >> Na^+ > Rb^+ > Cs^+ >> Li^+$ **[Neidle & Balasubramanian, 2006]**. It is commonly speculated that $Li^+$ ions disfavor the formation of guanine quadruplexes or even disrupt their structure. This is highly controversial and more theoretical analysis is needed to support this hypothesis. It seems, however, that rather than destabilizing, $Li^+$ has no net effect on the quadruplex structure **[Neidle & Balasubramanian, 2006]**. Among non-metal

monovalent cations, ammonium ions are known to be moderately potent stabilizers of the quadruplex structure **[Gros et al., 2007; Kankia & Marky, 2001]**. For example, the melting transition of d($G_2T_2G_2TGTG_2T_2G_2$), which is commonly known as TBA (thrombin binding aptamer), stabilized by $NH^{4+}$ ions is detected at about 29°C **[Kankia & Marky, 2001]**. Such stability is significantly lower than that observed for the potassium form of the quadruplex (~50°C at 100-200 mM $K^+$), yet comparable or mildly higher than that corresponding to TBA quadruplex induced by sodium ions (~20°C) **[Fialová et al., 2006]**. $NH^{4+}$ ions are rather bulky and therefore they occupy, similar to potassium ions, the position between two guanine tetrads **[Trajkovski et al., 2009]**. In mass spectroscopic studies of G-quadruplexes, $NH^{4+}$ ions substitute for other commonly stabilizing metal cations **[Collie et al., 2010; Gabelica et al., 2008; Gros et al., 2007; Rosu et al., 2010]**.

Divalent cations also have the propensity to promote the formation of G-quadruplexes. However, the mode of interaction of monovalent and divalent cations with nucleic acids is markedly different. Generally, the effect of divalent cations is noticeable at concentrations that are one to three orders of magnitude lower than those of monovalent ions **[Blume et al., 1997; Floris et al., 1999; Lee, 1990; Ottová et al., 2011]**. Although relatively low concentrations (~0.1-10 mM/L) of divalent cations strongly stabilize G-quadruplexes, an increase in their concentration above a certain threshold usually leads to a destabilization of the quadruplex structure. The disrupted quadruplex may then be replaced by another competing nucleic acid form such as a single-stranded helical structure or a triplex **[Blume et al., 1997; Floris et al., 1999; Lee, 1990]**. The stabilizing efficiency of alkaline earth metal cations can be written in the following order: $Sr^{2+} >> Ba^{2+} > Ca^{2+} > Mg^{2+}$ **[Hardin et al., 2001; Venczel & Sen, 1993]**. This order may slightly vary from case to case in a similar fashion as that for monovalent ions. In any case, $Sr^{2+}$ ions are known to be very potent stabilizers of the quadruplex structure with the efficiency exceeding that of the other divalent metal ions. In fact, the affinity of $Sr^{2+}$ ions to G-quadruplexes can be directly compared with that of $K^+$ ions. More interestingly, some G-quadruplexes like TBA quadruplex **[Kankia & Marky, 2001]** or the human telomeric quadruplex sequence $d(TTAGGG)_4$ **[Wlodarczyk et al., 2005]** are even markedly more stable in $Sr^{2+}$ than in $K^+$. A similarly strong or even greater propensity than $Sr^{2+}$ ions to facilitate the formation of G-quadruplexes was also found for $Pb^{2+}$ ions **[Smirnov & Shafer, 2000b]**. Interestingly, the affinity of $Pb^{2+}$

was found to be unusually high for TBA quadruplex and its cognate sequences as well as for the human telomeric sequence $d(AG_3(TTAG_3)_3)$, yet a similar effect was not observed for hairpin dimer quadruplexes, which may indicate the selectivity of $Pb^{2+}$ towards certain DNA sequences.

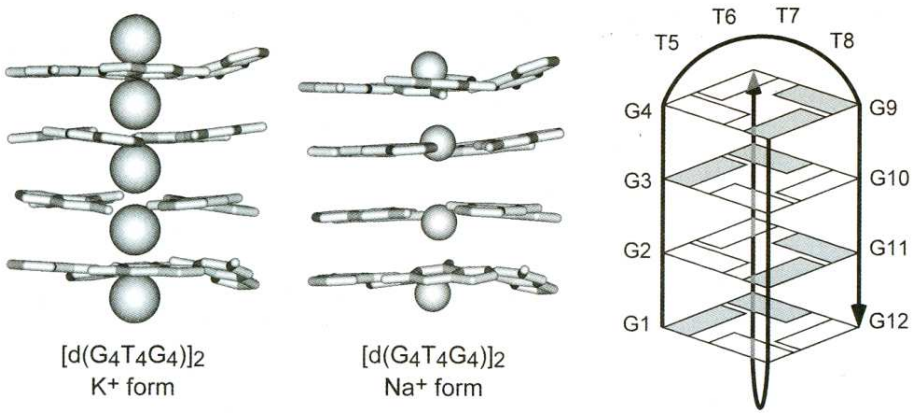



**Figure 14.** The comparison of X-ray structures of $d(G_4T_4G_4)$ stabilized either by potassium (PDB ID: 1JRN) or sodium (PDB ID: 1JB7). For clarity, only the guanine bases (stick representation) and cations (ionic radii) are shown. The schematic structure representing the orientations of DNA strands is depicted to the right of the crystal structures. White and gray rectangles are guanine residues with *anti* and *syn* glycosidic torsion angles, respectively. The structures illustrate the general fact that, potassium is never located within the G-tetrad plane, whereas sodium, due to its relatively smaller size, can be theoretically found anywhere within the central quadruplex channel. Adapted from ***[Neidle & Balasubramanian, 2006]***.

G-Quadruplex-forming sequences stabilized by different ions may or may not adopt distinct conformations, regardless of their thermal stability. For instance, the TBA quadruplex $d(G_2T_2G_2TGTG_2T_2G_2)$ melts at roughly 20°C in sodium and at 50°C in potassium (at ionic strengths close to physiological conditions) **[Fialová et al., 2006]**. Despite these differences in thermal stabilities, the overall fold of TBA is practically the same in both salts (neglecting slight changes in looping geometry for both quadruplex forms). By contrast, the Oxytricha telomeric repeat sequence $d(G_4(T_4G_4)_3)$ switches between two quadruplex conformations (antiparallel foldback and parallel tetramer structure) depending on the $[Na^+]/[K^+]$ ratio as shown by Raman spectroscopy **[Miura et al., 1995]**. Potassium-stabilized quadruplexes are

usually much more thermostable than those stabilized by sodium, but it is not always the case. For example, the stability of a short DNA human telomeric fragment $G_3(TTAG_3)_3$ is comparable in sodium and potassium **[Tomaško et al., 2008]**. Unlike TBA, the conformation of this quadruplex strongly depends on the cation type, especially at high DNA loads **[Renčiuk, 2009]**. Interestingly, some guanine quadruplexes are even more stable in $Na^+$ than in $K^+$. As an example may serve RNA/DNA hybrid quadruplexes derived from the Oxytricha telomere sequence $G_4T_4G_4$ **[Vondrušková et al., 2008]** or loop modifications of TBA sequence (our results; see the section 3.4). It is not unusual that G-rich oligonucleotides adopt more than one conformation at a time. A very good example of such polymorphism is the DNA fragment of the sequence from the human c-myc promoter: $AG_3TG_4AG_3TG_4$ **[Ambrus et al., 2005]**. This oligonucleotide adopts four distinct conformations differing in the configuration of loops (some guanines are either part of G-tetrads or are located within loops). It is proposed that dynamic equilibrium among these four loop conformers stabilizes the quadruplex structure.

The majority of naturally occurring G-rich sequences are paired with their complementary C-rich strands in the form of Watson-Crick double helix. The duplex to quadruplex competition is driven by a range of factors, with the temperature, strand concentration and pH being amongst the most important **[Halder & Chowdhury, 2005; Jaumot et al., 2006; Kumar & Maiti, 2004; Li et al., 2003; Phan & Mergny, 2002; Zhao et al., 2004]**. Generally, quadruplex formation, and a concomitant duplex disruption, is facilitated by acidic pH, higher salt concentrations and higher temperatures. The dissociation of duplex DNA is coupled to the formation of other interesting non-canonical structures known as I-motifs (somewhat incorrectly referred to as I-tetraplexes). I-motifs can be imagined as two intercalated right-handed parallel duplexes running in the opposite direction and consisting of hemiprotonated $C{\bullet}C^+$ pairs **[Benevides et al., 1996; Kang et al., 1995]** (see figure 15). The glycosidic torsion angles observed in the I-motif correspond predominately to the *anti* conformation; puckering of sugar rings is, unlike B-DNA, mostly $C_{3'}$-endo (with some residues adopting $C_{2'}$-endo pucker) **[Kang et al., 1995]**. The stability of the I-motif is enhanced with decreasing pH since the formation of this structure correlates with the presence of protonated cytosine bases **[Benevides et al., 1996]**. Low pH also adds to the G-quadruplex stability because its pK constant is within the physiological pH range (ca. 6.8), in contrast to WC pairs **[Hardin et al.,**

1992]. Phan and coworkers showed, on the basis of NMR experiments, that the formation time of duplex may be comparable to the dissociation time of I-motif and G-quadruplex, which indicates that the melting of folded structures may be the rate-limiting step for duplex formation [Phan & Mergny, 2002]. They also proposed that the duplex-quadruplex transition is only little influenced by the nature of G-quadruplex-stabilizing ions ($Na^+$, $K^+$). Quadruplex is favored over duplex at low strand concentrations [Halder & Chowdhury, 2005; Zhao et al., 2004] and destabilized by $Mg^{2+}$ ions [Li et al., 2003]. A comparative study of various G-rich DNA sequences reveals that the human telomere sequence $G_3(TTAG_3)_3$ has amongst other sequences, like the tetrahymena telomere sequence $G_4(T_2G_4)_3$ or the sequences derived from the c-myc promoter, the least propensity to form a G-quadruplex in the presence of a competing duplex [Risitano & Fox, 2003]. Quadruplex structures at unfavorable neutral or alkaline pH may be stabilized by the interactions with other molecules or by superhelical stress on duplex DNA [Gavathiotis et al., 2003; Phan & Mergny, 2002; Rangan et al., 2001].

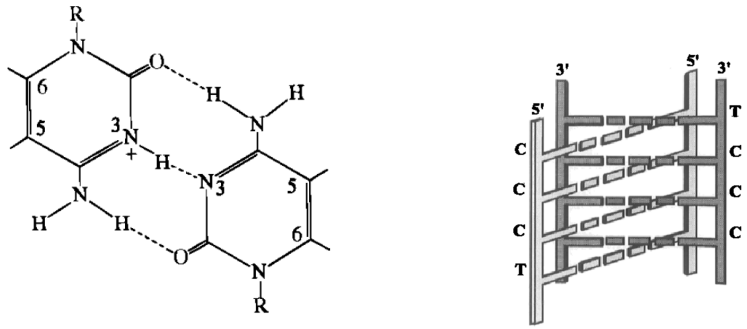



**Figure 15.** On the left: Hemiprotonated cytosine base pair. On the right: a schematic drawing illustrating the structure of I-motif for d(CCCT) oligonucleotide. Adapted from *[Benevides, 1996]*.

Structural polymorphism of G-quadruplexes and I-motifs can be a basis for DNA nanomachines [Alberti et al., 2006]. The functioning of such nanomachines is controlled through conformational changes of their structure induced by either external factors like pH, temperature or ionic composition of solution, or by consecutive introducing DNA “fuel” single strands. In the latter case, the first working step of a G-quadruplex-based nanomachine consists in mixing a G-

quadruplex with a cytosine-rich strand that has the region fully complementary to the G-quadruplex-forming sequence and a few bases long overhang. As a result, the original G-quadruplex is replaced with a duplex structure. In the second working step, the original G-quadruplex is restored at the expense of a newly formed "waste" GC duplex after adding another G-rich strand (different from the original G-quadruplex forming sequence) that is partly complementary to the previously added cytosine-rich strand. In this manner, one obtains a cyclic working "two stroke nanomotor" **[Alberti & Mergny, 2003; Wang et al., 2005]**.

Some other interesting nanotechnological applications based on self-assembly of G-rich DNA have been described **[Davis, 2004; Lena et al., 2009]**. They include, for example, an electronic nanoswitch that can be employed for the detection of potassium ions **[Wu et al., 2008]**, or a covalently modified intramolecular quadruplex folded within the hydrophobic phospholipid membrane that functions as an efficient $Na^+$ transporter **[Kaucher et al., 2006]**.

G-rich DNA oligonucleotides may spontaneously self-associate to form highly-ordered structures known as G-wires **[Marsh et al., 1995]** or frayed wires **[Protozanova & Macgregor, 1996]**. These structures, having the sequence $G_xN_y$ or $N_yG_x$ where $x \geq 10$, $y \geq 6$ and $N$ is either T or A, are readily stabilized by $Na^+$, $K^+$ or $Mg^{2+}$ ions, and once formed, they are extremely stable **[Abu-Ghazalah et al., 2010]**. Specifically, frayed wires are composed from several DNA strands joined through the formation of G-tetrads, and single-stranded regions with adenines that are free to make Watson-Crick bonds with T-rich oligonucleotides **[Protozanova & Macgregor, 1996]**. A direct observation of G-wires is possible by means of SPM (scanning probe microscopy) since these supramolecular structures are very long, ranging from nanometers to tenths of microns **[Marsh et al., 1995; Vesenka et al., 2007]**. As evidenced by Raman spectroscopy, guanines in these structures adopt mostly *anti* sugar puckers, which is consistent with the suggested parallel alignment of G-wire strands **[Abu-Ghazalah et al., 2010; Poon & Macgregor, 1999]**. The possibility of controlling the formation and orientation of G-wires (through the sequence, salt concentration, temperature, type of stabilizing ions, etc.) make these structures promising candidates for various scaffolding nanostructures or nanoscale molecular systems **[Spindler et al., 2010]**.

## 1.4 The stability, and kinetics of formation and dissociation of guanine quadruplexes

The laws of thermodynamics determine unequivocally which nucleic structures are energetically preferred at particular experimental conditions. However, the observed structures may not correspond to the thermodynamically most stable species because of long times needed for their formation. As compared to DNA or RNA hairpins, duplexes, triplexes or other "classical" nucleic acids structures, G-quadruplexes have relatively slow kinetics of folding and unfolding. In short, while the kinetics of intramolecular G-quadruplexes may be still rather fast, bimolecular and especially tetramolecular ones are noted for extremely slow kinetics **[Hardin et al., 2001; Lane et al., 2008; Mergny et al., 2005a]**. The folding and unfolding rate of intermolecular structures depends on oligonucleotide strand concentration and temperature. The following discussion will clarify the point.

Despite the relative paucity of data on kinetics of intramolecular guanine quadruplexes, it is often claimed that it is fast. The term "fast" in this context means that the structure adapts to varying experimental conditions (typically temperature, pH, DNA concentration or ionic composition) in much shorter time than the measurement itself. The timescale of rapid kinetics is on the order of seconds or minutes (at least with respect to the results provided by UV, CD spectroscopic and calorimetric techniques). Quick kinetics was observed, for instance, for human telomere related folds (~0.01s) **[Mergny et al., 1998]** or for TBA quadruplex and its various loop modifications **[Smirnov & Shafer, 2000a]**. However, some intramolecular guanine quadruplexes, such as that formed from the Oxytricha telomeric repeat $(G_4T_4)_4$, exhibit relatively slow kinetics of formation/dissociation. Half-life of the dissociation of the $(G_4T_4)_4$ quadruplex at 37°C is about 10 years in the presence of 50 mM $K^+$ **[Brown et al., 2005]**.

In addition to fast kinetics, intramolecular guanine quadruplexes are also characterized by the independence of their melting temperature on the nucleoside concentration. This finding generally holds for all types of intramolecular transitions, not only for G-quadruplexes **[Micura et al., 2001]**. The reversibility of heating and cooling thermal curves together with the independency of melting temperature on the DNA concentration is frequently used as a simple indicator of G-quadruplex intramolecularity within a given strand concentration range **[Jing et al., 1997;**

**Kankia & Marky, 2001; Kumar & Maiti, 2004; Risitano & Fox, 2003]**. If it is not the case, then a G-rich sequence may not adopt an intramolecular quadruplex or, in a more complicated case, it exists in multiple conformations (intramolecular as well as intermolecular).

Depending on the oligonucleotide sequence and concentration, the cations present and other milieu factors, intramolecular structures may be dominant conformations in solution **[Dailey et al., 2010; Miura et al., 1995]**. Multiple intramolecular conformations may have similar biophysical properties such as thermodynamic stability or electrophoretic mobility, which makes their discrimination rather difficult **[Dailey et al., 2010]**. The reversible melting curves (that is, those that correspond to thermodynamic equilibrium) obtained by means of various spectroscopic methods (UV, CD, NMR, Raman spectroscopy) or derived from calorimetric profiles can be used for the calculation of thermodynamic parameters. UV spectroscopy is the simplest and most widely used for this purpose.

The timescale of structural transitions as well as the stability of intermolecular quadruplexes is markedly different from that typical of intramolecular quadruplexes. Bimolecular and especially tetramolecular quadruplexes at low oligonucleotide concentrations or/and in the presence of low concentrations of stabilizing ions exhibit extremely slow kinetics of formation and dissociation ranging from several minutes to years **[Hardin et al., 2001; Mergny et al., 2005a; Wyatt et al., 1996]**. Hysteresis in melting curves obtained by means of routine spectroscopic techniques (UV and CD spectroscopy, fluorescence) or hysteresis in calorimetric profiles is an inevitable consequence of slow kinetics or slow instrumental response. Slowing down the rate of the temperature change may help to eliminate or significantly suppress hysteretic behavior, but this is not always workable if the kinetics is extremely slow. Slow kinetics rather complicates or even makes impossible determining the thermodynamic parameters by means of the classical van't Hoff approach.

The reaction order for the formation of tetramolecular guanine quadruplexes with respect to the oligonucleotide concentration is high (somewhere between 3 and 4) **[Bardin & Leroy, 2008; Mergny et al., 2005a; Wyatt et al., 1996]**. It explains why the association of G-rich strands into tetramers is such a slow process. For example, Wyatt and coworkers studied the kinetics of association and dissociation of parallel-stranded DNA tetramolecular quadruplexes formed from phosphorothioate and phosphodiester oligodeoxynucleotides with the sequence $T_2G_4T_2$ **[Wyatt et al.,**

**1996]**. They found that the strand association of the phosphorothiate tetramer was a fourth order reaction with the association rate constant increasing with decreasing temperature. The consequence of the high reaction order is that, at room temperature, only negligible strand association was detected at the oligonucleotide concentration of 100 μM, whereas the association reaction at 1 mM concentration was too rapid to be followed accurately. By contrast, quadruplex dissociation was a first order reaction with the dissociation rate constant increasing with increasing temperature. Wyatt et al. proposed that single strands are in equilibrium with dimers, which in turn give rise to tetramers (both steps are rate limiting consistently with the suggested reaction order 4). It seems plausible since it is highly unlikely that four strands associate at once (actually the reactions having the molecularity > 2 are very rare in chemistry). The suggested equilibrium must be necessarily reached very fast as they did not detect any dimer species. These conclusions about the formation pathway of tetramolecular quadruplexes are also consistent with those of Mergny et al. **[Mergny et al., 2005a]**.

Rather than by the association of transient dimeric species, tetramolecular quadruplexes may be formed by the gradual association of single strands, meaning that transient triplex species also come into question. This model is in line with the reaction order around 3 in KCl as determined by NMR spectroscopy **[Bardin & Leroy, 2008]**. However, no trimers were proven to exist. Interestingly, molecular dynamic simulations indicate that the possible dimer intermediates on the nanosecond time scale are not parallel-stranded duplexes connected one to another through Hoogsteen hydrogen pairing, but rather “cross-like” dimers formed by two approximately perpendicular strands **[Štefl et al., 2003]**.

The results of the extensive study of kinetics of tetramolecular quadruplexes provided by Mergny and coworkers can be summarized into several main points **[Mergny et al., 2005a]**. First, the strand association is greatly accelerated with the increasing ionic strength, yet the ionic strength itself has negligible effect on the quadruplex denaturation. Similar behavior is observed also for other types of nucleic acids structures, such as duplexes and triplexes. Second, the usually higher stability of quadruplexes in $K^+$ as compared to $Na^+$ is from the kinetical point of view the result of both, an increased folding rate and a decreased unfolding rate. It is consistent with the previously mentioned study of Wyatt and coworkers **[Wyatt et al., 1996]**. The same behavior was observed also for unimolecular quadruplexes

**[Halder & Chowdhury, 2005; Kumar & Maiti, 2004; Zhao et al, 2004]**. Third, $Mg^{2+}$ added either to $Na^+$ or $K^+$ solution accelerates quadruplex assembly but slows down dissociation with the resulting equilibrium constant being little affected. Fourth, quite surprisingly, RNA tetramolecular quadruplexes are generally more stable than their DNA counterparts. The increased stability is correlated with a faster association rate and a slower dissociation rate. Last but not least, the longer are guanine tracts, the faster is quadruplex assembly. Many tetramolecular quadruplexes with long guanine tracts do not denature even at temperatures around 90-100°C, if stabilized by $K^+$ or $Na^+$. Hence, it is not possible to determine correctly their dissociation rate constants.

From the foregoing discussion, it is clear that the interpretation of the experimental data and the estimation of thermodynamic parameters are hampered by the rather slow kinetics of intermolecular quadruplexes. Particular conformations may not correspond to their equilibrium states, but rather to kinetically trapped intermediates. In other words, the observed structures depend on the procedure of sample preparation, and not only on the milieu conditions. One often encounters the need to disrupt kinetically trapped or undesired G-quadruplex structures (that may persist at temperatures as high as 100°C for a long period of time) in order to stabilize the desired conformation or to study the kinetics of quadruplex assembly. The simplest way how to accomplish this is to add hydroxide to the sample and then to neutralize it with the appropriate amount of hydrochloric acid **[Lyonnais et al., 2002]**. Unwanted structures that are not thermostable enough can be removed easily by heating followed by cooling.

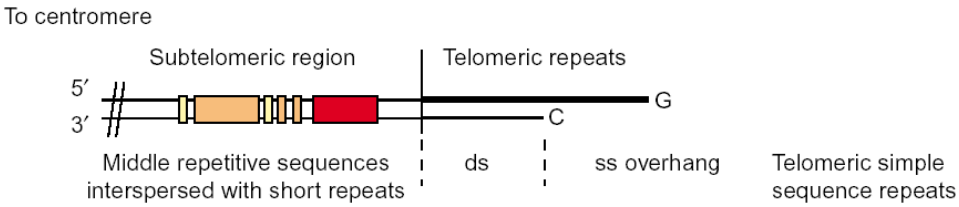
The cooling rate may determine which structure(s) are actually formed. It is reported, for instance, that the oligonucleotide $d(G_4T_2G_4)$ at higher strand concentrations may adopt a tetramolecular quadruplex instead of a thermodynamically more stable hairpin structure **[Balagurumoorthy et al., 1992]**. The hairpin structure is stabilized by heating the oligonucleotide sample, prepared at low DNA concentration in a buffer of low ionic strength, above its melting temperature and then cooling slowly to room temperature. It was also reported that repeated freezing/thawing the sample at low ionic strength can be used to stabilize a specific quadruplex conformation **[Peňázová & Vorlíčková, 1997]**.

## 1.5 Biological relevance of guanine quadruplexes

There are a great many G-rich sequences within the human genome with the potential to form guanine quadruplexes **[Huppert & Balasubramanian, 2005; Todd et al., 2005; Weitzmann et al., 1996]**. However, to date there is little evidence that these structures actually exist in vivo **[Duquette et al., 2004; Granotier et al., 2005; Lipps & Rhodes, 2009; Schaffitzel et al., 2001]**. Guanine quadruplexes may play both useful and harmful functions in vivo **[Arthanari & Bolton, 2001]**. The brief summary of putative biological functions of guanine quadruplexes in vivo is presented in the following paragraphs. The research at present is mainly focused on the elucidation of their role in telomere biology (see 1.5.1.). There is emerging evidence that quadruplexes might influence numerous biological processes including replication, transcription or recombination (see 1.5.2.). New G-rich oligonucleotides are continually devised and characterized with respect to their use as aptamers in both therapeutic and non-therapeutic applications (1.5.3).

### *1.5.1 Telomeric quadruplexes*

There are several remarkable differences among eukaryotic and prokaryotic chromosomes **[Bendich & Drlica, 2000]**. Chromosomes of eukaryotes are almost exclusively linear in contrast to circular chromosomes of prokaryotes. The ends of linear eukaryotic chromosomes are known as telomeres. The primary function of telomeres is to cap chromosomal ends and thus maintain the stability and integrity of the genome. They must be necessarily distinguished from double-strand breaks, which are considered by cellular machinery as damaged DNA and repaired mostly by **h**omologous **r**ecombination (HR) or **n**on-**h**omologous **e**nd-**j**oining (NHEJ) **[Jackson, 2001]**. Chromosomes are protected by telomeres from undesired recombinations, fusions and nuclease digestion **[Blackburn, 2001]**. Telomeres represent a solution to the end-replication problem **[Chakhparonian & Wellinger, 2003]**. They also play a role in the localization of chromosomes within the nucleus and in transcriptional repression of genes close to the chromosomal ends (TPE-**t**elomere **p**osition **e**ffect) **[Gilson et al., 1993]**. The telomeric DNA can be divided into the three distinct regions (see figure 16) **[Wellinger & Sen, 1997]**: (i) TAS (**t**elomere **a**ssociated **s**equences) DNA; (ii) double-stranded telomeric DNA and (iii) the single-stranded guanine-rich overhang.




| | | | | |
|---|---|---|---|---|
| *S. cerevisiae* | X, Y elements | 200–300 bp | 12–14 nt; S phase >25 nt | $(TG)_{1-6}\,TG_{2-3}$ |
| Ciliates | Various | 14–16 bp | 14–16 nt | |
| *T. termophila* | | | | $T_2G_4$ |
| *O. nova, E. crassus* | | | | $T_4G_4$ |
| Mammals | Various | | >100 nt | $T_2AG_3$ |
| *H. sapiens* | | 5–15 kb | | |
| *M. musculus* | | up to100 kb | | |
| Plants | Various | Variable, >50 kb in some cases | Either >20–30 nt or <12 nt | $T_3AG_3$ |
| *A. thaliana* | | | | |
| *Z. mais* | | | | |
| *N. tabacum* | | | | |

**Figure 16.** Organisation of telomeric DNA in different eukaryotes. Telomeric simple sequence repeats of the strand running in the direction 5'→3' towards the end of the chromosome are shown (the G-rich strand). Moderately repetitive elements are common to subtelomeric regions but they are nonessential and have not been assigned any clear function. The C-rich strand is denoted as C and the G-rich strand as G; ds = double strand, ss = single strand. Taken from reference ***[Chakhparonian & Wellinger, 2003]***.

There are two different TAS located in the subtelomeric region **[Wellinger & Sen, 1997]**. Middle-repetitive sequences of a few hundred to several thousand base pairs are known for plant, fungal, invertebrate and vertebrate species (including humans). Shorter tandem repeats of ~20 to 1000 base pairs are interspersed within the middle repetitive sequences. TAS DNA provides an alternative mechanism for maintenance of telomeres via recombinational or transpositional processes in the cases where telomerase is absent.

The double-stranded region of chromosome at its very end is composed of short tandem repeats, which are closely related among various species **[Wellinger & Sen, 1997]**. The telomeric repeats at the 3' chromosomal terminus are rich in guanine and almost always absent of cytosine (see table 4 for an overview of telomeric

repeats in various species). The human telomeric sequence TTAGGG is widely conserved among vertebrates **[Meyne et al., 1989]**. It is reported that the human telomeres contains at least three types of G-rich repeats **[Allshire et al., 1989]**. Some fungi have more complex repeats that are quite long (20-25 bp) and contain cytosine (see table 4). The number of telomeric repeats per telomere ranges from a few repeats in ciliates (20-30 bp in total) to several thousands in humans and mice (up to 100 kb) **[Chakhparonian & Wellinger, 2003; Wellinger & Sen, 1997]** (see figure 16).

| | |
|---|---|
| **Protozoa** | |
| Tetrahymena, Glaucoma, Oxytricha, Euplotes, Stylonychia | $T_2G_4$ |
| Paramecium | $T_2[T/G]G_3$ |
| Trypanosoma | $T_2AG_3$ |
| Giardia | $TAG_3$ |
| Plasmodium | $T_2[T/C]AG_3$ |
| Theileria | $T_{3-4}AG_3$ |
| **Slime moulds** | |
| Dictyostelium | $AG_{1-8}$ |
| Physarum, Didymium | $T_2AG_3$ |
| **Fungi** | |
| Neurospora, Podospora, Cladosporium, Histoplasma, Pneomocystis | $T_2AG_3$ |
| Cryptococcus | $T_2AG_{3-5}$ |
| Saccharomyces | $(TG)_{1-6}TG_{2-3}$ |
| Kluyveromyces | $ACG_2AT_3GAT_2AG_2TATGTG_2TGT$ |
| Candida albicans | $ACG_2ATGTCTA_2CT_2CT_2G_2TGT$ |
| Candida tropicalis 4443 | $A_2G_2ATGTCACGATCAT_2G_2TGT$ |
| Candida guillermondii | $ACTG_2TGT$ |
| Schizosaccharomyces | $T_{1-2}ACA_{0-1}C_{0-1}G_{1-6}$ |
| **Plants** | |
| Chlamydomonas | $T_4AG_3$ |
| Chlorella, Arabidopsis, Zea mays, Hordeum (Barley), Nicotiana | $T_3AG_3$ |
| Lycopersicum (Tomato) | $T_2(T/A)AG_3$ |
| **Invertebrates** | |
| Ascaris, Caenorhabditis | $T_2AG_2C$ |
| Parascaris | $T_2GCA$ |
| Bombyx (and other Lepidoptera) | $T_2AG_2$ |
| **Vertebrates** | |
| Home sapiens, many other verebrates | $T_2AG_3$ |

**Table 4.** Telomeric repeat sequences in various species. Adapted from reference ***[Wellinger & Sen, 1997]***

The terminal telomeric repeats form the single-stranded guanine-rich overhang, with the length of 100 or more nucleotides in humans (see figure 16). It is believed that the terminal duplex region and the G-rich overhang are not packaged into a nucleosomal structure **[Rhodes et al., 2002]** and thus provide binding sites for various telomeric proteins **[Olaussen et al., 2006; Rhodes et al., 2002; Shafer & Smirnov, 2001]**. Some of them are supposed to bind to the putative quadruplex structures at the G-rich overhang and even to facilitate G-quadruplex formation **[Fang & Cech, 1993; Paeschke et al., 2005]** or disruption **[Zaug et al., 2005]**. It is interesting that DNA double-strand breaks and telomeres share several common proteins participating in DNA repair **[Olaussen et al., 2006]**. How these proteins perform opposing roles (DNA repair and telomere maintenance) is not yet fully understood **[Lundblad, 2000]**. The G-rich overhang may be tucked into the duplex region of the telomere (forming thus the so-called t-loop structure), which is important for telomere capping **[Stansel et al., 2001; Yoshimura et al., 2004]**. Functional telomeres, called capped telomeres, protect chromosomal ends from being recognized as double-strand breaks. They are inaccessible to nucleases and telomerase. It was proposed that telomere can rapidly switch in a highly regulated manner among its capped and uncapped state **[Blackburn, 2001; Lydall, 2003]**.

Telomeres are maintained and their lengths are elongated by the enzyme telomerase. Telomerase synthesizes G-rich telomeric repeats from its RNA template using 3'-end of chromosome as a primer **[Blackburn, 1999; Cech, 2000]**. This synthesis is believed to occur concurrently with telomere replication **[Blackburn, 2001; Chakhparonian & Wellinger, 2003; Zhu et al., 1996]**. Telomerase template RNA is denoted as TERC or TR (**te**lomerase **R**NA **c**omponent; **h**TERC or **h**TR in humans). Although TERCs considerably vary in sequence and size among species, their secondary structure is widely conserved **[Blackburn et al., 1999; Cech et al., 2000]**. In humans, TERC is 445 nucleotides long with the template sequence 5'-$CUA_2C_3UA_2C$-3' that codes for the telomeric repeats $(T_2AG_3)_n$ **[Liu, 1999]**. A catalytic protein subunit of telomerase referred to as TERT (**te**lomerase **r**everse **t**ranscriptase, **h**TERT in humans) is conserved in evolutionary diverse organisms and distantly related to retroviral reverse transcriptases like HIV **[Cech, 2000]**. Telomerase is probably a tetramer enzyme consisting of two RNA (TERC) and two catalytic protein subunits (TERT), similarly to other reverse transcriptases **[Mergny et al., 2002; Olaussen et al., 2006]**.

Most somatic cell types lack telomerase (or show low levels of it) and their telomeres thus shorten (30 to 200 bp per population doubling in vitro or 50 to 105 bp per year in vivo **[Dahse et al., 1997]**) progressively with each cell division. Telomerase activity in normal adult cells is restricted to only germ cells, various kinds of stem cells and activated lymphocytes **[Granger et al., 2002; Liu, 1999; Shay & Bacchetti, 1997]**. Replicative senescence is a special case of cellular senescence in response to telomere shortening **[Hwang, 2002]**. It is not the average but rather the shortest telomere that may elicit senescence response **[Hemann et al., 2001]**. Human cells have telomeres ~10 kbp long **[Dahse et al., 1997; Kelland, 2005]** and undergo approximately 60-70 population doublings **[Granger et al., 2002]** until they senesce. Due to their altered biochemical properties, senescent cells may compromise proper functions of surrounding cells and thus contribute to various age-related pathologies. Accelerated telomere shortening is associated with various progeroid (premature aging) disorders such as Werner's syndrome, Bloom's syndrome, Hutchinson–Gilford progeria or dyskeratosis congenita.

Immortalized or cancer cells must inevitably find a way how to maintain critically short telomeres. Most cells solve this problem by expressing high levels of telomerase **[Blasco & Hahn, 2003]**. Cancer and germ cells retain constant telomere lengths, and telomeres of pluripotent stem cells shorten at slower rates than those of telomerase-negative somatic cells. Only a small fraction (10-15%) of human malignant tumors do not show detectable telomerase activity and use the ALT (**a**lternative **l**engthening of **t**elomeres) mechanism of telomere maintenance **[Blasco & Hahn, 2003; Muntoni & Reddel, 2005; Perrem et al., 2001]**. Unlike telomerase-positive cells, which usually have short telomeres, ALT cells have heterogeneous telomeres ranging from very short to extremely long. These are maintained via recombinational processes. Although telomerase may suppress ALT, both mechanisms of telomere maintenance can coexist in the same cells **[Perrem et al.; 2001]**.

Telomeres and telomerase have nowadays become very attractive targets for many anticancer drugs since high telomerase activity is almost exclusively limited to cancer cells with malfunctional telomeres. Anti-telomerase drugs down-regulate or/and impair proper functionality of the telomerase protein complex. They can target either hTERT, hTR or various telomerase-associated proteins. While hTR is ubiquitously expressed in many human tissues (healthy or cancerous), hTERT

expression is restricted (with some exceptions) to cancer cells. Consequently, hTERT is a better candidate than hTR, if one desire to specifically target tumor cells. Small molecules like porphyrins or peptides can interfere with hTERT transcription and thus inhibit tumor growth **[Grand et al., 2002]**.

Anti-telomere drugs inhibit tumor growth by disrupting telomere structure, rather than targeting telomerase alone. Such drugs can be especially useful against cancer cells with very short telomeres and can substitute for telomerase inhibitors in cases when cancer cells do not maintain telomeres by telomerase (10-15% of human cancers) and activate ALT (**a**lternative **l**engthening of **t**elomeres) mechanism of telomere maintenance. Conventional cytotoxic drugs interfere with replication or the catalytic action of telomerase by cross-linking neighboring bases at telomeres (cisplatin) or sensitizing telomeric sequences to cleavage (etoposide) **[Mergny et al., 2002; Olaussen et al., 2006; Redon et al., 2003]**. Guanine quadruplexes forming at telomeres affect the catalytic activity of telomerase or block its access to telomeres **[Fletcher et al., 1998; Kelland, 2005; Mergny et al., 2002; Olaussen et al., 2006; Patel et al., 2004; Salazar et al., 1996]**. A very large number of compounds have been found to stabilize guanine quadruplexes **[Balasubramanian & Neidle, 2009; Georgiades et al., 2010]** and thus potentially serve as anticancer agents. Most of these stabilizing ligands are polyaromatic molecules with one or more positive charges (see the figure 17). They include porphyrins **[Izbicka et al., 1999; Martino et al., 2009; Parkinson et al., 2007; Shi et al., 2001]**, acridines **[Sparapani et al., 2010]**, perylenes **[Pivetta et al., 2008]**, triazines **[Gomez et al., 2004]**, anthraquinones **[Read & Neidle, 2000]**, ethidium derivatives **[Breuzard et al., 2003]** and other diverse compounds **[Rosu et al., 2006]**.

Quadruplex ligands must meet several criteria, besides stabilizing telomeric quadruplexes, in order to be used in vivo as efficient and safe anticancer drugs. Primarily, each ligand should preferentially bind only to a desired G-quadruplex over other structures (different quadruplexes, triplexes, duplexes or single strands) **[Mergny et al., 2002]**. A low selectivity for G-quadruplexes over duplexes is associated with DNA topoisomerase I/II (enzymes helping to unwind double-helical DNA during transcription and replication) inhibition resulting in non-specific acute cell death **[Neidle & Read, 2001]**. Unfortunately, many of G-quadruplex stabilizing drugs are cytotoxic at concentrations required for inhibition of telomerase activity **[Balasubramanian & Neidle, 2009]**. Another important requirement is that

quadruplex ligands should be able to pass through cell membranes and distribute preferably in the nuclei of targeted cells. Many research findings indicate that at least porphyrins readily and preferentially accumulate in tumor tissues **[Izbicka et al., 1999]**.

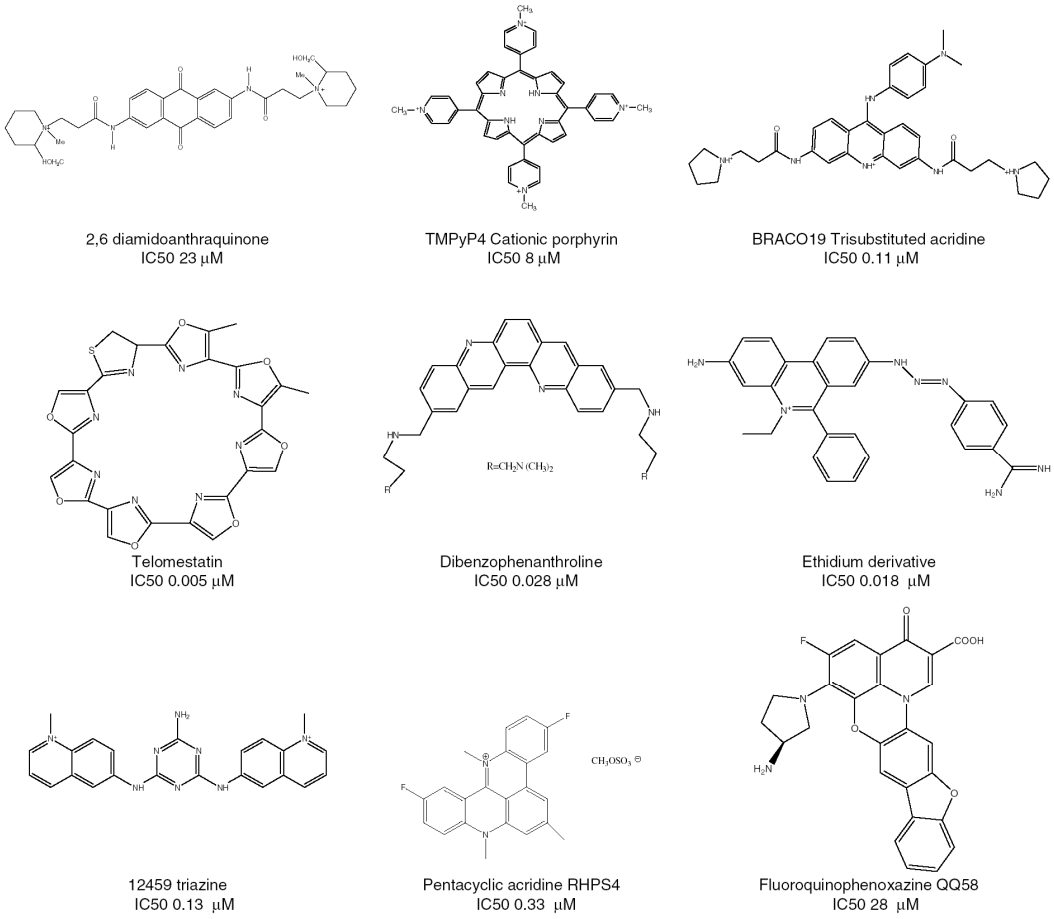



**Figure 17.** G-quadruplex stabilizing molecules. Values are for potency in cell-free assays to inhibit telomerase. Taken from reference ***[Kelland, 2005]***.

The mechanism behind the inhibition of telomerase activity by some anticancer agents is not clear **[Mergny et al., 2002; Seimiya et al., 2002]**. It is likely that, for some drugs, more than one mechanism may be involved. For example, the cationic porphyrin (TMPyP4) stabilizes both quadruplexes at telomeres and down-regulates hTERT (catalytic protein subunit of telomerase) transcription **[Grand et al., 2002]**. Anti-telomere and anti-telomerase drugs induce senescence or apoptosis after telomeres become seriously damaged and/or after they are shortened to a critical length. Consequently, these drugs can be only efficient against tumors with short

telomeres or in chemoprevention therapy. They are also supposed to have low toxicity since cancer cells usually have much shorter telomeres and proliferate much faster than normal somatic cells. The drawback of G-quadruplex ligands is that they may bind to non-telomeric guanine quadruplexes formed in normal tissues and that telomeres in cancer cells may be maintained via the ALT pathway **[Kelland, 2005]**. Anti-telomere and anti-telomerase therapeutic approaches will be in the future probably combined with conventional therapies like chemotherapy or radiotherapy since positive synergism has been already reported **[Kelland, 2005; Mergny et al., 2002; Olaussen et al., 2006; Shay & Wright, 2002; White et al., 2001]**.

### *1.5.2 Nontelomeric quadruplexes*

Non-canonical DNA structures like cruciforms, triplexes, hairpins, guanine quadruplexes or left-handed Z-DNA are implicated in diverse biological processes like transcription, replication or telomere maintenance **[Bacolla & Wells, 2004]**. Revealing structural features and true functions of these motifs is of therapeutic interest because non-B DNA conformations are linked to a variety of human pathologies such as cancer and many neurodegenerative disorders **[Bacolla & Wells, 2004]**. More and more scientific studies are now specifically directed at guanine quadruplexes showing convincingly that they are biologically relevant structures **[Arthanari & Bolton, 2001; Collie & Parkinson et al., 2011; Ghosal & Muniyappa, 2006; Shafer & Smirnov, 2001]**. Quadruplexes or other non-canonical structures compete with regular Watson-Crick helices, which predominate at physiological conditions in double-stranded regions of DNA. However, duplex DNA must be locally denatured to allow transcription to proceed. Each strand of DNA is then free to adopt alternative non-B DNA conformations. A number of factors influence quadruplex-duplex competition (see 1.3). Quadruplexes originating from duplex DNA regions can interact, much like those at telomeres, with various low molecular weight ligands such as porphyrins **[Cogoi et al., 2008; De Armond et al., 2005; del Toro et al., 2009; Sun et al., 2005]**. This interaction may result in altered gene expression and thus be of therapeutic value.

Quadruplexes formed at G-rich sites of regulatory regions of numerous genes may serve either as activators **[Catasti et al., 1996; Lew et al., 2000]** or repressors of gene transcription **[Cogoi et al., 2008; Seenisamy et al., 2005]**. Intramolecular quadruplexes within promoters of miscellaneous oncogenes including KRAS **[Cogoi**

et al., 2008], bcl-2 [Dai et al., 2006], c-kit [Fernando et al., 2006] or Rb (retinoblastoma) [Xu & Sugiyama, 2006] have been reported (see review [Qin & Hurley, 2008]). Most attention have been paid to the regulatory region of the c-myc oncogene [Ambrus et al., 2005; Grand et al., 2002; Halder & Chowdhury, 2005; Phan et al., 2004; Rangan et al., 2001; Seenisamy et al., 2005]. The members of the myc family of proto-oncogenes are known to function as transcriptional factors for a variety of genes regulating cellular proliferation [Dang, 1999]. Depending on the genetic context, c-myc can both activate and repress gene transcription. The c-myc gene is of interest as it is overexpressed in many human cancers. It is estimated that it contributes to one-seventh of U.S. cancer deaths [Dang, 1999]. Surprisingly, there is a link of c-myc to telomeres because c-myc is a transcription factor for the catalytic protein subunit of human telomerase, hTERT [Horikawa & Barret, 2003]. c-myc quadruplex is a mixture of four parallel propeller-like isomers with one isomer being predominant [Ambrus et al., 2005; Phan et al., 2004; Seenisamy et al., 2005]. Although G-quadruplexes are frequent transcriptional regulatory units of oncogenes, they may also influence transcription of other genes such as human insulin [Catasti et al., 1996; Lew et al., 2000], VEGF (vascular endothelial growth factor) [Sun et al., 2005], HIF-1α (hypoxia inducible factor 1α) [De Armond et al., 2005] or some muscle-specific genes [Etzioni et al., 2005; Yafe et al., 2005].

Microsatellites, also referred to as SSR (**s**imple **s**equence **r**epeats) or STR (**s**hort **t**andem **r**epeats), are tandem repeats of short sequences, typically consisting of 1 to 6 base pair long stretches [Toth et al., 2000]. They belong to the most polymorphic markers ever known [Agrawal & Khan, 2005]. Microsatellites consisting of three nucleotides have recently aroused intense interest among scientists since their abnormal expansion may result in about twenty severe hereditary neurological disorders, collectively known as **t**riplet (or **t**rinucleotide) **r**epeat **e**xpansion **d**iseases (TRED) [Mirkin, 2006; Richards, 2001; Wells et al., 2005]. However, tetrameric $(CCTG)_n$•$(CAGG)_n$, pentameric $(AATCT)_n$•$(AGATT)_n$ and even dodecameric $(C_4GC_4GCG)_n$•$(CGCG_4CG_4)_n$ repeats can also expand and cause disease [Mirkin, 2006]. Generally, there is a strong positive correlation between disease severity and the number of tandemly repeating base triplets. What mechanisms are behind repeat expansions and the nature in which they contribute to the disease pathology is not completely understood as yet and is an area of intense research. It is interesting that large intergenerational expansions of base triplets

associated with diseases are mostly restricted to humans [Sinden et al., 2002]. The propensity of tandem repeats for excessive expansion is nowadays explained by their ability to form a number of non-B DNA conformations such as slipped structures, hairpins, triplexes or quadruplexes [Mirkin, 2006; Sinden et al., 2002; Wells et al., 2005]. These non-canonical DNA structures interfere with many important biological processes such as DNA replication, repair, genetic recombination, transcription and translation [Jasinska et al., 2003; Mirkin, 2006; Richards, 2001; Wells et al., 2005]. Replication is most impaired of all processes [Wells et al., 2005]. Quadruplexes effectively block polymerases during DNA synthesis in potassium solutions [Usdin, 1998; Weitzmann et al., 1996]. Hairpins or slipped structures present during replication cause either repeat expansions or deletions depending on whether they are formed on the nascent or template strand, respectively [Mirkin, 2006; Shafer & Smirnov, 2001; Wells et al., 2005]. Expansion of triplet repeats is less likely if they are interrupted with nonrepeated sequences [Weisman-Shomer et al., 2000].

Quadruplex occurrence is not limited to telomeres, gene regulatory sites and some repetitive regions. Some other interesting examples of naturally occurring quadruplexes can be provided. A guanine quadruplex together with two hairpins flank the endonucleolytic cleavage site in a highly conserved part of the 3' untranslated region of **i**nsulin-like **g**rowth **f**actor **II** (IGF II) mRNAs [Christiansen et al., 1994]. These folded structures probably prevent the endonucleolytic cleavage site from adopting some stable RNA structure, which would make impossible IGF II posttranscriptional processing. Ribosomal DNA (rDNA) is naturally G-rich and thus has the propensity to form quadruplexes. Nucleolin, a protein which is implicated in rDNA transcription, replication or recombination, binds very strongly to rDNA through the G-quartet motif [Hanakahi et al., 1999]. Hairpin-like quadruplexes are formed within the G-rich DNA flap of the HIV-1 virus [Lyonnais et al., 2002]. The G-rich DNA flap arises during HIV-1 reverse transcription when two 99-nucleotides-long plus strands at the center of the viral genome overlap one another. This structure facilitates viral gene transduction. Under high salt concentrations in vitro, two HIV-1 RNA molecules self-associate via G-quartets [Sundquist & Heaphy, 1993]. Authors proposed that this also likely happens in vivo.

### *1.5.3 Guanine quadruplexes as aptamers*

Aptamers are short oligonucleotides (typically 10-100 bases) that have very high affinity and specificity for their desired targets **[Collie & Parkinson et al., 2011; Liu et al., 2009; Mairal et al., 2008]**. This specificity and affinity is similar to that for binding antibodies to antigens. Aptamers can be directed against diverse cellular targets including small molecules (amino acids, nucleotides, metabolites, hormones, antibiotics, etc.), proteins (antibodies, growth factors, enzymes, etc.), lipids, or nucleic acids. They are used in both therapeutic and non-therapeutic applications **[Collie & Parkinson et al., 2011; Davis, 2004; Liu et al., 2009; Mairal et al., 2008]**. Aptamers can be found in biosensors for diagnostics or separation of miscellaneous chemical compounds, they can be targeted against viral and bacterial proteins to impair their malicious functions, or they can be used to knock-out some genes in order to reveal their biological relevance. Unlike 'linear' antisense oligonucleotides, aptamers have a well-defined three-dimensional structure, which not only determines their binding properties, but also confers them greater resistance to nucleases and low toxicity in vivo. Aptamers are often chemically modified to modulate their biological properties and stability **[Petraccone et al., 2005; Sacca et al., 2005]**. They can be labeled to visualize the ligand-recognition process **[Nutiu & Li, 2004]**. Aptamers can be routinely prepared for almost any target molecule by the SELEX procedure (**S**ystematic **e**volution of **l**igands by **ex**ponential enrichment) **[Mairal et al., 2008]**.

Due to their well-defined three-dimensional geometry and conformational variability, guanine quadruplexes have been widely used as aptamers **[Collie & Parkinson et al., 2011; Davis, 2004]**. It is not surprising, in this respect, that the first therapeutic aptamer was G-rich. This aptamer was a 15-mer oligonucleotide (having the sequence d($G_2T_2G_2TGTG_2T_2G_2$)) selected against thrombin, which is an enzyme involved in blood clotting **[Collie & Parkinson et al., 2011]**. X-ray and NMR studies indicated that **t**hrombin-**b**inding **a**ptamer (TBA) is an intramolecular chair quadruplex **[Macaya et al., 1993]**. This is at variance with the results from the group of prof. Vorlíčková **[Fialová et al., 2006]** who argue that, in aqueous solutions, TBA has bimolecular rather than unimolecular folding. They have also shown that an elongation of the quadruplex loops by one or more thymines or an addition of the extra G-rich hexamer sequence at the 5' end causes that TBA folds intramolecularly.

In the Results and Discussion (section 3.4), we speculate on a dimeric structure of TBA composed of two intramolecular TBA quadruplexes. It was reported that, in the presence of thrombin, potassium may not be necessary for TBA to assume G-quadruplex conformation **[Baldrich & O'Sullivan, 2005]**. Thrombin then promotes TBA folding and behaves as a molecular chaperone.

DNA oligonucleotides seem to be promising as anti-HIV drugs, although they have not been so far approved for clinical use **[Jing & Xu, 2001]**. They can be directed against various viral components such as integrase, reverse transcriptase, envelope protein gp120 and genomic RNA **[Jing & Xu, 2001]**. According to the mechanism of their action, anti-viral oligonucleotides can be grouped as antisense, triplex and G-quartet inhibitors **[Jing & Xu, 2001]**. Although G-quartet inhibitors of HIV infection are mostly DNA oligonucleotides, G-rich RNA oligonucleotides have been also reported **[Fujihashi et al., 1994]**. Most of the known anti-HIV G-quartet aptamers are of a monomolecular type **[Bishop et al., 1996; de Soultrait et al., 2002; Jing et al., 1997; Jing et al., 2000; Urata et al., 2004]**. However, bimolecular **[Phan et al., 2005; Suzuki et al., 2002]**, tetramolecular **[Fujihashi et al., 1994; Koizumi et al., 2000; Wyatt et al., 1994]** or dimeric anti-HIV G-quartet aptamers **[Do et al., 2011]** have been also described. Tetramers, due to their very slow dissociation rate, can retain their biologically active form for a sufficiently long time to be useful as antiviral therapeutics **[Wyatt et al., 1996]**. G-quartet aptamers, though more stable than antisense oligonucleotides, are often modified to increase their stability and also to confer them higher anti-HIV activity. The modification where the natural phosphodiester linkages are replaced by phosphorothioate linkages is most frequent. Either some of the internucleoside linkages **[Jing et al., 1997]** or all of them are modified **[Suzuki et al., 2002]**. Moreover, also nucleobases can be modified **[Jing et al., 2000]**.

There are two main mechanisms how anti-HIV G-quartet oligonucleotides exert their inhibitory effect **[Tamura et al., 2000; Tarrago-Litvak et al., 2002]**. One mechanism involves their targeting to the envelope glycoprotein gp120, a protein responsible for viral binding and entry to cells **[Koizumi et al., 2000; Suzuki et al., 2002; Urata et al., 2004; Wyatt et al., 1994]**. The other mechanism is based on their binding to HIV-1 integrase and subsequent inhibition of HIV-1 integrase activity **[de Soultrait et al., 2002; Jing et al., 1997; Jing et al., 2000; Phan et al., 2005]**. HIV-1 integrase is an enzyme implicated in the integration of the viral

genome into the host genome [Craigie, 2001; Nair, 2002]. One of the anti-HIV G-quartet oligonucleotides, T30177 (denoted also as Zintevir or AR177), was already clinically tested in animals [Wallace et al., 2000]. Interestingly, some G-rich oligonucleotides that were originally devised as antisense drugs exert, in fact, their therapeutic effects by a certain non-antisense mechanism, which is related to G-quartet formation [Dapic et al., 2003]. G-rich DNA aptamers can also function as catalysts [Majhi & Shafer, 2006; Travascio et al., 1998]. Peroxidase activity of hemin (iron (III) protoporphyrin) under physiological conditions increases by two orders of magnitude when hemin is bound to the quadruplex aptamer structure. G-rich aptamers have been also used as stationary phases in capillary electrochromatography [Vo & McGown, 2004].

### *1.5.4 Interaction of guanine quadruplexes with biomolecules*

During last few decades, a considerable research effort has been directed to understanding how guanine quadruplexes interact with small molecular ligands [Monchaud & Teulade-Fichou, 2008]. As it follows from the previous overview, low molecular weight ligands stabilizing guanine quadruplexes at telomeres can be used as potential drugs (see the section 1.5.1). There are currently only few structural data (NMR or X-ray) available on ligand-quadruplex complexes [Haider et al., 2011]. On the basis of numerous studies, both structural (NMR, X-ray) and non-structural (UV, CD, fluorescence, microcalorimetry, photocleavage experiments, etc.), it can be concluded that there are basically three main ways in which small aromatic ligands interact with guanine quadruplexes. They stack on terminal guanine tetrads [Evans et al., 2007], bind into grooves [Cosconati et al., 2010] or locate within quadruplex loops [Cogoi et al., 2008]. Specially, some ligands have been shown to intercalate between outermost guanine tetrads of two quadruplexes stacking in an end-to-end manner [Clark et al., 2012; Fedoroff et al., 1998]. True intercalation of ligands between adjacent guanine tetrads of single quadruplex is hardly likely, probably because it may bring about distortions in the quadruplex geometry [Mita et al., 2006; Read & Neidle, 2000]. However, some works give some evidence in favour of intercalation [Anantha et al., 1998; Haq et al., 1999; Keating & Szalai, 2004; Wei et al., 2006; Yamashita et al., 2005].

Binding characteristics of DNA-interacting molecules can be influenced by modification of their structures. This can be nicely illustrated for the case of

porphyrins, the interaction of which with nucleic acids has been extensively studied for many years **[Shi et al., 2001]**. Three cationic porphyrins known as TMPyP2, TMPyP3, and TMPyP4 are depicted in figure 18. These porphyrins, especially TMPyP4 **[Martino et al., 2009; Parkinson et al., 2007]**, are the subject of many recent papers that help to clarify the nature of quadruplex-ligand interaction. Although each of the three porphyrins has four identical meso-aryl substituents, these substituents vary among individual porphyrins. TMPyP2, TMPyP3, and TMPyP4 bear N-methyl groups in the position orto, meta or para, respectively. The position of N-methyl groups within meso-aryl substituents determines the degree to which these substituents are free to rotate about bonds linking them to the aromatic porphyrine core. The closer to the porphyrine core N-methyl groups are, the more restricted due to steric reasons the rotation of meso-aryl substituents is. A constrained rotation makes it difficult for porphyrins to stack properly on guanine tetrads. It is reflected in the different propensity of TMPyP2, TMPyP3 and TMPyP4 to stabilize guanine quadruplexes and inhibit telomerase. TMPyP2 is the extreme case, with meso-aryl substituents practically unable to rotate. Not surprisingly, it has been shown by many studies that TMPyP2 is a poor quadruplex stabilizer and telomerase inhibitor **[De Armond et al., 2005; Han et al., 2001; Shi et al., 2001; Yafe et al., 2005]**. However, it was also reported that both TMPyP2 and TMPyP4 stabilize intramolecular guanine quadruplexes from human telomeric DNA to approximately the same extent **[Han et al., 1999]**. The different potency of TMPyP2 and TMPyP4 to inhibit telomerase may be related to their distinct quadruplex binding modes. While TMPyP2 binds to the loop region, TMPyP4 stacks on the terminal guanine tetrad **[Han et al., 1999]**.

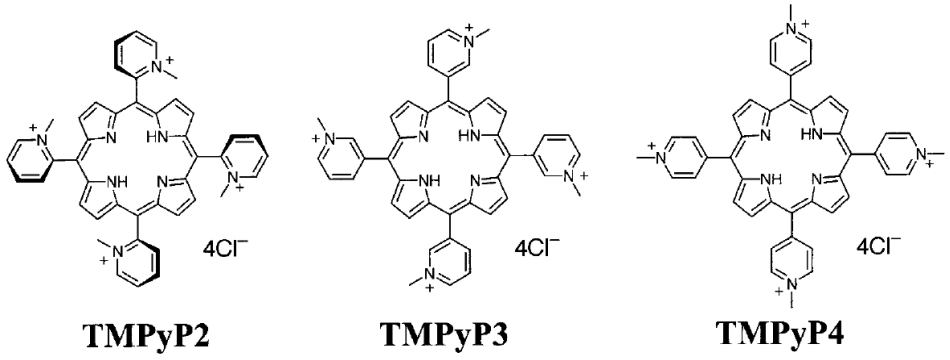



**Figure 18.** Structures of three cationic porphyrins: TMPyP2, TMPyP3, TMPyP4. Taken from reference *[Han et al., 2001]*.

Generally, the capacity of the porphyrins to stabilize quadruplexes and to inhibit telomerase can be expressed as TMPyP4 > TMPyP3 >> TMPyP2 **[Shi et al., 2001]**. However, TMPyP3 and TMPyP4 differ greatly in their ability to inhibit unwinding of quadruplexes by Sgs1 helicase. In the case of parallel folding, TMPyP3 is a stronger inhibitor of Sgs1 helicase than TMPyP4, but the opposite is true for antiparallel folding **[Han et al., 2001]**. In addition to the three mentioned porphyrins, many others can be synthesized **[Shi et al., 2001]**. They differ in the chemical structure of side chains (all chains can be different) or/and metalation. Generally, it holds that positively charged substituents are desired for telomerase inhibition. These substituents may be replaced and combined with hydrogen-bonding groups **[Shi et al., 2001]**. A core-modified expanded porphyrin analogue, $Se_2$SAP, was synthesized to have enhanced selectivity for intramolecular over intermolecular quadruplexes **[Seenisamy et al., 2005]**. Moreover, the photocytotoxicity of $Se_2$SAP is greatly reduced due to the presence of two selenium atoms in the core.

Selectivity as well as affinity of ligands towards guanine quadruplexes can be modulated by solution conditions, namely ligand and DNA concentration, temperature, ionic composition and pH **[Alzeer & Luedtke, 2010]**. Guanine quadruplexes may change their conformation **[Pradhan et al., 2011; Rezler et al., 2005; Seenisamy et al., 2005]** or even be destabilized upon ligand binding **[Weisman-Shomer et al., 2003; Yafe et al., 2005]**. The loop geometry determines to a great extent the affinity of a particular ligand for particular quadruplex conformations **[Campbell et al., 2009; Guan et al., 2011]**. Many molecular ligands such as pyridine dicarboxamide derivatives **[De Cian & Mergny, 2007]**, PIPER (perylene diimide derivative) **[Rangan et al., 2001]**, telomestatin (the metabolite produced by actinobacteria known as Streptomyces anulatus 3533-SV4; see figure 17) **[Kim et al., 2003; Rezler et al., 2005]**, the cationic porphyrin TMPyP4 **[Kim et al., 2003]**, a novel porphyrin with cationic side arms containing ammonium ions **[Murashima et al., 2008]**, a polyamine-anthracene conjugate **[Rodriguez et al., 2007]**, oxazine 750 dye **[Chen et al., 2009]** or a chiral metallo-supramolecular complex **[Zhao et al., 2011]** not only stabilize guanine quadruplexes but also accelerate or induce their formation. Interestingly, some G-quadruplex ligands promote quadruplex formation even in the absence of stabilizing monovalent cations **[Chen et al., 2009; Kim et al., 2003; Murashima et al., 2008; Rezler et al., 2005; Rodriguez et al., 2007; Zhao et al., 2011]**. G-quadruplex ligands that discriminate

between various guanine quadruplexes and that have low affinity for DNA duplexes can be used to prepare and/or stabilize desired quadruplex structures in solution **[Hamon et al., 2011; Rezler et al., 2005]**. A combination of several ligands may help to discriminate among many conformations (duplex, quadruplex) simultaneously present in solution **[Arthanari et al., 1998; Breuzard et al., 2003; Chen et al., 1996]**.

Specifically, TMPyP4 predominantly facilitates and then stabilizes intermolecular quadruplexes **[Kim et al., 2003]**. By contrast, telomestatin was shown to convert the preformed mixed parallel/antiparallel structure of the human telomeric quadruplex $d(T_2AG_3)_4$ to the intramolecular basket conformation **[Rezler et al., 2005]**. The ligand PIPER is interesting in that it facilitates quadruplex formation both from single strands and from Watson-Crick duplex DNAs **[Rangan et al., 2001]**. The preference of ligands that are structurally related to PIPER for quadruplexes over duplexes depends on their aggregation state which is extremely sensitive to pH (maximalized at the pH 7) **[Kern et al., 2002]**. It is possible that the preference for quadruplexes over duplexes, dependent on the pH, is not only limited to perylene analogues since, for instance, some porphyrins are also known to aggregate readily in aqueous solutions **[Dixon & Steullet, 1998; Kano et al., 1990; Pasternack et al., 1998]**.

Not only the structure but also the chirality of the ligand may be a determining factor for its interaction with guanine quadruplexes. For instance, only one enantiomer (having P helicity) from metallo- supramolecular cylinders $[M_2L_3]Cl_4$ (M = Ni or Fe) differentiates between G-quadruplex and duplex DNA **[Yu et al., 2008]**. Furthermore, the P-enantiomer induces the transition of the antiparallel form of the telomeric G-quadruplex $AG_3(T_2AG_3)_3$ in the presence of NaCl to hybrid parallel/antiparallel strand topology. It was also shown that only left-handed cyclic-helicene molecule has the binding preference for higher-order telomeric G-quadruplexes **[Xu et al., 2006]**. The above-mentioned telomestatin also exhibits enantioselectivity. The (S)-stereoisomer of telomestatin has four times higher propensity to induce parallel folding than its natural counterpart (R)-telomestatin **[Doi et al., 2011]**.

## 1.6 Experimental techniques used for studying guanine quadruplexes

A variety of experimental techniques are employed to probe biophysical properties and to determine the structure of G-quadruplexes **[Collie & Parkinson et al., 2011; Huppert, 2008; Keniry, 2001; Lane et al., 2008]**, as well as to monitor their interaction with small ligands **[Murat et al., 2011]**. Here, a brief overview of the application of the most common of these methods is given.

### *1.6.1 Absorption spectroscopy*

**U**ltra**v**iolet-**vis**ible (UV-vis) absorption spectroscopy is a routine method of optical spectroscopy for studying biological samples. Being relatively cheap and easy to use, UV-vis spectrometers now belong to standard equipment of biological and chemical laboratories. Importantly, samples for UV-vis absorption measurement can be prepared down to micromolar concentrations. Since the sixties, UV spectroscopy has been extensively used for studying thermal stabilities of many nucleic acid forms. Unfortunately, UV spectra are relatively broad and cannot easily distinguish among DNA or RNA forms with the slightly different sequences or conformations. Other methods, instead of UV spectroscopy, are therefore presently in use (see below) to characterize in detail the behavior of nucleic acids in solution. Nevertheless, UV spectroscopy is nowadays successfully applied for the determination of nucleic acid concentrations. Assuming the validity of Beer-Lambert law, the accuracy of the UV determined concentrations depends primarily on precise knowledge of the appropriate extinction coefficients **[Kallansrud et al., 1996]**. They are provided by NMR spectroscopy (most precisely) **[Cavaluzzi et al., 2004]** or by chemical means **[Murphy et al., 1996]**.

It is a common practice to monitor nucleic acid conformational changes, upon thermal or chemical denaturation of samples, by recording UV absorbance at one or more wavelengths (usually around 260 nm as nucleic acid bases strongly absorb in this region). Quadruplex folding and dissociation was followed by measuring absorbance at wavelengths ranging from 230 to 300 nm (see for example **[Alberti & Mergny, 2003; Green et al., 2003; Ren et al., 2002]**). The question arises as to what wavelength is most suitable for studies of quadruplex melting. Mergny and co-workers originally proposed that recording absorbance at 295 nm provide more

accurate melting curves than recording absorbance changes at wavelengths around 260 nm **[Mergny et al., 1998]**. Later they analyzed thermal difference spectra, which are defined as the difference between absorption spectra measured at high and low temperatures **[Mergny et al., 2005b]**. They claim that such spectra are specific for closely related structures and easily differentiate among various nucleic acid conformations (duplexes, triplexes, quadruplexes, I-motifs). These spectra can even distinguish between quadruplexes formed in $Na^+$ and $K^+$ solutions. Thermal difference spectra thus provide additional information to circular dichroism spectroscopy.

The example of thermal difference spectrum is given in figure 19 for the case of TBA quadruplex, d($G_2T_2G_2TGTG_2T_2G_2$). The thermal difference spectrum shows two characteristic bands (~274 nm, ~295 nm) reflecting the changes in the quadruplex structure upon melting. Generally, the disruption of the G-quadruplex structure (the molecularity is not important) is accompanied by hyperchromism at 240-245 nm, 270-275 nm and a strong hypochromism at 295 nm (unlike other DNA and RNA regular forms where the greatest change in absorbance occurs around 260 nm) **[Mergny et al., 1998; Mergny & Lacroix, 2003; Mergny et al., 2005a; Mergny et al., 2005b; Mergny & Lacroix, 2009]**. The change in absorbance at ~295 nm was frequently used by many researchers as an indicator of quadruplex thermal stability. The recommendations for conducting UV melting experiments with G-quadruplexes **[Mergny & Lacroix, 2009]** and the procedure for extracting thermodynamic parameters from UV melting curves **[Mergny & Lacroix, 2003]** was proposed by Mergny and Lacroix.

Although melting curves for G-quadruplexes are obtained by CD spectroscopy rather than absorption spectroscopy, the absorption in the visible region is often exploited for studying the interaction of G-quadruplexes with small aromatic molecules. This is possible since many G-quadruplex-interacting molecules (see figure 17) have absorption bands in the visible spectral range. Porphyrins, for example, have the strongest absorption bands between 400 and 700 nm. The Soret band of porphyrins is the most intense and most sensitive to interaction with other molecules **[Huang et al., 2000]**. Its position depends on the type of porphyrin substituents **[Huang et al., 2000]**. It is typically found at about 400 nm, but some porphyrins have been shown to have the Soret band above 600 nm **[Spence & Lash, 2000]**. In addition to the Soret band, free base porphyrins also have four weakly

intense absorption bands (referred to as Q bands) in the range approximately 500-700 nm. Metalated porphyrins have only two Q bands due their higher symmetry **[Huang et al., 2000]**. The interaction of porphyrins with DNA mainly manifests in spectral changes within the Soret band, namely in its bathochromic shift and hypochromicity (the compilation of data on the interaction of cationic porphyrins with various polyoligonucleotides and G-quadruplex-forming sequences is given in table 5). The character of these changes helps to reveal the nature of the porphyrin-DNA interaction. In general, it is considered that the interaction of porphyrins with guanine quadruplexes results in larger red shifts and larger hypochromicities of the Soret band than the interaction with WC duplexes or single strands. The intercalation features a substantial red shift (≥15 nm) and a hypochromicity (≥35 %) of the Soret band **[Keating & Szalai, 2004; Pasternack, et al. 1983]**. On the other hand, the groove binding is accompanied by a small red shift (≤8 nm) and either hyperchromicity or small hypochromicity (≤10 %) of the Soret band **[Keating & Szalai, 2004; Pasternack, et al. 1983]**.

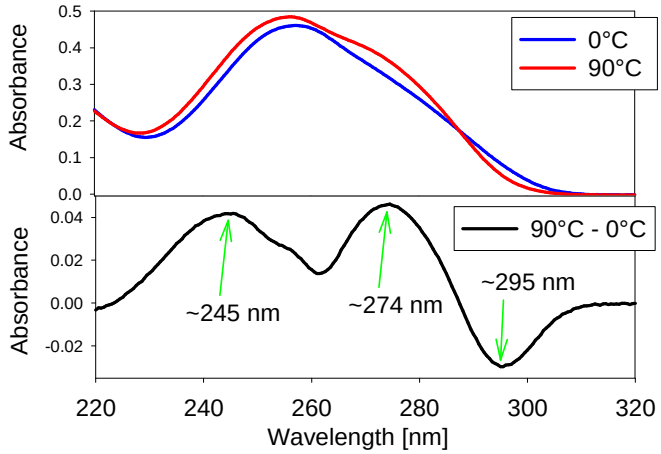



**Figure 19.** The comparison of the absorption spectra of the TBA sequence d($G_2T_2G_2TGTG_2T_2G_2$) recorded at low (0°C) and high (90°C) temperature. The difference spectrum between the spectrum at 90°C and 0°C depicted at the lower part of the figure illustrates the spectral changes accompanying the melting of the TBA quadruplex. A decrease in absorbance at ~295 nm and a concomitant increase at ~ 245 nm and ~274 nm observed upon melting is a signature of the quadruplex structure. Experimental conditions: 20 mM $Li^+$ cacodylate buffer, 100 mM $K^+$, pH 6.8.

| Porphyrin | Oligonucleotide | [a] Hypochromicity | Soret band shift | [b] DNA type | Reference |
|---|---|---|---|---|---|
| $H_2$TMPyP4 | ($TG_4T$) | 56.7 | 11 | tetrameric Q | [Wei et al., 2006] |
| | ($G_4T_4G_4$) | 62.8 | 13 | dimeric Q | |
| | $AG_3(T_2AG_3)_3$ | 61.1 | 13 | monomeric Q | |
| | $T_4G_4$ | 40.6 | 12 | tetrameric Q | [Anantha et al., 1998] |
| | $(CG)_2ATAT(CG)_2$ | 34.4 | 10 | WC duplex | |
| | poly(rA)•poly(rU) | 39 | 16 | WC duplex | [Uno et al., 2002] |
| | poly(rA)•poly(dT) | 38 | 16 | WC duplex | |
| | poly(rI)•poly(rC) | 46 | 11 | WC duplex | |
| | poly(rI)•poly(dC) | 39 | 15 | WC duplex | |
| | TTAGGG | 66 | 18 | tetrameric Q | [Mita et al., 2006] |
| CuTMPyP4 | $T_4G_4T_4$ | 50 | 9 | tetrameric Q | [Keating & Szalai, 2004] |
| | $T_4G_8T_4$ | 58 | 12 | tetrameric Q | |
| | $A_4C_4A_4$ | 25 | 5 | single strand | |
| | $A_4C_8A_4$ | 27 | 6 | single strand | |
| | poly(rA)•poly(rU) | 37 | 9 | WC duplex | [Uno et al., 2002] |
| | poly(rA)•poly(dT) | 33 | 10 | WC duplex | |
| | poly(rI)•poly(rC) | 25 | 5 | WC duplex | |
| | poly(rI)•poly(dC) | 23 | 8 | WC duplex | |

**Table 5.** Changes in the Soret band of G-quadruplex-forming oligonucleotides. [a]The hypochromicity values (%) are calculated in terms of the formula $[(\varepsilon_f-\varepsilon_b)/\varepsilon_f]\times100$, where $\varepsilon_f$ and $\varepsilon_b$ are the extinction coefficients (at the Soret band maxima) of the free and bound porphyrins, respectively. [b]Abbreviations: WC=Watson Crick, Q=quadruplex.

### *1.6.2 Circular dichroism*

CD spectroscopy **[Vorlíčková et al., 2012]** together with chemical probing methods (see 1.6.5) and NMR (see 1.6.6) are the major source of information about guanine quadruplexes and their interaction with other biomolecules. I-motif DNA tetraplexes have also been studied by CD spectroscopy **[Guo et al., 2007; Xu & Sugiyama, 2006]**. CD spectroscopy clearly differentiates between two main G-quadruplex topologies: parallel and antiparallel one (see figure 20). Parallel strand alignment is characterized by a positive maximum at ~260 nm and a negative minimum at ~240 nm **[Neidle & Balasubramanian, 2006]**. Antiparallel quadruplexes have a positive maximum at ~290 nm and a negative minimum at ~260 nm **[Neidle & Balasubramanian, 2006]**. Both quadruplex topologies also exhibit positive peak at 210 nm **[Masiero et al., 2010]**. The cytosine quadruplexes have a strong positive maximum of ellipticity around 290 nm **[Kypr et al., 2009]**. The Watson-Crick hairpin has a positive maximum at 286 nm **[Hardin et al., 1993; Keniry, 2001]**.

The classification of G-quadruplexes according to the strand orientation (parallel versus antiparallel) is frequently used by many researchers although it

provides only limited information about their structure (it was mentioned in the section 1.2 that there are 26 possible looping topologies). The usefulness and power of CD spectroscopy consists in its ability to quickly and easily characterize quadruplex samples at various environmental conditions, primarily temperature, ionic strength and pH. CD spectroscopy clearly distinguishes between G-quadruplex structures adopted in the presence of $K^+$ and $Na^+$ ions **[Dapic et al., 2003; Vorlíčková et al., 2012]**. The interpretation of CD spectra of G-quadruplexes can be problematic when two or more G-quadruplex species with distinct kinetics are simultaneously present in solution **[Phan & Patel, 2003; Rujan et al., 2005]**. In such cases, CD spectra have features of both the parallel and antiparallel folding. Multivariate statistical methods applied to CD or UV thermal melting curves of guanine quadruplexes can reveal the number of potential quadruplex conformers **[Antonacci et al., 2007; Jaumot et al., 2006; Petraccone et al., 2005]** and calculate thermodynamic data **[Gray & Chaires et al., 2011]**.

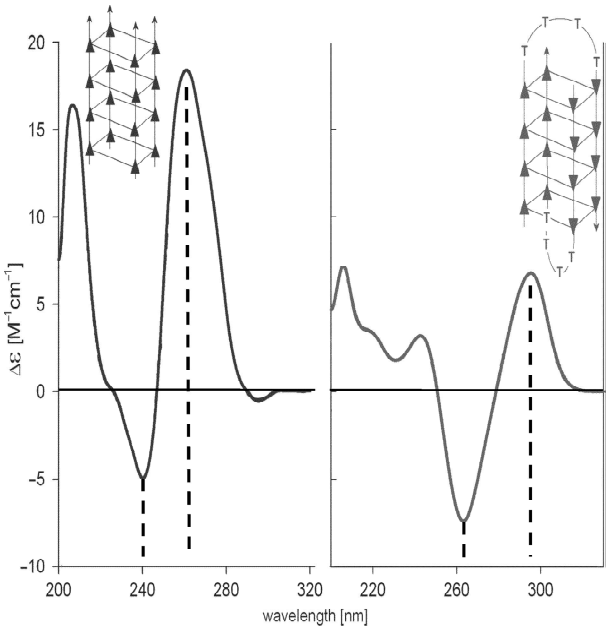



**Figure 20.** Characteristic CD spectra corresponding to two main quadruplex topologies. On the left: the parallel quadruplex $[d(G_4)_4]_4$ stabilized by 16 mM $K^+$. On the right: $Na^+$- induced antiparallel bimolecular quadruplex of $[d(G_4T_4G_4)]_2$. The triangles in the sketches indicate guanines and point in the 5´-3´ direction. Δε is expressed in terms of nucleotide molar concentration. Taken from reference ***[Masiero et al., 2010]***.

The empirical rules such as those relating the above-mentioned characteristic CD spectra of G-quadruplexes to parallel and antiparallel folding are mainly derived on the basis of the comparison of CD data with the known structures provided by NMR or X-ray analysis. However, these rules are not automatically applicable for all quadruplex-forming oligonucleotides **[Masiero et al., 2010]**. They are meaningless for the interpretation of CD spectra of monomeric guanosines derivatives, G-rich modified oligonucleotides or the oligonucleotides having the 3´-3´ and 5´-5´ inversion of polarity sites **[Masiero et al., 2010]**. For example, the "parallel" tetramolecular quadruplex $[d(TGG^{Br}GT)]_4$ ($G^{Br}$ denotes an 8-bromoguanosine residue) exhibits the same CD spectrum as the "antiparallel" bimolecular quadruplex $[d(G_3T_4G_3)]_2$ **[Masiero et al., 2010]**. CD spectra of G-quadruplexes are determined by the polarity of stacked G-quartets (see figure 21) rather than the relative orientations of the adjacent sugar-phosphate backbones (see figure 8). The polarity of neighboring G-quartets can be modulated by an intentional adjustment of the orientation of phosphate backbone **[Krauss et al., 2011; Pagano et al., 2008]**. The characteristic positive and negative CD bands of the parallel and antiparallel G-quadruplexes can be then explained in terms of distinct orientations of electronic transition dipoles for different stacking geometries **[Gray et al., 2008; Karsisiotis et al., 2011; Masiero et al., 2010]** (see figure 21).

CD spectroscopy in the visible range, as well as absorption spectroscopy, gives information about the mode of interaction of small aromatic molecules with guanine quadruplexes **[Chang et al., 2007; Chen et al., 1996; Keating & Szalai, 2004; Lubitz et al., 2007; Rezler et al., 2005; Sun et al., 2006; Yamashita et al., 2005]**. Although many G-quadruplex interacting ligands such as porphyrins are not chiral, they become structurally distorted and provide induced CD signal upon binding to chiral G-quadruplexes **[Keating & Szalai, 2004]**. The big advantage of CD spectroscopy over absorption spectroscopy is that it better separates spectral changes occurring in DNA and G-quadruplex ligands since many of them exhibit no or negligible CD signal in the UV absorption region of nucleic acids. While the ultraviolet portion of CD spectra (220-330 nm) reflects only the conformational properties of DNA in the presence of ligands, the visible part of CD spectra is exclusively associated with induced chirality due to the interaction of ligands with guanine quadruplexes. The visible part of CD spectra thus indirectly reflects the changes in the G-quadruplex geometry provoked by the presence of ligands.

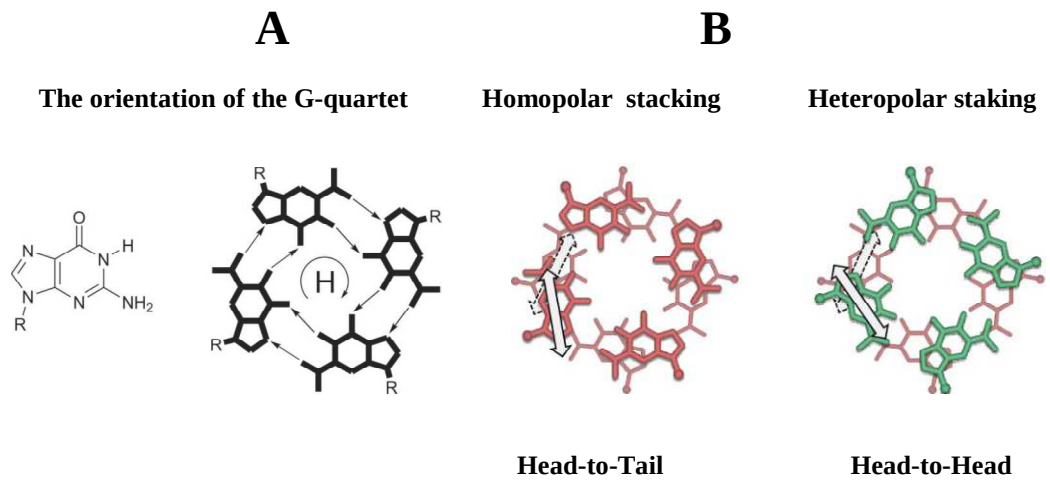



**Figure 21.** (**A)**: The guanine moiety and the G-quartet. The G-quartet shows its "head" (H) side (donor to acceptor H-bonding runs clockwise). The reverse side is referred to as "tail" (donor to acceptor H-bonding runs counter-clockwise).G-quartet can stack through the same (head-to-head or tail-to-tail) or the opposite (head-to-tail) face resulting in heteropolar or homopolar stacking, respectively. (**B)**: Top view of the heteropolar and homopolar stacking of two G-quarterts: the "head" and the "tail" sides of the G-quartets are represented in red and green, respectively. The double head arrows represent the electric transition moments corresponding to the absorption band at ~250 nm. Taken from reference *[Masiero et al., 2010]*.

### *1.6.3 Fluorescence*

Fluorescence methods have been extensively used for studying various aspects of guanine quadruplexes. They can provide information about thermodynamic stability and kinetics of quadruplexes **[Alberti & Mergny, 2003; Darby et al., 2002; Green et al., 2003; Li et al., 2005; Risitano & Fox, 2003; Xu & Sugiyama, 2006; Ying et al., 2003]** as well as about their interaction with other biomolecules **[Anantha et al., 1998; Arthanari et al., 1998; Juskowiak, 2011; Keating & Szalai, 2004; Kumar et al., 2008; Rachwal & Fox, 2007; Renčiuk et al., 2012; Wei et al., 2006; Yamashita et al., 2005]**. The obvious benefit of fluorescence methods over other spectroscopic techniques is their high sensitivity **[Darby et al., 2002]**. It means that samples for fluorescence measurements can have much lower concentrations (typically about two or three orders of magnitude lower) than samples used for UV and CD spectroscopy. Moreover, the conformational changes in nucleic acid structure are translated into large relative changes in fluorescence signal. The

definite advantage of fluorescence techniques over absorption and CD spectroscopy is their higher sample throughput. For instance, Darby and coworkers could record on their fluorimeter 32 melting profiles in parallel **[Darby et al., 2002]**.

Natural nucleic acids are not fluorescent or show only weak fluorescence for practical purposes. They can become fluorescent if fluorophores are introduced into their structure. Fluorophores may, however, alter biophysical properties of nucleic acid samples under investigation. A fluorescent analogue of adenine, 2- aminopurine, has been extensively used for about 40 years as a fluorophore in nucleic acids due to its high quantum yield and excitation profile shifted to the red, far from the absorbance region of protein and naturally occurring nucleobases **[Buscaglia et al., 2012; Li et al., 2005]**. This modified base pairs with thymine, minimally distorts DNA duplexes and its fluorescence is quenched upon stacking with other bases **[Li et al., 2005; Xu & Sugiyama, 2006]**. It has been used, for instance, to examine loop conformations in the human telomere quadruplex **[Li et al., 2005]** and to monitor quadruplex to duplex conversion of the G-rich oligonucleotide, with the sequence derived from the Rb gene, upon addition of the complementary C-rich oligonucleotide **[Xu & Sugiyama, 2006]**. Interestingly, although DNA sequences exhibit only very low fluorescence, it was reported in several studies that the fluorescence quantum yield of guanines is significantly increased upon the formation of the G-quadruplex structure **[Dao et al., 2011; Mendez & Szalai, 2009]**. This intrinsic fluorescence of guanines can be employed to follow the formation of G-quadruplexes without the need for targeting their structures with fluorescent dyes or incorporating modified guanine bases into G-rich DNA sequences.

FRET (Förster **r**esonance **e**nergy **t**ransfer) has been extensively used to probe secondary structures of nucleic acids **[Shirude & Balasubramanian, 2008]**. It is based on nonradiative energy transfer from a donor fluorophore, which captures light energy, to an acceptor fluorophore. FRET is employed to measure distances since the efficiency of FRET decreases approximately with the six power of the distance between the donor and acceptor molecules. To observe FRET, G-quadruplex-forming oligonucleotides are labeled at with a fluorophore (donor) at one end (fluorescein, cyanine dyes, pyrene, etc.) and quencher (acceptor) at the other (dabsyl, rhodamine derivatives, methyl red, perylene, etc.). Nevertheless, fluorophores and quenchers have also been attached to sugars or bases at oligonucleotide positions other than 5’ and 3’ ends **[Darby et al., 2002]**. FRET allows to study the folding and

denaturation of G-quadruplexes because the distance between two fluorophores depends on the particular structure (see figure 22) **[De Cian et al., 2007; Juskowiak, 2011; Kumar & Maiti, 2004; Renčiuk et al., 2012; Risitano & Fox, 2003; Shirude & Balasubramanian, 2008]**. Apart from guanine quadruplexes, FRET was also used to probe the formation of cytosine tetraplexes **[Choi et al., 2011; Mergny, 1999]**. The great difference between the FRET signal corresponding to an unordered G-rich oligonucleotide (in the absence of stabilizing cations) and an intramolecular quadruplex folding (in the presence of $K^+$) can be employed as a very sensitive sensor of potassium ions **[He at al., 2005; Nagatoishi et al., 2006; Ueyama et al., 2002]**.

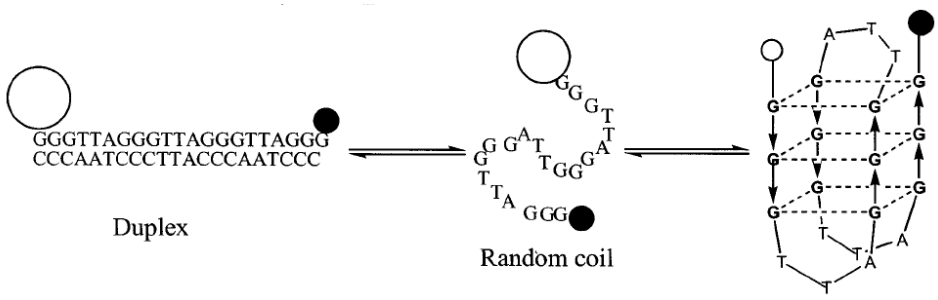



**Figure 22.** Schematic representation of the folding of intramolecular quadruplexes and the formation of duplexes with the complementary oligonucleotide. The fluorophore and quencher are represented by the open and closed circles, respectively. The duplex is shown on the left, the random coil is in the center, and the folded quadruplex is shown on the right. Although several different folded structures are possible, for simplicity only one antiparallel structure is shown. Taken from reference ***[Risitano & Fox, 2003]***.

Some quadruplex ligands such as ethidium derivatives or porphyrins are weak fluorophores. Fluorescence emission of these ligands increases significantly upon their interaction with DNA (see figure 23) **[Anantha et al., 1998; Arthanari et al., 1998; Keating & Szalai, 2004; Ren et al., 2002; Wei et al., 2006; Yamashita et al., 2005]**. The degree of fluorescence enhancement indicates the ligand binding mode. It is greatest for the intercalative binding because it allows an efficient FRET from DNA bases to ligands **[Anantha et al., 1998; Yamashita et al., 2005]** and shields ligands from solvent quenching **[Keating & Szalai, 2004; Wei et al., 2006]**.

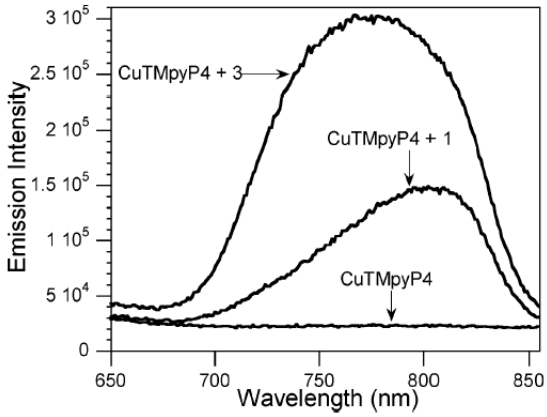



**Figure 23.** Emission spectra ($\lambda_{exc}$=434 nm) of 5 μM CuTMpyP4 alone, 5 μM CuTMpyP4 with 24 μM $T_4G_4T_4$ (denoted as 1) or 5 μM CuTMpyP4 with 24 μM $T_4G_8T_4$ (denoted as 3). All solutions contain 10 mM potassium phosphate and 50 mM KCl at pH 7.1. Taken from reference ***[Keating & Szalai, 2004]***.

### *1.6.4 Raman scattering*

Raman spectroscopy is far more sensitive to the secondary structure of nucleic acids than CD or UV-vis spectroscopy (see reviews **[Benevides et al., 2005; Thomas, 1999]**). It has been used since the sixties for studying both the structure of nucleic acids and their interaction with other biomolecules. Unlike calorimetric methods (see the section 1.6.7), Raman spectroscopy can differentiate between local and global melting effects **[Benevides et al., 2005]**. It provides information about the phosphodiester backbone conformation, sugar puckering, glycosidic torsion angles, base stacking interactions and hydrogen bonding. Unfortunately, as it is well known, Raman scattering is a very weak phenomenon. Samples of nucleic acids for non-resonant Raman spectroscopy therefore must be prepared at concentrations exceeding by two or three orders of magnitude those required for UV-vis or CD spectroscopy. This problem can be overcome by selecting the excitation wavelength around 260 nm, which is the spectral region where nucleic acid bases strongly absorb. For this choice, the intensity of Raman scattering of nucleic acids is strongly enhanced and fluorescence is usually negligible. The drawback of resonance Raman spectroscopy is that not all vibrational bands are significantly resonantly enhanced. **U**ltra**v**iolet **r**esonance **R**aman scattering (UVRR) spectroscopy does not provide any information about sugar and phosphate group vibrations (see figure 24) **[Benevides et al., 2005]**.

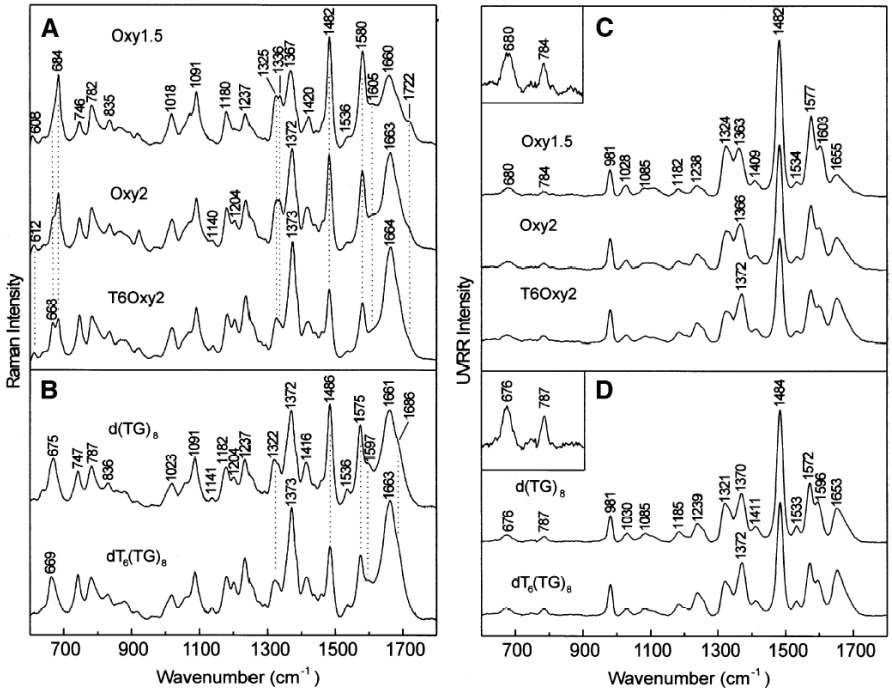



**Figure 24.** (**A**) Raman spectra of $d(G_4T_4G_4)$ (Oxy1.5), $d(T_4G_4)_2$ (Oxy2), $dT_6(T_4G_4)_2$ (T6Oxy2). (**B**) Raman spectra of $d(TG)_8$ and $dT_6(TG)_8$. In (A) and (B) solutions contained DNA at 30 mg/ml in 10 mM Tris pH 7.2, 1 mM EDTA and 50 mM NaCl and were maintained at 10°C. (**C**) UVRR spectra of Oxy1.5, Oxy2 and T6Oxy2. (**D**) UVRR spectra of $d(TG)_8$ and $dT_6(TG)_8$. In (C) and (D) the DNA solutions (10– 20 μM) also contained 25 mM $Na_2SO_4$ for use of the 981 $cm^{-1}$ band of $SO_4^{2-}$ as a reference intensity and wavenumber standard. The inset at the upper left in (C) and (D) show a 5-fold ordinate amplification of the 600–900 $cm^{-1}$ interval. Taken from reference *[Krafft et al., 2002]*.

Raman spectroscopy has thus far been rarely used to study guanine quadruplexes **[Abu-Ghazalah et al., 2010; Abu-Ghazalah et al., 2012; Breuzard et al., 2003; Krafft et al., 2002; Laporte & Thomas, 1998; Miura et al., 1995; Miura & Thomas, 1994; Miura & Thomas, 1995; Poon & Macgregor, 1999; Wei et al., 2006]**. To our knowledge, only one Raman study of the I-motif DNA structure was reported **[Benevides et al., 1996]**. There are several bands in Raman spectra of guanine quadruplexes that help to discriminate between their parallel and antiparallel conformation (for details, see supplementary text and supplementary table S1 in our paper: **[Palacký et al., 2012, submitted, in attachments, pages 195-220]**). The most important Raman lines include the markers of C2'-*endo/syn* (671 ± 2 $cm^{-1}$, 1326 ± 2

$cm^{-1}$) and C2'-*endo/anti* (686 ± 2 $cm^{-1}$, 1338 ± 2 $cm^{-1}$) conformation of guanosines **[Krafft et al., 2002]**. UVRR spectra of telomeric sequences have a characteristic band around 1603 $cm^{-1}$, which is assigned to Hoogsteen hydrogen bonding (N1-H...O6) in the quadruplex tetrad **[Krafft et al., 2002; Miura et al., 1995]**. The corresponding band for non-telomeric sequences is shifted to much lower frequencies (1596 $cm^{-1}$) **[Krafft et al., 2002]**. Hydrogen/deuterium and deuterium/hydrogen exchange of the amino and imino protons in the guanine tetrads of telomeric quadruplexes have been monitored by Raman spectroscopy **[Krafft et al., 2002; Miura et al., 1995]**. The rate of this exchange is very slow, which indicates the extremely high stability of quadruplex structures. Raman spectra of guanine quadruplexes are very sensitive to ionic strength and differ in $Na^+$ and $K^+$ solutions **[Laporte & Thomas, 1998; Miura et al., 1995; Miura & Thomas, 1994]**. Bimolecular and tetramolecular guanine quadruplexes or G-rich assemblies of high order (G-wires and frayed wires) usually predominate at high DNA concentrations typical of non-resonant Raman spectroscopy **[Abu-Ghazalah et al., 2010]**. Intramolecular quadruplexes, found preferentially at much lower concentrations, can be studied by UVRR spectroscopy **[Krafft et al., 2002; Wei et al., 2006]**.

There are basically three different ways how to study the interaction of guanine quadruplexes with other biomolecules. It can be done either by conventional non-resonant Raman spectroscopy **[Laporte & Thomas, 1998]**, resonance Raman spectroscopy **[Krafft et al., 2002; Wei et al., 2006]** or **S**urface **e**nhanced **R**aman **s**cattering (SERS) spectroscopy **[Breuzard et al., 2003; Wei et al., 2006]**. Non-resonant Raman spectroscopy uses an excitation wavelength, which is far from the absorbance maxima of interacting species. On the other hand, resonance Raman spectroscopy uses an excitation wavelength matching the absorbance maxima of either DNA or DNA ligands. While resonantly enhanced spectra of DNA reflect its response to ligand binding, resonantly enhanced spectra of DNA ligands give information about their structural changes upon binding to DNA. SERS spectroscopy can be used to investigate the interaction of small molecules like porphyrins with quadruplexes. It exploits the fact that Raman intensity of these DNA ligands is enhanced by several orders of magnitude when they come in close proximity of a rough metallic surface of colloid particles. Availability of DNA ligands (and thus the degree of SERS enhancement) to colloid particles depends on the mode of their interaction with quadruplexes.

### *1.6.5 Chemical probing techniques*

There are currently many chemical probing techniques available for the study of guanine quadruplexes. The simplest and most widely employed method is **p**oly**a**crylamide **g**el **e**lectrophoresis (PAGE) (see for example references **[Bauer et al., 2011; Han et al., 2001; Kejnovská et al., 2007; Poniková et al., 2011; Vorlíčková et al., 2005; Yu et al., 2006; Zemánek et al., 2005]**). Native PAGE separates molecules according to their mobilities in polyacrylamide gels under nondenaturating conditions. Generally, the electrophoretic mobilities of oligonucleotides in native PAGE gels are inversely proportional to their length **[Vorlíčková et al., 2005]**, the molecularity **[Vorlíčková et al., 2005]** and compactness **[Payet & Huppert, 2012]** of the electrophoresed DNA structures. One oligonucleotide may adopt more unimolecular **[Payet & Huppert, 2012]**, bimolecular **[Vorlíčková et al., 2005]** or tetramolecular **[Vorlíčková et al., 2005]** conformations differing in the electrophoretic mobilities. The electrophoretic mobility of G-rich oligonucleotides typically follows the order: higher order complexes (multimers or G-wires) >> tetramers >> dimers >> random coils > monomers **[Ambrus et al., 2005; Han et al., 2001; Marsh et al., 1995; Venczel & Sen, 1993; Vorlíčková et al., 2005; Zemánek et al., 2005]** (see figure 25). Single-stranded non-quadruplex conformations may move slower than intermolecular quadruplex structures **[Ambrus et al., 2005]**. Prior to gel loading, oligonucleotide samples are incubated at desired conditions for a certain period of time. However, it is not apparent whether conditions during nondenaturing PAGE retain DNA conformations formed during sample preparation **[Hardin et al., 2001]**. PAGE can be used to study the influence of metal cations **[Venczel & Sen, 1993]**, small molecular ligands **[Yu et al., 2008]** (see figure 25), proteins **[Salas et al., 2006]** or temperature **[Víglaský et al., 2010]** on quadruplex kinetics and stability.

PAGE is usually only an intermediate step in the analysis of nucleic acid samples. The extracted DNA or RNA from the band of interest can be the subject of detailed chemical analyses. G-rich oligonucleotides are most frequently probed by **dim**ethyl **s**ulfate (DMS) protection assay **[Cogoi et al., 2008; Seenisamy et al., 2005]**. This assay exploits the fact that DMS preferentially methylates guanines at $N_7$ atoms. The methylation preference of DMS for nucleic acid bases follows the order $G(N_7)>A(N_1)>C(N_3)$ **[Christiansen et al., 1994]**. Nucleic acid chains can be cleaved

at the positions of methylated guanines by piperidine. Denaturing (sequencing) electrophoresis of the cleaved fragments provides information about which guanines participate in the formation of G-quartets in guanine quadruplexes, because $N_7$ atoms of such guanines are poorly methylated by DMS.

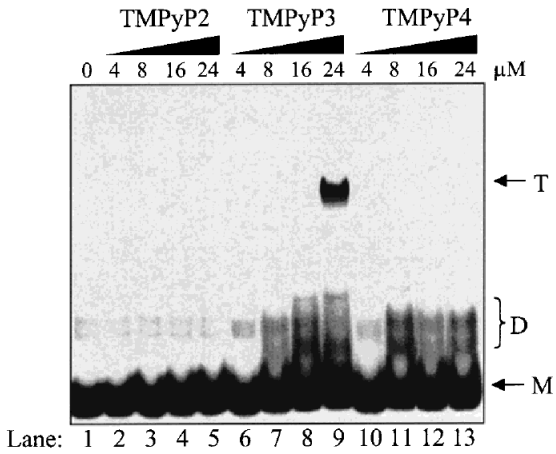



**Figure 25.** Effect of porphyrins on the facilitation of parallel G-quadruplex assembly illustrated by native PAGE. Porphyrins were titrated against $d(TTAGGG)_4$ in TE buffer containing 100 mM KCl (lane 1, control; lanes 2-5, TMPyP2; lanes 6-9, TMPyP3; lanes 10-13, TMPyP4). Major bands identified are monomer (M), dimer (D), and tetramer (T). Taken from reference ***[Han et al., 2001]***.

In addition to DMS, other chemicals can be used to probe quadruplex structure. KMnO4 **[Balagurumoorthy & Brahmachari, 1994; Usdin, 1998]** and $OsO_4$ **[Murchie & Lilley, 1992]** preferentially react with unpaired thymines such as those frequently found in loops of guanine quadruplexes. Diethylpyrocarbonate (DEPC) reacts particularly with purines (A>>G) at N-7 that are exposed to the solvent, either in single strands or Z-DNA conformation **[Balagurumoorthy & Brahmachari, 1994; Usdin, 1998]**. Consequently, DEPC can be used to determine whether N7 atoms of guanosines participate in Hoogsteen hydrogen pairing within guanine tetrads or whether they are unpaired and form a part of quadruplex loop regions. A combination of several nucleases with different specificities reveals even more about the secondary structure of G-rich nucleic acids **[Christiansen et al., 1994; Poon & Macgregor, 1999]**. Electrophoretic methods are also suitable for analysis of some quadruplex-ligand complexes. Some molecules that bind to guanine quadruplexes

such as porphyrins are photoactive. These molecules produce breaks in nucleic acid chains upon irradiation by intense ultraviolet light. Denaturing gel electrophoresis of the resulting cut fragments reveals the location of binding sites for a given G-quadruplex binding molecule **[Han et al., 2001; Vialas et al., 2000]**. Quadruplex-forming DNA sequences can be identified by a popular technique known as polymerase stop assay **[Duan et al., 2001; Weitzmann et al., 1996]**. The principle of this technique is apparent from the figure 26. Polymerase stop assay can be accomplished under physiological conditions and can differentiate between multiple quadruplexes in their mixture at concentrations far lower than those required for CD measurements.

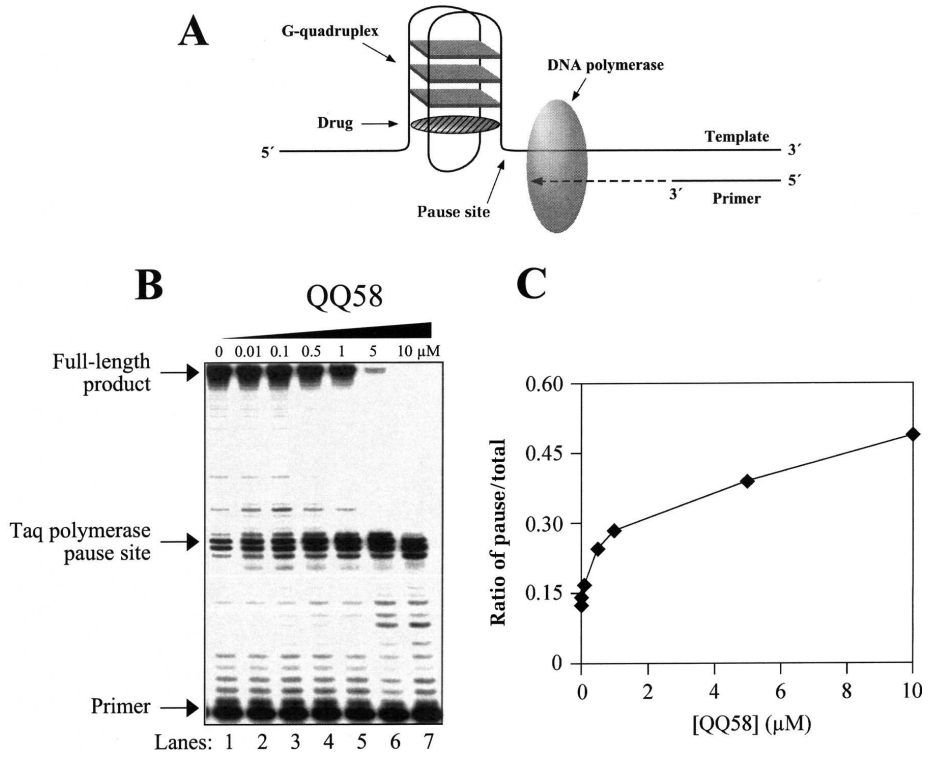



**Figure 26.** Inhibition of Taq polymerase with increasing concentrations of QQ58 (for the structure, see reference *[Duan et al., 2001]*). (**A**) Cartoon of the assay. (**B**) Autoradiogram of the sequencing gel showing enhanced DNA synthesis pausing at the G-quadruplex site with increasing concentrations of QQ58 (Lanes 1–7). The free primer, the pause site, and the full-length product are indicated. (**C**) Graphical representation of the quantification of the sequencing gel shown in B, showing the concentration of QQ58 to the ratio of intensity of the bands obtained for the pausing site/total intensity per lane. Taken from reference *[Duan et al., 2001]*.

### *1.6.6 Structural techniques: NMR and X-ray*

NMR spectroscopy **[Adrian et al., 2012; Gallagher, 2004; Keniry, 2001]** and X-ray spectroscopy **[Campbell & Parkinson et al., 2007; Haider et al., 2011; Neidle & Parkinson, 2008]** can be used to study the structure and dynamics of guanine quadruplexes and their complexes with other molecules. The common and major disadvantage of both techniques is that they are only applicable to samples at high concentrations (compared to UV-vis, CD or resonance Raman spectroscopy). NMR spectroscopy is more frequently used to study guanine quadruplexes than X-ray spectroscopy despite it is less accurate. It is because crystal structures may not fully correspond to solution structures and due to possible problems with crystallization. One dimensional $^{1}$H NMR spectra can differentiate between various nucleic acid conformations (see figure 27).

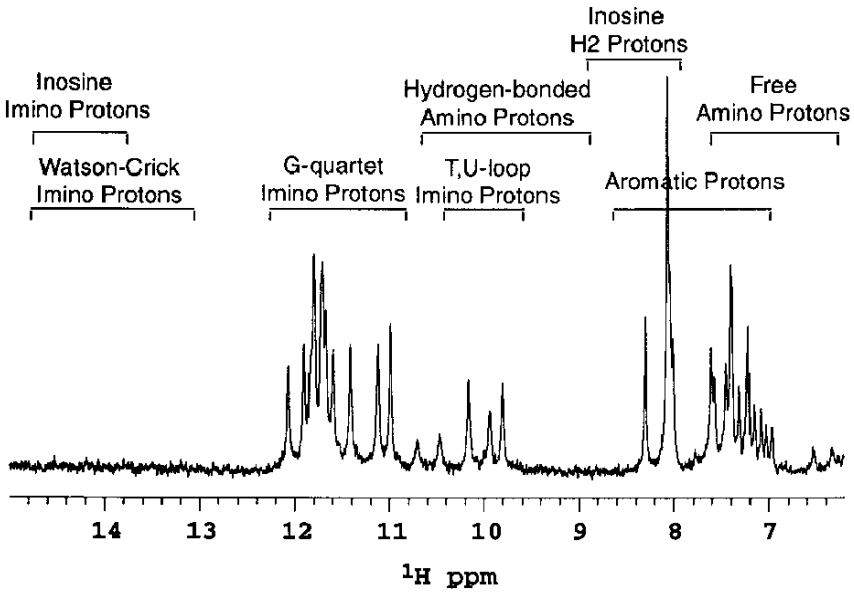



**Figure 27.** One-dimensional $^{1}$H NMR spectrum of the downfield region of $[d(G_3T_4G_3)]_2$ in $H_2O$ buffer. Overlaid on the spectrum are the approximate frequency ranges of the resonances of various protons observed in G-quadruplex structures. Taken from reference *[Keniry, 2001]*.

Hoogsteen hydrogen bonding in guanine tetrads is characterized by imino proton resonances in the range 10.5-12 ppm. $^{1}$H NMR spectroscopy provides information about the thermal stability and kinetics of folding and unfolding of guanine quadruplexes **[Phan & Patel, 2003]** and their interaction with drugs **[Mita**

et al., 2006]. The rate of hydrogen-deuterium exchange is usually very slow for imino protons in the inner guanine tetrads of quadruplexes [Keniry, 2001]. Its monitoring allows to study the folding and unfolding kinetics of guanine quadruplexes [Phan & Mergny, 2002]. Multidimensional NMR techniques together with molecular modeling studies have allowed the determination of the structures of many DNA (see for example references [Ambrus et al., 2005; Ambrus et al., 2006; Kuryavyi et al., 2000; Marusic et al., 2012; Phan et al., 2005; Schultze et al., 1994]) and some RNA [Cheong & Moore, 1992; Martadinata & Phan, 2009] guanine quadruplexes as well as the structures of several quadruplex-ligand complexes [Gavathiotis et al., 2003; Kern et al., 2002].

### *1.6.7 Microcalorimetry*

**D**ifferential **s**canning **c**alorimetry (DSC) and **i**sothermal **t**itration **c**alorimetry (ITC) are useful methods for studying biomolecular interactions and conformational changes of biomolecules [Jelesarov & Bosshard, 1999].

DSC is primarily used to study the thermal stabilities of guanine quadruplexes [Chaires, 2010; Lane et al., 2008]. DSC thermal profile is the dependence of the excess heat capacity on a linear rate of temperature change. It can be used to determine directly thermodynamic parameters for the transition between the folded and denaturated quadruplex structures without relying on a specific thermodynamic model. The calorimetric enthalpy is calculated as the area under DSC transition band [Bruylants et al., 2005; Chaires, 2008] (see figure 28b). It includes enthalpic contributions from all endothermic and exothermic processes (such as the disruption of hydrogen bonds) accompanying the studied transition [Bruylants et al., 2005]. The transition enthalpies can also be determined from the van't Hoff equation using the temperature dependence of equilibrium constants obtained by some of the spectroscopic methods (in this case, the transition enthalpy is often denoted as van't Hoff enthalpy). Such approach is model-dependent since the relations for calculating the equilibrium constants depend on the suggested reaction mechanism. Model-dependent enthalpy can also be obtained from DSC thermograms by fitting a selected model to them [Lane et al., 2008]. A discrepancy between the model-dependent spectroscopic (van't Hoff) or calorimetric enthalpies and model-independent enthalpies (obtained from the integration of DSC peaks) is frequently observed [Olsen et al., 2006]. There are several explanations for this. Primarily, the heat

capacity change $\Delta C_p$ (see figure 28b) between the folded and unfolded structures is often neglected, which introduces non-linearity into conventional van't Hoff analysis **[Chaires, 2008; Lane et al., 2008]**. In fact, a positive heat capacity change was observed for thermal transitions of some G-quadruplex structures **[Bončina et al., 2012; Majhi et al., 2007]**. To account for $\Delta C_p \neq 0$, the van't Hoff enthalpy and entropy terms are replaced by the more complicated relations taking into account the temperature dependence of enthalpy and entropy changes **[Bruylants et al., 2005; Chaires, 2008]**. The nonzero heat capacity change (see figure 28b) also complicates pretreatment of DSC thermograms consisting in the correction of baseline **[Lane et al., 2008]**. A large difference between calorimetric and van't Hoff thermodynamic parameters is frequently associated with the invalidity of a simple two-state model, which means that thermal transition from native to denaturated state involves multiply intermediates **[Gray et al., 2012]**.

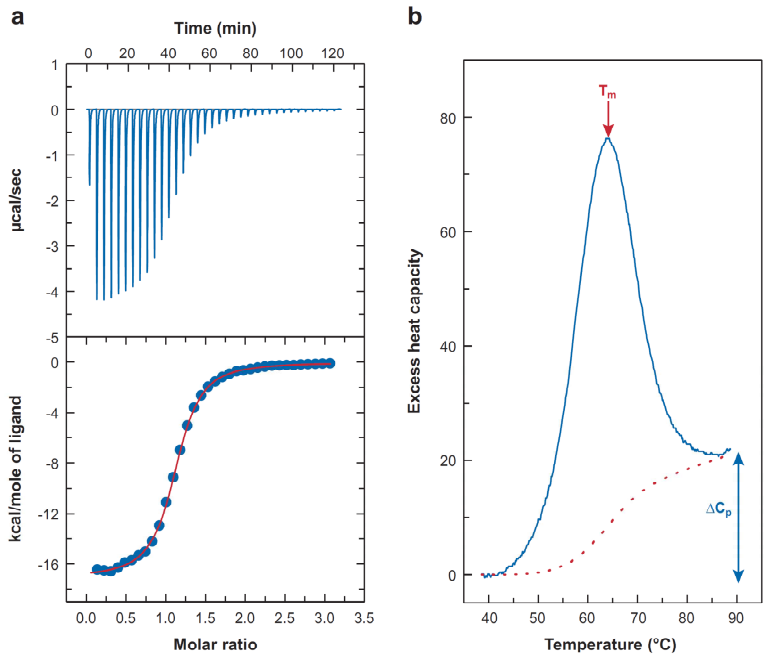



**Figure 28.** Representative ITC (**a**) or DSC (**b**) data. The ITC data are for the binding of 2´CMP to RNaseA. The top panel shows the primary titration data, and the lower panel shows the binding isotherm constructed from the primary data. The DSC transition shown is for the thermal denaturation of the protein ubiquitin.The diagram indicates the melting transition temperature $T_m$ and the heat capacity change ΔCp. Taken from ***[Chaires, 2008]***.

Isothermal titration calorimetry (ITC) is a calorimetric technique very suitable for the characterization thermodynamic properties of the interaction of guanine quadruplexes with drugs **[Pagano et al., 2009]**. ITC has the definite advantage over more frequently used fluorescence techniques that no labeling of interacting biomolecules is required. Another benefit of ITC is that, similarly to DSC, no thermodynamic model describing the interaction of a ligand with its binding partner is necessary. In a typical ITC experiment, the solution of one larger molecule (G-quadruplex, protein, lipid, etc.) is kept at constant temperature and titrated at regular intervals with a smaller ligand of interest. In each titration step, the precisely known (and the same) small volume of the ligand solution is added to the sample solution. The result of ITC measurement is a series of peaks, each of them corresponding to a heat change reflecting ligand binding to the substrate (see figure 28a). The raw ITC data are then converted into binding isotherm, which is plot of integrated heat for each ITC peak versus the molar ratio of ligand to substrate (see figure 28a). Binding isotherm is usually sigmoidal curve and its midpoint corresponds to the stoichiometry of the studied complex. ITC was used to determine the stoichiometry of several G-quadruplex complexes with drugs **[Bhadra & Kumar, 2011; Martino et al., 2009]** and various metal cations **[Kankia & Marky, 2001; Reshetnikov et al., 2011]**. In addition to stoichiometry, ITC can also provide binding constants and reaction enthalpies of bimolecular processes. Last but not least, the temperature dependence of reaction enthalpy can be used to calculate the heat capacity change corresponding to ligand binding on the receptor molecule **[Majhi et al., 2007]**.

### *1.6.8 Other techniques*

Besides the methods mentioned above, a wide variety of other methods have been applied to study G-rich nucleic acids (see table 6 for their basic description). These include surface plasmon resonance (SPR) **[del Toro et al., 2008; Dixon et al., 2005; Rezler et al., 2005; Zhao et al., 2004]**, scanning probe microscopy (SPM) **[Basnar et al., 2006; Neaves et al., 2009; Vesenka et al., 2007]**, electron microscopy **[Duquette et al., 2004; Li & Mirkin, 2005]**, analytical ultracentrifugation **[Li et al., 2005; Petraccone et al., 2010]**, photon correlation spectroscopy **[Wlodarczyk et al., 2005]**, mass spectroscopy **[Collie et al., 2010; Rosu et al., 2008; Vairamani & Gross, 2003; Wyttenbach & Bowers, 2007]**, electron paramagnetic resonance (EPR) **[Evans et al., 2007; Keating & Szalai,**

**2004; Singh et al., 2009]**, vibrational circular dichroism **[Andrushchenko et al., 2011; Nový et al., 2008; Petrovic & Polavarapu, 2008; Setnička et al., 2008; Urbanová, 2009]**, Fourier transform infrared spectroscopy (FTIR) **[Guzmán et al., 2006; Mondragon-Sanchez et al., 2004]** and molecular modeling approaches **[Haider et al., 2003; Heddi & Phan, 2011; Reshetnikov et al., 2011; Šponer et al., 2012]**.

| Method | Main usage |
|---|---|
| Surface plasmon resonance | Study of (un)folding of guanine quadruplexes and their interaction with other molecules. |
| Scanning probe microscopy | Observation of large G-rich aggregates such as G-wires. |
| Electron microscopy | Observation of G-loops, T-loops or G-rich aggregates. |
| Analytical ultracentrifugation | Estimation of the stoichiometry, hydrodynamic shapes and molecular weights of guanine quadruplexes. Precipitation of G-rich aggregates. |
| Photon correlation spectroscopy | Estimation of the stoichiometry and size of guanine quadruplexes |
| Mass spectroscopy | Separation of quadruplexes according to their mass and charge. Suitable for studying quadruplex-ligand complexes |
| Electron paragmagnetic resonance | Study of the interaction of quadruplexes with paramagnetic ions or molecules having unpaired electron spins such as some metalloporphyrines |
| Molecular modeling approaches | Refinement of the structures of guanine quadruplexes obtained by NMR or X-ray spectroscopy. The kinetics of formation and stability of guanine quadruplexes and their interaction with ligands. |

**Table 6.** An overview of methods used for studying guanine quadruplexes

# 2. Materials and methods

The characterization of nucleic acids properties in solution has received great attention at the Division of Biomolecular Physics at Institute of Physics of Charles University in Prague. A considerable part of research there is focused on the investigation of short modified DNA oligonucleotides, particularly with regard to their therapeutic potential as anti-proliferative agents **[Hanuš et al., 2004; Němeček et al., 2005]**, the interaction of nucleic acids complexes with porphyrins **[Mojzeš et al., 2003]**, proteins **[Kříž et al., 2012]** and other molecules **[Kočišová et al., 2010]**, and the estimation of thermal stabilities of short oligonucleotides in the presence of various metal cations **[Herrera & Štěpánek, 2010; Ottová et al., 2011; Vachoušek & Štěpánek, 2008]**. The main experimental technique at the Division of Biomolecular Physics includes, but it is not limited to, Raman spectroscopy. My first experience with the use of Raman spectroscopy date back to 2003-2005, when I was working on my diploma thesis dealing with the analysis of thermal stability and conformational transitions of DNA duplexes and triplexes arising from the hybridization of natural (poly(rA), poly(rU)) as well as modified DNA sequences **[Palacký, 2005]**. A range of the experimental and theoretical skills acquired during master study were also utilized during doctoral study.

The aim of the thesis is to characterize non-canonical DNA structures and their interaction with cationic porphyrins. The vast majority of the work has been devoted to the study of an interesting subclass of non-canonical structures, G-quadruplexes. Raman scattering, employed as the major experimental technique, was complemented with other generally known methods (circular dichroism, absorption spectroscopy, gel electrophoresis and microcalorimetry). Since the dissertation does not deal with methodological development, only the basic experimental set-up for each method is discussed (a short review of the applicability and the specifics of various methods for studying G-quadruplexes is given in the section 1.6). As much attention was paid to the development and application of advanced methods of data processing, a short introduction to SVD algorithm and its practical applications is given (see the section 2.2). First of all, the procedures essential or appropriate for the preparation of the samples of G-quadruplexes are briefly discussed.

## 2.1 Sample preparation

As it has been mentioned in the introduction, the structure of G-quadruplexes as well as their thermostability is strongly dependent on the nature of stabilizing cations. We therefore paid great attention to the preparation of buffers. Almost all experiments were conducted in potassium or sodium cacodylate or phosphate buffers.

We preferred phosphate buffer over cacodylic buffer for many optical experiments since cacodylate buffer is too toxic for routine use. For calorimetry and gel electrophoresis where a large volume of buffer is needed, we used exclusively phosphate buffer. A lot of our experiments were done on telomeric DNA using Raman spectroscopy where we were primarily interested in the structural transition between the *syn* and *anti* conformations of guanines. Unlike the cacodylate buffer, phosphate buffer was suitable for that since its Raman bands did not overlap the markers of glycosidic torsion angles.

Lithium cacodylate buffer is highly recommended for circular dichroism and ultraviolet absorption thermal experiments with G-quadruplexes **[Mergny & Lacroix, 2009]**. It has a negligible dependence of pH on temperature (unlike, for example TRIS or MES buffers) and its high toxicity prevents the growth of microorganism. For absorption, CD and Raman measurements at low nucleoside concentrations (< 20 mM), we used 20 mM lithium cacodylate buffers (prepared from pure cacodylic acid rather than the commonly available sodium cacodylate) with added 1-100 mM LiCl, NaCl or KCl. The choice of lithium cacodylate buffer is very suitable for the experiments intended to study the influence of stabilizing metal cations ($K^+$ or $Na^+$) on the G-quadruplex structure as a function of their concentration since it is established that $Li^+$ has little or no impact on the formation and stability of G-quadruplexes **[Neidle & Balasubramanian, 2006]**.

The HPLC-purified and lyophilized oligonucleotides with desired sequences were synthesized on demand by VBC-Biotech GmbH (Vienna, Austria), Generi-Biotech s.r.o. (Hradec Králové, Czech Republic), the Laboratory of Plant Molecular Physiology at the Faculty of Science of Masaryk University (Brno, Czech Republic) and Eurogentec (Seraing, Belgium). Cationic porphyrins were purchased from Frontier Scientific Inc. (Logan, Utah, USA) and TriPorTech GmbH (Lübeck, Germany).

The concentrations of DNA samples and porphyrins were determined spectrophotometrically using the appropriate extinction coefficients as specified in the original papers. The mixtures of DNA with porphyrins were always prepared immediately before measurement by mixing annealed DNA samples with porphyrins at a desired molar ratio. Stock solutions of both DNA and porphyrins were usually prepared in deionized water and kept at low temperature (0-5°C). A desired concentration of DNA or porphyrin was adjusted by evaporating water using rotary vacuum concentrator and subsequent adding appropriate amount of buffer.

The Raman study of conformational transitions of telomeric sequences induced by high DNA concentrations (see the section 3.2) had to be carried out at precise concentrations of either $Na^+$ or $K^+$. However, oligonucleotide samples at high nucleoside concentrations usually contain significant amount of counterions originating from DNA synthesis. The preparation of DNA samples for Raman measurements therefore required special treatment. First, stock solutions of oligonucleotides at high nucleoside concentrations (approximately 200 - 500 mM) were prepared by dissolving HPLC-purified, lyophilized oligonucleotide into deionized water. Removal or exchange of counterions was then achieved by repetitive (typically 3-5 times) washing out the DNA stock solution by desired buffer (with a given concentration of $Na^+$ or $K^+$ ions) or water using centrifugal filter devices (Amicon Ultra 3K, Millipore). DNA concentration in the concentrate was determined simply on the basis of knowledge of initial and starting centrifugal volumes, and verified by absorption spectroscopy.

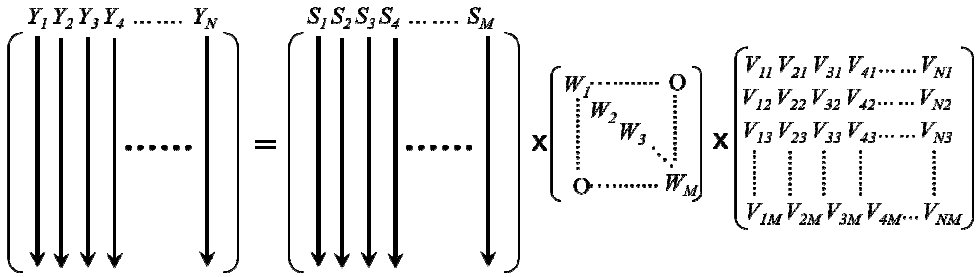
## 2.2 Data treatment

### *2.2.1 Singular value decomposition (SVD)*

Large datasets of absorption, Raman or CD spectra recorded as a function of one specific parameter such as temperature, nucleoside concentration or ionic strength were processed using singular value decomposition (SVD) algorithm. SVD approach facilitates the identification of subtle spectral changes that are normally hardly visible to the naked eye. It allowed the presentation of huge amount of data in a compact form and the separation of various competing processes in the sample. We used SVD algorithm for various purposes such as data visualization, baseline correction or noise removal.

Suppose now that the spectra having the common *x*-scale (corresponding, for instance, to absorbance or Raman shift values) are expressed as a matrix $\boldsymbol{Y}$ with $K$ rows (a number of data points comprising the x-scale) and $N$ columns (a number of spectra). The matrix component $Y_{kn} = Y_n(k)$ $(k \leq K;\ n \leq N)$ is the *k-th* data value for the *n-th* spectrum. Spectra $\boldsymbol{Y_i}$ $(i=1...N)$ are hence represented as columns of the data matrix $\boldsymbol{Y}$. The result of SVD procedure applied to the matrix $\boldsymbol{Y}$ can be written for our purposes as a product of the three matrices $\boldsymbol{Y}=\boldsymbol{SWV}^T$ (see relation (2.1)), where $\boldsymbol{W}$ is $M \times M$ diagonal matrix with singular values $W_j > 0$ $(j = 1...M)$ arranged in descending order, and $\boldsymbol{S}$, $\boldsymbol{V}^T$ (transpose of the matrix $\boldsymbol{V}$) are $K \times M$ and $M \times N$ semi-orthonormal matrices, respectively. The columns $\boldsymbol{S_j}$ $(j=1...M)$ of the matrix $\boldsymbol{S}$ are orthonormal and constitute the kind of artificial spectra, to which we henceforth refer as subspectra. Analogically to the data matrix $\boldsymbol{Y}$, the matrix component $S_{km} = S_m(k)$ $(k \leq K;\ m \leq M)$ of the matrix $\boldsymbol{S}$ denotes the *k-th* data value for the *m-th* subspectrum. Finally, the columns $\boldsymbol{V_j}$ $(j=1...N)$ of the matrix $\boldsymbol{V}$ (i.e. the rows of the transpose matrix $\boldsymbol{V}^T$) are mutually orthonormal and we denote them from now on simply as coefficients.

(2.1)



The SVD analysis of the data matrix $\boldsymbol{Y}$ can be now clearly interpreted. According to the equation (2.1), each experimental spectrum $\boldsymbol{Y_i}$ $(i=1...N)$ can be written as a linear combination of $M$ subspectra $\boldsymbol{S_j}$ $(j=1...M)$ weighted by the corresponding singular values $W_j$ $(j=1...M)$ and multiplied further by the values $V_{ij}$ (the *i-th* value for the coefficient $\boldsymbol{V_j}$) that represents the relative spectral contribution of the subspectrum $\boldsymbol{S_j}$ $(j=1...M)$ to the spectrum $\boldsymbol{Y_i}$ $(i=1...N)$.

The total number of components $M$ provided by SVD (the number of subspectra) is equal to the rank of data matrix $\boldsymbol{Y}$. It can be proved that the best approximation of the original spectra $\boldsymbol{Y_i}$ $(i=1...N)$ through the linear combination of $L$ spectral profiles $(L \leq M)$ is expressed as a sum of $L$ subspectra according to the relation (2.1) **[Malinowski, 2002]**. The matrix multiplication is in this case restricted to the $K \times L$, $L \times N$ and $L \times L$ submatrices of the full matrices $\boldsymbol{S}$, $\boldsymbol{V}^T$ and $\boldsymbol{W}$, respectively.

The factor dimension is the smallest possible number of subspectra that have to be linearly combined so that the original spectra are reproduced within the noise level. The trick of SVD is that the factor dimension is often significantly smaller than the rank of the data matrix $\boldsymbol{Y}$. In other words, an arbitrary large group of spectra can be written as a linear combination of just a few relevant subspectra. The noise is projected into higher, less statistically relevant, subspectra. SVD can be therefore employed as a useful means of data reduction (without significant loss of information), especially in the case of large spectral sets. The factor dimension can be estimated from the dependence of singular values $W_j$ versus the corresponding order $j$. It corresponds to the smallest possible value $j$ (usually below 0.5% of the maximal value) above which singular values decreases linearly with the increasing order $j$ (see figure 29) **[Malinowski, 2002]**.

The factor dimension equals to the number of all spectroscopically observable distinct species. It is, however, not always the same as the expected number of independent sample species. For instance, the number of independent sample species is three for a mixture of two substances that can form one specific complex (suppose that all three pure components are spectrally distinguishable). However, the factor dimension of this system may be greater than three due to the presence of various artifacts such as fluorescence background or wavenumber drift in Raman spectra, or thermal dependence of pure sample species (see the section 2.2.4). Once identified, the artifacts can be eliminated in subsequent experiments. A variety of mathematical approaches can also be applied to eliminate the artifacts from measured spectra.

### *2.2.2 Transformation of abstract SVD results to real world data*

SVD analysis is invaluable for the elucidation of spectral changes (especially small ones) within large spectral sets. The major drawback of the multivariate approach is that the subspectra provided by SVD are generally not real spectra related to pure sample species. Nevertheless, it is possible to express the spectra of pure sample species as a certain linear combination of the subspectra. Finding the coefficients of this linear combination (they are elements of a certain transformation matrix ***R***) is ambiguous without some additional assumptions relating the concentrations of pure sample species to other specific physico-chemical parameters such as thermodynamic parameters *(*$\Delta H$*,* $\Delta S$) or *pK* constants (depending on the studied phenomena). The simplified idea of the procedure for transforming the results of SVD (abstract subspectra) into the physically real spectra (and the determination of the corresponding species concentrations) can be briefly summarized (without going to unnecessary details or equations) in four basic steps:

**(1)** The whole process is iterative. It is a prerequisite to derive theoretical formulas that relate the concentrations of individual sample species in terms of physico-chemical parameters of interest such as thermodynamic parameters $\Delta H$ and $\Delta S$. The first step of the procedure consists in calculating the trial concentration profiles of individual sample species corresponding to some random values of physico-chemical parameters (it is best to work, if possible, with the parameters that are close to the suggested values for speeding the convergence of the iterative process). The trial concentration profiles can be calculated, for example, as a function of experimental temperatures (or pH values) using randomly selected values of thermodynamic parameters $\Delta H$ and $\Delta S$ (or pK constants).

**(2)** Once the trial concentration profiles are known, the next step is simple and straightforward. It consists in finding the transformation matrix ***R*** between the matrix ***C*** of the trial concentration profiles and the experimental matrix ***E*** constructed using the results of SVD analysis of the original data (the construction of the matrix ***E*** is explained hereinafter). Schematically, the searched transformation matrix ***R*** is a least-squares solution to the overdetermined system of equations ***C********R*** = ***E***.

**(3)** Now, the difference matrix is calculated between the matrix ***E*** and the matrix ***C**R*** (the multiplication of the matrix of the trial concentration profiles ***C*** by the trial transformation matrix ***R*** calculated in the step 2).

**(4)** In the last step, the norm **|*E-C**R*|** of the difference matrix from the step 3 is determined. It is simplest and most natural to choose the Euclidean norm of matrix (the square of sum of all matrix elements) for this purpose. The calculated norm of **|*E-C**R*|** is the measure of agreement between tentatively selected values of physico-chemical parameters and their real (searched) values. The steps 1-4 are repeated until the norm is below a given tolerance value. In such case, the trial concentration profiles, related to the optimized values of trial physico-chemical parameters (thermodynamic parameters, pk constants, etc.), should finally converge to the real (searched) concentration profiles, and the transformation matrix ***R*** corresponds to the useful matrix that relates the subspectra and coefficients (provided by SVD) to physically meaningful spectra of pure sample species and the corresponding concentrations.

The construction of the matrix **E** (introduced in the steps 2-4):

The matrix ***E*** is defined as a product of two matrices: $[\boldsymbol{V}^T]^D$ and $[\boldsymbol{W}]^D$ that correspond to the submatrices of the full matrices $\boldsymbol{V}^T$ and ***W*** (provided by SVD of the original experimental data according to relation (2.1)). The submatrices $[\boldsymbol{V}^T]^D$ and $[\boldsymbol{W}]^D$ are constructed in accordance with the expected number $D$ of pure sample species for the studied system. Specifically, the matrix $[\boldsymbol{V}^T]^D$ is $D \times N$ matrix (i.e. it contains only the first $D$ rows of the full matrix $\boldsymbol{V}^T$ or equally the first $D$ columns of the matrix ***V***) and the matrix $[\boldsymbol{W}]^D$ is $D \times D$ diagonal matrix with the first $D$ singular values $W_1, ....W_D$ on its diagonal (arranged in descending order).

The procedure just described allows to determine the values of key physico-chemical parameters that are associated with the concentrations of individual pure sample species. The found transformation matrix ***R*** can be used to calculate the pure spectra $Z_n$ $(n=1…D)$ of all $D$ expected sample species (normalized to unit concentration) as a linear combination of $D$ subspectra $S_j$ $(j = 1...D)$ obtained from SVD of the original data **[Hanuš et al., 1999]**:

(2.2) $$Z_n = \sum_{j=1}^{D} R_{nj} S_j \qquad n = 1 \ldots D$$

The concentrations of individual pure sample components (described by the matrix ***C***) are the solution of the overdetermined system of equations ***C**R** = ***E*** described previously (unlike the iterative step 2 in the procedure above, the transformation matrix ***R*** is now known and the concentration matrix ***C*** is calculated instead). The obtained concentration profiles should agree with the theoretical concentration profiles calculated using the optimized values of physico-chemical parameters.

### *2.2.3 Acid-base titration of sodium cacodylate*

The construction of the spectra of pure components can be demonstrated by a practical real-world example concerning acid-base titration of sodium cacodylate (12 Raman spectra, pH 2.4 – 10.8) using Raman spectroscopy. The first step is the analysis of the Raman spectra using SVD (see figure 29). The result of SVD indicates that the factor dimension of the Raman data is close to 2 (the values of residual error linearly decrease starting at the value of 2 of a tentative factor dimension). The estimated factor dimension of 2 for the acid-base titration of sodium cacodylate suggests the presence of two spectroscopically distinguishable species in the sample. SVD analysis also reveals the component $\boldsymbol{S_3}$ with the statistical weight of only ~1/25 of $\boldsymbol{S_2}$ (see figure 29). At this stage, it is difficult to interpret the component $\boldsymbol{S_3}$ but it may be a consequence of various artifacts due to, for example, small temperature drifts in the sample or the wavenumber scale instability. The subspectrum $\boldsymbol{S_3}$ (or higher $\boldsymbol{S_4}$, $\boldsymbol{S_5}$,...) may also reflect some other competing (but less relevant) phenomena that are not associated with the studied acid-base transition. Further analysis will be based on the assumption that the factor dimension corresponding to the studied acid-base transition is 2.

The factor dimension of 2 evidently corresponds to the equilibrium between the conjugate acid and the conjugate base of cacodylate acid. The point is to determine pure Raman spectra of the conjugate acid and the conjugate base. The transformation between “abstract world” of the artificial subspectra ($\boldsymbol{S_1}$, $\boldsymbol{S_2}$) and “real

world" of the spectra belonging to conjugate acid and conjugate base is done according to the above described procedure (see 2.2.2). For this calculation, theoretical relations for the concentrations of the conjugate acid and conjugate base are the points of departure. The equilibrium equation for deprotonation of cacodylic acid can be written:

(2.3) $$(CH_3)_2AsOOH \leftrightarrow H^+ + (CH_3)_2AsOO^-$$

The dissociation constant of cacodylic acid $K_a$ is written in terms of the concentrations of conjugate acid, conjugate base and solvated hydrogen ions:

(2.4) $$K_a = \frac{[H^+][(CH_3)_2AsOO^-]}{[(CH_3)_2AsOOH]}$$

The total concentration of cacodylic acid (given as a sum of the concentrations of its conjugate acid and conjugate base form) can be normalized to 1 without loss of generality:

(2.5) $$[(CH_3)_2AsOOH] + [(CH_3)_2AsOO^-] = 1$$

Considering equations (2.4) and (2.5), one can finally arrive at equations relating physico-chemical parameters (the dissociation constant $K_a$ of cacodylic acid) with the concentrations of both sample species.

(2.6) $$[(CH_3)_2AsOOH] = \frac{[H^+]}{[H^+] + K_a}$$
$$[(CH_3)_2AsOO^-] = \frac{[K_a]}{[H^+] + K_a}$$

The dissociation constant $K_a$ of cacodylic acid can be preferably expressed in terms of $pK_a$ constant:

(2.7) $$K_a = 10^{-pK_a}$$

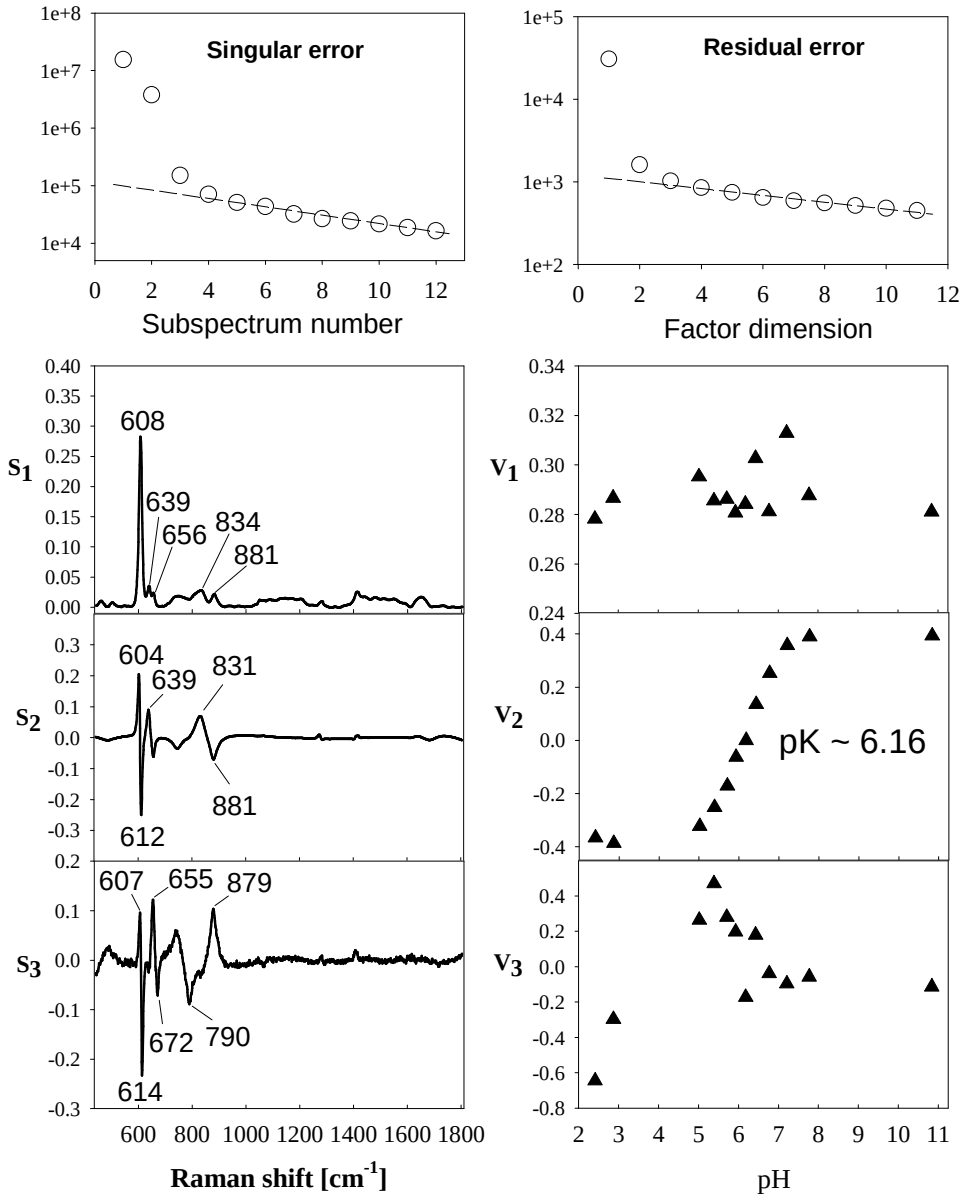



**Figure 29.** SVD analysis of a set of twelve pH-dependent Raman spectra of sodium cacodylate measured at 25°C. The plot of singular values (residual error) versus tentative factor dimension indicates the presence of two spectroscopically distinguishable species in the Raman spectra. The singular value for the subspectrum $S_3$ is only ~1/25 of that for the subspectrum $S_2$. The component $S_1$ is a kind of average spectrum with the coefficient $V_1$ reflecting predominately integral Raman intensity. The subspectrum $S_2$ describes the greatest change to the average spectrum $S_1$ caused by pH. It clearly reflects the acid-base transition as evident from the sigmoidal shape of the corresponding $V_2$ coefficient profile. The $pK_a$ value of the transition was determined to be 6.16 (see the text for details).

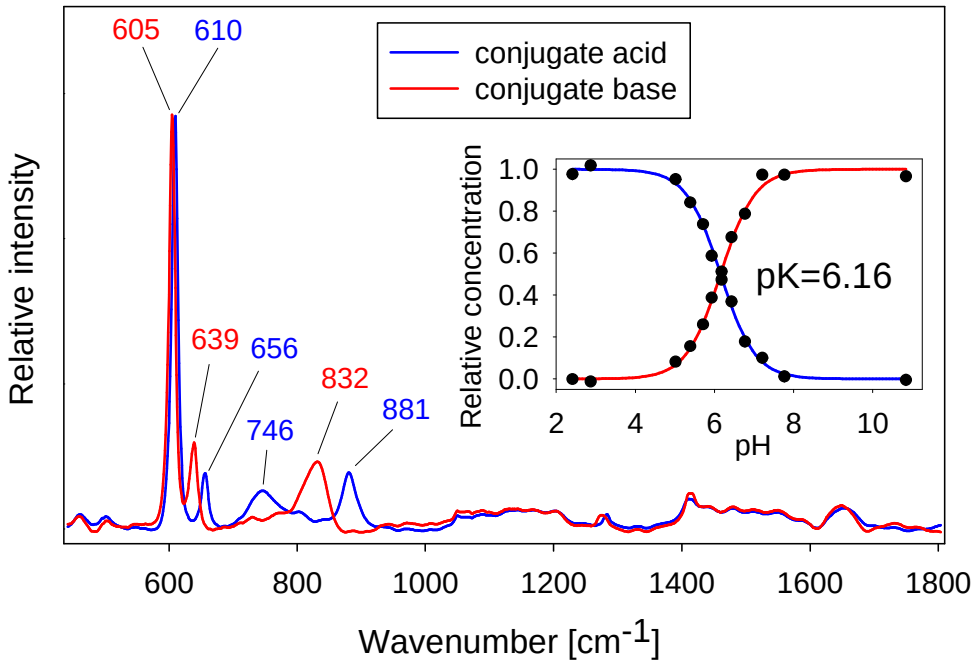



***Figure 30.*** Raman spectra of the conjugate acid and conjugate base of cacodylic acid at 25°C. These spectra are constructed as a linear combination of abstract subspectra obtained from SVD analysis of a set of pH-dependent Raman spectra of sodium cacodylate. The transformation between the artificial subspectra and the spectra of pure sample species is described in the section 2.2.2. Generally, to accomplish this transformation, it is necessary to estimate the factor dimension of data and to propose the theoretical model that predicts the concentrations of individual pure components. In this particular case the factor dimension is 2 (see figure 29) and the model is acid-base equilibrium between two conjugate forms of cacodylic acid. Theoretical concentrations of pure components calculated according to relations (2.6) considering the determined *pK* value of the transition (~ 6.16) and the concentrations corresponding to the experimental spectra (calculated using the transformation matrix ***R*** obtained using the procedure described in the section 2.2.2) are shown in inset as line and point graph, respectively.

The concentration of solvated hydrogen ions $H^+$ (more precisely $H_3O^+$) in equations (2.6) can be related to pH:

(2.8) $$[\mathrm{H}^+] = 10^{-pH}$$

The value of pH is a crucial experimental parameter that must be known for each spectrum from a set of pH-dependent Raman spectra. The analogical parameter is temperature in thermal melting experiments. Using equations (2.6), (2.7) and (2.8) and applying the procedure described in the section 2.2.2, it is possible to obtain the pure spectra of both components of conjugate acid-base pair (see figure 30). The calculated value of $pK_a$ constant from the procedure is 6.16. This value is in good agreement with the value reported previously **[Shin et al., 1997]** (when comparing the values, it is necessary to take into account that $pK_a$ value of cacodylic acid depends slightly on temperature).

The purpose of the presented analysis of pH-dependent Raman spectra of cacodylic acid was to illustrate the use of multivariate method (SVD) for data processing. The given example is straightforward and the results are easily interpreted. The same approach can be applied to more complicated multiply acid-base transitions associated with more than one dissociation constant $K_a$. In such cases, it is required to derive a set of equations (analogical to those for two protonated forms of cacodylic acid (2.6)) describing the concentrations of different protonated species in the sample. The procedure described in the section 2.2.2 is then applied to the appropriate submatrices (the dimensions of these matrices depend on the expected number of acid-base species) of the full matrices obtained from SVD of the original data.

### *2.2.4 Determination of thermodynamic parameters*

Van't Hoff thermodynamic parameters can be in principle determined from temperature-dependent spectra (absorption, Raman, etc.) according to the procedure described in the section 2.2.2. The procedure provides not only thermodynamic parameters (and the corresponding concentration profiles of pure sample species) but also provides the spectra of individual conformers (monomer, duplex, triplex, etc.). Unfortunately, it is not always applicable due to various reasons.

First, random fluctuations of signal intensity result in corrupted $\boldsymbol{V_1}$ profiles since these describe the biggest variation within spectra, which is usually associated with the oscillation of intensity. The relevant intensity change is usually smaller than random intensity change, especially for Raman spectra. In our case, it concerns inaccurate normalization of Raman spectra or uncontrolled evaporation of water from samples at higher temperatures in the case of absorption measurements.

Although, the addition of sealing mineral oil on the top of the sample can eliminate evaporation, we avoided it because we did not want to contaminate our samples with unknown cations and impurities.

Second, it is often difficult to estimate the number of relevant sample species. In the case of thermal experiments, the factor dimension is frequently greater than the number of expected sample species as a result of temperature dependence of their spectra. The small influence of temperature on the spectra of expected pure components cannot be taken into account in any way in the procedure described in the section 2.2.2. In fact, sample species with the spectra dependent on temperature cannot be exactly considered as pure components. They should be rather described as a linear combination of temperature-independent sub-components. It is difficult (if not impossible) to devise some theoretical model taking into account the number of species that exactly matches the factor dimension.

For the evaluation of thermodynamic parameters (especially in the above-mentioned problematic cases), we used different approach than the procedure described in the section 2.2.2. It is based on the fit of selected thermodynamic models to the coefficient profiles obtained from SVD analysis of experimental data **[Vachoušek & Štěpánek, 2008]**. First of all, the theoretical relations for the concentrations of individual sample species must be derived. The first step in this direction is to write the expressions for equilibrium constants in terms of the concentrations of constituent sample components.

Suppose that there are $N$ different species in the sample (we are interested in nucleic acid samples). Based on the selected thermodynamic model, it is then possible to write $N-1$ equations for equilibrium constants $K_j$ ($j=1...N-1$) and one equation for the total concentration of DNA in terms of the concentrations $[C_S]$ ($S=1...N$) of individual $N$ sample species. Altogether we have system of $N-1$ nonlinear equations for equilibrium constants and one linear equation for the total DNA concentration. The desired equations for the concentrations of individual species are obtained as the solution of parametric system of these $N$ equations with parameters being equilibrium constants. In this way, the concentrations $[C_S]$ ($S=1...N$) of $N$ sample species are expressed in terms of equilibrium constants $K_j$ ($j=1...N-1$) (the process of obtaining the theoretical relations for the concentrations of sample species will be demonstrated on three specific examples thereafter). Each equilibrium constant $K_j$ ($j=1...N-1$) is a function of thermodynamic temperature $T$

and is associated with its own pair of thermodynamic parameters (enthalpy change $\Delta H_j$ and entropy change $\Delta S_j$) according to van't Hoff equation (2.9). $R$ in equation (2.9) is molar gas constant ($R$ ~ 8.314 J/K/mol). Note that $C_{H20}$ is molar concentration of pure water (55.3 mol/L). In most cases, if not all, it represents only a small contribution to the entropy $\Delta S$ and can be neglected. In fact, it serves as the correction factor ensuring the dimensional correctness of Van't Hoff equation (2.9).

(2.9) $$K_j(T) = \frac{1}{C_{H20}} e^{-\frac{\Delta H_j - T\Delta S_j}{RT}}$$

The Van't Hoff thermodynamic parameters ($\Delta H_j$ and $\Delta S_j$; $j$=1...$N$-1) for all thermal transitions are obtained as the solution of the overdetermined system of $M$ nonlinear equations (2.10), each of them (for a given index $i$) corresponding to one experimental spectrum recorded at temperature $T_i$.

(2.10) $$V_{ij} = \sum_{S=1}^{S=N} P_S(T_i) * [C_S(T_i)] \qquad i = 1 \ldots . M$$

The term $V_{ij}$ in relation (2.10) represents the value of $j$-*th* coefficient $\boldsymbol{V_j}$ for the spectrum at temperature $T_i$. The concentrations of sample species $[C_S]$ ($S$=1...$N$) are a function of temperature $T$ and hence also a function of the searched Van't Hoff thermodynamic parameters since these are related to equilibrium constants $K_j$ ($j$=1...$N$-1) according to equation (2.9). Specifically, $[C_S(T_i)]$ is the concentration of the $S$-*th* sample species at temperature $T_i$. The $P_S$ are polynomial expressions in temperature $T$ that are associated with the $S$-*th* sample species ($S$=1...$N$) having the concentration $[C_S]$. Specifically, $P_S(T_i)$ is the value of $P_S$ at temperature $T_i$. We often assumed that all $P_S$ terms are linear in temperature. It is based on the assumption that spectra of "pure" sample species change linearly with temperature. It is justified when the flexibility of macromolecule and its interaction with solvent is not connected with significant changes in the average secondary or tertiary structure **[Němeček et al., 2009; Vachoušek & Štěpánek, 2008]**. Searching for the values of thermodynamic parameters is nothing more than fitting the expression on the right side of equation (2.10) to the thermal profile $\boldsymbol{V_j}$ ($j$-*th* coefficient from SVD).

The best precision of estimating thermodynamic parameters can be achieved by the simultaneous fit of two or more coefficient profiles **[Vachoušek & Štěpánek, 2008]**. To account for different statistical importance of individual $\boldsymbol{V_j}$ coefficient profiles, their values are weighted by a square root of the corresponding singular values $W_j$. This approach is not restricted to the coefficients corresponding to one specific spectral set but it is also applicable for the coefficients belonging to several different spectral sets. For example, it is possible to simultaneously fit a selected model to different coefficients $\boldsymbol{V_j}$ obtained from SVD analysis of absorption, CD and Raman spectra of a given sample. Such complicated fitting may account for different experimental conditions at which the sample is measured (such as DNA concentration) and may refine the values of thermodynamic parameters, which could result ambiguous in cases when coefficients $\boldsymbol{V_j}$ belong to only one spectral set.

The above-mentioned procedure for extracting thermodynamic parameters using SVD of spectroscopic data will be now clarified on three useful examples. These include two simple models that we frequently considered: intramolecular transition and bimolecular transition, and more complex model combining intramolecular and bimolecular transition. Each model is discussed separately below.

*A) Intramolecular model*

Intramolecular model is the simplest model that can be considered. This model is applicable to the conformational transitions occurring within one molecule such as premelting of duplex DNA **[Ottová et al., 2011]**, RNA hairpin opening **[Vachoušek & Štěpánek, 2008]** or the formation of intramolecular G-quadruplex **[Risitano & Fox, 2003]**. Suppose now, for simplicity, that the intramolecular transition proceeds from an unordered single-stranded oligonucleotide *S* to intramolecular G-quadruplex conformation *Q*:

(2.11) $$S \overset{K}{\leftrightarrow} Q$$

Equilibrium constant *K* is given in terms of the concentrations of an unordered single strand *[S]* and an intramolecular G-quadruplex *[Q]*:

(2.12) $$K = \frac{[Q]}{[S]}$$

Total concentration $C_T$ of DNA (in single strands) in the sample is written as:

$$C_T = [S] + [Q] \tag{2.13}$$

By combining equations (2.12) and (2.13), one can obtain relations (2.14) for the concentrations *[S]* and *[Q]* in terms of the equilibrium constant *K* and the total strand concentration $C_T$.

$$[S] = C_T - \frac{C_T * K}{1 + K}$$

$$[Q] = \frac{C_T * K}{1 + K} \tag{2.14}$$

The equilibrium constant *K* can also be expressed by Van't Hoff equation:

$$K(T) = \frac{1}{C_{H20}} e^{-\frac{\Delta H - T\Delta S}{RT}} \tag{2.15}$$

The Van't Hoff thermodynamic parameters $\Delta H$ and $\Delta S$ for the suggested intramolecular transition are acquired by solving the overdetermined system of *M* (*M* is equal to the number of measured spectra) nonlinear equations (2.16).

$$V_{ij} = (A_0 + A_1 T_i) * [Q(T_i)] + (B_0 + B_1 T_i) * [S(T_i)] \qquad i = 1, 2 \dots M \tag{2.16}$$

The system (2.16) is a special case of the system (2.10) that considers only two concentrations (*[S]* and *[Q]*) with the corresponding polynomial terms ($A_0$+$A_1T_i$ and $B_0$+$B_1T_i$) being linear in temperature *T*. Solving the system of nonlinear equations (2.16) is simply searching for the six parameters: four parameters $A_0$, $A_1$, $B_0$, $B_1$ (describing the temperature dependence of the spectra of pure sample species *[Q]* and *[S]*) and desired thermodynamic parameters $\Delta H$, $\Delta S$ (that are related to the concentrations *[Q]* and *[S]* through the relations (2.14) and (2.15)).

*B) Bimolecular model*

Bimolecular model describes the formation of dimeric structures from two molecular species. A typical example is the formation of duplex DNA **[Movileanu et al., 2002]**. Bimolecular model is also applicable to dimerization of guanine quadruplexes **[Collie et al., 2010; Trajkovski et al., 2012]**. Suppose (as an example) that two intramolecular G-quadruplexes $Q$ associate to form one dimeric guanine quadruplex $QQ$ (model neglects the structural reorganization of single strands to form quadruplex):

(2.17) $$Q + Q \overset{K}{\leftrightarrow} QQ$$

Equilibrium constant $K$ is expressed in terms of the concentrations of monomeric $[Q]$ and dimeric species $[QQ]$:

(2.18) $$K = \frac{[QQ]}{[Q]^2}$$

Total concentration $C_T$ of DNA (in single strands) in the sample can be written as:

(2.19) $$C_T = [Q] + 2 * [QQ]$$

By combining equations (2.18) and (2.19), we obtain the concentrations $[Q]$ and $[QQ]$ in terms of the equilibrium constant $K$ and the total strand concentration $C_T$:

(2.20)
$$[Q] = \frac{-1 + \sqrt{1 + 8C_T K}}{4K}$$

$$[QQ] = \frac{1 + 4C_T K - \sqrt{1 + 8C_T K}}{8K}$$

The formation of heteroduplex *AB* from the strand *A* and its complementary strand B is another important example of bimolecular reaction. Suppose a special case that the concentration ratio of two types of single strands is 1/1:

(2.21) $$A + B \overset{K}{\leftrightarrow} AB$$

The total sample concentration $C_T$ is then expressed in terms of the concentration of single strand species of one kind:

(2.22) $$C_T = [A] + [AB] = [B] + [AB]$$

Equilibrium constant *K* is expressed by equation (2.23).

(2.23) $$K = \frac{[AB]}{[A][B]} = \frac{[AB]}{(C_T - [AB])^2}$$

By combining equations (2.22) and (2.23), we get the concentration of single strand species *[A]* and the concentration of duplex *[AB]* in terms of the equilibrium constant *K* and the total sample concentration $C_T$:

(2.24)
$$[A] = [B] = \frac{-1 + \sqrt{1 + 4C_T K}}{2K}$$

$$[AB] = \frac{1 + 2C_T K - \sqrt{1 + 4C_T K}}{2K}$$

The van't Hoff thermodynamic parameters $\Delta H$ and $\Delta S$ corresponding to bimolecular model are determined in the exactly same way as those associated with intramolecular model (the system (2.16)), with expressions (2.14) being replaced by a little more complicated terms (2.20), (2.24) describing homodimer or 1:1 heteroduplex, respectively.

*C) The combination of intramolecular and bimolecular model*

This model takes into account both the intramolecular transition of monomeric species and their subsequent dimerization. We used this complicated model to study dimerization proposed for TBA quadruplex (see the section 3.4). Suppose that two intramolecular quadruplexes *Q* form a dimer *QQ* and that each intramolecular quadruplex Q originates from an unordered single strand *S:*

(2.25)
$$Q + Q \overset{K_1}{\leftrightarrow} QQ$$
$$S \overset{K_2}{\leftrightarrow} Q$$

Total concentration $C_T$ of DNA sample (in single strands) can be written as:

(2.26)
$$C_T = 2[QQ] + [Q] + [S]$$

Equilibrium constant $K_1$ for bimolecular and $K_2$ for unimolecular transition is defined by the following formulas:

(2.27)
$$K_1 = \frac{[QQ]}{[Q]^2} = \frac{[QQ]}{(C_T - 2[QQ] - [S])^2}$$
$$K_2 = \frac{[Q]}{[S]} = \frac{[Q]}{C_T - 2[QQ] - [Q]}$$

The equilibrium constants $K_1$ (for dimerization) and $K_2$ (for unimolecular transition) are associated with their own pair of thermodynamic parameters $\Delta H_1$, $\Delta S_1$ and $\Delta H_2$, $\Delta S_2$, respectively through van't Hoff equations:

(2.28)
$$K_1(T) = \frac{1}{C_{H20}} e^{-\frac{\Delta H_1 - T\Delta S_1}{RT}}$$
$$K_2(T) = \frac{1}{C_{H20}} e^{-\frac{\Delta H_2 - T\Delta S_2}{RT}}$$

By combining equations (2.26) and (2.27), the concentrations *[QQ]*, *[Q]*, *[S]* of all possible sample species can be expressed as a function of equilibrium constants $K_1$, $K_2$ and the total concentration $C_T$:

$$[QQ] = \frac{(1+K_2)^2 + 4C_T K1K2^2 - (1+K2)\sqrt{(1+K_2)^2 + 8C_T K_1 {K_2}^2}}{8K_1 {K_2}^2}$$

(2.29)

$$[Q] = \frac{\sqrt{(1+K_2)^2 + 8C_T K_1 {K_2}^2} - (1+K_2)}{4K_1 K_2}$$

$$[S] = \frac{\sqrt{(1+K_2)^2 + 8C_T K_1 {K_2}^2} - (1+K_2)}{4K_1 {K_2}^2}$$

The Van't Hoff thermodynamic parameters $\Delta H_1$, $\Delta S_1$, $\Delta H_2$, $\Delta S_2$ are obtained as the solution of the overdetermined system of nonlinear equations (2.30) that is analogical to the system (2.16). The only difference between the nonlinear systems (2.16) and (2.30) is that two transitions are considered instead of one (requiring the inclusion of one more linear term $C_0+C_1T_i$), and that relations (2.29) for the concentrations of individual sample species are much more complex than those described by expressions (2.14), (2.20) or (2.24). Solving the system of nonlinear equations (2.30) consists in least square optimizing the 10 parameters: 6 parameters originating from 3 polynomial terms: $A_0$, $A_1$, $B_0$, $B_1$, $C_0$, $C_1$ and 2 pairs of desired thermodynamic parameters $\Delta H_1$, $\Delta S_1$ and $\Delta H_2$, $\Delta S_2$ corresponding to bimolecular and intramolecular transition, respectively.

(2.30)

$$V_{ij} = (A_0 + A_1 T_i) * [Q(T_i)] + (B_0 + B_1 T_i) * [S(T_i)] + \\ +(C_0 + C_1 T_i) * [QQ(T_i)] \\ i = 1, 2 \ldots N$$

The figures 31, 32, 33 and 34 are an illustration of how selected thermodynamic models are fitted to thermal $\boldsymbol{V_j}$ profiles (coefficients from SVD). The coefficient $\boldsymbol{V_1}$ accounts for the greatest variability within data (such as the intensity

of Raman signal, hyperchromic change in absorption spectra upon DNA melting, or the change in elipticity of CD spectra). However, due to adverse effects (the evaporation of water from the sample at higher temperatures, poorly defined intensity of Raman spectra), the coefficient $V_1$ is not always informative. The coefficient $V_1$ well reflects large changes in absorbance upon thermal destabilization of DNA duplexes (figure 33) or big changes of CD spectra upon quadruplex melting (figure 31). It seems that the coefficient $V_2$ (and to a lesser degree $V_3$ or possibly $V_4$) generally best reflects minute changes in Raman and absorption spectra accompanying the thermal disruption of the G-quadruplex structure as evidenced by the sigmoidal shape of the coefficients $V_2$ in figures 32 and 34.

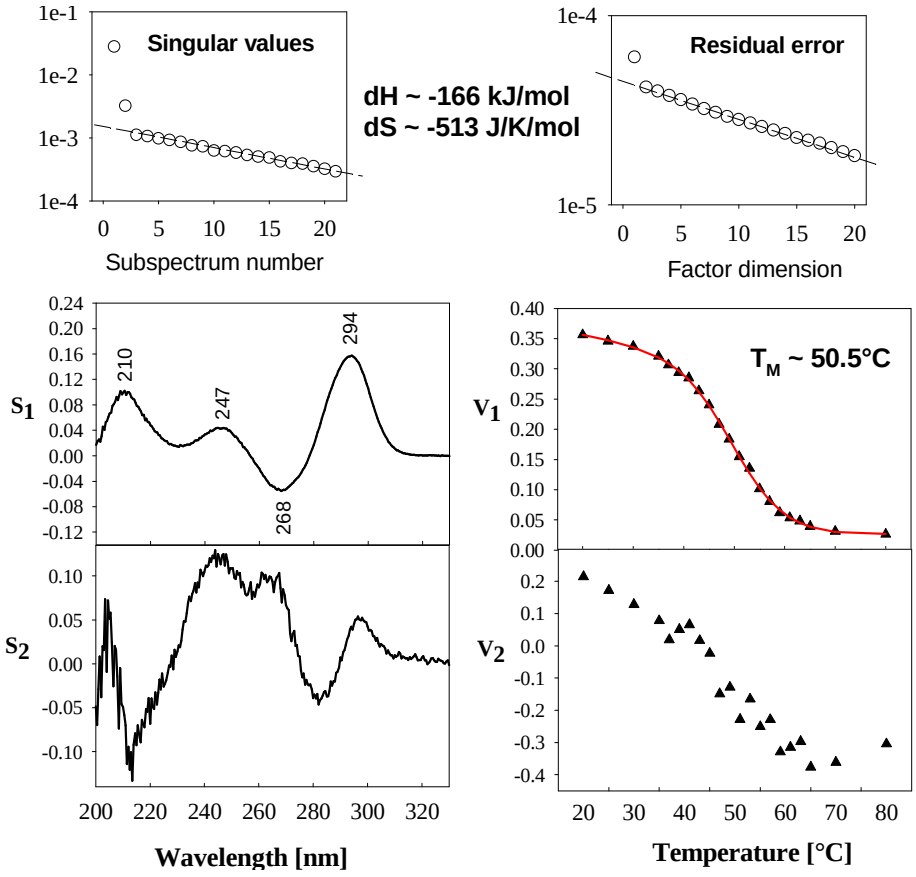



**Figure 31.** The SVD analysis of CD spectra of the TBA quadruplex (the sequence: 5'-d($G_2T_2G_2TGTG_2T_2G_2$)-3') recorded within 20-80°C. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM $K^+$, pH 6.8, 0.87 mM DNA strand concentration. Fit of intramolecular model (see the text for details) to the coefficient $V_1$ is depicted by red line. Thermodynamic parameters from the fit are listed in the upper part of the figure.

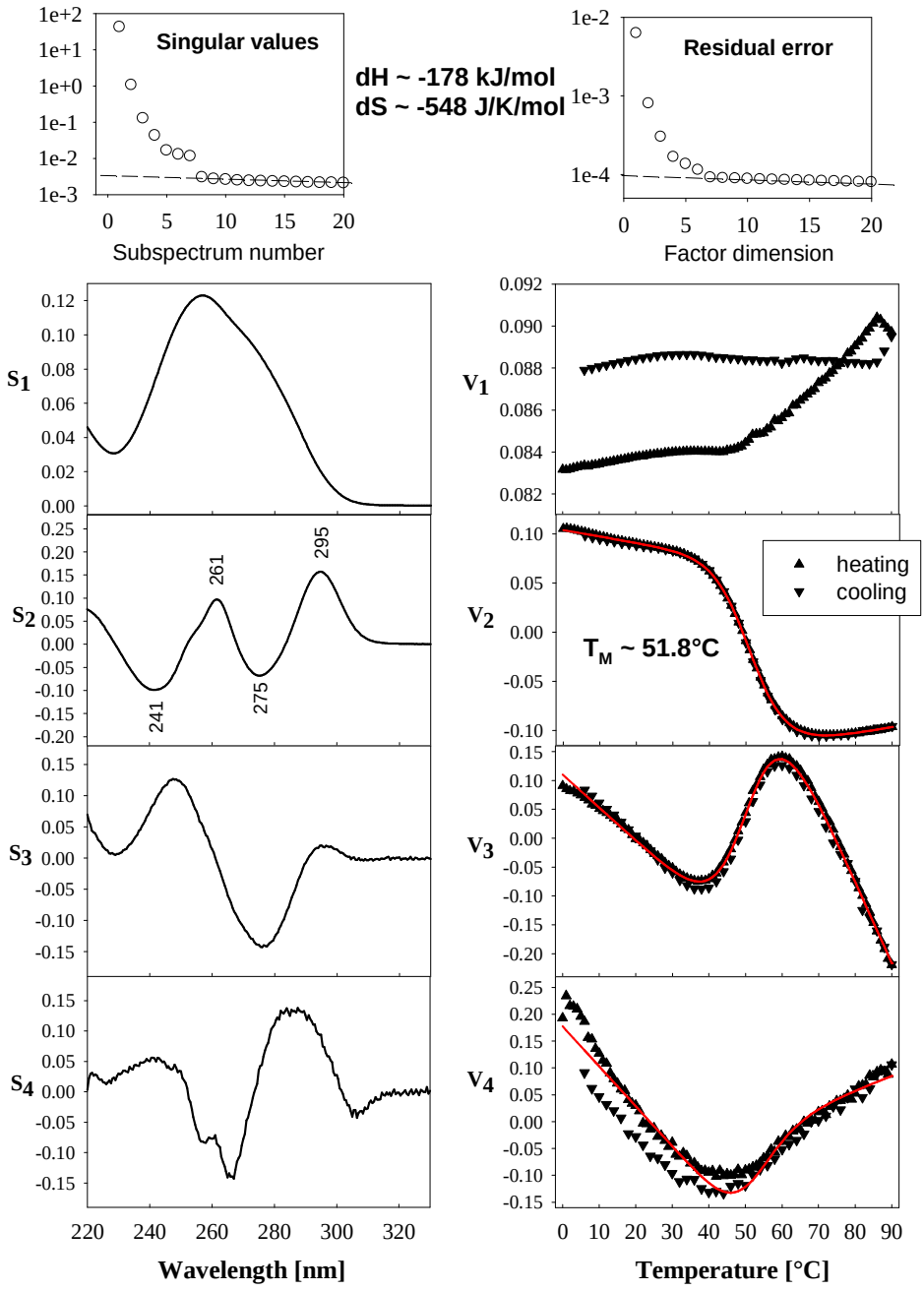



**Figure 32.** The SVD analysis of absorption spectra of the TBA quadruplex (the sequence: 5'-d($G_2T_2G_2TGTG_2T_2G_2$)-3') recorded within 0-90°C. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM $K^+$, pH 6.8, 3.2 μM DNA strand concentration. Joint fit of intramolecular model (see the text for details) to the coefficients $V_2$, $V_3$ and $V_4$ is depicted by red line. Heating and cooling profiles are reversible, which is usually considered as an indication of intramolecular quadruplex folding. Thermodynamic parameters from the fit are listed in the upper part of the figure.

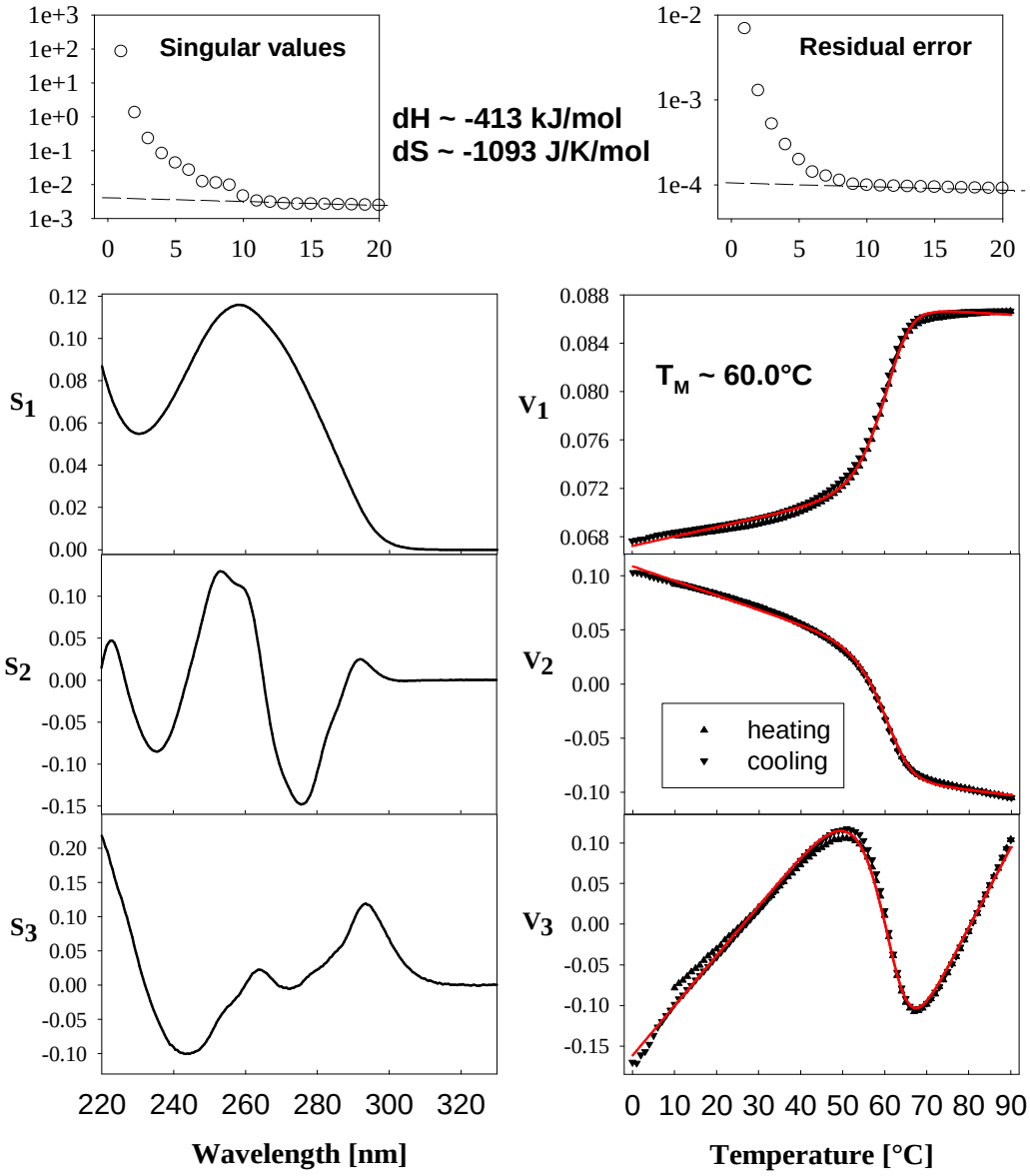



**Figure 33.** The SVD analysis of absorption spectra of 1:1 TBA heteroduplex (TBA sequence 5'-d($G_2T_2G_2TGTG_2T_2G_2$)-3' complexed with an equal amount of complementary C-rich strand) recorded within 0-90°C. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM $K^+$, pH 6.8, 6.0 μM total DNA strand concentration. Joint fit of bimolecular model (the non-autocomplementary case; see the text for details) to the coefficients $\boldsymbol{V_1}$, $\boldsymbol{V_2}$ and $\boldsymbol{V_3}$ is depicted by red line. Note: heating and cooling profiles are reversible indicating that the sample was at thermodynamic equilibrium during the experiment (it is a prerequisite for the determination of thermodynamic parameters from the thermal profiles). Thermodynamic parameters from the fit are listed in the upper part of the figure.

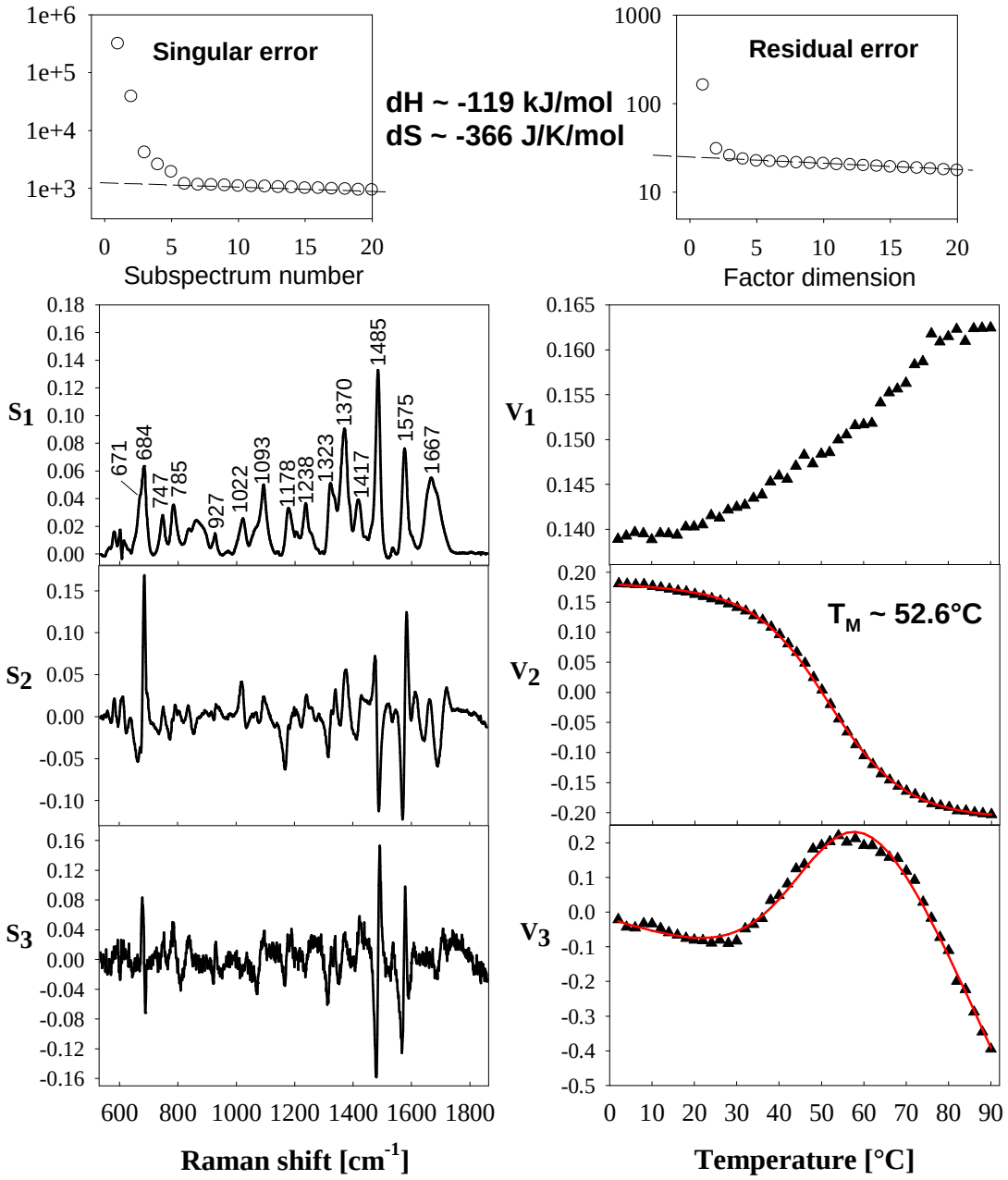



**Figure 34.** The SVD analysis of Raman spectra of TBA (the sequence: 5'-d($G_2T_2G_2TGTG_2T_2G_2$)-3') recorded within 2-90°C. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM $K^+$, pH 6.8, 11.3 mM DNA strand concentration. Joint fit of intramolecular model (see the text for details) to the coefficients $V_2$, $V_3$ is depicted by red line. Thermodynamic parameters from the fit are listed in the upper part of the figure.

## 2.3 Experimental techniques

My doctoral studies were done at the Division of Biomolecular Physics at Institute of Physics of Charles University in Prague. There, Raman spectroscopy was the main technique employed to study G-quadruplexes. At this institute, Raman spectroscopy was also complemented by absorption spectroscopy, microcalorimetry or molecular dynamics simulations (in collaboration with Dr. Barvík). Measurements of absorption thermal melting curves have been conducted in cooperation with the group of prof. Vorlíčková at Institute of Biophysics of the Academy of Sciences of the Czech Republic (Brno). Dr. Kejnovská from this group performed all electrophoretic analyses of G-quadruplex samples. CD measurements were done partly in the group of prof. Vorlíčková and partly in the group of Dr. Bednárová from Institute of Organic Chemistry and Biochemistry of the Academy of Sciences of the Czech Republic (Prague). NMR measurements were conducted by associate prof. Jan Lang, Mgr. Marián Grocký and Dr. Pavel Srb at Department of Low Temperature Physics of Charles University in Prague.

The absorption spectrometer (Varian Cary 4000) at the group of prof. Vorlíčková is equipped with temperature-controlled holder allowing the regulation of temperature within the range 0 – 90°C and placement up to six conventional quartz cuvettes having a path length of 1 cm. This convenient configuration allowed us to accomplish a huge number of experiments (>10000 spectra). Data processing using SVD approach was fast and reliable. Absorption spectroscopy was used not only for the routine determination of sample concentrations but also for the characterization of the formation of G-quadruplexes and their interaction with cationic porphyrins. The difference induced in absorption spectrum by adding $K^+$ to the sample of unstructured G-rich sequence (in water or $Li^+$) has the same pattern as the thermal difference spectra (see the section 1.6.1) **[Mergny et al., 2005b]**. Therefore, we frequently employed this feature as a means of preliminary characterization of the quadruplex samples (to check their quality).

CD spectrometer Jasco-815 (the group of Dr. Bednárová) has a PTC-423S/L Peltier type temperature-controlled system (1 - 90°C) that allowed for stirring the sample in 1cm quartz Hellma cells. CD spectrometer Jobin-Yvon CD6 (the group of prof. Vorlíčková) enabled us to measure temperature-dependent CD spectra in very thin quartz cuvettes having 0.01 cm path lengths. CD spectroscopy was used

primarily to discriminate between the parallel and antiparallel strand topology of guanine quadruplexes. It was also used to study the interaction of guanine quadruplexes with cationic porphyrins through the induced CD signal within the spectral region of their Soret band.

Raman experiments were carried out on a home-made Raman spectrometer. The spectra were recorded in the 90° scattering geometry on a multichannel Raman spectrograph (Jobin Yvon-Spex 270 M) equipped with a holographic notch-plus filter (Kaiser) to reject Rayleigh scattering and a liquid nitrogen-cooled CCD detector (Princeton Instruments). Most of Raman measurements were done in a temperature-controlled, hermetically-sealed quartz microcell (5 μL sample volume). Thermal heating/cooling experiments were conducted in the range of 2 – 96°C, with an accuracy of temperature control within ±0.5°C. Each sample was allowed to equilibrate for at least 5 min before collecting the spectrum. Wavenumber scale of Raman spectra was precisely calibrated using the emission spectra of a neon glow lamp recorded before and after each Raman measurement. The estimated precision of the calibration procedure was better than 0.1 $cm^{-1}$.

The excitation wavelength of 532 nm of a continuous-wave solid-state Nd:$YVO_4$ laser (Verdi 2, Coherent) was used in the non-resonant studies of G-quadruplexes. The radiation power at the sample was adjusted according to sample concentrations within the range 100-400 mW. In the case of DNA concentration-dependent measurements, Raman spectra were normalized to the peak height of the 1093 $cm^{-1}$ band associated with the $PO_2^-$ symmetric stretching mode, reported previously to be largely invariant to the melting of DNA duplexes **[Movileanu et al., 2002]**. If appropriate, e.g. for quantitative comparisons of the oligonucleotides differing in the length or in the number of specific nucleosides, normalized spectra were furthermore renormalized according to the parameter in question.

The interaction of guanine quadruplexes with the cationic porphyrin CuTMPyP4 was studied using excitation of 441.6 nm of a He-Cd laser (Liconix) with the average radiation power at the sample typically about 10 mW. The choice of CuTMPyP4 for many of our experiments was dictated by its high photostability **[Kruglik et al., 2001; McMillin & McNett, 1998]** and low fluorescence background, as compared to other studied porphyrins, enabling us to perform resonance Raman scattering measurements in parallel with absorption and CD experiments. The Raman contribution from the solvent was carefully subtracted and

the spectra were corrected for their non-Raman background by using advanced methods of factor analysis (see the results and discussion section).

In addition to optical spectroscopic techniques (Raman scattering, absorption and CD spectroscopy), other experimental methods (gel electrophoresis, microcalorimetry, NMR) and molecular dynamics simulations were employed.

Native polyacrylamide gel electrophoresis (PAGE) was done for a wide range of G-quadruplex samples as well as for some mixtures of G-quadruplexes with porphyrins at various conditions (defined by DNA concentration, temperature, the type of salt, ionic strength). PAGE was performed in a temperature-controlled electrophoretic apparatus (SE-600; Hoefer Scientific, USA) allowing the regulation of temperature between 2-40°C. Gel concentration was 16% (29:1 monomer to bis ratio, Applichem, Germany). About 2 μg of DNA was loaded into each lane of a 14×16×0.1 cm gel. Gels were stained with Stains All (Sigma-Aldrich) after electrophoresis and scanned using a Personal Densitometer SI, model 375-A (Molecular Dynamics, USA).

DSC (differential scanning calorimetry) experiments complemented optical thermal melting experiments in the determination of thermal stabilities of G-quadruplex samples. All calorimetric measurements were carried out on a commercial differential scanning calorimeter (model 6100 Nano-DSC II) from the company Calorimetric Science Corporation. The scanning rate was 1°C/min in most cases. The DSC profile of phosphate buffer was recorded and subsequently subtracted from sample profiles.

Molecular dynamics simulations were done with NAMD software package (2.7) for the TBA sequence d($G_2T_2G_2TGTG_2T_2G_2$). Starting structure of TBA for molecular simulation was the NMR structure with PDB entry 148d. The results of simulation were analyzed and visualized using VMD, chimera and ptraj modul of Amber software package. The structures were solvated by adding TIP3PBOX water molecules at a distance 10 angstrom around the DNA backbone. The final molecular dynamics trajectories were on 100 ns time scale. Molecular dynamics simulations were also used to study the interaction of TBA with the non-metalated cationic porphyrin $H_2$TMPyP4 (meso-tetrakis(4-(*N*-methylpyridiumyl))porphyrin).

Diffusion-ordered (DOSY) NMR experiment was employed on our demand by associate prof. Jan Lang, Mgr. Marián Grocký and Dr. Pavel Srb at Department of Low Temperature Physics of Charles University in Prague as an alternative

technique to gel electrophoresis to estimate the molecularity of G-quadruplexes (only two sequences TBA: d($G_2T_2G_2TGTG_2T_2G_2$), and T2/T3: d($G_2T_3G_2TGTG_2T_3G_2$) were investigated). NMR experiments were conducted on Bruker Avance 500 MHz spectrometer with TBI probe head in 5 mm Shigemi BMS sample tube. DNA samples were prepared at 20 mM nucleoside concentrations in $D_20$ with 100 mM $K^+$. Pulse sequence used in all measurements was double stimulated echo with bipolar gradients **[Jerschow & Muller, 1997]** and water suppression by excitation sculpting. This sequence is known to suppress many artifacts such as convection, eddy currents.

Key pulse sequence parameters:

P1 = 12.5 μs (1H 90 deg pulse)

P16 = 2000 μs (total length of the bipolar gradient pair: 4000 μs, sine-bell shape)

D20 = 0.065 s for 298.15 K measurements (total diffusion time 130 ms)

D20 = 0.15 s for 275.15 K measurements (total diffusion time 300 ms)

G = 58.4 $Gcm^{-1}$ gradient field strength.

Gas flow 800 l/h for temperature stabilization.

Acquired NMR data were analyzed by two independent methods. Firstly by direct fitting the Stejskal-Tanner equation **[Stejskal & Tanner, 1965]** and secondly by employing an inverse Laplace transform with maximum entropy method by program Gifa **[Pons et al., 1996]**.

# 3. Results and discussion

Here, a summary of the main results achieved during doctoral study is given. First, the methods developed for data processing are described (3.1). Then, the most important part of the thesis concerning the polymorphism of human telomeric quadruplexes conditioned by DNA concentration is presented (3.2). Afterwards, the results on the interaction of G-quadruplexes with cationic porphyrins are presented (3.3). Finally, a dimeric structure of the TBA quadruplex is discussed (3.4).

## 3.1 Development of methods for data processing

In the course of doctoral study, it emerged that there was urgent need for systematic, reliable methods for the treatment of large spectral datasets. Typical thermal melting experiment performed using absorption, CD or Raman spectroscopy produces 30-180 spectra. The differences between individual spectra within such large spectral sets are in many cases minute (especially in the case of Raman spectra) and it is not very practical to analyze them by naked eye. In fact, it is impossible for very large quantities of data. For example, to analyze time evolution of Raman spectra (e.g. slow binding kinetics), hundreds or thousands of spectra may be acquired. A widespread practice (not necessarily incorrect) to processs large datasets is to work only with a small portion of data. This can be achieved simply by discarding a large amount of 'unimportant' experimental data after measurement or by measuring only data of interest. For example absorption **[Mergny et al., 1998]**, fluorescence **[Risitano & Fox, 2003]** or CD melting experiments **[Olsen et al., 2009]** with G-quadruplexes are often conducted at one wavelength. However, recording melting curves at one wavelength may be precarious because potential artifacts, such as small drifts in the baseline with time or additive background owing to impurities or dust in the sample, can be hardly revealed.

In the materials and methods section, the concept of SVD as a multivariate method was introduced. The advantage of SVD over single- wavelength approaches is not only that SVD can extract information from the whole recorded spectral range but also its power to reduce the dimensionality of data without significant loss of information. Multivariate SVD-based methods can be used to determine the number of spectrally distinguishable quadruplex species **[Antonacci et al., 2007; Jaumot et**

**al., 2006; Petraccone et al., 2005]**, classify quadruplex structures **[Jaumot et al., 2009]** or determine their thermodynamical characteristics **[Gray & Chaires et al., 2011]**. SVD algorithm is due to its linear nature well suited to all kinds of spectroscopic problems **[Malinowski, 2002]**. The disadvantage of SVD may be long computation times when it is applied to very large datasets.

For convenient and fast processing of spectroscopic data, we have created a range of useful programs in programming environment Matlab® (version 7.4, Mathworks). Matlab is an easy-to use, high performance language for technical computing. It is especially designed for matrix operations, and SVD procedure is Matlab's buit-in routine. The most important SVD-based programs that we have created in Matlab include the program for manual correction of background and visualization of the results of SVD, the program for intensity normalization and background (solvent) subtraction exploiting the properties of water stretching vibrations, the program for visualization of the results of 2D imaging experiments performed by Raman microscope, the program for finding $pK_a$ constants for pH-dependent spectra (and the construction of the spectra of pure components). Other useful programs include, for instance, the program for interactive shift of x-scale, the program for spectra denoising by means of FFT, the program for the subtraction of solvent from FTIR spectra using $H_20$ combination band (2125 $cm^{-1}$) or the program for combing more spectral sets into one set and the adjustment of their x-scale.

All above-mentioned data processing methods are essentially based on simple operations with matrices or vectors. The principles of some of them will be briefly described. We can begin with the correction of background. The correction of spectral background using SVD is done in the following way (suppose, for simplicity, that we work with Raman spectra). First, the SVD of the original data is accomplished. The non-Raman backgrounds in the resulting subspectra are manually approximated by polynomials of low order or by cubic splines (see figure 35). Background-corrected spectra are then reconstructed (according to the basic relation (2.1) for SVD) from these background-corrected subspectra using the singular values and coefficients from SVD of the original data. The main advantage of this method of background correction is that it can be applied for very large spectral series. Instead of correcting baseline for each original spectrum separately (dataset can have hundreds to thousands spectra), it is sufficient to subtract non-Raman backgrounds only from a few subspectra having the greatest singular values since the subspectra

with small singular values contribute mainly to the white noise. By omitting noisy subspectra from the backward linear combination, one can obtain background- and noise-corrected spectra at the same time. The advantage of background correction of spectra is that the complexity of data is usually markedly reduced after the correction (i.e. factor dimension is smaller) and relevant spectral changes are thus easier to interpret. An example of background correction of a set of Raman spectra depending on DNA concentration is depicted in figure 36. Generally, the correction works well in many cases. However, it may sometimes happen that spectra have some regions slightly oversubtracted.

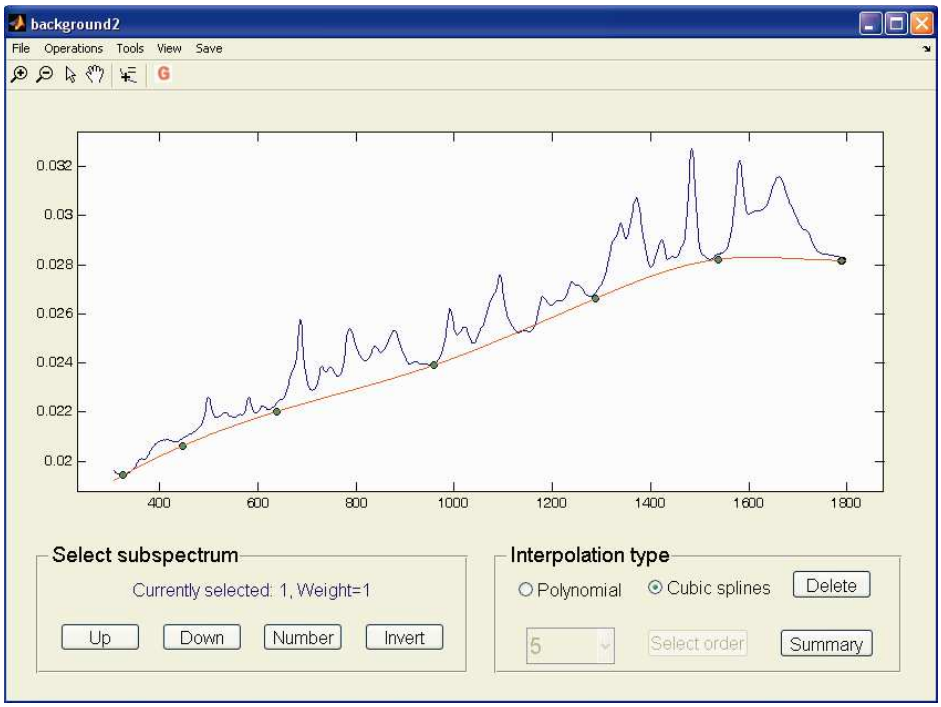



**Figure 35.** Screenshot of the program for the correction of background using SVD. The figure shows the first subspectrum $\boldsymbol{S_1}$ from SVD of a set of Raman spectra (depending on DNA concentration) of human telomeric quadruplex $G_3(TTAG_3)_3$ at 2°C in 240 mM $K^+$. A smooth curve approximating non-Raman background in the subspectrum $\boldsymbol{S_1}$ is obtained as polynomial or spline fit to the series of points selected as user's best estimate. Background-corrected spectra are obtained as a linear combination of only a few background-corrected subspectra using the relation 2.1 (singular values $W_j$ and coefficients $V_{ij}$ are taken from SVD of the original spectra). The final result of the correction of this set of Raman spectra is depicted in figure 36.

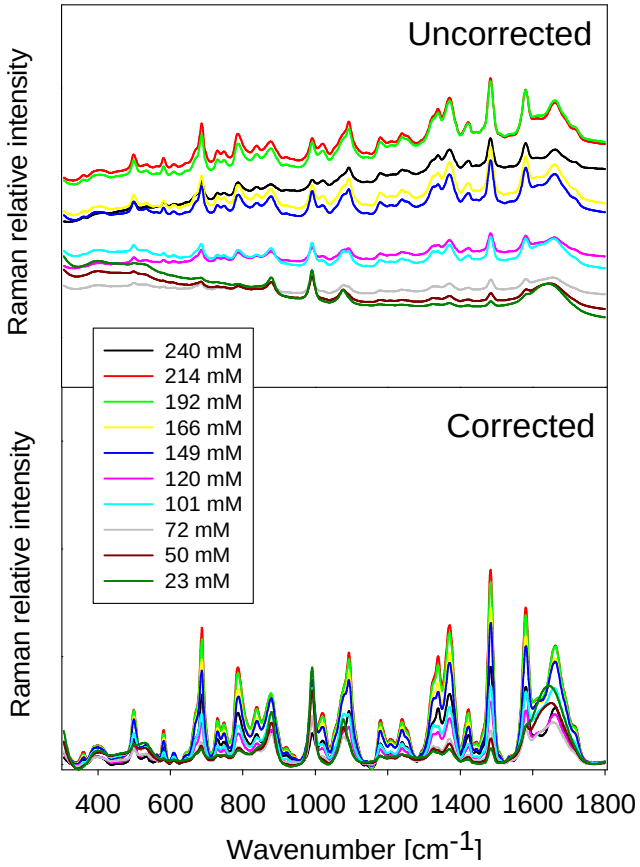



**Figure 36.** The correction of background using SVD for a series of DNA concentration-dependent Raman spectra of the human telomeric quadruplex $G_3(TTAG_3)_3$ at 2°C in 240 mM $K^+$. The values of DNA concentration are given in the legend.

For pre-processing the Raman spectra taken at different experimental and environmental conditions (e.g. DNA concentration, salt concentration, Raman excitation wavelength), there was the need for an effective, reliable and fast method for spectra normalization and solvent subtraction. Therefore, we have developed a semi-automated method based on SVD for intensity normalization, background correction and solvent subtraction that exploits the properties of water stretching vibrations. The method is described in detail in our paper **[Palacký et al., 2011, in attachments, pages 227-250]**. Here, we give only the basic idea of how it works. Our method is well applicable for aqueous solutions of analytes exhibiting small Raman intensity within the range of OH stretching vibrations as compared to that of solvent, which is the case of the majority of biological samples. The normalization of Raman intensity of the analyte is accomplished according to OH stretching band of

water (2700-3900 $cm^{-1}$). It is a good choice since this band is very strong as compared to other commonly used internal intensity standards such as the band at 1640 $cm^{-1}$ belonging to deformation vibration of water or the band at 981 $cm^{-1}$ belonging to $SO_4^{2-}$ ions **[Kříž et al., 2008]**. Another great advantage is that the OH stretching band does not overlap at all with the bands of commonly studied biological analytes such as protein or nucleic acids. Our method requires the reference set of solvent spectra be measured as a function of experimental parameter of interest (typically temperature or ionic strength) in order to subtract the solvent spectrum from analyte spectrum. The results of SVD analysis (within the region of the water stretching band) of the reference solvent spectra are then used to construct optimal artificial solvent spectra (within the entire spectral range) that are the best representation of undesired solvent contribution to the sample spectra. The background-corrected spectra are finally obtained by subtracting the artificial solvent spectra from the sample spectra.

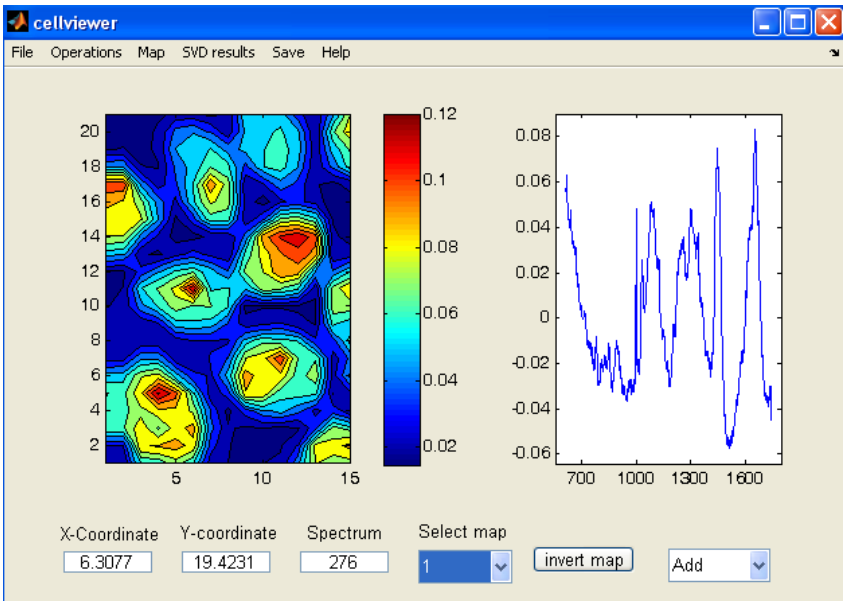



**Figure 37.** Screenshot of the program for visualization of the results of 2D imaging experiments.

Raman 2D mapping experiments typically include a large amount of data (hundreds to thousands of spectra) and their interpretation is therefore almost impossible without using multivariate methods. The results of these experiments can be easily visualized by displaying one specific parameter such as the intensity and

shift of a given band, the ratio of intensities of two bands, etc. However, the coefficients $\boldsymbol{V}_j$ originating from SVD are better choice since they comprise all the information within the spectra. SVD analysis of 2D imaging data can reveal the interdependence between various experimental parameters, possible artifacts, etc. Raman microspectropy can be used for mapping of cell and tissues (we reviewed this in detail in **[Mojzeš at al., 2011]**). The program in Matlab was written for the processing and visualization of imaging experiments conducted on a confocal Raman microspectrometer LabRam HR800 (Horiba) (see figure 37). As an example of usefulness of our programs for data treatment can serve the results of imaging of the cells of yeast vacuoles that have been published recently **[Bednárová at al., 2012, in attachments, page 221-226]**.

## 3.2 Human telomeric G-quadruplexes

The study of telomeric G-quadruplexes constitutes a substantial and possibly the most interesting part of my doctoral research. This study has been done in cooperation with the group of prof. Michaela Vorlíčková from the Department of CD Spectroscopy of Nucleic Acids at Institute of Biophysics of the Academy of Sciences of the Czech Republic (Brno). The paper on Raman study of human telomeric quadruplexes has been very recently accepted for publication **[Palacký et al., 2012, submitted, in attachments, pages 195-220]**.

Since the results are presented in the respective paper, only a brief overview will be given here. The novelty and importance of this work consist in unambiguous and convincing evidence that human telomere sequences can be switched to intramolecular parallel quadruplex in physiologically relevant $K^+$ solutions by the effect of high DNA concentration, without the addition of cosolutes to simulate molecular crowding. This suggestion revealed previously by CD spectroscopy has not been much accepted yet **[Renčiuk et al., 2009]**. Therefore, we decided to correlate the previous CD results with the new experiments using Raman spectroscopy. Raman spectroscopy has been increasingly used in the field of nucleic acids since it is a very sensitive technique to DNA conformational changes (see for example, **[Hanuš et al., 1999; Kříž et al., 2008; Movileanu et al., 2002; Ottová et al., 2011; Vachoušek & Štěpánek, 2008]**). However, the application of Raman spectroscopy for studies concerning G-quadruplex structures is rare (see, for instance, references: **[Abu-Ghazalah et al., 2012; Krafft et al., 2002; Miura &**

**Thomas, 1994; Miura et al., 1995]**). The great advantage of Raman spectroscopy is that it can be applied to a wide range of DNA concentrations ranging from aqueous solutions accessible to absorption and CD measurements up to highly concentrated gels and crystals. This means that Raman spectroscopy can bridge distinctions in the experimental conditions accessible to the particular methods that are the cause of still persisting discrepancies in structural interpretations of human telomere quadruplex arrangements **[Renčiuk et al., 2009]**. As an example serves the results of CD measurements, which are generally interpreted on the basis of high-resolution structures revealed by NMR, however at the concentrations of two orders of magnitude higher than those commonly encountered in CD spectroscopy, and thus often erroneously. Our Raman study clarifies this matter thoroughly and systematically. Furthermore, it shed a light on the quadruplex thermal melting and annealing effects at extremely high DNA concentrations, the information in principle inaccessible to absorption and CD spectroscopies due to the sample evaporation in ultrathin cells with detachable windows. Raman melting profiles obtained in the current study are the first and sole temperature profiles available to date for high-DNA quadruplex conditions. They can be inspiring for other investigators interested in determination of thermal stabilities and thermodynamic parameters of various quadruplex structures under the high-DNA concentrations.

The current searching for the real structure of telomere DNA under physiological conditions is associated with $K^+$-quadruplex structures conditioned by molecular crowding since DNA inside the cell is highly condensed. These crowding conditions not accessible to most common experimental methods are simulated by the addition of various cosolutes. However, the exact role of cosolutes in the quadruplex transitions from antiparallel to parallel arrangement is unclear, since it was shown that structural changes are rather a result of dehydration instead of the molecular crowding **[Miyoshi et al., 2006; Petraccone et al., 2012]**. Raman spectroscopy offers a unique possibility of studying the crowding effects induced solely by DNA itself, moreover in a methodologically consistent way.

The main achievements of the work on DNA telomeric sequences can be summarized in a few points. We studied the dependence of Raman spectra of $G_3(TTAG_3)_3$, the shortest four-repeat model of the human telomeric quadruplex, on DNA concentration. First, we show that the Raman spectra of $G_3(TTAG_3)_3$ at the lowest nucleoside concentrations are very similar in the presence of $Na^+$ and $K^+$,

which indicates the same antiparallel strand alignment in both salts. Raman spectra of thermodynamically equilibrated (slowly annealed) samples of $G_3(TTAG_3)_3$ stabilized by $K^+$ exhibit gradual spectral changes as a function of nucleoside concentration within the range 8-200 mM (see figure 38). Linear character of the interquadruplex transition revealed by careful processing of Raman data using SVD analysis can be interpreted as conformational heterogeneity and the coexistence of multiply quadruplex forms in solution. This polymorphism complicates the unambiguous structural identification of $G_3(TTAG_3)_3$ in solution by NMR and the correlation of the results provided by different experimental methods. The observed transition of $G_3(TTAG_3)_3$ in $K^+$ conditioned by DNA concentration reflects the switch from predominately antiparallel folding (at low DNA concentrations) to predominately parallel folding of $G_3(TTAG_3)_3$ (at high DNA concentrations). This interesting conclusion is made on the basis of the observation of C2'-*endo/syn* (671 ± 2 $cm^{-1}$, 1326 ± 2 $cm^{-1}$) and C2'-*endo/anti* (686 ± 2 $cm^{-1}$, 1338 ± 2 $cm^{-1}$) Raman markers of guanosines **[Krafft et al., 2002]** (see figure 38) that best reflect the studied conformational rearrangement. In addition to these key markers, the interquadruplex transition is associated with a range of other markers. The weak band at ~611 $cm^{-1}$ is a very interesting Raman marker since its intensity strongly decreases upon the switch from the antiparallel to parallel form of $G_3(TTAG_3)$. We speculate that the 611 $cm^{-1}$ band may be a universal marker distinguishing between side/diagonal and double-chain-reversal quadruplex loops.

Similar structural transition as that induced by increasing DNA concentration at a fixed concentration of $K^+$ is observed in the presence of increasing concentration of $K^+$ at a fixed concentration of $G_3(TTAG_3)_3$. We conclude that the propensity of the $G_3(TTAG_3)_3$ quadruplex to switch from the antiparallel to parallel arrangement upon the increase of DNA concentration is synergically amplified by increasing concentration of $K^+$. By contrast, the structure of $G_3(TTAG_3)_3$ quadruplex in $Na^+$ remains antiparalell regardless the concentrations of DNA and/or $Na^+$. Interestingly, for the modified telomeric sequence $AG_3(TTAG_3)_3$, it was reported that the transition to the parallel arrangement can be also induced by $Na^+$ concentration **[Abu-Ghazalah et al., 2012]**.

A crucial finding of our study is that the interquadruplex transition is extremely slow even at the highest nucleoside concentration for $K^+$-stabilized $G_3(TTAG_3)_3$ samples unless they are thermodynamically equilibriated (i.e. slowly annealed). In

fact, the structure of non-annealed $G_3(TTAG_3)_3$ in $K^+$ formed at extremely high nucleoside concentrations (more than 200 mM in nucleosides) is very similar to the thermodynamically equilibrated $G_3(TTAG_3)_3$ structures formed at low DNA concentrations (in $K^+$ or $Na^+$). In this sense, it is important to note that although the effect of sample heating and cooling is a key factor determining the folding topology of the telemeric quadruplex, it is not mentioned in some NMR studies **[Ambrus et al., 2006; Luu et al., 2006]**.

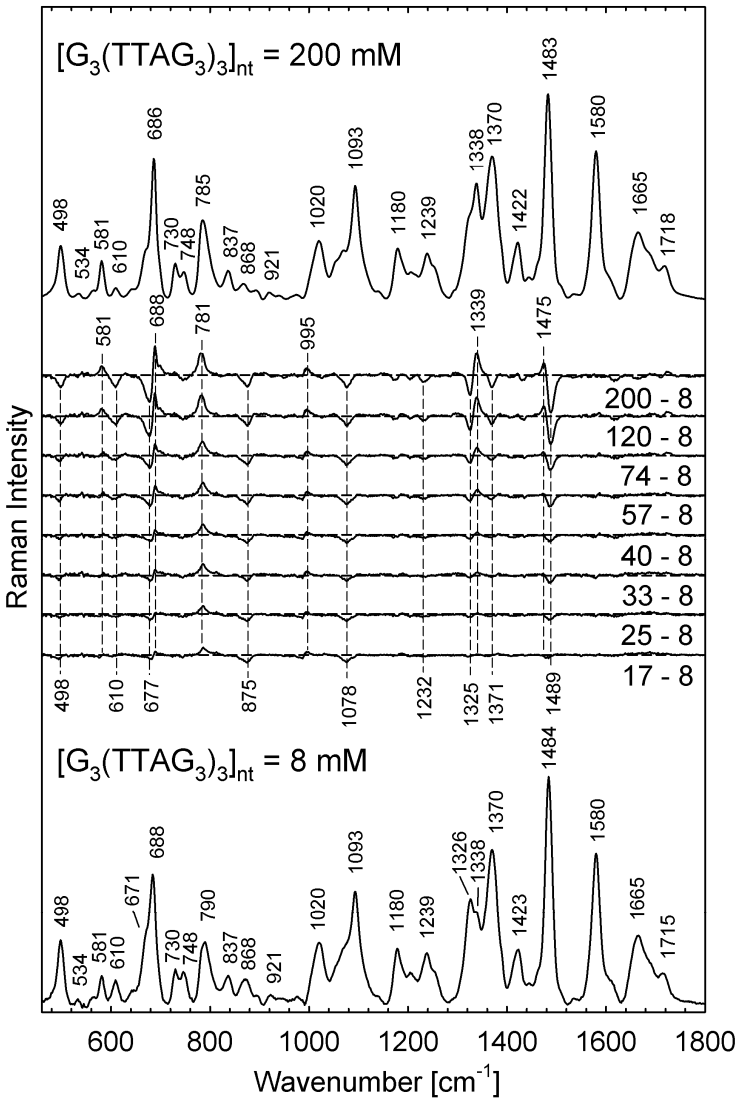



**Figure 38.** Raman spectra of $G_3(TTAG_3)_3$ in 200 mM $K^+$ (30 mM PBS, pH 6.8, t = 5°C) at the nucleoside concentrations of 8 mM (bottom trace) and 200 mM (top trace). Intermediate traces show the differences between the spectra at indicated concentrations and that at the lowest one.

The spontaneously adopted antiparallel structure of $K^+$-stabilized $G_3(TTAG_3)_3$ is not a thermodynamically most stable quadruplex arrangement. For example, when $G_3(TTAG_3)_3$ at high DNA concentration (330 mM in nucleosides) in 230 mM $K^+$ is heated to 95°C and then rapidly cooled down (in less than 1 min) to room temperature, partial (~50% according to differential Raman features) switching to the parallel form is achieved, as judged from the comparison with slowly-annealed samples. The hysteresis observed in thermal profiles corresponding to the 581, 611 and 1485 $cm^{-1}$ bands indicates that the antiparallel structure of $G_3(TTAG_3)_3$ switches to the parallel structure before the quadruplex is destabilized (see figure 39). This can be interpreted as a sign of thermally-induced conversion of intramolecular antiparallel conformers to more stable forms **[Víglaský et al., 2010]**.

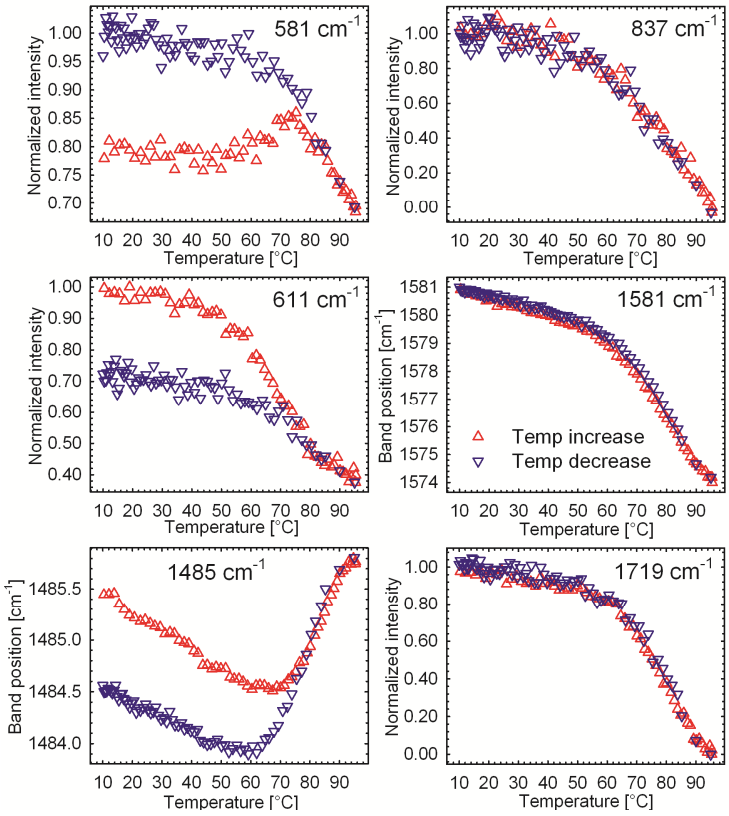



**Figure 39.** Heating and cooling Raman profiles of selected Raman bands sensitive (581, 611 and 1485 $cm^{-1}$) and insensitive (837, 1581 and 1719 $cm^{-1}$) to antiparallel-to-parallel switching of the $K^+$-stabilized $G_3(TTAG_3)_3$ quadruplex. Initially antiparallel $K^+$-quadruplex (330 mM), spontaneously formed on addition of 230 mM $K^+$, was heated from 10 to 95°C and then cooled down at the same rate of ca 1.5°C/min.

## 3.3 Interaction of guanine quadruplexes with cationic porphyrins

The interaction of cationic porphyrins as a class of small ligands with G-quadruplexes has been studied in depth by many researchers **[Georgiades et al., 2010]**. Porphyrins may both stabilize **[Bhattacharjee et al., 2011]** or destabilize G-quadruplex structures **[Weisman-Shomer et al., 2003]**. The modulation of structural properties of G-quadruplexes by porphyrins is of biological relevance **[Joachimi et al., 2007]**. In the present study, we have used cationic porphyrins as molecular probes of G-quadruplex DNA and its conformational changes. Although diverse optical methods (absorption, CD, fluorescence spectroscopy) as well as other methods (mass spectroscopy, calorimetry, molecular modeling, NMR, etc.) are available for this purpose **[Murat et al., 2011]**, there are only a few studies using Raman spectroscopy **[Wei et al., 2006]**. Therefore, we complemented common experimental techniques (absorption, CD spectroscopy) with Raman spectroscopy. The experiments were conducted with eight structurally related porphyrins depicted in figure 40. They differ from each other in the central metal cation ($Cu^{2+}$, $Zn^{2+}$ or no cation) and the structure of side chains (all side chains for a given porphyrin are the same).

According to the kinetics of spectral changes observed in the course of heating-cooling cycle, the studied G-quadruplexes can be classified into two groups. While some G-quadruplex samples exhibit no or small hysteresis, other show quite large hysteresis between heating and cooling thermal profiles (see figure 41). Generally, as already mentioned in the introduction part, hysteresis is a signature of slower kinetics of G-quadruplex formation. It is detected primarily in intermolecular (bimolecular or tetramolecular) G-quadruplex structures such as, for example, $G_3T_5G_3$ and $G_4T_5G_4$. Small hysteresis in the case of KRAS (d($AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2$); a regulatory element of the human KRAS promoter) could be considered as a sign of its intermolecularity. Nonetheless, it is likely a result of a slow equilibrium between two intramolecular G-quadruplex structures of KRAS **[Cogoi et al., 2008]**.

For the studied porphyrins (see figure 40), the Soret band is centered between 415-437 nm. During PhD study, most attention was paid to the porphyrin CuTMPyP4 and its non-metalated form $H_2$TMPyP4 (see table 5 for a summary of data on the interaction of CuTMPyP4 and $H_2$TMPyP4 with selected

polyoligonucleotides and G-quadruplexes). The visible part (350-500 nm) of the absorption spectra of quadruplex-porphyrin mixtures was analyzed by SVD approach. The data interpretation is not so straightforward. The factor dimension is often high (>3). However, SVD analysis can often reveal small differences between spectra. It is especially useful for finding the relation between large sets of spectra recorded at drastically different conditions such as the presence of different salts. The corresponding changes in real spectra are often small (the shift of the Soret band about ~1-2 nm) and/or it is difficult not to get lost in the large amount of data.

The evaluation of a large number of absorption spectra within the Soret band is demonstrated in figure 42. The spectra correspond to the 1:1 mixtures of the TBA quadruplex, d($G_2T_2G_2TGTG_2T_2G_2$), with the cationic porphyrin, CuTMPyP4, (see figure 40) prepared in the presence of LiCl, NaCl and KCl. The spectra show no hysteresis between heating and cooling. It implies that CuTMPyP4 binds to TBA in exactly the same way before and after the thermal denaturation (followed by renaturation) of the porphyrin-DNA complex. No hysteresis in the absorption spectra of TBA with (see figure 42) and without CuTMPyP4 (see figure 43) is strong evidence against bimolecularity of TBA **[Fialová et al., 2006]**. We will deal with the molecularity of TBA in the section 3.4.

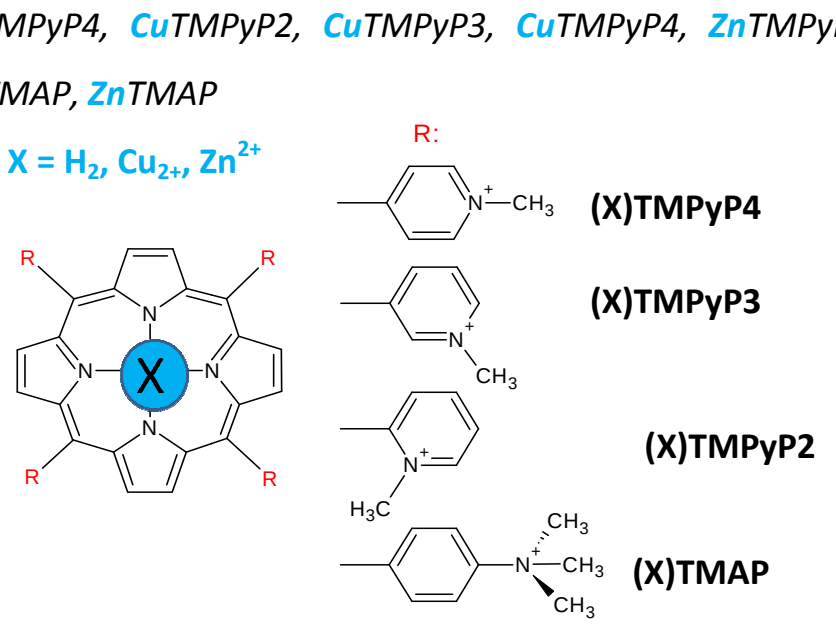



**Figure 40.** The schematic figure showing the structures of the cationic porphyrins used in our studies.

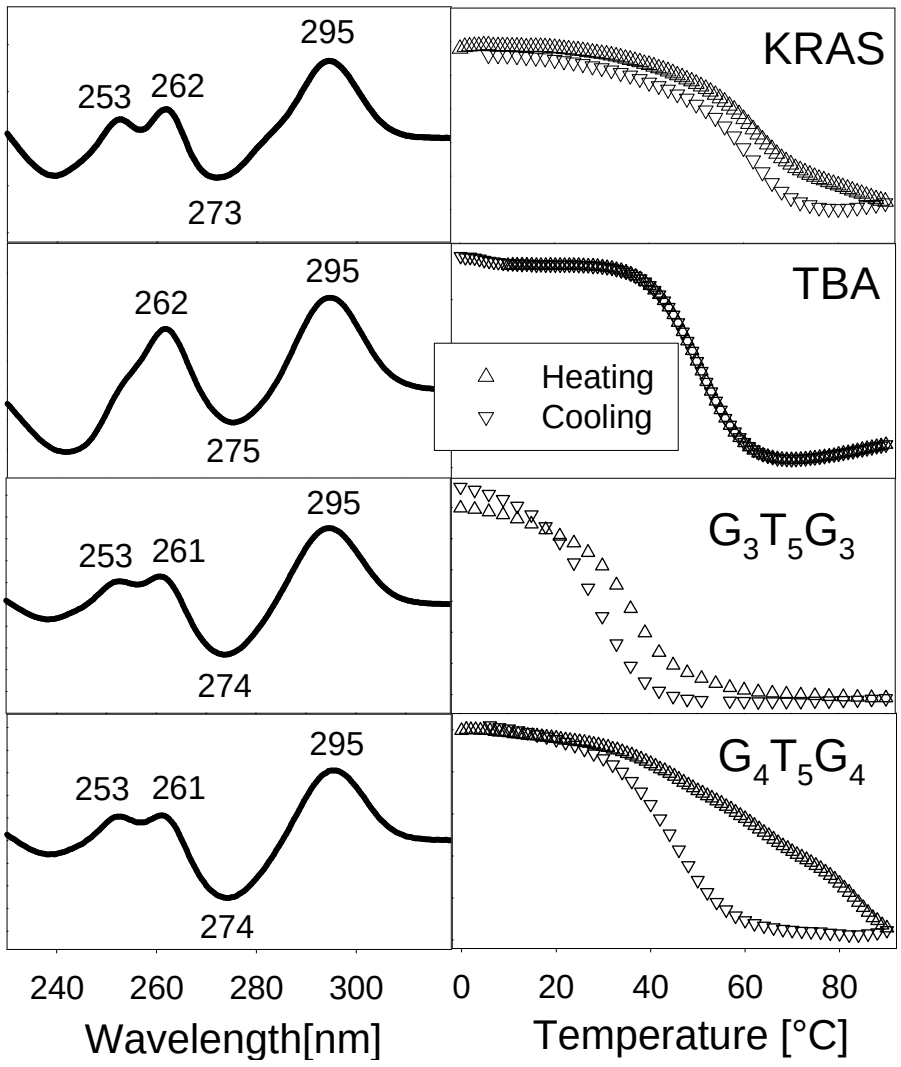



**Figure 41.** The SVD analysis of absorption spectra of four DNA oligonucleotides (KRAS, TBA, $G_3T_5G_3$, and $G_4T_5G_4$) recorded as a function of temperature. Only the subpectra $\boldsymbol{S_2}$ (black curves on the left part of the figure) with the corresponding $\boldsymbol{V_2}$ coefficients (scatter plots on the right part of the figure with up and down triangles indicating heating and cooling, respectively) are shown. The subspectra $\boldsymbol{S_2}$ are similar to thermal difference spectra (compare with figure 19; note that subspectra and coefficients are unambiguous except for sign). The sigmoidal shapes of the $\boldsymbol{V_2}$ coefficients nicely reflect thermal destabilization of G-quadruplexes. They indicate that the absorbance at 295 nm decreases with increasing temperature. Whereas no thermal hysteresis is observed for TBA, a moderate to a relatively large hysteresis is found for $G_3T_5G_3$ and $G_4T_5G_4$ and KRAS. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 3 μM oligonucleotide concentrations.

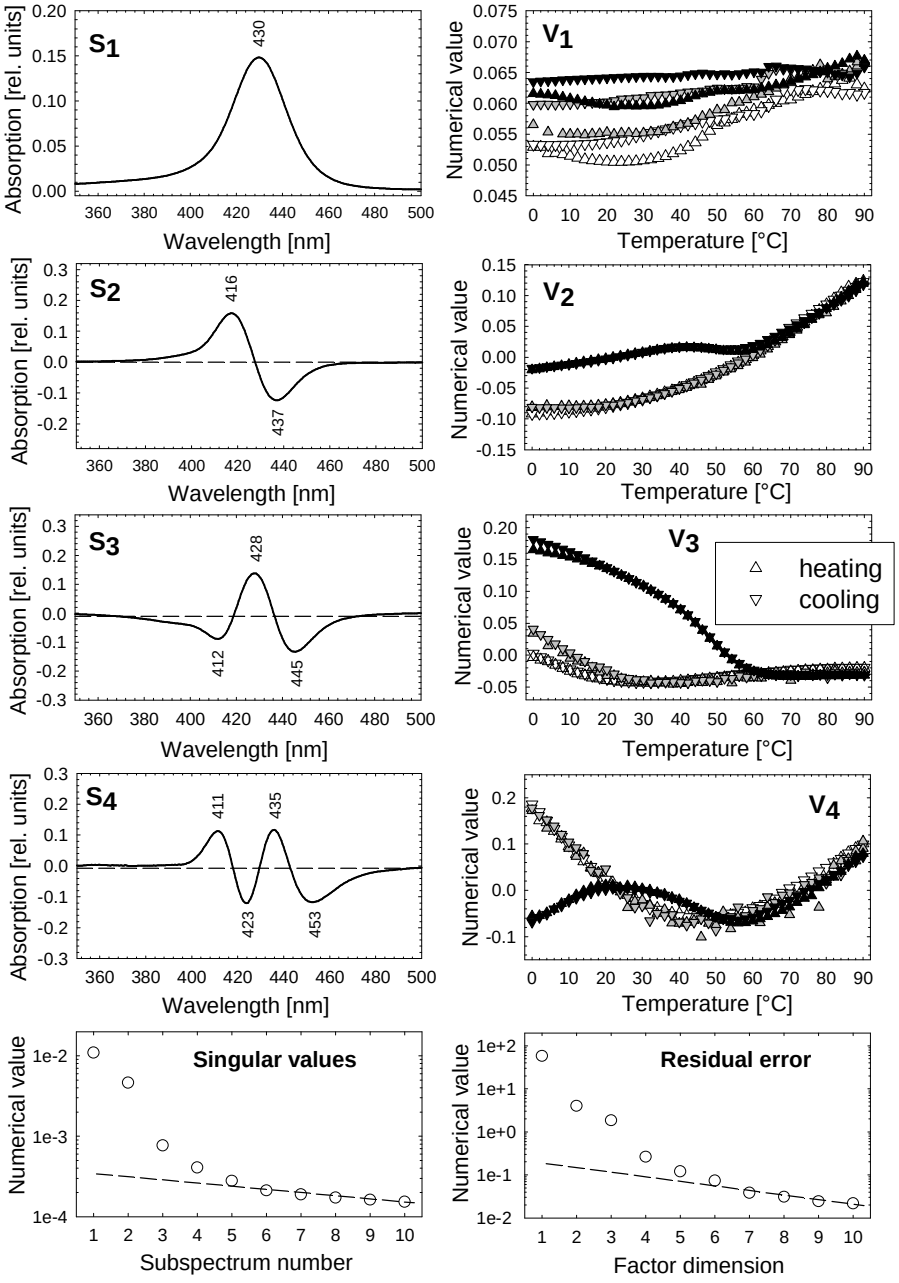



**Figure 42.** The joint SVD analysis of three sets of temperature-dependent absorption spectra within the Soret band spectral region corresponding to the mixture of TBA, d($G_2T_2G_2TGTG_2T_2G_2$), with the porphyrin, CuTMPyP4, (1:1) in 100 mM LiCl, NaCl and KCl. The $\boldsymbol{V_j}$ thermal profiles are colored white (LiCl), gray (NaCl) and black (KCl). Note: heating and cooling profiles are reversible indicating that the sample was at thermodynamic equilibrium during the experiment. Factor dimension for the full dataset is ~ 5. Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

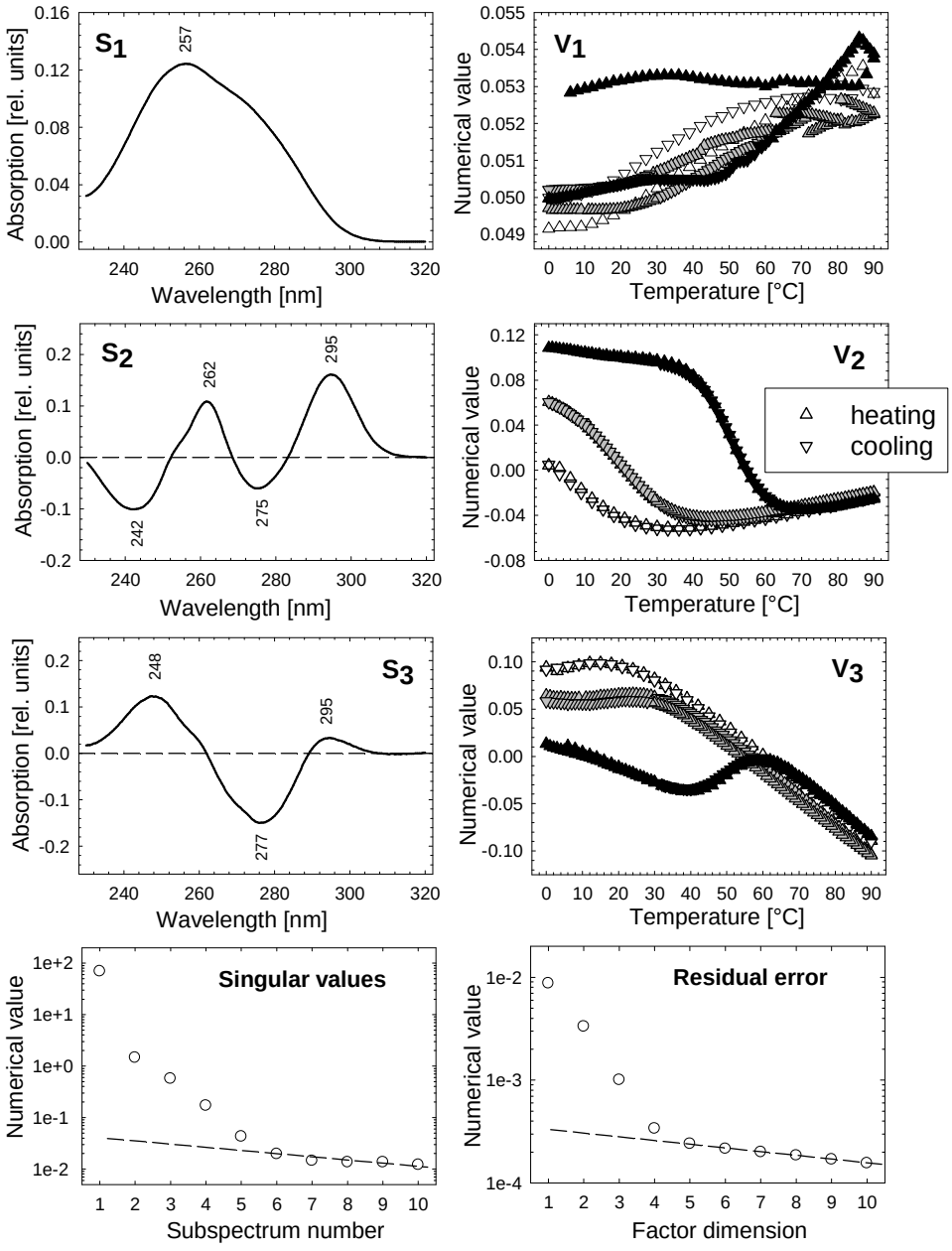



**Figure 43.** The joint SVD analysis of three sets of temperature-dependent absorption spectra corresponding to TBA, d($G_2T_2G_2TGTG_2T_2G_2$) without the presence of porphyrin, in 100 mM LiCl, NaCl and KCl. The $\boldsymbol{V_j}$ thermal profiles are colored white (LiCl), gray (NaCl) and black (KCl). Note: heating and cooling profiles are reversible indicating that the sample was at thermodynamic equilibrium during the experiment. The subspectrum $\boldsymbol{S_2}$ and the corresponding coefficient $\boldsymbol{V_2}$ best reflect thermal destabilization of the TBA quadruplex with increasing temperature. Factor dimension for the full dataset is ~ 5. Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

The joint SVD analysis of the absorption spectra of TBA in LiCl, NaCl and KCl shows that the thermal stability of TBA is highest in $K^+$ (~50°), significantly lower in $Na^+$ (20°C) and lowest in $Li^+$ (see figure 43). In parallel, the joint SVD analysis of the absorption spectra of the TBA-CuTMPyP4 mixture (1:1) clearly demonstrates that the porphyrin discriminates between the distinct G-quadruplex structures induced by the presence of different salts ($Li^+$, $K^+$, $Na^+$) and temperature (see figure 42). Distinct interaction of CuTMPyP4 with TBA in the presence of the three salts (LiCl, NaCl and KCl) manifests primarily in different shift (reflected mainly in the subspectrum $\boldsymbol{S_2}$) and broadening (reflected mainly in the subspectrum $\boldsymbol{S_3}$) of the Soret band (see figure 42).

The Soret band can also indirectly disclose slow kinetics of G-quadruplex formation (see figure 44). The SVD analysis of the 1:1 mixture of KRAS, $d(AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2)$, with CuTMPyP4 indicates that the binding mode of CuTMPyP4 to KRAS must be different before heating and after cooling the denaturated sample. This is best demonstrated by the divergence of heating and cooling $\boldsymbol{V_4}$ thermal profiles (see figure 44). The pronounced hysteresis observed for the KRAS-CuTMPyP4 mixture can be interpreted as the ability of the porphyrin to stabilize a specific type of G-quadruplex structure (depending on the annealing regime).

One can see that the hysteresis observed in the absorption spectra of KRAS is reproduced within the DNA region of the CD spectra (compare the coefficient $\boldsymbol{V_4}$ in figure 44 with the coefficient $\boldsymbol{V_2}$ in figure 45). The hysteresis shown in figure 45 could be interpreted as the irreversible transition of KRAS (in the presence of CuTMPyP4) from the parallel to mixed parallel/antiparallel quadruplex after thermal denaturation and subsequent annealing. The interconversion is characterized by a small increase of the positive faint band at ~290 nm concerted with a decrease of the positive band at ~ 262 nm and a decrease of the negative band at ~ 240 nm. The transition is best visualized by the subspectrum $\boldsymbol{S_2}$ and the corresponding coefficient $\boldsymbol{V_2}$ (see figure 45). Exactly the opposite transition induced by a different cationic porphyrin in the absence of stabilizing cations and buffer was observed for the tetrahymena telomeric sequence $d(T_2G_4)_4$ **[Murashima et al., 2008]**.

Porphyrins are normally achiral molecules but the chirality can be induced in them through the distorsion of their structures upon their interaction with nucleic acids or other biomolecules. Three main binding modes of porphyrins to DNA can be

distinguished: intercalation (negative CD), outside groove binding (positive CD) and outside stacking (bisignate CD) **[Balaz et al., 2005; Pasternack, et al. 1983]**. The majority of our samples exhibited either positive- or negative-induced CD bands, suggestive of the groove or intercalative binding mode of porphyrins to G-quadruplexes, respectively. For instance, we observed a weak negative CD band at ~436 nm for CuTMPyP4 bound to KRAS. The SVD analysis of temperature-dependent CD spectra of KRAS complexed with CuTMPyP4 in KCl within the spectral region of the Soret band indicates that this negative band exhibits analogical hysteretic behavior (see figure 46) as the Soret band (see figure 44) and the ultraviolet region of DNA absorption of CD spectra (see figure 45). The annealing of the thermally denaturated KRAS-CuTMPyP4 complex results in an irreversible increase in the negative ellipticity of the CD band at ~436 nm and its small red shift (see figure 46).

The above-mentioned classification of CD-induced spectra of porphyrins is doubtful since it is based on the interaction of porphyrins with canonical nucleic acids forms such as duplexes **[Pasternack, et al. 1983]**. Although the negative CD band at ~436 nm can be interpreted as a sign of intercalation of CuTMPyP4 within G-quartets of KRAS in KCl, it was proposed that $H_2$TMPyP4 (free base of CuTMPyP4) showing the same negative band preferentially binds within the loops of the KRAS quadruplex and stacks on the external G-tetrads **[Cogoi et al., 2008]**.

The resonance Raman spectra of porphyrins complexed with DNA reflect indirectly the information about the DNA structure and the possible binding modes of porphyrins to DNA. The definite advantage of resonance Raman spectroscopy is that the measurements can be performed at relatively low concentrations of porphyrins (and DNA) that are comparable to those commonly encountered in absorption or CD spectroscopy. The intensity of Raman spectra of porphyrins can be enhanced either resonantly by choosing the excitation line of laser within their absorption bands **[Pasternack et al., 2001; Uno et al., 2002]** or through the use of the SERS mechanism **[Wei et al., 2006]**. For resonance Raman measurements, we could use only the porphyrin denoted as CuTMPyP4 (Cu(II)-5,10,15,20-*Tetrakis*(1-methyl-4-pyridyl)) since other porphyrins exhibit high fluorescence and enhanced photodegradation in laser beam **[Kruglik et al., 2001; McMillin & McNett, 1998]**. The interaction of the porphyrin CuTMPyP4 with DNA (a fortiori with G-quadruplexes) has been little studied so far **[Evans et al., 2007]**, unlike its free base

counterpart $H_2$TMPyP4 **[del Toro et al., 2008; Haq et al., 1999]**. The excitation line of our laser (441.6 nm) is well suited for studies of DNA/CuTMPyP4 complexes since it is close to the Soret band of CuTMPyP4 (~424 nm). Resonance Raman spectroscopy of porphyrins has been never employed to probe the structure of guanine quadruplexes, and our work is the first of its kind.

The spectral variation within a group of resonance Raman spectra is often rather small. Consequently, sophisticated data processing methods had to be employed to extract useful information from data **[Palacký et al., 2011, in attachments, pages 227-250]**. It is surprising that such unrelated techniques as absorption, CD and Raman spectroscopy provide consistent results. It is nicely demonstrated for the TBA and KRAS sequence.

SVD analysis of resonance Raman spectra of CuTMPyP4 complexed with TBA (1:1) indicates that the structure of the CuTMPyP4-TBA complex changes reversibly with temperature (see figure 47). Resonance Raman spectroscopy clearly differentiates between the structures of TBA induced by different cations ($K^+$, $Na^+$ and $Li^+$), as best reflected in the $\boldsymbol{V_3}$ coefficient profile (see figure 47). SVD shows that all the Raman spectra are the same above ~55°C, meaning that CuTMPyP4 at higher temperatures interacts similarly with the TBA structures stabilized by KCl, NaCl or LiCl.

In contrast with TBA, the interaction of KRAS with CuTMPyP4 is distinctly different. Resonance Raman spectra of CuTMPyP4 in 1:1 mixture with KRAS exhibit considerable hysteresis between heating and cooling (see figure 48). It means that the initial structure of the CuTMPyP4-KRAS complex is irreversibly changed after its thermal denaturation followed by annealing.

In summary, the agreement between the results of common optical methods (compare for instance figures 42 and 47 or figures 44 and 48) and those provided by resonance Raman spectroscopy indicates that cationic porphyrins can be employed as sensitive Raman probes of the G-quadruplex structure. Although not all the results in our current study are new, our scientific contribution has methodological significance.

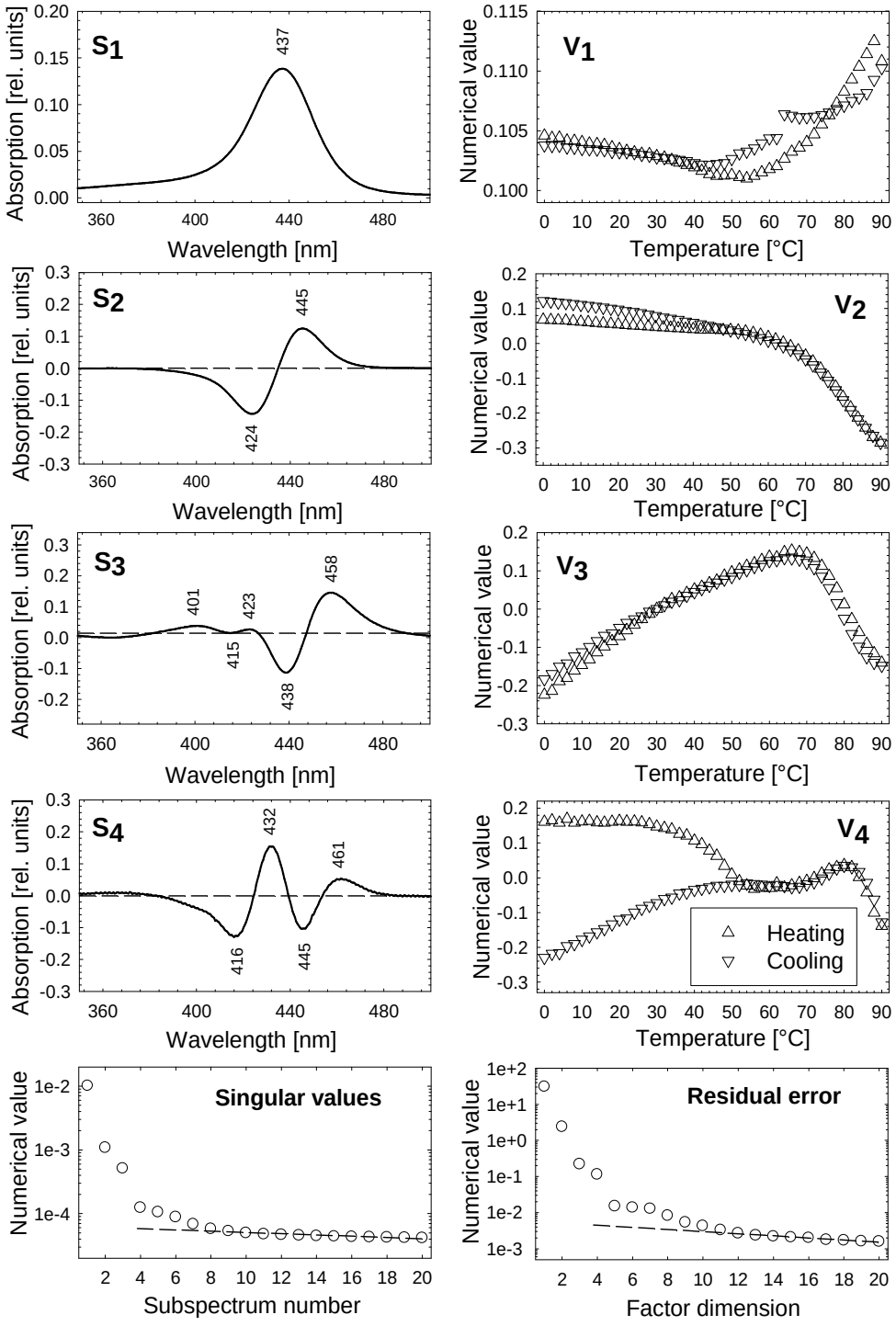



**Figure 44.** The SVD analysis of temperature-dependent absorption spectra within the Soret band spectral region corresponding to the mixture of KRAS, d($AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2$), with CuTMPyP4 (1:1) in 100 mM KCl. Factor dimension is ≥ 4-5. Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

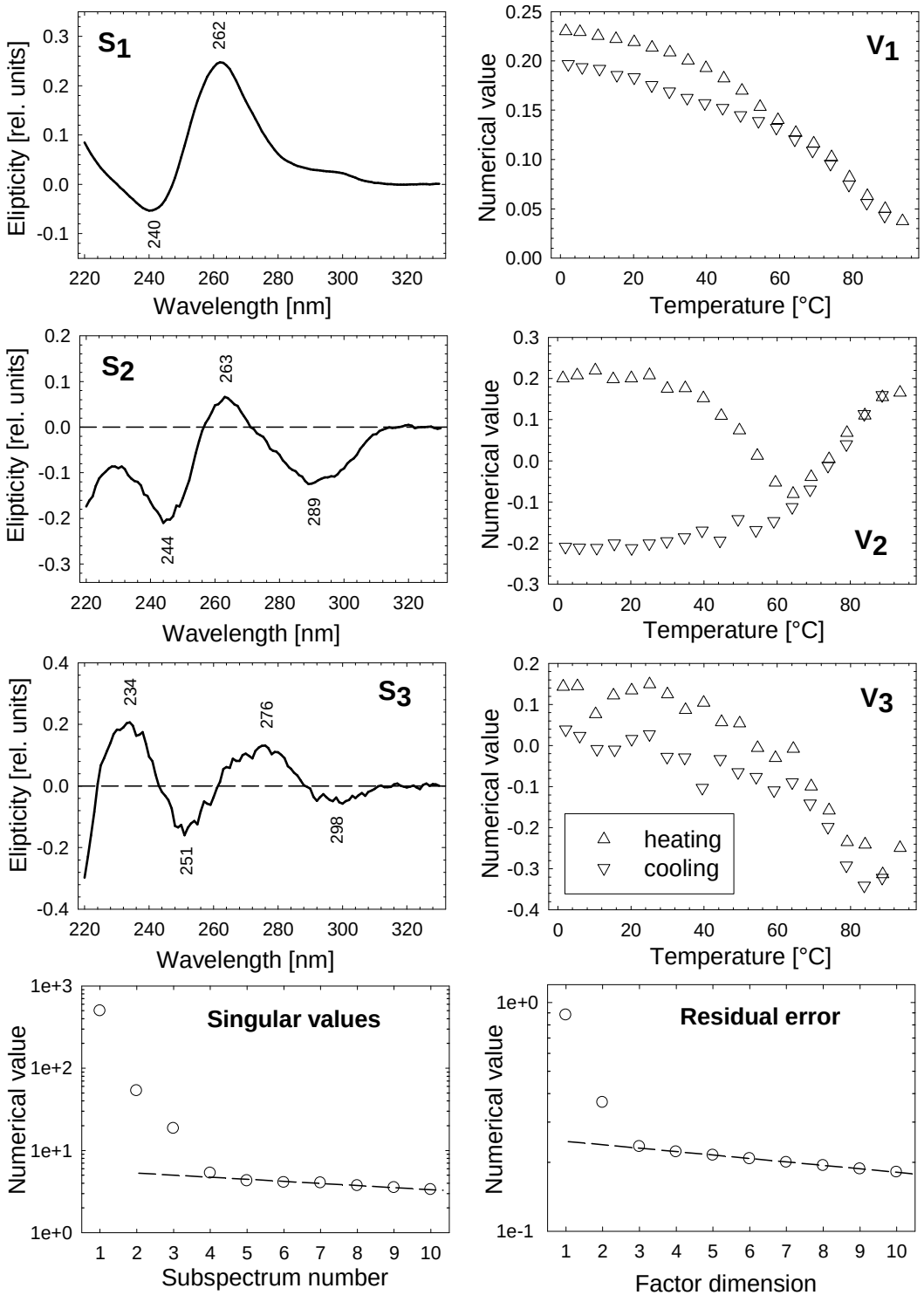



**Figure 45.** The SVD analysis of temperature-dependent CD spectra within the region of DNA absorption corresponding to the mixture of KRAS, d($AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2$), with CuTMPyP4 (1:1) in 100 mM KCl. Factor dimension is 3. Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

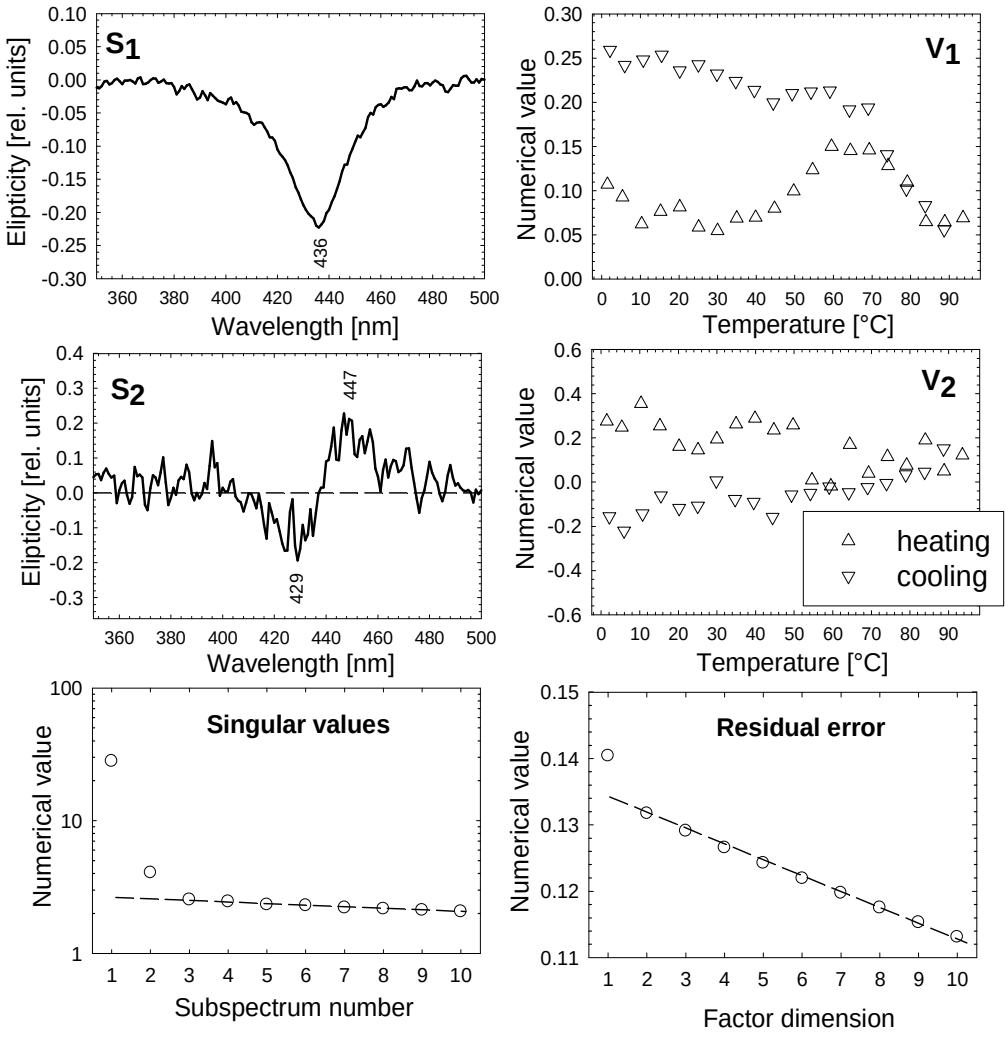



**Figure 46.** The SVD analysis of temperature-dependent CD spectra within the Soret band spectral region corresponding to the mixture of KRAS, d($AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2$), with CuTMPyP4 (1:1) in 100 mM KCl. Factor dimension is 2. Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

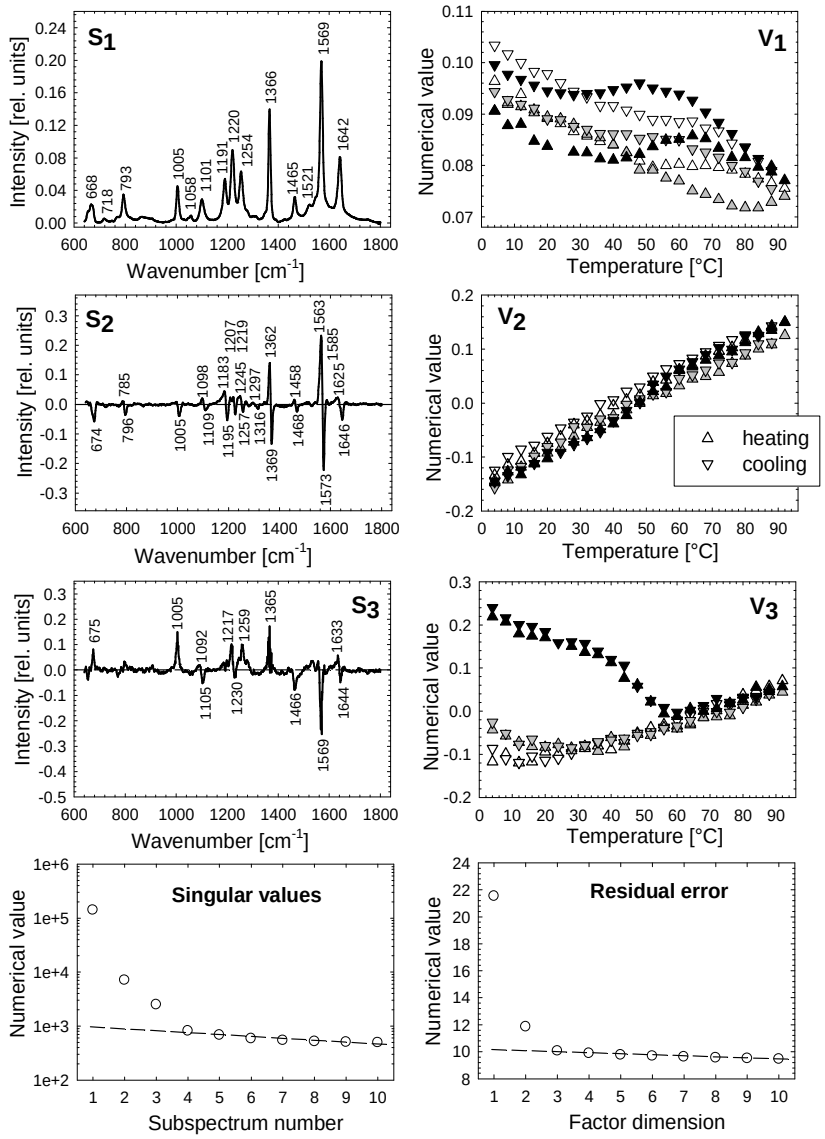



**Figure 47.** The joint SVD analysis of three sets of temperature-dependent resonance Raman spectra corresponding to the mixture of TBA with CuTMPyP4 (1:1) in 100 mM LiCl, NaCl and KCl. The $\boldsymbol{V_j}$ thermal profiles are colored white (LiCl), gray (NaCl) and black (KCl). The factor dimension is 3. The subspectrum $\boldsymbol{S_1}$ is a kind of average spectrum with the coefficient $\boldsymbol{V_1}$ reflecting differences in overall Raman intensity. The remaining two components $\boldsymbol{S_2}$ and $\boldsymbol{S_3}$ describe the greatest changes in the spectra and the corresponding coefficients $\boldsymbol{V_2}$ and $\boldsymbol{V_3}$ show no hysteresis between melting and cooling profiles. The coefficient $\boldsymbol{V_3}$ best describes the thermal stability of the TBA-CuTMPyP4 complex (it is apparent that the stability is highest in KCl and lowest in LiCl). Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

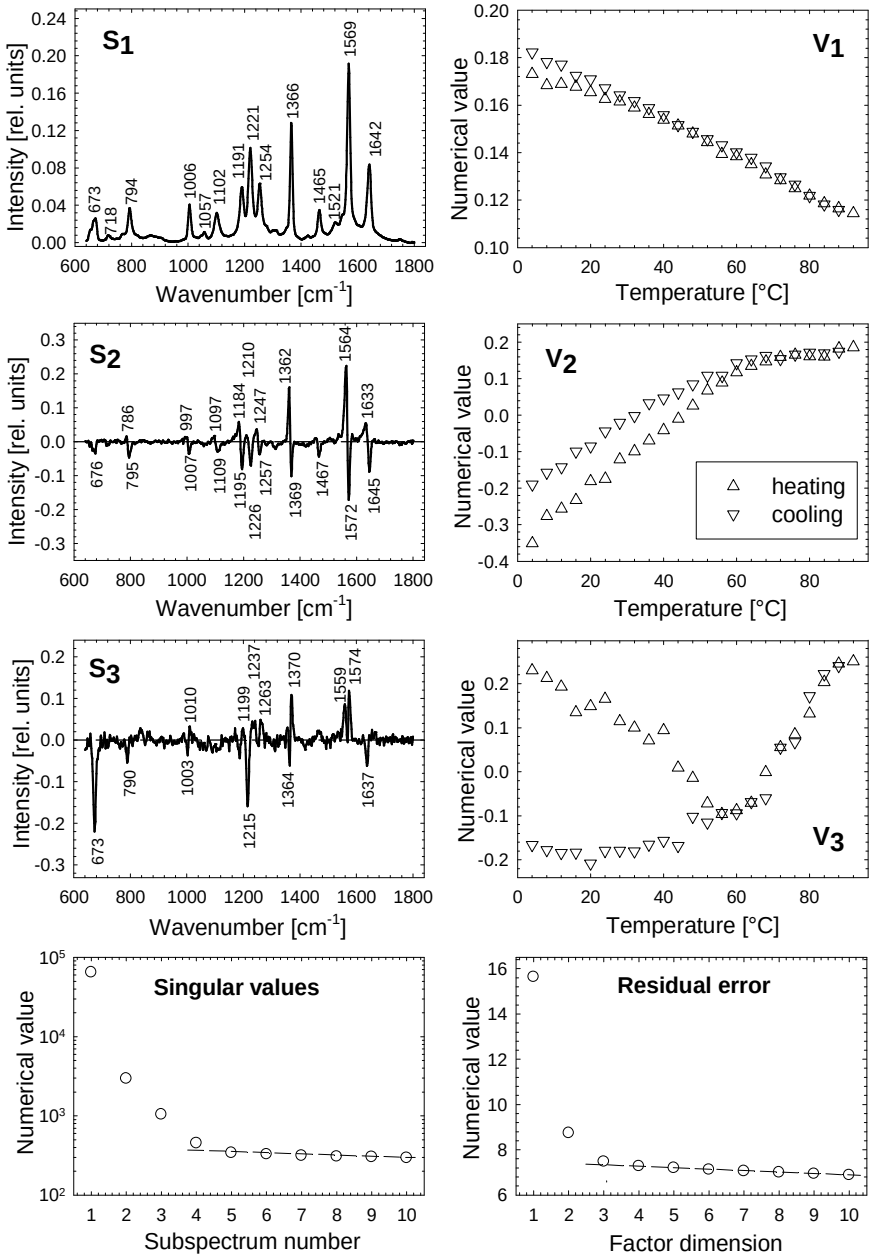



**Figure 48.** The SVD analysis of temperature-dependent resonance Raman spectra corresponding to the mixture of KRAS with CuTMPyP4 (1:1) in 100 mM KCl. The factor dimension is 3. The subspectrum ***S₁*** is a kind of average spectrum with the coefficient ***V₁*** reflecting differences in overall Raman intensity. The remaining two components ***S₂*** and ***S₃*** describe the greatest changes in the spectra and the corresponding coefficients ***V₂*** and ***V₃*** show significant hysteresis between melting and cooling profiles. The coefficient ***V₃*** best describes the thermal structural changes of the KRAS-CuTMPyP4 complex. Experimental conditions: 20 mM lithium cacodylate buffer, pH 6.8, 3 μM oligonucleotide concentration.

## 3.4 TBA quadruplex

We selected for the majority of our spectroscopic investigations a model sequence d($G_2T_2G_2TGTG_2T_2G_2$) known as TBA (**t**hrombin **b**inding **a**ptamer). The TBA quadruplex has only two guanine tetrads, which is rare for G-quadruplexes **[Lim et al., 2009b]**. The consequence of this is that TBA melts at only 50°C in 100 mM $K^+$, which is far below ~60°C of the corresponding heteroduplex (compare figures 31, 32 and 34 with figure 33). Moderate thermal stability of TBA is well suited for our spectroscopic melting experiments. TBA molecule serves not only as a good model system but the study has also biological relevance. TBA is at present amongst the most frequently studied DNA aptamers. It is the first single-stranded DNA molecule selected to bind specifically and with high affinity to protein that does not normally interact with nucleic acids **[Bock et al., 1992]**. The target protein, thrombin, is a protease implicated in blood clotting **[Crawley et al., 2007]**, and the study of its interaction with TBA is therefore very important.

The tertiary structure of TBA was resolved by both NMR **[Macaya et al., 1993; Schultze et al., 1994; Wang et al., 1993a; Wang et al., 1993b]** and X-ray spectroscopy **[Padmanabhan et al., 1993; Padmanabhan & Tulinsky et al., 1996]**. The proposed folding of TBA is an intramolecular chair-like quadruplex consisting of two edge-like TT loops and one TGT central loop with the adjacent phosphodiesther backbones running antiparallel to each other (see the scheme in figure 49). The structure of TBA quadruplex and/or its interaction with thrombin protein, other molecules and metal cations was further analyzed in depth by a range of methods including circular dichroism (see, for instance, ref. **[Kankia & Marky, 2001; Nagatoishi et al., 2011]**), absorption spectroscopy (see, for instance, references **[Kankia, 2004; Olsen et al., 2006a; Pasternak et al., 2011]**), calorimetry (DSC and ITC; see, for instance, references **[Haq et al., 1999; Majhi et al., 2007; Smirnov & Shafer, 2000a]**), gel electrophoresis **[Fialová et al., 2006; Tang & Shafer, 2006]**, NMR (see, for instance, references **[Mao & Gmeiner, 2005; Marathias & Bolton, 2000; Trajkovski et al., 2009]**), X-ray **[Krauss et al., 2012]**, fluorescence **[Baldrich & O'Sullivan, 2005; Kumar & Maiti, 2004]**, Raman scattering **[Pagba et al., 2010]**, infrared spectroscopy **[Mondragon-Sanchez et al., 2004]**, surface plasmon resonance **[del Toro et al., 2008]**, atomic force microscopy

**[Basnar et al., 2006]**, mass spectrometry **[Hong et al., 2010; Vairamani & Gross, 2003]** and molecular modeling **[Pagano et al., 2008; Reshetnikov et al., 2011]**.

Most studies are consistent with intramolecular folding of TBA quadruplex **[Macaya et al., 1993]**. However, the discrepancy between calorimetric and van't Hoff enthalpies **[Majhi et al., 2007; Smirnov & Shafer, 2000a]** and retarded electrophoretic mobility of TBA in native PAGE electrophoretic gels can be interpreted as a sign of intermolecularity (bimolecularity) of TBA **[Fialová et al., 2006]**. In our study, we attempt to reconcile these apparently contradictory results regarding the molecularity of TBA in aqueous solutions by probing the structure of TBA and its complexes with cationic porphyrins by means of several spectroscopic techniques including Raman, CD and NMR spectroscopy. We included in our investigation eight structurally related cationic porphyrins depicted in figure 40 and ten DNA oligonucleotide sequences (TBA and its nine related sequences) shown in table 7.

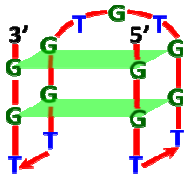



**Figure 49.** The scheme of the structure of TBA (**t**hrombin **b**inding **a**ptamer).

| Oligonucleotide designation | Oligonucleotide Sequence | Description |
|---|---|---|
| TBA | GGTTGGTGTGGTTGG | TBA – original sequence |
| TGT/T3 | GGTTGGT**T**TGGTTGG | TGT loop replaced by T3 |
| TGT/T4 | GGTTGGT**TT**TGGTTGG | TGT loop replaced by T4 |
| TGT/T5 | GGTTGGT**TTT**TGGTTGG | TGT loop replaced by T5 |
| TGT/T6 | GGTTGGT**TTTT**TGGTTGG | TGT loop replaced by T6 |
| T+ | GGT**T**TGGT**T**GT**T**GGT**T**TGG | T-regions extended by one T |
| T2+ | GGT**TT**TGGT**TT**GT**TT**GGT**TT**TGG | T-regions extended by two T |
| T2/T3 | GGT**T**TGGTGTGGT**T**TGG | T2 loops replaced by T3 |
| T2/T4 | GGT**TT**TGGTGTGGT**TT**TGG | T2 loops replaced by T4 |
| GTAGGT – TBA | **GTAGGT**GGTTGGTGTGGTTGG | 6 bases added to 5' end |

**Table 7.** The sequences of TBA and its related oligonucleotides used in the study. The difference of each TBA-related sequence with respect to the base TBA sequence is marked in red bold font.

Melting temperature should not depend on DNA concentration for intramolecular folding. As a test of intramolecularity of TBA, we therefore decided

to carry out a series of thermal melting experiments and to determine the stability of TBA as a function of oligonucleotide strand concentration. Our experiments were conducted over the range of TBA concentrations spanning more than 3 orders of magnitude (from ~ 3 μM to 11 mM in strands) through the combination of absorption, CD and Raman spectroscopy. Thermodynamic characterization of the TBA quadruplex is uncomplicated due to its reversible folding (see figure 43), even in the presence of porphyrins (see figures 42 and 47). We easily estimated the values of thermodynamic parameters and melting temperatures for TBA and the sequentially related oligonucleotides described in table 7. For details on the evaluations of thermodynamic parameters refer to the section 2.2.4. The all calculated thermodynamic data assuming intramolecular or bimolecular model are summarized in tables 8-12. In agreement with others, we found that the melting temperature of 50 ± 2°C for TBA stabilized by KCl is largely independent of DNA concentration (compare tables 8, 11 and 12). In fact, the thermal stability of TBA is sensitive to ionic strength and type of cation rather than to DNA concentration (see table 9 and 10) as the melting temperature is only ~ 30°C in 1 mM $K^+$ and ~20°C in 100 mM $Na^+$.

The van't Hoff parameters $\Delta H$ ~ 180 kJ/mol, $\Delta S$ ~ 550 J/K/mol determined under the assumption of intramolecular TBA folding and calculated from the absorption spectra recorded in 100 mM $K^+$ at relatively low strand concentrations (about 3 μM) are consistent with others **[Fialová et al., 2006; Majhi et al., 2007; Olsen et al., 2006b; Pasternak et al., 2011; Sacca et al., 2005; Smirnov & Shafer, 2000a; Tang & Shafer, 2006]**. However, the values of van't Hoff parameters determined in the same way from CD and Raman data obtained at much higher strand concentrations are different (compare tables 8, 11 and 12). In other words, our thermal data are incompatible with the unique pair of van't Hoff parameters $\Delta H$, $\Delta S$ for intramolecular transition. This interesting finding is best reflected in figure 50. One can see that the sigmoidal shape (steepness) of $\boldsymbol{V_2}$ profiles that best reflect the thermal stability of the TBA structure depends on DNA concentration.

It was proposed, based on retarded mobility of TBA in native PAGE gels, that TBA in aqueous solution forms a bimolecular quadruplex **[Fialová et al., 2006]**. The idea of bimolecularity of TBA seems to be interesting in view of its incompatibility with intramolecular model. However, again, absorption, CD and Raman thermal data

are inconsistent with bimolecularity (compare tables 8, 11 and 12). Furthermore, diffusion-ordered (DOSY) NMR study provided the value of ~ 130 $\mu m^2 s^{-1}$ for diffusion coefficient of TBA (see table 13). This value is the same (within experimental error) as that for the reference sequence T2/T3 (just adding one T in the edge-wise TT loops of the aptamer; see table 7), which migrates on native PAGE gels as expected for intramolecular rather than bimolecular species (as discussed later). Hence, the structure of the TBA quadruplex in solution cannot be described without contradiction in terms of either the intramolecular or bimolecular fold.

In effort to explain the spectroscopic and PAGE data, we proposed a more complicated model of TBA consisting in dimerization of two intramolecular TBA quadruplexes (the model was introduced in the section 2.2.4; pages 89-90). The idea of intermolecular associates of guanine quadruplexes is not new and has been recently proposed also for other quadruplexes **[Collie et al., 2010; Do et al., 2011; Haider et al., 2008; Kato et al., 2005; Trajkovski et al., 2012]**. On the one hand, a dimeric structure could mimic the properties of true bimolecular quadruplexes (where guanines in the tetrads are shared by two oligonucleotides) such as their retarded electrophoretic mobility but on the other hand the monomeric quadruplex constituents of a putative quadruplex dimer should largely retain their structural and thermodynamic properties. The thermal data for TBA obtained from all our spectroscopic and calorimetric measurements seems to be entirely consistent with this combined intramolecular/bimolecular model (see figure 51). Unfortunately, it was not possible to determine without ambiguity the values of van't Hoff enthalpies and entropies for both intramolecular transition and dimerization (bimolecular transition) because the fit of the suggested dimeric model to spectroscopic data produced many local minima corresponding to mutually different values of thermodynamic parameters.

The exact concentration profiles for three components of the system (intramolecular quadruplex, dimeric structure, unordered single strand) could be calculated once we knew the corresponding van't Hoff thermodynamic parameters (a possible scenario is depicted in figure 52). The dimeric model of TBA implies that at some lower temperatures (before complete denaturation of the TBA structure) a dimer of TBA intramolecular quadruplexes prevails over intramolecular TBA quadruplex species. The relative amount of the putative dimeric structure necessarily increases with increasing TBA concentration. In conformity with the previous work

**[Fialová et al., 2006]**, we believe that the newly proposed dimeric structure of TBA (at room temperature or lower) is present even at low DNA concentrations commonly encountered in absorptions and CD measurements.

| Oligonucleotide | ΔH [kJ/mol] | | Stand. Error[1] [kJ/mol] | | ΔS [J/K/mol] | | Stand. Error[1] [J/K/mol] | | Conc.[2] [μM] | $T_M$ [°C][3] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intra | Bimo | Intra | Bimo | Intra | Bimo | Intra | Bimo | | Intra | Bimo |
| TBA – I | -178 | -278 | 4 | 4 | -548 | -719 | 12 | 14 | 3.2 | 51.8 | 51.3 |
| TBA – II | -184 | -279 | 3 | 6 | -565 | -722 | 11 | 18 | 2.9 | 51.6 | 51.0 |
| TGT/T3 | -173 | -249 | 2 | 2 | -541 | -640 | 6 | 8 | 2.9 | 46.6 | 45.7 |
| TGT/T4 | -170 | -230 | 3 | 4 | -543 | -597 | 9 | 14 | 2.8 | 40.1 | 38.9 |
| TGT/T5 | -168 | -219 | 1 | 2 | -542 | -571 | 3 | 6 | 2.8 | 36.3 | 34.7 |
| TGT/T6 | -160 | -218 | 4 | 6 | -522 | -575 | 11 | 19 | 3.0 | 33.9 | 32.8 |
| TBA T+ | -207 | -302 | 2 | 4 | -648 | -805 | 5 | 14 | 2.9 | 47.1 | 46.5 |
| TBA T2+ | -216 | -308 | 4 | 7 | -700 | -859 | 14 | 23 | 3.0 | 35.6 | 35.2 |
| T2/T3 | -207 | -320 | 5 | 5 | -635 | -843 | 16 | 16 | 2.9 | 53.6 | 52.9 |
| T2/T4 | -217 | -317 | 3 | 6 | -692 | -870 | 8 | 18 | 2.8 | 41.1 | 40.8 |
| GTAGGT – TBA | -164 | -226 | 2 | 2 | -521 | -579 | 6 | 6 | 2.8 | 42.6 | 41.4 |
| 1:1 TBA duplex | - | -413 | - | 6 | - | -1093 | - | 17 | 3.0 | - | 60.0 |

**Table 8.** Van't Hoff thermodynamic parameters for TBA (2 independent experiments TBA-I and TBA-II) and its related sequences in 100 mM KCl determined from absorption measurements and corresponding to intramolecular (Intra) or bimolecular model (Bimo).

| Oligonucleotide | ΔH [kJ/mol] | | Stand. Error[1] [kJ/mol] | | ΔS [J/K/mol] | | Stand. Error[1] [J/K/mol] | | Conc.[2] [μM] | $T_M$ [°C][3] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intra | Bimo | Intra | Bimo | Intra | Bimo | Intra | Bimo | | Intra | Bimo |
| TBA – I | -97 | -118 | 1 | 2 | -329 | -268 | 4 | 6 | 2.8 | 23.3 | 17.2 |
| TBA – II | -92 | -109 | 1 | 2 | -309 | -238 | 4 | 5 | 3.2 | 22.8 | 15.8 |
| TGT/T3 | -102 | -122 | 1 | 2 | -348 | -284 | 3 | 5 | 2.9 | 20.5 | 14.4 |
| TGT/T4 | -159 | -222 | 2 | 3 | -508 | -574 | 7 | 10 | 2.8 | 39.3 | 38.3 |
| TGT/T5 | -179 | -257 | 2 | 5 | -568 | -679 | 5 | 14 | 2.8 | 41.6 | 40.7 |
| TGT/T6 | -193 | -281 | 5 | 8 | -614 | -754 | 15 | 25 | 3.0 | 41.4 | 41.0 |
| TBA T+ | -190 | -278 | 2 | 4 | -596 | -736 | 7 | 11 | 2.8 | 45.2 | 44.6 |
| TBA T2+ | -149 | -182 | 2 | 4 | -506 | -482 | 8 | 13 | 3.0 | 21.9 | 19.0 |
| T2/T3 | -124 | -164 | 2 | 2 | -409 | -407 | 5 | 5 | 2.8 | 29.6 | 27.4 |
| T2/T4 | -172 | -221 | 2 | 3 | -576 | -605 | 5 | 9 | 2.8 | 25.5 | 24.0 |
| GTAGGT – TBA | -115 | -144 | 2 | 3 | -384 | -347 | 5 | 8 | 2.7 | 27.2 | 23.1 |

**Table 9.** Van't Hoff thermodynamic parameters for TBA (2 independent experiments TBA-I and TBA-II) and its related sequences in 100 mM NaCl determined from absorption measurements and corresponding to intramolecular (Intra) or bimolecular model (Bimo).

| Salt | ΔH [kJ/mol] | | Stand. Error[1] [kJ/mol] | | ΔS [J/K/mol] | | Stand. Error[1] [J/K/mol] | | Conc.[2] [μM] | $T_M$ [°C][3] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intra | Bimo | Intra | Bimo | Intra | Bimo | Intra | Bimo | | Intra | Bimo |
| 1 mM K+ | -158 | -215 | 4 | 7 | -514 | -567 | 12 | 22 | 3.2 | 33.4 | 32.2 |
| 10 mM K+ | -168 | -245 | 4 | 7 | -530 | -637 | 12 | 22 | 3.1 | 43.6 | 42.9 |
| 1 mM Na+ | -106 | -127 | 2 | 3 | -370 | -315 | 6 | 8 | 3.1 | 13.1 | 6.2 |
| 10 mM Na+ | -100 | -115 | 1 | 1 | -347 | -274 | 3 | 4 | 3.1 | 14.0 | 6.7 |
| only buffer | -104 | -135 | 2 | 3 | -364 | -342 | 9 | 10 | 3.1 | 13.7 | 7.3 |

**Table 10.** Van't Hoff thermodynamic parameters for TBA in various salts determined from absorption measurements and corresponding to intramolecular (Intra) or bimolecular model (Bimo).

| Oligonucleotide | ΔH [kJ/mol] | | Stand. Error[1] [kJ/mol] | | ΔS [J/K/mol] | | Stand. Error[1] [J/K/mol] | | Conc.[2] [mM] | $T_M$ [°C][3] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intra | Bimo | Intra | Bimo | Intra | Bimo | Intra | Bimo | | Intra | Bimo |
| TBA | -166 | -239 | 11 | 16 | -513 | -650 | 33 | 49 | 0.87 | 50.5 | 48.8 |
| TGT/T3 | -135 | -216 | 16 | 24 | -419 | -587 | 52 | 76 | 1.6 | 48.6 | 46.5 |
| TGT/T4 | -187 | -285 | 18 | 27 | -596 | -825 | 58 | 88 | 1.6 | 40.0 | 39.2 |
| TGT/T5 | -167 | -240 | 9 | 15 | -539 | -689 | 29 | 47 | 1.4 | 37.2 | 36.3 |
| TGT/T6 | -164 | -227 | 11 | 17 | -538 | -655 | 34 | 55 | 0.90 | 32.5 | 31.4 |
| T+ | -212 | -303 | 37 | 51 | -667 | -863 | 118 | 159 | 0.82 | 44.4 | 43.9 |
| T2+ | -249 | -367 | 27 | 17 | -809 | -1100 | 86 | 54 | 0.58 | 34.3 | 34.1 |
| T2/T3 | -185 | -277 | 22 | 32 | -570 | -764 | 69 | 98 | 1.1 | 52.0 | 51.2 |
| T2/T4 | -200 | -296 | 14 | 21 | -642 | -859 | 44 | 68 | 1.0 | 38.7 | 38.0 |
| GTAGGT – TBA | -151 | -212 | 12 | 19 | -478 | -582 | 39 | 61 | 0.70 | 42.3 | 41.1 |

**Table 11.** Van't Hoff thermodynamic parameters for TBA and its related sequences in 100 mM KCl determined from CD measurements and corresponding to intramolecular (Intra) or bimolecular model (Bimo).

| Salt | ΔH [kJ/mol] | | Stand. Error[1] [kJ/mol] | | ΔS [J/K/mol] | | Stand. Error[1] [J/K/mol] | | Conc.[2] [mM] | $T_M$ [°C][3] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intra | Bimo | Intra | Bimo | Intra | Bimo | Intra | Bimo | | Intra | Bimo |
| 100 mM K+ | -119 | -180 | 2 | 3 | -366 | -486 | 6 | 10 | 11.3 | 52.6 | 50.4 |
| 100 mM K+ | -139 | -215 | 7 | 10 | -430 | -580 | 22 | 30 | 1.33 | 50.8 | 48.8 |
| 100 mM K+ | -138 | -203 | 5 | 9 | -425 | -540 | 16 | 26 | 1.33 | 51.9 | 50.5 |
| 100 mM K+ | -149 | -219 | 5 | 7 | -460 | -589 | 16 | 21 | 1.33 | 50.8 | 49.8 |
| 100 mM Na+ | -112 | -133 | 7 | 10 | -374 | -365 | 21 | 31 | 1.33 | 25.1 | 20.8 |
| no salt added | -91 | -111 | 4 | 6 | -303 | -311 | 13 | 20 | 19.7 | 28.3 | 22.1 |

**Table 12.** Van't Hoff thermodynamic parameters for TBA (at indicated salt conditions) determined from Raman measurements and corresponding to intramolecular (Intra) or bimolecular model (Bimo).

Notes for tables 8-12 with thermodynamic data:

[1]The standard error values obtained from the least square fit are more likely underestimated.

[2] DNA concentrations are in oligonucleotides.

[3]The melting temperatures for intramolecular and bimolecular model are calculated in conformity with the determined values of $\Delta H$, $\Delta S$. Their values are not so sensitive to the choice of thermodynamic model.

The suggested dimeric structure is spectroscopically (absorption, CD, Raman) or calorimetrically (see figure 53) hardly distinguishable from its monomeric quadruplex constituents. Consequently, we tried to differentiate between both monomers and dimers indirectly through their interaction with cationic porphyrins (see figure 40 for the list and schemes of the used porphyrins). Porphyrins may interfere with the anticoagulant activity of the TBA aptamer, which is of biological importance **[Joachimi et al., 2007]**. The relation between the sequence of TBA and its propensity to dimerize was investigated by including various TBA related sequences in the study (see table 7 for the list and description of these sequences).

The CD spectrum of 1:1 TBA-CuTMPyP4 mixture exhibits a relatively strong induced CD signal of the bisignate shape (-419 nm / +432 nm) within the spectral region of the porphyrin Soret band (~424 nm). This characteristic CD signal is also observed for the TBA mutant sequences (listed in table 7) having variously extended thymine loops but it is much weaker (except for the modification TGT/T3 having one guanine in the central loop replaced by thymine) and it disappears for longer edge-wise loops (see figure 54). By contrast, the CD spectra of 1:1 mixtures of TBA or its mutant sequences with free base porphyrin $H_2$TMPyP4 show no bisignate but only weak positive or zero induced CD signal (see figure 55), which is frequently considered as a sign of external groove binding of porphyrins to DNA **[Balaz et al., 2005; Pasternack, et al. 1983]**. The same is true for the positional isomers of CuTMPyP4 (CuTMPyP2 and CuTMPyP3) and $Zn^{2+}$ form of $H_2$TMPyP4 (see figures 56, 57 and 58). CuTMAP, another copper porphyrin, is the only other porphyrin included in our study (apart from CuTMPyP4) that provides a relatively strong bisignate signal (it has the reverse polarity compared to CuTMPyP4) upon the interaction with TBA (see figure 59). Non-metaled form of this porphyrin

($H_2$TMAP) and its $Zn^{2+}$ form (ZnTMAP) exhibit much weaker bisignate signals differing mutually in polarities (see supporting figures 60 and 61). The compilation of spectroscopic parameters for 1:1 mixtures of TBA with the above cationic porphyrins is tabulated for clarity in table 14.

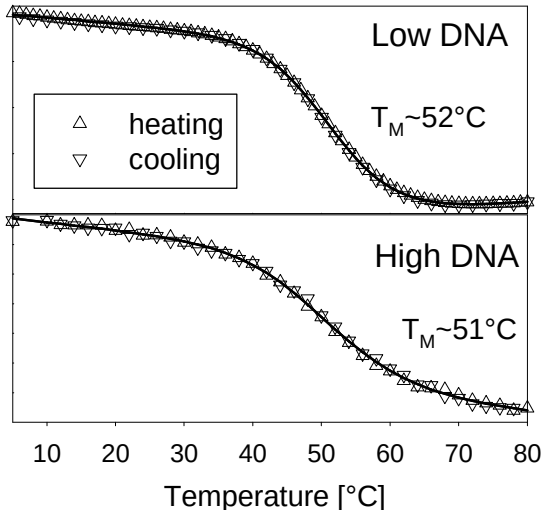



**Figure 50.** The effect of TBA concentration on the shape of $V_2$ thermal profiles. Heating and cooling profiles are superimposable at both concentrations indicating equilibrium conditions during the measurements. The upper part of the figure: $V_2$ coefficient profile from SVD of the absorption spectra measured at low DNA concentration (3.2 μM in strands). The lower part of the figure: $V_2$ coefficient profile from SVD of the Raman spectra measured at high DNA concentrations (1.3 mM in strands). The calculated melting temperatures considering intramolecular model are the same (within experimental error) at both concentrations. Nonetheless, the different steepness of $V_2$ coefficient curves at low and high DNA concentrations seems to contradict intramolecularity of TBA. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8.

It is unusual that the bisignate CD signal of CuTMPyP4 is still noticeable even at relatively low ratios of CuTMPyP4 to TBA (see figure 62). It completely disappears only at extremely low porphyrin loads (TBA/CuTMPyP4 > 10) when it turns to a positive band indicating external binding mode of CuTMPyP4 to TBA **[Balaz et al., 2005]**. The similar bisignate signal as for the TBA-CuTMPyP4 complex was observed for CuTMPyP4 bound to poly(dA) (-422 nm / +431 nm)

**[Pasternack et al., 1990]** or in the case of other porphyrin-quadruplex systems **[Murashima et al., 2008; Yamashita et al., 2005]**. It is generally attributed to self stacking of porphyrins externally bound to chiral matrix (polyoligonucleotide or polypeptide) (see for example references **[Lee et al., 2003; Pasternack et al., 1991; Uno et al., 2002]**). According to this interpretation, two or more CuTMPyP4 molecules should be crowding on one TBA quadruplex. However, there is no apparent reason for this. One would rather expect that two porphyrins repel each other due to their positive charges (formally 4+). Dimerization of TBA proposed above offers more plausible explanation of the CD spectra. It is conceivable that the TBA quadruplex have one or more binding sites for CuTMPyP4 coinciding with the regions implicated in dimerization. Consequently, porphyrins can come into close contact in a dimer and produce the bisignate CD signal. If a dimer split up into two intramolecular quadruplexes (their structure and the complex with CuTMPyP4 are not impaired), the bisignate CD signal disappears. Our dimeric model necessarily assumes that the structure of its constituents is not considerably affected and remains antiparallel. The antiparallel CD spectrum of a TBA dimer has ellipticity probably slightly higher than that for two isolated intramolecular TBA quadruplexes

| Method | FIT | | GIFA | |
|---|---|---|---|---|
| T [K] | TBA | T2/T3 | TBA | T2/T3 |
| 298.15 | 131 ± 4 | 130 ± 4 | 125 ± 6 | 126 ± 6 |
| 275.15 | 69 ± 5 | 70 ± 2 | 65.5 ± 6 | 66 ± 6 |

**Table 13.** The comparison of diffusion coefficients determined by NMR for TBA and T2/T3 (see table 7 for oligonucleotide sequences). The coefficients were obtained by either direct fitting ***[Stejskal & Tanner, 1965]*** or by maximum entropy analysis by program Gifa 4 ***[Pons et al., 1996]***. The fit of exponential Stejskal-Tanner equation to data at 298.15 K and 275.15 K gives identical diffusion coefficients for both oligonucleotides at each temperature. Inverse Laplace transform with maximum entropy provides the same diffusion coefficients for both samples. Consequently, no oligomeric species were found at either temperature. Theoretical hydro NMR calculation gives a value of diffusion coefficient of about 139 $um^2s^{-1}$ at 298 K, which is in good correlation with experimental results.

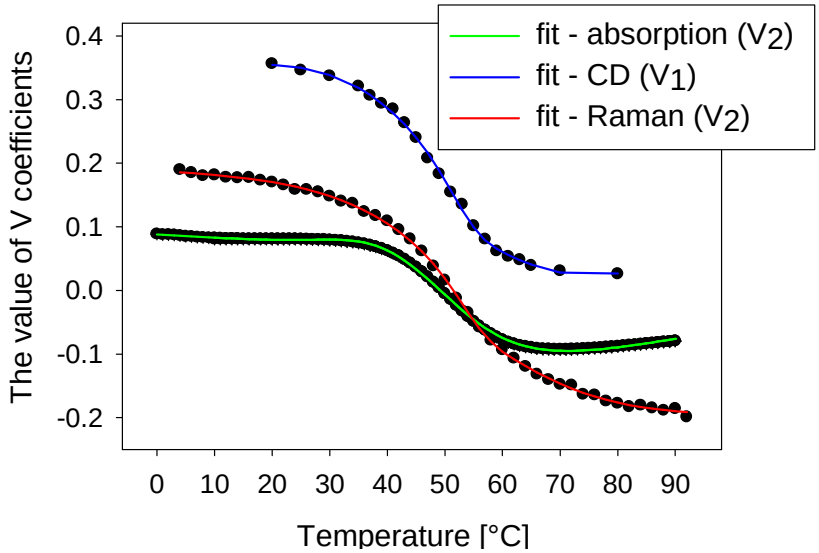



**Figure 51.** Joint fit of the combined intramolecular/bimolecular model to three coefficients profiles originating from SVD of absorption ($\boldsymbol{V_2}$ coefficient), CD ($\boldsymbol{V_1}$ coefficient) and Raman spectra ($\boldsymbol{V_2}$ coefficient) of TBA stabilized by 100 mM KCl.

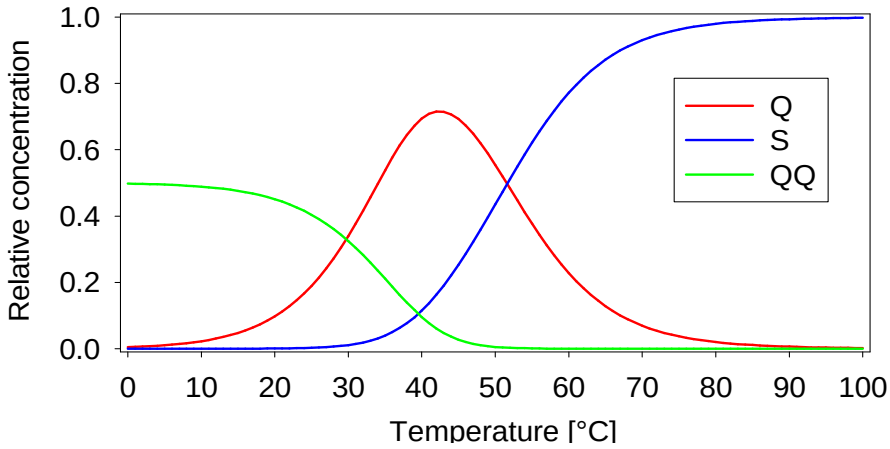



**Figure 52.** The relative concentration of intramolecular TBA quadruplex (Q), TBA dimeric structure (QQ) and unordered TBA single-stranded structure (S). The concentration profiles are calculated according relation (2.29) for 5μM total DNA strand concentration considering the van't Hoff thermodynamic parameters $\Delta H_1$ = -209 ± 12 kJ/mol, $\Delta S_1$ =-546 ± 36 kJ/K/mol (for dimerization) and $\Delta H_2$ = -131 ± 15 kJ/mol, $\Delta S_2$ =- 370 ± 44 kJ/mol (for intramolecular transition). The values correspond to one of the notable least square minimums of the joint fit in figure 51.

We speculate that thermal stabilities of the TBA dimer and TBA intramolecular quadruplex are different (see figure 52). We therefore investigated thermal denaturation of 1:1 TBA-CuTMPyP4 mixture in 100 mM KCl. The CD spectroscopy is most suitable for this purpose and provides the most convincing results since it separates best thermal changes in DNA and porphyrin. The heating profile derived from the ultraviolet part of the CD spectra reflects directly the stability of TBA quadruplex (see figure 63). The mid-point temperature of the transition is about 48°C, which means that CuTMPyP4 does not significantly affect the thermostability of TBA ($T_M$~50°C). The heating profile obtained from the visible part of the CD spectra reflects indirectly the stability of the suggested dimeric structure (see figure 64). Interestingly, the CD signal of CuTMPyP4 disappears at measurably lower temperatures (the mid-point temperature of the transition is ~39°C) than the signal of the G-quadruplex, which is perfectly in line with our hypothesis of dimeric structure of TBA (see figure 64). Furthermore, SVD analysis also identifies the second less statistically important CD component reflecting disordered TBA quadruplex (see figure 63).

Resonance Raman spectra of CuTMPyP4 provide the same conclusion about the thermal stability of the TBA-CuTMPyP4 complex as CD spectra, although their interpretation is less straightforward. The resonance Raman spectra of free CutTMPyP4 are quite different from the spectra of CutTPyP4 in complex with TBA (at low nucleoside concentration of 4 μM) at molar ratio [CuTMPyP4]/[TBA] = 1. They indicate that CutTMPyP4 prefers some binding mode that resembles the intercalation or end-stacking (molecular dynamics simulations seem to be, however, more consistent with the outside binding mode; see below) **[Schneider et al., 1988]**. The intercalation is unlikely as it would easily disrupt the structure of TBA quadruplex (having only two tetrads).

SVD processing of the temperature-dependent Raman data of the TBA-CuTMPyP4 mixture reveals minute differences between the spectra measured in the presence of $K^+$, $Na^+$ and $Li^+$. It has been already shown in the section 3.3 that the resonance Raman spectra of the TBA-CuTMPyP4 mixture at temperatures above approximately 55°C do not depend on the nature of metal cation present in solution (see figure 47). Since G-quadruplex conformation stabilized by $K^+$ (not by $Na^+$ or $Li^+$) is still present at 55°C and disappears completely above 70°C (see figure 50), it necessarily implies that the TBA structure in the presence of KCl at temperatures 55-

70°C interacts with CuTMPyP4 in the similar way as with unordered single strands in the presence of less stabilizing salts NaCl and LiCl (TBA melts in 100 mM $Na^+$ at ~20°C and it is even less thermostable in 100 mM $Li^+$). This clearly reflects different G-quadruplex structures of TBA (supposedly dimer and monomer) observed at lower versus higher temperatures. The mid-point temperature for the disintegration of the proposed dimeric complex in 100 mM KCl is roughly estimated as 40°C (based on the sigmoid-like profile of $\boldsymbol{V_3}$ in figure 47) in conformity with the CD results (see figure 64).

| Porphyrin | Melting temperature [°C][1] | Soret shift [nm] at room temperature | Induced CD extrema [nm] at 5°C |
|---|---|---|---|
| $H_2$TMPyP4 | 51 | 12 | +432 |
| CuTMPyP2 | 52 | 1 | +416 |
| CuTMPyP3 | 52 | 4 | +424 |
| CuTMPyP4 | 48 | 5 | -419<br>+432 |
| ZnTMPyP4 | 56 | 5 | +441 |
| $H_2$TMAP | 51 | 5 | +409<br>-433 |
| CuTMAP | 44 | 6.5 | +415<br>-424 |
| ZnTMAP | 51 | 4.5 | -422<br>+429 |

**Table 14.** Characteristics of 1:1 mixtures of TBA with cationic porphyrins

Notes for table 14

[1]The melting temperatures of TBA in its 1:1 porphyrin mixtures. For the mixture of TBA with CuTMPyP4, the melting temperature was determined precisely using CD spectra of TBA. In other cases, the melting temperature was determined based on DNA absorption bands, and the values are subject to undefined error as absorption spectra of DNA are partly obscured by low-wavelength contribution from porphyrins.

[2]The shift of the Soret band of porphyrins induced by their binding to TBA.

The result of SVD analysis of the temperature-dependent absorption spectra within the Soret band of the same 1:1 TBA-CuTMPyP4 mixture in 100 mM KCl is remarkably similar to that of resonance Raman spectra, and again the melting transition of the suggested TBA dimeric complex is roughly about 40°C (see $\boldsymbol{V_3}$ profile in figure 42).

The CD spectra of TBA mutants having modified loop sequences (see table 7) stabilized by $K^+$ in complex with CuTMPyP4 and other porphyrins have a similar shape corresponding to the antiparallel folding of TBA itself (see figures 54-61). It means that the average antiparallel-like structure is similar for all mutant G-quadruplexes and that their loops primarily determine the interaction with porphyrins and dimerization. The effect of thymine loops on the structure and dimerization of TBA was investigated by PAGE. Interestingly, the results of PAGE in 100 mM KCl at 2°C show that only TBA and the closely related sequence TGT/T3 (just one guanine in the TGT loop of TBA replaced by T) comigrate with their heteroduplexes (see figure 65A). The other oligonucleotides migrate more like single species. PAGE under the same conditions (2°C) in 100 mM NaCl (see figure 65B) gives consistent results. Strictly speaking, the electrophoretic mobility of TBA is about the same as that of its heteroduplex in KCl but slightly higher in NaCl. TGT/T3 in NaCl comigrates with its heteroduplex but it is slightly faster in KCl.

No easy or definite correlation can be drawn between the PAGE mobility of the TBA mutant sequences and the intensity of the bisignate CD signal of CuTMPyP4. However, it seems that the strong bisignate signal detected for 1:1 TBA or TGT/T3 mixtures with CuTMPyP4 (see figure 54) is apparently linked to the bimolecular-like electrophoretic migration of these oligonucleotides (and therefore to the suggested dimerization). Consequently, it seems that the edge-wise TT loops are more critical than the central TGT loop in the formation of the TBA dimer since their elongation causes a more noticeable decrease in the bisignate CD amplitude (see figure 54). TBA (and TGT/T3) in 100 mM KCl at temperature 40°C migrates much faster than at 2°C (compare figure 65 with figure 66). Its mobility is somewhere between monomeric and dimeric species. The higher mobility of TBA at elevated temperatures may correlate with the disintegration of the suggested dimeric structure detected indirectly by CuTMPyP4 (see figure 64). However, it is not completely clear why the electrophoretic band of TBA is not blurred as a consequence of

equilibrium interplay between monomeric, dimeric and single stranded forms of TBA expected at temperatures near the melting transition.

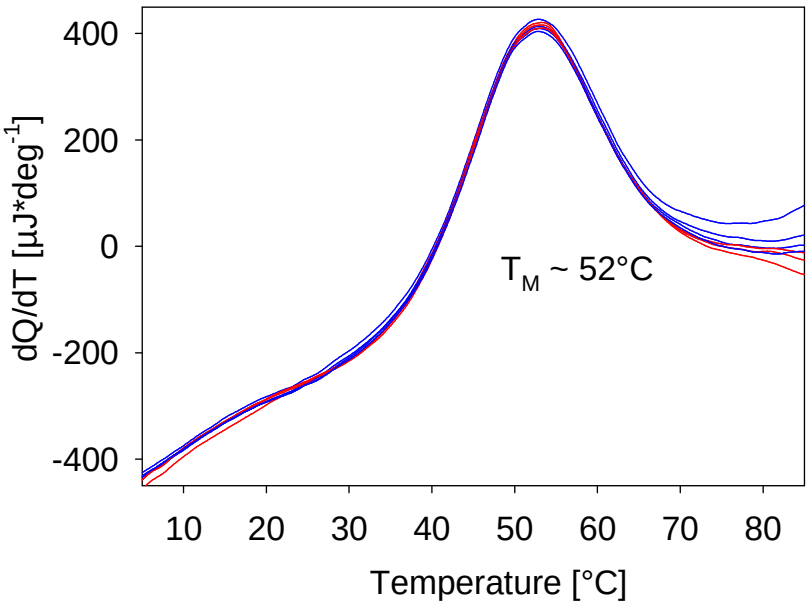



**Figure 53.** The results of DSC (differential scanning calorimetry) of TBA within 5-85°C. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 7.2, 0.3 mM DNA strand concentration. Heating and cooling scans are represented by red and blue curves, respectively. The DSC profile of buffer was subtracted. The rate of temperature change was 1°C/min for both heating and cooling. At such conditions, no hysteresis is visible. Note that only one broad band is visible. Consequently, deconvolution of the DSC band in the case of considering more than one transition is complicated or ambiguous.

The binding of CuTMPyP4 to TBA was nicely visualized by PAGE (see figure 67). Whereas the migration of TBA heteroduplex with added porphyrin (1:1) is practically the same as the migration of TBA heteroduplex alone, the migration of the TBA-CuTMPyP4 complex (1:1) is markedly slower than that of the TBA alone. Interestingly, the interaction of TBA with CuTMPyP4 or CuTMAP seems to be stronger and more specific than with other studied porphyrins, as can be demonstrated by slower and less blurred PAGE bands for the mixtures of TBA with CuTMPyP4 and CuTMAP than for the mixtures with the remaining porphyrins (see figure 68).

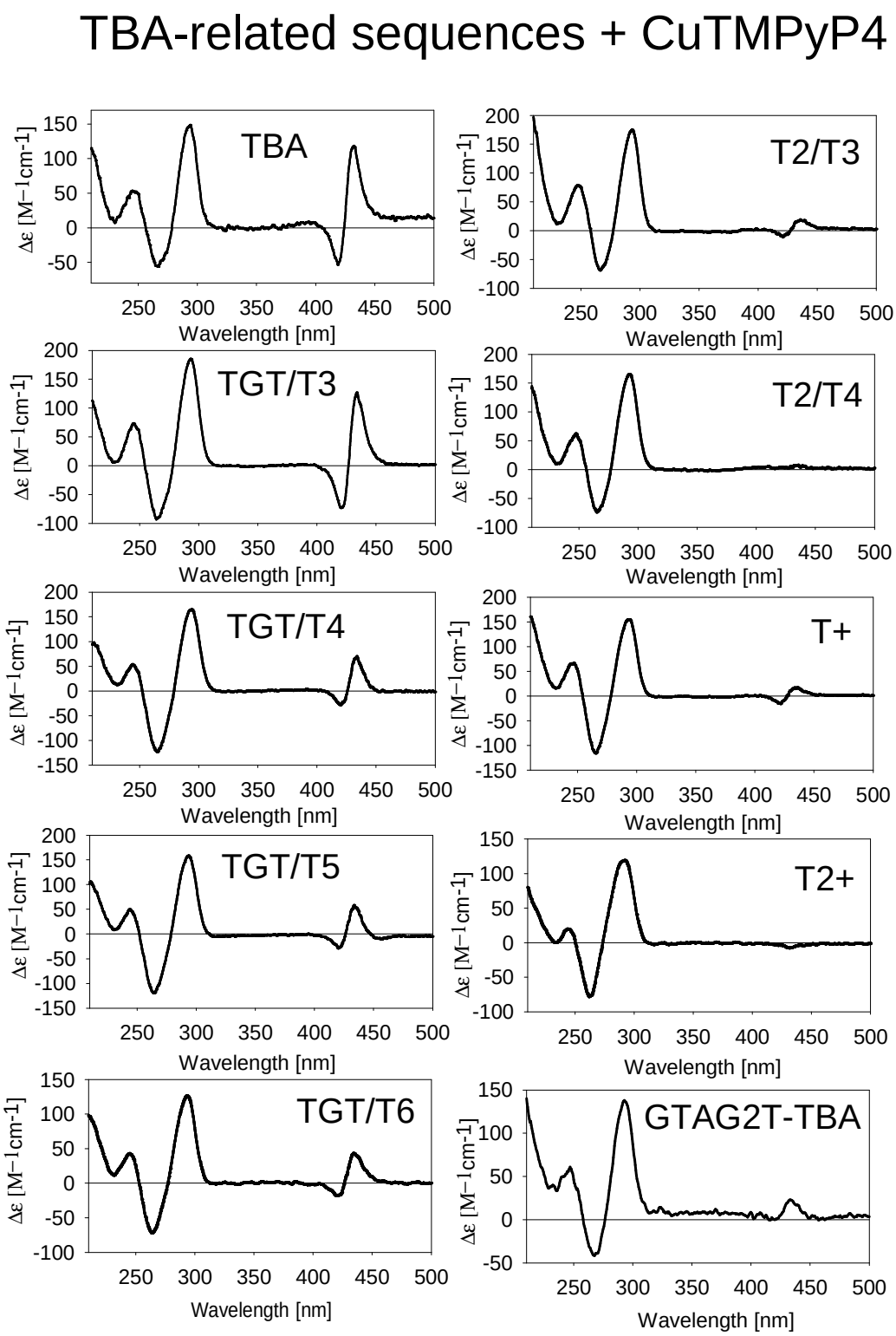



**Figure 54.** CD spectra profiles corresponding to mixtures of TBA related sequences with CuTMPyP4 (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

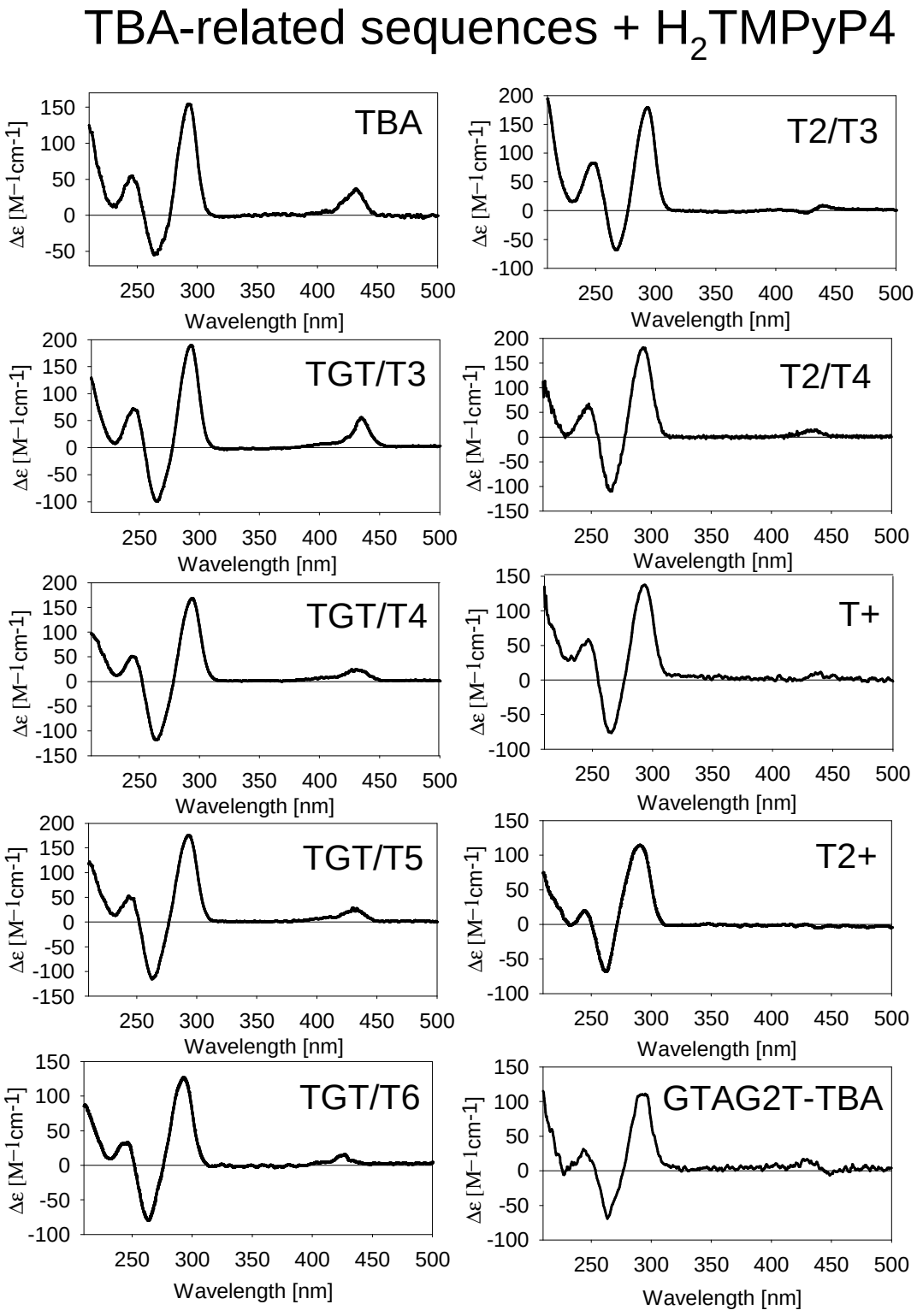



**Figure 55.** CD spectra profiles corresponding to mixtures of TBA related sequences with $H_2$TMPyP4 (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

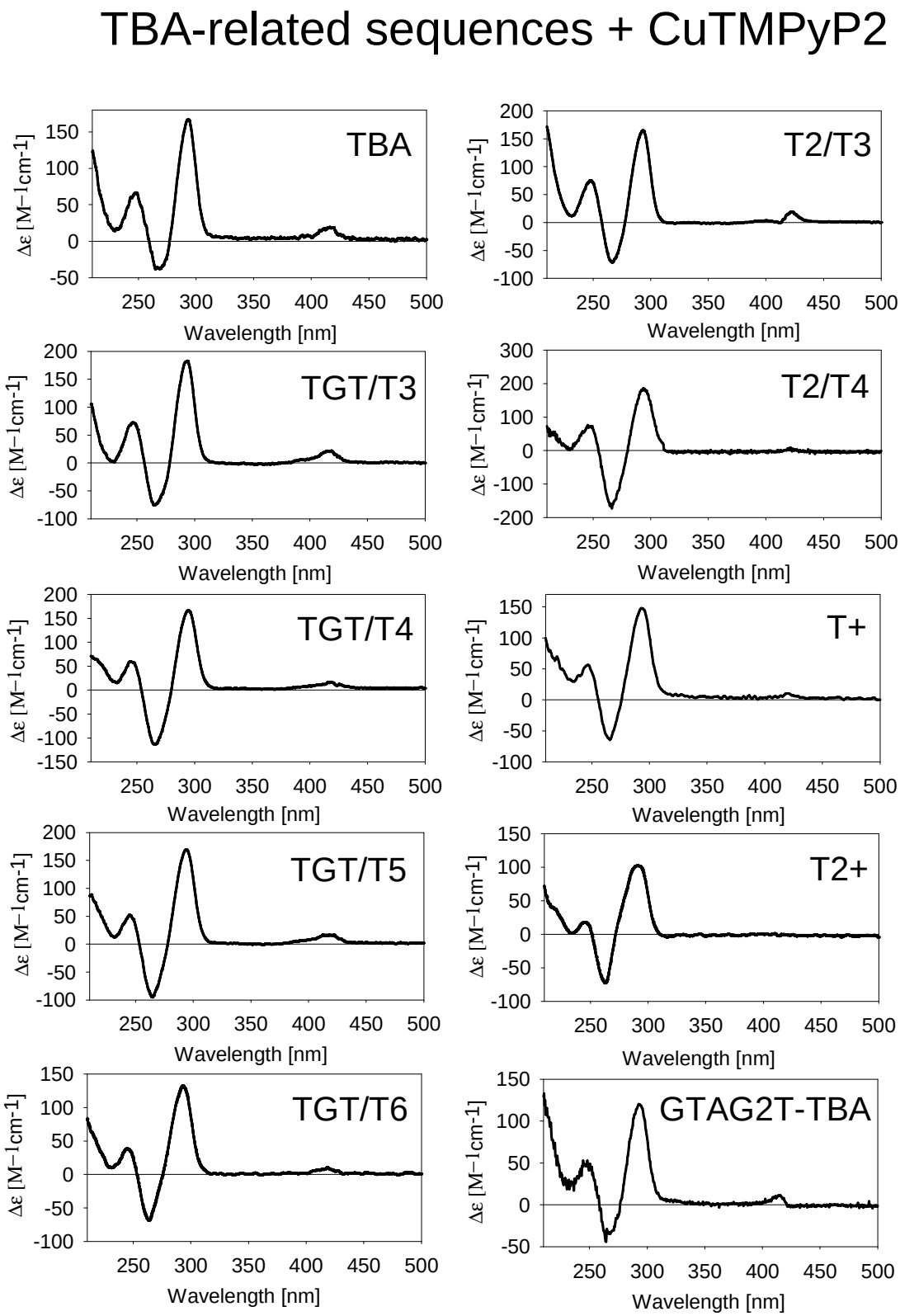



**Figure 56.** CD spectra profiles corresponding to mixtures of TBA related sequences with CuTMPyP2 (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

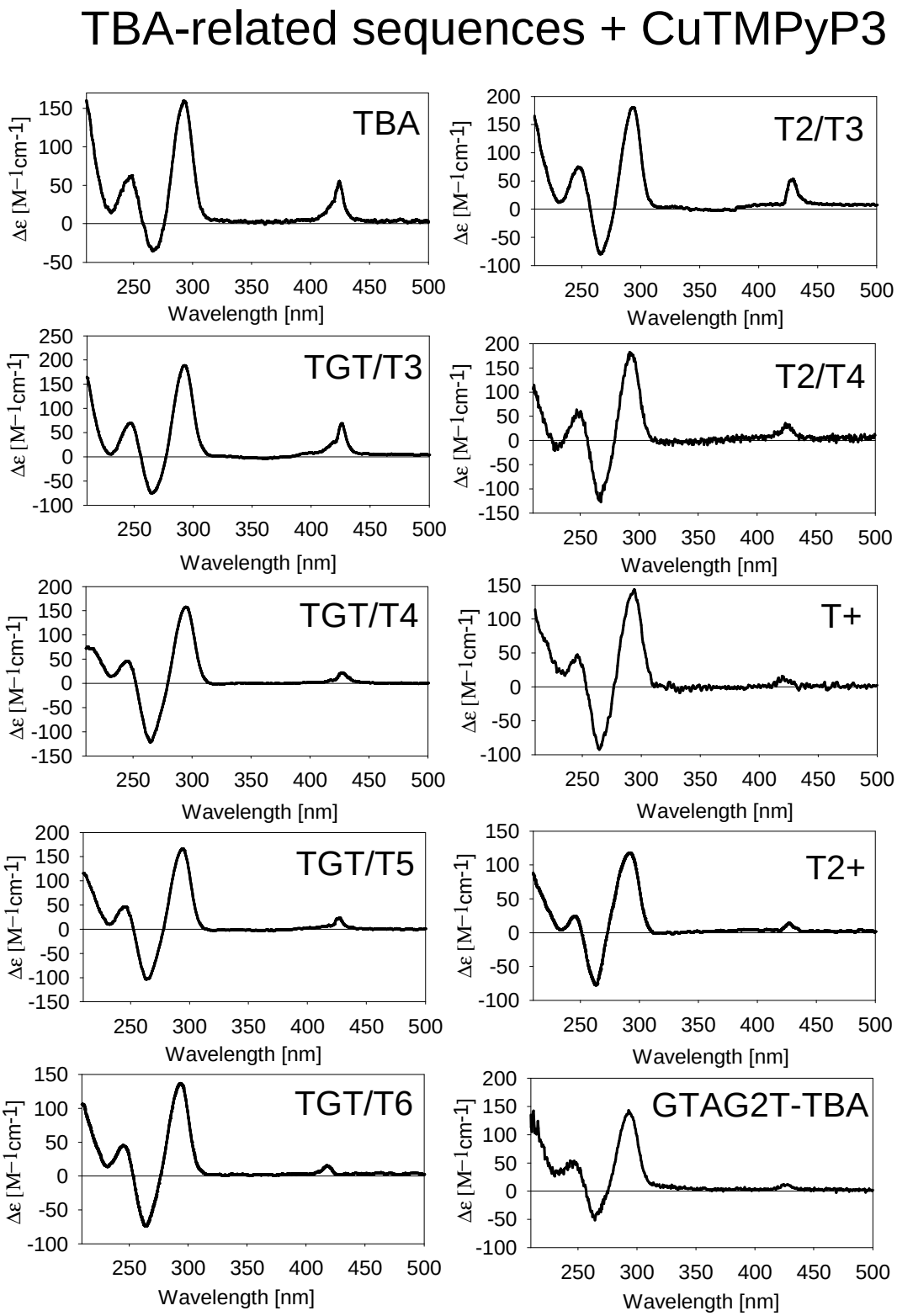



**Figure 57.** CD spectra profiles corresponding to mixtures of TBA related sequences with CuTMPyP3 (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

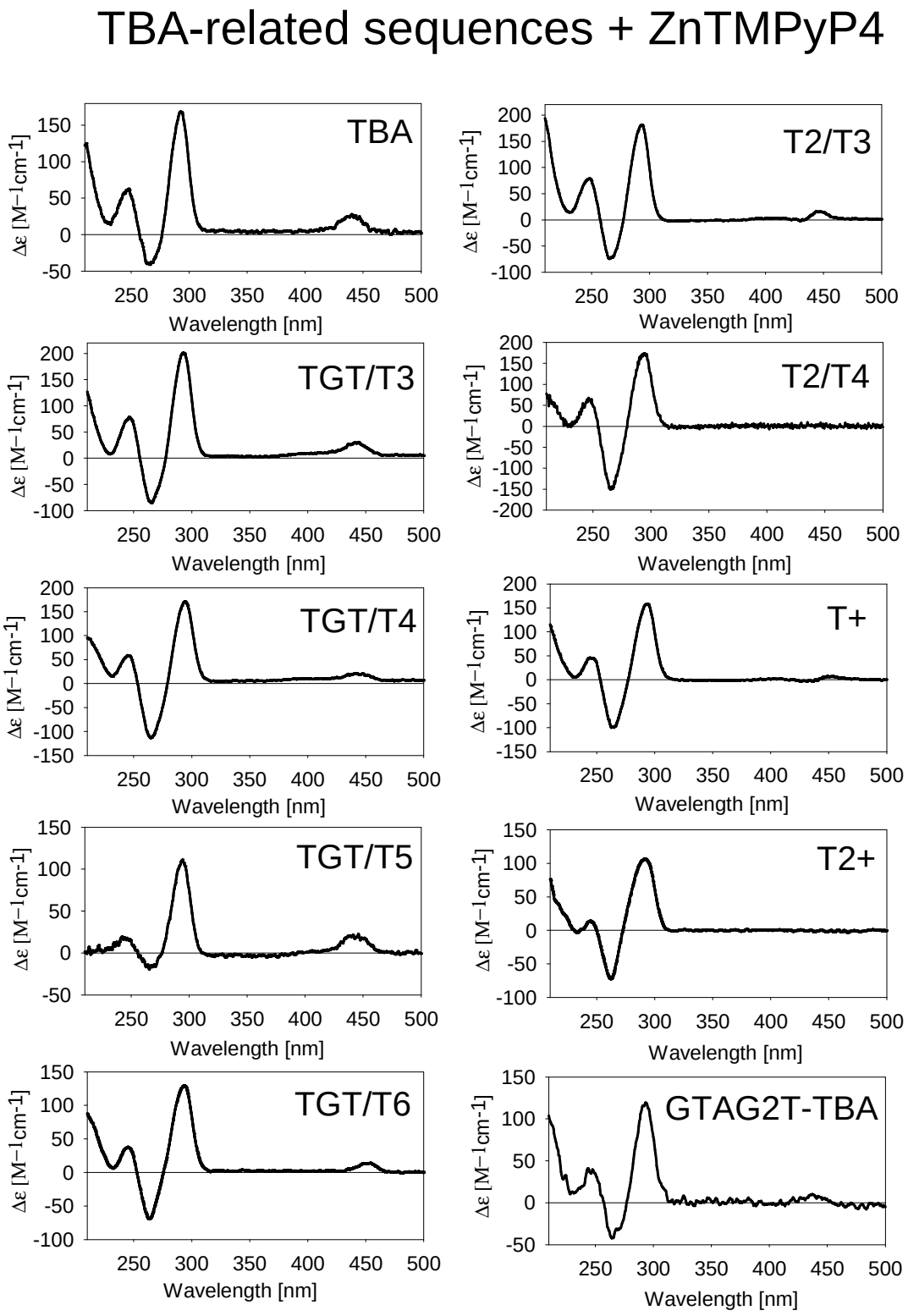



**Figure 58.** CD spectra profiles corresponding to mixtures of TBA related sequences with ZnTMPyP4 (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

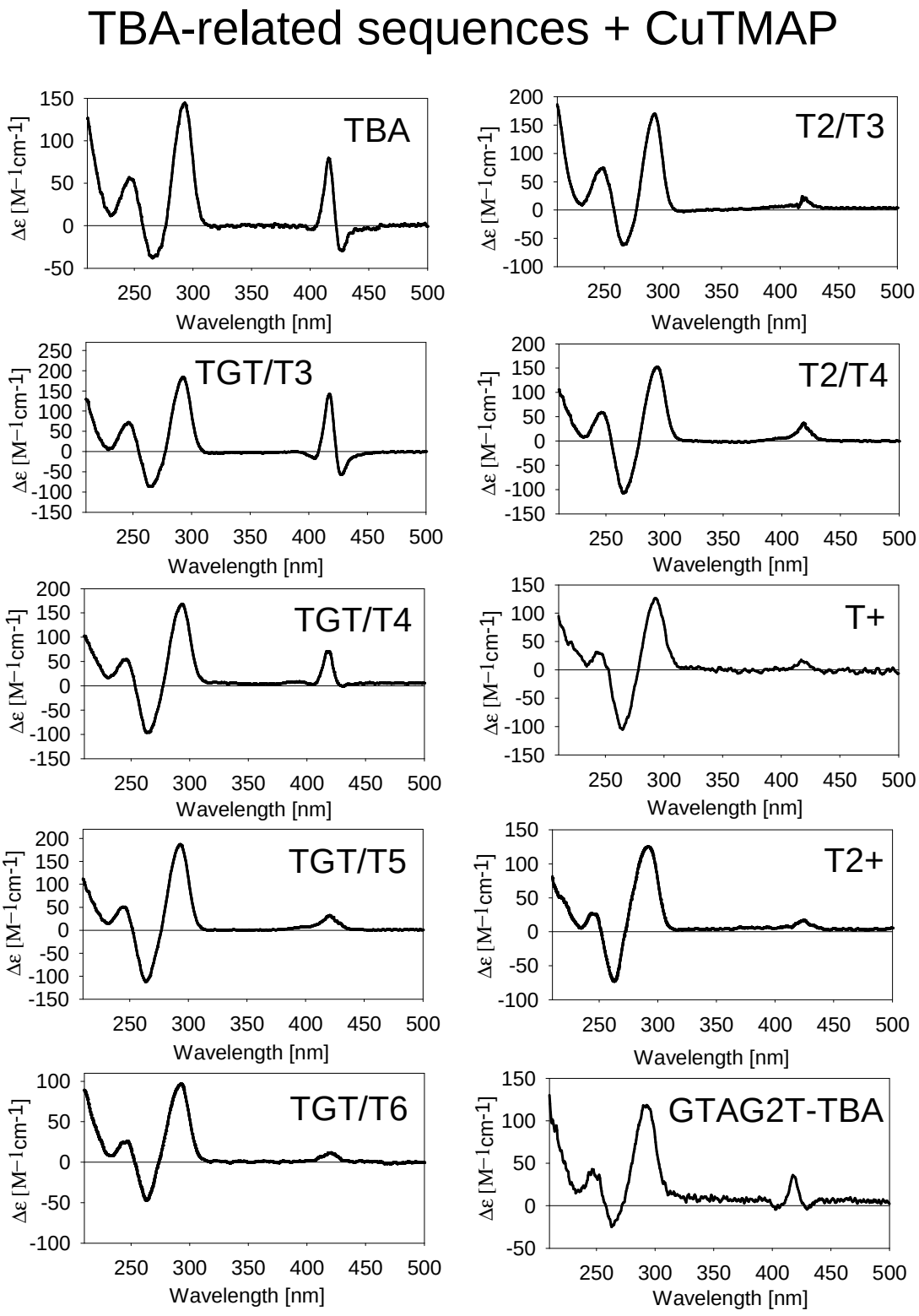



**Figure 59.** CD spectra profiles corresponding to mixtures of TBA related sequences with CuTMAP (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

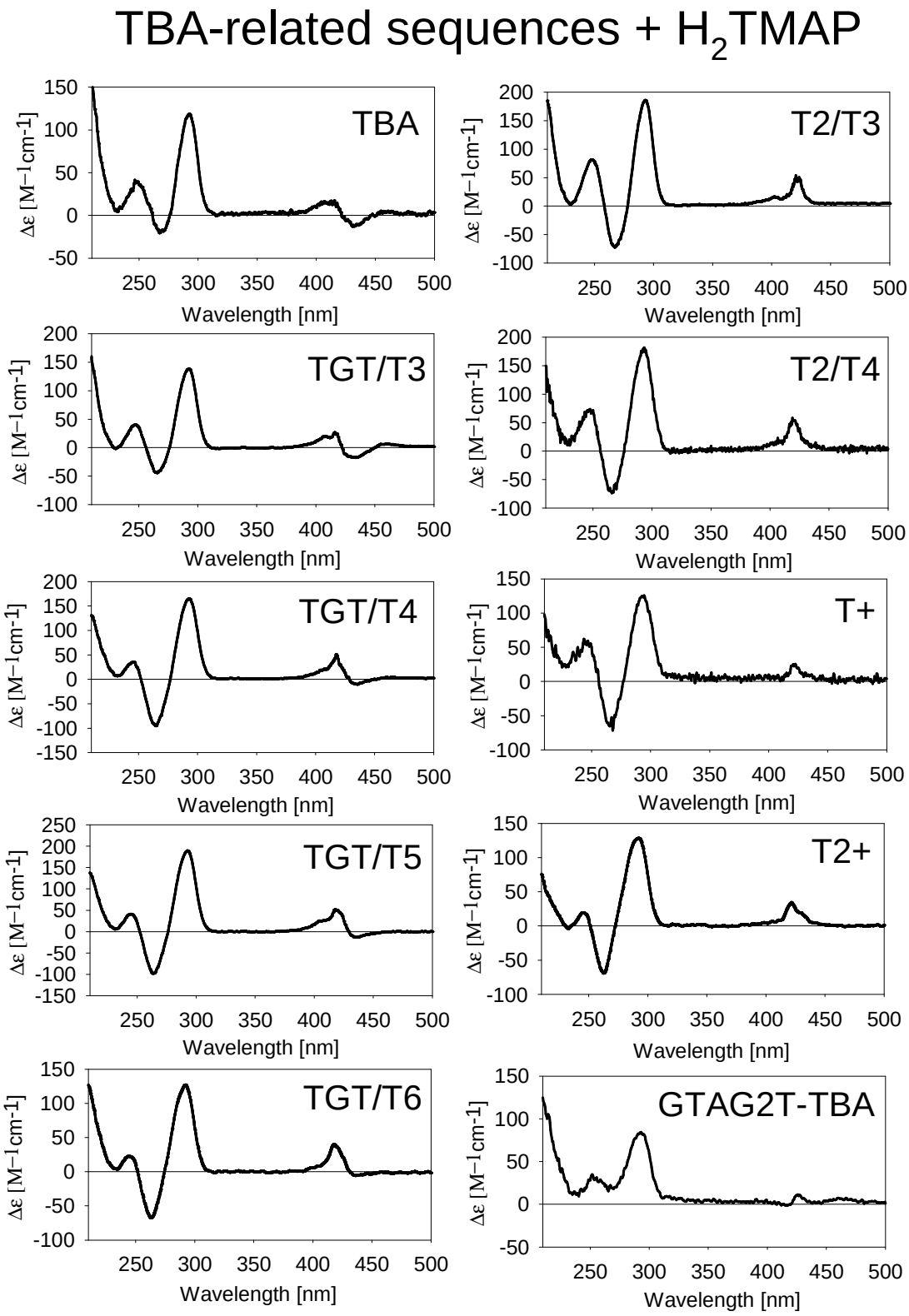



**Figure 60.** CD spectra profiles corresponding to mixtures of TBA related sequences with $H_2$TMAP (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration.

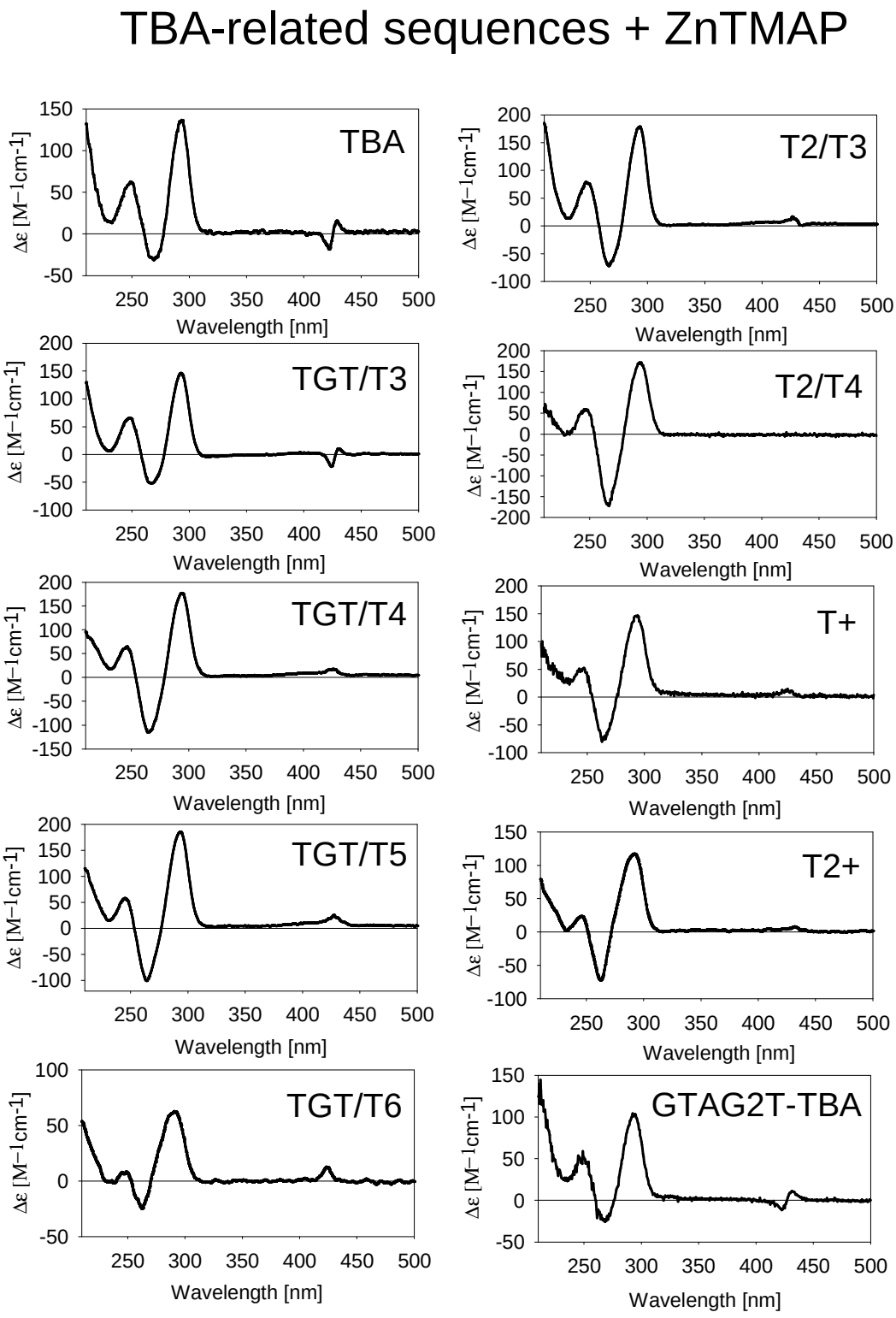



**Figure 61.** CD spectra profiles corresponding to mixtures of TBA related sequences with ZnTMAP (1:1). Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 5 μM DNA strand concentration

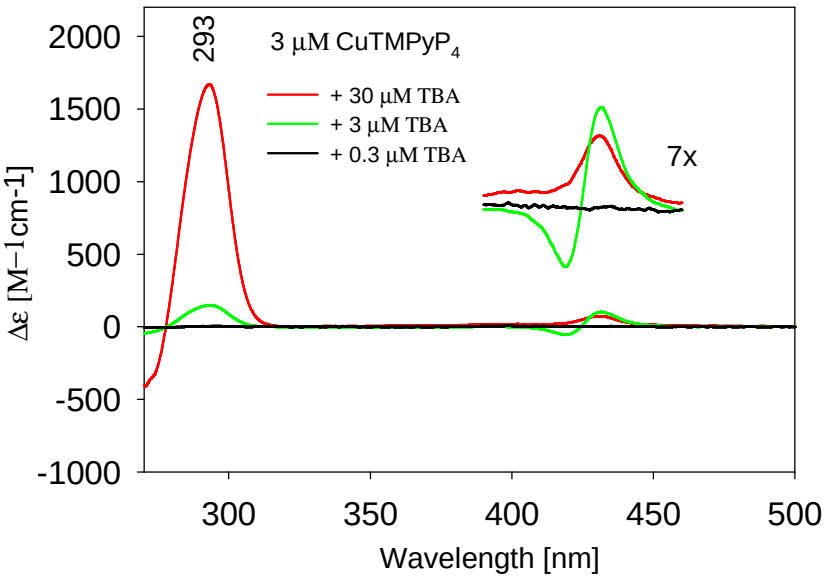



**Figure 62.** CD spectra of the TBA-CuTMPyP4 mixture for molar ratios r = [CuTMPyP4]/[TBA] equal to 0.1, 1 and 10. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, the concentration of CuTMPyP4 is fixed at 3 μM and the strand concentration of TBA is changing from 0.3, 3 to 30 μM. According to the classical interpretation of CD spectra of porphyrins complexed with DNA duplexes applied to G-quadruplexes, induced CD signal for r = 0.1 corresponds to CuTMPyP4 bound externally to TBA (within the grooves or dT loops). CD signal for r = 10 corresponds predominately to free CuTMPyP4, meaning that TBA does not form a complex with a large number of porphyrin molecules. The bisignate CD signal for r = 1 was interpreted in the past as a consequence of self-association CuTMPyP4 on chiral matrix. It is consistent with the notion that at least two molecules CuTMPyP4 form self-assembled dimers.

The role of thymine loops of TBA quadruplex in the dimerization was also investigated by analyzing thermal stabilities of its various loop mutants (see table 7 for the sequences). The results of thermal melting experiments are summarized in tables 8, 9 and 11. The measurements were done at low DNA concentrations in KCl (see table 8) and NaCl (see table 9) using absorption spectroscopy and at higher DNA concentration in KCl using CD spectroscopy (see table 11). Thermal data extracted from the absorption spectra are slightly more precise than those acquired from CD spectra because more absorption spectra were included in the SVD

analysis. The comparison of van't Hoff enthalpies, entropies and melting temperatures under the assumption of intramolecular model is depicted in figure 69. The most stable oligonucleotide in the study is T2/T3 (one T added to each edge-wise loop TT) together with TBA. The stability of most TBA mutants in $K^+$ is higher than in $Na^+$, as it is proper for G-quadruplexes. Nevertheless, it is interesting that TBA related sequences having long central loops (TGT/T5 and TGT/T6) are slightly more thermostable in NaCl than in KCl. It is interesting that this correlates with higher electrophoretic mobilies of these oligonucleotides in NaCl versus KCl. The big difference between thermostability of some oligonucleotide in KCl versus NaCl (and the same antiparallel fold in both salts) indicates that the structure of the loops is different in KCl and NaCl.

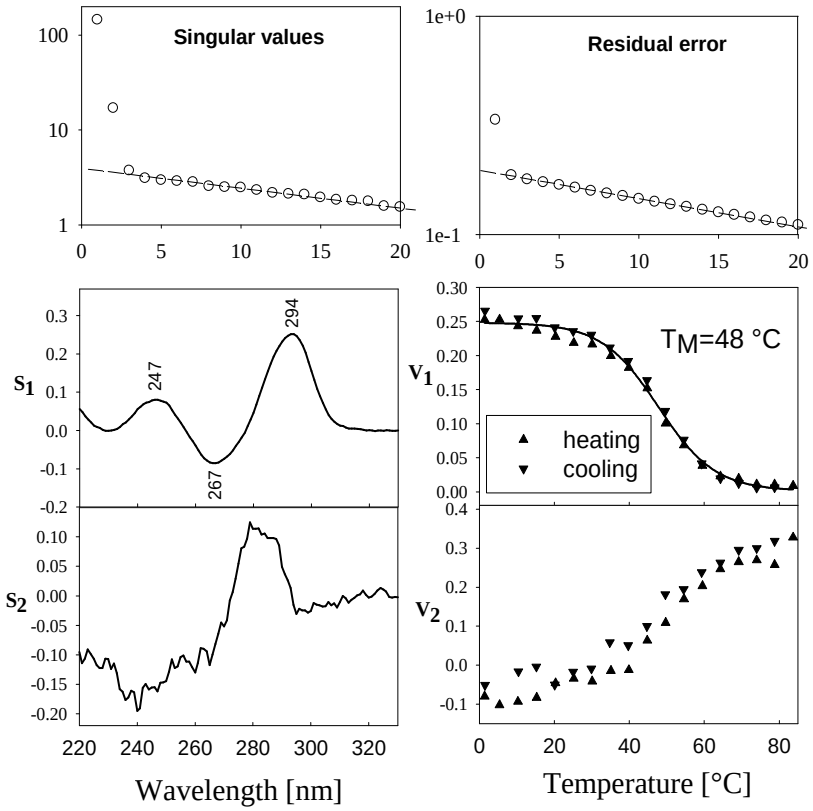



**Figure 63.** The result of SVD analysis within the ultraviolet part of the CD spectra corresponding to the TBA-CuTMPyP4 mixture for molar ratio 1:1 recorded within 1-84°C. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, 4 μM DNA strand concentration. CD signature of the antiparallel folding topology as well as melting temperature is not significantly altered upon binding CuTMPyP4 to TBA. Note that heating and cooling profiles are superimposable.

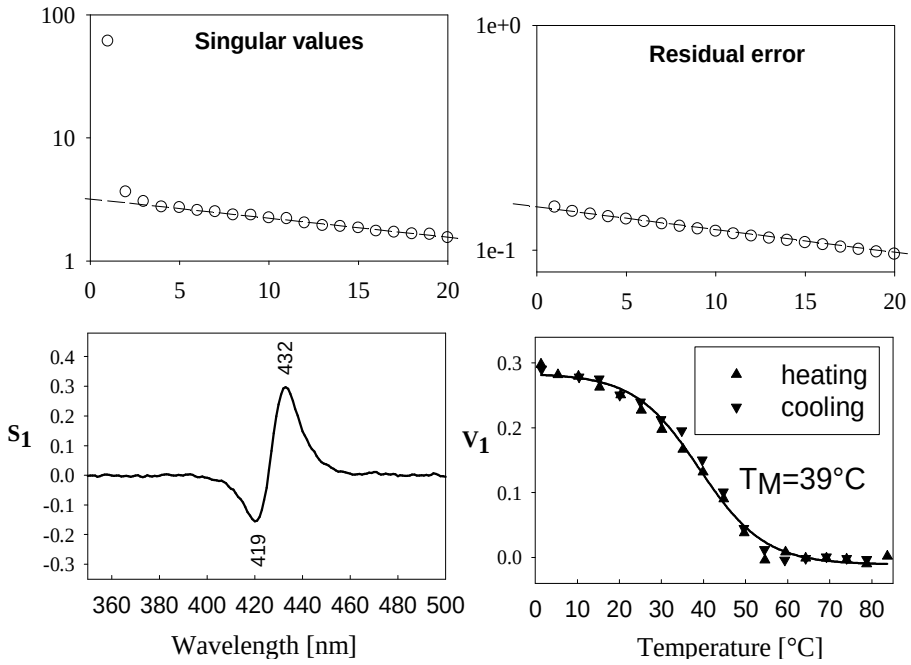



**Figure 64.** The result of SVD analysis within the visible part of the CD spectra corresponding to the TBA-CuTMPyP4 mixture for molar ratio 1:1 recorded in the temperature range 1-84°C. Experimental conditions are the same as in figure 63 (figures 63 and 64 correspond to one experiment). Note that heating and cooling profiles are superimposable. Thermal stability of bisignate CD feature is markedly lower than that of G-quadruplex (figure 63).

The formation of the proposed dimeric structure of TBA was also investigated by molecular dynamics simulations. It seems unlikely that such structure would be formed by the stacking of two TBA quadruplexes via their guanine tetrads in the same manner as it was proposed for other dimeric quadruplex structures. It is not possible because the guanine and thymine bases within the TGT loop stack on the G-quartet, as well as because of the flexibility of the TT loops revealed by molecular dynamics. Macaya et al. previously reported that, in addition to unimolecular G-quadruplex structure, bimolecular hairpin is also consistent with NMR connectivities **[Macaya et al., 1993]**. The hairpin structure was eventually excluded because it was not compatible with the NMR analysis of inosine derivatives of TBA **[Macaya et al., 1993]**. The results of our molecular dynamics simulation on 200 ns time-scale allow for the existence of dimeric structure (see figure 70) but do not exclude bimolecular hairpin quadruplex (see figure 71).

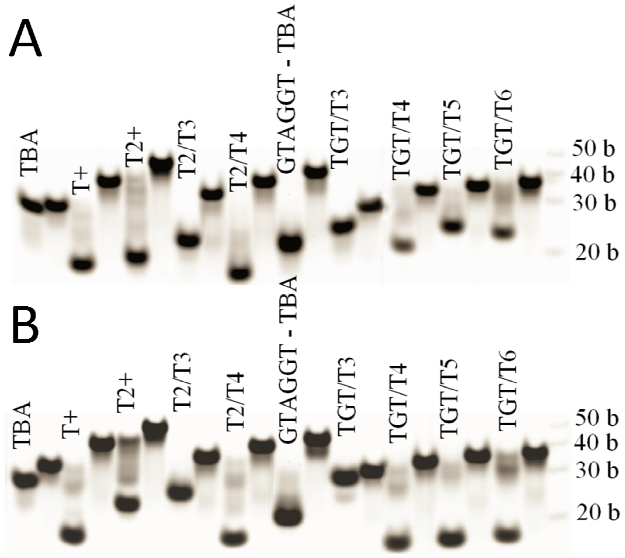



**Figure 65.** Native polyacrylamide gel electrophoresis performed in 10 mM potassium phosphate buffer (pH ~ 7) with added 100 mM KCl (**A**) or 100 mM NaCl (**B**). Annealed oligonucleotides (see table 7 for the sequences) were loaded at room temperature on the gel and electrophoresed at 2°C. The odd lanes represent the studied oligonucleotides and the even lanes contain the corresponding heteroduplexes. Only TBA and TGT/T3 migrate as bimolecular species and comigrate thus with their heteroduplexes. The mobility of T2/T3 and TGT/T5 (in KCl) is somewhere between that corresponding to monomeric and dimeric species.

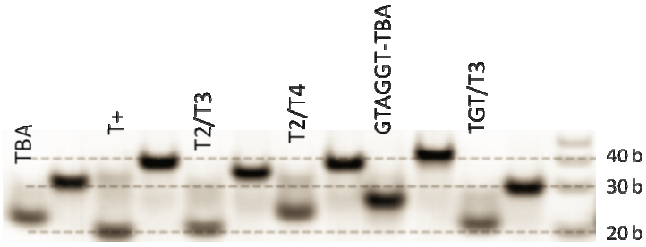



**Figure 66.** Native polyacrylamide gel electrophoresis performed in 10 mM potassium phosphate buffer (pH ~ 7) with added 100 mM KCl. Annealed oligonucleotides (see table 7 for the sequences) were loaded at room temperature on the gel and electrophoresed at 40°C. The odd lanes represent the studied oligonucleotides and the even lanes contain the corresponding heteroduplexes.

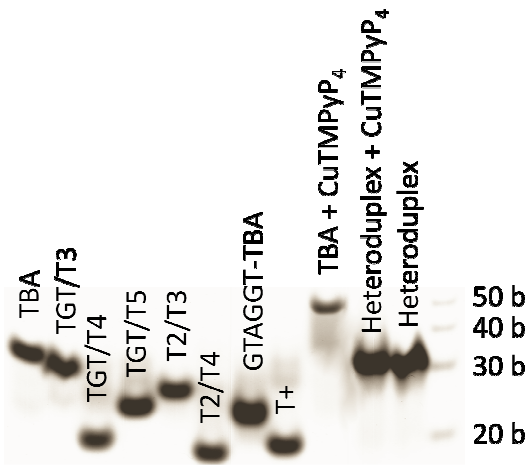



**Figure 67.** Native polyacrylamide gel electrophoresis performed in 10 mM potassium phosphate buffer (pH ~ 7) with added 100 mM KCl. The first eight lanes from left correspond TBA and its related oligonucleotides (refer to table 7 for the sequences). The last three lanes (near the ladder) contain from left a mixture of TBA with CuTMPyP4 (1:1), a mixture of TBA heteroduplex with CuTMPyP4 (1:1) and TBA heteroduplex. DNA samples were annealed prior to electrophoresis. Porphyrins were added to the annealed samples at room temperature just before loading them on the gel. The electrophoresis was conducted at 2°C.

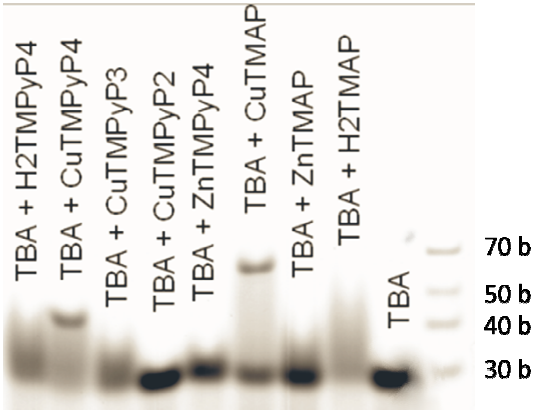



**Figure 68.** Native polyacrylamide gel electrophoresis performed in 10 mM sodium phosphate buffer (pH ~ 7) with added 100 mM NaCl. Last lane from left contains TBA. Other lanes contain 1:1 mixtures of TBA with the eight studied porphyrins (see figure 40), as indicated in the figure. DNA samples were annealed prior to electrophoresis. Porphyrins were added to the annealed samples at room temperature just before loading them on the gel. The electrophoresis was conducted at 2°C.

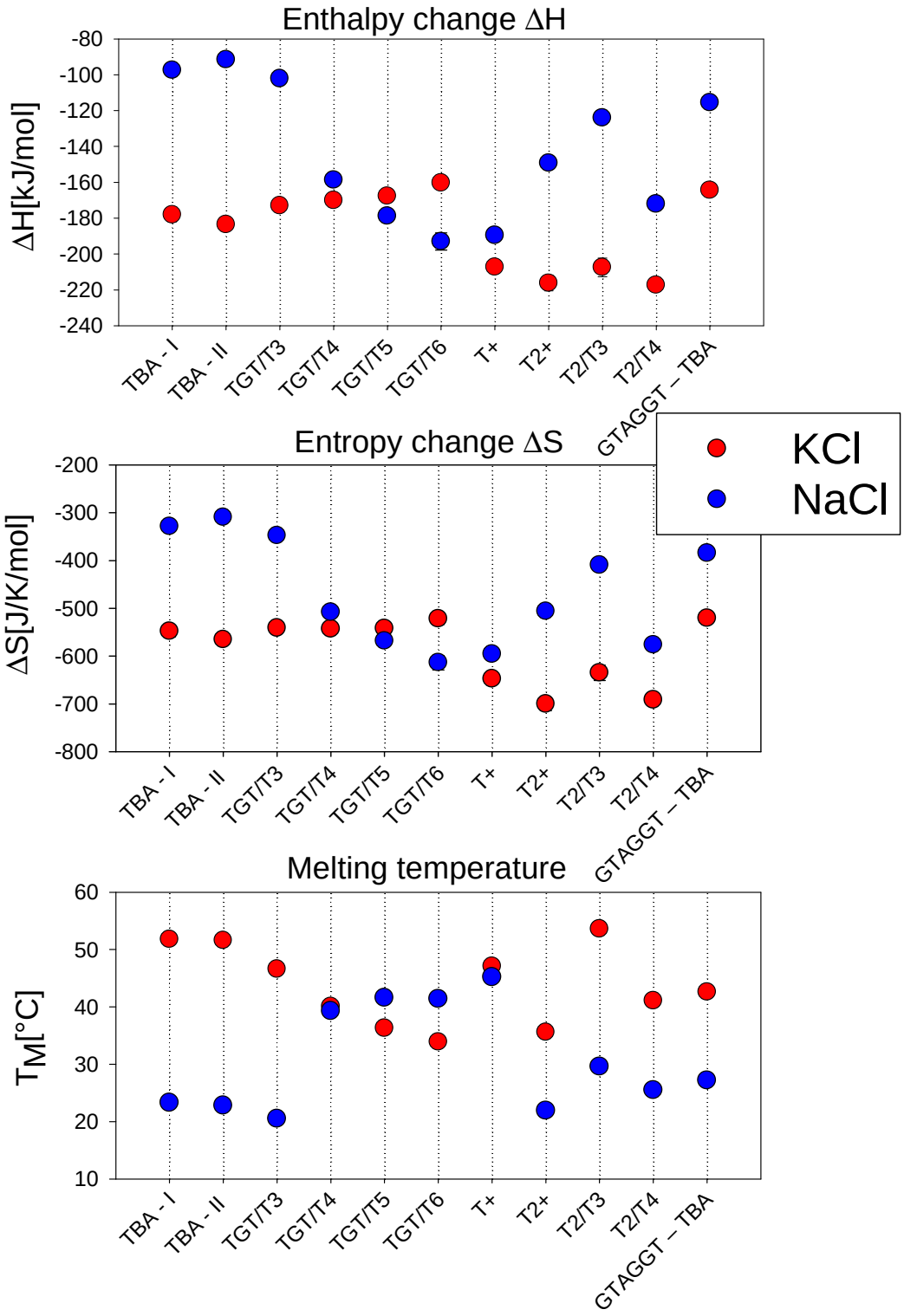



**Figure 69.** The comparison of thermodynamic data for TBA and its related sequences (see table 7) stabilized by 100 mM KCl versus 100 mM NaCl (extracted from tables 8 and 9). The results are obtained from the absorption measurements at low DNA concentrations (about 3 μM in strands) under the assumption of intramolecular model (see the section 2.2.4).

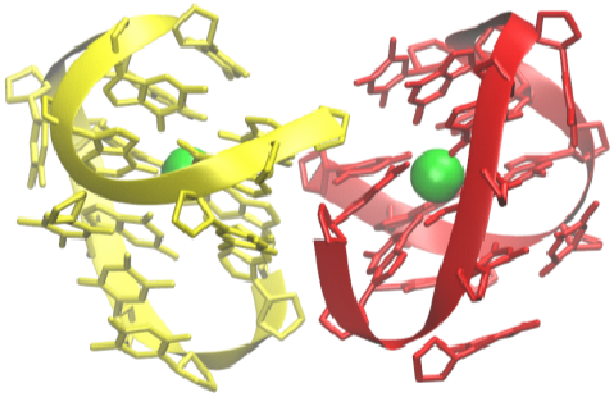


**Figure 70.** A snapshot of a dimeric structure of TBA from molecular dynamics simulation. A dimer was stable on 200 ns time scale. TBA molecules forming a dimer are depicted in yellow and red. Stabilizing potassium cations located within two G-tetrads are shown as green balls.

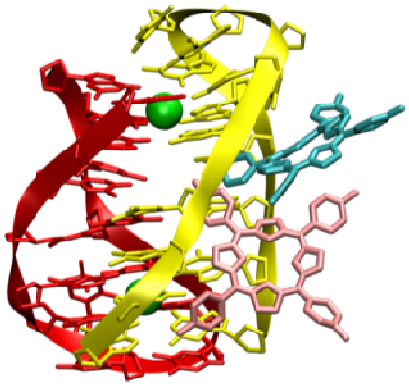


**Figure 71.** A snapshot of a bimolecular hairpin structure of TBA from molecular dynamics simulation. A hairpin structure was stable on 100 ns time scale. TBA molecules forming a dimer are depicted in yellow and red. Stabilizing potassium cations located within two G-tetrads are shown as green balls. Two porphyrins (depicted in cyan and pink color) can bind from the outside to a hairpin structure.

Long molecular simulations show that the free base porphyrin $H_2$TMPyP4 binds preferentially on the side of the TBA quadruplex and less frequently in the vicinity of the TGT loop. It corresponds to the stoichiometry between 1:1 and 1:2,

which seems to be consistent with the results reported by other researchers **[del Toro et al., 2008; Haq et al., 1999]**. Consequently, a dimer of TBA may be described as a sandwich-like structure (see figure 72) composed of two TBA quadruplexes glued together by 2 porphyrins (each porphyrin is initially bound to one TBA molecule; the stoichiometry 1:1). Using the method of continuous variation (Job plot) **[Haq et al., 1999]** applied to CD and absorption measurements (see figure 73), we have determined the stoichiometry of the TBA-CuTMPyP4 complex to be between 1:1 and 1:2, thus well corresponding to 2 TBA for 3 CuTMPyP4.

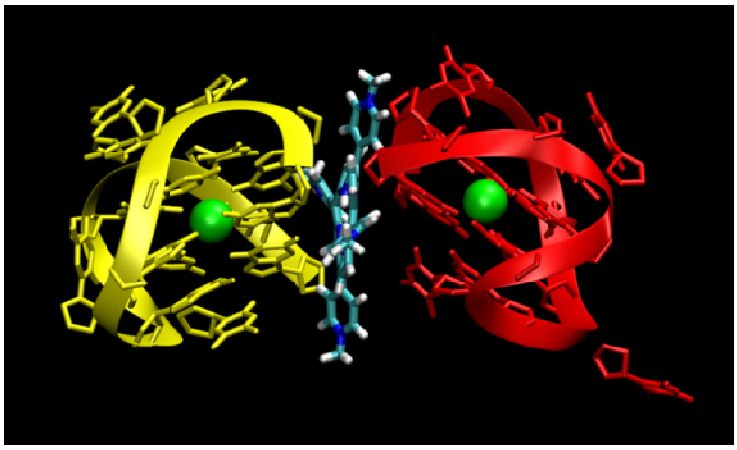


**Figure 72.** A snapshot of a sandwich-like structure of TBA from molecular dynamics simulation. A sandwich structure was stable on 100 ns time scale. TBA molecules forming a sandwich structure are depicted in yellow and red. Stabilizing potassium cations located within two G-tetrads are shown as green balls. One porphyrin in the figure is depicted as embedded between two G-quadruplexes. The interaction of two porphyrins that glue together two TBA molecules in a sandwich may result in the appearance of the bisignate CD signal.

In summary, we demonstrate spectroscopically that TBA quadruplex stabilized by $K^+$ or $Na^+$ may dimerize in aqueous solution. The bisignate shape of CD spectra of CuTMPyP4 in complex with TBA indirectly reflects the mutual interaction of monomeric TBA species in a putative dimer. Dimerization is exclusive to the TBA sequence because the related oligonucleotides with elongated thymine loops migrate on PAGE gels more like monomeric species. Dimeric structure is also consistent with long time-scale molecular dynamics simulations.

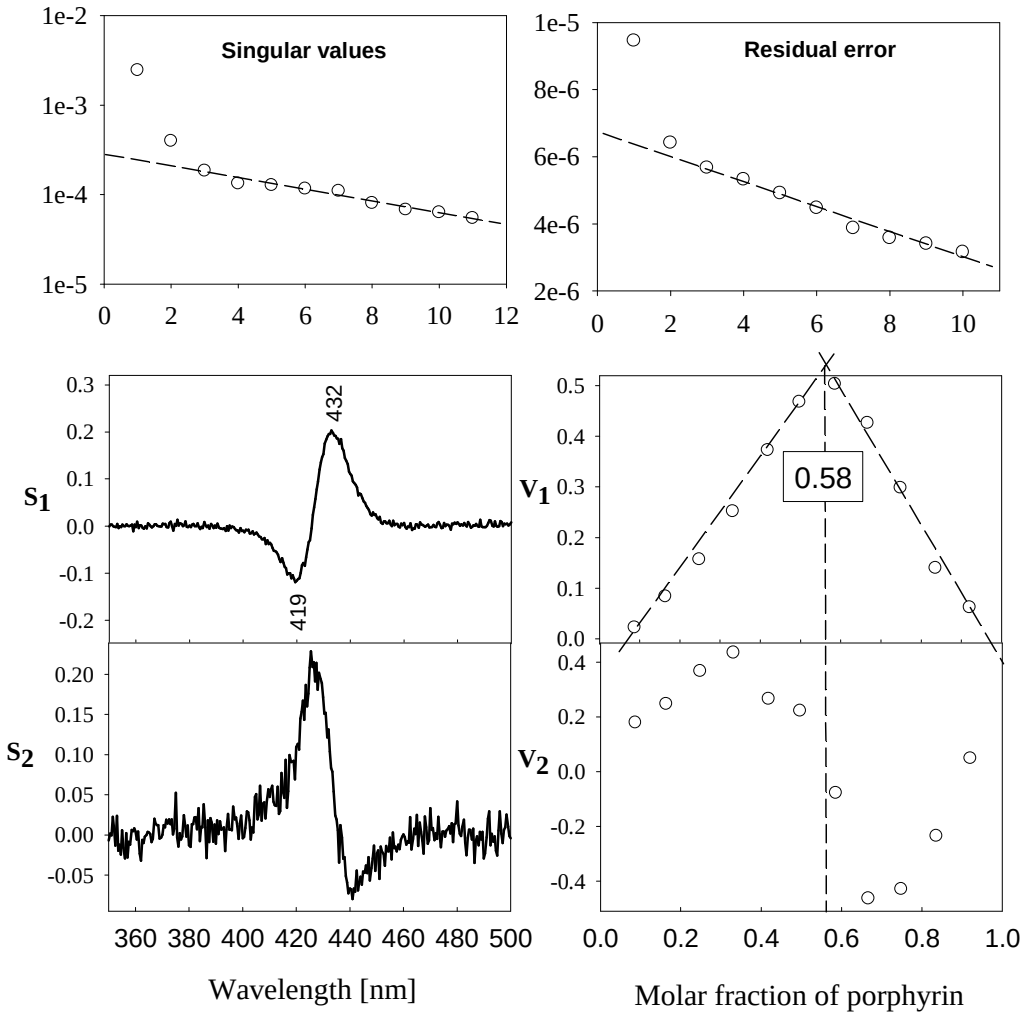



**Figure 73.** The method of continuous variation (Job plot) applied to CD spectra of the TBA-CuTMPyP4 complex gives molar fraction of CuTMPyP4 in the complex equal ~ 0.58, which is close to 0.6 corresponding to 2 TBA molecules per 3 CuTMPyP4. Experimental conditions: 20 mM lithium cacodylate buffer, 100 mM KCl, pH 6.8, temperature 5°C, the sum of the strand concentration of [TBA] and the ligand concentration [CuTMPyP4] is kept constant at 6 μM. The Job applied to the Soret band of CuTMPyP4 gives the same of stoichiometry.

# Conclusions

The subject of PhD work was to investigate the physical properties of non-canonical DNA structures. A great deal of the work was specifically focused on monitoring the conformational properties and thermal stability of unusual four-stranded nucleic acids structures commonly known as guanine quadruplexes or tetraplexes. The main idea was to employ cationic porphyrins as sensitive probes of DNA conformation and clarify their potential role in the formation and stabilization of guanine quadruplexes. A wide range of experimental techniques have been applied for this purpose. These include optical spectroscopic methods (absorption, CD and Raman spectroscopy), NMR, differential scanning calorimetry and gel electrophoresis. The main part of research was carried out at the Division of Biomolecular Physics at Institute of Physics of Charles University in Prague, and the Raman measurements conducted there on a home-made Raman spectrometer constitute the framework of the thesis.

Much of our scientific findings are the outcome of an intensive collaboration with the group of prof. Vorlíčková at Institute of Biophysics of the Academy of Sciences of the Czech Republic (Brno). Raman spectroscopy was successfully used in parallel with CD spectroscopy and native gel electrophoresis to monitor conformational polymorphism of human telomeric guanine quadruplexes in solution. The use of Raman spectroscopy in our study is not an autotelism. The method was applied as the best choice to elucidate the discrepancies still persisting about the real structure of human telomeric G-quadruplexes under the physiological conditions **[Renčiuk et al., 2009]**. Since DNA inside the cells is highly condensed, guanine quadruplex structures conditioned by molecular crowding should be studied under the nucleoside concentrations that are far higher than the concentrations appropriate for the most common experimental methods (absorption, CD, PAGE). However, there are doubts as to whether the G-quadruplex structures observed in the crystals can be transferred to aqueous solutions. Molecular crowding conditions in vitro are often simulated by the addition of various cosolutes but their effect on the G-quadruplex structure is largely unknown. It is questionable whether common crowding agents such polyethylene glycol or ethanol exert their effect on the DNA structure by the simple molecular crowding or whether it is rather a result of their

specific interaction with DNA or dehydration **[Miyoshi et al., 2006; Petraccone et al., 2012]**. As Raman spectroscopy is applicable to DNA samples at an ample range of nucleoside concentrations ranging from dilute aqueous solutions still accessible to absorption or CD spectroscopies up to highly dense gels or even crystals, it gives a unique possibility to probe the crowding phenomena induced by the presence of DNA solely. In our Raman study we simply demonstrate that the conformation of the human telomeric sequence $G_3(TTAG_3)_3$ annealed in the presence of $K^+$ ions depends on the DNA concentration. The quadruplex is gradually transformed from the predominantly antiparallel arrangement at low DNA concentrations to predominantly parallel arrangement known from crystals, probably passing through several hybrid (3+1) forms **[Renčiuk et al., 2009; Yue et al., 2011]** hardly discernible in solution by CD or other common spectroscopic methods. We found that the DNA concentration-dependent transition is synergically enhanced by increased $K^+$ concentration and no transition of this kind is observed for the $G_3(TTAG_3)_3$ quadruplex stabilized by $Na^+$. The work on human telomeric quadruplexes has been recently accepted for publication **[Palacký et al., 2012, submitted, in attachments, pages 195-220]**.

There are still uncertainties about the solution conformations of the most frequently studied guanine quadruplexes, especially if they are determined by the experimental techniques applicable at diametrically different conditions. This concerns the thrombin-binding aptamer 5'-d($G_2T_2G_2TGTG_2T_2G_2$)-3' (TBA). The structure of TBA has been described as a chair-like intramolecular G-quadruplex. This notion is directly supported by NMR **[Macaya et al., 1993]** or X-ray structures **[Padmanabhan et al., 1993]**. It was previously proposed, on the basis of bimolecular-like migration of TBA in native polyacrylamide gels, that TBA stabilized by $K^+$ or $Na^+$ forms a bimolecular quadruplex in solution **[Fialová et al., 2006]**. Nonetheless, our recent experimental data are better explained by some kind of dimeric arrangement of TBA in $K^+$ solution (see the section 3.4). The hairpin-type bimolecularity of TBA seems to be very unlikely as it contradicts the fact that the thermal cooling and heating profiles provided by absorption, CD and Raman spectroscopy are always reversible. In addition, the group of doc. Jan Lang (from Department of Low Temperature Physics of Charles University in Prague) excludes, on the basis of the diffusion coefficient determined by DOSY NMR, the presence of oligomeric species for TBA stabilized by KCl. Compelling indirect evidence in favor

of TBA dimerization comes from the systematic investigation of the interaction of TBA and similar sequences with cationic porphyrins (in total, ten DNA oligonucleotides and eight cationic porphyrins were included in the study). The most important result is that the copper porphyrin CuTMPyP4 in complex with TBA stabilized by KCl exhibits the pronounced bisignate CD spectrum. We detected the similarly strong bisignate CD signal only for the structurally related copper porphyrin CuTMAP. It is interesting that the bisignate CD spectrum of CuTMPyP4 disappears only at extremely low porphyrin loads (TBA/CuTMPyP4 > 10) when it changes to a positive band indicating external binding of CuTMPyP4 to the quadruplex. The bisignate CD signal of CuTMPyP4 complexed with TBA suggests a close proximity of two or more porphyrins due to their specific interaction in a putative TBA dimer. It is our crucial finding that the induced CD signal of CuTMPyP4 disappears at higher temperatures but some 9°C before the quadruplex is denaturated. The porphyrin thus indirectly indicates the presence of two different TBA structures: a suggested dimer present predominately at lower temperatures and an intramolecular G-quadruplex occurring at higher temperatures. We show that a dimeric structure is stable on a longer time-scale (100-200 ns) of molecular dynamics simulations. It seems that two TBA quadruplexes dimerize in a side-to-side manner since an end-to-end interaction via terminal G-tetrads is largely impeded by the stacking of the guanine and thymine bases in the central loop on the terminal guanine quartet and due to the conformational flexibility of loops. We demonstrate that the dimerization is reduced or lost for TBA related sequences with elongated thymine loops as they migrate on native PAGE gels more like monomeric species. As is evident, the study concerning the TBA quadruplex is very extensive. The key results have been already prepared for a publication in specialized biophysical journal.

Apart from the thrombin binding aptamer, we also investigated the interaction of other guanine quadruplexes with cationic porphyrins, although not in so much detail. Our results have methodological implications. G-quadruplex-porphyrin mixtures were characterized by Raman scattering, absorption and CD spectroscopy. We did not measure non-resonance Raman spectra of DNA but only resonance Raman spectra of the copper cationic CuTMPyP4 complexed with guanine quadruplexes. The use of the copper cationic porphyrin CuTMPyP4 for Raman measurements is the only option since other porphyrins exhibit low photostability and high fluorescence background **[Kruglik et al., 2001; McMillin & McNett,**

1998]. Our results convincingly demonstrate that copper cationic porphyrins are sensitive probes to the G-quadruplex structure. Using SVD approach, we demonstrate that the copper porphyrin CuTMPyP4 can differentiate between various quadruplex structures induced by the presence of different salts ($Li^+$, $Na^+$, $K^+$) or occurring at different temperature. We also show that resonance Raman spectroscopy of porphyrins can be conveniently employed to indirectly follow thermal denaturation and folding of guanine quadruplexes. Specifically, using resonance Raman spectroscopy, we could distinguish intermolecular and intramolecular guanine quadruplexes on the basis of their different kinetics of folding and denaturation that is reflected in the presence or absence of the hysteresis **[Mergny et al., 2005a]** between heating and cooling thermal profiles. The results of our experiments on the interaction of cationic porphyrins with guanine quadruplexes are in preparation for publication.

Spectroscopic experiments typically produce a large quantity of data. Furthermore, the differences between the spectra within large datasets are often very small. This is especially the case of Raman spectra. The inspection of each spectrum individually by naked eye is impractical, time-consuming and sometimes impossible. It is also often needed to compare the results of several experiments in a uniform and transparent manner. This can be done by using multivariate statistical techniques such as SVD (singular value decomposition). The definite advantage of these methods over the conventional methods is that the whole information within data can utilized (unlike, for instance, in the case of recording and/or processing absorption spectra at one wavelength) and that the dimensionality of data can be reduced without significant loss of information. During doctoral study, several useful programs based on the SVD algorithm have been written in Matlab® (version 7.4, Mathworks) environment to facilitate data processing and visualization. The most important programs include the program for manual correction of background and visualization of the results of SVD, the program for intensity normalization and background (solvent) subtraction exploiting the properties of water stretching vibrations (see the appropriate paper for details **[Palacký et al., 2011, in attachments, pages 227-250]**) and the program for visualization of the results of 2D imaging experiments performed by Raman microscope (see the results on imaging of the vacuoles of the yeast cells **[Bednárová at al., 2012, in attachments, page 221-226]** and review article **[Mojzeš et al., 2011]**).

# Bibliography

ABU-GHAZALAH R. M., IRIZAR J., HELMY A. S., MACGREGOR R. B. (2010) A study of the interactions that stabilize DNA frayed wires. *Biophysical Chemistry* **147**, 123-129.

ABU-GHAZALAH R. M., RUTLEDGE S., LAU L. W. Y., DUBINS D. N., MACGREGOR R. B., JR., HELMY A. S. (2012) Concentration-dependent structural transitions of human telomeric DNA sequences. *Biochemistry* **51**, 7357-7366.

ADRIAN M., HEDDI B., PHAN A. T. (2012) NMR spectroscopy of G-quadruplexes. *Methods* **57**, 11-24.

AGRAWAL S., KHAN F. (2005) Reconstructing recent human phylogenies with forensic STR loci: A statistical approach. *BMC Genetics* **6**, article number: 47.

ALBERTI P., BOURDONCLE A., SACCA B., LACROIX L., MERGNY J. L. (2006) DNA nanomachines and nanostructures involving quadruplexes. *Organic & Biomolecular Chemistry* **4**, 3383-3391.

ALBERTI P., MERGNY J. L. (2003) DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1569-1573.

ALLSHIRE R. C., DEMPSTER M., HASTIE N. D. (1989) Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic Acids Research* **17**, 4611-4627.

ALZEER J., LUEDTKE N. W. (2010) pH-Mediated fluorescence and G-quadruplex binding of amido phthalocyanines. *Biochemistry* **49**, 4339-4348.

AMBRUS A., CHEN D., DAI J. X., BIALIS T., JONES R. A., YANG D. Z. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Research* **34**, 2723-2735.

AMBRUS A., CHEN D., DAI J., JONES R. A., YANG D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-myc promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048-2058.

ANANTHA N. V., AZAM M., SHEARDY R. D. (1998) Porphyrin binding to quadruplexed $T_4G_4$. *Biochemistry* **37**, 2709-2714.

ANDRUSHCHENKO V., TSANKOV D., KRASTEVA M., WIESER H., BOUŘ P. (2011)

Spectroscopic detection of DNA quadruplexes by vibrational circular dichroism. *Journal of the American Chemical Society* **133**, 15055-15064.

ANTONACCI C., CHAIRES J. B., SHEARDY R. D. (2007) Biophysical characterization of the human telomeric (TTAGGG)(4) repeat in a potassium solution. *Biochemistry* **46**, 4654-4660.

ARTHANARI H., BASU S., KAWANO T. L., BOLTON P. H. (1998) Fluorescent dyes specific for quadruplex DNA. *Nucleic Acids Research* **26**, 3724-3728.

ARTHANARI H., BOLTON P. H. (2001) Functional and dysfunctional roles of quadruplex DNA in cells. *Chemistry & Biology* **8**, 221-230.

BACOLLA A., WELLS R. D. (2004) Non-B DNA conformations genomic rearrangements, and human disease. *Journal of Biological Chemistry* **279**, 47411-47414.

BALAGURUMOORTHY P., BRAHMACHARI S. K. (1994) Structure and stability of human telomeric sequence. *Journal of Biological Chemistry* **269**, 21858-21869.

BALAGURUMOORTHY P., BRAHMACHARI S. K., MOHANTY D., BANSAL M., SASISEKHARAN V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Research* **20**, 4061-4067.

BALASUBRAMANIAN S., NEIDLE S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Current Opinion in Chemical Biology* **13**, 345-353.

BALAZ M., DE NAPOLI M., HOLMES A. E., MAMMANA A., NAKANISHI K., BEROVA N., PURRELLO R. (2005) A cationic zinc porphyrin as a chiroptical probe for Z-DNA. *Angewandte Chemie - International Edition* **44**, 4006-4009.

BALDRICH E., O'SULLIVAN C. K. (2005) Ability of thrombin to act as molecular chaperone, inducing formation of quadruplex structure of thrombin-binding aptamer. *Analytical Biochemistry* **341**, 194-197.

BARDIN C., LEROY J. L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Research* **36**, 477-488.

BASNAR B., ELNATHAN R., WILLNER I. (2006) Following aptamer-thrombin binding by force measurements. *Analytical Chemistry* **78**, 3638-3642.

BAUER L., TLUCKOVA K., TOTHOVA P., VIGLASKY V. (2011) G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. *Biochemistry* **50**, 7484-7492.

BAUMRUK V., GOUYETTE C., HUYNH-DINH T., SUN J. S., GHOMI M. (2001) Comparison between CUUG and UUCG tetraloops: thermodynamic stability and

structural features analyzed by UV absorption and vibrational spectroscopy. *Nucleic Acids Research* **29**, 4089-4096.

BELL A. F., HECHT L., BARRON L. D. (1997) Vibrational Raman optical activity as a probe of polyribonucleotide solution stereochemistry. *Journal of the American Chemical Society* **119**, 6006-6013.

BENDICH A. J., DRLICA K. (2000) Prokaryotic and eukaryotic chromosomes: what's the difference. *BioEssays* **22**, 481-486.

BENEVIDES J. M., KANG C., THOMAS G. J., JR. (1996) Raman signature of the four-stranded intercalated cytosine motif in crystal and solution structures of DNA deoxycytidylates d(CCCT) and d(C8). *Biochemistry* **35**, 5747-5755.

BENEVIDES J. M., OVERMAN S. A., THOMAS G. J., JR. (2005) Raman, polarized Raman and ultraviolet resonance Raman spectroscopy of nucleic acids and their complexes. *Journal of Raman Spectroscopy* **36**, 279-299.

BHADRA K., KUMAR G. S. (2011) Interaction of berberine, palmatine, coralyne, and sanguinarine to quadruplex DNA: A comparative spectroscopic and calorimetric study. *Biochimica et Biophysica Acta – General Subject* **1810**, 485-496.

BHATTACHARJEE A. J., AHLUWALIA K., TAYLOR S., JIN O., NICOLUDIS J. M., BUSCAGLIA R., CHAIRES J. B., KORNFILT D. J. P., MARQUARDT D. G. S., YATSUNYK L. A. (2011) Induction of G-quadruplex DNA structure by Zn(II) 5,10,15,20-tetrakis (N-methyl-4-pyridyl) porphyrin. *Biochimie* **93**, 1297-1309.

BISHOP J. S., GUY-CAFFEY J. K., OJWANG J. O., SMITH S. R., HOGAN M. E., COSSUM P. A., RANDO R. F.**,** CHAUDHARY N. (1996) Intramolecular G-quartet motifs confer nuclease resistance to a potent anti-HIV oligonucleotide. *Journal of Biological Chemistry* **271**, 5698-5703.

BLACKBURN E. (1999) The telomere and telomerase: How do they interact? *Mount Sinai Journal of Medicine* **66**, 292-300.

BLACKBURN E. (2001) Switching and signaling at the telomere. *Cell* **106**, 661-673.

BLASCO M. A., HAHN W. C. (2003) Evolving views of telomerase and cancer. *Trends in Cell Biology* **13**, 289-294.

BLOOMFIELD V. A., CROTHERS D. M., TINOCO I., JR.: Nucleic acids structures, properties and functions, University science books, Sausalito (California) 2000, ISBN 0-935702-49-0.

BLUME S. W., GUARCELLO V., ZACHARIAS W., MILLER D. M. (1997) Divalent transition metals cations counteract potassium-induced quadruplex assembly of

oligo(dG) sequences. *Nucleic Acids Research* **25**, 617-625.

BOCK L. C., GRIFFIN L. C., LATHAM J. A., VERMAAS E. H., TOOLE J. J. (1992) Selection of single-stranded-DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564-566.

BONČINA M., LAH J., PRISLAN I., VESNAVER G. (2012) Energetic basis of human telomeric DNA folding into G-quadruplex structures. *Journal of the American Chemical Society* **134**, 9657-9663.

BORBONE N., AMATO J., OLIVIERO G., D'ATRI V., GABELICA V., DE PAUW E., PICCIALLI G., MAYOL L. (2011) d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Research* **39**, 7848-7857.

BREUZARD G., MILLOT J. M., RIOU J. F., MANFAIT M. (2003) Selective interactions of ethidiums with G-quadruplex DNA revealed by surface-enhanced Raman scattering. *Analytical Chemistry* **75**, 4305-4311.

BROWN N. M., RACHWALL P. A., BROWN T., FOX K. R. (2005) Exceptionally slow kinetics of the intramolecular quadruplex formed by the Oxytricha telomeric repeat. *Organic & Biomolecular Chemistry* **3**, 4153-4157.

BRUYLANTS G., WOUTERS J., MICHAUX C. (2005) Differential scanning calorimetry in life science: Thermodynamics, stability, molecular recognition and application in drug design. *Current Medicinal Chemistry* **12**, 2011-2020.

BUSCAGLIA R., JAMESON D. M., CHAIRES J. B. (2012) G-quadruplex structure and stability illuminated by 2-aminopurine phasor plots. *Nucleic Acids Research* **40**, 4203-4215.

CAMPBELL N. H., PARKINSON G. N. (2007) Crystallographic studies of quadruplex nucleic acids. *Methods* **43**, 252-263.

CAMPBELL N. H., PATEL M., TOFA A. B., GHOSH R., PARKINSON G. N., NEIDLE S. (2009) Selectivity in ligand recognition of G-quadruplex loops. *Biochemistry* **48**, 1675-1680.

CATASTI P., CHEN X., MOYZIS R. K., BRADBURY E. M., GUBTA G. (1996) Structure-function correlations of the insulin-linked polymorphic region. *Journal of Molecular Biology* **264**, 534-545.

CAVALUZZI M. J., BORER P. N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Research* **32**, article number: e13

CECH T. R. (2000) Life at the end of the chromosome: Telomeres and telomerase. *Angewandte Chemie - International Edition* **39**, 34-43.

CHAIRES J. B. (2008) Calorimetry and thermodynamics in drug design. *Annual Review of Biophysics* **37**, 135-151.

CHAIRES J. B. (2010) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS Journal* **277**, 1098-1106.

CHAKHPARONIAN M., WELLINGER R. J. (2003) Telomere maintenance and DNA replication: how closely are these two connected? *Trends in Genetics* **19**, 439-446.

CHANG C. C., CHIEN C. W., LI Y. H., KANG C. C., CHANG T. C. (2007) Investigation of spectral conversion of $d(TTAGGG)_4$ and $d(TTAGGG)_{13}$ upon potassium titration by a G-quadruplex recognizer BMVC molecule. *Nucleic Acids Research* **35**, 2846-2860.

CHEN Q., KUNTZ I. D., SHAFER R. H. (1996) Spectroscopic recognition of guanine dimeric hairpin quadruplexes by a carbocyanine dye. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 2635-2639.

CHEN M. L., SONG. G. T., WANG C. Y., HU D., REN J. S., QU X. G. (2009) Small-molecule selectively recognizes human telomeric G-quadruplex DNA and regulates its conformational switch. *Biophysical Journal* **97**, 2014-2023.

CHEONG C. J., MOORE P. B. (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **31**, 8406-8414.

CHOI J., KIM S., TACHIKAWA T., FUJITSAKA M., MAJIMA T. (2011) pH-induced intramolecular folding dynamics of i-Motif DNA. *Journal of the American Chemical Society* **133**, 16146-16153.

CHRISTIANSEN J., KOFOD M., NIELSEN F. C. (1994) A guanosine quadruplex and two stable hairpins flank a major cleavage site in insuline-like growth factor II mRNA. *Nucleic Acids Research* **22**, 5709-5716.

CLARK G. R., PYTEL P. D., SQUIRE C. J. (2012) The high-resolution crystal structure of a parallel intermolecular DNA G-4 quadruplex/drug complex employing syn glycosyl linkages. *Nucleic Acids Research* **40**, 5731-5738.

COGOI S., PARAMASIVAM M., SPOLAORE B., XODO L. E. (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Research* **36**, 3765-3780.

COLLIE G. W., PARKINSON G. N. (2011) The application of DNA and RNA G-

quadruplexes to therapeutic medicines. *Chemical Society Reviews* **40**, 5867-5892.

COLLIE G. W., PARKINSON G. N., NEIDLE S., ROSU F., DE PAUW E., GABELICA V. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *Journal of the American Chemical Society* **132**, 9328-9334.

COSCONATI S., MARINELLI L., TROTTA R., VIRNO A., DE TITO S., ROMAGNOLI R., PAGANO B., LIMONGELLI V., GIANCOLA C., BARALDI P. G., MAYOL L., NOVELLINO E., RANDAZZO A. (2010) Structural and conformational requisites in DNA quadruplex groove binding: Another piece to the puzzle. *Journal of the American Chemical Society* **132**, 6425-6433.

CRAIGIE R. (2001) HIV integrase, a brief overview from chemistry to therapeutics. *Journal of Biological Chemistry* **276**, 23213-23216.

CRAWLEY J. T. B, ZANARDELLI S., CHION C. K. N. K., LANE D. A. (2007) The central role of thrombin in hemostasis. *Journal of Thrombosis and Haemostasis* **5**, 95-101.

DA SILVA M. W. (2005) Experimental demonstration of T:(G:G:G:G):T hexad and T:A:A:T tetrad alignments within a DNA quadruplex stem. *Biochemistry* **44**, 3754-3764.

DA SILVA M. W. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry - A European Journal* **13**, 9738-9745.

DAHSE R., FIEDLER W., ERNST G. (1997) Telomeres and telomerase: biological and clinical importance. *Clinical Chemistry* **43**, 708-714.

DAI J., DEXHEIMER T. S., CHEN D., CARVER M., AMBRUS A., JONES R. A., YANG D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *Journal of the American Chemical Society* **128**, 1096-1098.

DAILEY M. M., MILLER M. C., BATES P. J., LANE A. N., TRENT J. O. (2010) Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Research* **38**, 4877-4888.

DANG C. V. (1999) c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Molecular and Cellular Biology* **19**, 1-11.

DAO N. T., HASELSBERGER R., MICHEL-BEYERLE M. E., PHAN A. T. (2011) Following G-quadruplex formation by its intrinsic fluorescence. *FEBS Letters* **585**, 3969-3977.

DAPIC V., ABDOMEROVIC V., MARRINGTON R., PEBERDY J., RODGER A., TRENT J. O., BATES P. J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Research* **31**, 2097-2107.

DARBY R. A. J., SOLLOGOUB M., MCKEEN C., BROWN L., RISITANO A., BROWN N., BARTON C., BROWN T., FOX K. R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Research* **30,** article number: e39.

DAVIS J. T. (2004) G-quartets 40 years later: From 5´-GMP to molecular biology and supramolecular chemistry. *Angewandte Chemie - International Edition* **43**, 668-698.

DE ARMOND R., WOOD S., SUN D., HURLEY L. H., EBBINGHAUS S. V. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* **44**, 16341-16350.

DE CIAN A., GUITTAT L., KAISER M., SACCA B., AMRANE S., BOURDONCLE A., ALBERTI P., TEULADE-FICHOU M. P., LACROIX L., MERGNY J. L. (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42**, 183-195.

DE CIAN A., MERGNY J. L. (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Research* **35**, 2483-2493.

DE SOULTRAIT V. R., LOZACH P. Y., ALTMEYER R., TARRAGO-LITVAK L., LITVAK S., ANDREOLA M. L. (2002) DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents. *Journal of Molecular Biology* **324**, 195-203.

DEL TORO M., BUCEK P., AVINO A., JAUMOT J., GONZALEZ C., ERITJA R., GARGALLO R. (2009). Targeting the G-quadruplex-forming region near the P1 promoter in the human BCL-2 gene with the cationic porphyrin TMPyP4 and with the complementary C-rich strand. *Biochimie* **91**, 894-902.

DEL TORO M., GARGALLO R., ERITJA R., JAUMOT J. (2008) Study of the interaction between the G-quadruplex-forming thrombin-binding aptamer and the porphyrin 5,10,15,20-tetrakis-(N-methyl-4-pyridyl)-21,23H-porphyrin tetratosylate. *Analytical Biochemistry* **379**, 8-15.

DIXON I. M., LOPEZ F., ESTEVE J. P., TEJERA A. M., BLASCO M. A., PRATVIEL G., MEUNIER B. (2005) Porphyrin derivatives for telomere binding and telomerase

inhibition. *Chembiochem* **6**, 123-132.

DIXON D. B., STEULLET V. (1998) Dimerization of tetracationic porphyrins: ionic strength dependence. *Journal of Inorganic Biochemistry* **69**, 25-32.

DO N. Q., LIM K. W., TEO M. H., HEDDI B., PHAN A. T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Research* **39**, 9448-9457.

DOI T., SHIBATA K., YOSHIDA M., TAKAGI M., TERA M., NAGASAWA K., SHIN-YA K., TAKAHASHI T. (2011) (*S*)-Stereoisomer of telomestatin as a potent G-quadruplex binder and telomerase inhibitor. *Organic & Biomolecular Chemistry* **9**, 387-393.

DUAN W. H., RANGAN A., VANKAYALAPATI H., KIM M. Y., ZENG Q. P., SUN D. K., HAN H. Y., FEDOROFF O. Y., NISHIOKA D., RHA S. Y., IZBICKA E., VON HOFF D. D., HURLEY L. H. (2001) Design and synthesis of fluoroquinophenoxazines that interact with human telomeric G-quadruplexes and their biological effects. *Molecular Cancer Therapeutics* **1**, 103-120.

DUQUETTE M. L., HANDA P., VINCENT J. A., TAYLOR A. F., MAIZELS N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & Development* **18**, 1618-1629.

ETZIONI S., YAFE A., KHATEB S., WEISMAN-SHOMER P., BENGAL E., FRY M. (2005) Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. *Journal of Biological Chemistry* **280**, 26805-26812.

EVANS S. E., MENDEZ M. A., TURNER K. B., KEATING L. R., GRIMES R. T., MELCHOIR S., SZALAI V. A. (2007) End-stacking of copper cationic porphyrins on parallel-stranded guanine quadruplexes. *Journal of Biological Inorganic Chemistry* **12**, 1235-1249.

FANG G., CECH T. R. (1993) Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein. *Biochemistry* **32**, 11646-57.

FEDOROFF O. Y., SALAZAR M., HAN H., CHEMERIS V. V., KERWIN S. M., HURLEY L. H. (1998) NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry* **37**, 12367-12374.

FERNANDO H., RESZKA A. P., HUPPERT J., LADAME S., RANKIN S., VENKITARAMAN A. R., NEIDLE S., BALASUBRAMANIAN S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene.

*Biochemistry* **45**, 7854-7860.

FIALOVÁ M., KYPR J., VORLÍČKOVÁ M. (2006) The thrombin binding aptamer GGTTGGTGTGGTTGG forms a bimolecular guanine tetraplex. *Biochemical and Biophysical Research Communications* **344**, 50-54.

FLETCHER T. M., SUN D., SALAZAR M., HURLEY L. H. (1998) Effect of DNA secondary structure on human telomerase activity. *Biochemistry* **37**, 5536-5541.

FLORIS R., SCAGGIANTE B., MANZINI G., QUADRIFOGLIO F., XODO L. E. (1999) Effect of cations on purine•purine•pyrimidine triple helix formation in mixed-valence salt solutions. *European Journal of Biochemistry* **260**, 801-809.

FUJIHASHI T., SAKATA T., KAJI A., KAJI H. (1994) Short, terminally phosphorylated oligoriboguanylic acids effectively inhibit cytopathicity caused by human immunodeficiency virus. *Biochemical and Biophysical Research Communications* **203**, 1244-1250.

GABELICA V., ROSU F., DE PAUW E., LEMAIRE J., GILLET J. C., POULLY J. C., LECOMTE F., GREGOIRE G., SCHERMANN J. P., DESFRANCOIS C. (2008) Infrared signature of DNA G-quadruplexes in the gas phase. *Journal of the American Chemical Society* **130**, 1810-1811.

GALLAGHER C. T. (2004) NMR Studies of nucleic acids as drug targets, *PhD thesis*, the University of Nottingham.

GAVATHIOTIS E., HEALD R. A., STEVENS M. F. G., SEARLE M. S. (2003) Drug recognition and stabilization of the parallel stranded DNA quadruplex $d(TTAGGGT)_4$ containing the human telomeric repeat. *Journal of Molecular Biology* **334**, 25-36.

GELLERT M., LIPSETT M. N., DAVIES D. R. (1962) Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 2013-2018.

GEORGIADES S. N., ABD KARIM N. H, SUTHARALINGAM K., VILAR R. (2010) Interaction of metal complexes with G-quadruplexes DNA. *Angewandte Chemie - International Edition* **49**, 4020-4034.

GHOSAL G., MUNIYAPPA K. (2006) Hoogsteen base-pairing revisited: Resolving a role in normal biological processes and human diseases. *Biochemical and Biophysical Research Communications* **343**, 1-7.

GILSON G., LAROCHE T., GASSER S. M. (1993) Telomeres and the functional architecture of the nucleus. *Trends in Cell Biology* **3**, 128-134.

GOMEZ D., LEMARTELEUR T., LACROIX L., MAILLIET P., MERGNY J. L., RIOU J. F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Research* **32**, 371-379.

GRAND C. L., HAN H. Y., MUNOZ R. M., WEITMAN S., VON HOFF D. D., HURLEY L. H., BEARSS D. J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Molecular Cancer Therapeutics* **1**, 565-573.

GRANGER M. P., WRIGHT W. E., SHAY J. W. (2002) Telomerase in cancer and aging. *Critical Reviews in Oncology Hematology* **41**, 29-40.

GRANOTIER C., PENNARUN G., RIOU L., HOFFSCHIR F., GAUTHIER L. R., DE CIAN A., GOMEZ D., MANDINE E., RIOU J. F., MERGNY J. L., MAILLIET P., DUTRILLAUX B., BOUSSIN F. D. (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Research* **33**, 4182-4190.

GRAY R. D., BUSCAGLIA R., CHAIRES J. B. (2012) Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *Journal of the American Chemical Society* **134**, 16834-16844.

GRAY R. D., CHAIRES J. B. (2011) Analysis and multimensional G-quadruplex melting curves. *Current Protocols in Nucleic Acid Chemistry*, Chapter: Unit17.4.

GRAY D. M., WEN J. D., GRAY C. W., REPGES R., REPGES C., RAABE G., FLEISCHHAUER J. (2008) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **20**, 431-440.

GREEN J. J., YING L., KLENERMAN D., BALASUBRAMANIAN S. (2003) Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap. *Journal of the American Chemical Society* **125**, 3763-3767.

GROS J., ROSU F., AMRANE S. (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Research* **35**, 3064-3075.

GUAN A. J., ZHANG E. X., XIANG J. F., LI Q., YANG Q. F., LI L., TANG Y. L., WANG M. X. (2011) Effects of loops and nucleotides in G-quadruplexes on their interaction with an azacalixarene, methylazacalix[6]pyridine. *Journal of Physical Chemistry B* **115**, 12584-12590.

GUO K., POURPAK A., BEETZ-ROGERS K., GOKHALE V., SUN D., HURLEY L. H. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the

proximal promoter region of the RET oncogene. *Journal of the American Chemical Society* **129**, 10220-10228.

GUZMÁN M. R., LIQUIER J., BRAHMACHARI S. K., TAILLANDIER E. (2006) Characterization of parallel and antiparallel G-tetraplex structures by vibrational spectroscopy. *Spectrochimica Acta Part A: Molecular and Bimolecular Spectroscopy* **64**, 495-503.

HAIDER S. M., NEIDLE S., PARKINSON G. N. (2011) A structural analysis of G-quadruplex/ligand interactions. *Biochimie* **93**, 1239-1251.

HAIDER S. M., PARKINSON G. N., NEIDLE S. (2003) Structure of a G-quadruplex–ligand complex. *Journal of Molecular Biology* **326**, 117-125.

HAIDER S., PARKINSON G. N., NEIDLE S. (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical Journal* **95**, 296-311.

HALDER K., CHOWDHURY S. (2005) Kinetic resolution of bimolecular hybridization versus intramolecular folding in nucleic acids by surface plasmon resonance: application to G-quadruplex/duplex competition in human c-myc promoter. *Nucleic Acids Research* **33**, 4466-4474.

HAMON F., LARGY E., GUÉDIN-BEAUREPAIRE A., ROUCHON-DAGOIS M., SIDIBE A., MONCHAUD D., MERGNY J. L., RIOU J. L. NGUYEN C. H., TEULADE-FICHOU M. P. (2011) An acyclic oligoheteroaryle that discriminates strongly between diverse G-quadruplex topologies. *Angewandte Chemie – International Edition* **50**, 8745-8749.

HAN H., LANGLEY D. R., RANGAN A., HURLEY L. H. (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. *Journal of the American Chemical Society* **123**, 8902-8913.

HAN F. X., WHEELHOUSE R. T., HURLEY L. H. (1999) Interaction of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *Journal of the American Chemical Society* **121**, 3561-3570.

HANAKAHI L. A., SUN H., MAIZELS N. (1999) High affinity interactions of nucleolin with G-G paired rDNA. *Journal of Biological Chemistry* **274**, 15908-15912.

HANUŠ J., CHMELOVÁ K., ŠTĚPÁNEK J., TURPIN P. Y., BOK J., ROSENBERG I., TOČÍK Z. (1999) Raman spectroscopic study of triplex-like complexes of polyuridylic acid with the isopolar, non-isosteric phosphonate analogues of diadenosine monophosphate. *Journal of Raman spectroscopy* **30**, 667-676.

HANUŠ J., NĚMEČEK D., ŠTĚPÁNEK J., TURPIN P. Y., KRÁLÍKOVÁ Š, BOK J., ROSENBERG I. (2004) Structural compatibility of novel nucleotide modifications with shortened linkages designed for antigene/antisense therapy. *Journal of Raman Spectroscopy* **35**, 418-425.

HAQ I., TRENT J. O., CHOWDHRY B. Z., JENKINS T. C. (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin. *Journal of the American Chemical Society* **121**, 1768-1779.

HARDIN C. C., CORREGAN M., BROWN B.A. 2$^{ND}$, FREDERICK L. N. (1993) Cytosine-cytosine$^{+}$ base pairing stabilizes DNA quadruplexes and cytosine methylation greatly enhances the effect. *Biochemistry* **32**, 5870-5880.

HARDIN C. C., PERRY A. G., WHITE K. (2001) Thermodynamic and kinetic characterization of the dissociation an assembly of quadruplex nucleic acids. *Biopolymers* **56**, 147-194.

HARDIN C. C., WATSON T., CORREGAN M., BAILEY C. (1992) Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG$_3$GCG). *Biochemistry* **31**, 833-841.

HAZEL P., HUPPERT J., BALUSUBRAMANIAN S., NEIDLE S. (2004) Loop-length dependent folding of G-quadruplexes. *Journal of the American Chemical Society* **126**, 16405-16415.

HE F., TANG Y. L., WANG S., LI Y. L., ZHU D. B. (2005) Fluorescent amplifying recognition for DNA G-quadruplex folding with a cationic conjugated polymer: A platform for homogeneous potassium detection. *Journal of the American Chemical Society* **127**, 12343-12346.

HEDDI B., PHAN A. T. (2011) Structure of human telomeric DNA in crowded solution. *Journal of the American Chemical Society* **133**, 9824-9833.

HEMANN M. T., STRONG M. A., HAO L. Y., GREIDER C. W. (2001) The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67-77.

HERRERA S. J. E., ŠTĚPÁNEK J. (2010) Raman study of magnesium induced conversion of polyU·polyA duplexes to polyU·polyA·polyU triplexes. *Spectroscopy: An International Journal* **24**, 445-448.

HONG E. S., YOON H. J., KIM B., YIM Y. H., SO H. Y., SHIN S. K. (2010) Mass spectrometric studies of alkali metal ion binding on thrombin-binding aptamer DNA. *Journal of the American Society for Mass Spectrometry* **21**, 1245-1255.

HORIKAWA I., BARRET J. C. (2003) Transcriptional regulation of the telomerase hTERT gene as a target for cellular and viral oncogenic mechanisms. *Carcinogenesis* **24**, 1167-1176.

HUANG X., NAKANISHI K., BEROVA N. (2000) Porphyrins and metalloporphyrins: Versatile circular dichroic reporter groups for structural studies. *Chirality* **12**, 237-255.

HUPPERT J. L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chemical Society Reviews* **37**, 1375-1384.

HUPPERT J. L., BALASUBRAMANIAN S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Research* **33**, 2908-2916.

HWANG E. S. (2002) Replicative senescence and senescence-like state induced in cancer-derived cells. *Mechanisms of Ageing and Development* **123**, 1681-1694.

IZBICKA E., WHEELHOUSE R. T., RAYMOND E., DAVIDSON K. K, LAWRENCE R. A., SUN D., WINDLE B. E., HURLEY L. H., VON HOFF D. D. (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Research* **59**, 639-644.

JACKSON S. P. (2001) Detecting, signalling and repairing DNA double-strand breaks. *Biochemical Society Transactions* **29**, 655-661.

JASINSKA A., MICHLEWSKI G., DE MEZER M., SOBCZAK K., KOZLOWSKI P., NAPIERALA M., KRZYZOSIAK W. J. (2003) Structures of trinucleotide repeats in human transcripts and their functional implications. *Nucleic Acids Research* **31**, 5463-5468.

JAUMOT J., ERITJA R., NAVEA S., GARGALLO R. (2009) Classification of nucleic acids structures by means of the chemometric analysis of circular dichroism spectra. *Analytica Chimica Acta* **642**, 117-126.

JAUMOT J., ERITJA R., TAULER R., GARGALLO R. (2006) Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand. *Nucleic Acids Research* **34**, 206-216.

JELESAROV I., BOSSHARD H. R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of bimolecular recognition. *Journal of Molecular Recognition* **12**, 3-18.

JERSCHOW A., MULLER N. (1997) Suppression of convection artifacts in stimulated-echo diffusion experiments. Double-stimulated-echo experiments. *Journal of Magnetic Resonance* **125**, 372-375.

JING N., DE CLERCQ E., RANDO R. F., PALLANSCH L., LACKMAN-SMITH C., LEE S., HOGAN M. E. (2000) Stability-activity relationships of a family of G-tetrad forming oligonucleotides as potent HIV inhibitors. A basis for anti-HIV drug design. *Journal of Biological Chemistry* **275**, 3421-3430.

JING N. J., RANDO R. F., POMMIER Y., HOGAN M. E. (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide. *Biochemistry* **36**, 12498-12505.

JING N., XU X. (2001) Rational drug design of DNA oligonucleotides as HIV inhibitors. *Current Drug Targets - Infectious Disorders* **2**, 79-90.

JOACHIMI A., MAYER G., HARTIG J. S. (2007) A new anticoagulant-antidote pair: Control of thrombin activity by aptamers and porphyrins. *Journal of the American Chemical Society* **129**, 3036–3037.

JUSKOWIAK B. (2011) Nucleic acid-based fluorescent probes and their analytical potential. *Analytical and Bionalytical Chemistry* **399**, 3157-3176.

KALLANSRUD G., WARD B. (1996) A comparison of measured and calculated single- and double-stranded oligodeoxynucleotide extinction coefficients. *Analytical Biochemistry* **236**, 134-138

KANG CH., BERGER I., LOCKSHIN C., RATLIFF R., MOYZIS R., RICH A. (1995) Stable loop in the crystal-structure of the intercalated 4-stranded cytosine-rich metazoan telomere. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 3874-3878.

KANKIA B. I. (2004) Optical absorption assay for strand-exchange reactions in unlabeled nucleic acids. *Nucleic Acids Research* **32**, article number: e154.

KANKIA B. I., MARKY L. A. (2001) Folding of the thrombin aptamer into G-quadruplex with $Sr^{+2}$: Stability, heat, and hydration. *Journal of the American Chemical Society* **123**, 10799-10804.

KANO K., TAKEI M., HASHIMOTO S. (1990) Cationic porphyrins in water. $^{1}H$ NMR and fluorescence studies on dimer and molecular complex formation. *Journal of Physical Chemistry* **94**, 2181-2187.

KARSISIOTIS A. I., HESSARI N. M., NOVELLINO E., SPADA G. P., RANDAZZO A., DA SILVA M. W. (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angewandte Chemie-International Edition* **50**, 10645-10648.

KATO Y., OHYAMA T., MITA H., YAMAMOTO Y. (2005) Dynamics and

thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAG$_n$) (n=3-5). *Journal of the American Chemical Society* **127**, 9980–9981.

KAUCHER M. S., HARREL W. A., DAVIS J. T. (2006) A Unimolecular G-Quadruplex that functions as a synthetic transmembrane $Na^+$ transporter. *Journal of the American Chemical Society* **128**, 38-39.

KEATING L. R., SZALAI V. A. (2004) Parallel-stranded guanine quadruplex interactions with a copper cationic porphyrin. *Biochemistry* **43**, 15891-15900.

KEJNOVSKÁ I., KYPR J., VORLÍČKOVÁ M. (2007) Oligo(dT) is not a correct native PAGE marker for single-stranded DNA. *Biochemical and Biophysical Research Communications* **353**, 776-779.

KELLAND L. R. (2005) Overcoming the immortality of tumour cells by telomere and telomerase based cancer therapeutics-current status and future prospects. *European Journal of Cancer* **41**, 971-979.

KENIRY M. A. (2001) Quadruplex structures in nucleic acids. *Biopolymers* **56**, 123-146.

KERN J. T., THOMAS P. W., KERWIN S. M. (2002) The relationship between ligand aggregation and G-quadruplex DNA selectivity in a series of 3,4,9,10-perylenetetracarboxylic acid diimides. *Biochemistry* **41**, 11379-11389.

KIM M. Y., GLEASON-GUZMAN M., IZBICKA E., NISHIOKA D., HURLEY L. H. (2003) The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Research* **63**, 3247-3256.

KOČIŠOVÁ E., PROCHÁZKA M., ŠTĚPÁNEK J., MOJZEŠ P. (2010) Interaction of porphyrin/oligonucleotide complex with liposomes studied by drop coating deposition Raman spectroscopy. *Spectroscopy: An International Journal* **24**, 197-200.

KOIZUMI M., AKAHORI K., OHMINE T., TSUTSUMI S., SONE J., KOSAKA T., KANEKO M., KIMURA S., SHIMADA K. (2000) Biologically active oligodeoxyribonucleotides. Part 12: $N^2$-methylation of 2'-deoxyguanosines enhances stability of parallel G-quadruplex and anti-HIV-1 activity. *Bioorganic and Medicinal Chemistry Letters* **10**, 2213-2216.

KRAFFT C., BENEVIDES J. M., THOMAS G. J., JR. (2002) Secondary structure polymorphism in Oxytricha nova telomeric DNA. *Nucleic Acids Research* **30**, 3981-3991.

KRAUSS I. R., MERLINO A., GIANCOLA C., RANDAZZO A., MAZZARELA L., SICA F. (2011) Thrombin–aptamer recognition: a revealed ambiguity. *Nucleic Acids Research* **39**, 7858-7867.

KRAUSS I. R., MERLINO A., RANDAZZO A., NOVELLINO E., MAZZARELLA L., SICA F. (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Research* **40**, 8119-8128.

KRUGLIK S. G., MOJZES P., MIZUTANI Y., KITAGAWA T., TURPIN P. Y. (2001) Time-resolved resonance Raman study of the exciplex formed between excited Cu-porphyrin and DNA. *Journal of Physical Chemistry B* **105**, 5018-5031.

KŘÍŽ M., NĚMEČEK D., TURPIN P. Y., ROSENBERG I., ŠTĚPÁNEK J. (2008) Structural effects of internucleotide linkage modification at mismatch site in DNA/RNA duplex: Raman study. *Vibrational Spectroscopy* **48**, 148-152.

KŘÍŽ M., SNÁŠEL J., KOPECKÝ V., PÁV O., ROSENBERG I., ŠTĚPÁNEK J. (2012) Structural changes of human RNase L upon homodimerization investigated by Raman spectroscopy. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1824**, 1039-1044.

KUMAR N., MAITI S. (2004) Quadruplex to Watson-Crick duplex transition of the thrombin binding aptamer: a fluorescence resonance energy transfer study. *Biochemical and Biophysical Research Communications* **319**, 759-767.

KUMAR N., SAHOO B., VARUN K. A. S., MAITI S., MAITI S. (2008) Effect of loop length variation on quadruplex - Watson Crick duplex competition. *Nucleic Acids Research* **36**, 4433-4442.

KURYAVYI V., KETTANI A., WANG W., JONES R., PATEL D. J. (2000) A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads. *Journal of Molecular Biology* **295**, 455-469.

KYPR J., KEJNOVSKÁ I., RENČIUK D. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Research* **37**, 1713-1725.

LANE A. N., CHAIRES J. B., GRAY R. D., TRENT J. O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research* **36**, 5482-5515.

LAPORTE L., THOMAS G. J., JR. (1998) Structural basis of DNA recognition and mechanism of quadruplex formation by the β subunit of the *Oxytricha* telomere binding protein. *Biochemistry* **37**, 1327-1335.

LEE J. S. (1990) The stability of polypurine tetraplexes in the presence of mono- and

divalent cations. *Nucleic Acids Research* **18**, 6057-6060.

LEE Y. A., LEE S., LEE H. M., LEE C. S., KIM S. K. (2003) Anomalous excitonic CD of meso-tetrakis(3-N-methylpyridiniumyl)porphyrin bound to poly[d(A-T)(2)]. *Journal of Biochemistry* **133**, 343-349.

LENA S., MASIERO S., PIERACCINI S., SPADA G. P. (2009) Guanosine hydrogen-bonded scaffolds: A new way to control the bottom-up realisation of well-defined nanoarchitectures. *Chemistry – A European Journal* **15**, 7792-7806.

LEW A., RUTTER W. J., KENNEDY G. C. (2000) Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12508-12512.

LI J., CORREIA J. J., WANG L., TRENT J. O., 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 Research* **33**, 4649–4659.

LI Z., MIRKIN C. A. (2005) G-quartet-induced nanoparticle assembly. *Journal of the American Chemical Society* **127**, 11568-11569.

LI W., MIYOSHI D., NAKANO S., SUGIMOTO N. (2003) Structural competition involving G-quadruplex DNA and its complement. *Biochemistry* **42**, 11736-11744.

LIM K. W., ALBERTI P., GUEDIN A., LACROIX L., RIOU J. F., ROYLE N. J., MERGNY J. L., PHAN A. T. (2009a) Sequence variant $(CTAGGG)_n$ in the human telomere favors a G-quadruplex structure containing a G·C·G·C tetrad. *Nucleic Acids Research* **37**, 6239-6248.

LIM K. W., AMRANE S., BOUAZIZ S., XU W. X., MU Y. G., PATEL D. J, LUU K. N., PHAN A. T. (2009b) Structure of the human telomere in $K^+$ solution: A stable basket – type G-quadruplex with only two G-tetrad layers. *Journal of the American Chemical Society* **131**, 4301-4309.

LIPPS H. J., RHODES D. (2009) G-quadruplex: *in vivo* evidence and function. *Trends in Cell Biology* **19**, 414-422.

LIQUIER J., GOUYETTE C., HUYNH-DINH T., TAILLANDIER E. (1999) Raman spectroscopy of nucleic acid triple helices. *Journal of Raman spectroscopy* **30**, 657-666.

LIU J. P. (1999) Studies of the molecular mechanisms in the regulation of telomerase activity. *FASEB Journal* **13**, 2091-2104.

LIU J. W., CAO Z. H., LU Y. (2009) Functional nucleic acid sensors. *Chemical Reviews* **109**, 1948-1998.

LUBITZ I., BOROVOK N., KOTLYAR A. (2007) Interaction of monomolecular G4-DNA nanowires with TMPyP: Evidence intercalation. *Biochemistry* **46**, 12925-12929.

LUNDBLAD V. (2000) DNA ends: maintenance of chromosome termini versus repair of double strand breaks. *Mutation Research – Fundamental and Molecular Mechanisms of Mutagenesis* **451** 227-240.

LUU K. N., PHAN A. T., KURYAVYI V., LACROIX L., PATEL D. J. (2006) Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold. *Journal of the American Chemical Society* **128**, 9963-9970.

LYDALL D. (2003) Hiding at the ends of yeast chromosomes: telomeres, nucleases and checkpoint pathways. *Journal of Cell Science* **116**, 4057-4065.

LYONNAIS S., HOUNSOU C., TEULADE-FICHOU M. P., JEUSSET J., LE CAM E., MIRAMBEAU G. (2002) G-quartets assembly within a G-rich DNA flap. A possible event at the center of the HIV-1 genome. *Nucleic Acids Research* **30,** 5276-5283.

MACAYA R. F., SCHULTZE P., SMITH F. W., ROE J. A., FEIGON J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3745-3749.

MAIRAL T., OZALP V. C., SANCHEZ P. L., MIR M., KATAKIS I., O'SULLIVAN C. K. (2008) Aptamers: molecular tools for analytical applications. *Analytical and Bioanalytical Chemistry* **390**, 989-1007.

MAJHI P. R., QI J. Y., TANG C. F., SHAFER R. H. (2007) Heat capacity changes associated with guanine quadruplex formation: An isothermal titration calorimetry study. *Biopolymers* **89**, 302-309.

MAJHI P. R., SHAFER R. H. (2006) Characterization of an unusual folding pattern in a catalytically active guanine quadruplex structure. *Biopolymers* **82**, 558-569.

MALINOWSKI E. R. Factor Analysis in Chemistry (Third Edition); John Wiley & Sons, Inc.: New York (2002). ISBN 0-471-13479-1.

MAO X., GMEINER W. H. (2005) NMR study of the folding-unfolding mechanism for the trombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *Biophysical Chemistry* **113**, 155-160.

MARATHIAS V. M., BOLTON P. H. (2000) Structures of the potassium-saturated, 2:1, and intermediate, 1:1, forms of a quadruplex DNA. *Nucleic Acids Research* **28**,

1969-1977.

MARSH T. C., VESENKA J., HENDERSON E. (1995) A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy. *Nucleic Acids Research* **23**, 696-700.

MARTADINATA H., PHAN A. T. (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in $K^+$ solution. *Journal of the American Chemical Society* **131**, 2570-2578.

MARTINO L., PAGANO B., FOTTICCHIA I., NEIDLE S., GIANCOLA C. (2009) Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *Journal of Physical Chemistry B* **113**, 14779-14786.

MARUSIC M., SKET P., BAUER L., VIGLASKY V., PLAVEC J. (2012) Solution-state structure of an intramolecular G-quadruplex with propeller, diagonal and edgewise loops. *Nucleic Acids Research* **40**, 6946-6956.

MASIERO S., TROTTA R., PIERACCINI S., DE TITO S., PERONE R., RANDAZZO A., SPADA G. P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Organic & Biomolecular Chemistry* **8**, 2683-2692.

MCMILLIN D. R., MCNETT K. M. (1998) Photoprocesses of copper complexes that bind to DNA. *Chemical Reviews* **98**, 1201-1219.

MENDEZ M. A., SZALAI V. A. (2009) Fluorescence of unmodified oligonucleotides: A tool to probe G-quadruplex DNA structure. *Biopolymers* **91**, 841-850.

MERGNY J. L. (1999) Fluorescence energy transfer as a probe for tetraplex formation: The i-Motif. *Biochemistry* **38**, 1573-1581.

MERGNY J. L., DE CIAN A., GHELAB A., SACCA B., LACROIX L. (2005a) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Research* **33**, 81-94.

MERGNY J. L., LACROIX L. (2003) Analysis of thermal melting curves. *Oligonucleotides* **13**, 515-537.

MERGNY J. L., LACROIX L. (2009) UV melting of G-quadruplexes. *Current Protocols in Nucleic Acid Chemistry*. Unit 17.1.

MERGNY J. L., LI L., LACROIX L., AMRANE S., CHAIRES J. B. (2005b) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Research* **33**, article number: e138.

MERGNY J. L., PHAN A. T., LACROIX L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Letters* **435**, 74-78.

MERGNY J. L., RIOU J. F., MAILLIET P., TEULADE-FICHOU M. P., GILSON E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Research*

**30**, 839-865.

MEYNE J., RATLIFF R. F., MOYZIS R. K. (1989) Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 7049-7053.

MICURA R., PILS W., HOBARTNER C., GRUBMAYR K., EBERT M. O., JAUN B. (2001) Methylation of the nucleobases in RNA oligonucleotides mediates duplex-hairpin conversion. *Nucleic Acids Research* **29**, 3997-4005.

MIRKIN S. M. (2006) DNA structures, repeat expansions and human hereditary disorders. *Current Opinion in Structural Biology* **16**, 351-358.

MITA H., OHYAMA T., TANAKA Y., YAMAMOTO Y. (2006) Formation of a complex of 5,10,15,20-tetrakis(N-methylpyridinium-4-yl)-21H,23H-porphyrin with G-quadruplex DNA. *Biochemistry* **45**, 6765-6772.

MIURA T., BENEVIDES J. M., THOMAS G. J., JR. (1995) A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *Journal of Molecular Biology* **248**, 233-238.

MIURA T., THOMAS G. J., JR. (1994) Structural polymorphism of telomere DNA: Interquadruplex and duplex-quadruplex conversions probed by Raman spectroscopy. *Biochemistry* **33**, 7848-7856.

MIURA T., THOMAS G. J., JR. (1995) Structure and dynamics of interstrand guanine association in quadruplex telomeric DNA. *Biochemistry* **34**, 9645-9654.

MIYOSHI D., KARIMATA H., SUGIMOTO N. (2006) Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. *Journal of the American Chemical Society* **128**, 7957-7963.

MOHANTY D., BANSAL M. (1993) Conformational polymorphism in G-tetraplex structures – strand reversal by base flipover or sugar flipover. *Nucleic Acids Research* **21**, 1767-1774.

MOJZEŠ P., KRUGLIK S. G., BAUMRUK V., TUIRPIN P. Y. (2003) Interactions of electronically excited copper(II)−porphyrin with DNA: Resonance Raman evidence for the exciplex formation with adenine and cytosine residues. *Journal of Physical Chemistry B* **107**, 7532-7535.

MOJZEŠ P., PALACKÝ J., BAUEROVÁ V., BEDNÁROVÁ L. (2011) Ramanova mikrospektroskopie a mapování buněk a tkání. *Československý časopis pro fyziku* **61**, 178 – 184.

MONCHAUD D., TEULADE-FICHOU M. P. (2008) A hitchhiker's guide to G-quadruplex

ligands. *Organic & Biomolecular Chemistry* **6**, 627-636.

MONDRAGON-SANCHEZ J. A., LIQUIER J., SHAFER R. H., TAILLANDIER E. (2004) Tetraplex structure formation in the thrombin-binding DNA aptamer by metal cations measured by vibrational spectroscopy. *Journal of Biomolecular Structure & Dynamics* **22**, 365–373.

MOVILEANU L., BENEVIDES J. M., THOMAS G. J., JR. (2002) Determination of base and backbone contributions to the thermodynamics of premelting and melting transitions in B DNA. *Nucleic Acids Research* **30**, 3767–3777.

MUNTONI A., REDDEL R. R. (2005) The first molecular details of ALT in human tumor cells. *Human Molecular Genetics* **14**, R191-R196.

MURASHIMA T., SAKIYAMA D., MIYOSHI D., KURIYAMA M., YAMADA T., MIYAZAWA T., SUGIMOTO N. (2008) Cationic porphyrins induced a telomeric DNA to G-quadruplex form in water. *Bioinorganic Chemistry and Applications* **2008**, article number: 294756.

MURAT P., SINGH Y., DEFRANCQ E. (2011) Methods for investigating G-quadruplex DNA/ligand interactions. *Chemical Society Reviews* **40**, 5293-5307.

MURCHIE A. I. H., LILLEY D. M. J. (1992) Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids Research* **20**, 49–53.

MURPHY J. H., TRAPANE T. L. (1996) Concentration and extinction coefficient determination for oligonucleotides and analogs using a general phosphate analysis. *Analytical Biochemistry* **240**, 273-282.

NAGATOISHI S., ISONO N., TSUMOTO K., SUGIMOTO N. (2011) Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie* **93**, 1231-1238.

NAGATOISHI S., NOJIMA T., GALEZOWSKA E., JUSKOWIAK B., TAKENAKA S. (2006) G quadruplex-based FRET probes with the thrombin-binding aptamer (TBA) sequence designed for the efficient fluorometric detection of the potassium ion. *Chembiochem* **7**, 1730-1737.

NAIR V. (2002) HIV integrase as a target for antiviral chemotherapy. *Reviews in Medical Virology* **12**, 179-193.

NEAVES K. J., HUPPERT J. L., HENDERSON R. M., EDWARDSON J. M. (2009) Direct visualization of G-quadruplexes in DNA using atomic force microscopy. *Nucleic Acids Research* **37**, 6269-6275.

NEIDLE S., BALASUBRAMANIAN S.: Quadruplex nucleic acids, The royal society of chemistry, Cambridge 2006, ISBN 10: 0-85404-374-8.

NEIDLE S., PARKINSON G. N. (2008) Quadruplex DNA crystal structures and drug design. *Biochimie* **90**, 1184-1196.

NEIDLE S., READ M. A. (2001) G-quadruplexes as therapeutic targets. *Biopolymers* **56**, 195-208.

NĚMEČEK D., OVERMAN S. A., HENDRIX R. W., THOMAS G. J., JR. (2009) Unfolding thermodynamics of the Δ-domain in the prohead I subunit of phage HK97: Determination by factor analysis of Raman spectra. *Journal of Molecular Biology* **385**, 628-641.

NĚMEČEK D., VAISOCHEROVÁ H., ŠTĚPÁNEK J., TUIRPIN P. Y. (2005) Structural features of a central mismatch in oligonucleotide hybrid duplexes visualized via Raman spectroscopy: Model system for evaluation of potential ‘‘antisense’’ drugs. *Biopolymers* **79**, 1-8.

NOVÝ J., BŐHM S., KRÁLOVÁ J., KRÁL V., URBANOVÁ M. (2008) Formation and temperature stability of G-quadruplex structures studied by electronic and vibrational circular dichroism spectroscopy combined with ab initio calculations. *Biopolymers* **89**, 144-152.

NUTIU R., LI Y. F. (2004) Structure-switching signaling aptamers: transducing molecular recognition into fluorescence signaling. *Chemistry - A European Journal* **10**, 1868-1876.

OLAUSSEN K. A., DUBRANA K., DOMONT J., SPANO J. P., SABATIER L., SORIA J. C. (2006) Telomeres and telomerase as targets for anticancer drug development. *Critical Reviews in Oncology Hematology* **57**, 191-214.

OLSEN C. M., GMEINER W. H., MARKY L. A. (2006a) Unfolding of G-quadruplexes: Energetic, and ion and water contributions of G-quartet stacking. *Journal of Physical Chemistry* **110**, 6962-6969.

OLSEN C. M., GMEINER W. H., MARKY L. A. (2006b) Interaction of $Cd^{2+}$ with G-quadruplexes containing $K^{+}$ or $Sr^{2+}$. *Journal of Biomedical Nanotechnology* **2**, 62-70.

OLSEN C. M., LEE H. T., MARKY L. A. (2009) Unfolding thermodynamics of intramolecular G-quadruplexes: Base sequence contributions of the loops. *Journal of Physical Chemistry B* **113**, 2587-2595.

OTTOVÁ P., ESPINOZA-HERRERA S. J., ŠTĚPÁNEK J. (2011) Magnesium effect on

premelting transitions in nucleic acids: DNA duplex and RNA hairpin models. *Journal of Molecular Structure* **993**, 324-327.

PADMANABHAN K., PADMANABHAN K. P., FERRARA J. D., SADLER J. E., TULINSKY A. (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *Journal of Biological Chemistry* **268**, 17651-17654.

PADMANABHAN K., TULINSKY A. (1996) An ambiguous structure of a DNA 15-mer trombin complex. *Acta Crystallographica Section D – Biological Crystallography* **52**, 272-282.

PAESCHKE K., SIMONSSON T., POSTBERG J., RHODES D., LIPPS H. J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature Structural & Molecular Biology* **12**, 847-854.

PAGANO B., MARTINO L., RANDAZZO A., GIANCOLA C. (2008) Stability and binding properties of a modified thrombin binding aptamer. *Biophysical Journal* **94**, 562-569.

PAGANO B., MATTIA C. A., GIANCOLA C. (2009) Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes. *International Journal of Molecular Sciences* **10**, 2935-2957.

PAGBA C. V., LANE S. M., WACHSMANN-HOGIU S. (2010) Raman and surface-enhanced Raman spectroscopic studies of the 15-mer DNA thrombin-binding aptamer. *Journal of Raman Spectroscopy* **41**, 241-247.

PALACKÝ J. (2005) Properties of novel modified oligonucleotides – Potential chemotherapeutics – Studied via spectroscopic methods, *Diploma thesis*, Charles University in Prague.

PALUMBO S. L., MEMMOTT R. M., URIBE D. J, KROTOVA-KHAN Y., HURLEY L. H., EBBINGHAUS (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Research* **36**, 1755-1769.

PARKINSON G. N., GHOSH R., NEIDLE S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry* **46**, 2390-2397.

PARKINSON G. N., LEE M. P. H., NEIDLE S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876-880.

PASTERNACK R. F., BRIGANDI R. A., ABRAMS M. J., WILLIAMS A. P., GIBBS E. J. (1990) Interaction of porphyrins and metalloporphyrins with single-stranded poly(dA). *Inorganic Chemistry* **29**, 4483- 4486.

PASTERNACK R. F., EWEN S., RAO A., MEYER A. S., FREEDMAN M. A., COLLINGS P. J., FREY S. L., RANEN M. C., DE PAULA J. C. (2001) Interactions of copper(II) porphyrins with DNA. *Inorganica Chimica Acta* **317**, 59-71.

PASTERNACK R. F., GIANNETTO A., PAGANO P., GIBBS E. J. (1991) Self-assembly of porphyrins on nucleic acids and polypeptides. *Journal of the American Chemical Society* **113**, 7799–7800.

PASTERNACK R. F., GIBBS E. J., VILLAFRANCA J. J. (1983) Interactions of porphyrins with nucleic acids. *Biochemistry* **22**, 2406-2414.

PASTERNACK R. F., GOLDSMITH J. I., SZÉP S., GIBBS E. J. (1998) A spectroscopic and thermodynamic study of porphyrin/DNA supramolecular assemblies. *Biophysical Journal* **75**, 1024-1031.

PASTERNAK A., HERNANDEZ F. J., RASMUSSEN L. M., VESTER B., WENGEL J. (2011) Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer. *Nucleic Acids Research* **39**, 1155-1164.

PATEL S. D., ISALAN M., GAVORY G., LADAME S., CHOO Y., BALASUBRAMANIAN S. (2004) Inhibition of human telomerase activity by an engineered zinc finger protein that binds G-quadruplexes. *Biochemistry* **43**, 13452-13458.

PAYET L., HUPPERT J. L. (2012) Stability and structure of long intramolecular G-quadruplexes. *Biochemistry* **51**, 3154-3161.

PEŇÁZOVÁ H., VORLÍČKOVÁ M. (1997) Guanine tetraplex formation by short DNA fragments containing runs of guanine and cytosine. *Biophysical Journal* **73**, 2054-2063.

PERREM K., COLGIN L. M., NEUMANN A. A., YEAGER T. R., REDDEL R. R. (2001) Coexistence of alternative lengthening of telomeres and telomerase in hTERT-transfected GM847 cells. *Molecular and Cellular Biology* **21**, 3862-3875.

PETRACCONE L., ERRA E., ESPOSITO V., RANDAZZO A., GALEONE A., BARONE G., GIANCOLA C. (2005) Biophysical properties of quadruple helices of modified human telomeric DNA. *Biopolymers* **77**, 75-85.

PETRACCONE L., GARBETT N. C., CHAIRES J. B., TRENT J. O. (2010) An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. *Biopolymers* **93**, 553-548.

PETRACCONE L., PAGANO B., GIANCOLA C. (2012) Studying the effect of crowding and dehydration on DNA G-quadruplexes. *Methods* **57**, 76-83.

PETROVIC A. G., POLAVARAPU P. L. (2008) The quadruplex-duplex structural

transition of polyriboguanylic acid. *Journal of Physical Chemistry B* **112**, 2245-2254.

PHAN A. T., KURYAVYI V., MA J. B., FAURE A., ANDREOLA M. L., PATEL D. J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 634-639.

PHAN A. T., MERGNY J. L. (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Research* **30**, 4618-4625.

PHAN A. T., MODI Y. S., PATEL D. J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *Journal of the American Chemical Society* **126**, 8710-8716.

PHAN A. T., PATEL D. J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *Journal of the American Chemical Society* **125**, 15021-15027.

PIVETTA C., LUCATELLO L., KRAPCHO A. P., GATTO B., PALUMBO M., SISSI C. (2008) Perylene side chains modulate G-quadruplex conformation in biologically relevant DNA sequences. *Bioorganic & Medicinal Chemistry* **16**, 9331-9339.

PONIKOVÁ S., TLUČKOVÁ K., ANTALÍK M., VÍGLASKÝ V., HIANIK T. (2011) The circular dichroism and differential scanning calorimetry study of the properties of DNA aptamer dimers. *Biophysical Chemistry* **155**, 29-35.

PONS J. L., MALLIAVIN T. E., DELSUC M. A. (1996) Gifa V4: A complete package for NMR data set processing. *Journal of Biomolecular NMR* **8**, 445-452.

POON K., MACGREGOR R. B., JR. (1999) Probing the structure of multi-stranded guanine-rich DNA complexes by Raman spectroscopy and enzymatic degradation. *Biophysical Chemistry* **79**, 11-23.

PRADHAN S. K., DASGUPTA D., BASU G. (2011) Human telomere d[$(TTAGGG)_4$] undergoes a conformational transition to the $Na^+$-form upon binding with sanguinarine in presence of $K^+$. *Biochemical and Biophysical Research Communications* **404**, 139-142.

PROTOZANOVA E., MACGREGOR R. B. (1996) Frayed wires: A thermally stable form of DNA with two distinct structural domains. *Biochemistry* **35**, 16638-16645.

QIN Y., HURLEY L. H. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions.

*Biochimie* **90**, 1149-1171.

RACHWAL P. A., FOX K. R. (2007) Quadruplex melting. *Methods* **43**, 291-301.

RANGAN A., FEDOROFF O. Y., HURLEY L. H. (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *Journal of Biological Chemistry* **276**, 4640-4646.

READ M. A., NEIDLE S. (2000) Structural characterization of a guanine-quadruplex ligand complex. *Biochemistry* **39**, 13422-13432.

REDON S., BOMBARD S., ELIZONDO-RIOJAS M. A., CHOTTARD J. C. (2003) Platinum cross-linking of adenines and guanines on the quadruplex structures of the $AG_3(T_2AG_3)_3$ and $(T_2AG_3)_4$ human telomere sequences in $Na^+$ and $K^+$ solutions. *Nucleic Acids Research* **31**, 1605-1613.

REN J. S., QU X. G., TRENT J. O., CHAIRES J. B. (2002) Tiny telomere DNA. *Nucleic Acids Research* **30**, 2307-2315.

RENČIUK D., KEJNOVSKÁ I., ŠKOLÁKOVÁ P., BEDNÁŘOVÁ K., MOTLOVÁ J., VORLÍČKOVÁ M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant $K^+$ solutions. *Nucleic Acids Research* **37**, 6625–6634.

RENČIUK D., ZHOU J., BEAUREPAIRE L., GUÉDIN A., BOURDONCLE A., MERGNY J. L. (2012) A FRET-based screening assai for nucleic acid ligands. *Methods* **57**, 122-128.

RESHETNIKOV R. V., SPONER J., RASSOKHINA O. I., KOPYLOV A. M., TSVETKOV P. O., MAKAROV A. A., GOLOVIN A. V. (2011) Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process. *Nucleic Acids Research* **39**, 9789-9802.

REZLER E. M., SEENISAMY J., BASHYAM S., KIM M. Y., WHITE E., WILSON W. D., HURLEY L. H. (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *Journal of the American Chemical Society* **127**, 9439-9447.

RHODES D., FAIRALL L., SIMONSSON T., COURT R., CHAPMAN L. (2002) Telomere architecture. *EMBO Reports* **3**, 1139-1145.

RICHARDS R. I. (2001) Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Human Molecular Genetics* **10**, 2187-2194.

RISITANO A., FOX K. R. (2003) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry* **42**, 6507-6513.

RODRIGUEZ R., PANTOS G. D., GONCALVES D. P. N., SANDERS J. K. M.,

BALASUBRAMANIAN S. (2007) Ligand-driven G-quadruplex conformational switching by using an unusual mode of interaction. *Angewandte Chemie – International Edition* **46**, 5405-5407.

ROSU F., GABELICA V., PONCELET H., DE PAUW E. (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry. *Nucleic Acids research* **38**, 5217-5225.

ROSU F., PAUW E. D., GABELICA V. (2008) Electrospray mass spectrometry to study drug-nucleic acids interactions. *Biochimie* **90**, 1074-1087.

ROSU F., PIROTTE S., DE PAUW E., GABELICA V. (2006) Positive and negative ion mode ESI-MS and MS/MS for studying drug–DNA complexes. *International Journal of Mass Spectrometry* **253**, 156-171.

RUJAN I. N., MELENEY J. C., BOLTON P. H. (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids research* **33**, 2022-2031.

SACCA B., LACROIX L., MERGNY J. L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Research* **33**, 1182-1192.

SAENGER W.: Principles of Nucleic Acid Structure, In Cantor, C. R. (ed.), Springer advanced texts in chemistry, Springer-Verlag. New York 1983, ISBN 3-540-90761-0.

SALAS T. R., PETRUSEVA I., LAVRIK O., BOURDONCLE A., MERGNY J. L., FAVRE A., SAINTOME C. (2006) Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Research* **34**, 4857-4865.

SALAZAR M., THOMPSON B. D., KERWIN S. M., HURLEY L. H. (1996) Thermally induced DNA.RNA hybrid to G-quadruplex transitions: possible implications for telomere synthesis by telomerase. *Biochemistry* **35**, 16110-16115.

SCHAFFITZEL C., BERGER I., POSTBERG J., HANES J., LIPPS H. J., PLUCKTHUN A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 8572–8577.

SCHNEIDER J. H., ODO J., NAKAMOTO K. (1988) Interactions of water-soluble metalloporphyrins with nucleic-acids studied by resonance Raman spectroscopy. *Nucleic Acids Research* **16**, 10323-10338.

SCHULTZE P., MACAYA R. F., FEIGON J. (1994) Three-dimensional solution structure

of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *Journal of Molecular Biology* **235**, 1532-1547.

SEENISAMY J., BASHYAM S., GOKHALE V., VANKAYALAPATI H., SUN D., SIDDIQUI-JAIN A., STREINER N., SHIN-YA K., WHITE E., WILSON D., HURLEY L. H. (2005) Design and synthesis of an expanded porphyrin that has selectivity for the c-myc G-quadruplex structure. *Journal of the American Chemical Society* **127**, 2944-2959.

SEIMIYA H., OH-HARA T., SUZUKI T., NAASANI I., SHIMAZAKI T., TSUCHIYA K., TSURUO T. (2002) Telomere shortening and growth inhibition of human cancer cells by novel synthetic telomerase inhibitors MST-312, MST-295, and MST-199. *Molecular Cancer Therapeutics* **1**, 657-665.

SETNIČKA V., NOVÝ J., BŐHM S., SREENIVASACHARY N., URBANOVÁ M., VOLKA K. (2008) Molecular structure of guanine-quartet supramolecular assemblies in a gel-state based on a DFT calculation of infrared and vibrational circular dichroism spectra. *Langmuir* **24**, 7520-7527.

SHAFER R. H., SMIRNOV I. (2001) Biological aspects of DNA/RNA quadruplexes. *Biopolymers* **56**, 209-227.

SHAY J. W., BACCHETTI S. (1997) A survey of telomerase activity in human cancer. *European Journal of Cancer* **33**, 787-791.

SHAY J. W., WRIGHT W. E. (2002) Telomerase: A target for cancer therapeutics. *Cancer Cell* **2**, 257-265.

SHI D. F., WHEELHOUSE R. T., SUN D., HURLEY L. H. (2001) Quadruplex-interactive agents as telomerase inhibitors: Synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *Journal of Medicinal Chemistry* **44**, 4509-4523.

SHIN T. W., KIM K., LEE I. J. (1997) Spectrophotometric determination of the acid dissociation constants for cacodylic acid and p-nitrophenol at elevated temperatures. *Journal of Solution Chemistry* **26**, 379-390.

SHIRUDE P. S., BALASUBRAMANIAN S. (2008) Single molecule conformational analysis of DNA G-quadruplexes. *Biochimie* **90**, 1197-1206.

SIMONSSON T. (2001) G-Quadruplex DNA structures-variations on a theme. *Biological Chemistry* **382**, 621-628.

SINDEN R. R., POTAMAN V. N., OUSSATCHEVA E. A., PEARSON C. E., LYUBCHENKO Y. L., SHLYAKHTENKO L. S. (2002) Triplet repeat DNA structures

and human genetic disease: dynamic mutations from dynamic DNA. *Journal of Biosciences* **27**, 53-65.

SINGH V., AZARKH M., EXNER T. E., HARTIG J. S., DRESCHER M. (2009) Human telomeric quadruplex conformations studied by pulsed EPR. *Angewandte Chemie - International Edition* **48**, 9728-9730.

SMIRNOV I., SHAFER R. H. (2000a) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry* **39**, 1462-1468.

SMIRNOV I., SHAFER R. H. (2000b) Lead is unusually effective in sequence-specific folding of DNA. *Journal of Molecular Biology* **296**, 1-5.

SPARAPANI S., HAIDER S. M., DORIA F., GUNARATNAM M., NEIDLE S. (2010) Rational design of acridine-based ligands with selectivity for human telomeric quadruplexes. *Journal of the American Chemical Society* **132**, 12263-12272.

SPENCE J. D., LASH T. D. (2000) Porphyrins with exocyclic rings. 14.[1] Synthesis of tetraacenaphthoporphyrins, a new family of highly conjugated porphyrins with record-breaking long-wavelength electronic absorptions. *Journal of Organic Chemistry* **65**, 1530-1539.

SPINDLER L., RIGLER M., DREVENŠEK-OLENIK I., HESSARI N. M., DA SILVA M. W. (2010) Effect of base sequence on G-wire formation in solution. *Journal of Nucleic Acids* **2010**, 8 pages.

STANSEL T., DE LANGE T., GRIFFITH J. D. (2001) T-loop assembly in vitro involves binding of TRF2 near the 3‘ telomeric overhang. *EMBO Journal* **20**, 5532-5540.

STEJSKAL E. O., TANNER J. E. (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics* **42**, 288-292.

SUN D., GUO K., RUSCHE J. J., HURLEY L. H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Research* **33**, 6070-6080.

SUN H. X., TANG Y. L., XIANG J. F., XU G. Z., ZHANG Y. Z., ZHANG H., XU L. H. (2006) Spectroscopic studies of the interaction between quercetin and G-quadruplex DNA. *Bioorganic & Medicinal Chemistry Letters* **16**, 3586-3589.

SUNDQUIST W. I., HEAPHY S. (1993) Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proceedings of the National Academy of Sciences of the United States of America*

**90**, 3393-3397.

SUZUKI J., MIYANO-KUROSAKI N., KUWASAKI T., TAKEUCHI H., KAWAI G., TAKAKU H. (2002) Inhibition of human immunodeficiency virus type 1 activity in vitro by a new self-stabilized oligonucleotide with guanosine-thymidine quadruplex motifs. *Journal of Virology* **76**, 3015-3022.

ŠPONER J., CANG X., CHEATHAM T. E. (2012) Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods* **57**, 25-39.

ŠTEFL R., CHEATHAM III T. E., ŠPAČKOVÁ N., FADRNÁ E., BERGER I. (2003) Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of substates. *Biophysical Journal* **85**, 1787-1804.

TAKAHAMA K., SUGIMOTO C., ARAI S., KUROKAWA R., OYOSHI T. (2011) Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's Sarcoma. *Biochemistry* **50**, 5369-5378.

TAMURA Y., YOSHIDA M., SUZUKI J., HIRATOU T., MIYANO-KUROSAKI N., TAKAI K., TAKAKU H. (2000) Properties of quadruplex oligonucleotides with anti-HIV-1 activity. *Nucleic Acids Symposium Series* **44**, 181-182.

TANG C. F., SHAFER R. H. (2006) Engineering the quadruplex fold: Nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *Journal of the American Chemical Society* **128**, 5966-5973.

TARRAGO-LITVAK L., ANDREOLA M. L., FOURNIER M., NEVINSKY G. A., PARISSI V., DE SOULTRAIT V. R., LITVAK S. (2002) Inhibitors of HIV-1 reverse transcriptase and integrase: Classical and emerging therapeutical approaches. *Current Pharmaceutical Design* **8**, 595-614.

THOMAS G. J., JR. (1999) Raman spectroscopy of protein and nucleic acid assemblies. *Annual Review of Biophysics and Biomolecular Structure* **28**, 1–27.

TODD A. K., JOHNSTON M., NEIDLE S. (2005) Highly prevalent putative quadruplex sequences motifs in the human DNA. *Nucleic Acids Research* **33**, 2901-2907.

TOMAŠKO M., VORLÍČKOVÁ M., SAGI J. (2008) Substitution of adenine for guanine in the quadruplex-forming human telomere DNA sequence G(3)(T(2)AG(3))(3). *Biochimie* **91**, 171-179.

TOTH G., GASPARI Z., JURKA J. (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* **10**, 967-981.

TRAJKOVSKI M., DA SILVA M. W., PLAVEC J. (2012) Unique structural features of interconverting monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myc gene. *Journal of the American Chemical Society* **134**, 4132-4141.

TRAJKOVSKI M., SKET P., PLAVEC J. (2009) Cation localization and movement within DNA thrombin binding aptamer in solution. *Organic & Biomolecular Chemistry* **7**, 4677-4684.

TRAVASCIO P., LI Y., SEN D. (1998) DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chemistry & Biology* **5**, 505-517.

UEYAMA H., TAKAGI M., TAKENAKA S. (2002) A novel potassium sensing in aqueous media with a synthetic oligonucleotide derivative. Fluorescence resonance energy transfer associated with guanine quartet-potassium ion complex formation. *Journal of the American Chemical Society* **124**, 14286-14287.

UNO T., AOKI K., SHIKIMI T., HIRANUMA Y., TOMISUGI Y., ISHIKAWA Y. (2002) Coppers insertion facilitates water-soluble porphyrin binding to rA•rU and rA•dT base pairs in duplex RNA and RNA•DNA hybrids. *Biochemistry* **41**, 13059-13066.

URATA H., KUMASHIRO T., KAWAHATA T., OTAKE T. (2004) Anti-HIV-1 activity and mode of action of mirror image oligodeoxynucleotide analogue of zintevir. *Biochemical and Biophysical Research Communications* **313**, 55-61.

URBANOVÁ M. (2009) Bioinspired interactions studied by vibrational circular dichroism. *Chirality* **21**, E215-E230.

USDIN K. (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Research* **26**, 4078-4085.

VACHOUŠEK J., ŠTĚPÁNEK J. (2008) Raman study of HIV TAR structural stability. *Spectroscopy: An International Journal* **22**, 267-277.

VAIRAMANI M., GROSS M. L. (2003) G-Quadruplex formation of thrombin-binding aptamer detected by electrospray ionization mass spectrometry. *Journal of the American Chemical Society* **125**, 42–43.

VENCZEL E. A., SEN D. (1993) Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry* **32**, 6220-6228.

VESENKA J., BAGG D., WOLFF A., REICHERT A., MOELLER R., FRITZSCHE W. (2007) Auto-orientation of G-wire DNA on mica. *Colloids and Surfaces B: Biointerfaces* **58**, 256-263.

VIALAS C., PRATVIEL G., MEUNIER B. (2000) Oxidative damage generated by an oxo-metalloporphyrin onto the human telomeric sequence. *Biochemistry* **39**, 9514-9522.

VÍGLASKÝ V., BAUER L., TLUČKOVÁ K. (2010) Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry* **49**, 2110-2120.

VO T. U., MCGOWN L. B. (2004) Selectivity of quadruplex DNA stationary phases toward amino acids in homodipeptides and alanyl dipeptides. *Electrophoresis* **25**, 1230-1236.

VONDRUŠKOVÁ J., KYPR J., KEJNOVSKÁ I., FIALOVÁ M., VORLÍČKOVÁ M. (2008) Guanine quadruplex formation by RNA/DNA hybrid analogs of Oxytricha telomere G(4)T(4)G(4) fragment. *Biopolymers* **89**, 797-806.

VORLÍČKOVÁ M., CHLÁDKOVÁ J., KEJNOVSKÁ I., FIALOVÁ M., KYPR J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Research* **33**, 5851-5860.

VORLÍČKOVÁ M., KEJNOVSKÁ I., BEDNÁŘOVÁ K., RENČIUK D., KYPR J. (2012) Circular dichroism spectroscopy of DNA: From duplexes to quadruplexes. *Chirality* **24**, 691-698.

VORLÍČKOVÁ M., KEJNOVSKÁ I., SAGI J., RENČIUK D., BEDNÁŘOVÁ K., MOTLOVÁ J., KYPR J. (2012) Circular dichroism and guanine quadruplexes. *Methods* **57**, 64-75.

WALLACE T. L., GAMBA-VITALO C., LOVEDAY K. S., COSSUM P. A. (2000) Acute, multiple-dose, and genetic toxicology of AR177, an anti-HIV oligonucleotide. *Toxicological Sciences* **53**, *63-70.*

WANG A. H. J., GESSNER R. V., VANDERMAREL G. A., VANBOOM J. H., RICH A. (1985) Crystal-structure of Z-DNA without an alternating purine-pyrimidine sequence. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 3611-3615.

WANG K. Y., KRAWCZYK S. H., BISCHOFBERGER N., SWAMINATHAN S., BOLTON P. H. (1993a) The tertiary structure of a DNA aptamer which binds to and inhibits thrombin determines activity. *Biochemistry* **32**, 11285–11292.

WANG K. Y., MCCURDY S., SHEA R. G., SWAMINATHAN S., BOLTON P. H. (1993b) A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry* **32**, 1899-1904.

WANG Y., ZHANG Y., ONG N. P. (2005) Speeding up a single-molecule DNA device

with a simple catalyst. *Physical Review E* **72**.

WEI C. Y., JIA G. Q., YUAN J. L., FENG Z. C., LI C. (2006) A spectroscopic study on the interactions of porphyrin with G-quadruplex DNAs. *Biochemistry* **45**, 6681-6691.

WEISMAN-SHOMER P., COHEN E., FRY M. (2000) Interruption of the fragile X syndrome expanded sequence d(CGG)(n) by interspersed d(AGG) trinucleotides diminishes the formation and stability of d(CGG)(n) tetrahelical structures. *Nucleic Acids Research* **28**, 1535-1541.

WEISMAN-SHOMER P., COHEN E., HERSHCO I., KHATEB S., WOLFOVITZ-BARCHAD O., HURLEY L. H., FRY M. (2003) The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence $d(CGG)_n$. *Nucleic Acids Research* **31**, 3963-3970.

WEITZMANN M. N., WOODFORD K. J., USDIN K. (1996) The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *Journal of Biological Chemistry* **271**, 20958-20964.

WELLINGER R. J., SEN D. (1997) The DNA structures at the ends of eukaryotic chromosomes. *European Journal of Cancer* **33**, 735-749.

WELLS R. D., DERE R., HEBERT M. L., NAPIERALA M., SON L. S. (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Research* **33**, 3785-3798.

WHITE L. K., WRIGHT W. E., SHAY J. W. (2001) Telomerase inhibitors. Trends in *Biotechnology* **19**, 114-120.

WLODARCZYK A., GRZYBOWSKI P., PATKOWSKI A., DOBEK A. (2005) Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. *Journal of Physical Chemistry B* **109**, 3594-3605.

WU Z. S., CHEN C. R., SHEN G. L., YU R. Q. (2008) Reversible electronic nanoswitch based on DNA G-quadruplex conformation: A platform for single-step, reagentless potassium detection. *Biomaterials* **29**, 2689-2696.

WYATT J. R., DAVIS P. W., FREIER S. M. (1996) Kinetics of G-quartet-mediated tetramer formation. *Biochemistry* **35**, 8002-8008.

WYATT J. R., VICKERS T. A., ROBERSON J. L., BUCKHEIT R. W., KLIMKAIT T., DEBAETS E., DAVIS P. W., RAYNER B., IMBACH J. L., ECKER D. J. (1994)

Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 1356-1360.

WYTTENBACH T., BOWERS M. T. (2007) Intermolecular interactions in biomolecular systems examined by mass spectrometry. *Annual Review of Physical Chemistry* **58**, 511-533.

XU Y., SUGIYAMA H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Research* **34**, 949-954.

XU Y., YAMAZAKI S., OSUGA H., SUGIYAMA H. (2006). The recognition of higher-order G-quadruplex by chiral cyclic-helicene molecules. *Nucleic Acids Symposium Series* **50**, 183-184.

YAFE A., ETZIONI S., WEISMAN-SHOMER P., FRY M. (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Research* **33**, 2887-2900.

YAMASHITA T., UNO T., ISHIKAWA Y. (2005) Stabilization of guanine quadruplex DNA by the binding of porphyrins with cationic side arms. *Bioorganic & Medicinal Chemistry* **13**, 2423-2430.

YING L. M., GREEN J. J., LI H. T., KLENERMAN D., BALASUBRAMANIAN S. (2003) Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14629-14634.

YOSHIMURA S. H., MARUYAMA H., ISHIKAWA F., OHKI R., TAKEYASU K. (2004) Molecular mechanism of DNA end-loop formation by TRF2. *Genes to Cells* **9**, 205-218.

YU H. Q., MIYOSHI D., SUGIMOTO N. (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. *Journal of the American Chemical Society* **128**, 15461-15468.

YU H. J., WANG X. H., FU M. L., REN J. S., QU X. G. (2008) Chiral metallo-supramolecular complexes selectively recognize human telomeric G-quadruplex DNA. *Nucleic Acids Research* **36**, 5695-5703.

YUE D. J. E., LIM K. W., PHAN A. T. (2011) Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *Journal of the American Chemical Society* **133**, 11462-11465.

ZAUG A. J., PODELL E. R., CECH T. R. (2005) Human POT1 disrupts telomeric G-

quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10864-10869.

ZEMÁNEK M., KYPR J., VORLÍČKOVÁ M. (2005) Conformational properties of DNA containing $(CCA)_n$ and $(TGG)_n$ trinucleotide repeats. *International Journal of Biological Macromolecules* **36**, 23-32.

ZHANG N., GORIN A., MAJUMDAR A., KETTANI A., CHERNICHENKO N., SKRIPKIN E., PATEL D. J. (2001) V-shaped scaffold: a new architectural motif identified in an A·(G·G·G·G) pentad-containing dimeric DNA quadruplex involving stacked G(*anti*)·G(*anti*)·G(*anti*)·G(*syn*) tetrads. *Journal of Molecular Biology* **311**, 1063-1079.

ZHAO C. Q., GENG J., FENG L. Y., REN J. S., QU X. G. (2011) Chiral metallo-supramolecular complexes selectively induce human telomeric G-quadruplex formation under salt-deficient conditions. *Chemistry – A European Journal* **17**, 8209-8215.

ZHAO Y., KAN Z. Y., ZENG Z. X., HAO Y. H., CHEN H., TAN Z. (2004) Determining the folding and unfolding rate constants of nucleic by biosenzor. Application to telomere G-quadruplex. *Journal of the American Chemical Society* **126**, 13255-13264.

ZHU X., KUMAR R., MANDAL M., SHARMA N., SHARMA H. W., DHINGRA U., SOKOLOSKI J. A., HSIAO R., NARAYANAN R. (1996) Cell cycle-dependent modulation of telomerase activity in tumor cells. *Proceedings of the National Academy of the United States of America* **93**, 6091-6095.

# List of tables

## List of abbreviations

CD – circular dichroism

DSC – differential scanning calorimetry

ITC – isothermal titration calorimetry

KRAS – a regulatory element of the human KRAS promoter (G-rich DNA sequence: $d(AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2)$)

PAGE – polyacrylamide gel electrophoresis

SERRS – surface enhanced resonance Raman spectroscopy

SERS – surface enhanced Raman spectroscopy

SVD – singular value decomposition

TBA – thrombin-binding-aptamer (G-rich DNA sequence: $d(G_2T_2G_2TGTG_2T_2G_2)$)

UV – ultraviolet

UVRR – Ultraviolet resonance Raman scattering

# Attachments:

# Publication in impacted journals

# Polymorphism of Human Telomeric Quadruplex Structure Controlled by DNA Concentration: A Raman Study

Nucleic Acids Research, accepted (2012)

Jan Palacký[1], Michaela Vorlíčková[2,3*], Iva Kejnovská[2,3], Peter Mojzeš[1*]

[1]Charles University in Prague, Faculty of Mathematics and Physics, Institute of Physics, Ke Karlovu 5, CZ-121 16 Prague 2, Czech Republic

[2]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, CZ-612 65 Brno, Czech Republic

[3]CEITEC – Central European Institute of Technology, Masaryk University, Kamenice 5, CZ-625 00 Brno, Czech Republic

Corresponding authors:
Peter Mojzeš, *mojzes@karlov.mff.cuni.cz*,
Michaela Vorlíčková, *mifi@ibp.cz*

Nucleic Acids Research Advance Access published November 27, 2012

*Nucleic Acids Research, 2012, 1–12*
*doi:10.1093/nar/gks1135*

# Polymorphism of human telomeric quadruplex structure controlled by DNA concentration: a Raman study

**Jan Palacký[1], Michaela Vorlíčková[2,3,*], Iva Kejnovská[2,3] and Peter Mojzeš[1,*]**

[1]Charles University in Prague, Faculty of Mathematics and Physics, Institute of Physics, Ke Karlovu 5, CZ-121 16 Prague 2, [2]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, CZ-612 65 Brno and [3]CEITEC—Central European Institute of Technology, Masaryk University, Kamenice 5, CZ-625 00 Brno, Czech Republic

Received September 27, 2012; Accepted October 22, 2012

## ABSTRACT

**DNA concentration has been recently suggested to be the reason why different arrangements are revealed for $K^+$-stabilized human telomere quadruplexes by experimental methods requiring DNA concentrations differing by orders of magnitude. As Raman spectroscopy can be applied to DNA samples ranging from those accessible by absorption and CD spectroscopies up to extremely concentrated solutions, gels and even crystals; it has been used here to clarify polymorphism of a core human telomeric sequence $G_3(TTAG_3)_3$ in the presence of $K^+$ and $Na^+$ ions throughout wide range of DNA concentrations. We demonstrate that the $K^+$-structure of $G_3(TTAG_3)_3$ at low DNA concentration is close to the antiparallel fold of $Na^+$-stabilized quadruplex. On the increase of $G_3(TTAG_3)_3$ concentration, a gradual transition from antiparallel to intramolecular parallel arrangement was observed, but only for thermodynamically equilibrated $K^+$-stabilized samples. The transition is synergically supported by increased $K^+$ concentration. However, even for extremely high $G_3(TTAG_3)_3$ and $K^+$ concentrations, an intramolecular antiparallel quadruplex is spontaneously formed from desalted non-quadruplex single-strand after addition of $K^+$ ions. Thermal destabilization or long dwell time are necessary to induce interquadruplex transition. On the contrary, $Na^+$-stabilized $G_3(TTAG_3)_3$ retains its antiparallel folding regardless of the extremely high DNA and/or $Na^+$ concentrations, thermal destabilization or annealing.**

## INTRODUCTION

Linear eukaryotic chromosomes are protected at both ends by telomeres (1). Telomeric DNA in humans consists of the $(TTAGGG)_n/(CCCTAA)_n$ repeats terminating on the 3′-end of the molecule in a single-stranded G-rich overhang (2), which folds into a G-quadruplex under suitable conditions (3). Telomeres play a key role in maintaining chromosomal integrity, control of DNA replication and protection against chromosome elongation by telomerase (4). Unregulated telomerase activity allows cancer cells to become immortal (5). As formation of G-quadruplexes at telomeres perturbs telomerase function (6), the quadruplex-stabilizing agents can suppress the proliferation of tumour cells (7,8). Se arching for these agents is of great importance for cancer treatment in human medicine (9). For a better efficiency, structural properties of telomeric quadruplexes under the physiologically relevant conditions, that is, at moderate concentration of $K^+$ ions and high local concentration of DNA, must be known properly.

Despite numerous studies, the structure adopted by human telomeric repeats under the biologically relevant conditions remains unclear. The first relevant structure solved at atomic resolution was reported by Wang and Patel in 1993 (10). Based on nuclear magnetic resonance (NMR) study and molecular dynamics simulations, the telomeric fragment $AG_3(TTAG_3)_3$ was shown to adopt an antiparallel basket-type structure in the presence of $Na^+$ (10). A year later, the same topology was evidenced by CD spectroscopy and chemical probing for quadruplexes formed by related telomeric sequences, $G_3(TTAG_3)_3$ and $(TTAG_3)_4$, in both $Na^+$ and $K^+$ solutions (11). On the contrary, radically different intramolecular parallel quadruplex with three propeller-shaped,

*To whom correspondence should be addressed. Tel: +420 541 517 188; Fax: +420 541 211 293; Email: mifi@ibp.cz
Correspondence may also be addressed to Peter Mojzeš. Tel: +420 221 911 471; Fax: +420 224 922 797; Email: mojzes@karlov.mff.cuni.cz

double-chain-reversal d(TTA)-loops was reported for the crystalline $K^+$-form of $AG_3(TTAG_3)_3$ (12). Platinum cross-linking studies revealed the formation of the basket-type antiparallel quadruplex of $AG_3(TTAG_3)_3$ in both $Na^+$ and $K^+$ solution (13), whereas $^{125}$I-radioprobing confirmed the basket-type of $AG_3(TTAG_3)_3$ in $Na^+$ solutions, and the antiparallel chair-type in the presence of $K^+$ (14).

Several studies also proposed that parallel and antiparallel quadruplexes could coexist in $K^+$ solutions, for example (15). On the other hand, parallel-strand alignment derived from crystallography (12) was excluded as the major biologically relevant conformation of the human telomere quadruplex in $K^+$ solution (16). In 2005, another type of quadruplex arrangement, so called '(3 + 1) structure' containing three parallel and one antiparallel strands, was suggested for $K^+$-stabilized $G_3(TTAG_3)_4$ containing one redundant $TTAG_3$ repeat, to explain peculiarities of its CD spectra (17). A year later, structure of the intramolecular (3 + 1) quadruplex scaffold was solved by NMR in $K^+$ solutions for $TTG_3(TTAG_3)_3A$, $TAG_3(TTAG_3)_3$ (18) and $A_3G_3(TTAG_3)_3AA$ (19); however, attempts to elucidate the $K^+$-structures of some related sequences bearing no or few residues attached to the minimal quadruplex-forming sequence $G_3(TTAG_3)_3$ failed because of their conformational heterogeneity. Another variant of the (3 + 1) topology with different order of the loop arrangements was discerned by NMR for $TAG_3(TTAG_3)_3TT$ stabilized by $K^+$ ions (20). Furthermore, depending on the nature and number of the extra residues flanking the minimal four-repeat $G_3(TTAG_3)_3$ unit, and the method applied, various quadruplex arrangements under the physiologically relevant $K^+$ conditions were reported, antiparallel (17,21), mixture of hybrid (3 + 1) and chair-type antiparallel (22), mixture of basket- and chair-type antiparallel (23) and others (24–26). There was even reported CD evidence of parallel arrangement, although in aqueous solution at high $K^+$ concentration (27). Parallel topology was later observed in the presence of organic solvents simulating crowding conditions present in cells, that is, polyethylene glycol (28,29) or ethanol (30). Recently, a new quadruplex type has been observed by NMR for $G_3(TTAG_3)_3T$ stabilized by $K^+$ ions (31). This stable structure ascribed to a basket-type quadruplex contains, however, only two G-tetrad layers capped with several layers of stacked guanines and adenines on both sides of the tetrad core.

From a great deal of experimental data available to date, it follows that surprisingly high polymorphism of the human telomere quadruplex could be a rather complex resultant of the exact sequence studied, DNA concentration, stabilizing cation type and concentration, dehydrating or molecular crowding conditions, as well as different protocols of sample preparation. Notably, DNA concentration differing by orders of magnitude for various experimental methods can be responsible for contradictory structures observed under otherwise comparable conditions. For instance, recent CD study has demonstrated that $K^+$-stabilized quadruplexes of several human telomeric fragments undergo structural transition on the increase of oligonucleotide concentration, even at the physiologically relevant $K^+$ concentrations (30). However, more detailed CD and absorption studies at high DNA concentrations are hindered by high optical densities of the samples.

In the present work, Raman spectroscopy is reintroduced as a practical and well-established, albeit yet underappreciated, method suitable for structural studies of G-quadruplexes. Although conventional, notably non-resonant Raman spectroscopy has proved useful for investigation of nucleic acids conformations [for overview see (32,33)], relatively few articles relate specifically to G-quadruplexes (e.g. 34–41). One of advantages of Raman spectroscopy consists in its applicability to DNA samples starting from the concentrations comfortably accessible by absorption and CD spectroscopies up to highly concentrated solutions inducing molecular crowding without need of cosolutes, gels and crystals studied by X-ray diffraction, with the possibility to acquire informational rich Raman melting profiles at high quadruplex concentrations, which are still missing. Raman spectroscopy can thus bridge, in a methodologically consistent way, an information gap between experimental methods restricted to low and high DNA concentrations solely. It can contribute to reconciliation of different $K^+$-quadruplex structures reported for aqueous solutions and clarify arrangement of human telomeric sequences under the physiologically relevant conditions.

## MATERIALS AND METHODS

### Samples

High-performance liquid chromatography-purified and lyophilized/highly concentrated oligonucleotides were purchased from VBC-Biotech (Vienna, Austria) and Generi Biotech (Hradec Králové, Czech Republic) in several consecutive batches as ammonium or methylammonium salts. Chemicals of analytical grade (Sigma-Aldrich) and deionized water (18 MΩ, Elga) were used for buffers. To adjust precisely type and concentration of the quadruplex-stabilizing cations in the samples differing in DNA concentration, the following procedure was used. Stock solutions of oligonucleotides (~200–500 mM in nucleosides) were prepared by dissolving lyophilized oligomer into deionized water. To replace/remove former countercations and small molecules initially present in the lyophilisate, appropriate amounts of the stock solution were repeatedly (typically 3–5 times) washed out by a phosphate buffer (PBS, 30–150 mM according to DNA concentration, pH 6.8) containing $Na^+$ or $K^+$ ions at desired concentrations (100–500 mM), using centrifugal filter devices (Amicon Ultra 3K, Millipore). Final concentration of $K^+$ or $Na^+$ in PBS was adjusted by addition of KCl or NaCl. The same centrifugal washing but by deionized water was used for preparation of the samples without alkali cations. Non-quadruplex arrangement of the fully desalted samples and the quadruplex formation after addition of alkali cations was checked out by ultraviolet (UV) absorption (42). The DNA concentrations were adjusted by controlling

the starting and final centrifuged volumes, as well as by subsequent dilution by an appropriate solvent. In addition to total and controllable solvent exchange and precise adjustment of the DNA concentration, repeated centrifugal filtration helped to remove trace contaminants causing fluorescence background in Raman spectra. The exact oligonucleotides concentration was determined by absorbance measurements of appropriately diluted samples at 95°C by using a Lambda 12 UV/VIS (Perkin-Elmer) or Unicam 5626 UV/VIS spectrophotometers and molar extinction coefficients calculated according to (43). To increase accuracy of the concentration determination of extremely concentrated DNA samples, especially important for DNA concentration series, relative concentrations were determined concurrently from Raman intensities normalized with respect to OH-stretching band of water molecules (44). If not stated explicitly otherwise, DNA concentrations are related to nucleosides throughout the present article. Quadruplex structures thermodynamically equilibrated in relevant solutions were prepared by a 15 min thermal denaturation at 95°C and subsequent slow annealing (~5 h) to room temperature. Heated but thermodynamically non-equilibrated structures were prepared by a fast cooling (from 95 to 20°C within <1 min), using expanding stream of compressed dry air. Moreover, structural changes in the course of thermal denaturation and annealing were monitored by Raman spectroscopy as a function of actual temperature in a temperature-controlled quartz microcell.

### Raman measurements

Raman spectra were excited with the 532 nm line of a continuous-wave solid-state Nd:$YVO_4$ laser (Verdi 2, Coherent) using the radiation power at the sample ranging from 100–400 mW (according to the sample concentration). The spectra were collected in the 90° scattering geometry on a multi-channel Raman spectrograph (Jobin Yvon–Spex 270 M) equipped with a holographic notch-plus filter (Kaiser) to reject Rayleigh scattering and a liquid nitrogen-cooled CCD detector (Princeton Instruments). Raman measurements were carried out in a temperature-controlled, hermetically sealed quartz microcell (5 µl sample volume) at somewhat reduced temperature [5–10°C, in comparison with room temperature used for CD and polyacrylamide gel electrophoresis (PAGE) experiments] to eliminate potential sources of artifacts specific for Raman scattering; however, such a slight temperature difference was proven to have no apparent effect on quadruplex structures and structural conclusions. Thermal denaturation/annealing experiments were conducted in the range of 2–96°C, with an accuracy of a temperature control within ±0.5°C. If not stated otherwise, the sample was equilibrated at the desired temperature for ~10 min before recording the spectrum. The wavenumber scales of Raman spectra were precisely calibrated using the emission spectra of a neon glow lamp taken before and after each Raman measurement. The estimated precision of the calibration procedure was better than 0.1 $cm^{-1}$. The Raman contribution from the solvent was carefully subtracted, and the spectra were corrected for their non-Raman background by using advanced methods of factor analysis (44). Spectra were normalized to the peak height of the 1093 $cm^{-1}$ band associated with the $PO_2^-$ symmetric stretching mode, reported previously to be largely invariant to the melting of DNA duplexes (45). If appropriate, for example, for quantitative comparisons of the oligonucleotides differing in the length or in the number of specific nucleosides, normalized spectra were furthermore renormalized according to the parameter in question.

### CD measurements

CD measurements were done in a Jobin-Yvon CD6 dichrograph (Longjumeau, France) in 1–0.001 cm path-length quartz cells (Hellma Analytics, Germany) placed in a temperature-controlled holder. The scan rate was 0.5 nm/s. CD signals are expressed as the difference in the molar absorption $\Delta\varepsilon$ of the right- and left-handed circularly polarized light. The molarities are related to nucleosides.

## PAGE

PAGE was performed in a temperature-controlled electrophoretic apparatus (SE-600; Hoefer Scientific, USA). Gel concentration was 16% (29:1 monomer to bis ratio, Applichem, Germany). About 2 µg of DNA was loaded into each lane of a 14 × 16 × 0.1 cm gel. Samples were electrophoresed at 20°C for 19 h at 30 V (~2 $Vcm^{-1}$). Gels were stained with Stains All (Sigma-Aldrich) after electrophoresis and scanned using a Personal Densitometer SI, model 375-A (Molecular Dynamics, USA).

## RESULTS AND DISCUSSION

### Equilibrated transition of annealed $K^+$-stabilized $G_3(TTAG_3)_3$ depends on DNA concentration

In the recent CD study (30), it was suggested that the antiparallel-to-parallel rearrangement of human telomeric sequences can be accomplished at physiologically relevant $K^+$ concentrations simply by increasing the concentration of DNA, without addition of organic cosolutes to simulate molecular crowding conditions. To provide methodologically independent evidence of the self-crowding effect throughout a wider range of DNA concentration, Raman spectra of thermodynamically equilibrated (slowly annealed) solutions of $G_3(TTAG_3)_3$ at DNA concentrations ranging from 8 to 240 mM and stabilized by 200–250 mM $K^+$ were acquired and compared with those prepared in a similar way in $Na^+$ buffers. As evident from differential spectra shown in Figure 1, initial quadruplex arrangement adopted by 8 mM $G_3(TTAG_3)_3$ at moderate concentration of 200 mM $K^+$ undergoes gradual structural changes on increase of DNA concentration. No similar spectral changes were observed for $G_3(TTAG_3)_3$ in $Na^+$ solutions (Supplementary Figure S3, top).

Differential Raman features of $K^+$-$G_3(TTAG_3)_3$ depicted in Figure 1 are mostly consistent with those reported on structural transition of *Oxytricha* telomeric

sequence $(T_4G_4)_4$ induced by increased concentration of $Na^+$ and $K^+$, and shown to be indicative for inter-quadruplex switching from intramolecular antiparallel fold to four-stranded parallel structure (35). As, to the best of our knowledge, human telomeric sequences have not been studied yet by conventional non-resonant Raman spectroscopy, assignment of Raman bands of $G_3(TTAG_3)_3$ and structural interpretation applied hereinafter is based on Raman studies of interquadruplex switching of *Oxytricha* telomeric sequences controlled by the type, concentration and molar ratio of $Na^+$ and $K^+$ cations (34–39), on relevant Raman studies of various non-quadruplex DNAs (32,33,45–51), as well as on Raman spectra of model oligonucleotides $(dG)_{15}$, $(dT)_{15}$ and $(dA)_{15}$ (Supplementary Figures S1 and S2). Raman markers used in the present work are listed and thoroughly discussed in the (Supplementary Table S1).

According to Raman markers indicative for *C2-endo/syn* (671 and 1326 $cm^{-1}$) and *C2-endo/anti* (686 and 1338 $cm^{-1}$) conformers of dG (32,34–36,39), structural

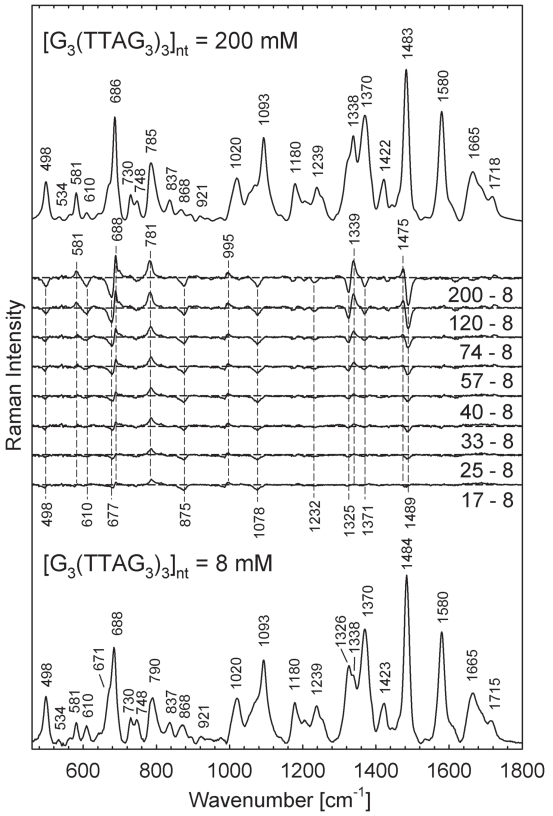



**Figure 1.** Raman spectra of $G_3(TTAG_3)_3$ in 200 mM $K^+$ (30 mM of PBS, pH 6.8, t = 5°C) at the nucleoside concentrations of 8 mM (bottom trace) and 200 mM (top trace). Intermediate traces show the differences between the spectra at indicated concentration and that of the lowest one.

transition of $K^+$-$G_3(TTAG_3)_3$ comprises gradual increase in the population of *C2-endo/anti*-dG at the expense of *C2-endo/syn*-dG (for detailed discussion see Supplementary Data), as can be documented by relative intensity increase of the Raman bands at 686 and 1338 $cm^{-1}$, accompanied by intensity decrease of the shoulder at 671 $cm^{-1}$ and the 1326 $cm^{-1}$ band (Figure 1 and Supplementary Figure S3, bottom). Furthermore, intensity decrease of relatively weak but important Raman marker of *C2-endo/syn*-dT at 611 ± 2 $cm^{-1}$ (32,36,39), sensitive to geometry of d(TTA)-loops (for detailed discussion see Supplementary Data), is consistent with quadruplex rearrangement from the fold incorporating side/diagonal loops to the quadruplex structure having loops organized in other geometry.

Characteristic changes in other Raman bands sensitive to *syn/anti*-dG ratio (~581 ± 1 $cm^{-1}$), structurally informative markers of hydrogen bonding within G-tetrads (~1483 ± 2 $cm^{-1}$), arrangement markers of d(TTA)-loops (~499 ± 2 and ~1371 ± 1 $cm^{-1}$) and markers sensitive to sugar-phosphate backbone (~788 ± 2 $cm^{-1}$) constitute differential pattern that can serve as Raman fingerprint of interquadruplex transition. Conformational switching concluded from Raman spectra corroborates structural transition inferred from the corresponding CD spectra indicating gradual changes from predominantly antiparallel (6.2 mM) to largely hybrid (3 + 1) (55 mM) and finally to predominantly parallel folding (30) at DNA concentration of 154 mM still accessible to CD measurements (Figure 2).

### Linear character of interquadruplex transition indicates conformational heterogeneity

To better understand progression of the changes induced by DNA concentration and to depict different impact of the $K^+$ and $Na^+$ ions, Raman spectra of DNA concentration series carefully prepared according to identical protocols from the same stock solution have been analysed together by a factor analysis (44). Using a singular value decomposition (SVD) approach (44), spectral variability within the normalized Raman dataset under consideration (11 and 11 Raman spectra of $G_3(TTAG_3)_3$ at the nucleotide concentrations 8–240 mM, stabilized by slightly higher concentration of 250 mM $Na^+$ and $K^+$ to favour parallel form, respectively) was decomposed into a set of orthonormal abstract factors (subspectra $S_j$), their statistical weights (singular values $W_j$) and coefficients $V_{ij}$ representing relative contributions of $S_j$ to the $i$-th experimental spectrum from the dataset (Figure 3). Despite somewhat abstract character of the subspectra and less common mode of presentation, practically constant values of the coefficients $V_{i1}$–$V_{i3}$ for $Na^+$-samples throughout full DNA concentration range tested here evidently means that their normalized Raman spectra are virtually identical. On the other hand, surprisingly linear dependency of all three coefficients $V_{i1}$–$V_{i3}$ on DNA concentration for $K^+$-samples suggests gradual structural transition (instead of sigmoidal one expected for conformational transition between two or three distinct forms), probably not completed even

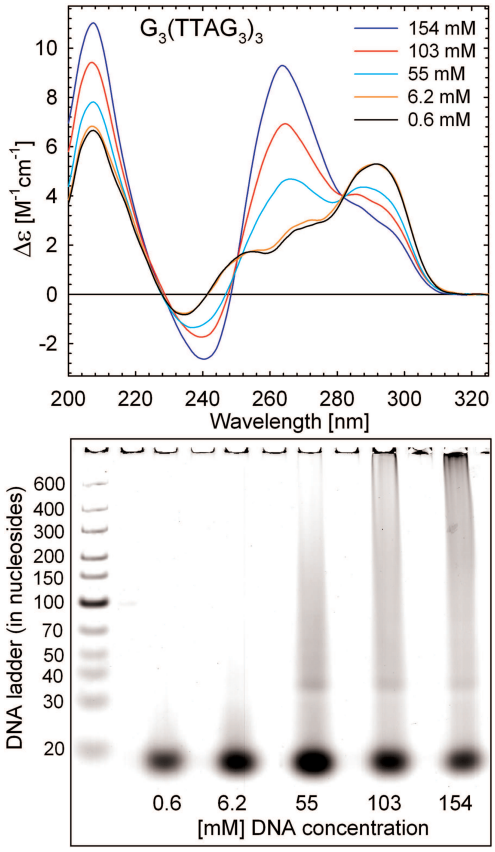



**Figure 2.** Dependence of CD spectra of $G_3(TTAG_3)_3$ on DNA concentration at 23°C. Each sample was slowly annealed after dilution to the final DNA concentration. Concentration of $K^+$ in all samples was 300 mM. Bottom: native PAGE of $K^+$-stabilized $G_3(TTAG_3)_3$ at indicated DNA concentrations. The PAGE was performed at 20°C in 30 mM of PBS containing 240 mM of $K^+$, pH 6.7.

at the highest DNA concentration of 240 mM studied in the series. Linear progression of the spectral changes in the range of DNA concentrations usual for NMR studies is consistent with conformational heterogeneity and coexistence of multiple quadruplex structures (18,19), the reason why attempts to elucidate $K^+$-structure of $G_3(TTAG_3)_3$ [or some related sequences bearing just few flanking nucleotides, e.g. $AG_3(TTAG_3)_3$] by NMR spectroscopy failed.

Furthermore, disclosure of three subspectra (according to singular values and residual errors shown in Figure 3) necessary for an adequate description of the spectral variability within full Raman dataset, corroborates persistence of the spectral differences between the $Na^+$- and $K^+$-structures beyond to the lowest DNA concentrations used in this experiment (~8 mM). Corresponding orthogonal differences extracted by the SVD from Raman dataset are expressed predominantly by a subspectrum $S_3$, as can be documented by substantial match with differential curve shown in Figure 4 (up), where the low-DNA spectra in both alkali cations are compared directly for better illustration. On the other hand, subspectrum $S_2$ can be related primarily to spectral differences between the high- and low-DNA structures of $K^+$-stabilized $G_3(TTAG_3)_3$, as evident from apparent similarity with high-DNA differential curve between the $K^+$ and $Na^+$ structures shown in Figure 4 (bottom).

Although the Raman spectra of $Na^+$- and $K^+$-stabilized $G_3(TTAG_3)_3$ at the lowest DNA concentration exhibit some spectral differences (Figure 4, top), these do not concern the crucial Raman markers of the *syn*/*anti*-dG conformers discussed earlier in the text. Only the minute flip of relative intensities between 674 and 687 $cm^{-1}$ bands of dG might be interpreted as a sign of slightly higher population of *C2-endo*/*anti*-dG in the $K^+$ form. However, the changes of another *syn*/*anti*-dG markers in the 1326/1338 $cm^{-1}$ region (i.e. appearance of 1332/1352 $cm^{-1}$ differential features) and opposite sign of difference at 1371 $cm^{-1}$ (Figure 4) are more complex and do not sustain changes in the *syn*/*anti* dG distribution or rearrangement of dT-loops expected for substantially different geometries. Furthermore, structurally sensitive band of symmetrical phosphodiester stretching mode at 789 $cm^{-1}$ exhibits similar intensity and shape in both $Na^+$- and $K^+$-forms at low DNA concentration (Figure 4, top), in contrast to substantial difference accompanying antiparallel-to-parallel transition at high-DNA content (Figure 4, bottom). Finally, a positive differential feature at 1488 $cm^{-1}$, the most distinct difference between Raman spectra of $K^+$- and $Na^+$-stabilized low-DNA folds of $G_3(TTAG_3)_3$ is inconsistent with antiparallel-to-parallel transition, for which a negative feature at the same position is expected (Figure 1). It can be concluded that our Raman study provides methodologically independent argument in favour of prevailing antiparallel topology of the $K^+$-$G_3(TTAG_3)_3$ at low DNA concentration, even though corresponding CD spectrum is dissimilar to that of $Na^+$-stabilized antiparallel structure (Figure 6). The CD spectrum of low-DNA $K^+$-$G_3(TTAG_3)_3$ has been previously interpreted as a sign of hybrid (3 + 1) form of $K^+$-$G_3(TTAG_3)_3$ (22,52), identifying its structure with the mixed (3 + 1) structures of $TTG_3(TTAG_3)_3A$, $TAG_3(TTAG_3)_3$ (18), $A_3G_3(TTAG_3)_3AA$ (19) and $TAG_3(TTAG_3)_3TT$ (20) in $K^+$ solutions resolved by NMR; however, at DNA concentrations by one or two orders higher that those studied by CD.

### Interquadruplex transition of $K^+$-$G_3(TTAG_3)_3$ is more effective at higher concentrations of $K^+$

Similar structural effect as induced by DNA concentration at a fixed concentration of $K^+$ can be seen on the increase of $K^+$ concentration at a fixed concentration of $G_3(TTAG_3)_3$. Representative Raman spectra of thermodynamically equilibrated (slowly annealed) solutions of 50 mM $G_3(TTAG_3)_3$, the concentration of the order common for NMR studies, in 100 and 500 mM of $K^+$ or $Na^+$ are shown in Figure 5. Corresponding CD spectra and native PAGE of the samples are presented in Figure 6.

According to CD, low-salt quadruplex structure of $G_3(TTAG_3)_3$ at moderate DNA concentration (Figure 6)

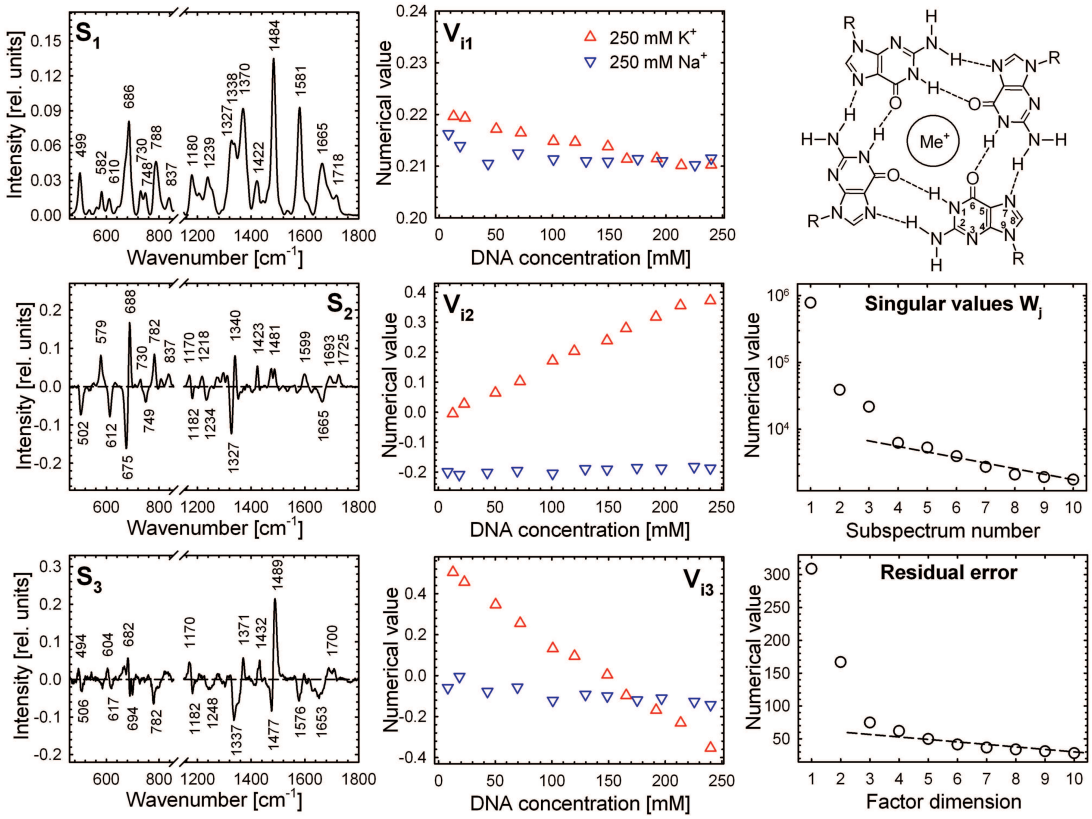



**Figure 3.** Dependence of normalized Raman spectra of $G_3(TTAG_3)_3$ on DNA concentration in the presence of 250 mM of $K^+$ or $Na^+$ (150 mM of PBS, pH 6.8, 5°C). Spectral differences between $K^+$- and $Na^+$-stabilized quadruplex structures and changes induced by increasing DNA concentration are expressed as abstract orthonormal factors obtained by SVD analysis (44). For better insight to real spectral differences and their extent, refer to Figure 4 and Supplementary Figure S3. Right top corner: scheme of a G-tetrad.

is identical to that of low-DNA $G_3(TTAG_3)_3$ at moderate concentration of $K^+$ (Figure 2). Corresponding Raman spectrum (Figure 5, bottom) corroborates structural similarity, exhibiting Raman features indicative for antiparallel structure in the same extent as for low-DNA $K^+$-$G_3(TTAG_3)_3$ in 250 mM $K^+$ (Figure 1). Increase in the $K^+$ concentration to 500 mM is followed by the same spectral changes signalling structural transition towards parallel form, as described above on interquadruplex transition induced by DNA concentration. Difference between high- and low-salt spectra of 50 mM $G_3(TTAG_3)_3$ (Figure 5, bottom) is virtually identical to Raman fingerprint pattern observed on DNA increase (Supplementary Figure S3, bottom). On the other hand, no remarkable structural changes occur for 50 mM $G_3(TTAG_3)_3$ on raising the $Na^+$ concentration (Figure 5, top). The effect of $K^+$ concentration is more profound for higher DNA concentrations, as will be shown later.

Rearrangement of $G_3(TTAG_3)_3$ from antiparallel to (3 + 1) structure is accompanied by a transition of one of the three side/diagonal d(TTA)-loops to double-chain-reversal conformation. In a fully parallel quadruplex, all three d(TTA)-loops have to be double-chain-reversal. However, on the interquadruplex transition induced by high DNA (Figure 1) or $K^+$ concentration (Figure 5), intensity of the dT-loop marker at $611 \pm 2\,cm^{-1}$ is reduced approximately to half of its initial value, whereas intensity of the *C2-endo*/*anti*-dG marker at $581 \pm 1\,cm^{-1}$ increases for ~20%. Consequently, a fraction of $G_3(TTAG_3)_3$ quadruplexes seems to preserve their loops in side/diagonal arrangement. The finding seems to be in qualitative agreement with CD spectra (Figures 2 and 6, top) that affirm partial preservation of (3 + 1) and/or even antiparallel form, as can be concluded from still intense positive CD shoulder at 293 nm. However, up to 20% of the $611\,cm^{-1}$ band intensity persists even under the conditions strongly favouring parallel quadruplex. As will be shown later, the band is still visible even at the highest concentrations of DNA and $K^+$ used in the present study, well beyond DNA concentrations for which CD spectra exhibit just CD features of parallel arrangement. For instance, in the Raman spectrum of 315 mM $G_3(TTAG_3)_3$ slowly annealed in the presence of 500 mM $K^+$ (Figure 8, bottom), normalized

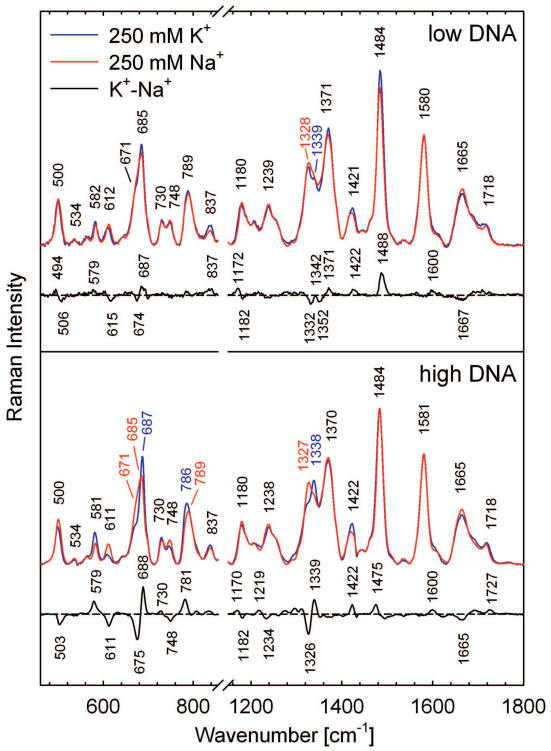



**Figure 4.** Detail comparison of the spectral differences between $K^+$- and $Na^+$-stabilized $G_3(TTAG_3)_3$ quadruplexes at low (8 mM, top) and high (240 mM, bottom) DNA concentration. Normalized Raman spectra correspond to lower and upper concentration limits of the series analysed by SVD in Figure 3.

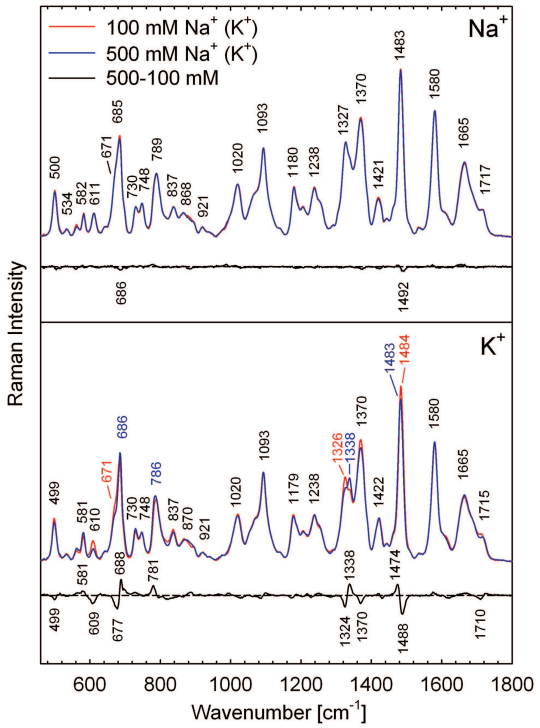



**Figure 5.** Raman spectra of 50 mM $G_3(TTAG_3)_3$ in the presence of 100 and 500 mM of $Na^+$ (top) or $K^+$ (bottom) at 10°C. Difference spectra between the high- and low-salt structures are shown to highlight cation-specific spectral variance. Differential Raman features are labelled according to their apparent maxima/minima that can differ from the maxima/minima of the respective Raman bands.

intensity of the 611 $cm^{-1}$ band still preserves ~20% of the intensity observed for 8 mM $G_3(TTAG_3)_3$ in 100 mM $K^+$, the later sample being considered as the least prone to interquadruplex transition. Beyond persisting conformational heterogeneity of the quadruplexes, the apparent quantitative disagreement between intensity of the 611 $cm^{-1}$ band, its structural assignment and real arrangement of the loops, may stem from diverse conformations of particular dT residues constituting d(TTA)- or $dT_4$-loops.

### Effect of thermal destabilization and annealing on interquadruplex transition

As demonstrated above, antiparallel-to-parallel transition of $G_3(TTAG_3)_3$ can be induced by increasing $K^+$ and/or DNA concentrations, both factors acting in synergy. However, thermal destabilization of the antiparallel folding spontaneously adopted by $G_3(TTAG_3)_3$ in the course of quadruplex formation under the high-DNA and/or high-$K^+$ conditions is a third prerequisite (30). Incidentally, protocols of sample preparation used in some NMR studies (18,19) do not mention heating or annealing. Effect of heating and annealing at high DNA concentrations, a task for which Raman spectroscopy comes exceptionally useful, was thus studied in detail.

As shown in Figure 7, desalted sample of $G_3(TTAG_3)_3$ prepared by repeated centrifugal washing with deionized water adopts non-quadruplex structure even at DNA concentration as high as 520 mM. Common quadruplex markers signalling tight base-pairing within G-tetrads (when located at 1484, 1581 and 1719 $cm^{-1}$) are shifted to the wavenumbers typical for dG hydrogen bonding with $H_2O$ (1486, 1578 and 1689 $cm^{-1}$). However, even in the absence of stabilizing cations, $G_3(TTAG_3)_3$ is probably ordered, as its spectrum contains conformational marker at 837 $cm^{-1}$, which is present with regular B-DNA-family backbone (32), and, except for the band at 1486 $cm^{-1}$, no Raman hypochromism expected for substantially destacked bases (32) is observed. After addition of $K^+$ ions (230 mM) but omitting annealing, the high-DNA $G_3(TTAG_3)_3$ spontaneously adopts antiparallel quadruplex structure, as can be documented by appearance of *syn*-dG markers (671 and 1327 $cm^{-1}$) and dT marker of lateral/diagonal d(TTA)-loops (612 $cm^{-1}$). The $K^+$-structure is close to the antiparallel $Na^+$-folding formed under the same DNA and $Na^+$ concentrations, as seen from the match of their Raman

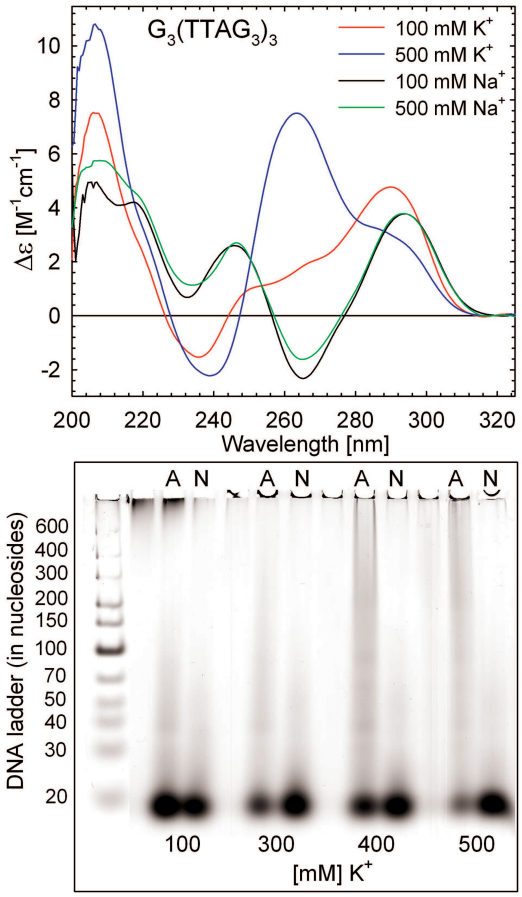



**Figure 6.** Top: CD spectra of 50 mM $G_3(TTAG_3)_3$ in the presence of 100 and 500 mM of $Na^+$ or $K^+$. Positive bands at ~245 and ~293 nm and negative band at ~265 nm affirm antiparallel foldback of the $Na^+$-stabilized $G_3(TTAG_3)_3$, regardless of the $Na^+$ concentration. Even though without negative band at ~265 nm, the CD spectrum of 100 mM $K^+$-stabilized $G_3(TTAG_3)_3$ was recently proven to be indicative for essentially the same antiparallel topology as adopted by $Na^+$-stabilized quadruplex (30), contradicting the previous interpretation attributing CD spectra of this shape to hybrid (3 + 1) form (18,19,22). Transition to the parallel structure via hybrid (3 + 1) forms is accompanied by appearance of a strong positive band at ~263 nm, decrease of the positive band at ~290 nm and the presence of negative band at ~240 nm (30), the CD features visible in the presence of 500 mM $K^+$. Bottom: native PAGE of 50 mM $G_3(TTAG_3)_3$ in 100–500 mM of $K^+$ before heating (N) and after slow annealing (A). The PAGE was performed at 20°C in 30 mM PBS containing 300 mM $K^+$ in total, pH 6.7.

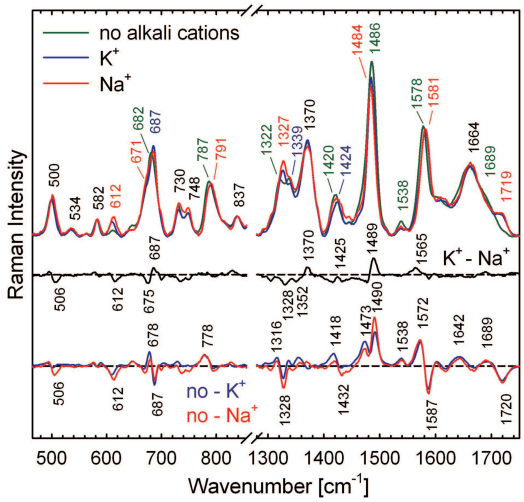



**Figure 7.** Spectral differences between non-quadruplex $G_3(TTAG_3)_3$ structure at extremely high DNA concentration (520 mM; no alkali cations present) and initial quadruplex structures adopted by highly concentrated $G_3(TTAG_3)_3$ (330 mM) after adjustment of $K^+$ or $Na^+$ concentrations to 230 mM, but before annealing.

spectra and respective differential curves (Figure 7, no-$K^+$ and no-$Na^+$ curves). Although Raman difference spectrum between 'unheated' $K^+$- and $Na^+$-stabilized quadruplexes at high-DNA (330 mM) reveals some spectral differences (Figure 7, $K^+$-$Na^+$ curve), they are virtually identical with those depicted in $K^+$–$Na^+$ difference spectrum of $G_3(TTAG_3)_3$ slowly annealed at low-DNA conditions of 8 mM (Figure 4, top). It can be suggested that, according to Raman markers, structure of $K^+$-stabilized $G_3(TTAG_3)_3$ quadruplex formed at extremely high DNA concentration before heating/annealing is virtually identical with that of the thermodynamically equilibrated $K^+$-structure formed at low-DNA conditions by a standard annealing. Raman spectra of freshly prepared unheated $K^+$-samples were found to be independent of $G_3(TTAG_3)_3$ concentration. However, if the unheated high-DNA solutions remain standing for several days at room temperature, Raman signs of antiparallel-to-parallel transition begin to appear slowly.

The spontaneously adopted antiparallel structure of $K^+$-stabilized $G_3(TTAG_3)_3$ is evidently not a thermodynamically most stable quadruplex arrangement, especially at extremely high DNA concentrations. For example, when heated to 95°C and rapidly (in <1 min) cooled down to room temperature, partial (~50% according to differential Raman features) switching to the parallel form is achieved, as evident from the comparison with a slowly annealed sample (Figure 8, top). The effect is synergically amplified by increasing concentration of $K^+$, as the difference between rapidly cooled and slowly annealed differential Raman spectra becomes smaller in 500 mM $K^+$ (Figure 8, bottom). By decreasing the $G_3(TTAG_3)_3$ concentration, the differences between unheated, rapidly cooled and slowly annealed Raman spectra decrease rapidly, becoming virtually indiscernible <100 mM DNA in 230 mM $K^+$, although yet still well apparent for 30 mM $G_3(TTAG_3)_3$ in 500 mM $K^+$. For $Na^+$-stabilized $G_3(TTAG_3)_3$, no such dependence on heating/annealing was observed.

To better understand thermal destabilization of initial antiparallel $K^+$-form of $G_3(TTAG_3)_3$ and interquadruplex switching under DNA concentrations inaccessible to CD

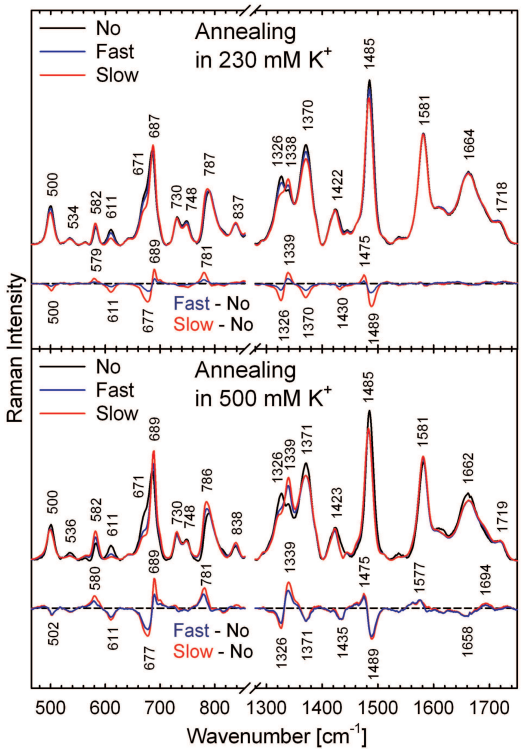



**Figure 8.** Effect of the annealing mode and $K^+$ concentration on $G_3(TTAG_3)_3$ quadruplex structure at extremely high DNA concentration. Raman spectra of 330 and 315 mM $G_3(TTAG_3)_3$ in the presence 230 and 500 mM of $K^+$, respectively, before thermal denaturation (No) and after fast (Fast) and slow (Slow) annealing. Differential spectra between thermally destabilized, thermodynamically equilibrated and initially adopted $K^+$-stabilized $G_3(TTAG_3)_3$ structures are shown to depict differences.

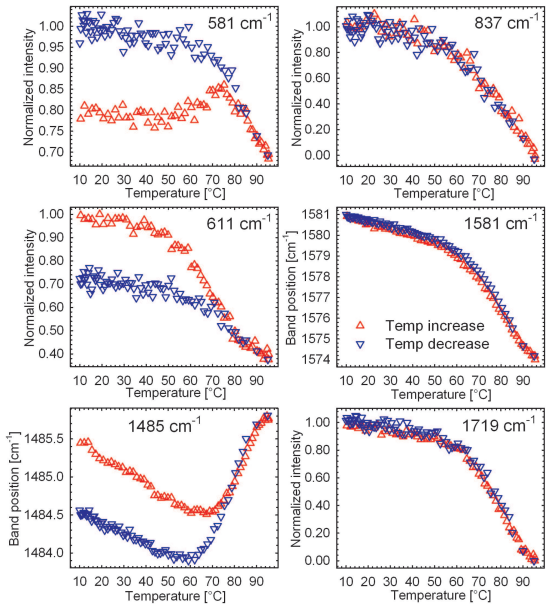



**Figure 9.** Heating and cooling Raman profiles of selected Raman bands sensitive (581, 611 and 1485 $cm^{-1}$) and insensitive (837, 1581 and 1719 $cm^{-1}$) to antiparallel-to-parallel switching of the $K^+$-stabilized $G_3(TTAG_3)_3$ quadruplex. Initially, antiparallel $K^+$-quadruplex (330 mM), spontaneously formed on addition of 230 mM $K^+$, was heated from 10°C to 95°C and then cooled down at the same rate of ~1.5°C/min.

or absorption measurements, annealing process was monitored by Raman scattering on the temperature increase and decrease. The 330 mM of $G_3(TTAG_3)_3$ in 230 mM of $K^+$ was heated continuously from 10°C to 95°C and then cooled down at the rate of ~1.5°C/min. The SVD analysis of the entire Raman dataset sorted out spectral changes common for an overall denaturation/renaturation process of both quadruplex structures ($S_2$, $V_{i2}$), and separated them from the features reflecting structural differences between respective quadruplexes ($S_3$, $V_{i3}$; Supplementary Figure S4). As evidenced by a close match of temperature dependencies of the $V_{i2}$ coefficients on the temperature increase and decrease, overall renaturation curve (giving rise to the parallel quadruplex) follows well denaturation curve of the antiparallel quadruplex (Supplementary Figure S4). On the contrary, orthogonal subspectrum $S_3$ and coefficients $V_{i3}$ clearly discriminate between antiparallel and parallel folds, except for the temperatures >77°C when Raman spectra measured on temperature increase and decrease become identical.

To better depict different temperature dependencies of the Raman features sensitive and insensitive to quadruplex arrangement, spectral parameters of selected bands are shown in Figure 9. Raman marker of regular B-DNA-family backbone (837 $cm^{-1}$) and general markers of interbase hydrogen bonding within G-tetrads (position and intensity of the ~1581 and ~1719 $cm^{-1}$ bands, respectively) affirm reversible melting of the sugar-phosphate backbone and similar strengths of the N2H and O6 hydrogen bonds within G-tetrads in $K^+$-stabilized antiparallel and parallel forms under the present DNA and $K^+$ concentrations. On the contrary, newly established *C2-endo/anti*-dG marker at 581 $cm^{-1}$, the 611 $cm^{-1}$ marker of the lateral/diagonal arrangement of d(TTA)-loops and an indicator of strong Hoogsteen hydrogen bonding at N7-dG (1485 $cm^{-1}$) exhibit more complicated temperature dependences (Figure 9). Convergence of the 581, 611 and 1485 $cm^{-1}$ melting curves of unheated antiparallel $G_3(TTAG_3)_3$ towards renaturation curves leading to formation of the parallel quadruplex at low temperatures (indicated by the 837, 1581 and 1719 $cm^{-1}$ bands; Figure 9) implies thermally induced structural switching of the antiparallel folding to a parallel-like arrangement that precedes complete quadruplex unfolding and complete disintegration of G-tetrads at higher temperatures. Denaturation/renaturation profiles exhibiting similar convergence were reported recently by CD spectroscopy for $K^+$-stabilized

$G_4(T_2G_4)_3$ *Tetrahymena* quadruplex (53), and interpreted as a sign of thermally induced conversion of intramolecular antiparallel conformers to more stable parallel forms. Poor reversibility observed during the first cooling of the melted sample was ascribed to the slow kinetics of the formation of antiparallel conformers (53). A similar preference for parallel structure after thermal destabilization of spontaneously preformed antiparallel quadruplex can be seen here for $K^+$-stabilized $G_3(TTAG_3)_3$. Once slowly annealed, $K^+$-stabilized parallel form of $G_3(TTAG_3)_3$ melts reversibly without any hysteresis, as can be evidenced by a close match of denaturation and renaturation profiles visualized by a factor analysis (Supplementary Figure S6).

It is worthwhile to note that under the extreme DNA and $K^+$ concentrations, unheated samples constitute clear and relatively non-viscous solution, which turns irreversibly into a solid gel after annealing. Increased viscosity or gelation after annealing was observed also for extremely high-DNA samples at moderate $K^+$ concentrations (200–300 mM). As the interquadruplex transition induced by DNA concentration assumes some physical contacts between $G_3(TTAG_3)_3$ molecules, a question about molecularity of the parallel form arises. According to electrophoretic mobilities, the samples identified by CD and Raman spectra as $K^+$-stabilized parallel quadruplexes migrate preferentially as monomers, although slowly migrating continuous smear indicates the presence of multi-molecular structures, especially in the samples annealed at the highest $K^+$ and DNA concentrations (Figures 2 and 6). Except for a weak band at the position corresponding to a bimolecular form visible in Figure 2, the smear lacks any structure. We suppose that the presence of the monomer band and slowly migrating, unstructured smear corresponds better to a wide distribution of the self-stacked associates consisting of intramolecular parallel quadruplexes (30) than to four-molecular, four-stranded parallel structures or multi-molecular, locally four-stranded parallel structures with out-of-register alignment of adjacent strands. As four- and multi-molecular parallel quadruplexes are more thermodynamically and kinetically stable than the higher-order self-associates of monomeric quadruplexes, their formation as well as disintegration should be less probable and slower than in the case of self-associates. The latter can be thus, under the diluting conditions of electrophoresis, disintegrated into the monomers much easier and faster than the four- and multi-molecular parallel quadruplexes, and it could better explain appearance of the monomer band in PAGE. In the case of molecular crowding induced by DNA itself, the self-association by stacking is a factor stabilizing parallel arrangement of each individual $G_3(TTAG_3)_3$ quadruplex. On the other hand, such a self-association requires quadruplexes with parallel or at least (3 + 1) intramolecular arrangement (30).

## CONCLUSIONS

It is generally accepted that quadruplex structures determined for crystalline states cannot be uncritically transferred to solutions. Less accepted is a fact that even quadruplex structures determined by NMR in solutions at relatively high DNA concentrations can differ from those existing in more dilute solutions investigated by absorption and CD spectroscopies.

By means of Raman spectroscopy that can be applied over wide range of DNA concentrations, we demonstrate here that quadruplex structure of $K^+$-stabilized $G_3(TTAG_3)_3$, the shortest four-repeat model of human telomeric quadruplex, depends on DNA concentration, as well as on the mode of sample preparation. Complex interplay of DNA and $K^+$ concentrations with other factors affecting quadruplex structure, for example, presence or absence of thermal destabilization and annealing in protocols, storage conditions and time elapsed between the sample preparation and measurement, could be the reason why distinct $K^+$-stabilized quadruplex structures of human telomeric sequences were reported by various experimental methods using DNA concentrations differing by orders of magnitude. Raman spectroscopy could be of great benefit in the current effort to better understand quadruplex structural properties under the crowding conditions existing in cells.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–6.

## FUNDING

Grant Agency of the Charles University in Prague [3222/2008]; Grant Agency of the Czech Republic [P202/09/0193, P205/12/0466]; Ministry of Education of the Czech Republic [MSM0021620835]; 'CEITEC—Central European Institute of Technology' [CZ.1.05/1.1.00/02.0068] European Regional Development Fund. Funding for open access charge: Grant Agency of the Czech Republic [P202/09/0193].

*Conflict of interest statement*. None declared.

## REFERENCES

1. Blackburn,E.H. (2001) Switching and signaling at the telomere. *Cell*, **106**, 661–673.
2. Londono-Vallejo,J.A., DerSarkissian,H., Cases,L. and Thomas,G. (2001) Differences in telomere length between homologous chromosomes in human. *Nucleic Acids Res.*, **29**, 3164–3171.
3. Murchie,A.I.H. and Lilley,D.M.J. (1994) Tetraplex folding of telomere sequences and the inclusion of adenine bases. *EMBO J.*, **13**, 993–1001.
4. Harley,C.B., Futcher,A.B. and Greider,C.W. (1990) Telomeres shorten during aging of human fibroblasts. *Nature*, **345**, 458–460.
5. Kim,N.W., Piatyszek,M.A., Prowse,K.R., Harley,C.B., West,M.D., Ho,P.L.C., Coviello,G.M., Wright,W.E., Weinrich,S.L. and Shay,J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011–2015.
6. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.

7. Zimmermann,S. and Martens,U.M. (2007) Telomeres and telomerase as targets for cancer therapy. *Cell Mol. Life Sci.*, **64**, 906–921.
8. De Cian,A., Lacroix,L., Douarre,C., Temime-Smaali,N., Trentesaux,C., Riou,J.F. and Mergny,J.L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
9. Neidle,S. (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, **277**, 1118–1125.
10. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. *Structure*, **1**, 263–282.
11. Balagurumoorthy,P. and Brahmachari,S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
12. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
13. Redon,S., Bombard,S., Elizondo-Riojas,M.A. and Chottard,J.C. (2003) Platinum cross-linking of adenines and guanines on the quadruplex structures of the $AG_3(T_2AG_3)_3$ and $(T_2AG_3)_4$ human telomere sequences in $Na^+$ and $K^+$ solutions. *Nucleic Acids Res.*, **31**, 1605–1613.
14. He,Y.J., Neumann,R.D. and Panyutin,I.G. (2004) Intramolecular quadruplex conformation of human telomeric DNA assessed with $^{125}$I-radioprobing. *Nucleic Acids Res.*, **32**, 5359–5367.
15. Ying,L.M., Green,J.J., Li,H.T., Klenerman,D. and Balasubramanian,S. (2003) Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA*, **100**, 14629–14634.
16. 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.
17. Vorlickova,M., Chladkova,J., Kejnovska,I., Fialova,M. and Kypr,J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.*, **33**, 5851–5860.
18. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in $K^+$ solution: an intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
19. Ambrus,A., Chen,D., Dai,J.X., Bialis,T., Jones,R.A. and Yang,D.Z. (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.
20. Phan,A.T., Luu,K.N. and Patel,D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in $K^+$ solution. *Nucleic Acids Res.*, **34**, 5715–5719.
21. Qi,J.Y. and Shafer,R.H. (2005) Covalent ligation studies on the human telomere quadruplex. *Nucleic Acids Res.*, **33**, 3185–3192.
22. Xu,Y., Noguchi,Y. and Sugiyama,H. (2006) The new models of the human telomere d[$AGGG(TTAGGG)_3$] in $K^+$ solution. *Bioorgan. Med. Chem.*, **14**, 5584–5591.
23. Chang,C.C., Chien,C.W., Lin,Y.H., Kang,C.C. and Chang,T.C. (2007) Investigation of spectral conversion of $d(TTAGGG)_4$ and $d(TTAGGG)_{13}$ upon potassium titration by a G-quadruplex recognizer BMVC molecule. *Nucleic Acids Res.*, **35**, 2846–2860.
24. Matsugami,A., Xu,Y., Noguchi,Y., Sugiyama,H. and Katahira,M. (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a $K^+$ solution. *FEBS J.*, **274**, 3545–3556.
25. Antonacci,C., Chaires,J.B. and Sheardy,R.D. (2007) Biophysical characterization of the human telomeric $(TTAGGG)_4$ repeat in a potassium solution. *Biochemistry*, **46**, 4654–4660.
26. Dai,J.X., Carver,M., Punchihewa,C., Jones,R.A. and Yang,D.Z. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in $K^+$ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.
27. Rujan,I.N., Meleney,J.C. and Bolton,P.H. (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res.*, **33**, 2022–2031.
28. Miyoshi,D., Karimata,H. and Sugimoto,N. (2005) Drastic effect of a single base difference between human and *Tetrahymena* telomere sequences on their structures under molecular crowding conditions. *Angew. Chem. Int. Ed. Engl.*, **44**, 3740–3744.
29. 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.
30. Renciuk,D., Kejnovska,I., Skolakova,P., Bednarova,K., Motlova,J. and Vorlickova,M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant $K^+$ solutions. *Nucleic Acids Res.*, **37**, 6625–6634.
31. Lim,K.W., Amrane,S., Bouaziz,S., Xu,W.X., Mu,Y.G., 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.
32. Benevides,J.M., Overman,S.A. and Thomas,G.J. Jr (2005) Raman, polarized Raman and ultraviolet resonance Raman spectroscopy of nucleic acids and their complexes. *J. Raman Spectrosc.*, **36**, 279–299.
33. Thomas,G.J. Jr (1999) Raman spectroscopy of protein and nucleic acid assemblies. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 1–27.
34. Miura,T. and Thomas,G.J. Jr (1994) Structural polymorphism of telomere DNA: interquadruplex and duplex-quadruplex conversions probed by Raman spectroscopy. *Biochemistry*, **33**, 7848–7856.
35. Miura,T., Benevides,J.M. and Thomas,G.J. Jr (1995) A phase diagram for sodium and potassium-ion control of polymorphism in telomeric DNA. *J. Mol. Biol.*, **248**, 233–238.
36. Miura,T. and Thomas,G.J. Jr (1995) Structure and dynamics of interstrand guanine association in quadruplex telomeric DNA. *Biochemistry*, **34**, 9645–9654.
37. Laporte,L. and Thomas,G.J. Jr (1998) Raman spectral studies of nucleic acids part LXVI—structural basis of DNA recognition and mechanism of quadruplex formation by the β subunit of the *Oxytricha* telomere binding. *Biochemistry*, **37**, 1327–1335.
38. Laporte,L. and Thomas,G.J. Jr (1998) A hairpin conformation for the 3′ overhang of *Oxytricha nova* telomeric DNA. *J. Mol. Biol.*, **281**, 261–270.
39. Krafft,C., Benevides,J.M. and Thomas,G.J. Jr (2002) Secondary structure polymorphism in *Oxytricha nova* telomeric DNA. *Nucleic Acids Res.*, **30**, 3981–3991.
40. Abu-Ghazalah,R.M., Irizar,J., Helmy,A.S. and Macgregor,R.B. (2010) A study of the interactions that stabilize DNA frayed wires. *Biophys. Chem.*, **147**, 123–129.
41. Pagba,C.V., Lane,S.M. and Wachsmann-Hogiu,S. (2011) Conformational changes in quadruplex oligonucleotide structures probed by Raman spectroscopy. *J. Biomed. Opt.*, **2**, 207–217.
42. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
43. Gray,D.M., Hung,S.H. and Johnson,K.H. (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol.*, **246**, 19–34.
44. Palacký,J., Mojzeš,P. and Bok,J. (2011) SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations. *J. Raman Spectrosc.*, **42**, 1528–1539.
45. Movileanu,L., Benevides,J.M. and Thomas,G.J. Jr (2002) Determination of base and backbone contributions to the thermodynamics of premelting and melting transitions in B DNA. *Nucleic Acids Res.*, **30**, 3767–3777.
46. Weisz,K., Leitner,D., Krafft,C. and Welfle,H. (2000) Structural heterogeneity in intramolecular DNA triple helices. *Biol. Chem.*, **381**, 275–283.
47. Thomas,G.J. Jr, Benevides,J.M., Overman,S.A., Ueda,T., Ushizawa,K., Saitoh,M. and Tsuboi,M. (1995) Polarized Raman spectra of oriented fibers of A-DNA and B-DNA: anisotropic

and isotropic local Raman tensors of base and backbone vibrations. *Biophys. J.*, **68**, 1073–1088.
48. Movileanu,L., Benevides,J.M. and Thomas,G.J. Jr (1999) Temperature dependence of the Raman spectrum of DNA. Part I–Raman signatures of premelting and melting transitions of Poly(dA–dT)•Poly(dA–dT). *J. Raman Spectrosc.*, **30**, 637–649.
49. Peticolas,W.L., Kubasek,W.L., Thomas,G.A. and Tsuboi,M. (1987) Biological applications of Raman spectroscopy. In: Spiro,T.G. (ed.), *Raman Spectra and the Conformations of Biological Macromolecules.* John Wiley & Sons, New York, pp. 81–133.
50. Benevides,J.M. and Thomas,G.J. Jr (1988) A solution structure for Poly(rA).Poly(dT) with different furanose pucker and backbone geometry in rA and dT strands and intrastrand hydrogen bonding of adenine 8CH. *Biochemistry*, **27**, 3868–3873.
51. Serban,D., Benevides,J.M. and Thomas,G.J. Jr (2003) HU protein employs similar mechanisms of minor-groove recognition in binding to different B-DNA sites: demonstration by Raman spectroscopy. *Biochemistry*, **42**, 7390–7399.
52. Gray,R.D., Li,J. and Chaires,J.B. (2009) Energetics and kinetics of a conformational switch in G-quadruplex DNA. *J. Phys. Chem. B*, **113**, 2676–2683.
53. Víglaský,V., Bauer,L. and Tlučková,K. (2010) Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats. *Biochemistry*, **49**, 2110–2120.

# Polymorphism of Human Telomeric Quadruplex Structure Controlled by DNA Concentration: A Raman Study

## Supplementary data

Nucleic Acids Research, accepted (2012)

Jan Palacký[1], Michaela Vorlíčková[2,3*], Iva Kejnovská[2,3], Peter Mojzeš[1*]

[1]Charles University in Prague, Faculty of Mathematics and Physics, Institute of Physics, Ke Karlovu 5, CZ-121 16 Prague 2, Czech Republic

[2]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, CZ-612 65 Brno, Czech Republic

[3]CEITEC – Central European Institute of Technology, Masaryk University, Kamenice 5, CZ-625 00 Brno, Czech Republic

Corresponding authors:
Peter Mojzeš, *mojzes@karlov.mff.cuni.cz*,
Michaela Vorlíčková, *mifi@ibp.cz*

**Detailed interpretation of Raman markers diagnostic of quadruplex structures**

Structural interpretation of the conventional non-resonant Raman spectra of $G_3(TTAG_3)_3$ quadruplexes is based on the ultimate Raman studies of interquadruplex switching between antiparallel and parallel forms of *Oxytricha* telomeric sequence $(T_4G_4)_4$ controlled by the type, concentration, and molar ratio of $Na^+$ and $K^+$ cations (34, 35). Raman markers established there, refined in subsequent Raman studies (36-39) and catalogued (32, 39) are listed in Supplementary Tab. S1. Since human telomeric sequences incorporate also dA residues, proper assignment of the $G_3(TTAG_3)_3$ spectra has required analysis of dA contribution missing in the previous Raman quadruplex studies. Raman spectra of homooligonucleotides $(dG)_{15}$, $(dT)_{15}$ and $(dA)_{15}$ in $Na^+$- and $K^+$-PBS have been acquired (Supplementary Fig. S1) and used for spectral interpretation (Supplementary Fig. S2).

Quadruplex Raman markers can be assorted according to structural information that they provide. There are **(i)** markers signaling formation of G-tetrads and constitution of the quadruplex, **(ii)** markers sensitive to conformation of dG residues distinguishing between *C2'-endo/anti* and *C2'-endo/syn* conformers, **(iii)** the bands sensitive to the arrangement of dT-loops, and **(iv)** markers of overall geometry and regularity of the sugar-phosphate backbone.

**(i) Markers sensitive to formation of G-tetrads**

Raman signatures of G-tetrad formation concern structurally informative dG bands located at ~1483±2, 1581±1 and 1719±3 $cm^{-1}$. The bands are sensitive to the strength of hydrogen bonding at dG-N7, dG-N2H and dG-O6, respectively (see G-tetrad scheme in Fig. 3), distinguishing thus stronger hydrogen bonding of dG within G-tetrads from weaker bonding to surrounding $H_2O$.

Major contribution to the intensity of the most intense Raman band of $G_3(TTAG_3)_3$ at ~1483±2 $cm^{-1}$ comes from vibrations of dG residues, although there are some minor contributions from dA and dT bands at ~1486 and 1483 $cm^{-1}$, respectively (Supplementary Fig. S2). The band is sensitive to Hoogsteen hydrogen bonding between dG-N7 serving as an acceptor and the donor dG-N2H group (32). In single-strands and B-DNA duplexes, the band typically occurs at ~1486 – 1490 $cm^{-1}$ due to weak hydrogen bonding with $H_2O$, and downshifts for ~ 6 – 9 $cm^{-1}$ upon formation of Hoogsteen hydrogen bonds within G-tetrads (34, 39). According to the present study on $G_3(TTAG_3)_3$, exact position of ~1483±2 $cm^{-1}$ band is sensitive to interquadruplex transition induced by $K^+$ and DNA concentration, as well as to the type of stabilizing cation. The band located at 1484 $cm^{-1}$ in 100 mM $K^+$ slightly downshifts (~1 $cm^{-1}$) and decreases in intensity (~10% hypochromism) as the concentration of $K^+$ increases to 500 mM (Fig. 5). Similar changes can be observed on the antiparallel-to-parallel interquadruplex transition induced by high DNA concentration (Fig. 1 and Supplementary Fig. S3). On the other hand, the same relative downshift and decrease in intensity at ~1483 $cm^{-1}$ can bee seen in Raman spectrum of $(dG)_{15}$ in 200 mM $K^+$ in comparison to 200 mM $Na^+$ solution (Supplementary Fig. S1), albeit $(dG)_{15}$ evidently forms an intermolecular parallel four-stranded quadruplex (36) with both cations. Additional downshift of the ~1483±2 $cm^{-1}$ band may thus indicate minor changes in

the microenvironment of the five-membered imidazole ring of dG due to dG-syn/dG-anti glycosyl reorientation, as well as cation-dependent differences in hydration of the grooves or in the stacking of G-tetrads. However, as demonstrated by Raman melting profile of $K^+$-stabilized $G_3(TTAG_3)_3$ quadruplex (Fig. 9 and Supplementary Fig. S6), thermal dependency of the position of 1484 $cm^{-1}$ band exhibits rather complicated behavior impeding simple and straightforward interpretation of the downshift.

Quadruplex band at ~1581±1 $cm^{-1}$ is sensitive to hydrogen bonding at exocyclic dG-N2H donor site. The band exhibits downshift to ~1574 $cm^{-1}$ on the quadruplex melting (Fig. 9 and Supplementary Fig. S5). For $G_3(TTAG_3)_3$ adopting a single-strand non-quadruplex structure (e.g. in the absence of stabilizing cations), the band is located at 1578 $cm^{-1}$ (Fig. 7). According to our study using solvent- and background-corrected, normalized Raman spectra with precisely calibrated wavenumber scales, the position and intensity of the ~1581±1 $cm^{-1}$ is virtually insensitive to the type of stabilizing cation, as well as to interquadruplex transition, since no significant spectral changes are observed on the $K^+/Na^+$ replacement (Fig. 4), and on the antiparallel-to-parallel transition induced by increased DNA or $K^+$ concentration (Figs. 1 and 5). Relative insensitivity of the ~1581±1 $cm^{-1}$ band to interquadruplex switching contrasts with high sensitivity of the nearby located ~1483±2 $cm^{-1}$ band.

Information on the formation of G-tetrads and on the strength of hydrogen bonds can be drawn also from a weak but clearly resolvable shoulder at ~1719±3 $cm^{-1}$ that is sensitive to participation of the dG-O6 in the interbase hydrogen bonding with dG-N1H (39). In non-quadruplex structures the band is located at ~1688±3 $cm^{-1}$ (Fig. 7). Intensity of the band seems to be insensitive to the interquadruplex transition (Fig. 9), and can thus serve as an indicator of general quadruplex structure.

**(ii) Markers sensitive to *C2'-endo/anti* and *C2'-endo/syn* conformation of dG**

Raman bands at 686±2 and 1338±2 $cm^{-1}$ were ascribed to *C2'-endo/anti* conformers of dG (32, 34-36, 39). On the other hand, shoulder at 671±2 and the band at 1326±2 $cm^{-1}$ indicate *C2'-endo/syn*-dG. Structural transition of antiparallel-to-parallel $G_3(TTAG_3)_3$ comprises increase in the population of *C2'-endo/anti-dG* at the expense of *C2'-endo/syn-dG*, as can be illustrated by relative intensity increase of the bands at 686 and 1338 $cm^{-1}$, accompanied by intensity decrease of the shoulder at 671 $cm^{-1}$ and the 1326 $cm^{-1}$ band (Fig. 1). Although there is no doubt that *syn/anti*-dG Raman markers qualitatively indicate antiparallel-to-parallel transition of $K^+$-stabilized $G_3(TTAG_3)_3$, identification of the initial and final structures and accurate quantification of the number of dG residues concerned in the transition simply from Raman spectra are neither straightforward nor simple. As *C2'-endo/syn* and *C2'-endo/anti* populations of dG residues in antiparallel $G_3(TTAG_3)_3$ are equal whereas parallel structure consists exclusively of *C2'-endo/anti*-dG conformers, one could expect twofold intensity increase at 686 and 1338 $cm^{-1}$ and disappearance of the bands at 671 and 1326 $cm^{-1}$ upon *complete* transition from fully antiparallel to fully parallel folding. On the other hand, antiparallel and (3+1) quadruplex structures, differing in a single dG residue switched from *C2'-endo/syn* to *C2'-endo/anti*, are expected to be distinguishable due to 6:6 and 5:7 intensity ratio of the corresponding Raman markers, respectively. Unfortunately, because of the complicated overlap with neighboring Raman bands (Supplementary Fig. S1 and S2) and resulting ambiguity in the absolute non-

Raman background, reliable decomposition of the *syn*-dG/*anti*-dG markers into separate bands and quantification of their *absolute* intensities is affected by a substantial uncertainty. Although *relative* intensity changes can be evaluated rather accurately from differential Raman spectra, the baseline for quantification of the number of dG residues involved in a transition is rather unsure. Consequently, Raman-based conclusions on the $K^+$-structures of $G_3(TTAG_3)$, especially differentiation between antiparallel and (3+1) $K^+$-forms, have to be supported by a careful and critical comparison with Raman spectra of the $Na^+$-stabilized, undoubtedly antiparallel $G_3(TTAG_3)_3$, taking into account particularities of vibrational spectroscopy (i.e. possible non-uniformity of Raman cross-sections of the respective *syn*-dG/*anti*-dG vibrational modes).

**(iii) Markers sensitive to arrangement of dT-loops**

The band at 611±2 $cm^{-1}$ was initially assigned as a marker of *C2'-endo/syn*-dG (34), since its intensity strongly decays on the antiparallel-to-parallel quadruplex transition of $(T_4G_4)_4$ (35). Later, on the basis of protium-deuterium exchange (36), the band was reassigned to dT conformer preferred within lateral $dT_4$-loops of the antiparallel foldback of *Oxytricha* $(T_4G_4)_4$ (36, 39). No suitable precursor of the band can be found in Raman spectra of $(dA)_{15}$, $(dG)_{15}$ and $(dT)_{15}$ (Supplementary Figs. S1 and S2). Furthermore, the 611±2 $cm^{-1}$ band was reported not to appear in non-telomeric sequences $d(TG)_8$ and $dT_6(TG)_8$ containing $dT_n$ tracts, but organized in other structures than dT-loops (38, 39). The band is missing also in other dT-containing structures such as A- and B-DNA duplexes (36, 39). On the other hand, it appears at 615 $cm^{-1}$ in the Raman spectra of intramolecular triple helices consisting of purine and pyrimidine tracts interconnected by $dT_4$ loops (46). It is thus speculated (36, 39) that the 611±2 $cm^{-1}$ band is a marker of an unusual *C2'-endo*/*syn*-dT conformer, occurrence of which was suggested to be restricted exclusively to lateral (edgewise) or diagonal loops of foldback quadruplexes or in other dT-loop-containing DNA structures. Disappearance of the 611±2 $cm^{-1}$ band indicates rearrangement of the loops to double-chain-reversal orientation typical for intramolecular parallel quadruplex. If the band assignment (32, 36, 39) is valid, then, according to our results on $G_3(TTAG_3)_3$, it could serve as a more universal Raman marker distinguishing between side/diagonal and double-chain-reversal loop arrangement also for dT-loops incorporating dA residue.

Since the 611±2 $cm^{-1}$ band is located in the low-wavenumber region containing few relatively separate bands with simple non-Raman baseline and negligible overlaps, it is suitable for quantitative analysis. The low-wavenumber region below 650 $cm^{-1}$ was only partly exploited in the previous Raman studies (32, 34-36, 39). In addition to the 611±2 $cm^{-1}$ marker, the region contains four relatively separate bands at 499±2, 534±1, 564±1 and 581±1 $cm^{-1}$. Two of them (499±2 and 581±1 $cm^{-1}$) exhibit intensity variations on the interquadruplex transition induced by increase of $K^+$ and DNA concentration (Figs. 1, 3, 4 and 5). Although the spectral changes of the relatively weak low-wavenumber bands may seem to be tiny, they are highly reproducible and constitute characteristic pattern easily recognizable in differential Raman spectra whenever the well-established Raman markers 671/686 $cm^{-1}$ and 1326/1338 $cm^{-1}$ indicate *syn/anti* transition of dG residues.

For the $Na^+$-stabilized antiparallel $G_3(TTAG_3)_3$, the bands at 581±1 and 611±2 $cm^{-1}$ are of comparable heights. Similar intensity ratio can be seen in the Raman spectra $K^+$-stabilized $G_3(TTAG_3)_3$ under the

conditions favoring antiparallel quadruplex structure (low-$K^+$, low-DNA). However, detailed comparison of the putative $Na^+$- and $K^+$-antiparallel folds reveals slight differences, e.g. downshift (1 $cm^{-1}$), narrowing and intensity decrease (~15%) of the 611±2 $cm^{-1}$ band in the $K^+$-form. On the interquadruplex transition induced by high $K^+$ and DNA concentration, intensity of the 611±2 $cm^{-1}$ band is reduced, whereas intensity of the 581 $cm^{-1}$ band increases.

The 499 ± 2 $cm^{-1}$ band was formerly assigned to the scissoring mode of the phosphodioxy ($PO_2^-$) group (47). Its shape and intensity can be satisfactorily explained as a simple linear combination of the band precursors present in the spectra of the corresponding homooligonucleotides (Supplementary Fig. S2) with the main contribution coming from dT residues. The band is slightly more intense at lower $K^+$ concentrations or in the presence of $Na^+$ (regardless of the $Na^+$ concentration), probably reflecting higher flexibility of the sugar-phosphate backbone of d(TTA)-loops in the antiparallel $G_3(TTAG_3)_3$ backfold. Unlike the 499 $cm^{-1}$ band, the principal contribution to the band at 581 $cm^{-1}$ comes from dG residues (Supplementary Figs. S1 and S2). Its intensity increases on the interquadruplex transition induced by high concentration of $K^+$ or DNA, and follows increase of 686 and 1338 $cm^{-1}$ markers of *C2'-endo/anti*-dG. Albeit weak, the change evidently indicates quadruplex switching to parallel structure. Slight but apparent intensity decrease of the 499 $cm^{-1}$ band accompanied by comparable intensity increase of the 581 $cm^{-1}$ band can be suggested as a new signature of the antiparallel-to-parallel quadruplex transition.

The band at ~1370 $cm^{-1}$ is dominated by the most prominent Raman band of dT (at 1376 $cm^{-1}$ in single-strand $(dT)_{15}$, Supplementary Fig. S1), however with some contributions from dG and dA providing their own precursors at 1363 and 1380 $cm^{-1}$, respectively (Supplementary Fig. S1). Intensity of the dT band near ~1375 - 1380 $cm^{-1}$ in B-DNA duplexes was previously reported to increase with increasing hydrophobicity of the environment of dT methyl groups (51). Consequently, an intensity decrease at 1370 $cm^{-1}$ observed in $G_3(TTAG_3)_3$ on the increase of $K^+$ may be interpreted as easier accessibility of the solvent molecules to dT methyl groups in double-chain-reversal d(TTA)-loops of (3+1) or parallel quadruplexes than in the lateral/diagonal loops of the antiparallel fold. Although real molecular mechanism explaining intensity change of the dT band may be different, we suggest that it reflects rearrangement of d(TTA)-loops and can serve as Raman marker of the loop arrangement.

**(iv) Markers sensitive to overall geometry and regularity of the sugar-phosphate backbone**

Raman spectra of $Na^+$- and $K^+$-stabilized $G_3(TTAG_3)_3$ quadruplexes display DNA phosphate markers at 837±1 and 1093±1 $cm^{-1}$ (Fig. 1). Positions of both markers indicate that phospodiester (O-P-O) and phosphodioxy ($PO_2^-$) groups have the local geometries similar to canonical B-DNA (32, 34). Whereas position and shape of the 837±1 $cm^{-1}$ band seem to be insensitive to interquadruplex transition, band broadening, intensity decrease and finally disappearance is observed on the quadruplex melting at increased temperatures (Fig. 9). On the other hand, intensity of the 1093±1 $cm^{-1}$ band associated with $PO_2^-$ symmetric stretching mode remains largely unchanged on the quadruplex melting, and can serve as intensity standard for Raman melting studies (45).

Further Raman marker characteristic for regular B-DNA-family backbone (32, 34), but diagnostic for antiparallel-to-parallel transition of $G_3(TTAG_3)_3$, is a broad band with a maximum at 788±2 $cm^{-1}$ (Figs. 1, 4 and 5). The band consists of spectral contributions from dG, dT as well as dA residues (51) (Supplementary Figs. S1 and S2). Previously it was attributed to symmetrical stretching mode of the phosphodiester (O-P-O) group (48-50). The band gradually downshifts from ~788 to 786 $cm^{-1}$ (with an intensity increase at the short-wavenumber side) on the antiparallel-to-parallel transition induced by high concentration of DNA or $K^+$ (Fig. 1, 4 and 5). It is likely that the sharp differential feature located at ~781 $cm^{-1}$ in all differential spectra between parallel and antiparallel $G_3(TTAG_3)_3$ forms is primarily due to rearrangement of the quadruplex sugar-phosphate backbone involving mainly dG (and to a lesser extent dT) residues since the same feature can be found in differential Raman spectra describing antiparallel-to-parallel transition of *Oxytricha* quadruplex $(T_4G_4)_4$ (35) containing no dA residues.

**Reliability of spectral differences revealed in Raman spectra**

Although spectral changes of several Raman bands of moderate and weak intensities may seem to be too small to be identified reliably, their statistical significance is assured by the mode of data acquisition and treatment used in the present work. Instead of visual inspection and identification of the differences in few intentionally selected Raman spectra obtained under particular physico-chemical conditions, multivariate SVD methods (44) are applied to extensive Raman datasets to take into consideration also inherent spectral variabilities caused by uncontrollable variations of experimental conditions (day-to-day fluctuations, different signal-to-noise ratio, uncertainties in the background- and solvent-correction, different batches used for sample preparation). Representative Raman spectra represent average spectra extracted from several measurements realized on different days under the identical experimental conditions, often by using oligonucleotides from different syntheses. By means of advanced multivariate SVD methods (44), inherent variabilities within datasets consisting of several Raman spectra belonging to the same particular structure or state, were evaluated for their statistical significance. Spectral differences between Raman spectra of two different structures were considered as significant merely in the case when corresponding spectral variability between two datasets was significantly greater (at least 5x) than the spectral variability within a dataset describing identical structure. Instead of comparing two intentionally selected spectra taken under different physico-chemical conditions, spectral differences were recognized by the multivariate SVD methods in the spectral series (consisting of dozens of spectra) obtained with gradually varied intensive variable (concentration, ionic strength, temperature, time). Continuous dependence of the spectral difference as a function of an intensive variable was regarded as a further sign of reliability and evidence of real existence of structural differences behind. For a presentation in a more common mode, Raman spectra corresponding to the edges of the interval of intensive variable under the study have been extracted and their difference used to better highlight spectral differences.

**Table S1.** Raman markers diagnostic of quadruplex structures

| Marker position [$cm^{-1}$] | Assignment | Diagnostic of telomeric fold or structure |
|---|---|---|
| 499 ± 2 | sc $PO_2^-$, dT, dG | stronger in ap-Q |
| 581 ± 1 | dG | stronger in p-Q |
| 611 ± 2 | $dT_4$ and d(TTA) loops | lateral or diagonal loops in ap-Q or hairpin |
| 671 ± 2 | C2'-*endo/syn*-dG | stronger shoulder in ap-Q |
| 686 ± 2 | C2'-*endo/anti*-dG | stronger in p-Q |
| 781 ± 1 | O-P-O, dG, dT, dA | stronger in p-Q |
| 837 ± 1 | O-P-O | marker of regular B-type bk |
| 1093 ± 1 | $PO_2^-$ | intensity standard |
| 1322 | C2'-*endo/syn*-dG | non-Q |
| 1326 ± 2 | C2'-*endo/syn*-dG | stronger in ap-Q |
| 1338 ± 2 | C2'-*endo/anti*-dG | stronger in p-Q |
| 1371 ± 1 | dT | stronger in ap-Q |
| 1483 ± 2 | dG-N7 strong Hoogsteen H-bond | ap-Q or p-Q; downshift and hypochromism in p-Q |
| ~1486 | dG-N7 weak H-bond to $H_2O$ | non-Q |
| ~1578 | dG-N2H H-bond to $H_2O$ | non-Q |
| 1581 ± 1 | dG-N2H interbase H-bond | ap-Q or p-Q |
| 1688 ± 2 | dG-O6 H-bond to $H_2O$ | non-Q |
| 1719 ± 3 | dG-O6 interbase H-bond | ap-Q or p-Q |

Data compiled from ref. (32, 34-36, 39, 46-51) and this work.

Abbreviations: ap-Q, antiparallel quadruplex; p-Q, parallel quadruplex; non-Q, non-quadruplex; bk, backbone; sc, scissoring vibration

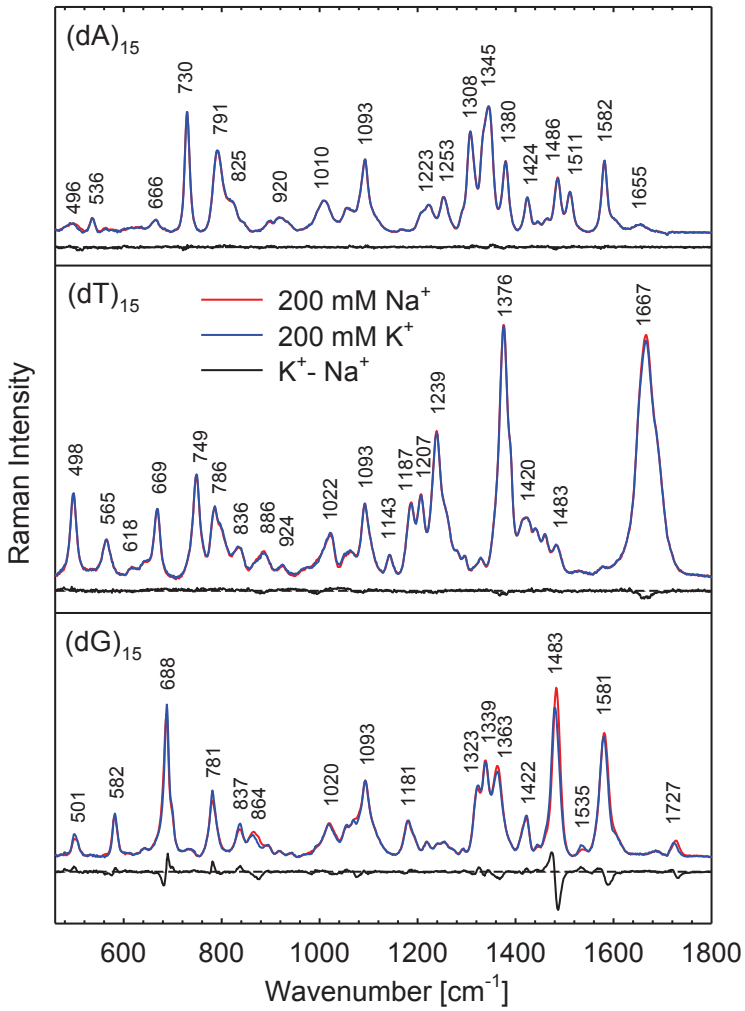



**Figure S1.** Normalized solvent- and background-corrected Raman spectra of $(dG)_{15}$, $(dT)_{15}$ and $(dA)_{15}$ in 30 mM PBS (pH 6.8) containing 200 mM of $Na^+$ or $K^+$ in total. Oligonucleotide concentration was 1.3 mM (in the strand, 20 mM in nucleoside) and the spectra were taken at 5°C. Difference spectra between $K^+$ and $Na^+$-solutions are shown to highlight cation-specific spectral variance. Under the conditions used, $(dA)_{15}$ and $(dT)_{15}$ adopt single-stranded B-DNA helices of similar structures in both alkali solutions, whereas $(dG)_{15}$ evidently forms four-stranded intermolecular parallel quadruplex exhibiting structural differences between $Na^+$- and $K^+$-stabilized forms.

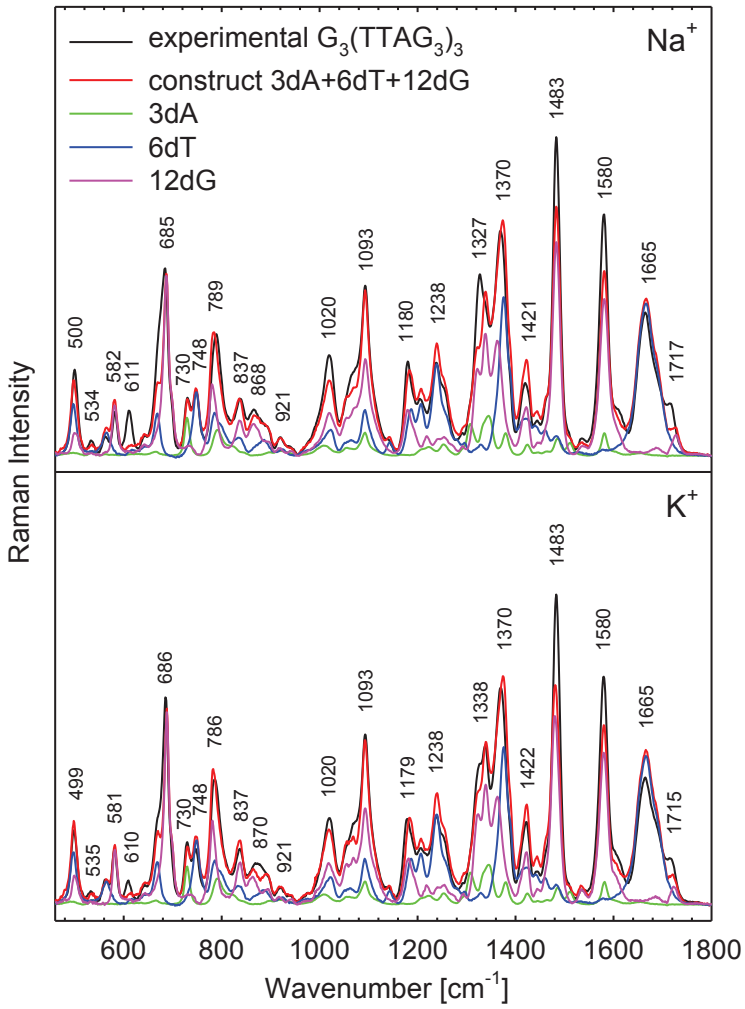



**Figure S2.** Comparison of experimental Raman spectra of $G_3(TTAG_3)_3$ in 500 mM $Na^+$ (top) or $K^+$ (bottom) and the synthetic spectra constructed as linear combinations of the Raman spectra of $(dG)_{15}$, $(dT)_{15}$ and $(dA)_{15}$ (in 200 mM $Na^+$ or $K^+$; Fig. S1). Proportional contributions from constituent nucleotides are shown to depict origin of individual Raman bands. Positions of the bands relate to the experimental spectrum.

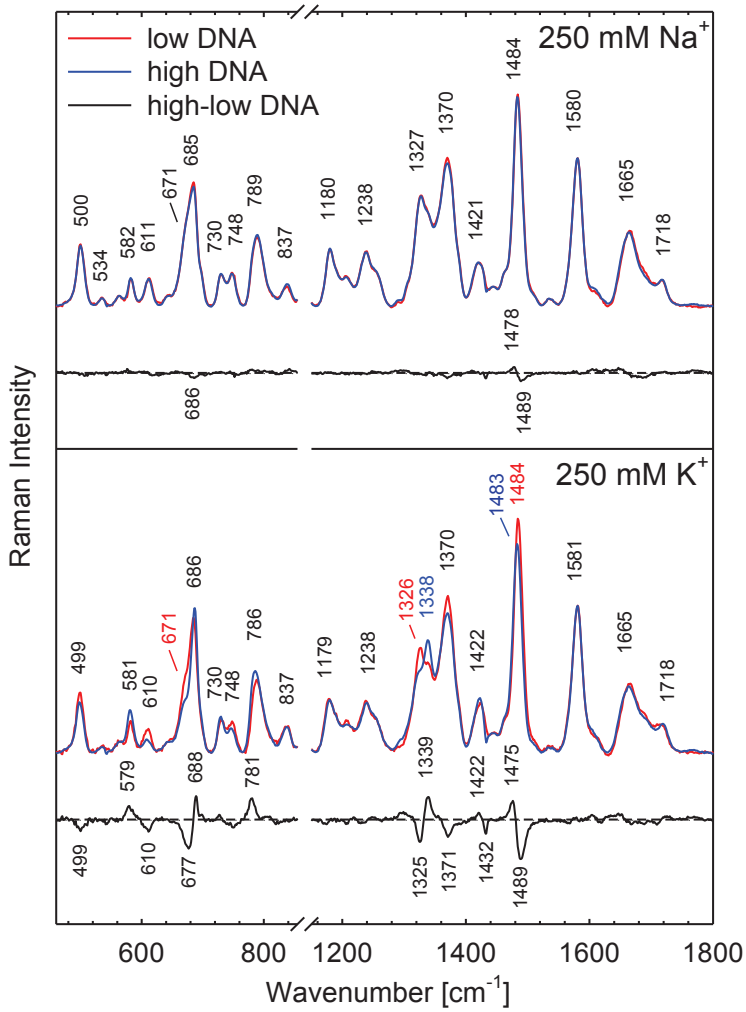



**Figure S3.** Cation-specific effect of high DNA concentration on the quadruplex structure. Normalized Raman spectra of $G_3(TTAG_3)_3$ at low (~8 mM) and high (~240 mM) DNA concentration in the presence of 250 mM of $Na^+$ (top) and $K^+$ (bottom) are compared in detail.

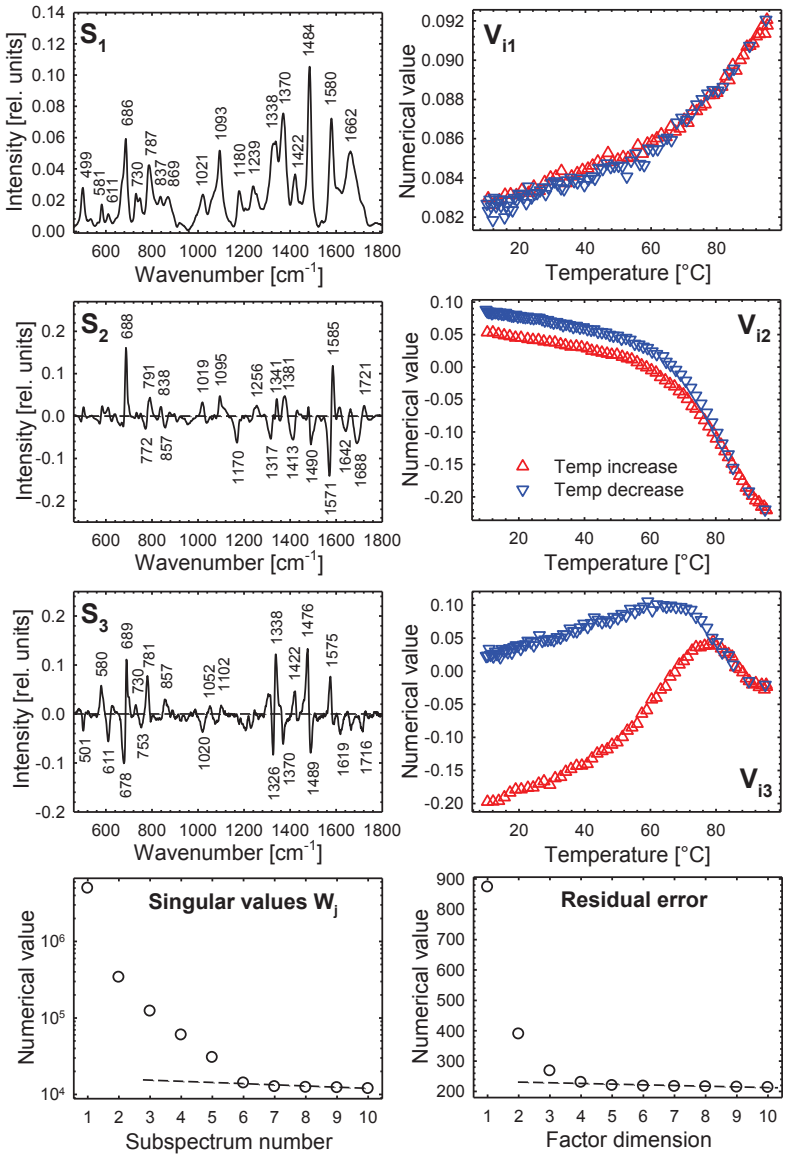



**Figure S4.** Factor analysis of the spectral changes observed in the course of annealing. $K^+$-stabilized (230 mM $K^+$) antiparallel quadruplex $G_3(TTAG_3)_3$ (330 mM in nucleosides) was heated continuously from 10 to 95°C and then cooled down. Rate of temperature increase/decrease was ca 1.5°C/min. Only the most important orthogonal spectral components describing differences related preferentially to overall denaturation of the quadruplex structures ($S_2$, $V_{i2}$) and to interquadruplex transition ($S_3$, $V_{i3}$) are shown for clarity.

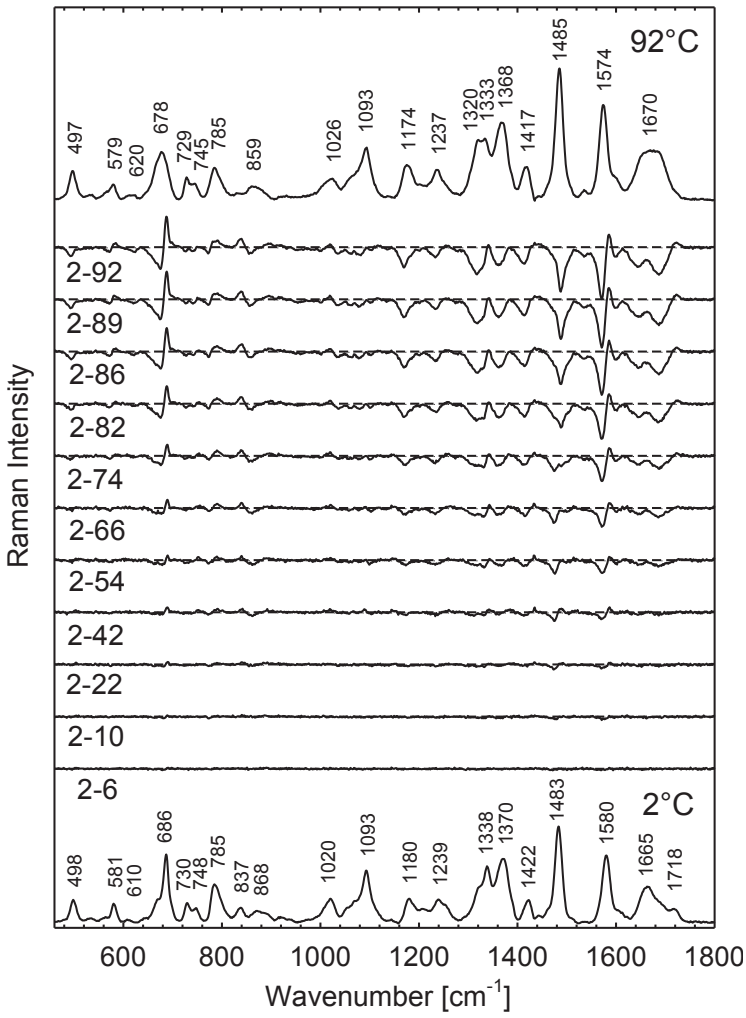



**Figure S5.** Spectral changes induced by thermal melting of preformed (by slow annealing) parallel $K^+$-stabilized quadruplex $G_3(TTAG_3)_3$. DNA concentration of 250 mM (in nucleoside), in 250 mM $K^+$ (30 mM PBS, pH 6.8). Intermediate traces show the differences between the spectrum of fully ordered quadruplex (2°C) and selected spectra taken at the temperatures indicated. Factor analysis of the full dataset comprising 46 temperature-equilibrated Raman spectra taken on the temperature increase and decrease is shown in Fig. S6.

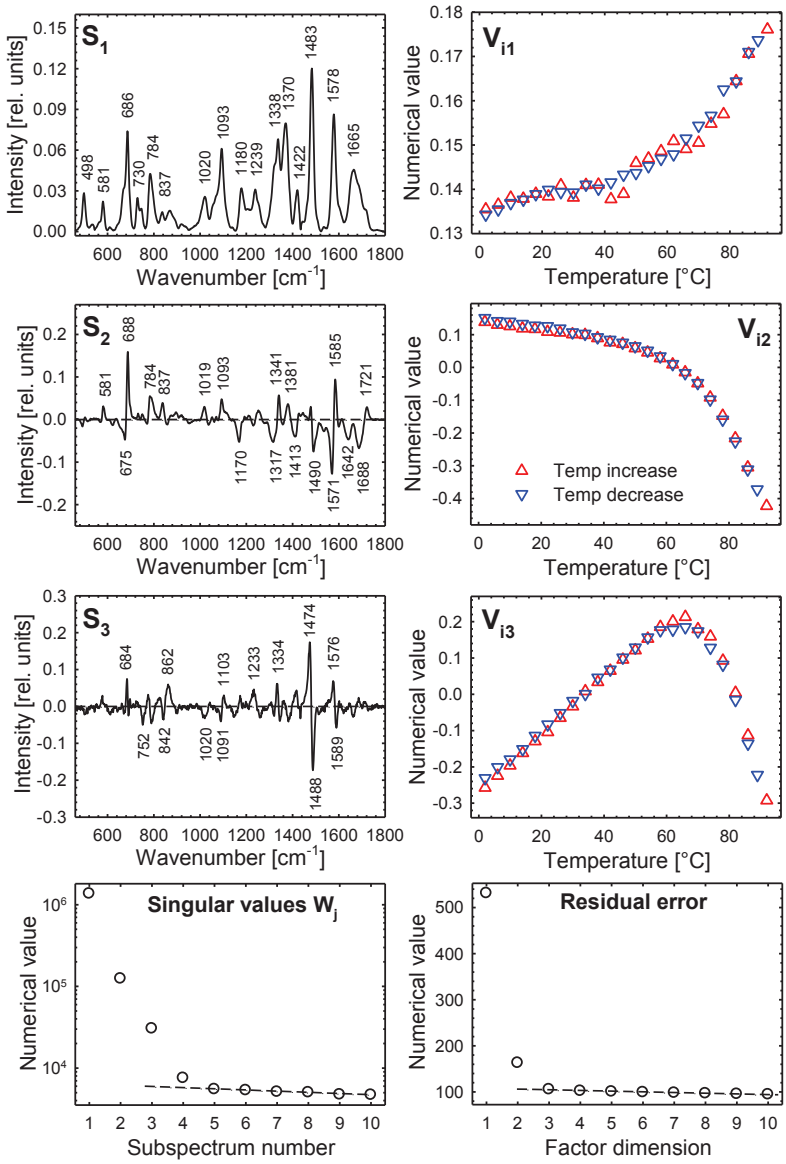



**Figure S6.** Factor analysis of the spectral changes induced by thermal melting of preformed (by slow annealing) $K^+$-stabilized parallel quadruplex $G_3(TTAG_3)_3$. DNA concentration of 250 mM (in nucleoside), in 250 mM $K^+$ (30 mM PBS, pH 6.8), temperature range 2 – 92°C. Analysis comprises dataset of 46 temperature-equilibrated Raman spectra taken on the temperature increase as well as decrease. Selected spectra are shown in Fig. S5 to better depict temperature-induced spectral changes. The third orthogonal spectral component $S_3$ resolved by SVD is needed to describe the specific temperature dependence of the 1483 $cm^{-1}$ band which differs from thermal behavior of the other bands.

# Raman Microspectroscopy of the Yeast Vacuoles

Spectroscopy: An International Journal 27 (2012) 503-507

Lucie Bednárová[1], Jan Palacký[2], Václava Bauerová[1], Olga Hrušková-Heidingsfeldová[1], Iva Pichová[1] and Peter Mojzeš[1]

[1] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo Náměstí 2, CZ-166 10 Prague 6, Czech Republic

[2] Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, CZ-121 16 Prague 2, Czech Republic

Correspondence should be addressed to Lucie Bednárová, bednarova@uochb.cas.cz

Hindawi Publishing Corporation
Spectroscopy: An International Journal
Volume 27 (2012), Issue 5-6, Pages 503–507
doi:10.1155/2012/746597

# Raman Microspectroscopy of the Yeast Vacuoles

Lucie Bednárová,[1] Jan Palacký,[2] Václava Bauerová,[1] Olga Hrušková-Heidingsfeldová,[1] Iva Pichová,[1] and Peter Mojzeš[2]

[1] *Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo Nám. 2, 166 10 Prague 6, Czech Republic*
[2] *Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, 121 16 Prague 2, Czech Republic*

Correspondence should be addressed to Lucie Bednárová, bednarova@uochb.cas.cz



**Abstract.** In the present work, real ability of a confocal Raman microspectroscopy to monitor chemical composition of the vacuoles within living yeast cells was investigated and critically assessed. Simple, economical, and practical protocols of the yeast immobilization suitable for less laborious, high-throughput, and spatially resolved Raman measurements were tested for their possible impacts on physiological states and viability of the cells. We have demonstrated that, acquiring Raman spectra from statistically sound sets of immobilized cells and employing advanced multivariate methods for spectral analysis, the chemical composition of the yeast vacuoles can be reliably studied. The most easily and accurately quantifiable seems to be the concentration of polyphosphates which can be unambiguously identified due to unmistakable Raman features. Our approach can be useful for routine, label-free, and noninvasive monitoring of the chemical composition of the vacuoles of living yeasts exposed to various stress factors, the information important in biomedical research of pathogens.

Keywords: Raman microspectroscopy, living cell, yeast, vacuole, chemical composition, polyphosphate, *Candida albicans*

## 1. Introduction

The yeast *Candida albicans* is an opportunistic human pathogen representing serious threat to immunocompromised individuals. It exhibits considerable metabolic flexibility, which enables the fungus to colonize host niches as diverse as skin, blood, oral cavity, or vaginal mucosa. *C. albicans* can survive under severe nutrition limitations and escape host immune defense [1]. Recent studies have shown significance of vacuoles for adaptation of *C. albicans* to these changing environments. Vacuoles serve as stores of amino acids, ions, and numerous metabolites. They are involved in pH homeostasis and osmoregulation, and they play a crucial role in digestion of proteins and recycling of nutrients [2, 3]. A *C. albicans* mutant that lacked recognizable vacuoles was not able to kill host macrophages [4]. Information about the chemical composition of the vacuoles within living cells exposed to various external factors could thus be of great importance for development of novel antifungal strategies.

In the last decade, confocal Raman microspectroscopy has been repeatedly shown as promising contactless and nondestructive method suitable for determination of intracellular chemical composition and chemical imaging of living cells [5]. However, attempts to employ Raman microspectroscopy for routine use in microbiology should cope with numerous technical and methodological problems, for example, adequate fixation assuring both immobility and viability of the cells during exposure to laser beam, compromise between requirements for high spatial resolution and inherent weakness of Raman signal, long acquisition times needed to improve signal-to-noise ratio, discrimination of spectral contributions from cultivation media and supporting materials, acquisition of statistically sound datasets and their treatment, as well as interpretation of complicated Raman spectra in the terms of concentrations of chemical constituents.

In the present work, applicability of the confocal Raman microspectroscopy for determination of chemical composition of the vacuoles within living yeasts was tested and critically assessed. Using both microscope slides and coverslips functionalized by poly-L-lysine, simplified and less laborious immobilization of the yeast cells was achieved. The mounts were found to be suitable for high-throughput, spatially resolved Raman measurements. As the immobilized yeasts immediately stopped their cell cycle due to the washing with water, they remained alive during several hours, and it was possible to acquire Raman spectra from vacuoles of numerous cells. Using advanced multivariate methods [6, 7], mean Raman spectra of the vacuoles throughout the yeast cultures were extracted and compared. Among all specific compounds localized in this compartment, vibrations of polyphosphate appeared to be the most easily and accurately quantifiable characteristic. We show that Raman microspectroscopy can be used in future experiments as a tool to analyze vacuoles of *Candida* yeasts cultivated under different nutrition conditions, exposed to various stress factors, and/or existing in different physiological states.

## 2. Materials and Methods

The *C. albicans* strain used for this work was clinical isolate HE 169 obtained from the mycological collection of the Faculty of Medicine, Palacky University, Olomouc, Czech Republic. The yeast cells were incubated at 37°C in a nutrient-rich YPD medium (2% peptone, 2% glucose, 1% yeast extract). Aliquots of the cell suspension were gently centrifuged (500 ×g for 2 min), washed two times with water, and finally resuspended into the fivefold less amount of water. Few microliters of the suspension were applied to a poly-L-lysine-coated slide, covered with a poly-L-lysine-coated coverslip, and sealed with a clear nail polish. When applying appropriate amount of the cell suspension and gently pressing the coverslip against the slide, combination of the spatial restriction with chemical fixation immobilized the cells in a single layer. The cells were thus in contact with the slide as well as with the coverslip. To eliminate autofluorescence of the glass substrate, fluorescence-free borofloat slides and coverslips (nexterion, schott) were used.

Raman spectra were taken on confocal Raman microspectrometer LabRam HR800 (Horiba) with a He-Ne laser excitation 632.8 nm (3–8 mW at the sample). To obtain well spatially-resolved Raman spectra restricted to the vacuole volume, the 100x oil-immersion objective (NA 1.40) was used in combination with relatively small confocal pinhole (200 $\mu$m) to eliminate Raman signal from the coverslip and the slide. Raman spectra in the range 500–3900 $cm^{-1}$ were collected using acquisition time of 60 s per a cell. Typically, Raman spectra of 50–80 different cells from the same mount were collected. Advanced background- and medium-correction procedures based on singular value decomposition (SVD) methods

were used for spectral treatment [7]. Statistical evaluation of datasets and visualization of their spectral variability were based on the factor scores provided by SVD analysis. Finally, the spectra were normalized using the band of water OH-stretching vibrations at $\sim$3400 $cm^{-1}$ as an internal intensity standard.

### 3. Results and Discussion

The main problem in Raman microspectroscopy and imaging of nonadherent living cells is their immobilization. The cells should be attached to the solid surface; however immobilization should not influence negatively their physiological status or spectral measurements. In the case of living cells it is also important that the immobilization procedure is fast and does not represent excessive stress. From that point of view, the protocols relying on the immobilization in polymer matrices (agar, agarose, pectin, gelatin) at elevated temperatures or use of high-viscosity media were found unsuitable for routine Raman experiments due to the risk of thermal shock or spectral contribution from immobilizing medium. As an alternative, we have tried to immobilize *Candida* at the slides modified by poly-L-lysine. The cells were either left in a cultivation medium or washed by a pure water, dropped on a slide, and observed directly by a water-immersion objective or covered with a coverslip and observed by an oil-immersion objective. However, immobilization by a contact with a single functionalized surface proved to be insufficient for Raman measurements. Even slow motion of the cell during the acquisition, moreover, often escalated by optical tweezers effect of the laser beam completely disrupted spatial resolution. Therefore, combination of the slides and coverslips both functionalized by collagen or poly-L-lysine was tested, the latter being proved to be more efficient. Besides better stabilization due to tight confinement between two functionalized surfaces, higher density of the fixed cells in the field of view and their arrangement into a single layer allowed semiautomated data collection throughout large set of individual cells.

As shown in Figure 1, several yeast intracellular structures can be readily distinguished in the video image; however some of them are of virtually identical shapes and shades, for example, lipidic droplet (E) and round protein-containing structures (C and D). Nevertheless, they can be clearly discerned due to different Raman spectra (Figure 1), without need of staining. Chemical composition of the intracellular structures can be furthermore analyzed in detail, although reliable spectral assignment and decomposition into the spectra of pure constituents could be difficult.

In the case of vacuoles, we have found that Raman spectra can be useful for their identification as well as for analysis of their chemical composition. To average cell-to-cell variability and to extract representative spectrum throughout the cell culture, Raman spectra from vacuoles of different cells have been acquired. Measurements at the same mount producing dozens of Raman spectra did not last more than 1.5 hours in total, the time corresponding to generation time of *Candida* under the conditions of cultivation. All data collected from the cells cultivated under the similar conditions and supposed to have similar physiological status were treated by the SVD procedure [7] to extract typical Raman spectrum (Figure 1, B), along with representative spectral variability. According to repeated experiments, usually only two orthogonal spectral components were disclosed in datasets, suggesting fair chemical uniformity of the cells in cultures. The cells differ especially in the relative concentration of the main components, not in their absence. The detailed spectral analysis revealed Raman features of amino acids, lipids, and polyphosphates, the principal constituents of *Candida* vacuoles. In particular the polyphosphates exhibiting typical Raman bands at $\sim$688 $cm^{-1}$ and $\sim$1154 $cm^{-1}$ were found in vacuoles at considerably high concentration. We suggest that polyphosphate band at 1154 $cm^{-1}$ can serve as a marker of the

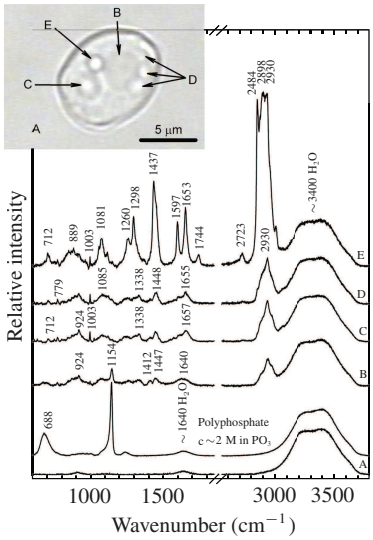



**Figure 1:** Intracellular structures apparent on the video image of *Candida* yeast and respective spatially-resolved Raman spectra of cultivation medium (A), vacuole (B), protein structure I (C), protein structures II (D), and lipidic structure (E). Spectra are normalized using intensity of the water OH-stretching band at 3400 $cm^{-1}$ as an intensity standard. Raman spectrum of 2 M aqueous solution of polyphosphate is shown to highlight its contribution to the spectrum of vacuole.

vacuoles since it was not detected in the Raman spectra of other *Candida* compartments. Though the cells coming from the same culture were—at least according to the video image—in a similar physiological status, the absolute concentration of the polyphosphates in their vacuoles was found to vary more than the concentration of other constituents. It was determined quantitatively by means of properly normalized intensity of the 1154 $cm^{-1}$ band with respect to the intensity of the OH-stretching band of water at 3400 $cm^{-1}$ (Figure 1). Actual concentration of the polyphosphates in the vacuoles of *C. albicans* cultivated under the standard conditions in YPD medium was found to vary from few dozens up to several hundreds of mM (expressed as a concentration of $PO_3^-$). The differences in the polyphosphate concentration could result from differences in the physiological status of the investigated cell. Raman microspectroscopy thus provides the information that cannot be obtained *in situ* from living cells by other methods.

## 4. Conclusions

In the present work, immobilization protocol suitable for Raman microspectroscopy of non-adherent cells was tested. It was demonstrated that with the properly immobilized *Candida* cells it is possible to collect spatially-resolved Raman spectra of individual compartments within living cells. The spectra can serve for unambiguous identification of intracellular compartments of similar appearance as well as for investigation of their chemical content. We report that the yeast vacuoles can be reliably

identified by the polyphosphate Raman bands at 688 and 1154 $cm^{-1}$, the spectral features missing in other yeast compartments. Using polyphosphate Raman bands normalized with respect to the water signal, actual polyphosphate concentration can be measured inside the vacuoles of living cells. Raman microspectroscopy will be furthermore employed for investigation of the *C. albicans* vacuoles and their chemical content affected by nutrition, stress, or phase of cell cycle conditions.

## Acknowledgments

Ministry of Education of the Czech Republic (MSM0021620835), the Grant Agency of the Czech Republic (P208/10/0376 and 310/09/1945), and Research Project Z 40550506 are acknowledged for financial support.

## References

[1] J. Kim and P. Sudbery, "Candida albicans, a major human fungal pathogen," *Journal of Microbiology*, vol. 49, no. 2, pp. 171–177, 2011.
[2] S. C. Li and P. M. Kane, "The yeast lysosome-like vacuole: endpoint and crossroads," *Biochimica et Biophysica Acta*, vol. 1793, no. 4, pp. 650–663, 2009.
[3] V. Veses, A. Richards, and N. A. Gow, "Vacuoles and fungal biology," *Current Opinion in Microbiology*, vol. 11, no. 6, pp. 503–510, 2008.
[4] G. E. Palmer, M. N. Kelly, and J. E. Sturtevant, "The Candida albicans vacuole is required for differentiation and efficient macrophage killing," *Eukaryotic Cell*, vol. 4, no. 10, pp. 1677–1686, 2005.
[5] M. Diem, M. Romeo, S. Boydston-White, M. Miljković, and C. Matthäus, "A decade of vibrational micro-spectroscopy of human cells and tissue (1994–2004)," *Analyst*, vol. 129, no. 10, pp. 880–885, 2004.
[6] E. R. Malinowski, *Factor Analysis in Chemistry*, Wiley, New York, NY, USA, 2002.
[7] J. Palacký, P. Mojzeš, and J. Bok, "SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations," *Journal of Raman Spectroscopy*, vol. 42, no. 7, pp. 1528–1539, 2011.

# SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations

Journal of Raman Spectroscopy 42 (2011) 1528-1539

Jan Palacký, Peter Mojzeš and Jiří Bok

Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, CZ-121 16 Prague 2, Czech Republic

Correspondence should be addressed to Peter Mojzeš, mojzes@karlov.mff.cuni.cz

Research Article

Journal of RAMAN SPECTROSCOPY

Received: 30 July 2010 Accepted: 21 December 2010 Published online in Wiley Online Library: 23 February 2011

(wileyonlinelibrary.com) DOI 10.1002/jrs.2896

# SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations

**Jan Palacký, Peter Mojzeš* and Jiří Bok**

**A semiautomated method combining intensity normalization with effective elimination of the solvent signal and non-Raman background is presented for Raman spectra of biochemical and biological analytes in aqueous solutions. The method is particularly suitable for rapid and effortless preprocessing of extensive datasets taken as a function of gradually varied physicochemical parameters, e.g. analyte and/or ligand concentration, temperature, pH, pressure, ionic strength, time, etc. For intensity normalization, the strong Raman OH stretching band of water in the range of 2700–3900 $cm^{-1}$ recorded together with the analyte spectrum in the fingerprint region below 1800 $cm^{-1}$ is employed as internal intensity standard. Concomitant dependences of the solvent Raman spectra are taken into account and, in some cases, turned into advantage. Once the Raman spectra of the solvent are acquired for a particular range of the parameter varied, solvent contribution can be subtracted correctly from any analyte spectrum taken within this range. The procedure presented can be efficiently applied only for the analytes having their own Raman signal in the range of OH stretching vibrations much weaker than that of the solvent. However, this is the case for a great number of biochemical and biological samples. Accuracy, reliability and robustness of the method were tested under the conditions of spontaneous Raman, resonance Raman and surface-enhanced Raman scattering. Serviceability of the method is demonstrated by several real-world examples.** 

*Supporting information may be found in the online version of this article.*

**Keywords:** Raman spectroscopy; intensity normalization; background correction; solvent subtraction; singular value decomposition

## Introduction

Introduction and dissemination of multivariate statistical methods in treatment of Raman spectra during past two decades has influenced experimental approaches, methodologies and requirements on specific parameters and quality of the spectral data. Instead of looking for and presenting obvious spectral differences between a few selected Raman spectra taken under definite conditions, more complex and less apparent spectral changes can be discerned, studied and interpreted as a function of a particular physicochemical parameter (intensive variable[1]), gradually varied over a given range. Thermal-induced structural transitions of biomolecules,[2,3] acid–base titrations[4,5] or kinetic studies[6,7] monitored by Raman spectroscopy can be taken as an archetype of particular but frequent Raman experiments when a biological analyte is dissolved in an aqueous buffer to rather low concentration, and spectral changes are studied as a function of temperature, pH and time, respectively. Instead of laborious visual inspection and manual identification of spectral changes, extensive series of Raman spectra can be analyzed by multivariate statistical methods[8] (factor analysis, principal component analysis). However, besides the signal from the analyte, the raw Raman spectra include also solvent contribution and non-Raman background, both usually evolving with the intensive variable. In addition, absolute intensity of the Raman signal can fluctuate in its own way, hindering or complicating proper use of multivariate methods. Hence, correct quantitative analysis of the analyte spectral changes requires preprocessing of the Raman datasets,[9] notably intensity normalization, subtraction of the solvent signal and correction for the non-Raman background.

In principle, internal or external standards can be used for intensity normalization of Raman spectra.[10] Both approaches have their advantages, disadvantages and limits of applicability. Concerning biological analytes dissolved in aqueous buffers, internal standards seem to be more suitable and more frequently used.[10] Two types of internal standards are chosen: either Raman bands of chemical species constituting an integral part of the sample, e.g. solvent peaks or particular Raman bands of the analyte itself,[2,11] or bands of a foreign compound added to the sample *ad hoc*, e.g. $ClO_4^-$, $NO_3^-$, $SO_4^{2-}$, $SeO_4^{2-}$ or $PO_4^{2-}$ ions.[11,12] In both cases, an ideal internal standard should provide intense, clearly resolvable Raman features not coinciding with other Raman bands and remaining unaffected by the physicochemical parameter varied over the en-

* *Correspondence to: Peter Mojzeš, Faculty of Mathematics and Physics, Institute of Physics, Charles University in Prague, Ke Karlovu 5, CZ-121 16 Prague 2, Czech Republic. E-mail: mojzes@karlov.mff.cuni.cz*

*Faculty of Mathematics and Physics, Institute of Physics, Charles University in Prague, CZ-121 16 Prague 2, Czech Republic*

Journal of RAMAN SPECTROSCOPY

tire range of interest. In addition, the internal standard should not be photochemically active or significantly absorb the excitation or Raman scattered light.[12] Furthermore, foreign internal standard added to the sample should not interact with the analyte or affect its microenvironment. It is by no means an easy task to find an ideal standard satisfying all the aforementioned requirements. However, proper choice of standard is crucial since it can strongly influence results of multivariate analysis of Raman spectra.[13]

Because in aqueous solutions of biomolecules the water molecules are naturally present at a molar concentration (55.6 M) by orders greater than that of the analyte, their Raman spectrum can be a good candidate for internal standard.[9] Water molecules provide a broad, weak band centered at 1640 $cm^{-1}$ and a broad, strong band centered at 3400 $cm^{-1}$. Despite its weakness, it is a common practice to use as a standard the 1640 $cm^{-1}$ band belonging to deformation vibration,[4,9,11] as its position, intensity and line shape are less sensitive to ionic strength and temperature (or anything else that affects hydrogen bonding) than those of the stronger 3400 $cm^{-1}$ band. However, this approach provides less satisfactory results for analytes having their own Raman signal between the 1550 and 1700 $cm^{-1}$. For example, nucleic acids and proteins contribute to that region considerably by their relatively broad carbonyl and amide I bands, respectively, complicating correct differentiation between the solvent and the analyte signal. Intensity normalization and proper subtraction of the solvent contribution from the 1550 to 1700 $cm^{-1}$ region is an important task for many biomolecular studies relying on the Raman information overlapped by the 1640 $cm^{-1}$ water band.

Normalization to the intense Raman band belonging to the water OH stretches is less common because of its sensitivity to water hydrogen bonding. On the other hand, in the range 2700 to 3900 $cm^{-1}$, there is no risk of considerable interference with the Raman bands of biological analytes, notably due to their lower intensities, downshifted positions and markedly narrower line shapes. Therefore, presupposing detail knowledge of the solvent spectrum dependence on the intensive variable, the OH stretching band can be used as an internal standard and the solvent contribution properly subtracted. In the present work, a method based on factor analysis and combining intensity normalization to the OH stretching band with effective elimination of the solvent signal is proposed for semiautomated preprocessing of extensive Raman datasets.

## Experimental

Cacodylic acid, LiOH, KOH, NaOH, LiCl, KCl and NaCl (analytical grade, Sigma–Aldrich) and deionized water (18 mΩ, Elga) were used for preparation of cacodylic aqueous buffers (pH ~6.8). The effect of the ionic strength was studied on the $Li^+$-, $K^+$- and $Na^+$-cacodylic buffers with addition of respective chlorides. In order to evaluate the effect of pH on the Raman spectrum of water, appropriate volumes of HCl, LiOH, KOH and NaOH solutions were added to deionized water and the resulting pH was determined by a pH meter.

The method was tested using cacodylic buffer (100 mM, pH ~6.8), where the cacodylate ions themselves constitute the analyte with a simple Raman spectrum relatively insensitive to the intensive variables studied (temperature and ionic strength). Furthermore, two biomolecules dissolved in the $Na^+$-cacodylic buffer (20 mM, pH ~6.8, $\mu \sim 100$ mM) to form a concentration series were used as representative analytes. 5′-Uridine monophosphate (5′-UMP, Sigma–Aldrich) and Cu(II)-5,10,15,20-*tetrakis*(1-methyl-4-pyridyl) porphyrin (CuTMPyP4, Frontier Scientific) were selected because of their low propensity to form self-aggregates in aqueous solutions of low ionic strength ($\mu \sim 100$ M).[14–17] CuTMPyP4 was chosen as an illustrative case of an absorbing analyte since its resonance Raman spectra can be comfortably excited at the onset of the Soret band ($\lambda_{max} = 424$ nm)[14] by the 441.6 nm line of an He–Cd laser. Concentration series (5′-UMP: $6.0 \times 10^{-2}$ to $5.7 \times 10^{-1}$ M, CuTMPyP4: $9 \times 10^{-6}$ to $3.9 \times 10^{-5}$ M) were prepared by diluting the respective stock solutions with the buffer. Actual concentrations were checked by a Lambda 12 UV–vis spectrophotometer (Perkin–Elmer) using extinction coefficients $\varepsilon_{262} = 9.78 \times 10^3$ $M^{-1}$ $cm^{-1}$ and $\varepsilon_{424} = 2.31 \times 10^5$ $M^{-1}$ $cm^{-1}$ for 5′-UMP[18] and CuTMPyP4,[14] respectively. Both analytes were found to obey Beer's law over the respective concentration ranges.

Applicability of the method to extensive SERS datasets was tested by using two SERS-active systems based on a borohydride-reduced Ag colloid[19,20] and CuTMPyP4 as an analyte (total concentration of $1 \times 10^{-8}$ M). The systems differed in the concentration of colloidal particles, one being 5 times more concentrated than the other one, and exhibiting different dependences of their SERS intensities on exposure time. SERS datasets were acquired at constant temperature (20 °C) as a temporal series of 150 spectra with short accumulation times (2 s).

All Raman spectra presented here were excited with the 441.6 nm (~10 mW at the sample, polarized He–Cd laser, Liconix) and collected in the 90° scattering geometry using a multichannel Raman spectrograph (Jobin Yvon–Spex 270 M) equipped with a holographic notch-plus filter (Kaiser) to reject Rayleigh scattering and a liquid nitrogen-cooled charge-coupled device (CCD) detector (1340 pixels in spectral axis, Princeton Instruments). Polarization geometry X(Z,X+Z)Y[1,21] was used to obtain all-polarization spectra. A scrambler was placed in front of the slit to compensate for the polarization-dependent transmissivity of the spectrograph. The combination of a grating with lower spectral dispersion and blue excitation permitted us to place at the CCD detector fingerprint parts of the analyte spectra (starting from ~650 $cm^{-1}$) concurrently with the 3400 $cm^{-1}$ band of water. All measurements included in the common treatment were taken at fixed parameters of the instrument (grating position, slit width) to ensure a constant response function. To minimize absorption of the laser intensity and Raman scattering in absorbing samples, the beam was placed close to the exit window of the cell.

In the case of temperature-dependent Raman spectra, measurements were carried out in temperature-controlled quartz cells of various volumes (5–100 μl). Spectra were taken as a function of temperature in the range 1–95 °C, with an accuracy of ±0.5 °C. Before measuring a spectrum, the sample was stabilized for 5 min at a given temperature. Reversibility of the temperature-induced changes was tested by changing the temperatures in a reverse order. The wavenumber scales of the Raman spectra were precisely calibrated using the emission spectra of a neon glow lamp taken before and after each Raman measurement. The estimated precision of the calibration procedure was better than 0.1 $cm^{-1}$.

All mathematical processing was performed using the codes written in Matlab, version 7 (The MathWorks).

## Results

To illustrate our procedure step by step, 37 Raman spectra of 100 mM $Na^+$-cacodylate (pH ~6.8) taken in the temperature range 2–90 °C (19 and 18 spectra on temperature increase and decrease, respectively) will be used as an analyte dataset (Fig. 1).



Journal of RAMAN SPECTROSCOPY



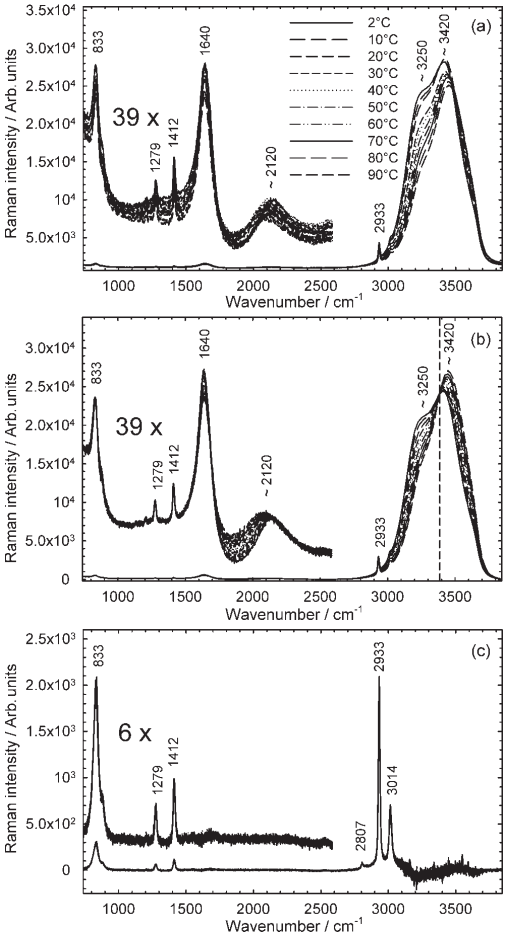



**Figure 1.** Temperature-dependent Raman spectra of 100 mM $Na^+$-cacodylic buffer. Only selected raw (a), background-corrected and SVD-normalized (b) and solvent-subtracted (c) spectra are shown for clarity. It is worth noting that the SVD-normalized water stretching bands intersect in an isosbestic point at 3387 $cm^{-1}$ (b).

Temperature series of 37 spectra of deionized water taken in a similar way represents corresponding solvent reference dataset (SRD; Fig. S1, Supporting Information). Some treatment steps will be demonstrated using dataset consisting of 101 (28, 45 and 28) temperature-dependent Raman spectra of 20 mM cacodylate buffer containing 100 mM total concentration of $Li^+$, $Na^+$ and $K^+$ ions, respectively. The same dataset will be used as SRD for concentration series of 5′-UMP and CuTMPyP4.

### Factor analysis

Our approach of non-Raman baseline correction, normalization and subtraction of the solvent contribution is based on factor analysis (singular value decomposition (SVD)),[8,22] a mathematical procedure projecting experimental spectra from dataset $[Y_i(\nu)]$ into orthonormal set of functions $[S_j(\nu)]$ (denoted hereafter as subspectra) according to Eqn (1), where singular values $W_j$ are statistical weights of $S_j(\nu)$ and elements $V_{ij}$ (denoted hereafter as coefficients) represent relative spectral contributions of $S_j(\nu)$ to the spectrum $Y_i(\nu)$. Typical results of SVD are shown in Figs 2 and 3. The minimum number of subspectra that have to be linearly combined (Eqn 1) so that the original spectral information is retained within the noise level determines the dimensionality of data (factor dimension). The factor dimension can be guessed from the plot of the singular values $W_j$ *versus* their corresponding order $j$ (Fig. 3). Small singular values $W_j$ (usually below 0.5% of the maximal value) that decrease linearly with the order $j$ are related to subspectra describing only white noise.[8,22]

$$Y_i(\nu) = \sum_j W_j V_{ij} S_j(\nu) \quad (1)$$

All the operations concerning non-Raman baseline correction, normalization and subtraction of the solvent from individual spectra $Y_i(\nu)$ are afterwards realized in the space of subspectra $S_j(\nu)$, coefficients $V_{ij}$ and singular values $W_j$. Finally, the all-corrected and normalized spectra are obtained as linear combinations of the adequately corrected subspectra.

The following requirements must be met so that Raman spectra can be normalized and corrected for the solvent contribution by using the properties of water valence vibrations: (1) the analyte dataset $[Y_i(\nu)]$ must be recorded simultaneously in the fingerprint region along with the region of water stretches (2700–3900 $cm^{-1}$); (2) the water stretching region must not be significantly affected either by the analyte signal or by strong fluorescence; (3) SRD $[Y_i^s(\nu)]$ recorded on the same apparatus and under equivalent experimental conditions (notably spectral range and resolution) must be available for the spectral and intensive variable ranges containing as their subsets respective ranges of the analyte dataset $[Y_i(\nu)]$.

In the practical implementation, the Matlab function *svd* (based on the LAPACK library) was used here to perform the SVD calculations. Our approach requires that all the spectra (analyte and SRD) be SVD-processed jointly as a single dataset (Fig. 2). For that reason, their wavenumber scales have to be properly calibrated and unified in their limits, number of points and spacing. Using codes written in Matlab, even datasets consisting of thousands of spectra with hundreds of spectral points can be processed easily and in a user-friendly environment.

### Baseline correction

Various approaches to baseline removal have been reviewed and compared thoroughly.[23,24] These include manual baseline correction,[25] baseline fitting,[26] derivative methods[27] or data filtering (Fourier and wavelet transforms).[27,28] Nevertheless, methods based on factor analysis, principal component analysis or SVD seem to be rare.[23] Our SVD-based method is similar to the procedure described by Glenn *et al*.[29] However, there are some differences in the approximation of the background shapes within orthogonal subspectra. Our approach consists in identification of the non-Raman backgrounds in the subspectra $S_j(\nu)$ and their approximation by polynomials or cubic splines $P_j(\nu)$ [Eqn (2)]. For instance, polynomials shown in Fig. 2 were fitted to the series of points selected manually as user's best estimates[25]; nevertheless more sophisticated methods can be employed for that purpose.[26] Afterwards, the baseline-corrected spectra $Y_i'(\nu)$ are constructed



Journal of RAMAN SPECTROSCOPY

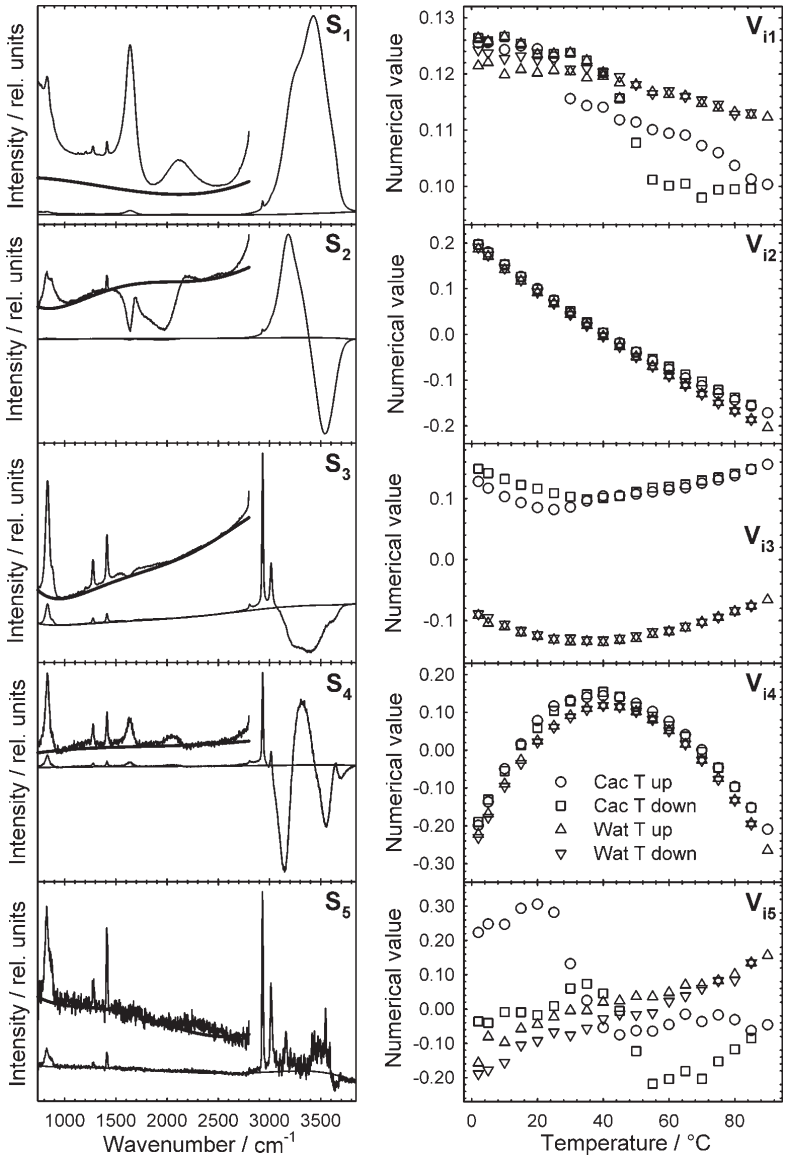



**Figure 2.** Subspectra $S_j$ and coefficients $V_{ij}$ from factor analysis of the dataset consisting of temperature-dependent Raman spectra of 100 mM $Na^+$-cacodylate (Cac) and spectra of deionized water (Wat) as their respective SRD. Smooth curves approximating non-Raman backgrounds in the subspectra were obtained as polynomial fits to the series of points selected manually as user's best estimates. Fingerprint regions with expanded intensity scale are shown for clarity.

according to Eqn (3), using singular values $W_j$ and coefficients $V_{ij}$ from the SVD of raw spectra. The main advantage of this approach consists in its simplicity and applicability for large spectral series, since SVD concentrates variability of the non-Raman background in a few subspectra with greatest singular values. Depending on the complexity of Raman dataset (expressed by its factor dimension), the non-Raman background can be eliminated from large datasets *en bloc*, estimating its shape by smooth, noiseless curves only in a few relevant subspectra. The baseline-corrected spectra obtained by this way retain complete spectral information except for the non-Raman background. Moreover, if the subspectra with small statistical weights (contributing only to the white noise)



Journal of RAMAN SPECTROSCOPY



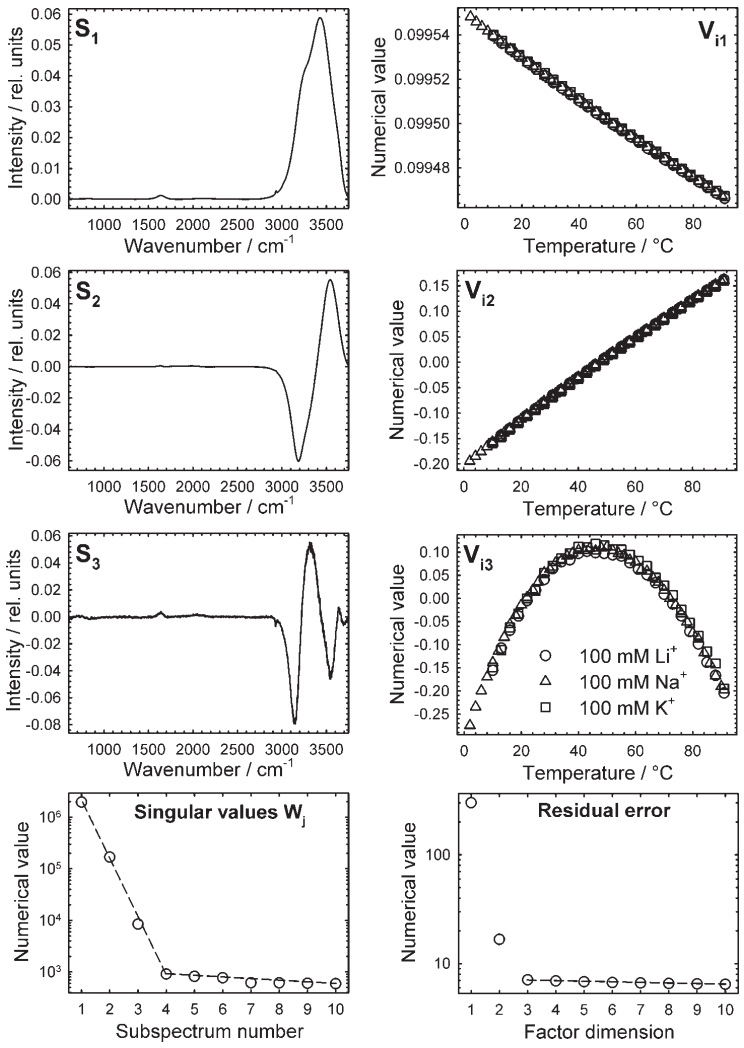



**Figure 3.** The SVD analysis of the 101 (28, 45, 28) background-corrected and SVD-normalized spectra of 20 mM cacodylate buffer adjusted to ionic strength $\mu = 100$ mM by addition of LiCl, NaCl and KCl, respectively. As evident from the sharp break of linear dependences of singular values $W_j$ with subspectrum number $j$, only first three subspectra are of real significance. Factor dimension of the three can be deduced also from the plot of residual error.

are not taken into linear combination, the background- and noise-corrected spectra are obtained at once.

$$Y_i(\nu) = \sum_j W_j V_{ij} S_j(\nu) = \sum_j W_j V_{ij} [S'_j(\nu) + P_j(\nu)] \quad (2)$$

$$Y'_i(\nu) = \sum_j W_j V_{ij} [S_j(\nu) - P_j(\nu)] = \sum_j W_j V_{ij} S'_j(\nu) \quad (3)$$

**Spectra normalization and solvent subtraction**

Before normalization, joint SVD-based baseline correction of both analyte $[Y_i(\nu)]$ and SRD $[Y_i^s(\nu)]$ datasets should be performed to compensate for slight non-Raman baseline variations, notably in the range of water stretches. The procedure is illustrated in Fig. 2, where results of joint SVD analysis of analyte dataset and SRD are shown along with polynomial estimates of non-Raman backgrounds in the first five subspectra. Afterwards, the background-corrected $[Y'_i(\nu)]$ and $[Y_i^{s\prime}(\nu)]$ datasets are reconstructed according to Eqn (3) and their 2700–3900 $cm^{-1}$ regions are SVD-analyzed again. Full-range background-corrected datasets are subsequently normalized (SVD-normalized) according to the first coefficients $V_{i1}$ obtained from SVD analysis restricted to the water stretching band. Except for intensive variables that evidently do not alter



Journal of RAMAN SPECTROSCOPY

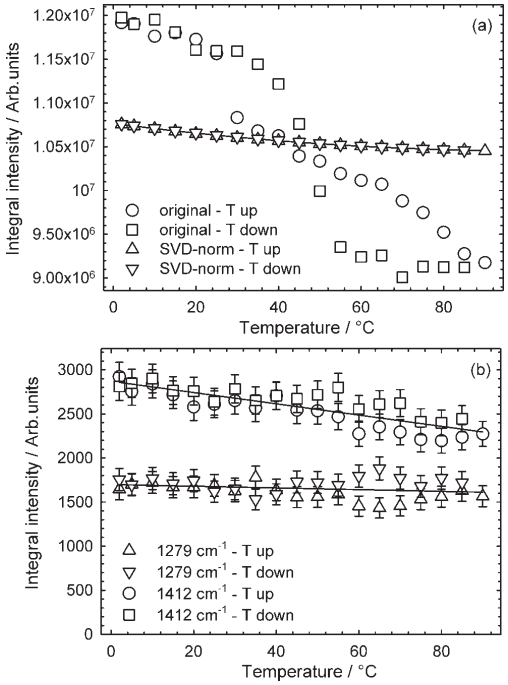



**Figure 4.** Integral intensities of the water stretching band (a) determined from background-corrected spectra of 100 mM $Na^+$-cacodylate prior to (original) and after SVD-normalization (SVD-norm). Integral intensities of the cacodylate bands (b) are from the same SVD-normalized dataset. Intensities are shown separately on the temperature increase (*T* up) and decrease (*T* down).

the shape of the water band (e.g. analyte/ligand concentration, time evolution), normalization to the coefficients $V_{i1}$ cannot be simply rationalized by a universally accepted presupposition on the properties of Raman band of water stretches. As evident from Fig. 1b, background-corrected and SVD-normalized spectra intersect in an isosbestic point at 3387 $cm^{-1}$ and the integral intensity of the water band calculated from the SVD-normalized spectra exhibits smooth and monotonic dependence on the temperature (Fig. 4a). Such SVD-normalization removes stochastic fluctuations, replacing them probably with systematic deformation of the analyte intensities that copies dependence of the water band with the intensive variable. The deformation can be subsequently corrected by further renormalization, using theoretically predicted or experimentally determined dependence of the water properties with the intensive variable. Evidently, normalization based on the water stretching band is reasonable only for analytes having negligible signal in the 2700–3900 $cm^{-1}$ range.

Once the datasets are SVD-normalized (Fig. 1b), subtraction of the solvent contribution can be accomplished. The subtraction procedure consists of three steps. At first, SVD analysis of the background-corrected and SVD-normalized SRD [$Y_i^{s'}(\nu)$] is carried out again for the full spectral range. According to our experience with several commonly used aqueous buffers and frequently varied physicochemical parameters (temperature, ionic strength, pH, time), factor dimension of the background-corrected SRD is rarely greater than three (Fig. 3). Consequently, all spectral changes due to the intensive variable can be expressed as a linear combination of three subspectra. By fitting the profiles of the corresponding coefficients $V_{ij}$ with polynomials, solvent spectrum for any variable value within the range can be easily constructed. As illustrated by temperature-induced changes in 20 mM cacodylic buffer (Fig. 3), coefficients $V_{ij}$ are at the most quadratic in temperature. The solvent spectrum $Y_T^{s'}(\nu)$ at any arbitrary temperature *T* can then be constructed according to Eqn (4), where $S_j^{s'}(\nu)$ are the solvent subspectra with the corresponding weights $W_j$, and $a_j$, $b_j$ and $c_j$ are the coefficients obtained from quadratic fits of the $V_{ij}$:

$$Y_T^{s'}(\nu) = \sum_{j=1}^{3}(c_jT^2 + b_jT + a_j)\, W_jS_j^{s'}(\nu) \qquad (4)$$

The last step in the algorithm consists in subtraction of the proper solvent spectra $Y_T^{s'}(\nu)$ from the spectra $Y_i'(\nu)$. Optimal solvent elimination is achieved by minimizing the least-square difference (minimization restricted to the range of water stretching band) between the $Y_T^{s'}(\nu)$ constructed according to Eqn (4) and the background-corrected spectrum $Y_i'(\nu)$. The only parameter optimized is the temperature *T* since other parameters in Eqn (4) are invariant for a particular SRD. The temperature obtained is then used to construct the appropriate solvent spectrum in the whole spectral range. Excellent agreement between temperatures determined by direct measurement and those obtained from the least-square procedure (Fig. S3) can be used as a supporting argument in favor of reliability of our solvent subtraction method. As seen from Fig. 1c, along with water contribution in the 2700–3900 $cm^{-1}$ range, other water features (bands at 1640 and 2120 $cm^{-1}$) are suppressed quite effectively. Temperature-dependent integral intensities of selected cacodylate Raman bands are shown in Figs 4b and S2.

### Concentration series of biological analytes

To demonstrate accuracy, reliability and robustness of the method under the conditions of nonresonant and resonant excitation, Raman datasets of aqueous solutions of 5′-UMP (Fig. 5a) and CuTMPyP4 (Fig. 6a) have been used as representatives of nonabsorbing and absorbing analytes, respectively. Both analytes were excited by the same 441.6 nm line since by using this short-wavelength excitation in combination with spectral dispersion of the spectrograph at our disposal, sufficiently broad spectral range (640–3740 $cm^{-1}$) was fitted onto the CCD detector to cover fingerprint and hydrogen-stretching regions. Although both datasets were acquired at constant temperature, the full set of the temperature-dependent solvent spectra (20 mM $Na^+$-cacodylate, $\mu \sim 100$ mM; Fig. 3) was used for the procedure as SRD, fitting the temperature of each sample as an unknown parameter. In both cases, the fitted temperatures were found to lie within the ±0.5 °C interval around the experimental values.

As evident from final SVD-normalized and all-corrected spectra (Figs 5b and 6b), the solvent contribution was suppressed excellently throughout the spectral range for both analytes. Even though factor analysis disclosed two orthogonal components in the 5′-UMP and CuTMPyP4 final datasets (Figs S4 and S5, respectively), their singular values differ significantly ($W_1/W_2 \sim 66$ and 113, respectively) and spectral features of $S_2(\nu)$ seem to reflect only minute uncertainties in the wavenumber scales. It means that, by neglecting $S_2(\nu)$, the Raman spectra of both analytes can be



Journal of RAMAN SPECTROSCOPY



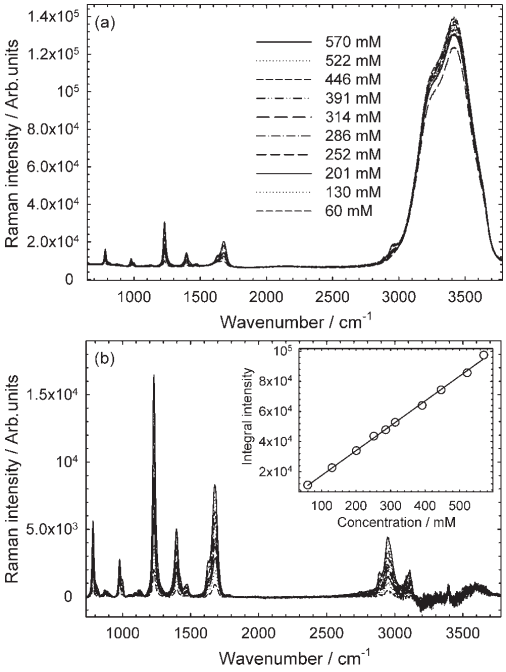



**Figure 5.** Raw (a) and SVD-normalized, all-corrected (b) Raman spectra of the 5′-UMP concentration series. Concentration dependence of the total Raman intensity of 5′-UMP in the fingerprint region is shown as an inset.

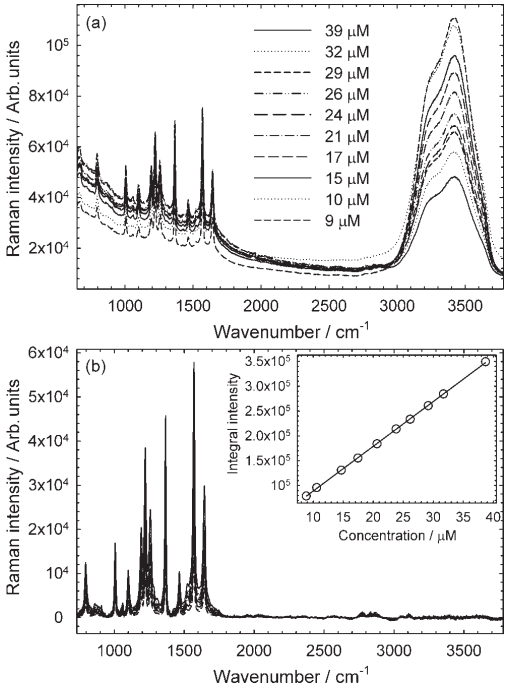



**Figure 6.** Raw (a) and SVD-normalized, all-corrected (b) Raman spectra of the CuTMPyP4 concentration series. Concentration dependence of the total Raman intensity of CuTMPyP4 in the fingerprint region is shown as an inset.

approximated (throughout the concentration ranges used here) simply by their first subspectra $S_1(\nu)$. In that case, dependence of the SVD-normalized Raman intensity as a function of the concentration can be expressed using absolute values of the coefficients $V_{i1}$. As seen in Figs S4 and S5, those dependences are fairly linear for both analytes, indicating reliability of the method used for intensity normalization.

### Solvents with various ionic strength and pH

Further intensive variables frequently used in biological studies are ionic strength and pH. It is known that the presence of electrolytes, particularly in high concentrations, strongly affects the shape of water stretching band.[30–37] Similar to temperature measurements, it was difficult to find a biological analyte exhibiting no spectral dependence on the ionic strength. Finally, 100 mM cacodylate ions in solutions with ionic strength adjusted by LiCl, NaCl and KCl were used to demonstrate that even in that case our SVD approach provides reliable results. As shown in Fig. 7b, water stretching band of background-corrected and SVD-normalized spectra intersect in an isosbestic point at 3374 $cm^{-1}$. Solvent-subtracted spectra (Fig. 7c) evidence nice suppression of the water features throughout the spectral range. Details of the cacodylate spectral changes induced by increasing ionic strength are seen in Figs 8 and S6–S9.

Concerning aqueous solvents of various pH, their water vibrations should be affected by the presence of $H_3O^+$ and $OH^-$ ions as well as by counter-ions from acids, bases and salts used for pH adjustment.[36] However, since in Raman studies of biomolecules the pH usually ranges between 2 and 12, concentrations of $H_3O^+$ and $OH^-$ ions are too low to affect significantly the shape of the water band. As seen in Fig. 9, background-corrected and SVD-normalized Raman spectra of water with pH adjusted by HCl and NaOH between 1.3 and 11.1 are not distinguishable from the Raman spectrum of pure water. It means that, at least in limited ranges, pH itself is an intensive variable with no appreciable effect on the water stretches.[37] It does not mean, however, that buffering agents used to maintain pH cannot have their own effect on the shape of the water band, especially when used at higher concentrations.

### Quantitative analysis of SERS spectra from silver hydrosols

The SVD normalization based on the water stretching band may be helpful for quantitative comparison of the SERS intensities from hydrosols.[19,20] As demonstrated in Fig. 10, our procedure can be applied routinely to strongly opalescent colloids to obtain all-corrected and properly normalized SERS spectra. Owing to SVD normalization and solvent subtraction, it was possible to compare quantitatively exposure-time dependences of the SERS intensities of CuTMPyP4 from two SERS-active systems. Though the temporal series of the SERS spectra were measured at fixed and known temperatures, temperature-dependent spectra of deionized water (Fig. S1) were used as their respective SRD to obtain optimal solvent subtraction. As seen in Fig. S10 (Supporting Information),



Journal of RAMAN SPECTROSCOPY

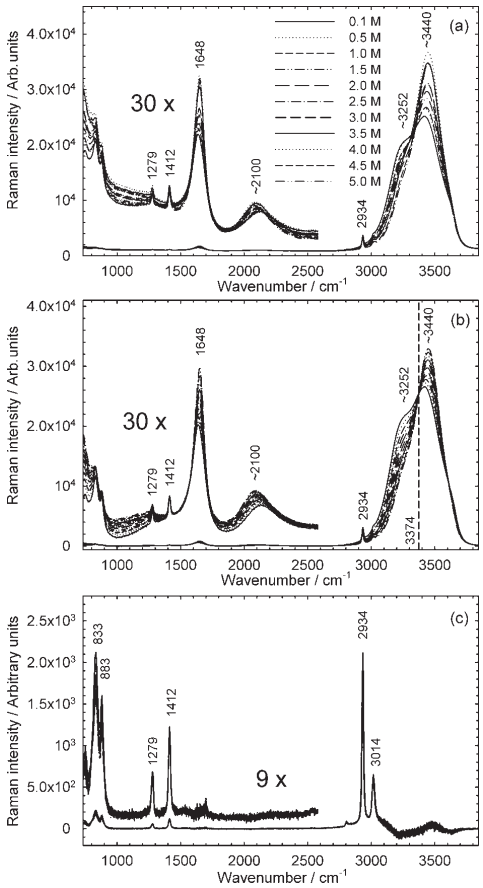



**Figure 7.** Background-corrected (a), SVD-normalized (b) and solvent-subtracted (c) Raman spectra of 100 mM $Na^+$-cacodylate as function of ionic strength adjusted by addition of NaCl. Clear isosbestic point at 3374 $cm^{-1}$ is indicated.

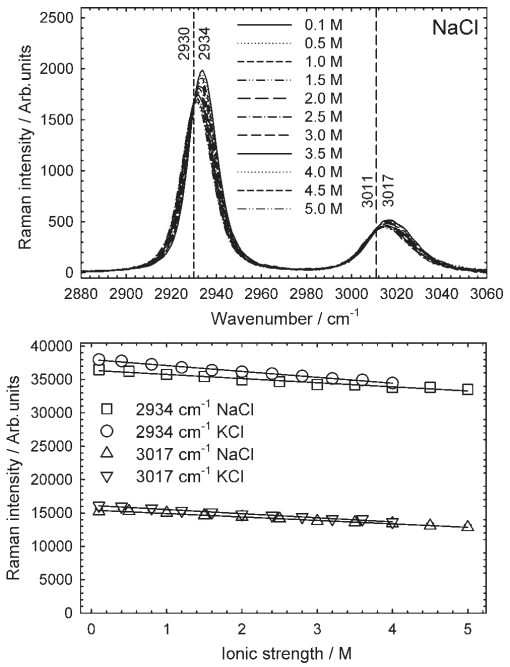



**Figure 8.** Detail view of spectral changes of the 2934 and 3017 $cm^{-1}$ cacodylate bands induced by increasing NaCl concentration (data from Fig. 7). Both bands are clearly downshifted on the ionic strength increase. SVD-normalized and all-corrected spectra intersect in two isosbestic points located at 2930 and 3011 $cm^{-1}$. Ionic-strength dependences of integral intensities of 100 mM $Na^+$-cacodylate are presented along with the intensities of 95 mM $K^+$-cacodylate to demonstrate the sensitivity and reliability of the normalization procedure. Linear fits to experimental points are shown to highlight intensity differences that correspond well to tiny difference in cacodylate concentrations.

the exposure-time dependence of the SERS-active systems A can be satisfactorily approximated by a single-exponential rise to maximum, disclosing some 17% intensity increase from its initial value with a time constant of 133 s. On the contrary, the time evolution of the SERS-active systems B have to be fitted as double-exponential intensity decay to 47% of its initial value with time constants $\tau_1 = 11$ s and $\tau_2 = 115$ s. Although detailed explanation of such different exposure-time dependences is out of the scope of the present work, it would be worthwhile to indicate that it probably relates to different structure of the SERS-active sites formed in the respective colloidal systems and concomitant differences in the light-assisted adsorption/desorption of the analyte.

## Discussion

Intensity normalization with respect to water Raman signal eliminates the necessity to collect absolute Raman intensities for quantitative analysis,[1,38] which is an uphill job for real-world samples, especially when using real-world experimental setups. As intensity of the analyte signal is related to that of the solvent, normalization is independent of hardly controllable parameters such as stability of excitation intensity, fixed collection geometry, reproducible detection efficiency of the entire setup, constant scattering volume, absorption effects, or changes of the sample density and refractive index with the intensive variable. The most fundamental question of the applicability of the water stretching band as Raman intensity standard can thus be narrowed to the existence of some intensity invariants with respect to the intensive variables. Such serviceable invariants could be intensity of the band at maximum, integral intensity or intensity at isosbestic point within the water band.

Water is a highly structured liquid with an extensive network of hydrogen bonds. Broad and highly asymmetric profile of the stretching band reflects the intricate structure of the network and associated dynamics that affects the vibrational wavenumbers and Raman cross-sections of the hydrogen-bonded and non-hydrogen-bonded OH oscillators. The band is often approximated by a linear combination of several spectral components, typically by 2–6 Gaussian or Lorentzian profiles. In spite of numerous



Journal of RAMAN SPECTROSCOPY



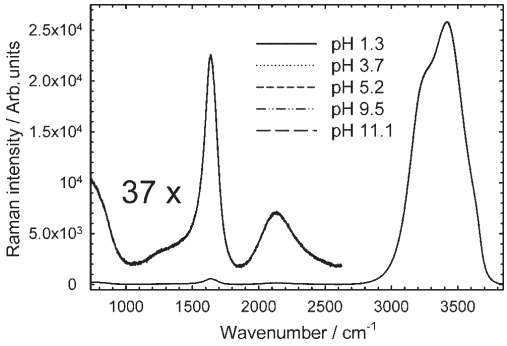



**Figure 9.** Background-corrected and SVD-normalized Raman spectra of deionized water with pH adjusted by addition of HCl and/or NaOH.

studies, no definite consensus was achieved on the number of components that should be taken into account, since their structural interpretation is largely empirical.[33,39–41] Spectral components are generally assigned, in conformity with the water model used, to different cluster structures[42] or hydrogen-bonding schemes.[43] Similar lack of agreement persists on the relationship between Raman intensities of the OH-stretching spectral components and intensive variables affecting hydrogen bonding, namely temperature and presence of electrolytes. Though temperature and ionic-strength dependences of water Raman spectra have been subjects of intense research over several decades,[1,21,30–35,39–41,44–46] there are still controversies about the absolute Raman intensity preservation and/or existence of isosbestic points within the water stretching band.

As a rough approximation, intensity of the low-wavenumber region of the stretching band of liquid water decreases on temperature increase, whereas that in the high-wavenumber region increases. In general, addition of an electrolyte to water produces similar structural and spectral changes as an increase in temperature,[31,32] and the concept of structural temperature, i.e. temperature at which pure water would have effectively the same inner structure as the water in an electrolyte solution at given temperature, can be useful in some cases.[33,43] Despite minor dissimilarities,[43] for the purpose of SVD-normalization both the temperature and ionic strength effects can be considered in a similar way.

More profound investigation of the water Raman spectra by multivariate methods was out of the scope of the present paper. Nevertheless, it is worthwhile to note that factor analysis of SVD-normalized, all-corrected datasets revealed that temperature-dependent Raman spectra (at least in the range of 2–90 °C) of pure water or neutral pH aqueous buffers could be sufficiently well described by using only three orthogonal spectral components (Fig. 3), the first two being substantial. Moreover, within a wide range of ionic strengths (0.1–5 M), no difference was evidenced between ionic strength dependences induced by LiCl, NaCl or KCl. For instance, remarkable agreement of temperature dependences regardless of the cation type is evident from joint factor analysis of the SVD-normalized, all-corrected datasets containing 100 mM LiCl, NaCl or KCl (Fig. 3). The finding seems to be in accordance with the more significant impact of anions than cations on the water structure.[31,35,43]

Using the sophisticated method of Raman instrumentation enabling the collection of absolute Raman spectra[38] for various polarization geometries, precise isosbestic points were reported by Walrafen[1,21,45] for pure water in the temperature range of 3–85 °C. The exact position of the isosbestic point was shown to be confined to the 3370–3524 $cm^{-1}$ interval depending on the polarization geometry, the upper and lower limit belonging to the pure $45\alpha^2$ and $6\beta^2$ component, respectively.[1] Isosbestic points at intermediate wavenumbers were shown to arise simply from linear combinations of pure $\alpha^2$ and $\beta^2$ spectra due to different polarization geometries. The limits of the interval were suggested to define the upper and lower wavenumbers of the peaks belonging to stretching vibrations of hydrogen-bonded and non-hydrogen-bonded water molecules, respectively.[1] Referring to existence of temperature-insensitive isosbestic point at the 3425 $cm^{-1}$ (backscattering, unpolarized X(Z,X+Z)X geometry), Raman intensity of the narrow band centered at the isosbestic point has been proposed as a practical measure of the amount of liquid water in clouds for the Raman lidar technique.[46]

Although the presence of isosbestic points in the absolute Raman spectra was questioned later,[30] clear isosbestic points often arise after normalization of the water bands to their integral intensities[30,41] or after intensity corrections for temperature dependences of water density and polarizability.[33] Recently, using a combined experimental and theoretical approach, asymmetric profile of the water stretching band and experimentally reconfirmed existence of temperature-invariant isosbestic point were interpreted at the microscopic level.[47]

### Intensive variables not affecting water stretching band

In the case of intensive variables having no effect on the shape of water stretching band, the precondition of SVD normalization is fulfilled in a trivial way. Since water band preserves its shape to the smallest details (except for the negligible contribution from the hydrogen-bond stretches of the analyte/ligand), factor analysis restricted to the 2700–3900 $cm^{-1}$ regions of the properly background-corrected spectra should reveal only single spectral component $S_1(\nu)$. All the spectra in the dataset can differ solely in the water band intensity and can be thus normalized with respect to the coefficient $V_{i1}$. Analyte/ligand concentration series, pH dependences or temporal evolutions studied at constant temperatures can serve as typical examples. As seen in Fig. 6, neither considerable reduction of the excitation intensity along the laser beam path in the absorbing analyte nor absorption of the resonance Raman signal by the sample layer between the scattering volume and the exit window of the cell deforms linear dependence of the SVD-normalized Raman intensity on the analyte concentration. Strictly speaking, absorption of the Raman signal may modify relative intensities of the analyte bands because they are absorbed with different efficiencies due to the wavelength dependence of the molar absorptivity. Such an effect can be easily demonstrated using a more extended interval of CuTMPyP4 concentrations (data not shown). In this case, for the more concentrated CuTMPyP4 samples, the low-wavenumber Raman bands are less intense than the high-wavenumber ones, and the factor analysis reveals a second orthogonal spectral component having a characteristic difference-spectrum shape.

A similar effect associated with wavelength-dependent absorption in the fingerprint region can be seen in the SERS spectra of CuTMPyP4 taken from two different SERS-active systems (Fig. 10). The small but detectable second spectral component $S_2(\nu)$ that



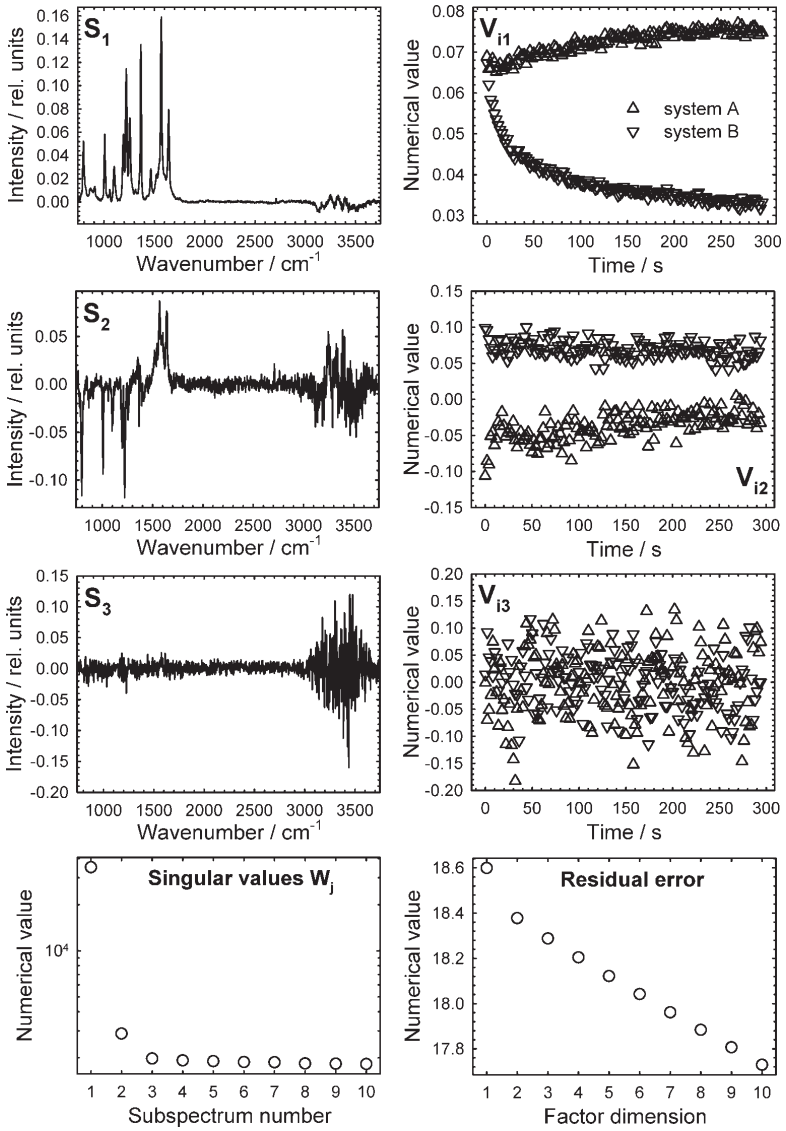



**Figure 10.** Factor analysis of the SVD-normalized, all-corrected SERS spectra of CuTMPyP4 in silver colloid. Exposure-time dependences of two SERS-active systems are compared quantitatively. Full-range spectra were decomposed to show considerable suppression of the solvent signal.

represents – in addition to different exposure-time dependences expressed by $S_1(\nu)$ and $V_{i1}$ – further orthogonal difference between the systems A and B is caused not by the analyte itself but by different intensities (and slightly different shapes) of the surface plasmon absorption bands ($\lambda_{max} \sim 390$ nm) of respective colloidal systems.[20] Under the 441.6 nm excitation, the more effective absorption of the lower-wavenumber Raman bands by optically 5× more dense system A is responsible for typical difference-spectrum appearance of the $S_2(\nu)$, exhibiting Raman bands with negative and positive intensities bellow and above ~1400 $cm^{-1}$, respectively (Fig. 10).

Other factors complicating routine use of SVD-normalization (even in the case of intensive variables having no direct effect on the water stretches) could be considerable absorption or fluorescence of the analyte/ligand in the spectral range of the water Raman band. Such an absorption, if really strong, deforms the shape of the water Raman bands in the sample dataset, whereas the bands in the corresponding SRD remain unchanged. After the



Journal of RAMAN SPECTROSCOPY



solvent-subtraction procedure described above, residual baseline of characteristic shape remains. The extent of the residuum can be used for the estimation of the normalization error and/or for its correction. On the other hand, analyte/ligand fluorescence increases the non-Raman background under the water Raman band. Even though it can be effectively suppressed by the SVD-based background correction, increased stochastic noise decreases the accuracy and reliability of the SVD-normalization as well as solvent subtraction procedures.

#### Intensive variables affecting water stretching band

In the case of intensive variables that considerably influence the shape of the water stretching band, our approach of solvent subtraction was proven to be, in general, as effective, accurate and reliable as for variables conserving the shape of water band, albeit with inherent limitations given by absorption/fluorescence of the samples. As documented by Figs 1c and 7c, when using appropriate SRD and fitting procedure, solvent contribution can also be eliminated from temperature- and ionic-strength-dependent datasets, respectively. Nevertheless, correctness of the SVD-normalization seems to be less straightforward since it is based on the isosbestic point revealed by factor analysis in an automated way, without any prior knowledge about its actual position. The problem consists in the fact that without absolute Raman spectra it is impossible to verify whether the SVD-determined isosbestic points are really insensitive to the intensive variables or not. It is the Achilles' heel of our SVD-normalization method. Correctness of the premise can be justified only *post facto* by reasonably simple dependences of the normalized analyte Raman spectra, provided that the analyte itself is sufficiently insensitive to the intensive variable.

As shown in Fig. S1, despite rigorously constant conditions of excitation intensity, collection geometry and detection efficiency,[38] properly background-corrected spectra of deionized water taken as a function of temperature neither retain their surface area nor intersect in an isosbestic point within the range of water stretches. The most preserved parameter seems to be the height of the band.[33] According to our vast experience with real-word temperature-dependent Raman spectra of aqueous buffers taken under realistic experimental conditions, integral intensities of the water stretching bands determined from background-corrected spectra often oscillate randomly (up to ±10% for the same temperature and fixed experimental arrangement) around some kind of temperature dependence, exhibiting integral intensity decrease of ~25% on going from 2 to 90 °C (Fig. 4a). However, on applying the SVD-normalization to the same datasets, water bands intersect in an isosbestic point around 3387 $cm^{-1}$ (Figs 1b and S1b) and their integral intensities exhibit smooth and monotonic dependence on the temperature (Fig. 4a) that can be approximated by a quadratic function (decrease of ~3% on going from 2 to 90 °C). Consequently, integral intensities of the analyte bands from SVD-normalized and all-corrected datasets can reveal their own temperature dependences. As shown in Fig. 4b, the weak cacodylate band at 1279 $cm^{-1}$ (at 100 mM concentration comprises merely ~4% of integral intensity of the nearby water deformation Raman band at 1640 $cm^{-1}$, Fig. 1b) seems to be almost temperature independent (~5% linear decrease), whereas the slightly more intense 1412 $cm^{-1}$ band exhibits 22% linear decrease on going from 2 to 90 °C. Similarly, the slight linear (~7%) and more apparent quadratic (~24%) intensity decrease of the more intense hydrogen-bond stretching bands of cacodylate located at 2933 and 3015 $cm^{-1}$ can be revealed, respectively (Fig. S2).

In the case of dataset with variable ionic strength, the SVD-normalized water stretching bands also form an isosbestic point at 3374 $cm^{-1}$ (Fig. 7b). After routine solvent subtraction, a detailed inspection of all-corrected cacodylate spectra (Fig. 7c) indicates a different impact of the increasing ionic strength on its hydrogen-bond stretching bands (Fig. 8) and on the pair of vibrational bands in the fingerprint region (Fig. S7). As the concentration of NaCl increases from 0.1 to 5.0 M, the 2933 and 3015 $cm^{-1}$ bands undergo peak-wavenumber downshifts, intensity decrease and narrowing half-widths, both forming their own isosbestic points at 2930 and 3011 $cm^{-1}$, respectively (Fig. 8). Prior the SVD-normalization, such fine details were completely lost in stochastic fluctuations. Owing to the SVD-normalization and solvent subtraction, the difference in cacodylate concentration as small as 5% can be reliably detected from integral intensities of strong (Fig. 8) as well as weak (Figs S7 and S8) cacodylate Raman peaks.

Providing that it is the integral intensity of the water stretching band that is conserved on temperature or ionic strength increase, SVD-normalized datasets can be easily renormalized. As shown in Fig. S1, subsequent renormalization to the unit area shifts the isosbestic point from 3387 to 3370 $cm^{-1}$, slightly affecting temperature dependences of the cacodylate bands (Fig. S2). It is worthwhile to note that the position of the isosbestic point in the renormalized spectra exactly matches the lower limit of the interval of isosbestic points reported by Walrafen[1,21,45] for various polarization geometries. A similar effect can be seen for the ionic strength dependences (Figs S6, S8 and S9) after the renormalization to unit area. Though slight differences are evident between SVD-normalized and renormalized intensities, it is impossible to decide which normalization is more correct. In any case, on the shortage of information about the isosbestic point actually independent on the intensive variable, the point disclosed by the SVD-normalization should be taken as reasonable option.

## Conclusion

Intensity normalization, background correction and solvent subtraction are essential prerequisites for any further treatment of Raman spectra in order to obtain reliable quantitative information. They are important for routine treatment of extensive Raman datasets by means of advanced multivariate methods. Dozens of approaches to this complex issue have been proposed. The method described in this paper, particularly suitable for biomolecules in aqueous solvents and exposed to gradually varied physicochemical parameters, is novel in two respects. Firstly, the non-Raman background correction, intensity normalization and solvent subtraction are based on the factor analysis, which is a powerful technique rarely employed for this purpose.[23] Secondly, the water stretching bands are shown to be a reliable internal intensity standard even for intensive variables significantly affecting their shape. The method only requires that the analyte spectra and solvent reference spectra be taken simultaneously in the fingerprint as well as in the water stretching regions. Although the applicability of the method is limited to aqueous solvents and particular classes of problems, the practical benefit could be considerable because of its simplicity, reliability and robustness.

#### Acknowledgements

This work was supported by the Grant Agency of the Charles University in Prague (3222/2008), the Grant Agency of the Czech



Journal of RAMAN SPECTROSCOPY

Republic (P208/10/0941), and the Ministry of Education of the Czech Republic (MSM0021620835).

### Supporting information

Supporting information may be found in the online version of this article.

## References


[1] G. E. Walrafen, M. S. Hokmabadi, W.-H. Yang, *J. Chem. Phys.* **1986**, *85*, 6964.
[2] L. Movileanu, J. M. Benevides, G. J. Thomas Jr, *J. Raman Spectrosc.* **1999**, *30*, 637.
[3] D. Němeček, S. A. Overman, W. Roger, R. W. Hendrix, G. J. Thomas Jr, *J. Mol. Biol.* **2009**, *385*, 628.
[4] V. Kopecký Jr, P. Mojzeš, J. V. Burda, L. Dostál, *Biopolymers (Biospectroscopy)* **2002**, *67*, 285.
[5] T. Oconnor, S. Mansy, M. Bina, D. R. McMillin, M. A. Bruck, R. S. Tobias, *Biophys. Chem.* **1982**, *15*, 53.
[6] T. Miura, G. J. Thomas Jr, *Biochemistry* **1995**, *34*, 9645.
[7] M. Procházka, P.-Y. Turpin, J. Štepánek, J. Bok, *J. Mol. Struct.* **1999**, *482–483*, 221.
[8] E. R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons, Inc.: New York, **2002**.
[9] M. J. Pelletier, *Appl. Spectrosc.* **2003**, *57*, 20A.
[10] R. N. Favors, Y. N. Jiang, Y. L. Loethen, D. Ben-Amotz, *Rev. Sci. Instrum.* **2005**, *76*, 033108.
[11] M. Kříž, D. Němeček, P.-Y. Turpin, I. Rosenberg, J. Štěpánek, *Vib. Spectrosc.* **2008**, *48*, 148.
[12] S. Song, S. A. Asher, *Biochemistry* **1991**, *30*, 1199.
[13] J. R. Beattie, J. V. Glenn, M. E. Boulton, A. W. Stitt, J. J. McGarvey, *J. Raman Spectrosc.* **2009**, *40*, 429.
[14] R. F. Pasternack, E. J. Gibbs, J. J. Villafranca, *Biochemistry* **1983**, *22*, 2406.
[15] K. Kano, H. Minamizono, T. Kitae, S. Negi, *J. Phys. Chem. A* **1997**, *101*, 6118.
[16] R. F. Pasternack, L. Francesconi, D. Raff, E. Spiro, *Inorg. Chem.* **1973**, *11*, 2606.
[17] K. H. Scheller, H. Sigel, *J. Am. Chem. Soc.* **1983**, *105*, 5891.
[18] M. J. Cavaluzzi, P. N. Borer, *Nucleic Acids Res.* **2004**, *32*, e13.
[19] M. Sladkova, B. Vlckova, P. Mojzes, M. Slouf, C. Naudin, G. Le Bourdon, *Faraday Discuss.* **2006**, *132*, 121.
[20] M. Procházka, P. Mojzeš, B. Vlčková, P.-Y. Turpin, *J. Phys. Chem. B* **1997**, *101*, 3161.
[21] G. E. Walrafen, M. R. Fisher, M. S. Hokmabadi, W.-H. Yang, *J. Chem. Phys.* **1986**, *85*, 6970.
[22] J. Hanuš, K. Chmelová, J. Štepánek, P.-Y. Turpin, J. Bok, I. Rosenberg, Z. Točík, *J. Raman Spectrosc.* **1999**, *30*, 667.
[23] G. Schulze, A. Jirasek, M. M. L. Yu, A. Lim, R. F. B. Turner, M. W. Blades, *Appl. Spectrosc.* **2005**, *59*, 545.
[24] M. N. Leger, A. G. Ryder, *Appl. Spectrosc.* **2006**, *60*, 182.
[25] A. Jirasek, G. Schulze, M. M. L. Yu, M. W. Blades, R. F. B. Turner, *Appl. Spectrosc.* **2004**, *58*, 1488.
[26] C. A. Lieber, A. Mahadevan-Jansen, *Appl. Spectrosc.* **2003**, *57*, 1363.
[27] P. A. Mosier-Boss, S. H. Lieberman, R. Newbery, *Appl. Spectrosc.* **1995**, *49*, 630.
[28] Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, J. Hu, *Chemom. Intell. Lab. Syst.* **2007**, *85*, 94.
[29] J. V. Glenn, J. R. Beattie, L. Barrett, N. Frizzell, S. R. Thorpe, M. E. Boulton, J. J. McGarvey, A. W. Stitt, *FASEB J.* **2007**, *21*, 3542.
[30] T. A. Dolenko, I. V. Churina, V. V. Fadeev, S. M. Glushkov, *J. Raman Spectrosc.* **2000**, *31*, 863.
[31] S. A. Burikov, T. A. Dolenko, V. V. Fadeev, A. V. Sugonyaev, *Laser Phys.* **2005**, *15*, 1175.
[32] S. A. Burikov, T. A. Dolenko, V. V. Fadeev, I. I. Vlasov, *Laser Phys.* **2006**, *17*, 1255.
[33] A. M. P. Neto, O. Sala, *Braz. J. Phys.* **2004**, *34*, 137.
[34] K. Furic, I. Ciglenecki, B. Cosovic, *J. Mol. Struct.* **2000**, *550*, 225.
[35] P. N. Perera, B. Browder, D. Ben-Amotz, *J. Phys. Chem. B* **2009**, *113*, 1805.
[36] J. J. Max, M. Trudel, C. Chapados, *Appl. Spectrosc.* **1998**, *52*, 234.
[37] H. Wang, C. Wang, C. K. Mann, T. J. Vickers, *Appl. Spectrosc.* **1998**, *52*, 240.
[38] G. E. Walrafen, *J. Chem. Phys.* **2004**, *120*, 4868.
[39] R. Li, Z. Jiang, F. Chen, H. Yang, Y. Guan, *J. Mol. Struct.* **2004**, *707*, 83.
[40] D. M. Carey, G. M. Korenowski, *J. Chem. Phys.* **1998**, *108*, 2669.
[41] A. V. Kargovsky, *Laser Phys. Lett.* **2006**, *3*, 567.
[42] R. Ludwig, *Angew. Chem. Int. Ed.* **2001**, *40*, 1808.
[43] Y. Marcus, *Chem. Rev.* **2009**, *109*, 1346.
[44] P. Terpstra, D. Combes, A. Zwick, *J. Chem. Phys.* **1990**, *92*, 65.
[45] W. B. Monosmith, G. E. Walrafen, *J. Chem. Phys.* **1984**, *81*, 669.
[46] D. N. Whiteman, G. E. Walrafen, W.-H. Yang, S. H. Melfi, *Appl. Opt.* **1999**, *38*, 2614.
[47] J. D. Smith, C. D. Cappa, K. R. Wilson, R. C. Cohen, P. L. Geissler, R. J. Saykally, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14171.



# SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations

## Supplementary data

Journal of Raman Spectroscopy 42 (2011) 1528-1539

Jan Palacký, Peter Mojzeš and Jiří Bok

Institute of Physics, Faculty of Mathematics and Physics, Charles University in Prague, Ke Karlovu 5, CZ-121 16 Prague 2, Czech Republic

Correspondence should be addressed to Peter Mojzeš, mojzes@karlov.mff.cuni.cz

## Supporting Information

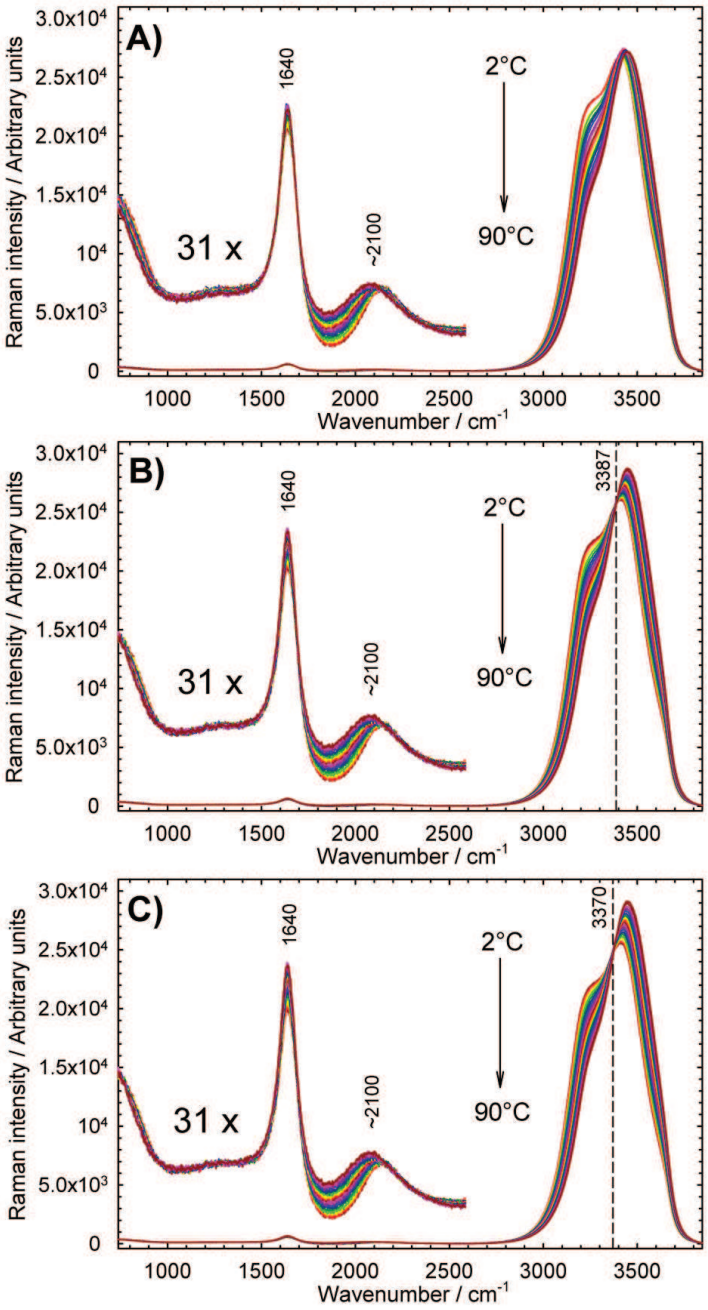



**Fig. S1.** Background-corrected temperature-dependent Raman spectra of deionized water (A). After SVD-normalization to $V_{i1}$, water stretching bands intersect in an isosbestic point at 3387 $cm^{-1}$ (B). Subsequent renormalization to unit area of the water stretching bands shifts the isosbestic point to 3370 $cm^{-1}$ (C).

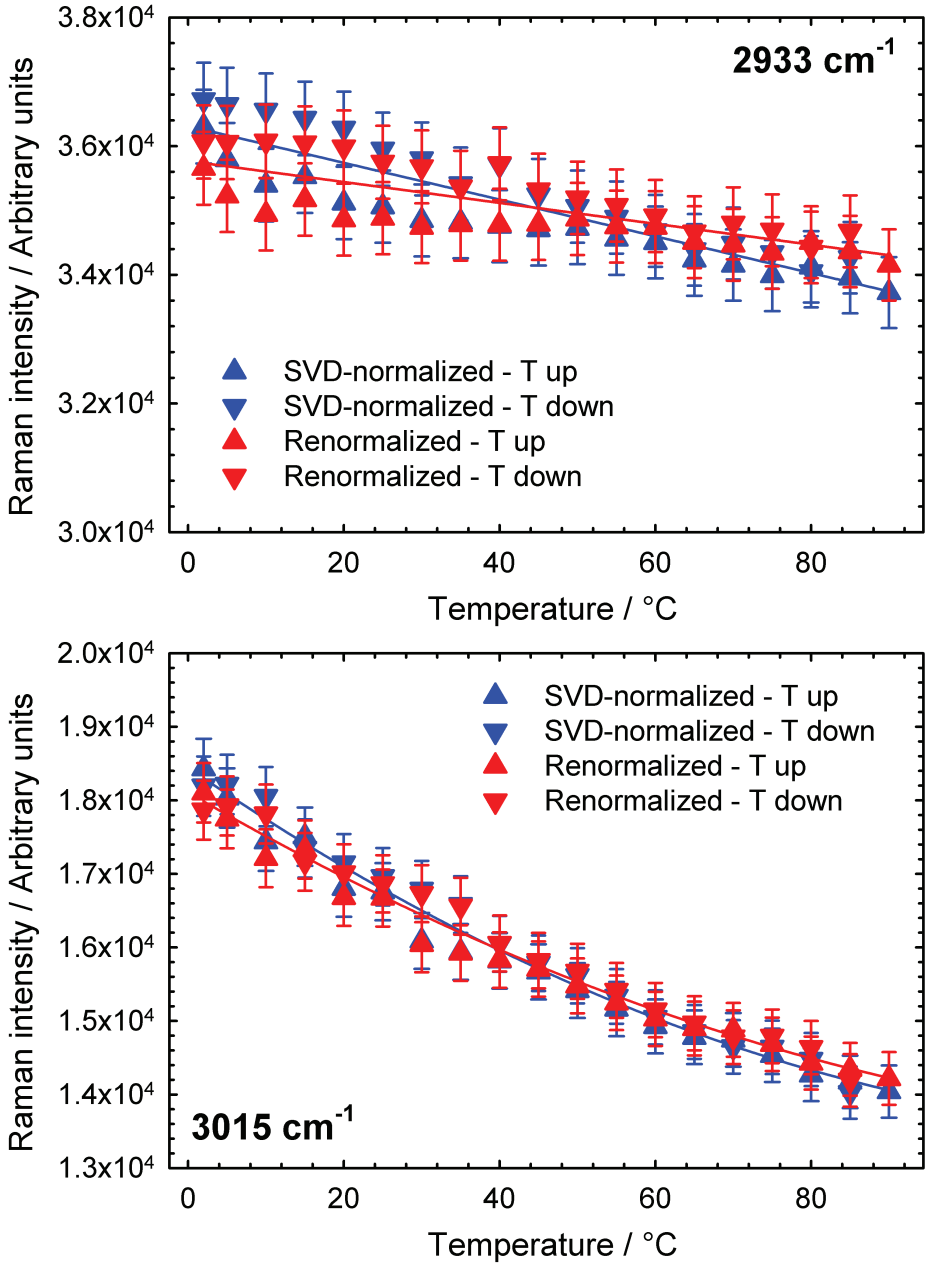



**Fig. S2.** Integral intensities of the 2933 and 3015 $cm^{-1}$ cacodylate bands as a function of temperature. SVD-normalized and renormalized (to unit area of the water stretching bands) intensities are shown separately on temperature increase and decrease. Experimental points are fitted to linear (2933 $cm^{-1}$) and quadratic (3015 $cm^{-1}$) dependences.

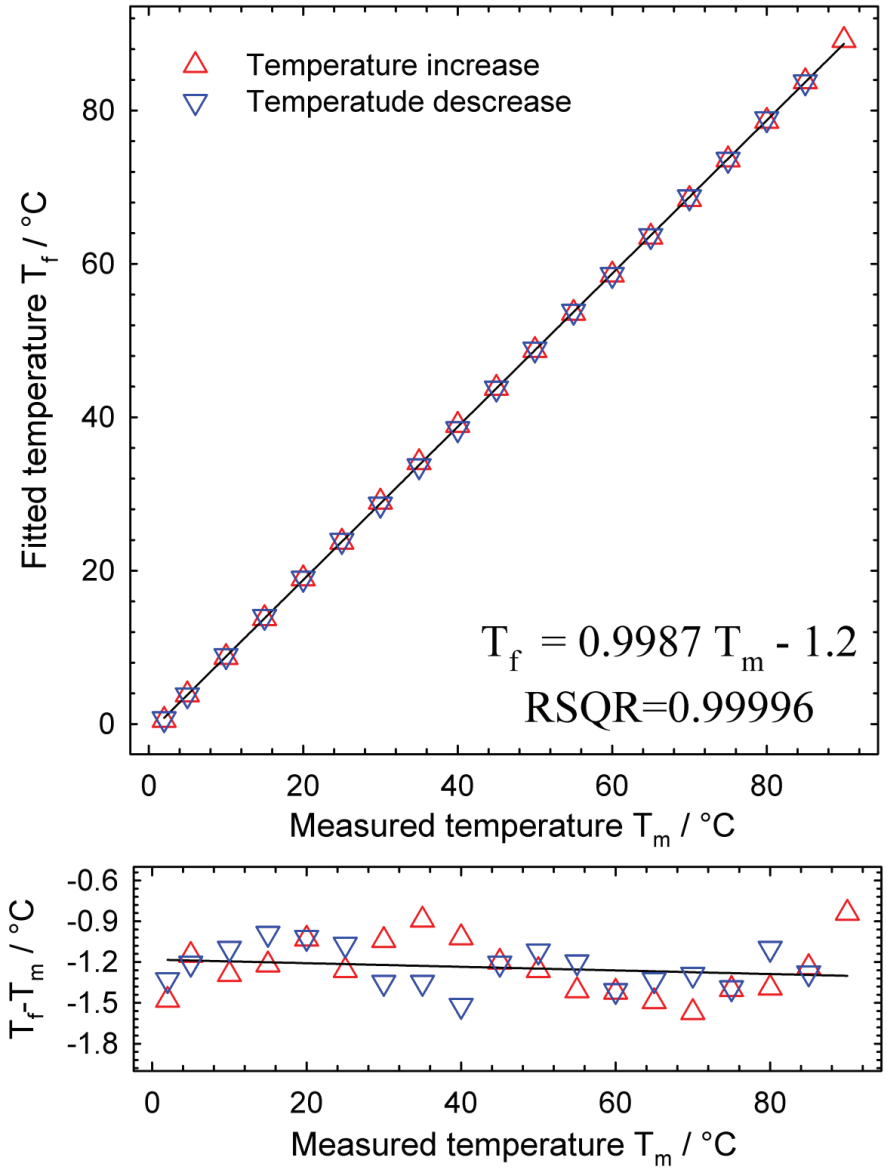



**Fig. S3.** Comparison of the sample temperatures $T_m$ measured directly with the fitted temperatures $T_f$ obtained by the least-square procedure (Eq. 4) applied to the dataset from Fig. 2. Coefficients of linear fit, the RSQR and differences between the fitted and measured temperatures approve reliability of the procedure.

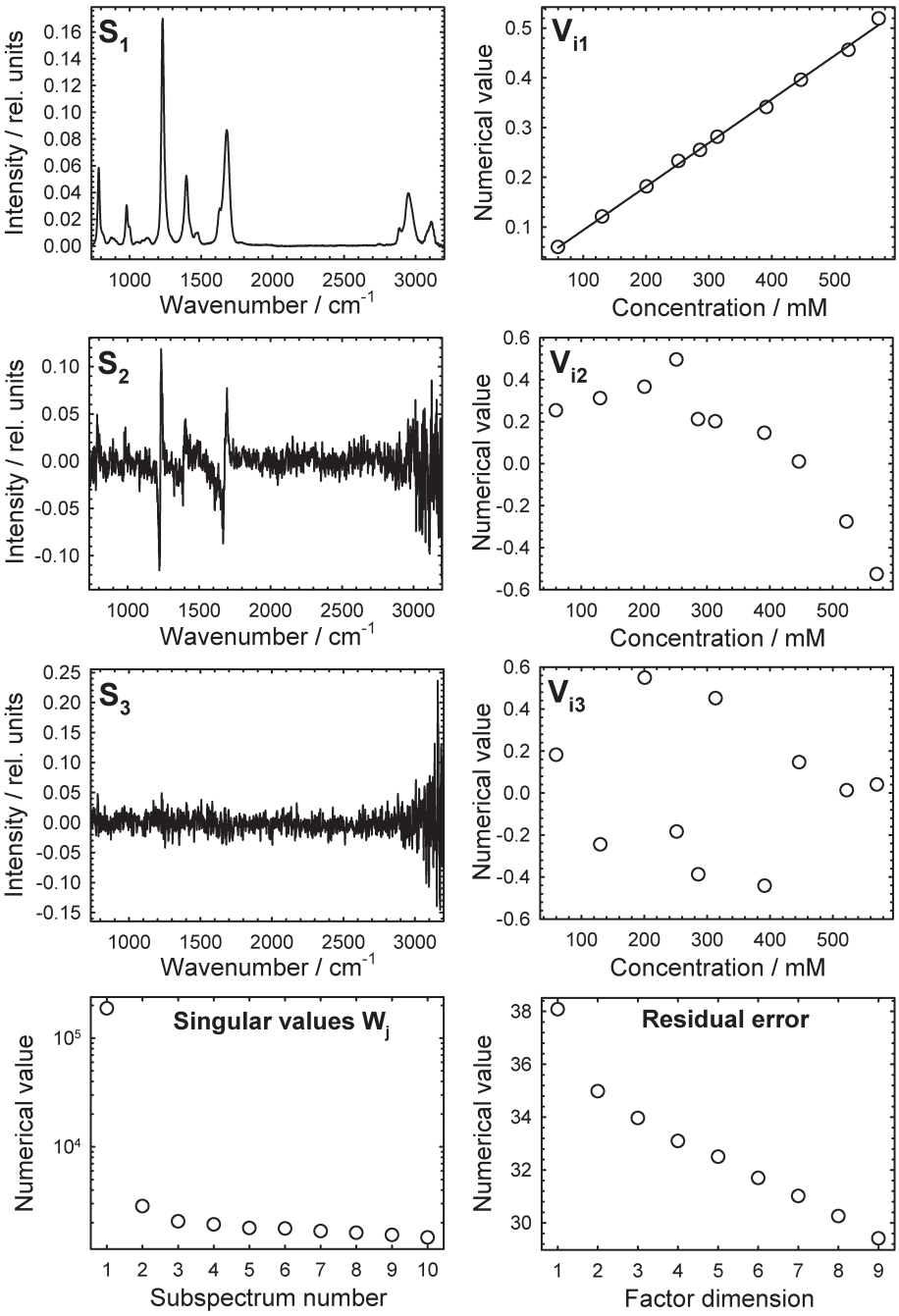



**Fig. S4.** First three orthogonal spectral components disclosed by factor analysis in the SVD-normalized, all-corrected Raman spectra of the concentration series of 5'-UMP (dataset from Fig. 5B).

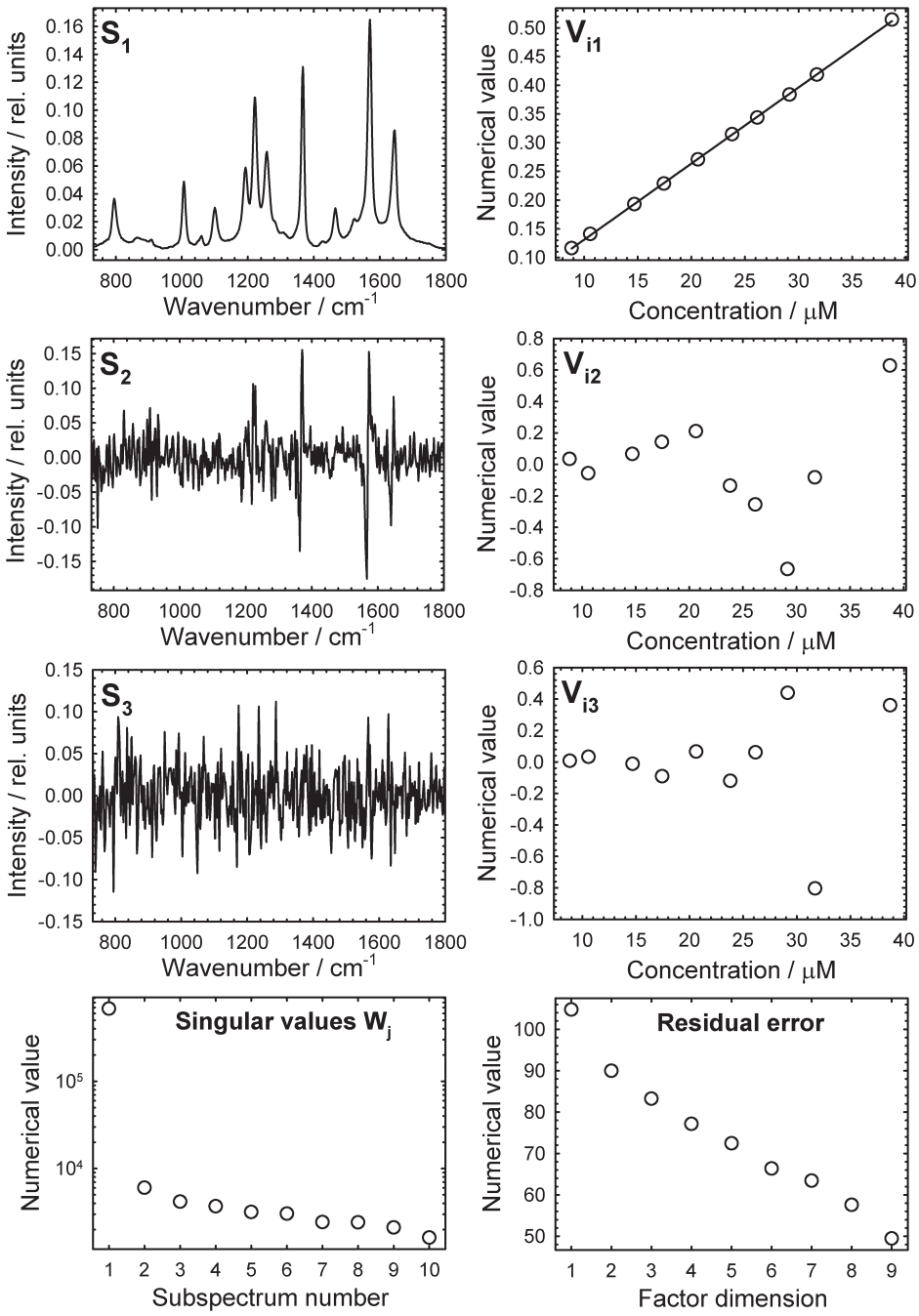



**Fig. S5.** First three orthogonal spectral components disclosed by factor analysis in the SVD-normalized, all-corrected Raman spectra of the concentration series of CuTMPyP4 (dataset from Fig. 6B).

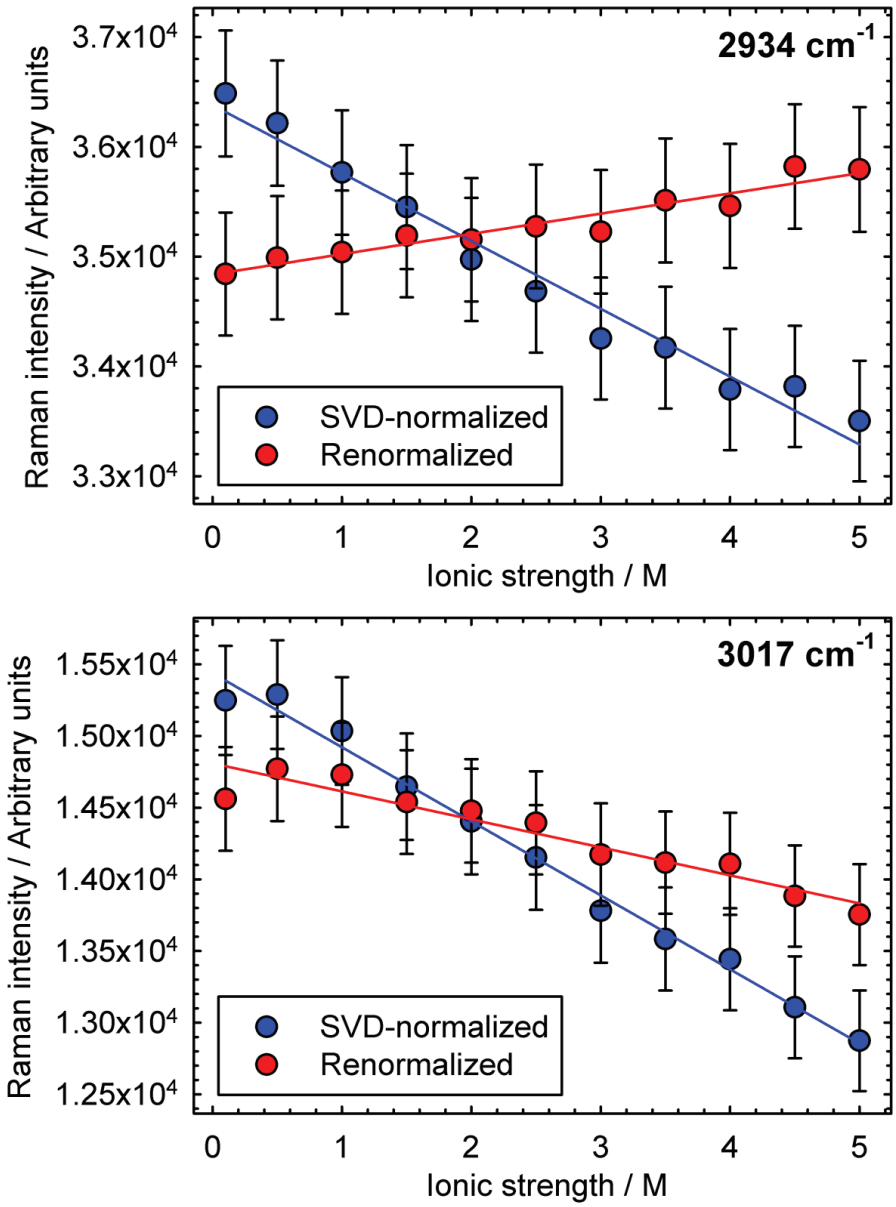



**Fig. S6.** Details of the ionic-strength dependences of integral intensities of 2934 and 3017 $cm^{-1}$ cacodylate bands. The SVD-normalized and renormalized (to unit area of the water stretching bands) intensities are compared. The ionic-strength dependences were approximated by linear fits.

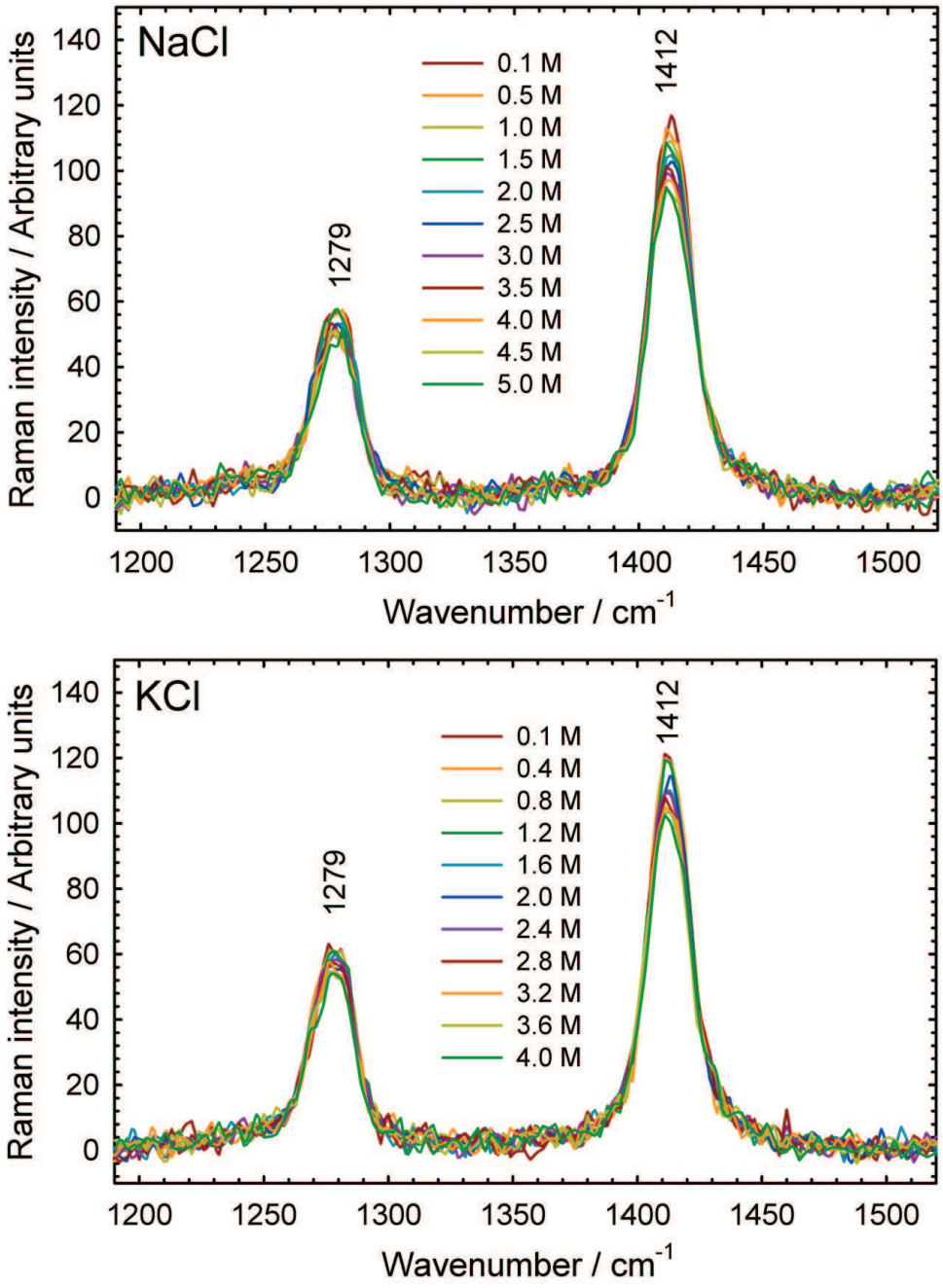



**Fig. S7.** Detail view of spectral changes of the 1279 and 1412 $cm^{-1}$ cacodylate bands induced by increasing NaCl or KCl concentrations. SVD-normalized spectra are shown. Contrary to 2934 and 3017 $cm^{-1}$ bands (Fig. 8), no changes in the position nor shape of the 1279 and 1412 $cm^{-1}$ bands were observed on the ionic strength increase.

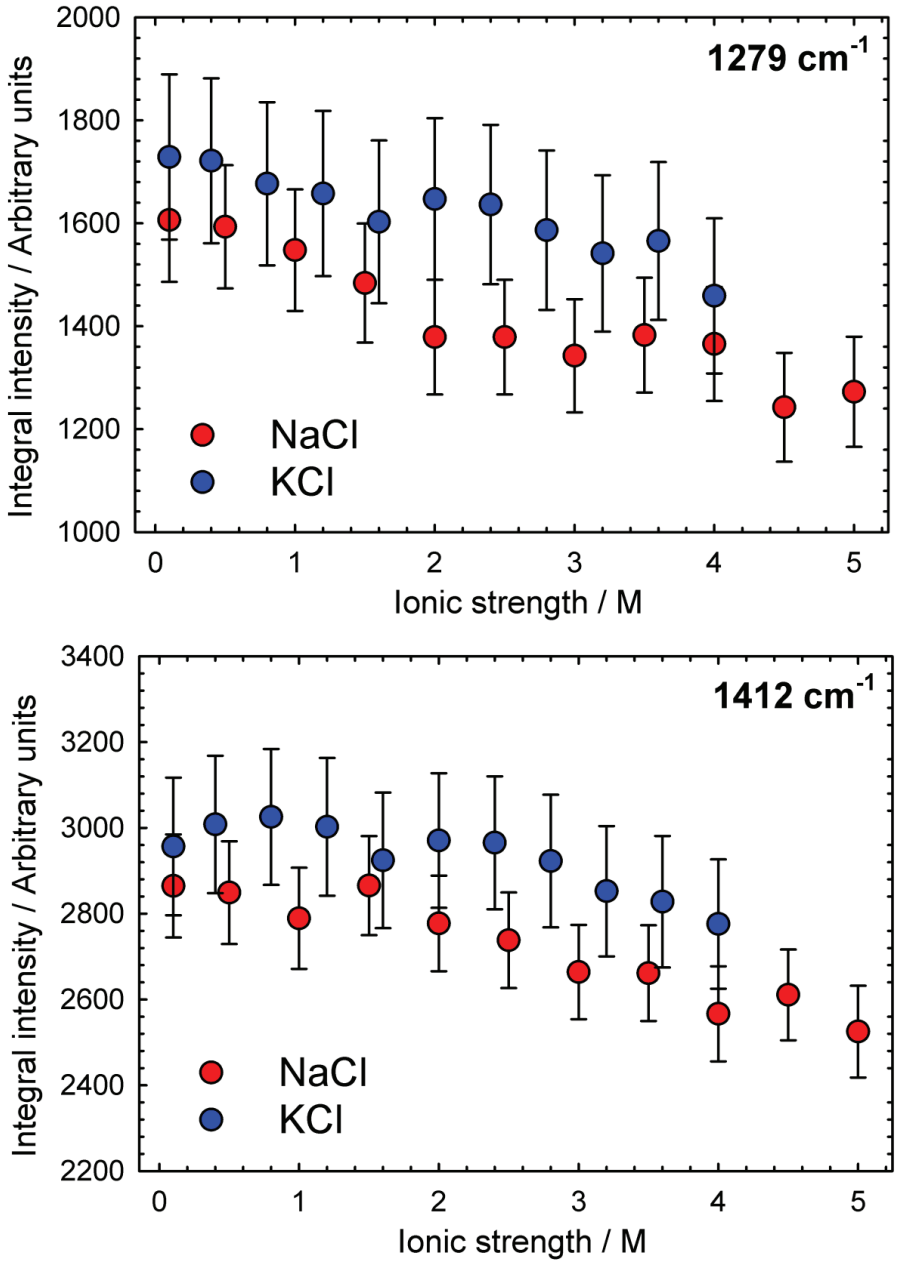



**Fig. S8.** Ionic-strength dependences of the SVD-normalized integral intensities of 1279 and 1412 $cm^{-1}$ cacodylate bands (data from Fig. S7). Concentration of cacodylate ions in NaCl and KCl series are100 and 95 mM, respectively.

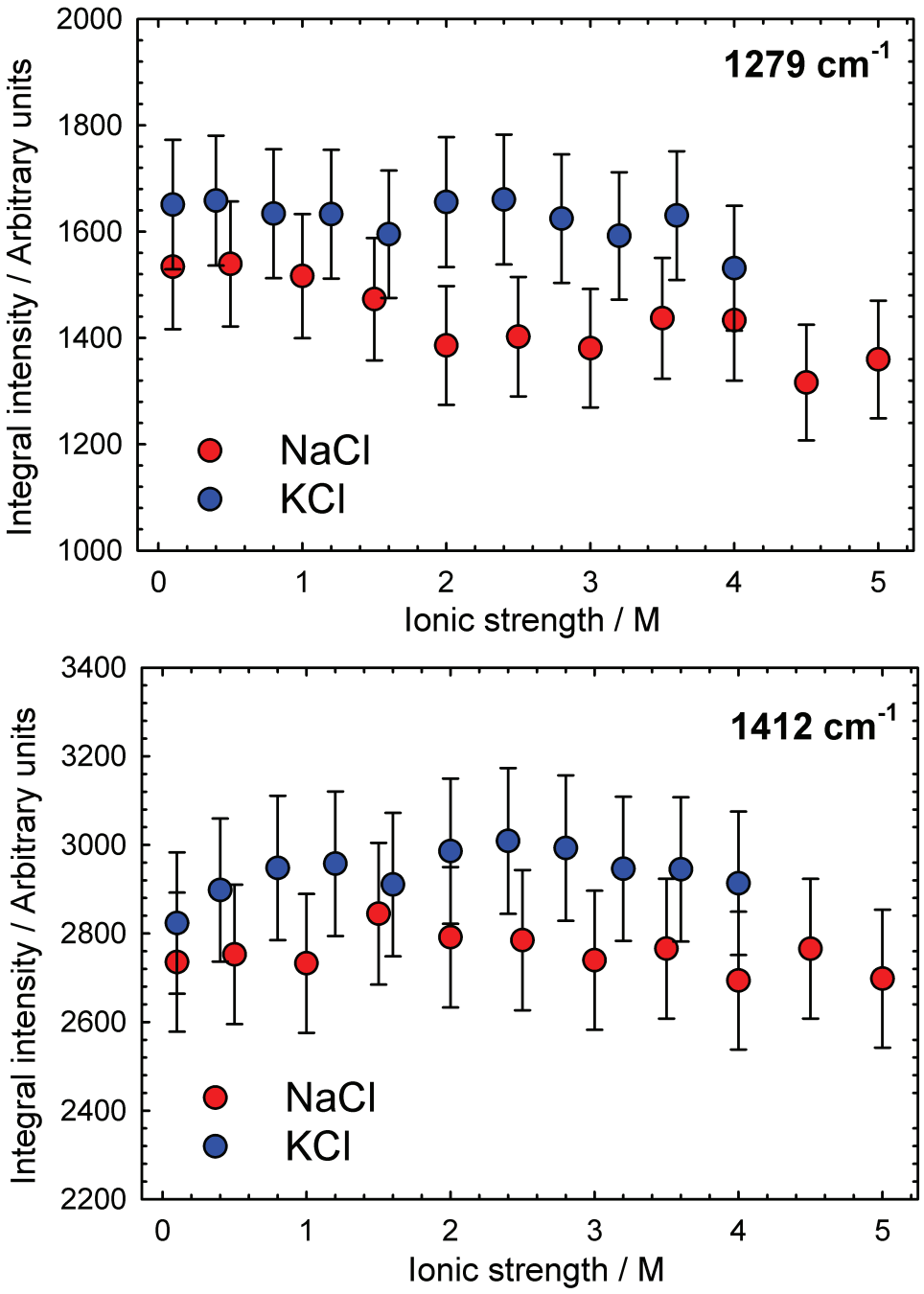



**Fig. S9.** Ionic-strength dependences of the cacodylate bands from Fig. S8 renormalized to unit area of the water stretching band.

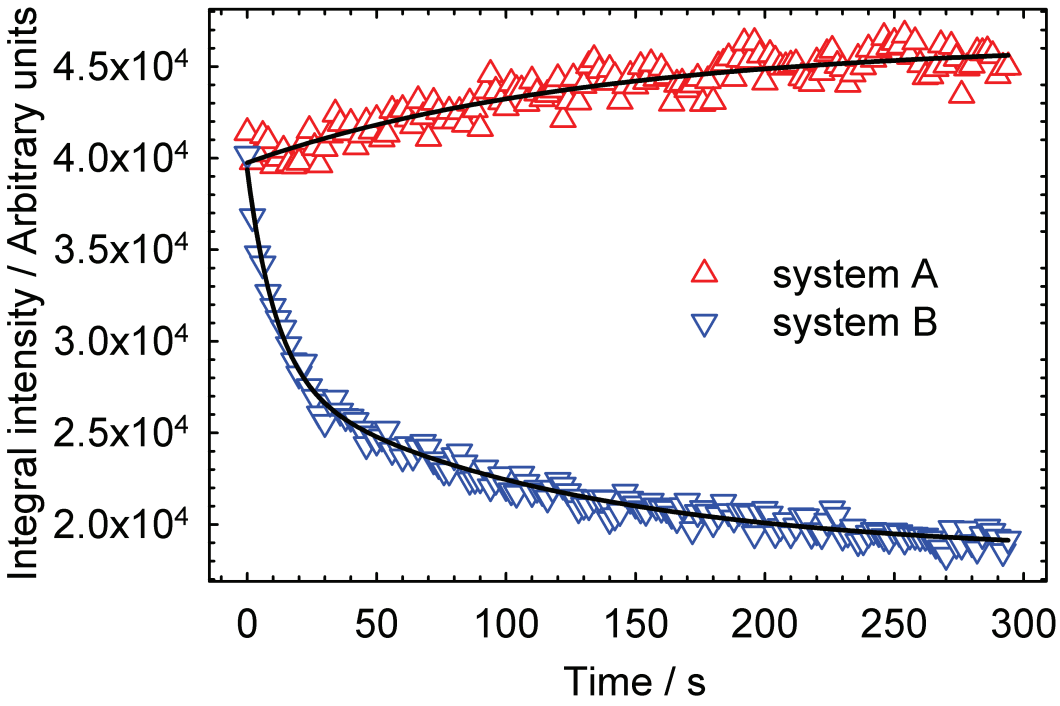



**Fig. S10.** Quantitative comparison of the exposure-time dependencies of integral SERS intensities from fingerprint regions of datasets shown on Fig. 10. The exposure-time dependency of the SERS-active systems A can be satisfactorily fitted by a single-exponential rise to maximum, disclosing a 17% intensity increase from its initial value with a time constant of 133 s. On the contrary, time evolution of the SERS-active systems B have to be fitted as double-exponential intensity decay to 47% of its initial value with time constants $\tau_1$ = 11 s and $\tau_2$ = 115 s. The estimated inaccuracy of time constants was better than ±15%.

# List of all results

## *International peer-reviewed journals and proceedings*

1. PALACKÝ JAN, VORLÍČKOVÁ MICHAELA, KEJNOVSKÁ IVA, MOJZEŠ PETER (2012) Polymorphism of human telomeric quadruplex structure controlled by DNA concentration: a Raman Study. *Nucleic acids research, submitted*.
2. BEDNÁROVÁ LUCIE, PALACKÝ JAN, BAUEROVÁ VÁCLAVA, HRUŠKOVÁ-HEIDINGSFELDOVÁ OLGA, PICHOVÁ IVA, MOJZEŠ PETER (2012) Raman microspectroscopy of the yeast vacuoles. *Spectroscopy: An International Journal* 27, 503-507.
3. PALACKÝ JAN, MOJZEŠ PETER, BOK JIŘÍ (2011) SVD-based method for intensity normalization, background correction and solvent subtraction in Raman spectroscopy exploiting the properties of water stretching vibrations. *Journal of Raman spectroscopy* 42, 1528-1539.
4. MAŠEK VLASTIMIL, MOJZEŠ PETER, PALACKÝ JAN, BOK JIŘÍ, ANZENBACHER PAVEL: Binding of platinum complexes to DNA monitored by Raman spectroscopy, In: Champion P. M., Ziegler L. D. (Eds.), *Proceedings of the XXII$^{nd}$ International Conference on Raman Spectroscopy (ICORS 2010)*, Boston, The United States of America, August 8-13, 2010, 416-417, ISBN: 978-0-7354-0818-0.
5. MOJZEŠ PETER, PALACKÝ JAN, MAREK VÍT, LÁSKOVÁ BARBORA: Interaction of copper porphyrin with guanine quadruplexes studied by resonance Raman and SERRS spectroscopies, In: Withnall R., Chowdhry B. Z. (Eds.), *Proceedings of the XXI$^{st}$ International Conference on Raman Spectroscopy (ICORS 2008),* Book of Abstracts, London, United Kingdom of Great Britain and Northern Ireland, August 17-22, 2008, IM Publications LLP (2008) 912-913, ISBN: 978-1-906715-00-7.

## *Domestic journals*

MOJZEŠ PETER, PALACKÝ JAN, BAUEROVÁ VÁCLAVA, BEDNÁROVÁ LUCIE (2011) Ramanova mikrospektroskopie a mapování buněk a tkání. *Československý časopis pro fyziku* 61, 178 – 184

## *Publications in preparation*

1. PALACKÝ JAN, VORLÍČKOVÁ MICHAELA, KEJNOVSKÁ IVA, BARVÍK IVAN, BEDNÁROVÁ LUCIE, GROCKÝ MARIÁN, SRB PAVEL, LANG JAN, ŠTĚPÁNEK JOSEF, MOJZEŠ PETER, Spectroscopic study of the TBA quadruplex and its interaction with cationic porphyrins.
2. PALACKÝ JAN, VORLÍČKOVÁ MICHAELA, KEJNOVSKÁ IVA, BEDNÁROVÁ LUCIE, MOJZEŠ PETER, Monitoring the interaction of the copper porphyrin with G-quadruplexes by Raman scattering

## *Conference proceedings:*

1. MOJZEŠ PETER, ŠLOUFOVÁ IVANA, PEKSA VLASTIMIL, PALACKÝ JAN, VLČKOVÁ BLANKA: Kinetics studies of SERS hot-spots formation, *31st European Congress on Molecular Spectroscopy*, Conference abstract, Cluj-Napoca, Romania, August 26 – August 31, 2012
2. KEJNOVSKÁ IVA, PALACKÝ JAN, MOJZEŠ PETER, SÁGI JÁNOS, RENČIUK DANIEL, KYPR JAROSLAV, VORLÍČKOVÁ MICHAELA: DNA concentration determines the folding of the human telomere quadruplex, *Third International Meeting on G-Quadruplex and G-assembly*, Book of Abstracts, Sorrento, Italy, June 28 – July 1, 2011
3. MAŠEK VLASTIMIL, MOJZEŠ PETER, PALACKÝ JAN, BOK JIŘÍ, ANZENBACHER PAVEL: Formation of cisplatin-DNA adducts monitored by Raman spectroscopy, *14thEuropean Conference on the Spectroscopy of Biological Molecules (ECSBM 2011)*, Conference abstract, Coimbra, Portugal, August 29 – September 3, 2011
4. SPÁČIL DUŠAN, MOJZEŠ PETER, PALACKÝ JAN, VLČKOVÁ BLANKA: SERRS of a cationic metalloporphyrin in systems with non-aggregated and aggregated silver nanoparticles: Electromagnetic and molecular resonance contributions, *14th European Conference on the Spectroscopy of Biological Molecules (ECSBM 2011)*, Conference abstract, Coimbra, Portugal, August 29 – September 3, 2011
5. BEDNÁROVÁ LUCIE, PALACKÝ JAN, OLGA HRUŠKOVÁ-HEIDINGSFELDOVÁ, VÁCLAVA BAUEROVÁ, MOJZEŠ PETER: Raman microspectroscopy of the yeast vacuoles, *14th European Conference on the Spectroscopy of Biological Molecules (ECSBM 2011)*, Conference abstract, Coimbra, Portugal, August 29 – September 3, 2011
6. PALACKÝ JAN, MOJZEŠ PETER: Dimeric structure of quadruplex-forming thrombin-binding aptamer in complexes with cationic porphyrin, In: Cupane A. (Ed.), *XIII European Conference on the Spectroscopy of Biological Molecules (ECSBM 2009)*, Book of Abstracts, Palermo, Italy, August 28 – September 2, 2009, PB 51
7. MOJZEŠ PETER, PALACKÝ JAN, LÁSKOVÁ BARBORA: Adsorbate-induced aggregation of silver colloids and irreversible loss of SERS signal due to irradiation, In: Cupane A. (Ed.), *XIII European Conference on the Spectroscopy of Biological Molecules (ECSBM 2009)*, Book of Abstracts, Palermo, Italy, August 28 – September 2, 2009, PA 84
8. MAŠEK VLASTIMIL, MOJZEŠ PETER, PALACKÝ JAN, ANZENBACHER PAVEL: Cisplatin to DNA binding kinetics monitored by Raman spectroscopy, In: Cupane A. (Ed.), *XIII European Conference on the Spectroscopy of Biological Molecules (ECSBM 2009)*, Book of Abstracts, Palermo, Italy, August 28 – September 2, 2009, PB 48
9. PALACKÝ JAN, MOJZEŠ PETER, KEJNOVSKÁ IVA: The Interaction of the TBA Quadruplex with Cationic Porphyrins, COST action MP0802 – training school, Ischia, Italy, October 27 – 29, 2009, p. 24