

**Time-Resolved NMR-Spectroscopic Studies of
Conformational Dynamics in DNA G-Quadruplexes**

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

J. Tassilo Grün

geboren in Wetzlar

Frankfurt am Main

2021

(D30)

Vom Fachbereich Biochemie, Chemie und Pharmazie der Johann Wolfgang Goethe-Universität als
Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Prof. Dr. Harald Schwalbe

Prof. Dr. Alexander Heckel

Datum der Disputation: 24.06.2021

Table of Contents

List of Abbreviation	iv
1 Summary and Overview	
1.1 Summary	1
1.2 Zusammenfassung	3
2 General Introduction	
2.1 Structure and Dynamics of DNA G-Quadruplexes	
2.1.1 Molecular Structure and Conformations of DNA.....	9
2.1.1 Non-Canonical DNA Structures.....	11
2.1.2 Canonical Structural Polymorphism.....	14
2.1.3 Non-Canonical Structural Polymorphism.....	15
2.1.4 Conformational Dynamics	17
2.2 Folding Dynamics of DNA G-Quadruplexes	
2.2.1 Folding Energy Landscapes and Folding Pathways	19
2.2.2 Experimental Approaches to Measure Folding Kinetics.....	21
2.2.3 Folding Kinetics of DNA G-Quadruplexes.....	22
2.2.4 Comparison to RNA G-Quadruplexes.....	23
2.2.5 Mutual Exclusive Formation of DNA i-motifs and G-Quadruplexes.....	23
2.3 Function and Targetability of DNA G-Quadruplexes	
2.3.1 Occurrence of G-Quadruplex Forming Sequences in the Human Genome	24
2.3.2 Transcriptional Regulation of G-Quadruplexes and Binding-Interactions	26
2.3.3 Oxidative Damage under Cellular Stress Conditions.....	30
2.3.4 Strategies in Anti-Cancer Treatment.....	31
2.4 The cMYC Proto-Oncogene Promoter	
2.4.1 The cellular Myelocytomatosis Protooncogene	33
2.4.2 Nuclease Hypersensitive Element III ₁	35
2.4.3 Structures and Dynamics of the cMYC G-Quadruplex.....	37
2.4.4 Targeting the cMYC G-Quadruplex.....	40
2.5 NMR Methods to Study Structure and Dynamics in Nucleic Acids	
2.5.1 Structure Determination of Nucleic Acids with Nat. Abundance Isotopes....	43
2.5.2 The NMR-Timescale: Methods to investigate Dynamics.....	45
2.5.3 Real-Time NMR: Non-Equilibrium Dynamics	47
2.5.4 Conformational Selection with Photolabile Protecting Groups.....	49

3 Materials and Methods

3.1 DNA Oligonucleotides and Sample Preparation	
3.1.1 Overview of DNA Oligonucleotides.....	53
3.1.2 Synthesis of Photocaged DNA Oligonucleotides	54
3.2 CD Spectroscopy.....	58
3.3 Non-Linear Regression and Global Fitting of Kinetic Data.....	58
3.4 NMR Spectroscopy	59

4 G-Register Exchange Dynamics

4.1 Overview	61
4.2 K ⁺ -Induced Folding into G-register Isomers	62
4.3 Light-Induced Refolding of Caged G-Register Isomers	
4.3.1 Conformational Selection of G-Register Isomers	64
4.3.2 Refolding into a Two-State G-Register Ensemble	67
4.3.3 Refolding into a Full Set G-Register Ensemble.....	68
4.4 G-Register Polymorphism in the <i>hTERT</i> Promoter G-Quadruplex.....	69
4.5 Activation Energies of G-Register Exchange Transitions	71
4.6 Folding Initiation in Pre-Equilibrated K ⁺ -Solution	
4.6.1 Dynamics of Thermally-Induced (Un)-Folding of G-Register Isomers	73
4.6.2 Light-Induced Folding into G-register Isomers.....	75
4.7 Conclusions	76

5 Spare-Tire Exchange Dynamics

5.1 Overview	79
5.2 Light-Induced Folding of Caged Spare-Tire-Isomers	
5.2.1 Trapping Compl. Unfolded Spare-Tire Isomers with Multiple Photocages ..	80
5.2.2 Light-Induced Folding Kinetics under Pre-Equilibrated K ⁺ Conditions	81
5.2.3 Intermediate Formation and Folding Pathway for <i>cMYC</i> -1234.....	84
5.3 Temperature Dependence of Folding.....	86
5.4 Light-Induced Refolding of Caged Meta-Stable Minor Conformations	
5.4.1 Conformational Selection.....	88
5.4.2 Refolding Kinetics.....	89

5.5 Conclusions	91
6 The Dynamic Behaviour of the <i>cMYC</i> Promoter G-Quadruplex	
6.1 Folding Pathways for the 22 nt <i>cMYC</i> Full Length Sequence	
6.1.1 K ⁺ -Recruitment During Folding	93
6.1.2 Parallel Reaction Pathways Accelerate Folding (G-Register Isomers)	94
6.1.3 Enlarged Conformational Space Decelerates Folding (<i>Spare-Tire</i> Isomers) ..	95
6.2 Folding intermediates	95
6.3 Refolding Across Different Transitory Ensembles	97
6.4 Non-Canonical Structural Polymorphism as Challenge	98
6.5 Model for <i>cMYC</i> G4 (Re)-Folding including Non-Canonical Polymorphism ..	99
6.6 Conclusion and Outlook	100
References	101
Appendix	
6.1 HMBC and Deuterium Exchange NMR Spectrum	121
6.2 Plots of Time-Resolved 1D ¹ H NMR Spectra for <i>Spare-Tire</i> G4 Folding	122
6.3 Python Script for Kinetic Fits	123
Acknowledgements - Danksagung.....	131
Curriculum Vitae	133

List of Abbreviations

A	Adenosine	MS	Mass spectrometry
b(p)	Base (pair)	n.a.	not assigned
C	Cytosine	n.d.	not determined
CD	Circular Dichroism	NMR	Nuclear Magnetic Resonance
COSY	Correlation Spectroscopy	NOESY	Nuclear Overhauser Spectroscopy
DMSO	Dimethyl sulfoxide	ns	number of scans
DNA	Deoxyribonucleic Acid	nt	Nucleotide
dNTP	deoxy-Nucleoside-tri-Phosphate	OD	optical density
ds/ss	double/single stranded	PAGE	Poly Acrylamide Gel Electrophoresis
DSS	2,2-Dimethyl-2-silapentane-5-sulfonic acid	PDB	Protein Data Bank
EDTA	Ethylene Diamine Tetra Acetic acid	PNA	peptide nucleic acid
ESI	Electrospray-Ionization	PPG	photolabile protecting group
<i>et al.</i>	<i>et alia</i> , and others	ppm	Parts Per Million
FRET	Förster Resonance Energy Transfer	PQS	Putative Quadruplex Sequence
G	Guanosine	RNA	Ribonucleic Acid
G4	G-Quadruplex	rt	Room temperature (298 K)
HSQC	Heteronuclear Single Quantum Coherence	T	Thymidine
HMBC	Heteronuclear Multiple Bond Correlation	TEA	Tetraethylammonium [(C ₂ H ₅) ₄ N ⁺]
HPLC	High Performance Liquid Chromatography	TH	Thermal Hysteresis
K	Kelvin	TOCSY	Total Correlation Spectroscopy
K_d	Dissociation constant	TRIS	Tris-(hydroxymethyl)-aminomethane
K-P_i	potassium phosphate buffer	TSS	Transcription start site
MALDI	Matrix-assisted Laser Ionization	U	Uridine
		UV-vis	Ultraviolet-visible

“One of the defining characteristics of a living system is the ability of even the most intricate of its component molecular structures to self-assemble with precision and fidelity. Uncovering the mechanisms through which such processes take place is one of the grand challenges of modern science.”

Sir Christopher M. Dobson, 2003^{1,2}

1 Summary and Overview

1.1 Summary

The present thesis *Time-Resolved NMR-Spectroscopic Studies of Conformational Dynamics in DNA G-Quadruplexes* deals with the detailed investigation of important structural dynamics in non-canonical secondary structure elements in nucleic acids.

G-rich DNA sequences in the human genome can form non-canonical secondary structures that deviate from the double stranded Watson-Crick helix. In presence of monovalent cations, G-residues are able to form tetrads via G-G Hoogsteen base pair interactions that stack to structures known as G-quadruplexes (G4, chapter 2.1). G4s are highly dynamic structures able to adopt numerous conformations. The pronounced polymorphism and the complex folding energy landscapes of DNA G4s lead to the co-existence of different folded conformations (chapter 2.2). The involved folding and refolding dynamics, described by their transition kinetics, remain largely enigmatic. The intrinsic dynamic and volatile nature of DNA G4s is a crucial feature for their vital roles in cell regulation and gene expression (chapter 2.3). Malfunction of the G4 biological functions is strongly linked to rare hereditary diseases, epigenetic regulation and oncogenesis. DNA G4s therefore have evolved as valuable targets in structure-based drug design. Understanding the conformational dynamics of G4s is crucial to enable and aid such therapeutic approaches. The human *cMYC* oncogene is a key proliferation driver in numerous cancer types. A nuclease hypersensitive element (NHE-III₁) in the *cMYC* promoter region is able to fold into a polymorphic G4 ensemble that regulates up to 90% of *cMYC* transcription levels (chapter 2.4). The *cMYC* NHE-III₁ G4 ensemble features two different kinds of newly discovered non-canonical polymorphism: (i) the sequence has more than four G-rich tracts (“*spare-tires*”) close in sequence that can isomerize by involving different G-tracts into their most stable structure; and (ii) in addition, two of the individual G-tracts contain more than three consecutive G-residues. Different G-residues within one G-tract can participate in the tetrad formation opening the possibility of **G-register** isomerism. The kinetics and dynamics of G4 formation (folding) and transformation (refolding) featuring both kinds of non-canonical polymorphism have remained completely elusive. This thesis presents a novel approach to investigate this kinetics with atomic resolution using time-resolved NMR methods (chapter 2.5). A general method to prepare and trap unfolded (conformational suppression) or isolated folded (conformational selection) states utilizing photolabile protecting groups is presented.

Chapter 4 examines the dynamics of G-register isomers. The kinetics of folding (starting from unfolded states) and refolding (starting from isolated folded states) have been investigated. The refolding kinetics for two different G4 pairs are presented, each pair is related via G-register isomerism: (i) two co-existing *all-parallel* G4s found in the *cMYC* promoter; and (ii) two co-existing hybrid and parallel G4s found in the *hTERT* promoter.

1.1 Summary

K⁺-induced folding into a two-state G-register pair reveals a kinetic partitioning mechanism. Both isomers are folded concurrently, with a kinetic overshoot of one isomer. Subsequent to initial folding, slow refolding kinetics ($\sim 0.9 \text{ h}^{-1}$) have been observed for the relaxation towards conformational equilibrium. A careful analysis of the involved kinetics and apparent activation energy barriers supports the hypothesis of inherently different refolding mechanisms for the *hTERT* and *cMYC* G4s. The time-resolved NMR data are accompanied by experimental data from a biophysical method that is based on analysis of thermal hysteresis for thermal (un)-folding of the respective G4 oligonucleotides.

Chapter 5 explores the dynamics of *spare-tire* isomers. The kinetics of all relevant folding and refolding trajectories for three *cMYC spare-tire* G4s (1234, 1245, and 2345) each involving four of the five possible strands have been unravelled. The folding kinetics are vastly different, with the fastest conformation (2345) folding four times ($\sim 400 \text{ h}^{-1}$) faster than the slowest conformation (1234). Interestingly, the temperature dependence is markedly different for each of the three isomers. For 1234, a positive apparent activation energy barrier is observed. Folding of this conformation involves the formation of *off-pathway* intermediates as kinetic traps. The intermediate state was spectroscopically scrutinized and a reasonable structural model could be proposed, based on experimental findings. 2345 shows optimized fast kinetics proceeding close to funnel-like, without observable kinetic trapping. The negative folding activation barrier indicates a mainly entropically driven folding process. 1245 features a long internal loop. The folding kinetics of 1245 show an unprecedented non-Arrhenius temperature dependence. It was shown that the 1245 conformation is a rare interjacent case for the two limiting regimes on the folding energy landscape observed for 1234 (pronounced kinetic partitioning) and 2345 (funnel-like folding). Refolding kinetics between the different *spare-tire* isomers complete the analysis of the individual pathways along the conformational energy landscape of the entire *cMYC* G4 ensemble. Refolding is slow ($0.16 - 0.30 \text{ h}^{-1}$) and results into a complete refolding into the 2345 conformation, which allows denoting 1234 and 1245 as long-lived *meta-stable* conformations.

Chapter 6 presents a comprehensive discussion of all experimental data within this thesis. The discussion further evaluates the findings of this thesis in the context of the recent literature. The bias in the comparison of crucial experimental conditions (e.g. K⁺-recruitment or possible pre-folded states) highlight the great advantages of the presented methodological approach over previously reported studies on G4 structural dynamics. Finally, a kinetic model is proposed that allows the disentanglement of key kinetic steps in for all relevant folding and refolding trajectories within the *cMYC* G4 ensemble. This model explains the different observed kinetics based on structural considerations for intermediate states and possible transitory ensembles.

1.2 Zusammenfassung

Die vorliegende Arbeit *Zeitaufgelöste NMR-spektroskopische Untersuchung konformationeller Dynamiken in DNA G-Quadruplexen* befasst sich mit der detaillierten biophysikalischen Untersuchung wichtiger strukturdynamischer Eigenschaften von nicht-kanonischen Nukleinsäure Sekundärstrukturelementen.

Im Genom aller eukaryotischer Lebewesen, insbesondere dem menschlichen Genom finden sich DNA-Sequenzabschnitte, die überdurchschnittlich Guanosin (G)-reich sind. Diese poly-G Abschnitte sind nicht zufällig im Genom verteilt, sondern häufen sich vermehrt in Genabschnitten, die besonders wichtig für die Regulation der Genexpression sind. G-reiche DNA-Sequenzen können unter geeigneten Umständen alternative Sekundärstrukturen ausbilden, die von der doppelsträngigen, kanonischen Watson-Crick Konformation abweichen. In Anwesenheit monovalenter Kationen können sich G-Nukleotide in einer Tetrade über Hoogsteen Interaktionen anlagern. Diese Tetraden können sich stapeln und dadurch sogenannte G-Quadruplexe (G4) ausbilden.

G-Quadruplexe haben durch ihre ungewöhnliche Struktur Einfluss auf die Transkription und Replikation, sowie DNA-Protein Interaktionen und die Genomstabilität insgesamt. G4 Strukturelemente sind dabei sehr dynamisch im zellulären Umfeld, sie werden ständig gefaltet, umgefaltet und entfaltet. Eine Vielzahl molekularer Mechanismen steuert und reguliert die Ausbildung von G4 und deren Einfluss auf die gesunde Zell-Homöostase. Eine Fehlregulierung dieser G4 Mechanismen ist assoziiert mit unter anderem der Pathogenese einiger Erbkrankheiten, epigenetischen Einflüssen, sowie der Karzinogenese verschiedener Tumorerkrankungen. Die Untersuchung der Dynamiken von Faltung und Umfaltung von G4, die an der Genregulation beteiligt sind, ist daher von fundamentaler Bedeutung.

Das menschliche *cMYC* Gen wird typischerweise als proto-Onkogen bezeichnet. Es kodiert für einen unspezifischen Transkriptionsfaktor, der bei einer Vielzahl von systematischen und soliden Tumorerkrankungen stark überexprimiert wird. Die zelluläre Konzentration des Genprodukts kann zu 90% über ein G4 *cis*-Element in der Promotorregion reguliert werden; das Protein selbst ist jedoch nicht mit herkömmlichen niedermolekularen Wirkstoffen angreifbar. Aufgrund dieser Erkenntnisse hat sich das G4 *cis*-regulierende Element im *cMYC* Promotor als vielversprechendes Wirkstoffziel für strukturbasiertes Wirkstoffdesign etabliert.

Der *cMYC* G4 hat die Möglichkeit verschiedene Konformationen einzunehmen, was auch für viele andere G4 gilt. Dieser Polymorphismus zeichnet sich typischerweise durch verschiedene Faltungstopologien aus. Im Falle des *cMYC* G4 kann man zusätzliche, nicht-konventionelle Formen der konformationellen Isomerie finden. Zum einen gibt es die Möglichkeit, dass bei einem G4, der aus drei Tetraden und vier intramolekularen Strangabschnitten (dreistöckiger G4) besteht, einzelne Strangabschnitte mehr als drei konsekutive G-Nukleotide besitzen. Dadurch können sich Faltungs-

1.2 Zusammenfassung

Isomere bilden, die sich durch Verschieben des Strangs relativ zum verbleibenden dreistöckigen Tetradengerüst ergeben. Man spricht von **G-Register Isomeren**. Eine zweite Möglichkeit der Strukturisomerie ergibt sich, wenn in einer Nukleotidsequenz mehr als vier G-reiche Strangabschnitte aufeinander folgen. Jeweils vier dieser Strangabschnitte können in unterschiedlicher Weise kombiniert werden, um ein G4 Isomer auszubilden. In jedem dieser so zustande gekommenen G4 verbleibt ein (oder mehrere) G-reicher Strangabschnitt, der im konkreten Isomer nicht zur Faltung verwendet wird. Diese zusätzlichen G-Stränge werden daher auch Ersatzräder (engl. *spare-tires*) genannt; man erhält ***spare-tire Isomere***.

Obwohl diese Formen des Polymorphismus, deren biologischer Kontext und die biophysikalischen Konsequenzen in Arbeiten von C. Burrows (2015) und A. Mittermaier (2016) erstmals umfassend beschrieben wurden, gab es bis zum Ausgangspunkt dieser Arbeit keine Kenntnisse über deren strukturelle Dynamik, den Faltungswegen und den zugrundeliegenden molekularen Mechanismen. Die Erkenntnisse, die über die konformationellen Dynamiken des *cMYC* G4 Elements gewonnen wurden und hier vorgestellt werden, sind zum einen grundlegender Natur. Die Faltung von G4 ist ebenso wie beispielsweise die Faltung von Proteinen ein komplexer und faszinierender Vorgang, der weder in der theoretischen noch in der experimentellen Beschreibung bislang hinreichend verstanden ist. Darüber hinaus erweitern die beschriebenen konkurrierenden Faltungskinetiken direkt das Verständnis der biologischen Funktionsweise und Aufrechterhaltung der Funktionsfähigkeit des *cMYC* G4 Elements in zellulärer Umgebung. In der Literatur gibt es zahlreiche Hinweise darauf, dass die *cMYC* G4 Sequenz unter zellulären Stressbedingungen leicht durch Oxidation der G-Nukleobasen geschädigt oder durch Mutationen verändert werden kann. Zudem wird das *cMYC* G4 Element von zahlreichen Proteinen erkannt und gebunden, wobei die Bindungsaffinitäten Konformations-spezifisch sehr stark variieren. Der ausgeprägte Polymorphismus ist also Teil der Regulation und Funktionalität des *cMYC* G4 Elements, wodurch die Adaption an eine veränderte zelluläre Umgebung gewährleistet wird.

Zeitaufgelöste Kernspinresonanz (engl. nuclear magnetic resonance, *NMR*) Spektroskopie ist eine bestens geeignete Methode, um die Dynamik von Biomakromolekülen mit atomarer Auflösung zu studieren. Die Proben können dabei unter physiologischen Bedingungen und auf einer breiten Zeitskala untersucht werden. Typischerweise werden Echtzeit-Messungen solcher Dynamiken als Relaxationsprozess eines Nicht-Gleichgewichtszustands zurück in einen Gleichgewichtszustand untersucht. Um solche Experimente durchführen zu können, braucht es geeignete Herangehensweisen für die Präparation eines Nicht-Gleichgewichtszustands. In dieser Arbeit wird eine neu erarbeitete Strategie vorgestellt, die es erlaubt, Einblick in die Faltungs- und Umfaltungskinetiken eines dynamischen Konformations-Ensembles nicht-konventioneller Strukturisomere der *cMYC* G4 DNA-Sequenz zu erhalten.

1.2 Zusammenfassung

Der Vollängensequenzabschnitt (hier: 22-mer) des *cMYC* G4 Elements umfasst fünf G-reiche Strangabschnitte, für diesen Abschnitt wurden drei Konformationen mit paralleler Faltungstopologie berichtet (D. Yang 2005, 2011, 2019). Im thermodynamischen Gleichgewicht unter physiologischen Bedingungen bildet sich fast ausschließlich diejenige Konformation, die sich aus den vier 3'-terminalen G-Strangabschnitten bildet. Zunächst wurde die Faltung dieser prädominanten *cMYC* Konformation untersucht, wobei eine 18-mer Oligonukleotidsequenz ohne den 5'-terminalen G-Strangabschnitt verwendet wurde. Diese Konformation liegt in einem Gleichgewicht zweier co-existenter Sub-Konformationen vor, die sich formal durch einfaches Verschieben eines Strangabschnitts ergeben (G-Register Isomere). Hierfür wurde ein ungefalteter Zustand erzeugt, indem K^+ -freie Bedingungen präpariert wurden. Durch sehr schnelle Zugabe von K^+ -Ionen kann die Faltung *in situ* induziert werden. Es zeigte sich, dass die beiden Sub-Konformationen parallel gebildet werden, wobei eines der G-Register Isomere zunächst kinetisch begünstigt wird und sich danach durch langsame Umfaltung das Gleichgewicht der Isomere einstellt.

Um tiefergehenden Einblick in die Umfaltung zu erhalten, wurde die Umfaltungskinetik direkt untersucht. Hierzu wurden photolabile Schutzgruppen (engl. *Photocages*) positionsspezifisch an bestimmten G-Nukleobasen (O^6 -(**R**)-NPE) angebracht. Die Schutzgruppen blockieren die Basenpaar-Interaktionen des Nukleotids, wodurch dieses sich nicht mehr an einer Tetradenbildung beteiligen kann. Die Photocages wurden jeweils an den Nukleotiden eingeführt, die nur in jeweils einem der G-Register Isomere an der Tetradenbildung beteiligt sind. Durch diese gezielte Destabilisierung konnten die Isomere getrennt und im gefalteten Zustand isoliert werden. Die so erhaltenen Konformationen wurden umfassend spektroskopisch charakterisiert. Es ergaben sich keine Änderungen im Vergleich zu den literaturbekannten NMR-Strukturen. Nach *in situ* Laser-Lichtanregung werden die Photocages abgespalten, wodurch das native, nicht modifizierte Oligonukleotid zurückerhalten wird. Es konnte beobachtet werden, dass sich das natürliche Gleichgewicht der Isomere zurückbildet und die Kinetik des Vorgangs konnte dekonvolviert werden.

Dieser Ansatz wurde ausgeweitet auf eine weitere G4 DNA-Sequenz, die in zwei unterschiedlichen G-Register Isomeren vorliegen kann (*hTERT*). Da sich hier (*hybrid* vs. *parallel*), anders als bei *cMYC* (*parallel* vs. *parallel*), die beiden Konformationen durch größere strukturelle Unterschiede auszeichnen, sollten Rückschlüsse auf vermutete unterschiedliche Umfaltungsmechanismen gezogen werden. Die Ratenkonstanten der Umfaltung beider G4 Systeme weichen jedoch nicht signifikant voneinander ab, was zunächst überraschend scheint. Da sich die Aktivierungsenergie aus dem geschwindigkeitsbestimmenden Schritt der gesamten Umfaltung ergibt, konnte nicht direkt von der apparenten Umfaltungskinetik auf nachgelagerte, schnellere kinetische Schritte geschlossen werden.

Ein Vergleich der kinetischen Daten der Echtzeit-NMR Experimente mit Experimenten, die auf Hystereseeffekten bei thermischer (Ent)-Faltung beruhen, ergab jedoch wesentliche energetische

1.2 Zusammenfassung

Unterschiede für die jeweiligen Umfaltungsprozesse. Dadurch konnten unterschiedliche Übergangszustände skizziert werden, die sich durch einen unterschiedlichen Grad der Entfaltung auszeichnen. Während der *hTERT* G4 fast vollständig entfaltet und anschließend in eine zweite Konformation rückfaltet, gleicht die Umfaltung des *cMYC* G4 vielmehr einem Verrücken des einzelnen Stranges, wobei die Gesamtstruktur weitestgehend erhalten bleibt. Die apparente Aktivierungsenergie beschreibt dabei das initiale Brechen der Wasserstoffbrückenbindungen für einen der G-Strangabschnitte.

Der Ansatz, das konformationelle Gleichgewicht durch Photocages transient zu stören, wurde daraufhin weiterentwickelt. Mehrere Photocages wurden an Nukleobasen in zentraler Position einzelner G-Strangabschnitte angebracht. Dadurch konnte eine ausreichende Destabilisierung erreicht werden, die die Faltung jedweder G4 Strukturen unterbindet. Somit wurde ein ungefalteter Zustand erzeugt, der unter ansonsten frei wählbaren, physiologischen Bedingungen besteht. Durch *in situ* Photolyse der Schutzgruppen konnte so die Licht-induzierte G4 Faltung unter konstanten Puffer- und Temperaturbedingungen untersucht werden. Dieser Ansatz wurde auf die Untersuchung der Faltungswege, die zu verschiedenen *spare-tire* Isomeren führen, fokussiert. Hierfür wurde die 22-mer Vollängensequenz des *cMYC* G4 Elements mit allen fünf G-reichen Strangabschnitten (von 5'-3', abgekürzt: 1-2-3-4-5) verwendet. Neben der prädominanten *cMYC* Konformation (mit den G-Strangabschnitten: 2-3-4-5, *cMYC*-2345) können so zusätzlich zwei weitere Konformationen gebildet werden, die im Gleichgewichtszustand unter physiologischen Bedingungen nur geringfügig populiert sind (*cMYC*-1234 und *cMYC*-1245).

Es zeigen sich signifikant langsamere Faltungskinetiken für die beiden Alternativ-Konformationen im Vergleich zur prädominanten Konformation. Eine temperaturabhängige Untersuchung dieser Kinetiken und anschließende Arrhenius-Analyse zeigte, dass den drei möglichen Konformationen drei fundamental unterschiedliche Faltungsmechanismen zugrunde liegen. Die prädominante Konformation *cMYC*-2345 zeigt eine negative Aktivierungsenergie für die spontane Faltung. Dieser Prozess scheint daher entropisch getrieben zu sein. Die Möglichkeit paralleler Faltungswege, über G-Register Sub-Konformationen erhöht die konformationelle Entropie dieses Konformations-Ensembles zusätzlich. Dadurch wird die Faltungskinetik deutlich beschleunigt; die Energielandschaft gleicht einem Faltungstrichter. (Trichter-artige Faltung). Die Bezeichnung „Trichter-artig“ zur Beschreibung des Faltungsweges der prädominanten Konformation darf dabei nicht missverstanden werden. Denn auch hier konnten parallele, konkurrierende Faltungswege kinetisch differenziert werden, die zu weiteren Sub-Konformationen (G-Register Isomeren) führen. Der Kippunkt für die distinkten Faltungswege der Sub-Konformationen liegt jedoch an einem weit fortgeschrittenen Zeitpunkt der Faltung. *cMYC*-1234 hingegen verzweigt sich bereits zu einem frühen Zeitpunkt in separate Faltungswege und bildet ein langlebiges Intermediat, das sowohl NMR-, als auch CD-spektroskopisch nachgewiesen werden konnte. Die CD-Signatur des Intermediats zeigt, dass die

1.2 Zusammenfassung

Konformation in einer relativ zueinander *anti-parallelen* Anordnung (5'-3'-Richtung) der G-Strangabschnitte faltet. Da die stabile Konformation aller untersuchten *cMYC* G4 *spare-tire* Isomere jedoch in jedem Fall eine *parallele* Anordnung aller beteiligten G-Strangabschnitte relativ zueinander aufweist, muss dieses Intermediat folglich wieder aufgebrochen werden. Die apparente Aktivierungsenergie ist daher positiv und repräsentiert die Umfaltungsbarriere, die aus dieser kinetischen Falle überwunden werden muss. Der Faltungsmechanismus kennzeichnet sich durch ein komplex verzweigtes Netzwerk verschiedener Makrozustände entlang der Faltungswege; die Faltung ist enthalpisch getrieben mit ausgeprägter kinetischer Beteiligung (engl. *kinetic partitioning*) aller Faltungswege. *cMYC-1245* zeigt ein unvorhergesehenes non-Arrhenius Verhalten. Diese Konformation hat eine ungewöhnliche Schleifenanordnung (1:6:1), mit einer besonders langen internen Schleife. Das führt dazu, dass der Faltungsmechanismus bei höheren Temperaturen mehr dem Trichter-artigen Mechanismus der prädominanten Konformation gleicht; bei niedrigeren Temperaturen bilden sich zunehmend kinetische Fallen aus, die unter Energieaufwendung aufgebrochen werden müssen. Die Faltung der *cMYC-1245* Konformation ist daher ein seltener Grenzfall zwischen den beiden Regimen (*rein entropisch, Trichter-artig* vs. *kinetic partitioning*). Dieses non-Arrhenius Verhalten deutet auf eine Faltungsabfolge hin, die der von DNA/RNA-Haarnadelstrukturen ähnelt. Die Ergebnisse verdeutlichen daher die komplexen und multiplen Faltungsmechanismen für G4. Die flachen konformationellen Energieoberflächen bzw. Energielandschaften führen dazu, dass unterschiedliche Faltungswege genutzt werden können und, dass sich eine Vielzahl unterschiedlicher Konformationen abseits eines globalen thermodynamischen Minimums ausbilden kann.

Um die Dynamiken zwischen den lokalen Minima der alternativen Konformationen (*cMYC-1234* und *cMYC-1245*) und der thermodynamisch begünstigten Hauptkonformation *cMYC-2345* zu untersuchen, wurde die bereits für G-Register Isomere genutzte Photocage-Strategie angewandt. Durch positionsspezifische Anbringung nur einzelner Photocages, konnten einzelne G-Strangabschnitte blockiert werden. Der Ansatz war erfolgreich, um die einzelnen, vollständig gefalteten Konformationen *cMYC-1234* und *cMYC-1245* in einer *meta*-stabilen Weise zu isolieren. Direkt nach der Photolyse-Reaktion ergibt sich dadurch ein Zustand, in dem ein vollkommen unmodifiziertes DNA-Oligonukleotid mit der natürlichen Sequenz des *cMYC* G4 in einer gefalteten Nicht-Gleichgewichtssituation vorliegt. Die exklusive Population dieser Faltungszustände unter nativen, physiologischen Bedingungen konnte so auf einzigartige Weise demonstriert werden. Beide Konformationen falten sich über mehrere Stunden hinweg zur Hauptkonformation um. Die Umfaltungskinetiken der beiden Konformationen weichen dabei stark voneinander ab, bei vergleichbarer Aktivierungsenergie für diesen Umfaltungsprozess. Die vergleichbaren Aktivierungsenergien weisen erneut darauf hin, dass der gleiche geschwindigkeitsbestimmende Schritt, nämlich die Brechung der Wasserstoffbrückenbindungen einer der G-Strangabschnitte,

1.2 Zusammenfassung

zugrunde liegt. Die unterschiedlichen Umfaltungskinetiken, sowie der Vergleich zur Umfaltung der *cMYC* und *hTERT* G-Register Isomere, erlauben eine Skizzierung des Übergangszustands. *cMYC*-1234 gleicht dem Umfaltungsmechanismus der G-Register Verschiebung; es wird nur einer der G-Strangabschnitte aufgebrochen und ausgetauscht. *cMYC*-1245 hingegen gleicht einem Umfaltungsmechanismus, bei dem ein deutlich höherer Entfaltungsgrad notwendig ist. Die langsamere Umfaltungskinetik lässt sich dadurch erklären, dass sich ein teilentfalteter Zustand neu orientieren muss, um in die Hauptkonformation übergehen zu können. Da die Rückfaltung des Ausgangszustands jedoch vermutlich hinreichend schnell verläuft, ist diese Konformation primär kinetisch gefangen.

Zusammenfassend kann festgestellt werden, dass es insgesamt erstmalig gelungen ist, die Kinetiken der wesentlichen Faltungs- und Umfaltungswege entlang der konformationellen Energielandschaft des *cMYC* G4 Elements zu untersuchen. Das komplexe, dynamische Zusammenspiel aller relevanten, nicht-konventionellen isomeren G4 Strukturen konnte entwirren und umfassend experimentell beschrieben werden. Der dafür weiterentwickelte Ansatz über konformationelle Selektion mit Hilfe photolabiler Schutzgruppen hat dabei experimentelle Einblicke erlaubt, die bislang nicht zugänglich waren. Die Strukturen und Faltungszustände, die mit den chemisch modifizierten Oligonukleotiden erhalten und isoliert wurden, sind umfassend spektroskopisch untersucht worden und konnten anhand der in der Literatur bekannten Strukturen verifiziert werden. Die Anwendung verschiedener spektroskopischer Ansätze und deren Kombination mit weiteren biophysikalischen Methoden hat eine Methoden-unabhängige Validierung der erhaltenen kinetischen und thermodynamischen Daten ermöglicht.

2 General Introduction

2.1 Structure and Dynamics of DNA G-Quadruplexes

*“DNA neither cares nor knows. DNA just is.
And we dance to its music.”*

Richard Dawkins, *River out of Eden*³

2.1.1 Molecular Structure and Conformations of DNA

Deoxyribonucleic acid (DNA) is now long known as the carrier of genetic information in all living organisms.⁴⁻⁶ The legendary discovery of its molecular structure by Franklin, Wilkins, Watson and Crick has been one of the major breakthroughs in modern science and laid the foundation for molecular biology in the way we understand it today.⁷⁻⁹ Though the architecture of DNA molecules up to the macroscopic level is of vast complexity, it can be disassembled to only four distinguishable primary building blocks. Figure 1 shows the DNA nucleotides/nucleobases adenosine/adenine (A), cytidine/cytosine (C), guanosine/guanine (G) and thymidine/thymine (T) and their respective atom-wise numbering. All nucleobases are attached to a deoxyribose ring via an N-glycosidic bond. These nucleosides are linked via phosphodiester bonds and constitute the primary structure of a DNA oligo-/polymer.

Each of the nucleotides shows a specific hydrogen-donor/acceptor pattern, that allows interactions between nucleobases. In their structural model, Watson and Crick proposed the most stable interactions to be formed by purine-pyrimidine base pairs A-T and G-C (Figure 1). This arrangement leads to the formation of *anti*-parallel, helical double-strands with distinct structural parameters. The native, right-handed conformation of the DNA is known as B-form, since Rosalind Franklin found a dehydrated conformation, which she referred to as A-form.¹⁰ The DNA conformational alphabet was further extended with the discovery of a rare C-form¹¹ (Li⁺-DNA) and the left-handed Z-form^{12,13} as deviating helical conformations.

In a first approximation, this seems to be the end of the story – in terms of molecular structure; the complexity of structures for genomic DNA is less diverse compared to the cognate ribonucleic acid (RNA). Basically, genomic DNA is *in situ* synthesized as a double strand and thus *born* in its thermodynamic most stable conformation. Other than that, most of the RNA molecules usually appear as single stranded oligomers, which allows entering plenty of folding pathways along the energy landscape. In this regard, RNA is by nature much more prone to form different secondary structures such as stem-loops and hairpins and complex tertiary structures.¹⁴⁻¹⁸

2.1 Structure and Dynamics of DNA G-Quadruplexes

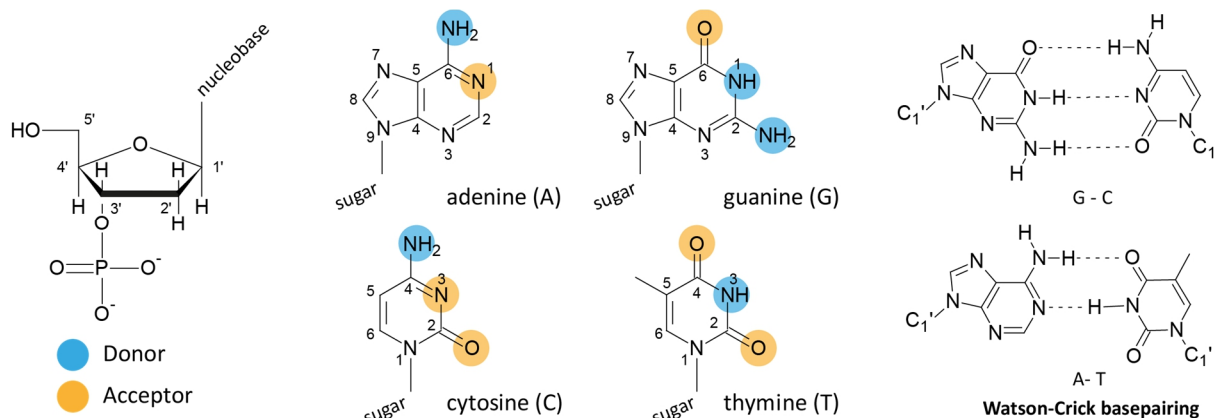


Figure 1: Nucleotides as basic building blocks of DNA: 2'-deoxyribose with 3'-phosphate and 5'-OH group that allow polymerization via phosphodiester bonds (left). Purine (adenine, guanine) and pyrimidine nucleobases (cytosine, thymine) (middle) are attached via 1'-N^{1/9}-glycosidic bonds. Donor (blue)/acceptor (orange) patterns at the Watson-Crick interface are indicated. Watson-Crick base pairs between purine-pyrimidine nucleobases (A-T and G-C, right)

For single stranded DNA oligomers, the game changes and many secondary structures become possible, similar to those known for RNA. Many DNA oligomer sequences have been designed/selected (mostly with **S**ystematic **E**volution of **L**igands by **E**Xponential Enrichment, SELEX)^{19–21} especially to obtain certain structures of all kind or structure-related functions such as DNA hairpins²², DNazymes^{23–26} or DNA-Aptamers^{27–29}. However, some non-B-form DNA conformations (Figure 7) are also found naturally e.g. three-way junctions^{30–33}, Holliday junctions^{34,35,36}, cruciform DNA^{37–40}, hairpin DNA^{41,42} or looped DNA (D-, R-, T-loops)^{16,43}.

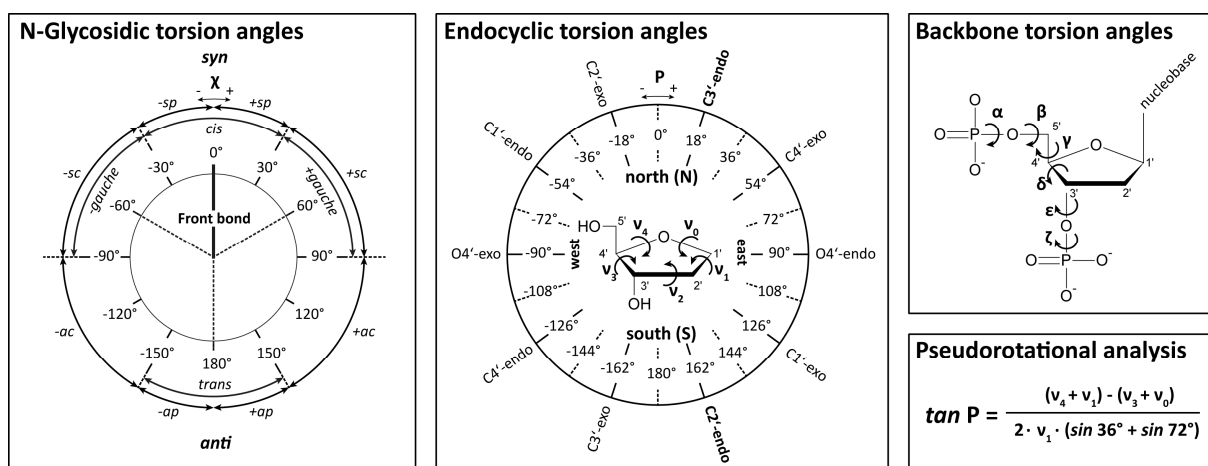


Figure 2: Nomenclature and definition of the torsion angles for N-glycosidic bond (χ) rotation (left, Newman projection), endocyclic pseudorotation/(sugar pucker) (middle) and backbone rotation (upper right). Figures adapted according to cited references.^{44,45} Abbreviations for Klyne-Prelog notation: *sp*=synperiplanar, *sc*=synclinal, *ac*=antiperiplanar, *ap*=antiperiplanar. **P** is defined as pseudorotational angle (lower right).

Figure 2 shows an overview for the parameters that determine the DNA conformation depending on the rotation or *pseudorotation* about given torsion angles between the ribose and the nucleobase, within the ribose ring and along the DNA backbone. The thermodynamic most favourable conformations for the N-glycosidic bond angle (*syn/anti*) and the sugar pucker (2'-endo/3'-endo) are shown in Figure 3.

2.1 Structure and Dynamics of DNA G-Quadruplexes

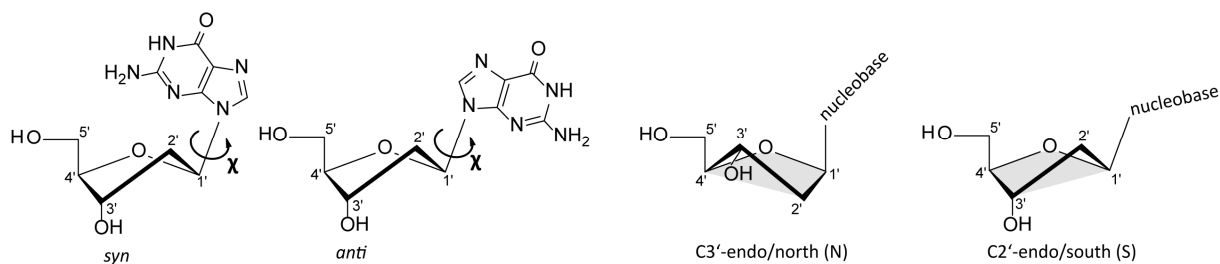


Figure 3: Preferred N-glycosidic conformations: *syn* ($0^\circ < \chi < \pm 90^\circ$) and *anti* ($\pm 90^\circ < \chi < 180^\circ$), (left); and preferred sugar pucker 3'-endo (N) or 2'-endo (S), (right). B-DNA adopts the C2'-endo/(S) conformation, while A-DNA adopts the C3'-endo/(N) conformation. In duplex DNA N-glycosidic torsion angles are all *anti*. The sugar pucker affects the orientation of the nucleobase relative to the phosphate backbone.

2.1.1 Non-Canonical DNA Structures

Certain alignments of these parameters and the phosphate backbone make other interaction sites at the nucleobases accessible and alternative, non-canonical base pairing patterns become possible.^{46,47} Hoogsteen hydrogen bonding⁴⁸ is of special importance and creates an own set of possible base pairs (Figure 4) that can also form in canonical duplex DNA.⁴⁹ In regions with homo-purine (Pu) and homo-pyrimidine (Py) strands, the formation of different non-canonical *paranemic* DNA structures have been discussed.^{50,51} In mirrored tandem repeats one Py-strand can fold back (3'-Py: H- γ 3 or 5'-Py: H- γ 5)⁵² and form an intramolecular triplex via Hoogsteen interactions, known as H-DNA (Figure 7).^{39,52-57}

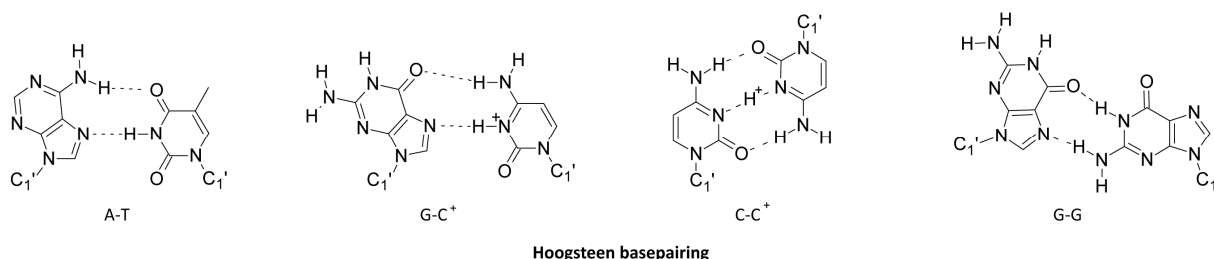


Figure 4: Hoogsteen base pairs between purine-pyrimidine (A-T and G-C⁺), pyrimidine-pyrimidine (C-C⁺) and purine-purine (G-G) nucleobases. The formation of hemiprotonated base pairs that involve C⁺ is pH-dependent (pK_a(N3)=4.58, isolated cytosine).⁵⁸

Already back in 1962 Gellert *et al.*⁵⁹ could show that GMP in presence of monovalent cations forms planar G-tetrads via Hoogsteen interactions. In G-rich DNA sequences the planarity of the G-tetrads helps to stack upon each other stabilized by π - π -interactions to form structures known as G-quadruplexes (G4), first reported by Sen and Gilbert in the late '80s.^{60,61} The recruitment of monovalent cations is a pre-requisite for the stability of this G4 structure, due to electrostatical reasons (Figure 5, Figure 7).⁶²⁻⁶⁵ Evidence that these DNA G4-structures really do form *in vivo* under cellular conditions is undisputed.⁶⁶⁻⁶⁸ The strong evidence for *in vivo* RNA G4 formation⁶⁹⁻⁷¹ however is still controversially discussed.⁷²

2.1 Structure and Dynamics of DNA G-Quadruplexes

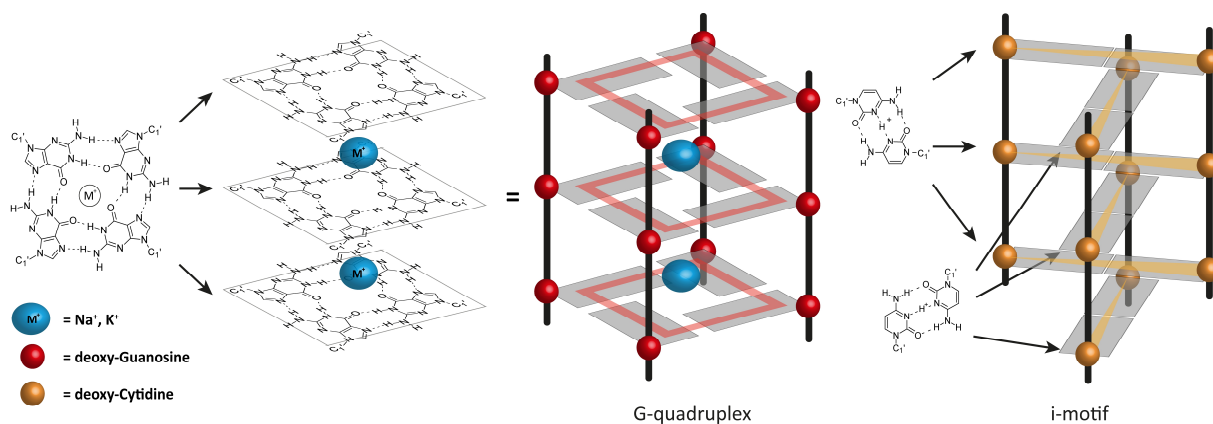


Figure 5: Three G-tetrads form a G-quadruplex (G4) structure via π - π stacking and stabilization with monovalent cations (mainly Na^+ or K^+), (**left**). Schematic representations of four-stranded DNA structures (**right**) nucleotides are represented as coloured spheres with planar nucleobases; hydrogen-bond connectivities are indicated.

Most comparable to G4s, Cytidines can form four-stranded structures, known as i-motif.^{73–82} The overall structure⁸³ is build up by stepwise alternating diagonal hemi-protonated C^+-C base pairs. The base pairs are stacked with offset, so that from a top view each 2+2 crossing base pairs define a jagged tetrad (Figure 5, Figure 7). DNA i-motif structures have also been detected in human cells.⁸⁴

Canonical B-form DNA is by far the most important non-regulatory conformation, since it is the conformation of chromosomal DNA that is adopted to guarantee stability for the genes. For gene regulation and additional functionality in gene expression G-quadruplexes have now emerged as most important non-B-form DNA conformation. This is highlighted by (December 2020) more than 8731 publications in the *Web of Science* (Clarivate Analytics; Figure 6, left), increasing relative share of publications in the nucleic acid field of research (Figure 6, right) and more than 421 structures in the Protein Data Bank (RCSB PDB, “G-quadruplex”).

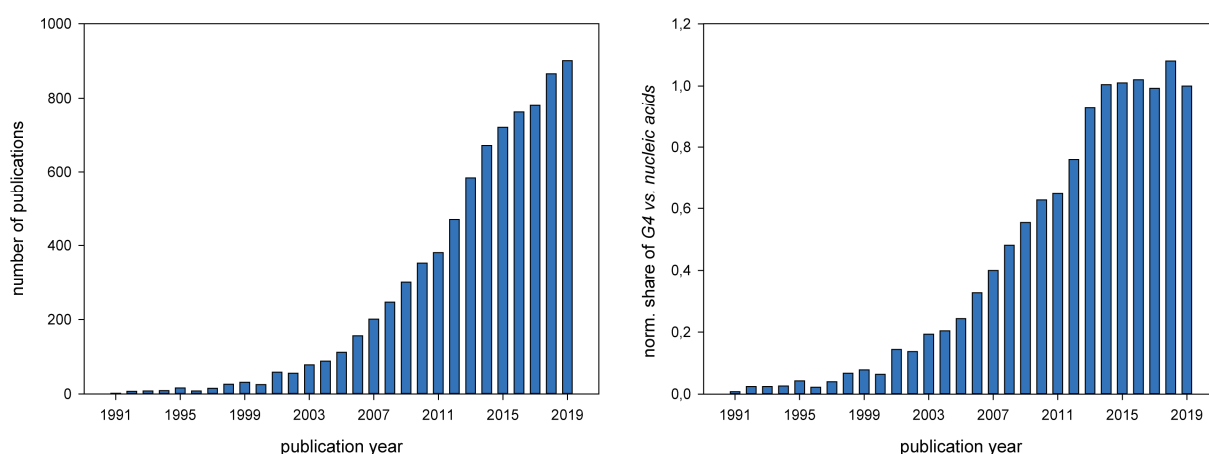


Figure 6: Number of G-quadruplex related publications from 1991-2019 (“G-quadruplex”). **Left**: total numbers, **right**: in relation to the total number of nucleic acid publications (“nucleic acids, or DNA, or RNA”, normalized share).

2.1 Structure and Dynamics of DNA G-Quadruplexes

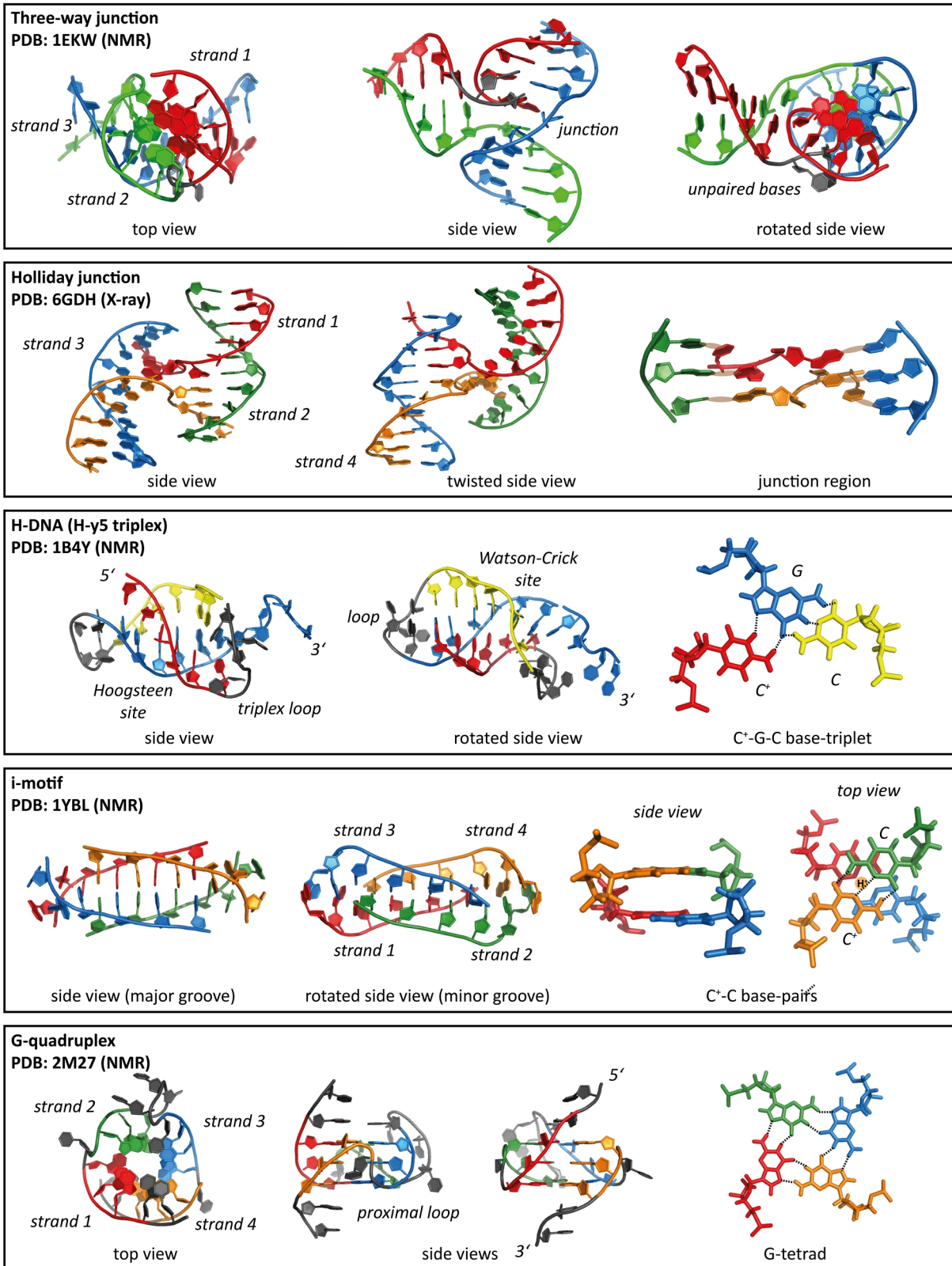


Figure 7: Overview of non-canonical (non-B-form) DNA structures. Representative PDB-structures are shown for three-way junctions (PDB: 1EKW)³⁰, four-way junctions (Holliday junction, PDB: 6GDH)³⁶, triplex DNA (H-DNA, PDB: 1B4Y)⁵² and tetraplex DNA (i-motif, PDB: 1YBL⁷⁵; G-quadruplex, PDB: 2M27⁸⁵).

2.1 Structure and Dynamics of DNA G-Quadruplexes

2.1.2 Canonical Structural Polymorphism

“Folding” of nucleic acids in general, describes the process of formation of secondary and tertiary interactions that define the structure of the macromolecule. A folded state can thus be defined as a state that represents a thermodynamic minimum on the energy landscape. This state can still be highly dynamic and is not necessarily the most stable state, but in principle, it can be described with defined coordinates. In the context of G-quadruplexes, folding usually refers to the topology of a specific conformation. In this way, the directionality of the DNA-backbone and the relative orientation of the four strands to each other defines the G-quadruplex fold. G-quadruplexes can be formed both inter- and intramolecular which results in tetra-, bi- or unimolecular G-quadruplexes. Irrespective of their molecularity, herein, a G-quadruplex is defined as a four-stranded structure, referring to the four G-tracts that constitute the G-quadruplex core structure. These strands can be oriented in an *all-parallel*, (3+1) or hybrid 2+2) *anti-parallel* conformation, which force different nucleobase conformations and tetrad polarities (Figure 8); in the case of uni-/intramolecular G-quadruplexes, the linking loop sequences are forced in distinct arrangements.⁸⁶ The structural key features depicted in Figure 8 divide the conformational landscape in fundamentally separated subsets of conformations. These canonical G4 structures are in part predictable from sequence.⁸⁷

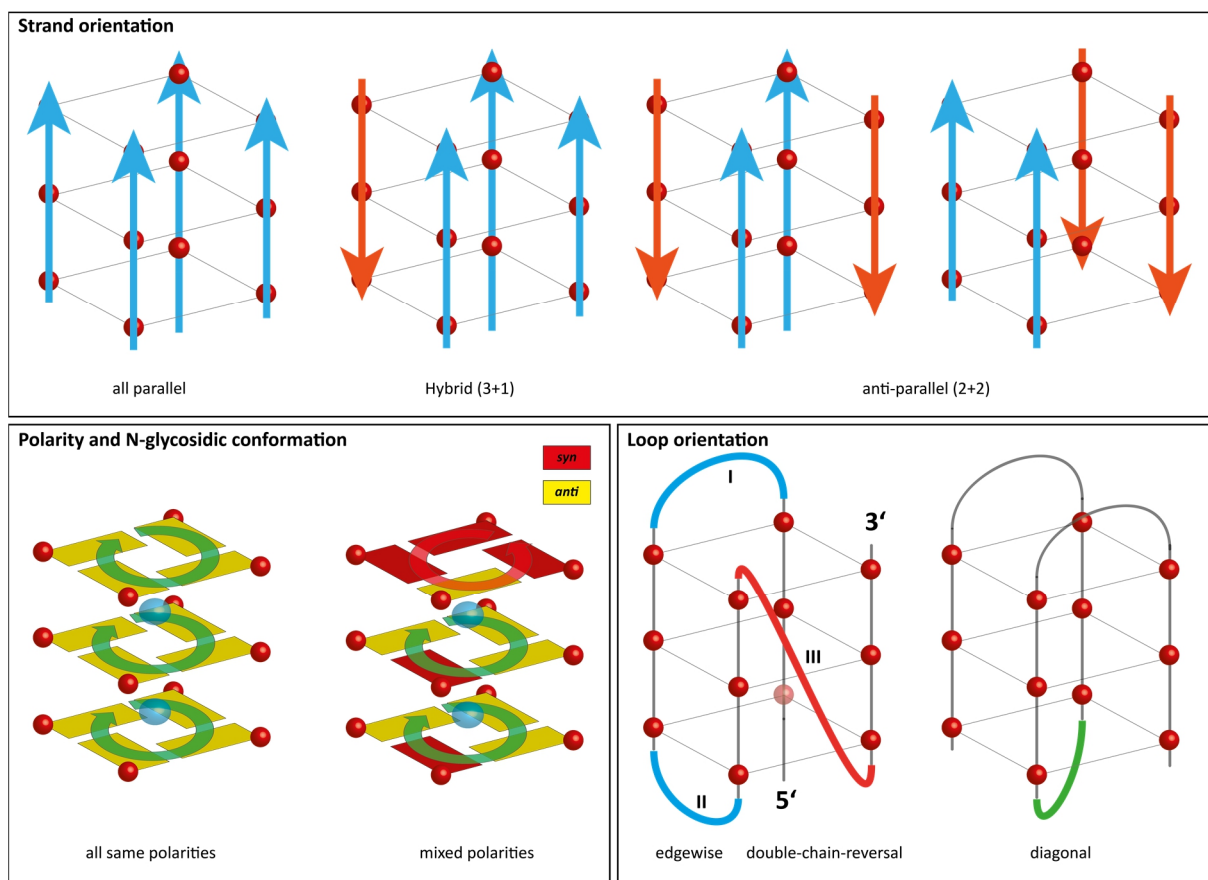


Figure 8: Canonical structural features that determine a G4 structure: strand orientation/directionality (relative orientation shown in orange/blue), tetrad polarity and N-glycosidic conformation (*all-parallel* G4s are in *all-anti* conformation) and loop orientation. The three constituting loops are numbered in 5'-3' direction.

2.1 Structure and Dynamics of DNA G-Quadruplexes

The depicted ($n=3$)-layered motifs are the most common G4-motif. In general however G4s with $n \geq 2$ ⁸⁸ layers can form stable structures, depending on the nucleotide sequence ($n=2$ ⁸⁹⁻⁹², $n=4$ ^{93,94}).

2.1.3 Non-Canonical Structural Polymorphism

Besides the canonical structural features that define the G-quadruplex structure in terms of folding topology, sugar conformation and loop orientation, additional non-canonical features add layers of complexity to the structural polymorphism.⁹⁵⁻⁹⁸ Especially along G-rich sequences that do not fit the consensus G-quadruplex sequence (2.3.1) additional structural isomerism can occur (Figure 9), which has broadened the definition of G4 structural parameters.

G-tracts with more than n ($n+x$, with n = number of tetrads in the folded G4) subsequent G-residues can be shifted relative to the tetrads to incorporate different G-residues into the G4-core (**G-register shift**). This kind of isomerism is possible for many reported G4 sequences (e.g. *cMYC*⁹⁹⁻¹⁰¹, *VEGF*^{85,101} and *hTERT*^{102,103}). Poly-G stretches of up to 30 nucleotides found e.g. in *C. elegans*^{104,105} have been shown to adopt distinct parallel G4 structures.¹⁰⁶ Stable G4 structures with $n-1$ G-tracts have been reported recently. Here, the G4-core lacks a single position in a tetrad, but the remaining triad is stable enough to maintain a stable G4 structure.^{107,108} This **G-vacancy sites** can be filled up with guanine metabolites which mimics an ideal binding pocket with specific molecular recognition.^{109,110}

If more than four G-stretches are nearby in a DNA sequence, also different G-tracts can be incorporated into the G4 structure. This feature is known as **spare-tire isomerism** and also has been reported for many G4 sequences (*cMYC*¹¹¹⁻¹¹³, *VEGF*¹¹³). If the number of G-tracts increases in longer G-rich sequences, multiple, stacked¹¹⁴ G-quadruplexes can be formed or potentially switch between hairpin-G-quadruplex arrangements. This has been examined especially for the *hTERT* core promoter.^{102,103,115-117} Further increasing the number of G-tract repeats can result in the formation of multimers,¹¹⁸ called G-wires.^{119,120} These nanostructures have been visualised with atomic force microscopy (AFM) in *Tetrahymena* telomeres with $G_4T_2G_4$ repeats.¹²¹ The concatenation of stacked G-quadruplexes can be achieved in different ways and results in highly polymorphic types of polymer structures.

Finally, peculiar loop and strand arrangements yield non-canonical structural features like bulges¹²²⁻¹²⁵, hairpin loops^{117,126}, hairpin loops within a bulge¹²⁷ or snap-back motifs^{91,128-131}. All of the diverse canonical and non-canonical G-quadruplex structures share a high similarity in the overall G4-core constitution. However, the schematic representation neglects the fact that G4s are indeed helical structures. Typically they are right-handed, but more recently structures of left-handed G4s have been reported (Z-G4).¹³²⁻¹³⁵

2.1 Structure and Dynamics of DNA G-Quadruplexes

The group of J. Plavec has reported on a series of novel tetrahelical structures from tandem repeats of alternating GGG and GCG tracts with non-canonical G-A and G-G base pairs as well as from AGCGA-rich DNA with GAGA- and GCCG-quartets.¹³⁶⁻¹³⁹ They also characterized a G-quadruplex structure from G_4C_2 repeats that features different stacked C-C base pairs.⁹⁴ The formation of G4s in expanded G_4C_2 hexanucleotide repeats in the *C9orf72* gene is strongly linked to neurodegenerative diseases as amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD).^{140,141}

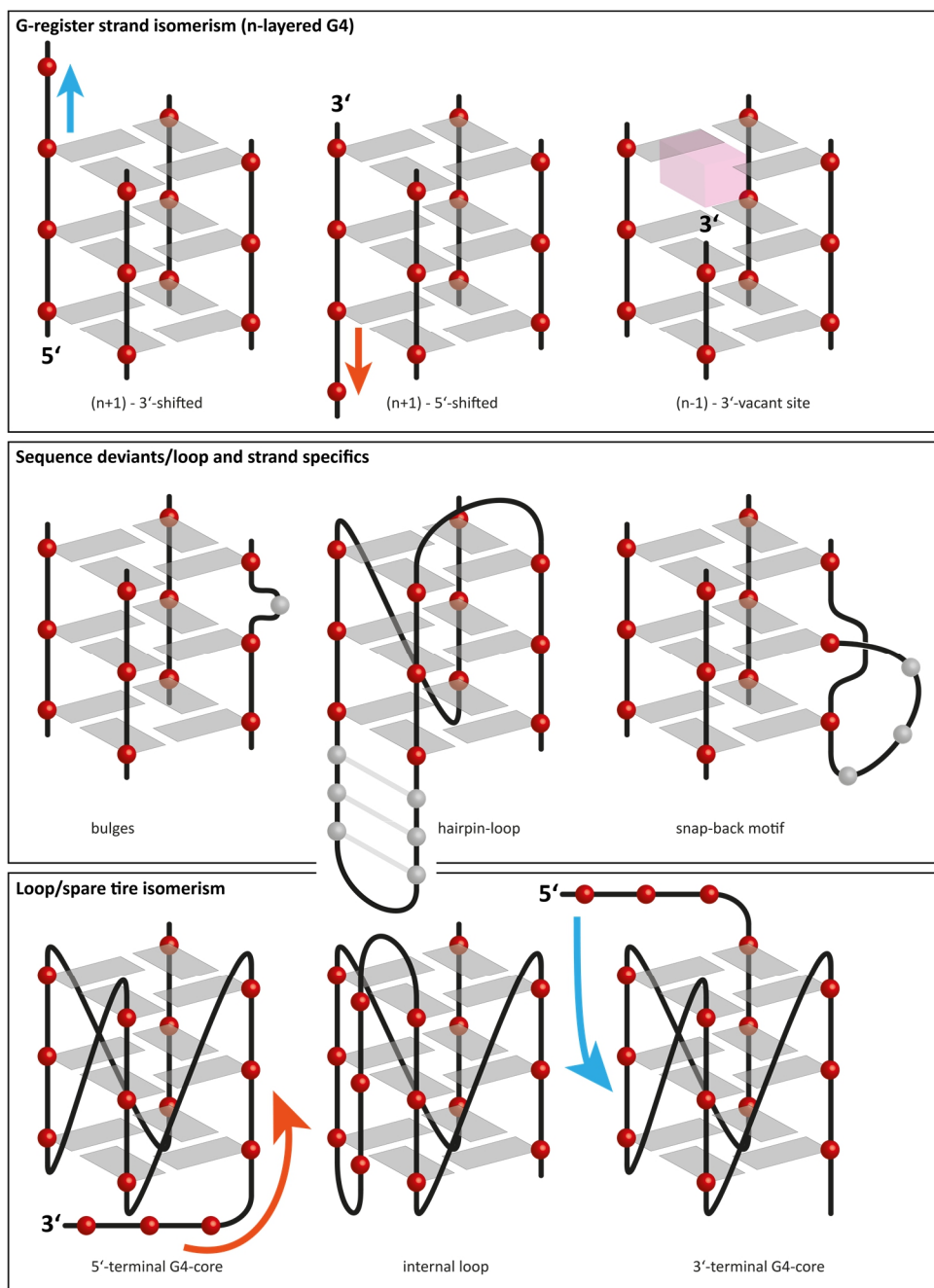


Figure 9: Non-canonical structural features that broaden the regular conformational space of G4 structures. **Upper part:** G-stretches that have $(n+x)$ G-nucleotides can form $(x+1)$ n -layered isomers by formally shifting the G-registers; $(n-1)$ -G-stretches can form G4s with G-vacancy sites in the respective G-tetrad. **Middle:** rare structural elements that can be found in G4 motifs. **Lower part:** Loop isomerism in G4 forming sequences with $4+x$ G-stretches.

2.1 Structure and Dynamics of DNA G-Quadruplexes

2.1.4 Conformational Dynamics

The pronounced structural polymorphism is linked to inherent conformational dynamics (Figure 10).¹⁴² Refolding and topology exchange has been investigated for the human telomeric G4^{143,144} as well as for the human telomerase promoter G4 (*hTERT*)^{102,145,146}. After initial folding in two different hybrid conformations (see also chapter 2.2.3), the telomeric G4s refold via an ensemble of partially unfolded states into the conformational equilibrium.¹⁴⁴ The partially unfolded states remain as long-lived intermediates for several hours. The *hTERT* G4 co-exists in two conformations, a hybrid and a parallel. The parallel conformation is populated nearly exclusively at higher temperatures, while at ambient temperatures the hybrid conformation is slightly more populated (60%). Nußbaumer *et al.*¹⁴⁵ inserted 8-¹³C-guanosines into conformation-specific positions and monitored the relaxation back to the conformational equilibrium after a temperature-jump with BEST-TROSY ¹H-¹³C-HSQC. They found a slow refolding rate for the interconversion of the two conformations (0.72 h⁻¹, 90 mM K⁺, 298 K).

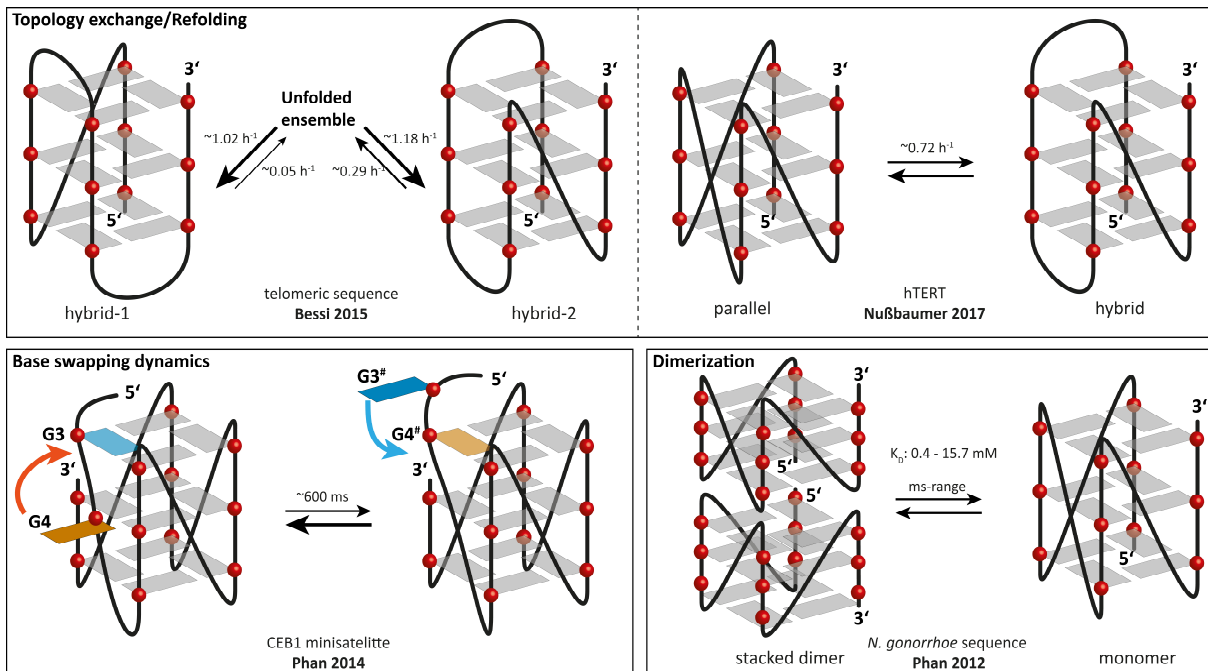


Figure 10: Different types of conformational transitions in folded G4s: **topology exchange** that involves complete refolding between two different conformations, **base swapping** dynamics of G-residues in a G-tetrad and **dimerization** dynamics between a stacked dimer and the folded monomer.

Phan *et al.*⁹¹ have studied the conformational dynamics of a stacked dimeric G4 from the human CEB1 minisatellite.⁹¹ Using NOESY and ROESY experiments they found conformational exchange at the 5'-end in the millisecond timescale (~600 ms). They propose a model for base swapping dynamics of two G-nucleotides that exchange in tetrad formation. Such guanine-flipping dynamics are involved in the unwinding mechanism of G4s by RecQ helicases.¹⁴⁷ In another study, the group of A.T. Phan investigated the dynamics of a parallel conformation that dimerizes via tetrad stacking on the 5'-5' interface.^{114,148} The K_D for the dimer formation is concentration dependent ($[DNA]$ and $[K^+]$) and sequence specific, but overall in the low mM-regime (approximately in the range of milliseconds).

2.1 Structure and Dynamics of DNA G-Quadruplexes

Loop dynamics of a CEB25 G4 structural ensemble have been analyzed in the sub- μ s regime with MD simulations using NMR RDC-restraints.¹⁴⁹ Conformational dynamics that are associated to G-register exchange and *spare-tire* exchange have been described as a consequence of thermodynamic and biological considerations.^{101,113} However, interconversion dynamics for these non-canonical structural features remain largely elusive and will be investigated and discussed in detail as a result of this thesis. The main thermodynamic parameters that have been reported so far on dynamics of G-register isomers are summarized in the following section. The observation of thermal hysteresis in the melting and annealing curves for an ensemble of G-register isomers in the *cMYC* G4 forming sequence (see chapter 2.4) allowed an indirect approximation of their exchange rates.^{100,101} Figure 11 shows an overview of the main parameters (ΔG_{25} , K_{ex} and ΔH reported at 5 mM $[K^+]$, E_a at 2 mM $[K^+]$). Chapter 4 of this thesis presents experimental results for G-register exchange dynamics in the context of promoter G4 sequences (in particular *cMYC*).

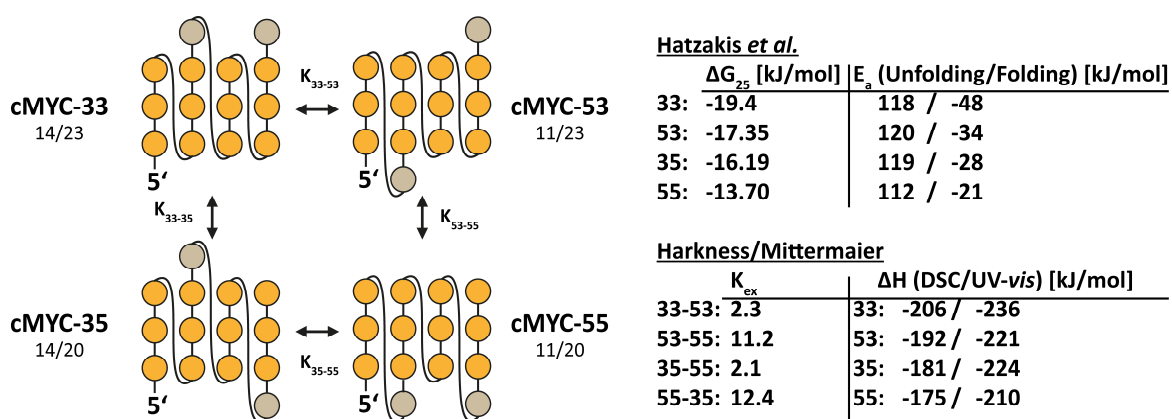


Figure 11: **Left:** schematic representation of possible G-register isomers in the *cMYC* G4 (adapted from Harkness and Mittermaier¹⁰¹). The numbering shows both the G-register nomenclature by Harkness and Mittermaier used throughout this thesis and their equivalent by Hatzakis *et al.*¹⁰⁰. **Right:** Summary of the main experimental findings for ΔG_{25} , K_{ex} , ΔH (5 mM $[K^+]$) and E_a (2 mM $[K^+]$).

Immediately prior to the submission of this thesis, Mittermaier *et al.* published a paper that outlines the consequences of an increasing length of telomeric repeat $(TTAGGG)_n$ sequences with increasing number of adjacent G-tract.¹⁵⁰ They describe a thermodynamic and kinetic folding frustration that arises from a competitive incorporation of G-tracts into contiguous G4s with negative cooperativity. The simulated folding frustration for a 32-repeat sequence ($n = 32$) is based on global analysis of rapid thermal melting experiments with thermal hysteresis (as also discussed above). The frustrated folding energy landscape results in partially unfolded and potentially misfolded parts of the DNA chain. This observation has implications for possible related conformational dynamics that would be required to straighten up the entangled G4 chain. It is conceivable that this process involves a step-by-step *spare-tire* exchange refolding that evolves through the DNA chain. Chapter 5 of this thesis presents experimental results for *spare-tire* exchange dynamics in the context of promoter G4 sequences.

2.2 Folding Dynamics of DNA G-Quadruplexes

*“Thus, a pathway of folding means
that there exist a well-defined sequence of events
which follow one another [...]”*

Cyrus Levinthal, How to fold gracefully^{151,152}

2.2.1 Folding Energy Landscapes and Folding Pathways

The observation that protein folding is faster than hypothesized from theoretical considerations (*Levinthal's paradoxon*) led to the proposal of a folding funnel.^{1,151,152} In the funnel-like folding energy landscapes of proteins, a deep well represents the native state of a folded protein (Figure 12). The fast and apparently barrier-free folding of proteins seems contradictory, given the complex nature of protein structures.^{153,154}

However, it is exactly this structural complexity and the residue diversity (20 amino acids in proteins) that is now commonly assumed to guide an optimized folding. Thus, the native structure of proteins is likely to be pre-defined by its primary sequence (*Anfinsen's dogma*).^{155,156} Following this, folding energy landscapes of non-canonical DNA structures must certainly be more complex, because only four residues (in G-quadruplexes only one) determine the specificity of local contacts. G-quadruplex forming oligonucleotides are built from G-rich sequences; hence, the poor nucleotide dispersion is self-defining.¹⁵⁷

Indeed, folding of nucleic acids is fundamentally different and folding of DNA G-quadruplexes in general is a slower process compared to proteins. The folding energy landscapes of nucleic acid oligonucleotides are not funnel-like but rough with flatter wells (Figure 12). In particular, the folding energy landscapes of DNA G-quadruplexes are complex, with numerous competing basins of attraction. This results in a pronounced structural polymorphism, the coexistence of different folded states and different competing folding pathways. Thus, in many cases it is not appropriate to speak about “native states”, since many nucleic acid sequences are prone to adopt several competing native folds.¹⁵⁷

The coexistence of different folded states is a clear sign of a *kinetic partitioning* folding mechanism for G-quadruplexes. However, even in the absence of concurrent conformations multiple folding pathways must be assumed. This follows from the consideration that no restraints do apply for the conformational space during folding that would limit the complexity of possible sub-states. Some intermediate states are long-lived and give clear implications for kinetic traps during folding; others are short lived and have not been detected directly in experimental studies.¹⁵⁷

2.2 Folding Dynamics of DNA G-Quadruplexes

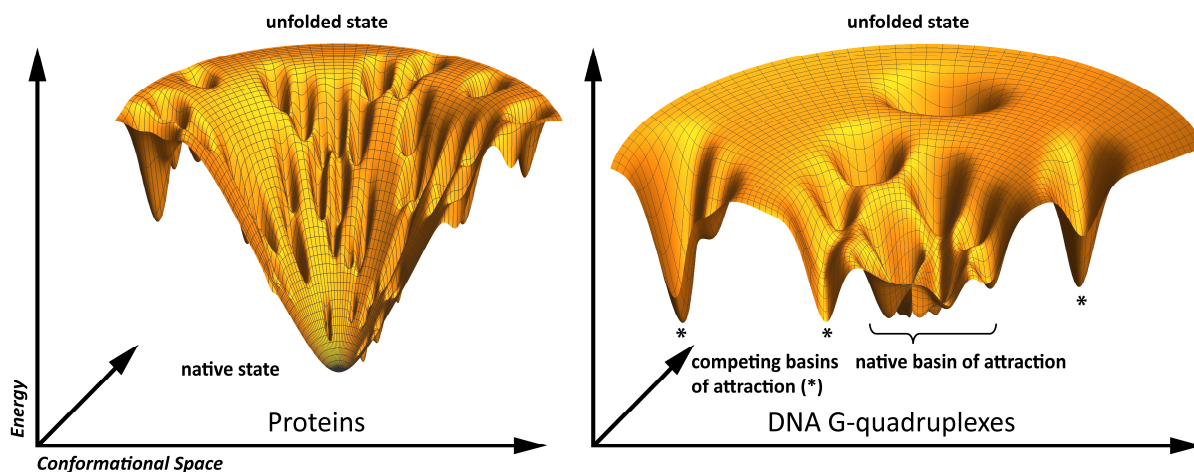


Figure 12: Representation of folding energy landscapes. **Left:** Folding-funnel as it is commonly proposed for protein folding. **Right:** Folding energy landscape with flat wells and competing basins of attraction, proposed for G-quadruplex folding. It is commonly assumed that proteins and nucleic acids follow folding pathways along their energy landscapes. Folding dynamics are typically described as funnel-like (**left**, proteins) or with kinetic partitioning (**right**, oligonucleotides).

Here, it must be noted that the definition of unfolded, partially folded, folded and misfolded states is an arbitrary definition, since any given oligonucleotide orientation is just a Boltzmann distributed ensemble over the available conformational space under given conditions at thermodynamic equilibrium. One has to be aware that any folding or refolding process is affected by external conditions and the physical nature of its starting point. This holds especially true for ensembles of unfolded states that can be prepared in different ways and lack a clear definition.^{157–160} From an energetic perspective, thermal denaturation marks a complex ensemble of high-entropy states, while e.g. force unfolding yields low-entropy states. The complexity of unfolded states can also be visualized easily with the *syn*- and *anti*-patterns of G4 forming oligonucleotides. In an unfolded oligonucleotide, every G-residue can adopt either of the two glycosidic conformations, which sums to $2^{12} = 4096$ combinations in a 3-layered G4. If a certain combination is productive or not, meaning if it is *on*- or *off*- the folding pathway, depends on the thermodynamic stability of the finally folded conformation. Hence, a description of the entire folding process is a seemingly impossible task for computational methods, as the starting point in these simulations is hard to define. Nevertheless, both experimental and theoretical studies have proposed numerous folding intermediates.^{161–163} Again, it depends on the underlying definition, if those are *off*- or *on*-pathway.

Recent progress in computational studies and MD simulations has helped to describe and outline the conformational energy landscapes and trajectories of DNA G4 with major contributions from the group of J. Šponer.^{157,161–166} Experimental studies to investigate G4 folding have been published using mass spectrometry^{167,168}, CD-spectroscopy^{169–171}, NMR-spectroscopy^{144,172}, smFRET microscopy^{173,174} and force spectroscopy with magnetic tweezers^{175,176}.

2.2 Folding Dynamics of DNA G-Quadruplexes

2.2.2 Experimental Approaches to Measure Folding Kinetics

Many experimental approaches to investigate folding kinetics of G-quadruplexes rely on rapid mixing with monovalent cations to induce isothermal folding.¹⁷⁷ Otherwise, thermal denaturation allows investigating folding processes under otherwise physiological conditions. Kinetic information on (un)-folding events of G-quadruplexes can be obtained from thermal hysteresis, when samples are rapidly heated or cooled.^{100,178,179} Isothermal unfolding kinetics as a transition from G4 to duplex DNA can be investigated by addition of the complementary strand (complement trapping).^{170,180–182} A method that allows investigating (un)-folding kinetics both isothermally and in the presence of monovalent cations is mechanical unwinding. This has been demonstrated with force microscopy on oligonucleotides that are attached to magnetic tweezers.¹⁷⁶

However, results from different folding experiments have a limited comparability (see Table 1 for overview). The reason for this is the lack of a clear definition for an unfolded state. Obviously, the nature of an unfolded state can be drastically different: (i) Thermal denaturation or high-pressure unfolding represent the population of thermodynamically excited states.^{183,184} (ii) Chemically unfolded states in the absence of monovalent cations or due to pH changes can alter the conformational energy landscape itself.^{159,160} (iii) Unfolding with external mechanical tension restricts the flexibility of oligonucleotides and constrains predefined orientations.^{175,176,185}

While mechanical unfolding ensures complete unfolding in a linearly stretched strand, chemically altered conformational energy landscapes can cause the population of alternative folds like hairpins or pre-folded states.^{158–160,186,187} In all cases, the experimental conditions determine different starting points for folding pathways that result in different folded conformations. Furthermore, oligonucleotides have significantly greater flexibility at their termini than DNA sequences in a genomic context. A study on G-triplex formation has investigated the influence of proximal DNA and found increased complexity for the dynamics of tethered oligonucleotides.¹⁸⁸ The DNA flanking regions also have an impact on the kinetics of *anti*-parallel vs. parallel topologies.¹⁸⁸ This might be especially important for the evaluation of smFRET-derived kinetics, where often the dyes are attached at double stranded overhangs.¹⁸⁹ The influence of different metal cations on the folding kinetics has also been studied in detail.^{190,191}

Table 1: Comparison of different methods to induce folding of G4s. All methods modulate the folding energy landscape and thermodynamic aspects during folding in a different way.

	<i>constant [K+]</i>	<i>isothermal</i>	<i>flexibility/tumbling</i>
<i>mixing/chemical</i>	-	+	+
<i>hysteresis/thermal</i>	+	-	+
<i>force/mechanical</i>	+	+	-

2.2 Folding Dynamics of DNA G-Quadruplexes

2.2.3 Folding Kinetics of DNA G-Quadruplexes

The timescale of folding kinetics for DNA G-quadruplexes ranges from milliseconds to minutes.^{171,177} The folding behaviour is highly sequence dependent and is sensitive to ion-concentration and buffer conditions.^{191,192} For many G4 forming sequences, multi-pathway folding¹⁷³ has been reported and parallel folding into different conformations has been observed.¹⁷⁴ In general, folding of G4s is multiphasic and is supposed to involve intermediates like *anti*-parallel hairpins^{165,170,171}, triplexes^{193–195} and (n-1)-tetrad conformations.^{167,196–198}

Folding of the human telomeric sequence (TTAGGG)_n and its sequence variants have been studied most extensively, both experimentally^{144,159,167,169,171,173–175,188,199–201} and theoretically^{202–204}. The telomeric G4 adopts at least two different hybrid conformations under physiological conditions.^{205–210} K⁺-induced folding revealed a kinetic partitioning mechanism, where one hybrid conformation is kinetically favoured. The conformational equilibrium is reached only after days and partially unfolded states have been observed as long-lived intermediates.¹⁴⁴ This highlights the importance of kinetic studies, since non-equilibrium G4 conformation can have lifetimes that exceed the biological relevant timescales (~20 ms/nt for DNA replication and ~200 ms/nt for transcription).^{171,211} This means that from a chemists point of view, there are kinetic and thermodynamic products of G4 folding.²⁰¹

The kinetics of G-quadruplex folding are very sensitive to ionic conditions ([K⁺] in the following). In principle, studies at physiological conditions should be most reliable to evaluate kinetic aspects, which is however difficult for experimental reasons (2.2.2). Dimerization during folding is a critical aspect that can bias the folding kinetics at higher [K⁺].²¹² Kinetics of tetramolecular G-quadruplex assembly have been studied.^{213–215} In these intermolecular studies a stepwise mechanism was proposed with a fast monomer-dimer equilibrium followed by an intermediate triplex formation that seems to be stabilized especially at higher K⁺ concentration.²¹⁴ Folding pathways inspired by such stepwise strand recruitments are often transferred to explain G4 folding kinetics, but this has to be treated with caution, since the kinetics and thermodynamics of intermolecular assemblies are inherently different from intramolecular folding. At very low K⁺ concentrations (<1 mM or substoichiometric) a drastic effect on the kinetics is observed, since the recruitment of K⁺-ions gets rate-limiting and determines the overall folding pathways.¹⁶⁷ These effects on folding acceleration saturate however at approx. 2-3 mM concentrations of K⁺,¹⁷⁷ whereas the thermal stability of G4s still increases at much higher [K⁺].¹⁹⁶

The geometry of the loops is a critical parameter for the thermal stability of G4s and determines the folding topologies that are preferentially adopted (2.1.2).^{216,217} The loops also have a considerable impact on the folding kinetics^{100,216} and can shift the folding process towards different pathways via pre-folded arrangements.^{159,160,218} An effect on accelerated folding has been observed especially for hairpin-forming loops.^{127,158,192,219}

2.2 Folding Dynamics of DNA G-Quadruplexes

2.2.4 Comparison to RNA G-Quadruplexes

RNA G4s generally show reduced structural polymorphism.^{220,221} They tend to adopt *all*-parallel conformations, where all sugars are in *anti*-configuration.^{222,223} Chemical modifications like 8-bromoguanosine can be used to flip the glycosidic conformation and force RNA G4s into *anti*-parallel conformations.^{224,225} In line with that, folding kinetics of RNA G-quadruplexes are markedly different and are considered to be much faster compared to their corresponding DNA sequences.^{171,172} The human telomeric-repeat containing RNA (short: TERRA)^{226–231} sequence is one of a few examples that has been investigated in aspect of RNA folding and compared with its homologous DNA sequence.^{171,172} The drastically faster RNA-G4 folding has been explained with a reduced conformational space that results from pre-oriented sugar configurations. The more complex (and stable) *syn-anti* patterns in DNA G4s lead to an increased number of unproductive microstates and potentially misfolded conformations.

2.2.5 Mutual Exclusive Formation of DNA i-motifs and G-Quadruplexes

The folding of i-motifs and G-quadruplexes is a comparably complex process. The rapid pH-induced folding of i-motifs has been investigated with time-resolved NMR and revealed a kinetic partitioning mechanism.^{76–78,80} Within this study⁷⁶ it was shown that two coexisting i-motif conformations fold in parallel with a kinetically favoured conformation ($\sim 0.9 \text{ min}^{-1}$ vs. 2.0 min^{-1}). The subsequent refolding process to reach conformational equilibrium is slower by about two orders of magnitude.

Under *ex vivo* conditions, it was found that in many cases the formation of i-motifs and G-quadruplexes is mutually exclusive.^{232–234} A steric hindrance and differences in their different requirements for stability at physiological conditions impedes a simultaneous formation.²³⁵ However, the i-motif structure itself is highly dynamic. It can switch between hairpin and i-motif conformations in the C-rich strand, which affects the formation of G-quadruplexes in the complementary strand and the recognition by protein binding partners (such as hnRNP-LL).^{81,236–239} The influence of chemical and mechanical factors on i-motif and G4 formation has been studied in detail for the G4 forming sequence in the *cMYC* promoter (NHE-III₁, 2.4.2),²⁴⁰ and the general requirements for a simultaneous formation and the possibility for a co-existence have been reviewed recently.²³⁵ Unsurprisingly, the population of either of the structures is dependent on ion-concentrations and pH, but interestingly also on factors like superhelicity and molecular crowding.^{234,239,241} Most important to note is that the equilibrium between G4, i-motif and duplex DNA is highly competitive and dynamic. Under *in vivo* conditions, the effects of e.g. negative superhelical-stress (2.4.3), protein binding (2.3.2) and oxidative stress (2.3.3) greatly affect the interplay of non-canonical structures in the complementary G-/C-rich strands.

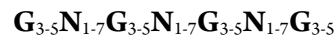
2.3 Function and Targetability of DNA G-Quadruplexes

*“If G-quadruplexes form so readily in vitro,
Nature will have found a way of using them in vivo”*

Sir Aaron Klug, more than three decades ago²⁴²

2.3.1 Occurrence of G-Quadruplex Forming Sequences in the Human Genome

G-rich sequences that are able to form G-quadruplexes are spread throughout the entire human genome,^{243–247} particularly in certain regulatory regions or functional domains, especially gene promoters of proto-oncogenes,^{243,248,249} immunoglobulin switch regions,^{250,251} telomeric^{210,228,252} and subtelomeric tandem repeats (minisatellites)^{91,253} and 5'-untranslated regions (5'-UTR)^{98,254}. Furthermore, their formation is linked to CpG islands²⁵⁵ and chromatin modifications in the human genome.^{256–259} The consensus definition for G4 motifs:



yields ~376.000 potential G4 motifs in the human genome. However, every bioinformatic algorithm has a trade-off for prediction to balance between false-positives and false-negatives. The above consensus definition does not account for non-canonical features of G4 polymorphism (2.1.3). A refined, sophisticated prediction tool called G4hunter^{260,261} found ~700.000 G4 motifs, which is in line with 716.310 distinct G4 structures that have been identified with a high-throughput sequencing method that is able to detect G4s in human genomic DNA.²⁶² Out of this, 451.646 G4s have not been predicted by the consensus definition,^{249,263} which emphasizes both the overall significance of G4s in the human genome as well as the outstanding importance on non-canonical G4 structures.

Following this enlarged G4 definition, human subtelomeric tandem repeats (minisatellites) and their variations in *Saccharomyces cerevisiae* have come into focus for G4 function.^{91,96,253,264–267} The single stranded overhang in human telomeres consists of tandem repeats of the sequence:



The formation, dynamics, structure and function of G4s in human telomeres, their impact on genome stability and their interaction with telomerase have been extensively studied^{252,268–273} and strategies to target telomeric-G4-structures^{205–210} have been discussed controversially in this context.^{274–289} The interested reader will find an astonishing plethora of papers and discussions in the literature, of which even a short summary would go beyond the scope of this discussion. For an updated overview, these most recent reviews on this topic are recommended.^{290–293}

2.3 Function and Targetability of DNA G-Quadruplexes

Besides bioinformatic prediction, G4 structures in the human genome have been mapped and detected with various approaches²⁴⁶ using high-resolution sequencing (G4-seq)²⁶², chromatin immunoprecipitation coupled to G4-specific single-chain antibody detection and high-throughput sequencing (G4-ChIP-seq²⁹⁴, qG4-ChIP-seq²⁹⁵).

The first *in vivo* detection (nuclear staining) of telomeric G4 structures has been reported in ciliates (*Stylonychia lemnae*) with single chain variable fragments of an *in vitro* selected antibody.^{246,296,297} Detection in fixed human cells has been reported with G4-specific antibodies BG4⁶⁶, 1H6²⁹⁸ and D1²⁹⁹ using immunofluorescence microscopy.²⁴⁶

In-cell NMR has been used to investigate G-quadruplexes in living cells (in particular *Xenopus laevis* oocytes).³⁰⁰⁻³⁰⁶ Degradation of G4 forming oligonucleotides by endogenous nucleases is retarded, which elongates the obtainable time frame for NMR experiments. In-cell conditions allow the evaluation of the structural integrity (or structural changes) of G4 conformations under physiological relevant conditions with molecular crowding. The problems arising from broad signal linewidths and poor spectral resolution that are typically encountered for in-cell NMR spectra, can be evaded by using e.g. ¹⁹F NMR³⁰⁷⁻³⁰⁹ (or ³¹P)^{304,310} on chemically modified oligonucleotides.

More recently, small-molecule optical probes³¹¹ have been used to visualize G4 structures in live cells (IMT, ⁶⁷BMVC,³¹² template-assembled synthetic G-quartets: N-TASQ³¹³) even at the single molecule level.⁶⁸ A fluorescent probe (DAOTA-M2) was used in combination with **F**luorescence **L**ifetime **I**maging **M**icroscopy (FLIM) to visualize G4 structural dynamics in live cells.³¹⁴

The advancements and efforts in the past decade²⁴⁶ towards mapping, detection, imaging and visualization of G4s *in cellulo* and in live cells now gives an astonishingly convincing view about the importance of G4 formation in the human genome. Today, there is no doubt anymore about the existence of G4s in human chromosomal DNA. Approaches like qG4-ChIP-seq (using G4-specific antibodies) and FLIM (using G4-specific fluorescent probes) now enable the investigation of G4 occurrence in relation to e.g. epigenetics, pathogenesis and carcinogenesis.^{259,315}

Besides the potential role of DNA G-quadruplexes in the telomeric repeats, their function is discussed in basically all contexts that are related to genomic DNA.^{247,251,259,315} This includes in particular replication³¹⁶⁻³¹⁹, epigenetic processes such as CpG methylation^{255,320-323} or alterations to the chromatin structure^{257,258,294}, genomic instability^{96,253} and transcription^{324,325}. Within the following chapters, the focus will be on DNA G4s as transcription regulatory elements. The group of S. Balasubramanian has recently published a comprehensive review²⁵⁹ (mini review in *Trends in Chemistry*)³¹⁵ that covers the functions of both DNA and RNA G-quadruplexes.^{259,315}

2.3 Function and Targetability of DNA G-Quadruplexes

2.3.2 Transcriptional Regulation of G-Quadruplexes and Binding-Interactions

In a common, yet oversimplified picture of the regulatory function of G4s in gene promoters, transcription suppression (transcriptional *off-switch*) is proposed due to sterical hindrance and stalling of the RNA polymerase II (RNAP-II). However, even if further protein interactions are left unconsidered, the basic principles of how G4 can affect transcription efficiency are certainly more complex. In the context of transcription regulation, different modes of protein interactions have been discussed and similar effects are expected for small-molecules interactions. The concept of transcription regulation (*on/off-switch*) via small-molecule ligand targeting to G-quadruplexes did drive the field in the recent decades (see chapters 2.3.4 and 2.4.4). Unfortunately, regulation of genomic G-quadruplexes is not as straightforward as chemists like to imagine it. One aspect of a simple “*actio-reactio*” picture is, however, very well true: DNA G-quadruplexes *can* be represented as the starting point of the entire downstream cascades in cellular regulation – with all consequences to the complex cellular circuits. This chapter will therefore focus on the basic principles of G4 recognition. Two general types of interactions with G-quadruplex motifs will be differentiated: **stabilizing/folding inducing** and **destabilizing/unfolding facilitating**. Figure 13 shows an overview of the different modes of G4-regulated, altered transcription. The role of the complementary strand that can potentially form an i-motif is not discussed herein. To avoid a suggestive depiction, the strand is completely ignored in the Figure. Chapters 2.1.1 and 2.2.5 give implications for this issue. Here, it is important to note that the distribution of G4 forming sequences downstream of the transcription start site (TSS) is highly asymmetric in the sense and anti-sense strand (coding and non-coding strand).³²⁶ The location on either of the strands plays a crucial role for RNAP-II blocking, and can not only suppress but also activate transcription. The same holds true for the processing of G4s via helicases.

G4-binding proteins

Protein binding to G4 forming sequences has ambiguous effects on transcription levels: Some proteins have been reported to stabilize the G4 fold, while others inhibit G4 formation by binding to the single stranded DNA. G4-binding proteins can either inhibit the interaction with other proteins or recruit further proteins to form complexes. A prime example of a G-quadruplex binding and stabilizing protein is the nucleolar phosphoprotein **nucleolin**.^{327–333} Nucleolin was first found to bind G-G paired DNA³³⁴ and then reported to bind the *cMYC* promoter G4³³⁵ and later also other promoter G4s (LTR promoter^{336,337}). Binding is mediated via the C-terminus that contains RGG-repeats.³³⁸ There is growing evidence that nucleolin can differentiate between different conformations and preferentially binds long-looped G4 conformations.^{339,340} This has strong implications for the recognition of different folded isomers (e.g. *spare-tire* isomers) as has been outlined for the *cMYC* promoter G4 (see chapters 2.4.2 and 2.4.3).^{341,342} Recently, the interactions with RNA G4s have been used to target nucleolin.³⁴³⁻³⁴⁵

2.3 Function and Targetability of DNA G-Quadruplexes

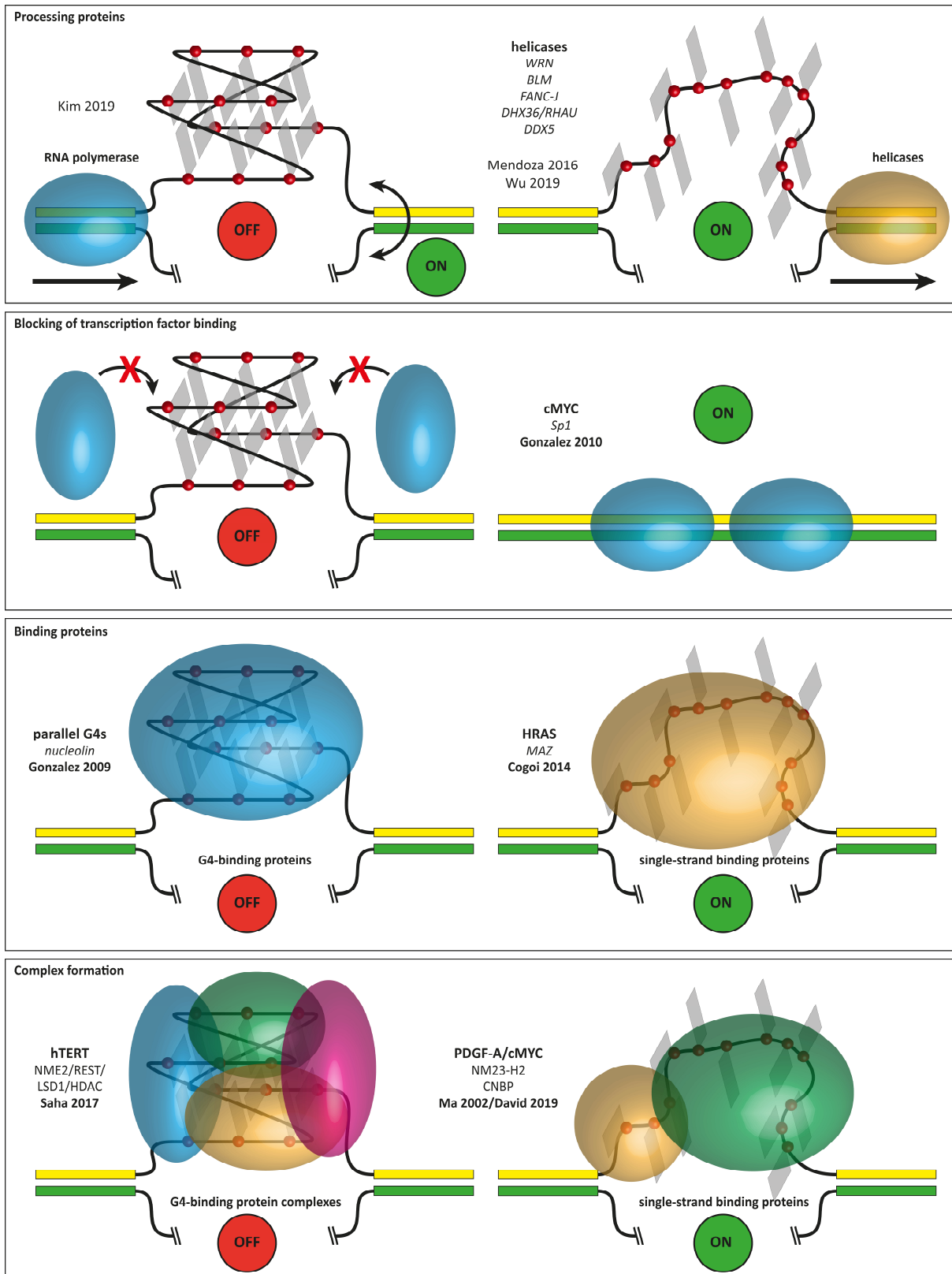


Figure 13: Different mechanisms of transcription regulation with G4-interacting proteins. **Processing proteins:** Classical model that shows stalling of RNA polymerase (RNAPII) during transcription due to sterical reasons (inverse effects for G4s in the coding/non-coding strand).³²⁴ The G4-fold cannot be resolved by the RNAPII itself, but helicases³⁴⁶ like WRN, BLM, FANC-J, RHAU or DDX5³⁴⁷ are effective G4 unwinders. **Binding proteins:** Proteins that bind and stabilize G4s can downregulate the transcription, while proteins that bind the unfolded single strand (and/or facilitate the unfolding) can have an activating effect. **Complex formation:** Some G4 binding proteins recruit additional proteins to form complexes, which can act by either silencing or activating.

2.3 Function and Targetability of DNA G-Quadruplexes

A manifold of G4-binding proteins have been reported (e.g. Zuo1³⁴⁸, IFI16³⁴⁹, Lia3³⁵⁰, LARK³⁵¹, SLIRP³⁵², CNBP^{353,354}) and reviewed recently.^{355,356} The analysis of shared motifs³⁵⁷ and the general interaction with RGG-domains^{358–360} have been analyzed and a powerful tool to explore G4-interacting proteins called G4IPDB (a web-based database with more than 200 entries) is now available.³⁶¹

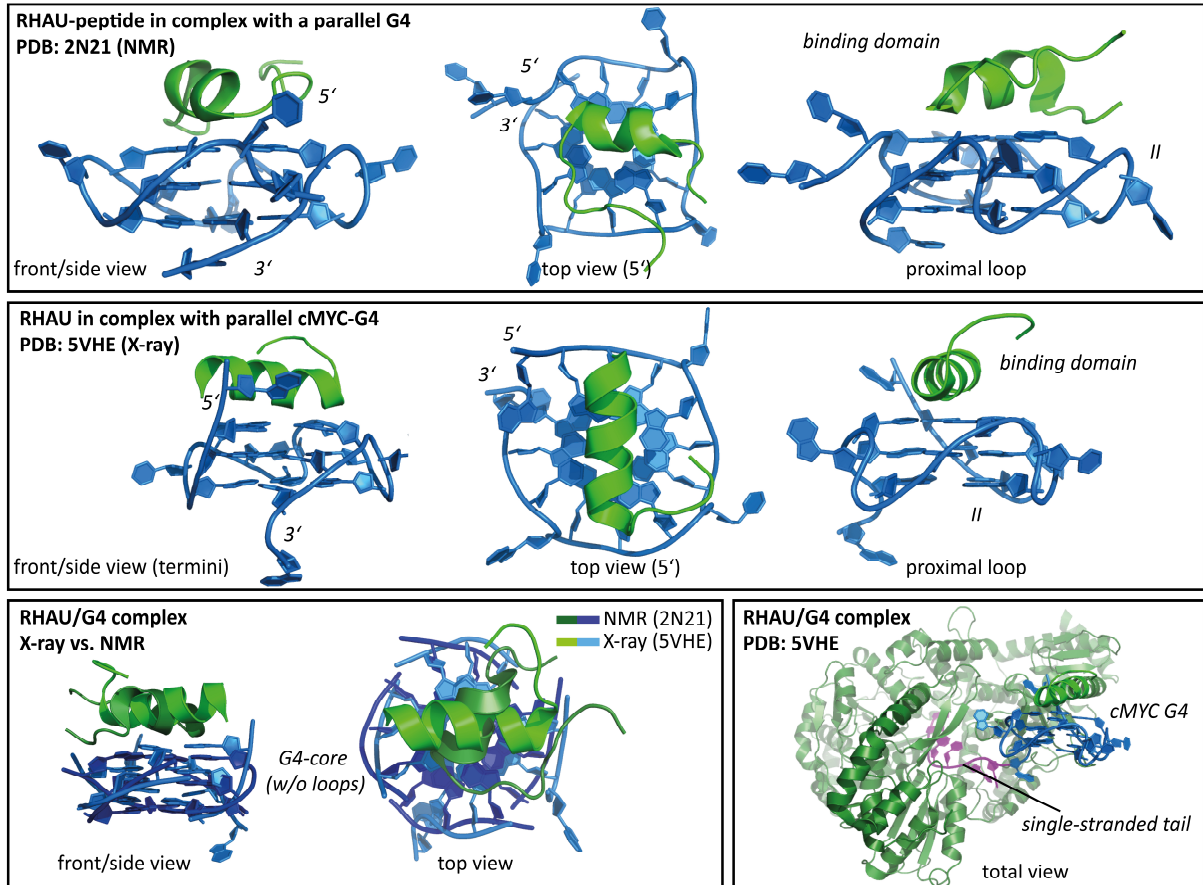


Figure 14: Details of the PDB-structures of the RHAU-peptide (binding motif) from NMR spectroscopy (PDB: 2N21)³⁶² and the whole RHAU helicase from X-ray crystallography (PDB: 5VHE, detail of binding domain, entire protein complex at the bottom-right)³⁶³. The comparison of the structures shows different orientation of the binding domain at the top-tetrad (nearly perpendicular arrangement).

G4-unwinding proteins and helicases

Some interacting proteins have been shown to facilitate G4 unfolding or directly bind to the unfolded single-strand and thereby disturb G4 formation (MAZ^{364,365}, NM32-H2^{366–368}). Recent studies suggest that CNBP-binding acts in a comparable manner.³⁵⁴ Information on the detailed binding modes of these proteins remains ambiguous^{369,370} and NM32-H2 has also been discussed in a different context as mediator for epigenetic suppression of the *hTERT* promoter (recruiting a REST-LSD1 complex).³⁷¹ Different helicases have been reported (DEAH-helicases: DHX36/RHAU^{362,363}, SF2 helicase: FANC-J^{298,372–374}, SF1b helicase (*S. cerevisiae*): Pif1^{253,264–266,375,376}, RecQ-helicases^{147,377–379}: BLM^{380–384}, WRN^{385,386}, RecQ5³⁸⁷, RecQL4^{388–390}) that are able to process and unwind G-quadruplexes.^{346,391} G4 unwinding ability is vastly different for the members of the RecQ helicase family, with BLM and WRN showing the highest efficiency. For some helicases a topology-specific unwinding activity has been

2.3 Function and Targetability of DNA G-Quadruplexes

shown.^{380,392} Interestingly, some are strictly unable to unwind G4s (RecQ1)³⁹³, which has implications for the recognition of distinct templates. Helicase-deficiencies in general are linked to severe hereditary diseases such as Bloom-syndrome (age-unrelated cancer predisposition, BLM-deficient)³⁹⁴, Werner-syndrome (pre-mature aging phenomena, WRN-deficient)³⁹⁵ or Rothmund-Thomson syndrome (accelerated aging, RecQL4-deficient)³⁹⁶. The main cause of this is a chromosome instability that is also associated to the inability to resolve G-quadruplexes in the genome. How the helicases recognize specifically the G4-fold still remains largely elusive. Figure 14 shows the binding of RHAU helicase to different parallel G4 conformations. Interestingly, the peptide binding domain of RHAU is arranged differently in a NMR-structure (PDB: 2N21, peptide only)³⁶² compared to the X-ray-structure (PDB: 5VHE, whole protein)³⁶³. More recently, the DEAD-box helicase DDX5 has been linked directly to transcriptional activation of *cMYC*.³⁴⁷ DDX5 is overexpressed in cancer cells and might have an important function as *G4-resolvase* that contributes to the high expression levels of *MYC*. Sterical hindrance of unprocessed G4s is further important for the binding of transcription factors that bind to B-form, double-stranded DNA (e.g. Sp1^{397,398}). The formation of G4s thus blocks the binding sites of transcription factors and thereby act as transcriptional silencer.³⁶⁹

Small-molecule binding

Around 2000, the first studies were published that report on specific promoter G4-ligand binding (cationic porphyrins).³⁹⁹⁻⁴⁰¹ Since then, a wide range of ligand scaffolds has been proposed, synthesized and studied for specific G-quadruplex targeting.^{402,403} A discussion of small-molecule structural classes and strategies to bind specific G-quadruplexes is far beyond the scope of this thesis and the interested reader will find countless comprehensive reviews in the literature. (see some non-comprehensive, exemplary references: selected studies^{280,404-410}, more recent studies⁴¹⁰⁻⁴¹³ and for further analysis of G4-ligands the web-based database G4LDB can be used⁴¹⁵) Figure 15 shows different structures of ligand-G4 complexes taken from the PDB that all feature the most common binding mode for G4-ligands, which is stacking to the tetrads via large π -electron surfaces. The NMR-structure of an Au(III)-ligand in complex with a telomeric hybrid-2 G4 (PDB: 5MVB)⁴¹⁶ shows an interesting rearrangement of the capping structures that are required to accommodate the ligand. An important point to mention here is that all of these studies aim to stabilize the G4-fold and modulate its function by shifting the thermodynamic equilibrium towards a folded state. The group of D. Monchaud recently conceptualized an alternative, orthogonal approach that aims to destabilize certain G4 conformations.⁴¹⁷ The different strategies that have been proposed to target G4s with small molecules show the diversity of possible effects that result from interfering with G4-regulation.

2.3 Function and Targetability of DNA G-Quadruplexes

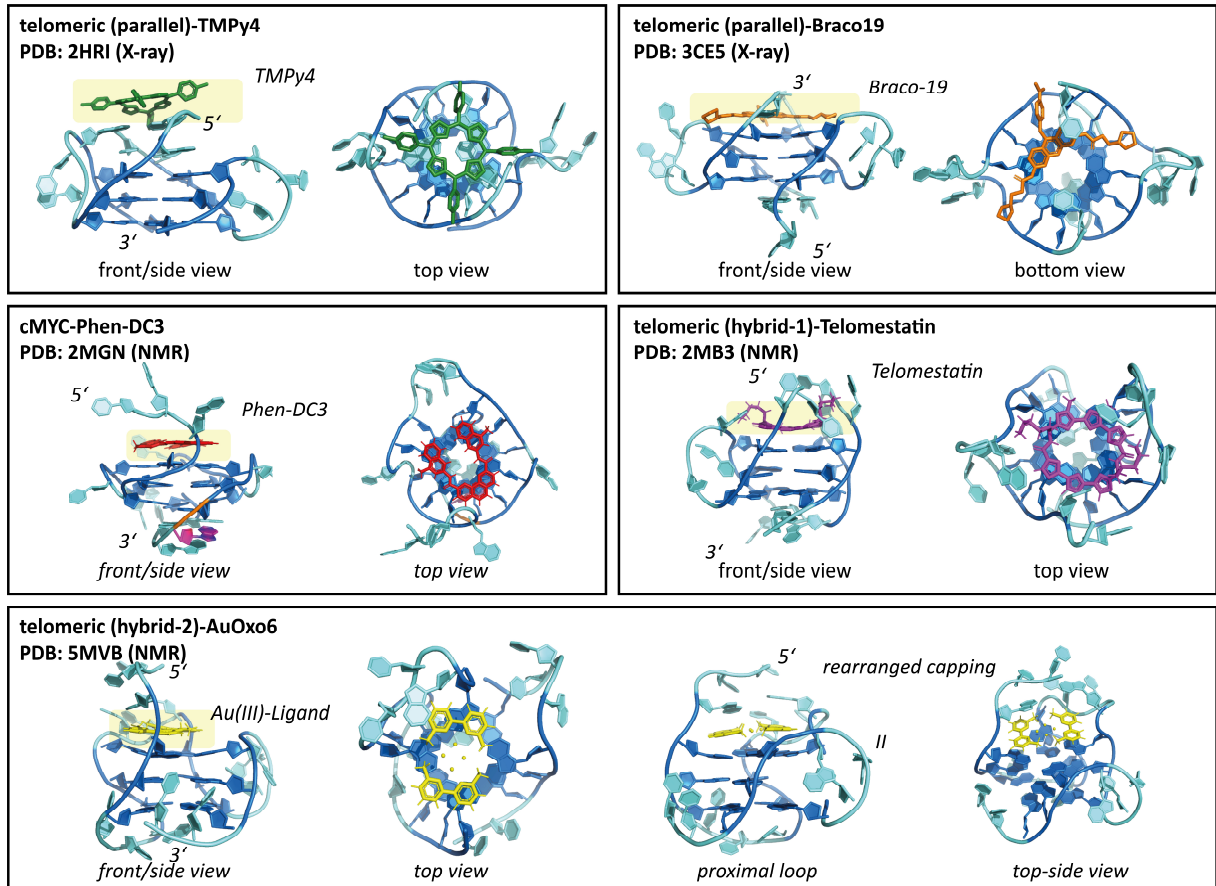


Figure 15: Details of the PDB-structures of G-quadruplex-ligand complexes that feature π -stacking to the top ($5'$) or bottom ($3'$) tetrads as a common binding mode. The complexes show a dimeric, parallel telomeric G4 with TMPy4^{401,418,419} (PDB: 2HRI)⁴²⁰ and Braco19^{286,421} (PDB: 3CE5)⁴²²; a cMYC G4 with Phen-DC3^{423,424} (PDB: 2MGN)⁴²⁵; a telomeric hybrid-1 G4 with Telomestatin⁴⁰⁹ (PDB: 2MB3)⁴²⁶ and a telomeric hybrid-2 G4 with an Au(III)-ligand (PDB: 5MVB, AuOxo6)^{416,427–429}.

2.3.3 Oxidative Damage under Cellular Stress Conditions

Guanine is the most electron rich from all four DNA nucleobases and therefore the most prone to oxidation. In G-rich runs this effect is even increased, since due to π -stacking of Gs the ionization potential of the $5'$ -G is significantly lowered with increasing length of the G-tract. Duplex DNA has the ability to funnel electron holes (oxidation sites) through the π -stacked nucleobases. These electron holes can migrate through the DNA duplex until they reach the site that is most prone to oxidation, which will likely be a $5'$ -G of G-runs. This has been hypothesized to serve as *cathodic protection* of genes⁴³⁰ or epigenetic mark to modulate gene expression.⁴³¹ The consequences of oxidative damage to G-residues and G-quadruplex formation in the genome have been conceptualized and reported in a series of publications from C. Burrows in the last years^{113,254,431–439} and have been recently reviewed.⁴⁴⁰ 8-oxo-guanine (OG) is the major oxidation product that results from a two-electron oxidation (Figure 16) with reactive oxygen species (ROS). OG can be further oxidized (four-electron hyperoxidation) by ROS to 5-guanidinohydantoin (Gh) or spiroiminodihydantoin (Sp). Oxidized G-residues with lesions can be repaired by OG-glycosylase (OGG1) or NEIL-glycosylases⁴⁴¹ that

2.3 Function and Targetability of DNA G-Quadruplexes

initiate the base excision repair (BER). Those oxidized species alter the Hoogsteen interface of G-residues and disturb the tetrad formation in G-quadruplexes.⁴⁴²⁻⁴⁴⁴ The N-H7 imino proton enables the formation of A-G Hoogsteen base pairing and the O8 causes a slight sterical clash, if base paired with C. These effects cause a mutagenic potential (during replication) towards a G-to-T transversion, when the OGs are not repaired.

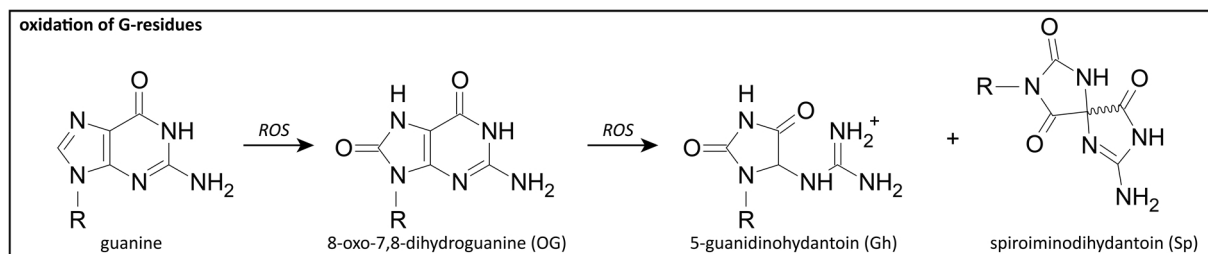


Figure 16: Major oxidation products for the reaction of guanine with reactive oxygen species (ROS). 8-oxo-7,8-dihydroguanine (OG) as oxidation product of a two-electron oxidation; 5-guanidinohydantoin (Gh) and spiroiminodihydantoin (Sp) are the hyperoxidation products of a four-electron oxidation.

Many G-quadruplex forming regions in oncogene promoters possess more than four G-rich tracts that are close by in sequence.¹¹³ The consequences of this for a possible non-canonical structural polymorphism have been discussed in chapter 2.1.2. Burrows and co-workers have outlined a possible evolved function for these *spare-tire* G-tracts in response to oxidative damage in one of the G-rich tracks.^{113,438} They propose a mechanism that describes a *maintaining of function* for G4s. The G-tract that carries a lesion is excluded from G4 formation and one of the *spare-tires* participates instead in G4 folding as surrogate. The so exposed oxidized residues are accessible to the BER machinery. This mechanism has also been utilized for an approach with an intermolecular *spare-tire* surrogate (using a PNA) to recover the G4 functionality.⁴⁴⁵ Further, OG sites in folded G4s (shown for the polymorphic KRAS G4⁴⁴⁶⁻⁴⁴⁸) have been shown to be recognition sites for the MYC associated zinc finger protein (MAZ).^{364,365} MAZ binds to oxidized G-residues in the loops of folded G4s and facilitates unfolding and subsequent activation for BER. This finding has implications for the recognition of oxidized Gs in the loops, that can emerge from G-register isomers or G-quadruplexes with G-vacancy sites (2.1.3).

2.3.4 Strategies in Anti-Cancer Treatment

The stabilizing effect of small-molecule binding has now developed to a strategy that aims to target specifically G-quadruplexes in the human genome for anti-cancer treatment.^{276,282,449-451} The first reports on G-quadruplex targeting mainly aimed for telomeres and telomerase inhibition.²⁷⁴ This strategy is discussed controversially and the biological implications for it have shifted, since the functionality and structural assembly of human telomeres still remains largely elusive. One of the main ideas was to block telomerase interactions by facilitating G-quadruplex formation in the telomeres. This prevents the limitless replicative potential of cancer cells and re-enables senescence and

2.3 Function and Targetability of DNA G-Quadruplexes

apoptosis. However, the alteration of telomere capping structures and shelterin complexes overrule this effect and lead to genomic instability and the activation of DNA repair machineries.^{284–288,449}

One of the very first studies on G-quadruplex formation in the *cMYC* promoter showed that expression levels can be significantly reduced after binding of a small molecule.⁴⁵² This strategy has now evolved and G-quadruplexes in gene promoters are recognized as *cis*-acting elements that can be directly targeted to regulate transcription levels.^{54,90}

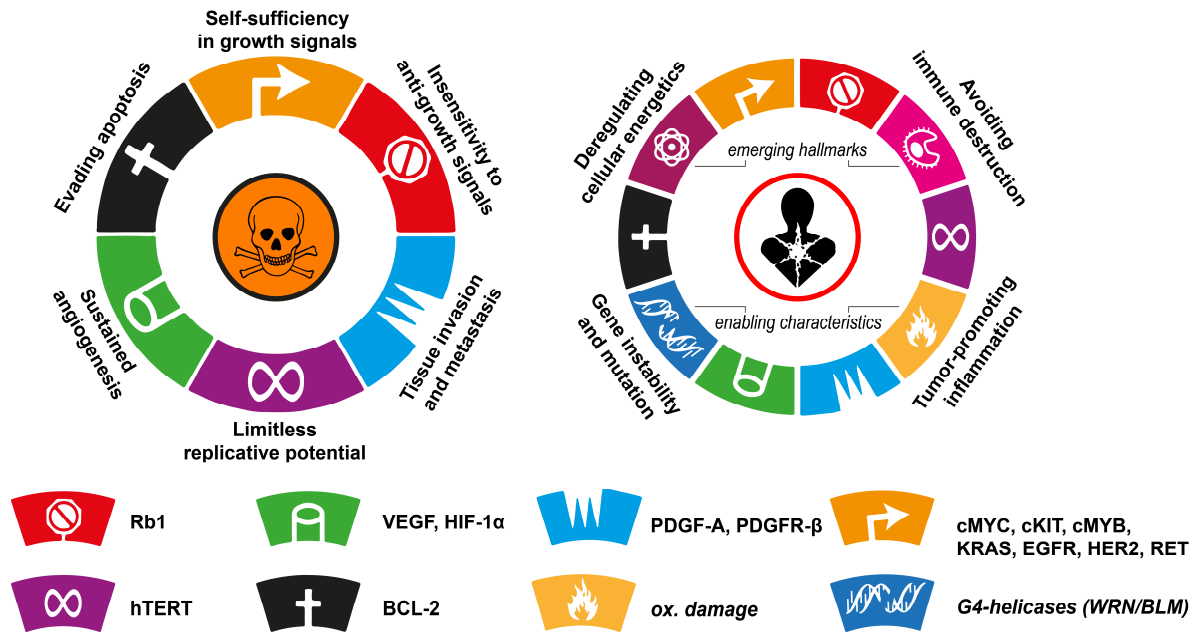


Figure 17: The six “classical” hallmarks of cancer according to Hanahan and Weinberg 2000 (**left**).⁴⁵³ Refined and enlarged definition (“next generation”, 2011), including *enabling characteristics* and *emerging hallmarks* (**right**).^{454,455} (**bottom**) Overview of genes with reported regulation through promoter G-quadruplexes or mechanisms that affect G4 regulation (italic).^{239,456–459} Schematic depiction adapted from Hanahan and Weinberg.^{453,454}

An analysis of the gene ontology of promoters with G4 forming sequences reveals that the enrichment of promoter G4s is unequally distributed among different gene functions.^{244,249} Gene ontology categories with significantly higher frequencies of G4 forming sequences are linked to e.g. transcription factor activity, RNAPII transcription factor activity, neurogenesis, kinase activity and cell differentiation, while categories like immune response, nucleic acid binding, protein biosynthesis, ribosome-related or antigen binding have significantly lower frequencies of G4 forming sequences. This aspect of genomic G4-distribution highlights the general relevance for G4 regulation in cancer cells and their importance as molecular targets.⁴⁵⁹ Structural validation of G-quadruplex forming sequences has been reported for the promoters of many important genes that can be assigned to all different hallmarks of cancer (Figure 17).^{239,244,450,453} Among them are (non-comprehensive): *cMYC*^{99,342,460,461}, *BCL-2*^{126,217,462–464}, *hTERT*^{102,103,115,116}, *VEGF*^{85,408}, *cMYB*⁹⁰, *cKIT*^{130,465}, *KRAS*^{365,446,447,466}, *EGFR*²¹⁸, *PDGF-A*^{368,467}, *PDGF-β*^{109,131,468}, *HIF-1α*⁴⁶⁹, *HER2*⁸⁹, *RB1*^{470,471} and *RET*³²⁵.

2.4 The cMYC Proto-Oncogene Promoter

*“I am, of course, most ignorant about all things biological,
but I imagine most (X-ray) people start that way.”*

Rosalind Franklin, response to John Randall 1951⁴⁷²

2.4.1 The cellular Myelocytomatosis Protooncogene

Information within this chapter on molecular structure, biological function, oncogenic potential and targeting of *MYC* have been extensively reviewed in the last years.^{473–479} *MYC* is a family of protooncogenes that was found after the discovery of oncogenic retroviruses that cause myelocytomatosis (*v-myc* oncogene). The *MYC*-family has three members, namely *cMYC* (or just *MYC*), *N-MYC* and *L-MYC* that encode for transcription factors. The C-terminal domains of *MYC*-members have basic-helix-loop-helix (bHLH) and leucine zipper motifs. This C-terminal domain remains largely unstructured until dimerization with other transcription factors such as the myc-associated factor X (*MAX*). The bHLH enables the binding to DNA, while the leucine zipper allows dimerization. Figure 18 shows the X-ray structure of a *MYC-MAX* heterodimer bound to DNA (PDB: 1NKP).⁴⁸⁰ The unstructured N-terminal domain is responsible for transcription regulation (*transactivation*). It contains highly conserved *myc-boxes* and forms complexes with other factors that induce structured folding upon binding.

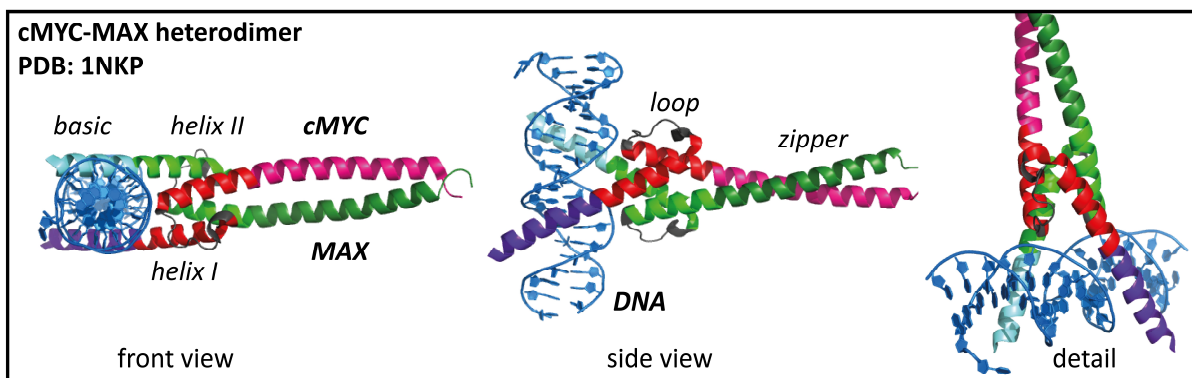


Figure 18: Crystal structure of a *cMYC-MAX* heterodimer bound to DNA (PDB: 1NKP). *cMYC* is shown in red/violet colours and *MAX* in green/cyan colours. The structure of the C-terminal domain comprises a basic-helix-loop-helix (bHLH) and a leucine zipper motif.

MYC, as a *super* or *master* transcription factor plays many important roles and regulates more than 15% of the human genome. A major function of *MYC* is cell proliferation, growth and differentiation as well as apoptosis. In this function, it has a high transformation potential and is a key driver in tumorigenesis. However, strategies for small-molecule drug development to target *MYC* are extremely challenging, mainly because it lacks a specific active binding site. Hence, *MYC* is considered *undruggable*.⁴⁸¹

2.4 The cMYC Proto-Oncogene Promoter

Usually, the expression of *MYC* is highly regulated through the tight control of its promoter region. However, *MYC* is found deregulated in a wide range of nearly 70% of human cancers with elevated levels of *MYC*-expression. Burkitt's lymphoma, a highly aggressive B-cell non-Hodgkin lymphoma, is a prime example for the devastating effect of *cMYC* misregulation. Here, chromosomal translocation to an ectopic promoter leads to an activation of the *cMYC* oncogene, thereby making it one of the fastest growing tumour in humans.⁴⁸²

A plethora of molecular mechanisms maintains cellular homeostasis of *MYC*, many of them are still poorly understood.^{473,483,484} What is known is that the promoter is controlled and regulated by a large number of *cis*-regulatory elements, intra- and extracellular signals and signal pathways associated to an array of transcription factors that directly bind to the *cMYC* promoters:

“Everything regulates MYC. From stem cells and proliferation to senescence and cell death, MYC participates in almost every crucial decision of almost every cell. Hundreds of extracellular and intracellular signals, operating through an array of transcription factors, chromatin modifiers/remodelers, and regulatory RNAs (recruited to or synthesized at the c-myc locus), are brought to cis-elements vicinal to the promoter or strewn across a still poorly delineated chromosomal domain and are all somehow integrated to set the physiological levels of c-myc mRNA.”

David Levens, “You Don't Muck with MYC” (2010)⁴⁸⁴

Hence, unravelling the regulation, function and role of *MYC* has remained an “enduring enigma” (Eisenman, 2001)⁴⁸³ since its discovery in 1979. Due to its outstanding significance in tumorigenesis and the impracticality of a direct *MYC* targeting, a focus on the gene promoter region has evolved. Soon after sequencing the *cMYC* locus⁴⁸⁵ a nuclease S1 analysis revealed several altered DNA secondary structures in slow equilibrium with a canonical B-form helix in the promoter region.⁴⁸⁶ DNase 1 mapping of this region then localized several nuclease hypersensitive sites.⁴⁸⁷ At one of these sites, the nuclease hypersensitive element III₁ (NHE-III₁) located -142 to -115 bp upstream of the P₁ promoter a high strand asymmetry was found with a purine (Pu, G-rich) and a pyrimidine (Py, C-rich) strand. In the late '80s it was hypothesized that a regulation via triplex formation might be a plausible mechanism that could explain the vulnerability against nucleases at these sites. Two structures have been proposed that involve either a complex tandem H-DNA with two intramolecular Py-Pu-Py triplexes⁴⁸⁸ or an intermolecular Pu-Py-Pu hybrid-triplex.⁵⁴ Targeting this site with anti-sense RNAs indeed repressed transcription *in vitro*.⁴⁸⁹ The role of interactions with transcription factors was discussed for both models.^{490,491} However, Py-Pu-Py triplexes require low pH^{488,491} and Pu-Py-Pu triplexes require 5-20 mM magnesium ions^{489,492,493}. Both requirements do not fit physiological conditions.

2.4 The cMYC Proto-Oncogene Promoter

In the mid '90s it was found that the structural dynamics in the *cMYC* promoter region are highly ion-dependent and that especially physiological concentrations of K^+ inhibit a potential triplex formation and competitive triplex/quadruplex equilibria have been proposed.⁴⁹² Finally, in 1998 the formation of a tetraplex;⁴⁹³ and in 2000 the formation of an i-tetraplex⁴⁹⁴ in the complementary strand of the *cMYC* NHE-III₁ region has been proposed.

However, even after the discovery of the G-quadruplex formation in the *cMYC* promoter, a possible biological function of triplex (inter- and intramolecular) formation is still in discussion.^{495–497} The promoter can be targeted with triplex-forming oligonucleotides (TFOs) that have been shown to downregulate the transcription and this effect is particularly intensified in cancer cells.^{498–503} TFO-targeting strategies can benefit from an increased cellular uptake and improved intracellular stability of homologous G4 forming oligonucleotides (also shown for VEGF⁵⁰⁴).⁵⁰² Strategies based on siRNAs and anti-sense oligonucleotides have also been developed to target not the genomic DNA, but inhibit the *MYC* mRNA.⁵⁰⁵ Further, the promoter region is a mutagenic hotspot;^{495,497} the mutations and single nucleotide polymorphism (SNP) have been mapped and linked to strong effects on transcription levels.^{506,507} This has severe implications for the delicate equilibrium of all possible non-canonical DNA conformations and these mechanisms have to be considered in the general context of the biological function and targetability of G-quadruplexes in oncogene promoters (chapters 2.3.4 and 2.4.4). In conclusion, this chapter cannot be summarized better than with this quote from a recent review:

“Thus, the c-myc promoter is still something of a black box and only one point is certain: regulation of the c-myc promoter is extremely complex with a lot of redundancy, many feedback loops, and several cross-regulatory circuits involved.”

Inken Wierstra and Jürgen Alves, *The c-myc Promoter: Still MysterY and Challenge* (2008)⁴⁷³

2.4.2 Nuclease Hypersensitive Element III₁

The *cMYC* promoter region (Figure 19) has four promoters (P_0 , P_1 , P_2 and P_3) and seven nuclease hypersensitive elements (NHE). 75-90% of transcription initiates at P_2 , 10-25% at P_1 and less than 10% at P_0 and P_3 . The NHE-III₁ spans a 27 bp (Pu27) sequence -142 to -115 bp upstream of the P_1 promoter. The NHE-III₁ has been shown to control up to 90% of total transcriptional activation of *cMYC*.^{369,452,473} Transcriptional regulation through the NHE-III₁ is believed to be primarily mediated through G-quadruplex formation^{99,452,493,508} in the G-rich strand (Pu27), but the role of i-motif formation in the complementary C-rich strand (Py27) has also been discussed.^{83,241,509–511} This chapter will focus on models on the G4-mediated transcription control (see also chapter 2.3.2 for general considerations).

2.4 The cMYC Proto-Oncogene Promoter

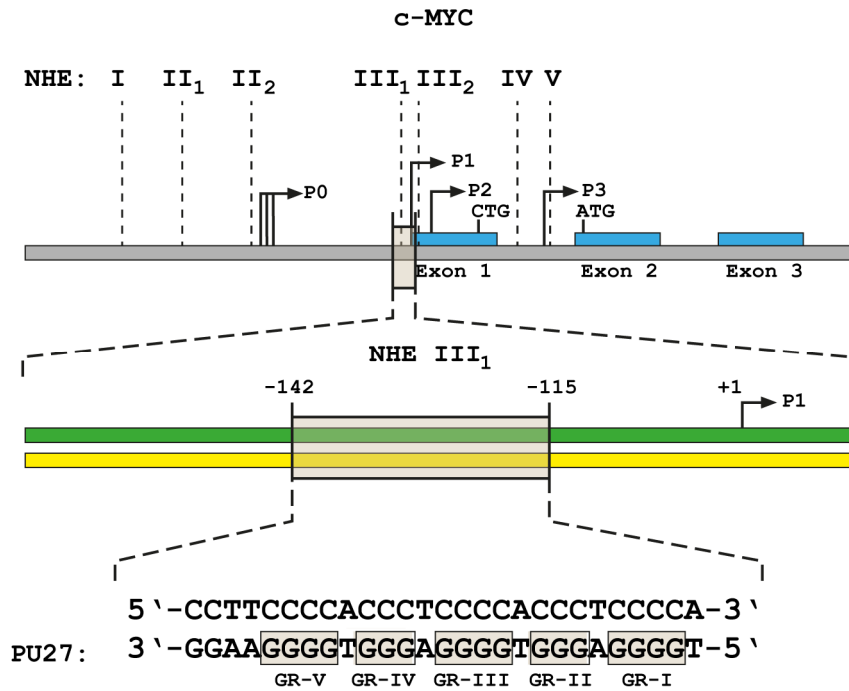


Figure 19: Depiction of the *cMYC* promoter region including nuclease hypersensitive elements (NHE) I-V, promoters 0-3 and transcriptional start sites. G-runs I-V in the NHE-III₁ are numbered from 5'-to-3' direction Figure has been adapted according to the publications from the Hurley group.^{99,452}

The group of L. Hurley has outlined a model for the transcriptional regulation of the NHE-III₁ in the last two decades.^{369,456,512} Figure 20 shows a schematic representation of the regulatory model that includes the binding of different proteins. The NHE-III₁ is a high affinity-binding site for the transcription factor **Sp1**.^{369,397,398} In cells that express low levels of *MYC* the promoter is not occupied by Sp1, but serum stimulation induces *MYC* transcription and results in binding of Sp1.^{369,513} **CNBP**^{353,354} unfolds the G4, binds the Pu27 single strand and has been shown to induce *MYC* expression.^{359,369,514} **hnRNP K** acts as transcription factor, binds the Py27 single strand and activates transcription.^{369,515,516} **NM23-H2** has been shown to bind both the single-stranded Pu27 and Py27 and drive the unfolding of the G4 structure.^{366,367,369,517-519} **Nucleolin** recognizes and binds the G-quadruplex structure and stabilizes the folded conformation, which silences the *MYC* transcription (see also chapter 2.3.2).^{335,338,369}

Connecting the dots of the interactions of all mentioned NHE-III₁-binding proteins allows sketching the regulatory model shown in Figure 20. In earlier studies, it was also proposed that **hnRNP A1** binding to Pu27 is involved in *cMYC*-NHE-III₁ regulation, but due to its weak affinity in competitive assays is now considered not to take part in these processes.^{448,493,520} The role of hnRNP A1 is still discussed for transcriptional regulation in the *KRAS* promoter.⁴⁴⁸

In this model, a pre-requisite for G-quadruplex formation is a local unwinding with strand separation, in particular because of negative supercoiling. Negative supercoiling is typically induced during transcription,^{456,521} which implies that G4 formation is favoured after transcription is already initiated.

2.4 The cMYC Proto-Oncogene Promoter

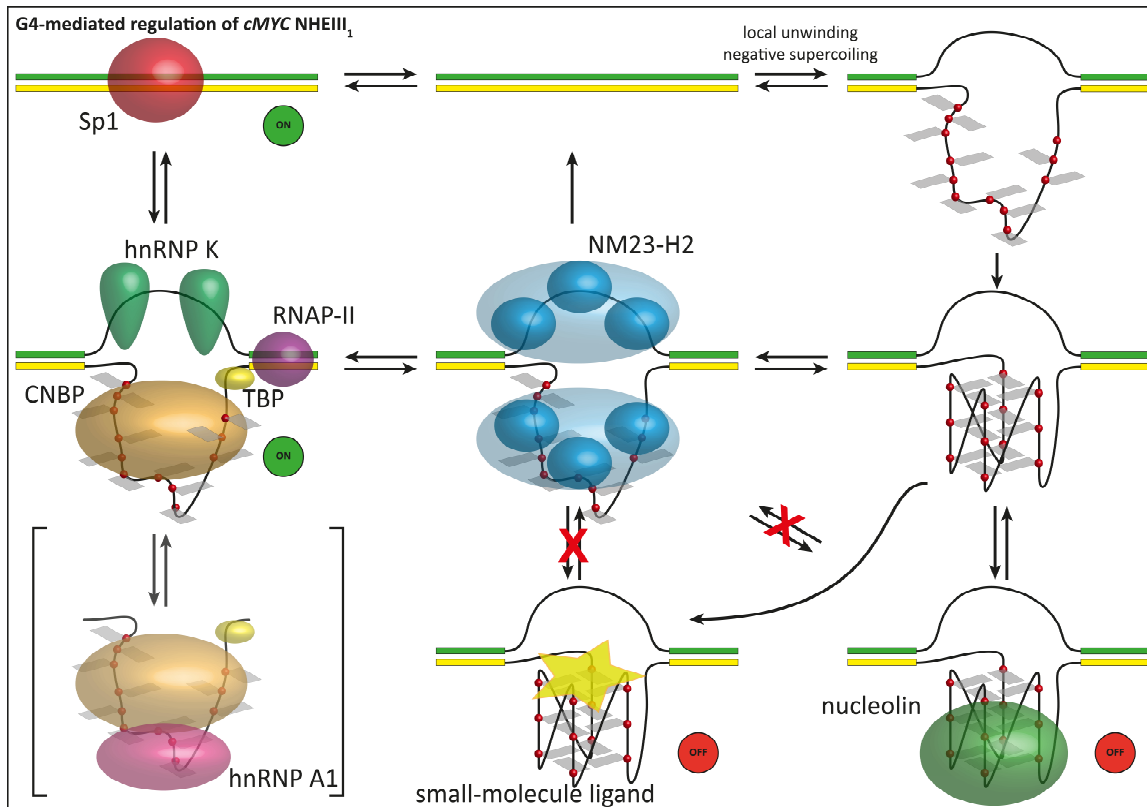


Figure 20: Schematic representation for the model of G4-mediated transcriptional regulation of the NHE-III₁, mainly proposed by Hurley and co-workers.³⁶⁹ **(left)** Transcription factors like Sp1 bind the double stranded B-form DNA and initiate transcriptional activation. A complex of CNBP and TBP (TATA-box) binding protein and hnRNP K binding to the complementary strand enables recruitment of RNA Polymerase-II (RNAPII). Initial studies by Simonsson *et al.* proposed the involvement of hnRNP A1 binding,⁴⁹³ but the binding affinity of hnRNP A1 is comparably weak.^{448,520} NM23-H2 facilitates the unfolding of the G4-fold and enables reformation of the double strand or recruitment of other proteins. **(right)** Negative superhelicity induces strand separation and G4 formation. The G-quadruplex can be processed by destabilizing proteins or can be bound by nucleolin, which then acts as transcriptional suppressor. Binding of small-molecule ligands interferes with protein binding and stabilizes the G4 in a non-protein-bound state.

2.4.3 Structures and Dynamics of the cMYC G-Quadruplex

2.4.3.1 High-Resolution Structures

Within the first report on G-quadruplex formation in the *cMYC* promoter in 2002,^{452,508} two *anti*-parallel structure models were proposed based on DMS-footprinting, namely *chair* or *basket* conformations. This model has been revised in 2004, based on simple CD-spectra and gel-electrophoresis to find that the 3'-terminal part (2345) exclusively forms parallel structures.⁹⁹ These findings already allowed to distinguish the formation of four possible loop-isomers (G-register isomers) for the 2345-conformation. It may be permitted to give a comment at this point from a spectroscopist's or even from a structural biologist's perspective. These two papers have been published from the same authors and claims were made about the structure-based targetability with small-molecule ligands. The false and misleading proposals for structural models without experimental evidence could have been avoided by measuring just a simple CD-spectrum already in the first place. Therefore, this highly cited and nevertheless groundbreaking paper emphasizes the need for at least a basic structural characterization of involved G-quadruplex conformations.

2.4 The cMYC Proto-Oncogene Promoter

Initial NMR assignments for the 27-mer long G-quadruplex forming sequence in the NHE-III₁ region have been made by Phan *et al.* in 2004.¹¹¹ They described two propeller-type, parallel conformations and identified them as conformations 2345 and 1245 (no PDB structures). Yang *et al.* published the first high-resolution NMR structure for cMYC-2345 (PDB:1XAV, 33 G-register isomer) in early 2005.⁴⁶⁰ Later in 2005, Phan *et al.* published a non-canonical G-quadruplex structure with snap-back motif for the 2345 conformation.¹²⁸ A G-to-I mutation in G-tract 3 here is compensated by a replacement with a 3'-terminal G-nucleotide snapped back into the 3'-lower G-tetrad (Figure 21). The NMR structures for cMYC-1234 (2011)⁴⁶¹ and cMYC-1245 (2019)³⁴² then have also been published from the group of D. Yang (Figure 22).

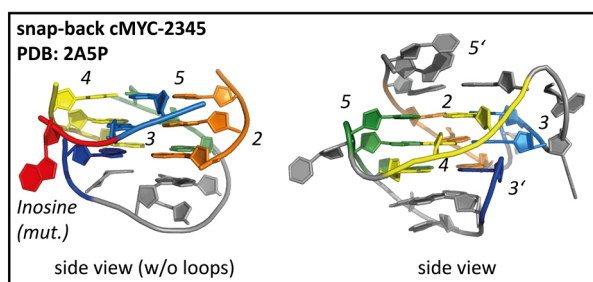


Figure 21: Structure of a parallel snap-back conformation adopted by cMYC-2345 (PDB: 2ASP).¹²⁸ G-tracts are shown coloured (2: orange, 3: blue, 4: yellow, 5: green) loops and capping residues are shown in grey.

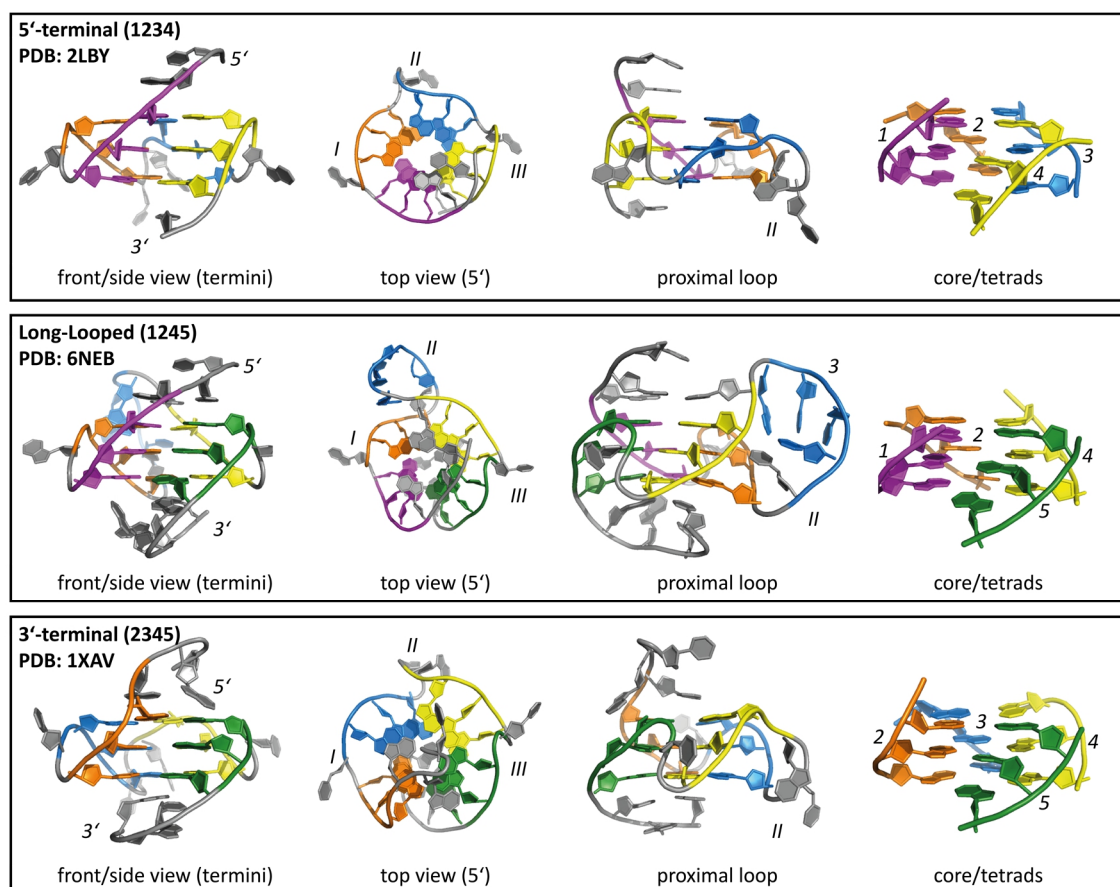


Figure 22: High-resolution NMR-structures of all three relevant G-quadruplex conformations in the Pu27 strand of the cMYC NHE-III₁. All structures feature an *all*-parallel strand orientation, but involve different G-tracts in the tetrad/G4-core formation: a 5'-terminal G4 (1234, PDB: 2LBY)⁴⁶¹, a long-looped G4 (1245, PDB: 6NEB)³⁴² and a 3'-terminal G4 (2345, PDB: 1XAV)⁴⁶⁰. G-tracts are shown in: 1: violet, 2: orange, 3: blue, 4: yellow and 5: green.

2.4 The cMYC Proto-Oncogene Promoter

Finally, in 2018 a structure based on X-ray crystallography was published for cMYC-2345 (Figure 23, RMSD = 2.35 Å, PDB = 6AU4).⁵²² Major deviations for the crystal structure compared to the NMR derived structure concern the 5'- and 3'-terminus. The NMR structure shows a capping coordination that stacks on the top and bottom tetrads, while in the crystal structure the termini are directed outwards. The differences highlight the flexibility of the termini compared to the G4-core. The distal orientation in the crystal structure, however, is most likely caused by packing interactions and dimer stacking. The stabilizing and concealing capping arrangement seen in the NMR structure might be the better reflection of the G4 structure adapted by a DNA oligonucleotide, but the restricted flexibility in a longer sequential context might hamper its formation.

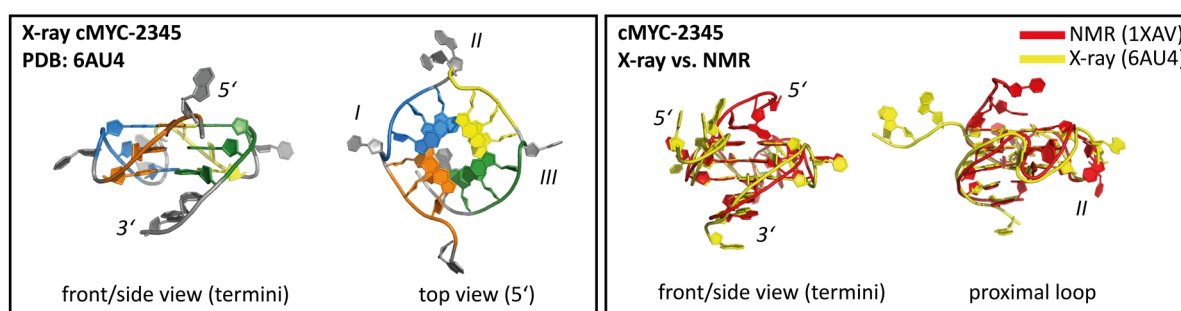


Figure 23: Comparison of the NMR (PDB: 1XAV)⁴⁶⁰ and X-ray (PDB: 6AU4)⁵²² structures. The NMR-structure shows defined capping structures with stacked residues from the 5'-terminus, while the crystal-structure shows more flexible termini that are directed outwards from the G4-core.

2.4.3.2 Thermodynamic Stabilities and Structural Dynamics

For the four loop-isomers (*G-register* isomers), a detailed thermodynamic characterization has been reported based on thermal denaturation (CD melting) and kinetics for folding and unfolding have been derived based on hysteresis.^{99,100} Noteworthy, both studies find melting temperatures (T_m) for each of the single isolated *G-register* mutants that are lower than for the *G-register wt*. For a thermodynamic comparison of *spare-tire* isomers cMYC-1245 and cMYC-1234 also calorimetric studies have been published,^{112,523} which are in line with reported T_m s based on CD melting for all reported cMYC structures.^{111,342,460,461} The plethora of sequence modifications, buffer and DNA concentrations make a detailed comprehensive overview and comparison of all thermodynamic parameters nearly impossible. Furthermore, at physiological $[K^+]$, T_m values are typically >90 °C. Nevertheless, across all published data, a clear trend of melting temperatures for the isolated, stabilized conformations can be drawn as follows:

spare-tire isomers: 1245<1234<<2345<wt; **G-register** isomers (2345): 55<35<53<33<wt

In general, thermodynamic parameters are typically used to evaluate e.g. the effects of ligand binding. However, such thermodynamic analysis is ambiguous and has to be evaluated carefully to prevent false implications.⁵²⁴ This holds particularly true for G-quadruplex forming oligonucleotides that are itself

2.4 The cMYC Proto-Oncogene Promoter

prone to polymorphism and structural dynamics.⁵²⁴ Already the first structural investigations for the cMYC G-quadruplex revealed the importance of a dynamic interplay of possible conformational isomers. Luciferase expression assays showed a two-fold increase in expression levels for individual (mutated) *G-register* isomers (five-fold increase for complete G4-destabilization).⁹⁹ A more recent study that investigated the expression levels with a dual-luciferase assay in various cancer cell lines (MCF-7, T47D, MDAMB 231, HeLa, AGS) found an increased level of promoter activity for mutations that restrict the structural polymorphism of the G4 forming element (1234-only: ~1.5-3 fold, 2345-only: ~2-4 fold).³⁴¹

The role of negative supercoiling has revealed strong implications for a shift of adopted conformations from 2345 towards 1234.^{456,521} Transcriptionally induced negative supercoiling has significant effects in the promoter region up to more than 1000 bases upstream of the TSS. It can affect not only the NHE-III₁ (-100 bases to P₁) but also causes strand separation on the far-upstream element (FUSE, -17000 to P₂).⁴⁵⁶

Sengupta *et al.* showed different affinities of G4-binding proteins that interact with the NHE-III₁ (2.4.2) towards different mutations and truncated sequences that preclude the formation of *spare-tire* isomers.³⁴¹ This study strongly supports a refined model of the transcriptional regulation of the NHE-III₁ that accounts for the *spare-tire* polymorphism, which is in line with the drastically different affinities of the G4-binding protein nucleolin (2.3.2) towards especially the long-looped conformation 1245.^{340,342}

2.4.4 Targeting the cMYC G-Quadruplex

The cMYC transcription factor has a high significance in cancer progression and is still considered undruggable (2.4.1). Different strategies have been proposed to overcome the main obstacles in therapeutic anti-MYC treatment.^{478,525-529} cMYC is a prime example for targeting a promoter G-quadruplex with small-molecule ligands as novel strategy in anti-cancer treatment (2.3.4).^{512,530} As discussed above, most common binding motif is stacking on the top and bottom tetrads. For cMYC a NMR structure of a 2:1 complex with a fluorescent carbazole derivative shows exactly this binding motif (Figure 24).⁵³¹ In this regard it remains challenging to generate binding specificity for a certain G-quadruplex.⁵³² Many different scaffolds have evolved e.g. cationic porphyrins⁵³³, carbazole derivatives,⁵³⁴ benzofuranes,⁵³⁵ benzothiazoles,⁵³⁶ imidazole-benzothiazole conjugates,⁵³⁷ thiazole peptides,⁵³⁸ triazoles,⁵³⁹ 5-nitro-indole derivatives,⁵⁴⁰ binaphthyl-amines,⁵⁴¹ dansyl-guanosine conjugates⁵⁴² and more complex conjugates like a four-leaf clover diaryl-substituted imidazole/carbazole⁵⁴³ and hybrid molecules that link a quadruplex-specific binding site with a duplex minor groove binder (Figure 24).⁵⁴⁴ With this hybrid linker approach, it is possible to generate sequence-specific binding in combination with high G-quadruplex affinity (low nM).⁵⁴⁵

2.4 The cMYC Proto-Oncogene Promoter

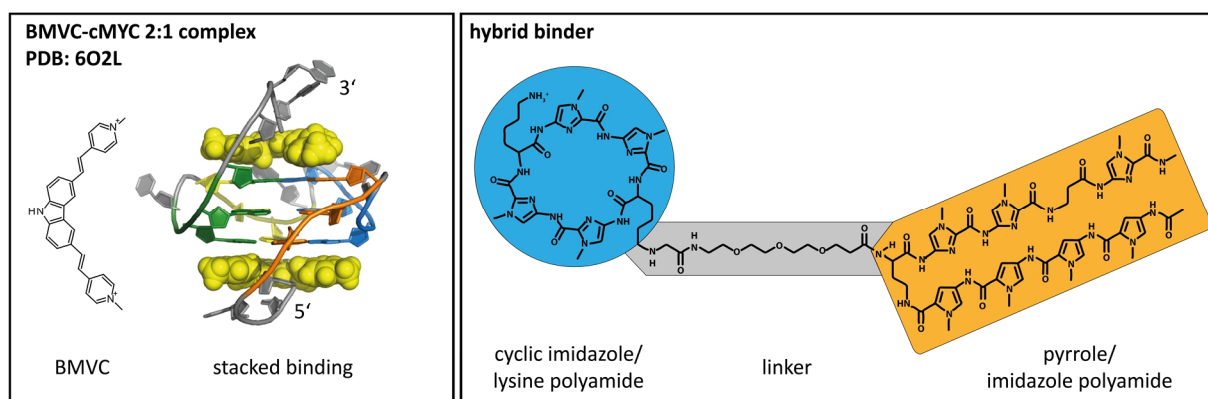


Figure 24: **(left)** High-resolution NMR-structure (PDB: 602L)⁵³¹ of a 2:1 complex of a *cMYC* G4 with BMVC. **(right)** Chemical structure of a hybrid G4-binding molecule. The blue part (cyclic imidazole) is able to recognize and bind the G4-structure with high affinity. The orange part (pyrrole) can be used to generate sequence specific binding of the contiguous flanking DNA segments.⁵⁴⁴

Methods like rational drug design,⁵⁴⁰ small molecule microarrays⁵³⁵ or spectroscopic screening^{543,546} have been used to find these binders and initially evaluate them. One *cMYC* G-quadruplex interacting compound (CX-3543⁵⁴⁷, *Quarfloxin*) has entered phase II clinical trials, but was finally discontinued due to *MYC*-independent effects.^{528,548} Recent pre-clinical studies have found GQC-05⁵⁴⁹, Cz1⁵⁵⁰, IZCZ-3⁵⁴³, DC-34⁵⁵¹, Stauprimidine⁵⁵² and BMH-21⁵⁵³ as promising drug candidates.⁵²⁹ The G4 stabilizer APTO-253 inhibits *MYC* expression and has entered phase I clinical trials.^{554–557}

The compound CX-5461 is a promising possible chemotherapeutic agent that can act as *MYC* repressor.⁵⁵⁸ The cytotoxic effects have been discussed mainly in RNA polymerase I inhibition,^{559,560} but also in topoisomerase I and II inhibition,^{561,562} activating the DNA damage response,⁵⁶³ and CX-5461 is a G4 stabilizer.^{548,564} A disturbed *cMYC* G4 maintenance is a likely explanation for all of the observed major molecular mechanism for the observed subsidiary molecular mechanisms. CX-5461 has entered phase I clinical trials.⁵⁶⁵

Alternative approaches that aim to target the G4 forming sequence of the NHE-III₁ use different oligonucleotide (ODN) or peptide strategies. The 12 aa long peptide KR12C (a derivative of the human cathelicidin peptide LL-37)⁵⁶⁶ binds the *cMYC* G4 in the nanomolar range in cells and promotes apoptosis.⁵⁶⁷ Triplex-forming oligonucleotides (TFOs) can be used to target the Py27 strand.^{498–503} These strategies benefit from increased lifetimes of the ODN after cellular uptake, presumably due to better nuclease resistance of G4 forming ODNs.⁵⁰² Application of Pu27 to different leukaemia cell lines resulted in up to 90% reduction of transcription levels and 60% decrease in protein expression.⁵⁰² Recently, a G-quadruplex-forming RNA motif has been discussed to target nucleolin, which has been suggested as a possible biomarker in lung cancer prognosis.³⁴⁵ Strategies like this could be potentially exploited to saturate G4-binding proteins.

2.4 The cMYC Proto-Oncogene Promoter

2.5 NMR Methods to Study Structure and Dynamics in Nucleic Acids

*“I have not yet lost that sense of wonder, and of delight,
that this delicate motion should reside in all ordinary things around us,
revealing itself only to him who looks for it.”*

E.M. Purcell, Nobel lecture 1952⁵⁶⁸

2.5.1 Structure Determination of Nucleic Acids with Natural Abundance Isotopes

For initial structural screening of G-quadruplex topologies and folding patterns, a combination of NMR-spectroscopic evaluation with CD-spectroscopy is very useful.⁵⁶⁹ CD spectroscopy provides an easy access to determine the overall folding topology and homogeneity of folded states in G4s that help to analyse and assign the ¹H-NMR spectra.^{570–574} The Gabelica group has also developed a mass-resolved CD-setup that allows to disentangle polymorphic G4 species.⁵⁷⁵

Experimental approaches to obtain structural information of nucleic acids with nuclear magnetic resonance (NMR)-spectroscopy mainly rely on inter-proton distance restraints from nuclear Overhauser effect spectroscopy (NOESY).^{576–581} A manifold of homo- and heteronuclear NMR-experiments has been developed that help for the assignment of the relevant proton (in particular N-H1 imino proton) signals.^{582–584} In addition to that, many NMR-experiments are available that yield information on base pair patterns or structural restraints on dihedral angles (via scalar couplings, *abbr.* J-couplings) and coordinates for the relative orientation (via residual dipolar couplings, RDC). The strategies and advantages for these experimental approaches have been extensively reviewed.^{582–584} In particular for DNA G-quadruplexes, NMR is the most important method for structure determination at atomic resolution.^{310,585,586} From 277 PDB-structures (“G-quadruplex *and* DNA”, December 2020) 154 (56%) have been solved by NMR and 123 (44%) by X-ray crystallography (NMR for: “nucleic acids”: 38%; total: 8%). However, most heteronuclear 2D-correlated experiment relies on isotope enrichment for the NMR-active (spin-1/2) nuclei ¹³C and ¹⁵N. Isotope labelled RNA (both selectively and uniformly) is easily accessible via *in vitro* transcription using T7 RNA-Polymerase and the respective isotope labeled NTPs.^{582,584} Preparation of isotope labeled, single stranded DNA via enzymatic methods in appropriate amounts for NMR investigations is not routinely established. For DNA oligonucleotides, isotope labeling is introduced typically via solid-phase oligonucleotide synthesis. Different *in vitro* methods have been proposed and reported, but did not find widespread application for NMR-studies.^{587–593} Therefore in many cases NMR-spectroscopic studies on DNA G-quadruplexes are limited to samples with a natural abundance⁵⁹⁴ of ¹³C (1.1%) and ¹⁵N (0.4%),

2.5 NMR Methods to Study Structure and Dynamics in Nucleic Acids

which challenges the applicability of sophisticated, multidimensional heteronuclear experiments due to the low sensitivity.

Figure 25 shows the G4 relevant correlations and connectivities that can be obtained from 2D-NOESY spectra and heteronuclear multiple bond correlation (HMBC) ^{13}C experiments that work at natural abundance.^{310,595,596} Intra-tetrad H1-H1 and H1-H8 are indicated, expected NOE cross-peak intensity for inter-tetrad correlations are given in Table 2.³¹⁰ Sequential correlations for H8-H1' along the nucleotide chain (5'-3') yield information on the N-glycosidic conformation (*syn/anti*) and differ for *syn-anti*, *anti-syn* or *anti-anti* steps, Figure 25 shows correlations along an *anti-anti* step. The relevant J_{CH} -couplings for H-C-H correlations are given in Hz.^{595,596}

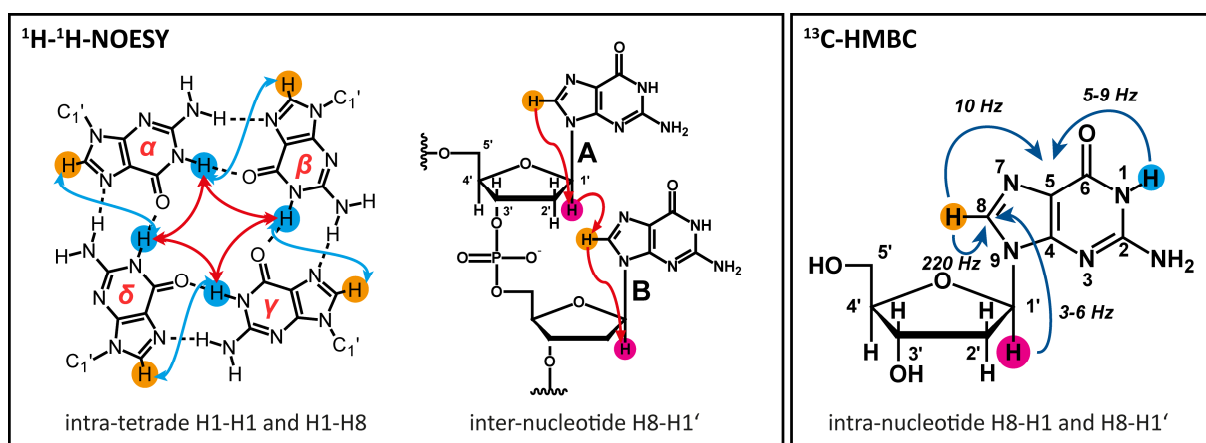


Figure 25: Relevant correlations and connectivities from NMR experiments at natural abundance. NOE connectivities (**left**) and J_{CH} -couplings (**right**) help for spectral assignment of the relevant cross-peaks in 2D-NOESY spectra.

Table 2: Intra- and inter-tetrad NOE-connectivities: expected medium (3.2 – 4.6 Å) and weak (4.6 – 6.0 Å) cross-peak intensities are indicated. Table was taken from³¹⁰ (PDB entries: 2GKU (hybrid) and 143D (*anti-parallel*)).

		same polarity (<i>anti-anti</i>)		opposite-polarity (<i>syn-anti</i>)		opposite-polarity (<i>anti-syn</i>)	
		G_{α}		G_{α}		G_{α}	
(n+1)		H1	H8	H1	H8	H1	H8
G_{α}	H1	medium		weak		medium	
	H8		weak		weak		
G_{β}	H1						
	H8						medium
G_{γ}	H1			weak			
	H8						
G_{δ}	H1	weak	weak	medium		medium	
	H8						

Guanine amino groups are essential for hydrogen bond formation in G-tetrads (Figure 25). They have rotational freedom around the C-N² bond with a rotational frequency between 0.2 and 1.1 kHz (in the millisecond time regime), which affects the linewidths of the signals. Lineshape analysis of amino proton signals from stacked guanines involved in G-quadruplex formation yields useful information on rotation rates and serves as probe for local structural dynamics.⁵⁹⁷ Recently, ^{13}C -detected experiments

(“amino-NOESY”) have been developed that make use of the amino group dynamics.^{583,598} The sensitivity of these amino-experiments requires ^{13}C isotope labelling, but nevertheless these experiments will be very useful especially for G4s.

2.5.2 The NMR-Timescale: Methods to investigate Dynamics

Methods to investigate dynamics with NMR on broad timescales in particular for application on nucleic acids have been extensively reviewed,^{599–604} most recently in a comprehensive overview from the group of K. Petzold⁶⁰⁵.

Spin-relaxation is a non-spontaneous process that is caused by magnetic field fluctuations. Depending on the source of these fluctuations, distinct relaxation mechanisms can be differentiated. The main source of fluctuating fields is molecular tumbling, which slows down with increasing molecular size. T_1 -relaxation (spin-lattice relaxation, R_1) is caused by fluctuations only in the (x,y)-direction (static frame), while T_2 -relaxation (spin-spin relaxation, R_2) is influenced by fluctuations in *any* direction. T_1 therefore depends on fluctuating fields near the Larmor-frequency ν_0 , while T_2 is affected by fluctuating fields at any frequency. Relaxation is thus (i) dependent on the magnetic field (B_0), which correlates with the nuclear Larmor-frequency ν_0 of the respective spins; and (ii) dependent on the molecular size and temperature, which both affect the molecular tumbling. Assuming that molecular motion is isotropic, the rotational correlation time τ_c for a molecule is defined as the average time it takes to rotate through one radian. τ_c therefore is a measure of the average, size-dependent molecular tumbling.⁶⁰⁶ With increasing molecular size (hence increasing τ_c) R_1 reaches an optimum (shortest relaxation time, field dependent) at $\nu_0^{-1} \approx \tau_c$, and afterwards gets longer again. R_2 instead is steadily decreasing with increasing τ_c and is also affected by chemical exchange and conformational dynamics. The linewidth of NMR signals is proportional to T_2^{-1} , thus dynamics in the time regime of τ_c modulate the linewidth and cause severe line broadening.^{607–609}

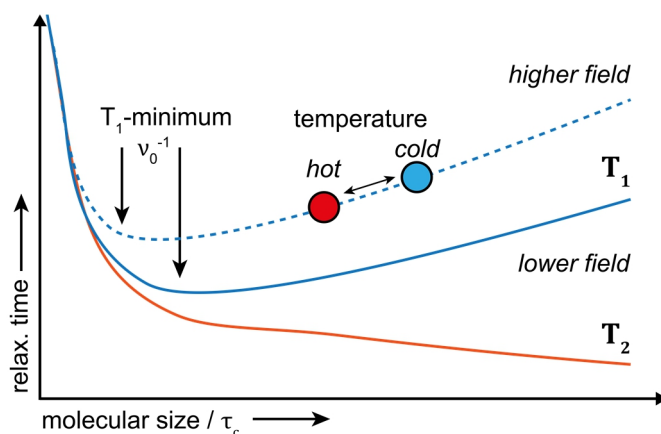
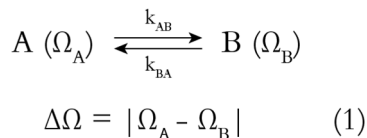


Figure 26: Qualitative trends for relaxation rate constants T_1 and T_2 in dependency of the rotational correlation time τ_c , magnetic field strength and temperature. τ_c is field-independent, thus increasing magnetic field strength shifts ν_0 (thereby $\Delta\Omega$) in relation to k_{ex} for a given molecule. Figure was adapted from literature references.^{607–609}

2.5 NMR Methods to Study Structure and Dynamics in Nucleic Acids

The NMR-timescale (Figure 27) can be roughly divided in three time regimes for dynamics with the following definition:⁶⁰⁵ In a given system in dynamic exchange between two states A and B with the chemical shifts Ω_A and Ω_B , A and B have a chemical shift difference of $\Delta\Omega$ (1). The exchange rate between A and B (k_{ex}) is the sum of the forward (k_{AB}) and reverse (k_{BA}) rate constant (2).



$$k_{\text{ex}} = k_{AB} + k_{BA} \quad (2)$$

In this system, the time regimes are commonly defined as: **slow exchange regime**, if $k_{\text{ex}} \ll \Delta\Omega$ (seconds and slower); **intermediate exchange regime**, if $k_{\text{ex}} \approx \Delta\Omega$ (μs - ms); and **fast exchange regime**, if $k_{\text{ex}} \gg \Delta\Omega$ (ps - ns).⁶⁰⁵ As discussed above, the time regimes are dependent on τ_c and do vary, depending on the molecular size. For a given molecular size, the regimes can be slightly shifted with experimental parameters, when varying temperature (k_{ex} is temperature-dependent) or magnetic field strength ($\Delta\Omega$ in Hz is field-dependent).

Dynamic processes faster than τ_c (e.g. bond vibration, libration, angle fluctuation) are averaged out in solution state NMR and do not cause distinguishable states that can be directly detected. Information on dynamics in this time regime can be obtained from R_1 and R_2 relaxation analysis and NOEs.^{602,610-612} Dynamic processes in the range of/slightly slower than the molecular tumbling rate (supra- τ_c , ns- μs) have for long remained a blind spot for NMR.⁶¹³ This gap now can be probed with experimental approaches that make use of residual dipolar couplings (RDCs)^{614,615} or cross-correlated relaxation (CCR)⁶¹⁶⁻⁶¹⁸ rates. In the μs - to ms - and s -regime **C**arr-**P**urcell-**M**eiboom-**G**ill (CPMG)⁶¹⁹⁻⁶²¹ sequences and **C**hemical **E**xchange **S**aturation **T**ransfer (CEST)^{622,623} are the most important experimental approaches to investigate dynamics. In the ms to s regime **E**Xchange **S**pectroscopy (EXSY)^{624,625} can be used to investigate the dynamics of spins that are connected by chemical exchange. In the time regime of seconds and slower, the acquisition of single experiments is faster than the dynamics of the observed system. Here, the information can be obtained in **real-time** by recording a series of spectra.⁶²⁶⁻⁶²⁸

Independent of the strategy that is used to prepare coherent dynamics for **time-resolved** NMR measurements (see following chapter) the fast acquisition and sensitivity of the NMR experiments remains a limiting factor. This holds particularly true for multidimensional, mostly heteronuclear-correlated (^{13}C , ^{15}N) experiments. Recent advances⁶²⁹ in acquisition (SOFAST and BEST)⁶³⁰⁻⁶³³ and ultrafast methods^{634,635}, sensitivity enhancement (hyperpolarized water)⁶³⁶⁻⁶⁴⁰ and combinations of ultrafast acquisition with hyperpolarization,^{641,642} have narrowed the timeframe of observable real-time dynamics down to seconds.^{643,644}

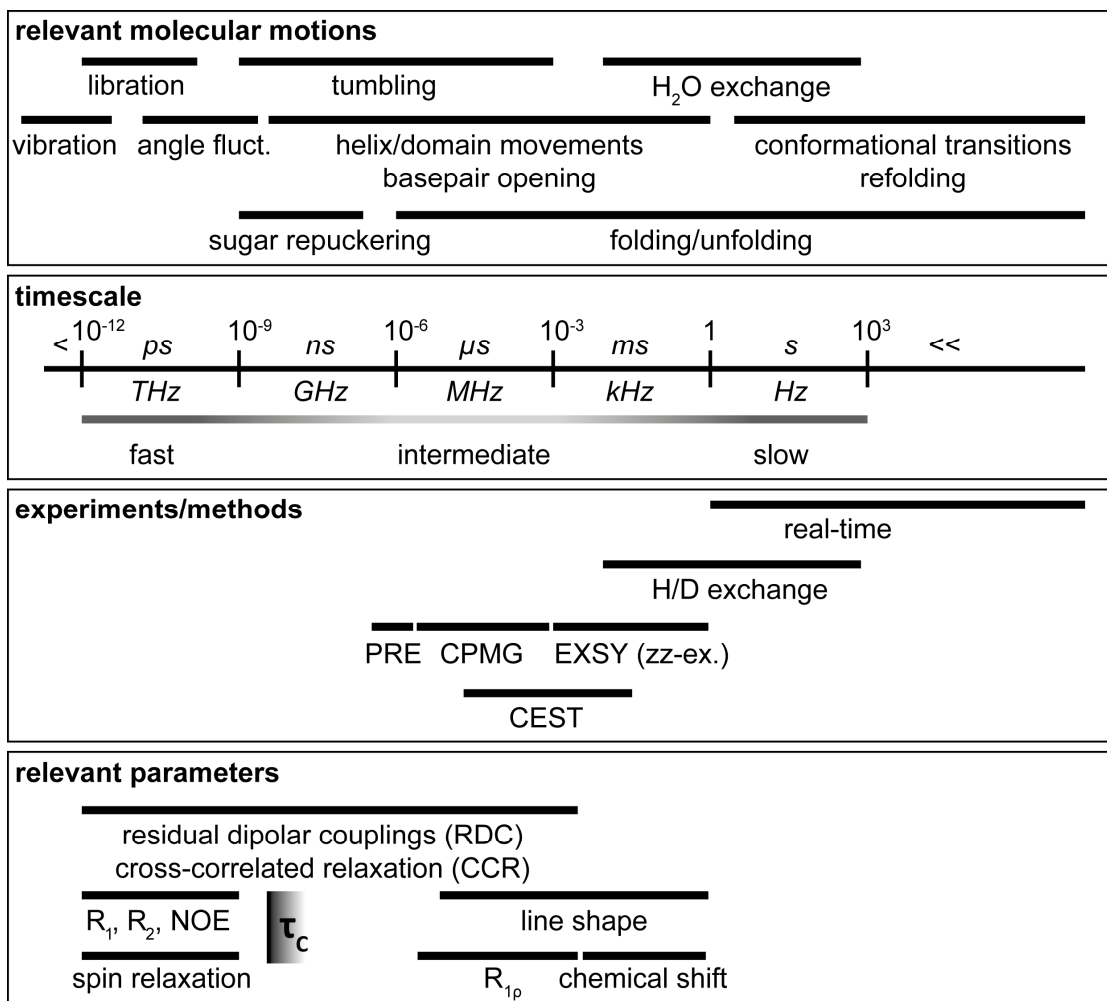


Figure 27: The NMR-timescale^{602,605,613}, overview of: the **relevant molecular motions** for nucleic acids in solution, **timescale and frequencies** with indicated exchange-regime (fast, intermediate, slow), NMR **methods and experiments** to probe directly specific dynamic processes, and **relevant parameters** in NMR experiments that influence and modulate the shape and intensity of NMR signals. $R_{1\rho}$ is defined as relaxation in the rotating frame, measurement of $R_{1\rho}$ is strictly rather a method than a parameter.^{645,646}

2.5.3 Real-Time NMR: Non-Equilibrium Dynamics

Two different types of dynamics shall be differentiated in this chapter: reversible and irreversible events. The folding of nucleic acids follows an irreversible trajectory, while steady-state refolding or e.g. domain motions have reversible trajectories. In general, NMR spectroscopy detects a large ensemble of spins in concentrated samples. This is different to single-molecule methods like **Fluorescence Resonance Energy Transfer (smFRET)**^{647,648} microscopy, where single events can be detected. This allows for the observation of both kind of dynamics. In solution state NMR, all dynamic changes within a molecule are ensemble averaged. Time-resolved studies of both kind of structural dynamics thus require coherent progression of the underlying molecular motions. To create a coherent evolution, different ways to prepare non-equilibrium, excited or *meta*-stable states are feasible. Time-resolved NMR typically tracks the relaxation of these states back to their thermodynamic equilibrium or steady states. While this thesis was written, a review with contributions from me (the author of this thesis) was published as preprint.⁶²⁸

Temperature or pressure change

The intensive properties temperature (T) or pressure (p) influence the structural and physicochemical properties of biomacromolecules, which makes them suitable variables to investigate thermodynamic and kinetic parameters.⁶⁴⁹ High pressure or temperature results in e.g. structural changes (pressure-related stabilization or unfolding; or thermal denaturation)^{183,650–652}, or disassembly/disaggregation of molecular complexes⁶⁵³. A steady application of high pressures^{654–656} (up to 3 kbar⁶⁵⁷) forces biomolecular systems to populate excited states and shift conformational equilibria, which yields useful, otherwise hidden⁶⁵⁸ information.^{184,659,660}

Technical setups for rapid *in situ* changes in temperature (**T-jump**: microwave-heating^{661,662}, laser-induced (MAS ssNMR)^{663,664}, rf-heating^{665,666} or capacitively coupled⁶⁶⁷) or pressure (**p-jump**^{668–673}) have been reported that can be used to investigate the relaxation back to thermodynamic equilibrium. Only a limited number of T-jump NMR experiments has been reported for biomolecules. Akasaka *et al.* have pioneered an experiment for temperature-induced folding of RNase A at low pH ($\Delta T = 16\text{ }^{\circ}\text{C}$ in 6 s).⁶⁷⁴ More recently, application of T-jump NMR for protein folding has been demonstrated for the investigation of cold-denatured Barstar ($\Delta T = 20\text{ }^{\circ}\text{C}$ in 500 ms).⁶⁷⁵ The Bax group used p-jump NMR experiments to study ubiquitin protein folding with pressure jumps up to 2.5 kbar.^{669,676,677} Numerous studies also applied high-pressure and p-jump experiments to study β -amyloid aggregation.^{653,678–680}

Rapid-mixing

Rapid-mixing measurements allow the investigation of kinetics that can be initiated by different means.⁶⁸¹ Yushmanov and Furó reported on a rapid-mixing setup with stopped-flow design⁶⁸²; a device for rapid-mixing of components within a NMR-sample has been reported by Mok *et al.*⁶⁸³ and has been refined recently⁶⁸⁴. Using rapid-mixing, protein (re-)folding has been investigated e.g. after dilution of a chemically denatured (guanidinium chloride) state⁶⁸⁵ (vice versa for unfolding⁶⁸⁶) or pH-change^{687,688}.^{689–691} Folding of nucleic acids after rapid-mixing has been investigated e.g. for RNA riboswitches by inducing their formation with addition of their specific ligand,⁶⁹² for ribozymes by addition of Ca^{2+} ,⁶⁹³ for tRNA maturation after addition of cell extract,⁶⁹⁴ for a DNA i-motif with induced pH-change⁷⁶ and for a telomeric DNA G-quadruplex by inducing its formation with addition of K^{+} .¹⁴⁴

Photochemical trapping

Instead of rapid mixing, a binding event can also be initiated after releasing a trapped (caged) ligand with light. This has been demonstrated for ligand-induced riboswitch folding after *in situ* photochemical release of photocaged hypoxanthine.⁶⁹⁵ For protein folding a similar approach has been reported, utilizing a light-trigger; therefore a photolabile nitrophen chelator has been used to release Ca^{2+} and study the ion-induced refolding of α -lactalbumin.^{696,697}

2.5.4 Conformational Selection with Photolabile Protecting Groups

Light is an excellent trigger to obtain selective (spatio-) temporal control over chemical processes in a given experimental setup. Different chemical strategies have been developed to make use of light absorption as a versatile tool for reaction control.^{698–701} These tools can be reversible *photoswitches*^{702–705} that react to light (e.g. in *E/Z* isomerization along a -N=N- double bond in azobenzene dyes or a -C=C- double bond within stilbene chromophores) or irreversible photocleavable moieties (*photosensitive/photolabile protecting groups, PPGs*)⁷⁰⁶ (Figure 28). As they help to block molecular interactions and suppress reactivity at distinct sites in a molecule temporarily, photolabile protecting groups have been named *photocages*; their photolytic “release” hence is called *uncaging*.

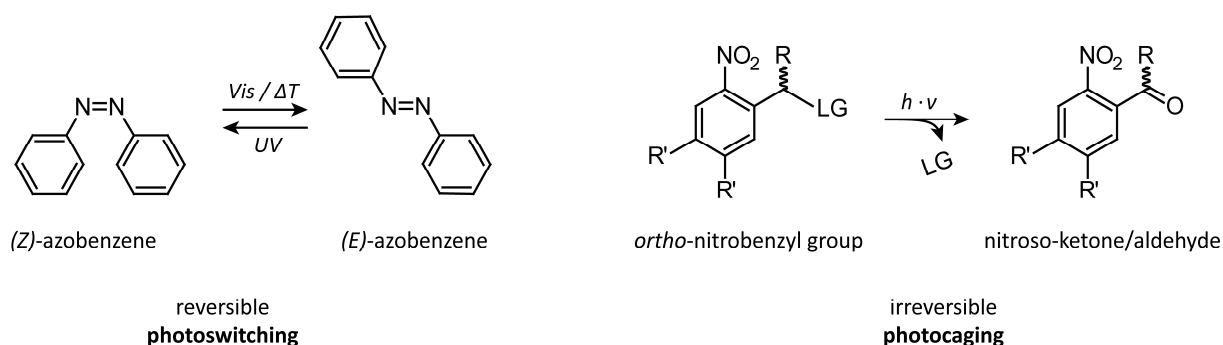


Figure 28: Paradigm reaction schemes for reversible (photoswitches, left) and irreversible (photosensitive protecting groups/photocages, right) interactions with light. Azobenzene and *ortho*-nitrobenzyl derivatives have established the largest relevance for widespread applications. [Substitutions (non-comprehensive): [R= H (**O**r**t**h**o** **N**itro **B**enzyl), Methyl (**N**itro **P**henyl **E**thyl), COOH (**C**arboxy **N**itro **B**enzyl), (...); R' = H, -OMe, -CH₂OCH₂-]

Photolabile protecting groups for use in nucleic acid/sugar chemistry have first been reported in 1977.⁷⁰⁷ This very first study already introduced *ortho*-nitrobenzyl (ONB) (Figure 28) as photocleavable group to block cAMP (Figure 29). The ONB scaffold is synthetically easy accessible and can be fine-tuned for light-stability or wavelength specificity, hence it has been commonly established in widespread applications. After photocleavage ONB (R=H) yields nitrosoaldehydes, nitro-phenyl-ethyl (NPE, R=Me) yields nitrosoketones, which are less cell-toxic. NPE can also be deprotected faster than ONB and therefore is typically the cage of choice.

Photocages can be attached to backbone phosphate groups^{708–710}, nucleobases^{29,711–716} or ribose rings at the 2'-OH^{717,718}; also photocleavable linkers have been introduced both in the phosphate backbone^{719,720} and at nucleobases⁷²¹. All these strategies (Figure 29) can be used to block molecular interactions and thereby influence the structure and functions of DNA/RNA.

Photocaging of nucleobases prevents base pairing interactions at the Watson-Crick and/or Hoogsteen interface in two ways: (i) sterically and (ii) by locking the T/U: N³-C⁴-O⁴ and G: N¹-C⁶-O⁶ bond in an enolic form (-N=C-O-), which alters the hydrogen-bond donor/acceptor pattern. This approach has been used to destabilize certain secondary or tertiary conformations in RNA and DNA oligonucleotides.^{29,712,713,716}

2.5 NMR Methods to Study Structure and Dynamics in Nucleic Acids

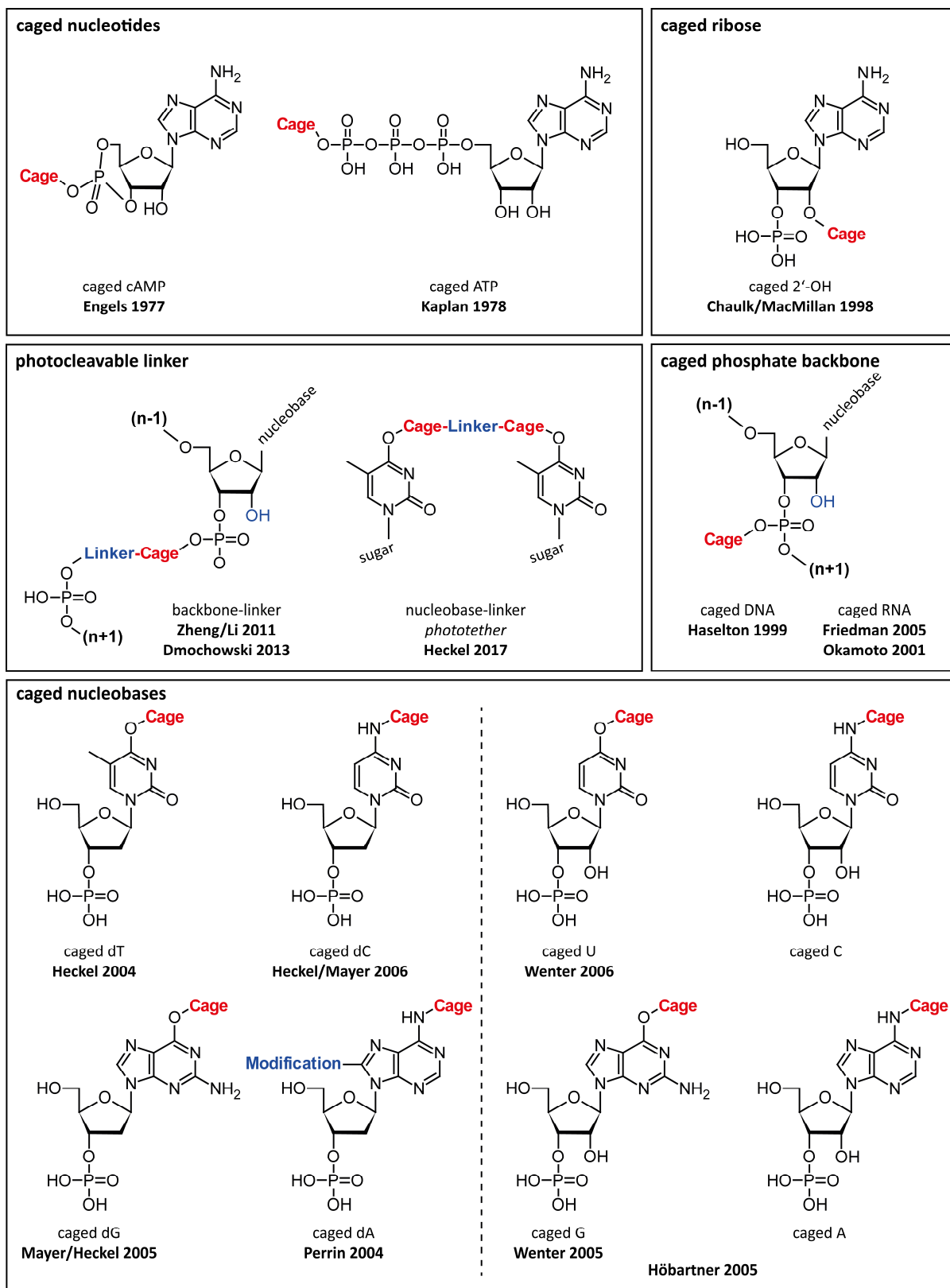


Figure 29: Overview of caging strategies for DNA/RNA nucleosides: early approaches of caged-ATP^{707,722}, 2'-caged ribose^{717,718}, caged phototethers^{719–721}, caged phosphate backbone^{708–710} and the entire set of caged nucleotides for DNA^{711,714–716} and RNA^{712,713,723}.

Blocking biological function does not require a complete distortion of the functional fold of nucleic acids, but can be achieved even within their native conformation and single photocages can be sufficient to block e.g. ribozyme cleavage.⁷¹⁷ A sufficiently large destabilization with reasonably sparse

2.5 NMR Methods to Study Structure and Dynamics in Nucleic Acids

incorporation of photolabile protecting groups that prevents secondary structure formation of oligonucleotides has been first demonstrated for a thrombin binding aptamer (TBA, Figure 30).²⁹ TBA adopts a two-tetrad G-quadruplex structure in its binding competent state²⁹; suppression of a three-tetrad G-quadruplex formation then also has been achieved for an oligonucleotide from a telomeric sequence⁷¹⁶. However, thermodynamically, photocaging base pair interactions only causes local structure disturbance^{724–726} and a complete suppression of secondary or even tertiary conformations remains challenging.⁷¹³ In this context, alternative strategies that are making use of cyclized oligonucleotides via phototethers are powerful improvements for conformational caging (Figure 30).⁷²¹

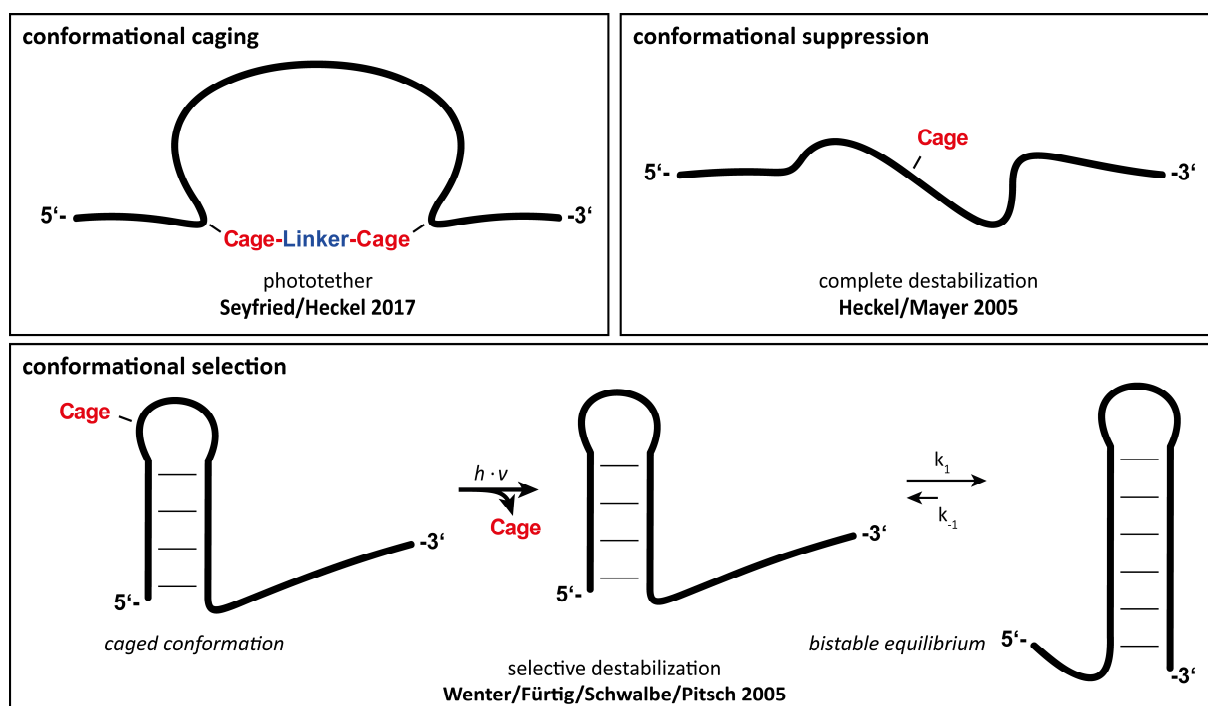


Figure 30: Different caging strategies for DNA/RNA oligonucleotides: **conformational caging** using phototethers⁷²¹, **conformational suppression** or destabilization^{29,716} and **conformational selection** or trapping with specific site-directed, local destabilization⁷¹².

The flat conformational energy-landscapes for nucleic acids (2.2.1) often causes the formation of competing conformations (2.1.2, 2.1.3). A well-studied example^{627,712,723,727–729} for this are two 20-mer RNA oligonucleotide sequences (I^{730} , II^{731}) that can each form a bistable two-state equilibrium system of two coexisting hairpin conformations. The two conformations are separated by only a small enthalpy difference [$(\Delta H_I = -5.9 \text{ kcal}\cdot\text{mol}^{-1})^{712}$; $(\Delta H_{II} = -5.9 \text{ kcal}\cdot\text{mol}^{-1})^{723}$] and incorporate different residues into base pair formation. In their first of its kind publication, Wenter *et al.*⁷¹² selected a single conformation by attaching a photocage to a residue that is in the loop for one conformation, but in the base paired stem for the other conformation. This selective destabilization (Figure 30) led to a homogenous population of a single folded conformation. A quantitative photolysis enabled a complete relaxation to the equilibrium populations.

General Introduction

3 Materials and Methods

3.1 DNA Oligonucleotides and Sample Preparation

All used DNA oligonucleotides have been HPLC-purified, desalted via centrifugation over a vivaspin (Sartorius) centrifugal concentrator with 1 kDa molecular-weight cut-off, lyophilized and dissolved in ddH₂O. Salt-free samples for K⁺-induced folding have been repeatedly heated to 95 °C and diluted with 5 M LiCl solution. Concentrations have been determined with UV/vis absorption at 260 nm. Molar extinction coefficients (ϵ_{260}) were calculated using a nearest-neighbour method. Unmodified DNA oligonucleotides have been purchased by *eurofins Genomics*, Ebersberg (Germany) and HPLC-purified by the manufacturer. Modified (photocaged) DNA oligonucleotides have been synthesized in the group of Prof. A. Heckel. Dean-Paulos Klötzner (^{DK}, synthesis 1) or Anja Blümmler (^{AB}, synthesis 2).

3.1.1 Overview of DNA Oligonucleotides

Table 3: DNA oligonucleotide sequences used for the experiments presented in this thesis. (X = (R)-NPE-dG), (Synthesis: DK = Dean-Paulos Klötzner, AB = Anja Blümmler), (f = caged in folded state, uf = caged in unfolded state), G-to-T mutations in bold, G-tracts underlined.

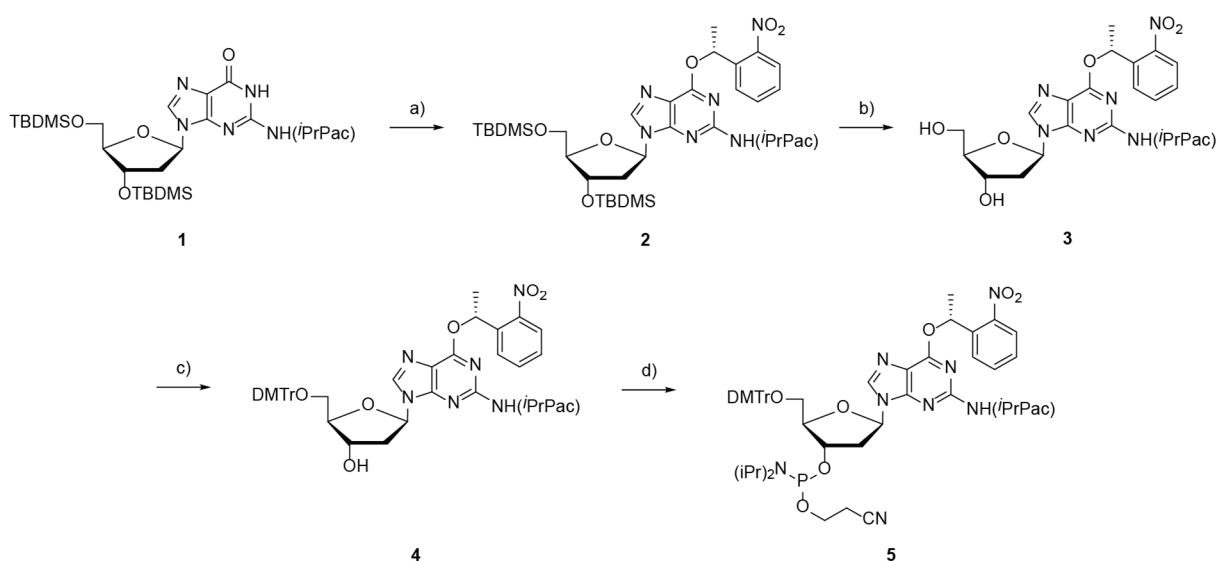
Origin	Length	Conformation	Modification	Sequence	
<i>cMYC</i>	18	2345-53	--	AGGGT <u>TGGG</u> AGGGTGGGT	
		2345-33	--	AGGGTGGGTAGGGTGGGT	
		2345-X3	--	AGGGTGGGGAGGGTGGGT	
		2345-wt	--	AGGGTGGGGAGGGTGGGG	
		1234	--	TGGGAGGGT <u>TGGG</u> AGGGT	
	22	wt-53	--	TGGGAGGGT <u>TGGG</u> AGGGTGGGT	
		1234	--	TGGGAGGGT <u>TGGG</u> AGGGT TTTT	
		1245	--	TGGGAGGGT TTTT AGGGTGGGT	
		2345	--	TTTT AGGGT <u>TGGG</u> AGGGTGGGT	
	18	2345-53 (f)	1xNPE ^{DK}	AGGGT <u>TGGG</u> AGGGTGGGT	
		2345-33 (f)	1xNPE ^{DK}	AGGGTGGGTAGGGTGGGT	
		2345-53' (f)	2xNPE ^{DK}	AGGGT XGGG AGGGTGGG X	
		2345-53 (uf)	3xNPE ^{DK}	AGGGT TGXGAGXGTGXGT	
		2345-X3 (uf)	3xNPE ^{AB}	AGGGTGG XGAGXGTGXGT	
		1234 (uf)	3xNPE ^{AB}	TGGGAG XGTGXGAGXGT	
	22	1234 (f)	1xNPE ^{AB}	TGGGAGGGT <u>TGGG</u> AGGGT GXGT	
		1245 (f)	1xNPE ^{AB}	TGGGAGGGT TGXG AGGGTGGGT	
		wt-53 (uf)	3xNPE ^{AB}	TGGGAGGGT TGXGAGXGTGXGT	
		1234 (uf)	2xNPE ^{AB}	TGGGAGGGT TGXGAGXGTTTTT	
		1245 (uf)	3xNPE ^{AB}	TGGGAG XGTTTTT AG XGTGXGT	
		2345 (uf)	3xNPE ^{AB}	TTTT AGGGT TGXGAGXGTGXGT	
	<i>hTERT</i>	20	53, parallel (f)	--	A <u>TGGG</u> AGGGTCTGGGAGGGC
			35, hybrid (f)	--	AGGGTATGGGCTGGGAGGGC
			53, parallel (f)	2xNPE ^{DK}	AXGGG AGGG XCTGGG AGGGC
35, hybrid (f)			2xNPE ^{DK}	AGGG XXGGG CTGGGAGGGC	

3.1.2 Synthesis of Photocaged DNA Oligonucleotides

The protocols presented here in the following are kindly provided by the respective synthesist that have conducted the synthesis as collaboration partners within the presented projects. The protocols are taken from the Supporting Information of Grün *et al.* **2020**⁷³² and Grün *et al.* **2021**⁷³³, where also further details and analytics can be found. In general, chemical synthesis is an fascinating process.^{734,735}

Synthesis 1, performed and reported by Dean-Paulos Klötzner (Grün *et al.* 2020)

The reactions presented were performed under argon atmosphere using dry solvents. NMR analysis of new compounds were performed on Bruker AV500 and DRX600 MHz spectrometers. Assignment of proton chemical shifts was done via ¹H-¹H-COSY experiments. For flash chromatography silica gel 60 by Macherey-Nagel was used. Compound **1** was prepared according to Mayer *et al.*⁷¹⁶.



Scheme 1: Synthesis of ^(R)-NPE dG phosphoramidite. a) (*S*)-1-(2-Nitrophenyl)ethanol (prepared as described in the literature⁷³⁶), PPh₃, DEAD (40 wt% in toluene), THF, 65%; b) TBAF, HOAc, THF, 78%; c) 4,4'-Dimethoxytrityl chloride, pyridine, 82%; 2-Cyanoethoxy-*N,N*-diisopropylaminochlorophosphine, (*i*Pr)₂NEt, DCM, 71%.

Synthesis of compound 2 (65% yield):

4.09 g DEAD solution (40 wt% in toluene, 9.39 mmol, 1.5 eq.) was added dropwise to a solution of 4.21 g of compound **1** (6.26 mmol, 1 eq.), 1.15 g (*S*)-1-(2-nitrophenyl)ethanol (6.89 mmol, 1.1 eq.) and 2.46 g PPh₃ (9.39 mmol, 1.5 eq.) in 30 mL dry THF. The solution was stirred for 60 minutes at room temperature, concentrated and diluted with 200 mL DCM and 200 mL brine. The aqueous layer was extracted with 100 mL DCM (2x). The combined organic phases were dried over MgSO₄. The solvent was evaporated and the residue purified via column chromatography (cyclohexane/EtOAc 3:1 → 1:1) to give **2** as a yellowish foam.

Synthesis of compound 3 (78% yield):

3.3 g of compound **2** (4.02 mmol, 1 eq.) were dissolved in 40 mL dry THF. 1.38 mL Acetic acid (24.11 mmol, 6 eq.) and 12.06 mL TBAF solution (1 M in THF, 12.06 mmol, 3 eq.) were added. The reaction mixture was stirred at room temperature for 23 hours. The solvent was evaporated and the residue purified via column chromatography (DCM/MeOH 99:1 → 92:8) to give **3** as a yellowish foam.

Materials and Methods

Synthesis of compound 4 (82% yield):

1.87 g of compound **3** (3.12 mmol, 1 eq.) were dissolved in 50 mL dry pyridine. 1.28 g 4,4'-Dimethoxytrityl chloride (3.79 mmol, 1.2 eq.) were added in portions. The reaction mixture was stirred at room temperature for 3 hours. The solvent was evaporated and the residue diluted with 200 mL DCM and washed with 200 mL saturated sodium bicarbonate solution. The aqueous layer was extracted with 100 mL DCM (2 x). The combined organic layers were dried over MgSO₄ and evaporated. The residue was purified via column chromatography (DCM/MeOH/Et₃N 98:1:1) to give **4** as a yellowish foam.

Synthesis of compound 5 (71% yield):

1 g of compound **4** (1.12 mmol, 1 eq.) and 950 μ L (*i*Pr)₂N₂Et (5.59 mmol, 5 eq.) were dissolved in 15 mL dry DCM. 529 mg 2-Cyanoethoxy-*N,N*-diisopropylaminochlorophosphine (2.23 mmol, 2 eq.) were added. The reaction mixture was stirred at room temperature for 1 hour. The solution was diluted with 150 mL DCM and washed with 150 mL saturated sodium bicarbonate solution. The aqueous layer was extracted with 100 mL DCM (2x). The combined organic layers were dried over MgSO₄ and evaporated. The residue was purified via column chromatography (cyclohexane/EtOAc 4:1 \rightarrow 1:1 \rightarrow 1:2) to give **5** as a yellowish foam. The column was packed with cyclohexane/EtOAc 4:1 + 0.5% Et₃N while the eluent was prepared without Et₃N.

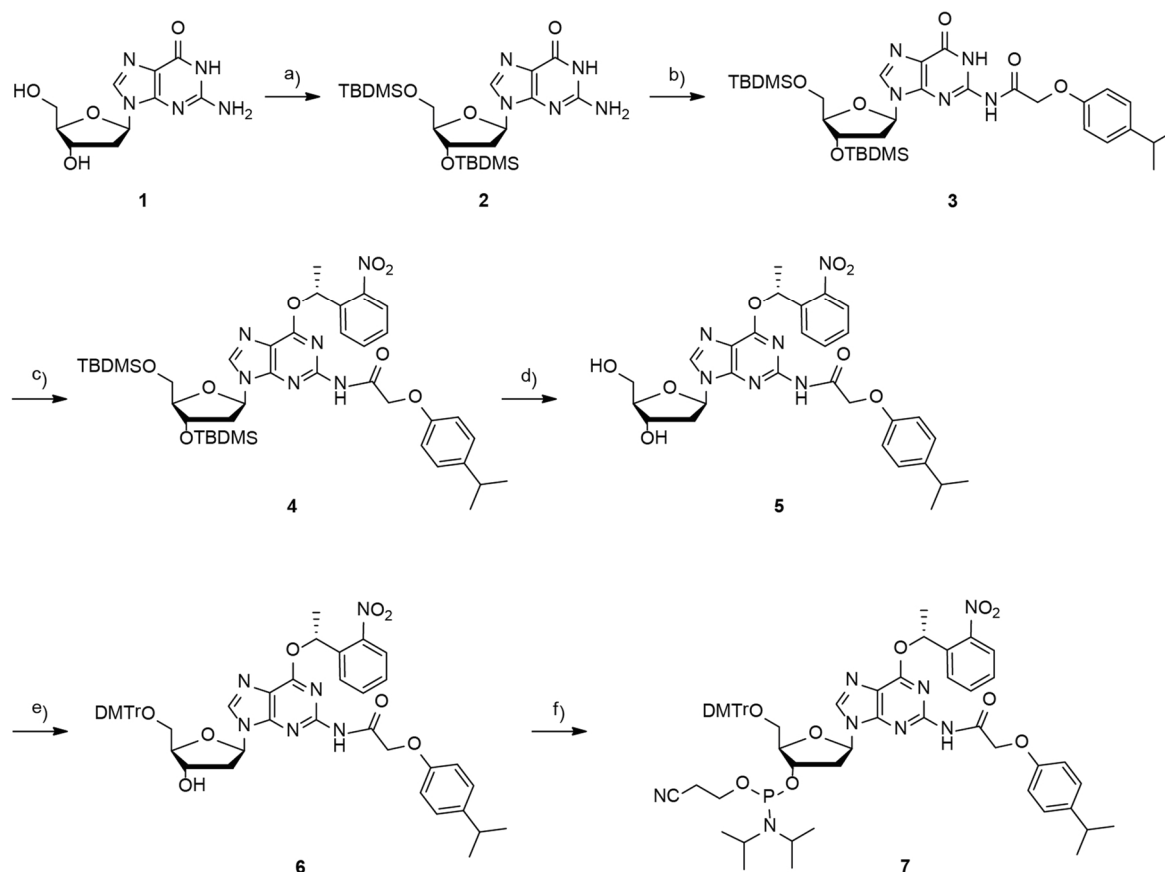
Oligonucleotide synthesis

Photocaged oligonucleotides were synthesized on an Äkta Oligopilot in 18 μ mol (S1), 16 μ mol (S2), 13 μ mol (S3), 16 μ mol (S4) and 19 μ mol (S5) scales. The following reagents were purchased from Sigma Aldrich: DMT-dG-(tac)-; DMT-dA-(tac)-; DMT-dC-(ac)-; and DMT-dT phosphoramidites and Fast Deprotection Cap 2 containing 4-*tert*-butylphenoxyacetic anhydride. 5-(Ethylthio)-1*H*-tetrazole was used as activator. 500 Å DMT-dT-CPG from Glen Research, 500 Å DMT-dC-(tac)-CPG from Millipore, 500 Å CUTAG CPG and 500 Å DMT-dG-(tac)-CPG from Sigma Aldrich were used. All caged oligonucleotides were synthesized DMTr-ON. Deprotection of S1, S2, S4 and S5 was performed using 6.5 mL NH₄OH for 1.5 hours at 65 °C while S3 was deprotected using 6.5 mL *t*-BuNH₂:water (1:3) for 4 hours at 60 °C. After deprotection the solvent was evaporated and the crude product purified via RP-HPLC using a MultoKrom 100 – 5 C18 column (dimensions: 20 · 250 mm, gradient: 5-40% MeCN in 0.1 M TEAA buffer pH = 7 in 44 min, flow: 10 mL/min) on an HPLC from Young Lin Instruments with SP930D pumps and a UV730D detector. After evaporation of the solvent, the DMTr group was removed with 80% HOAc for 20 min at room temperature. The resulting oligonucleotides were purified again via RP-HPLC using the conditions mentioned before.

Synthesis 2, performed and reported by Anja Blümmler (Grün *et al.* 2021)

All reactions were performed under argon atmosphere using dry solvents purchased from *Acros Organics* or *Sigma Aldrich*. Reagents were purchased from *Acros Organics*, *Sigma Aldrich*, *ChemPur*, *TCl*, *Alfa Aesar* or *Carbosynth* and used without further purification. For flash chromatography the used silica gel was purchased from *Macherey-Nagel* (particle size: 40-63 μ m), solvents were of technical grade. NMR spectra were recorded on *Bruker DPX250*, *AV400* and *AV500* instruments at ambient temperature. The synthesis of the (*R*)-NPE protected 2'-deoxyguanosine phosphoramidite was performed according to literature procedure.^{716,736}

Materials and Methods



Scheme 2: Overview of the synthesis of ^(R)-NPE dG phosphoramidite **7**. a) TBDMS-Cl, imidazole, DMF, rt, quantitative; b) (4-isopropylphenoxy)acetyl chloride, pyridine, 0 °C to rt, 50%; c) (*S*)-1-(2-nitrophenyl)ethanol (prepared according to the literature procedure(34)), PPh₃, DEAD (40 wt% in toluene), THF, 0 °C to rt, 42%; d) TBAF, AcOH, THF, rt, 93%; e) 4,4'-dimethoxytrityl chloride, pyridine, 0 °C to rt, 91%; f) 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, DIPEA, DCM, rt, 68%.

Synthesis of compound 2 (quantitative yield):

5.00 g 2'-deoxyguanosine **1** (18.7 mmol, 1.0 eq), 8.46 g *tert*-butyldimethylsilyl chloride (56.1 mmol, 3.0 eq) and 8.92 g imidazole (130.8 mmol, 7.0 eq) were dissolved in 100 mL DMF and stirred at room temperature. After 1 h, a white solid began to precipitate. The mixture was stirred overnight. The reaction was quenched by the addition of 150 mL EtOH and concentrated under reduced pressure. The residue was dissolved in 150 mL EtOAc and washed with dest. H₂O, 1 M HCl and saturated aqueous NaHCO₃ solution. The organic layer was dried over Na₂SO₄ and the solvent was removed under reduced pressure to give **2** as a white solid. The product was used for the following reaction without further purification.

Synthesis of compound 3 (50% yield):

3.50 g of the TBDMS protected nucleoside **2** (7.1 mmol, 1.0 eq) were dissolved in 25 mL pyridine. The solution was cooled to 0 °C before 1.83 mL (4-isopropylphenoxy)acetyl chloride (10.6 mmol, 1.5 eq) were added dropwise. The orange colored reaction mixture was stirred at 0 °C for 1 h and at room temperature for 4 h. After quenching the reaction by the addition of 100 mL MeOH, the solvent was removed under reduced pressure. The crude product was purified by column chromatography (SiO₂, CH₂Cl₂/MeOH 98:2 → 9:1) to give **3** as a brownish solid.

Synthesis of compound 4 (42% yield):

2.00 g of **3** (3.0 mmol, 1.0 eq), 0.746 g (*S*)-1-(2-nitrophenyl)ethanol (4.5 mmol, 1.5 eq) and 1.17 g PPh₃ (4.5 mmol, 1.5 eq) were dissolved in 35 mL THF. 0.7 mL DEAD solution (40 wt% in toluene, 4.5 mmol, 1.5 eq) were added dropwise.

Materials and Methods

The reaction mixture was stirred at room temperature for 3 d and concentrated under reduced pressure. The residue was diluted with 200 mL CH₂Cl₂ and washed with brine. The organic layer was dried over Na₂SO₄, the solvent removed under reduced pressure and the crude product purified by column chromatography (SiO₂, cyclohexane/EtOAc 95:5 → 3:1). Product **4** was isolated as a pale yellowish foam.

Synthesis of compound 5 (93% yield):

820 mg of **4** (0.99 mmol, 1.0 eq) were dissolved in 15 mL THF and 0.34 mL acetic acid (6.0 mmol, 6.0 eq) and 3.0 mL tetrabutylammonium fluoride solution (1 M in THF, 3.0 mmol, 3.0 eq) were added. The reaction mixture was stirred at room temperature for 16 h. The solvent was removed under reduced pressure and the crude product was purified by column chromatography (SiO₂, CH₂Cl₂ → CH₂Cl₂/MeOH 95:5) to give **5** as a yellowish foam.

Synthesis of compound 6 (91% yield):

550 mg of **5** (0.93 mmol, 1.0 eq) were dissolved in 15 mL pyridine and cooled to 0 °C. 378 mg 4,4'-dimethoxytrityl chloride (1.1 mmol, 1.2 eq) were added slowly. The ice bath was removed and the yellow reaction mixture was stirred at room temperature for 3 h. The solution was concentrated under reduced pressure and the residue diluted with 100 mL CH₂Cl₂. The organic layer was washed with saturated aqueous NaHCO₃ solution and dried over Na₂SO₄. The solvent was removed under reduced pressure. The crude product was purified by column chromatography (SiO₂, CH₂Cl₂/Et₃N 99:1 → CH₂Cl₂/MeOH 95:5) to give **6** as a yellowish foam.

Synthesis of compound 7 (68% yield):

To a solution of 753 mg of **6** (0.84 mmol, 1.0 eq) and 733 µL (*i*Pr)₂NEt (4.2 mmol, 5.0 eq) dissolved in 15 mL CH₂Cl₂ 376 µL 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (1.7 mmol, 2.0 eq) were added. The reaction mixture was stirred at room temperature for 2 h. The solution was diluted with 150 mL CH₂Cl₂ and washed with saturated aqueous NaHCO₃ solution. The organic layer was dried over Na₂SO₄ and the solvent removed under reduced pressure. The crude product was purified by column chromatography (SiO₂, hexane/EtOAc/Et₃N 95:4:1 → hexane/EtOAc 95:5 → 4:1 → 1:1 → 1:2) to give **7** as a yellowish foam.

Oligonucleotide synthesis:

All photolabile modified oligonucleotides were synthesized on an ABI392 DNA/RNA synthesizer from Applied Biosystems. 0.3 M BTT from *emp BIOTECH* was used as activator. All oligonucleotides were synthesized and deprotected under *UltraMILD* conditions (Pac-dA-CE, *i*PrPac-dG-CE, Ac-dC-CE, dT-CE and dT SynBase™ 1000Å CPG purchased from *Glen Research* and *Linktech*) in 1 µmol scales. Pac₂O was used as capping reagent. **short-2345** and **wt-1245 (caged)** were synthesized in DMTr-On while the remaining DNAs were synthesized in DMTr-Off mode. Deprotection was performed according to a protocol from *Glen Research* using 400 µL concentrated ammonium hydroxide for 2 h at room temperature. After evaporation of the solvent at 4 °C, the crude product was purified *via* RP-HPLC using an *Agilent 1200 series* instrument with an *XBridge Peptide BEH C18 OBD Prep Column* (300 Å, 5 µm, 10x250 mm, 4.0 mL min⁻¹, 60 °C, gradient: 5-45.5% MeOH in 400 mM hexafluoro-2-propanol, 16.3 mM NEt₃ buffer pH = 8.2 in 20 min) from *Waters*. The solvent was evaporated at 4 °C using a vacuum concentrator (*SpeedVac™* from *Thermo Fisher*). The DMTr-group was cleaved by incubation of the DNA with 80% acetic acid for 20 min at room temperature. After evaporation at 4 °C, the resulting DNA was again RP-HPLC purified using the same conditions as mentioned before. The solvent was removed using a vacuum concentrator. Finally, all samples were coevaporated several times with ultrapure water and lyophilized.

3.2 CD Spectroscopy

All CD-spectra and CD-melting curves have been recorded on a JASCO spectropolarimeter J-810 at indicated temperatures. Spectra have been smoothed using a Savitzky-Golay filter. Time-resolved CD data were recorded on a Jasco J-810 spectropolarimeter in a 2 mm cuvette at 285 K using 10 μ M DNA in 5 mM K-P_i-buffer (phosphate buffer: 38.5% KH₂PO₄ + 61.5% K₂HPO₄) at pH 7.0. Laser irradiation (355 nm, 4 W) within the cuvette was achieved via glass fiber connection to a laser set up (Paladin Advanced 355-8000) (Figure 31). Kinetics were recorded at 290 nm (1 s data interval) and in the range between 250 – 300 nm for control (6 s per spectrum resolution).

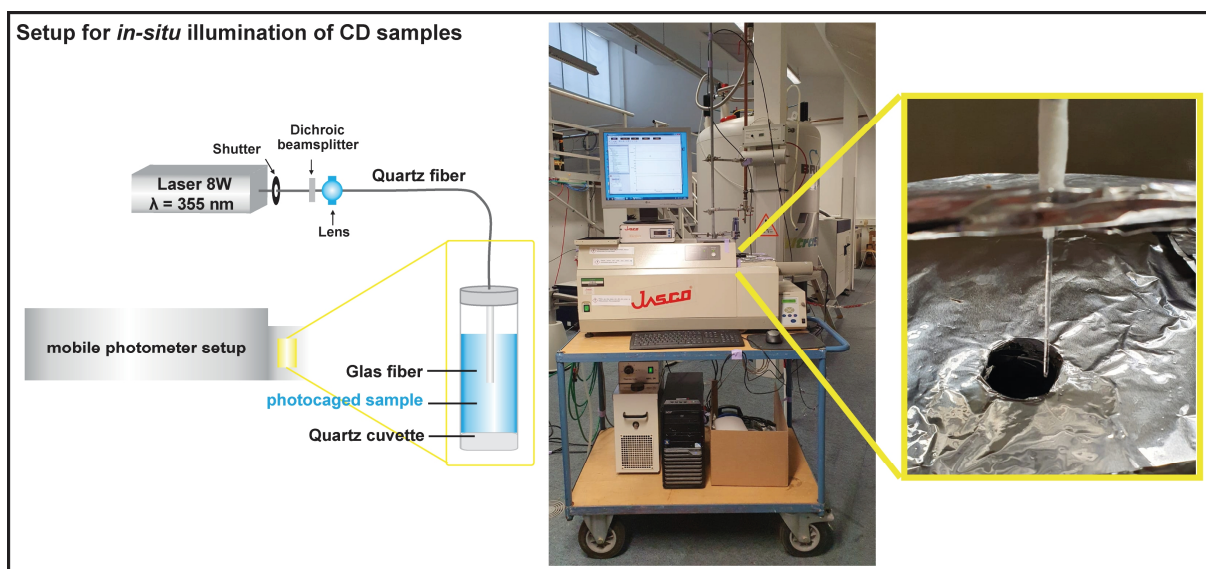


Figure 31: Setup for *in-situ* illumination of CD samples. Original setup design, idea and photos by J. Tassilo Grün.

3.3 Non-Linear Regression and Global Fitting of Kinetic Data

Kinetic traces have been processed in TopSpin 3.5pl7 (Bruker BioSpin) as either single signal intensity (“slice” – “extract column” or as sum-integral (“proj” – “sum”) over a defined region. Traces then have been exported as .ascii-files and normalized for further analysis.

Non-linear regression of kinetic data was achieved with mono- or bi-exponential non-linear regression using the software SigmaPlot 12.5 (Systat Software Inc.).

For complex kinetic fits, numerical solutions for the system of differential equations were obtained from a python script (Appendix, commented script). The script was written by Dominik Brey, as intern under the supervision of me (the author of this thesis). The respective signal intensities of the NMR-experiments were normalized according to the mean equilibrium signal intensity. Global fit of the data was achieved with differential evolution using global optimization according to the respective kinetic model. Error estimation was estimated from a bootstrap method. Models were evaluated based on F-test analysis and residual plots.

3.4 NMR Spectroscopy

All time-resolved NMR spectra were recorded on a Bruker AVIII HD 700 MHz spectrometer equipped with a 5 mm z-axis TXI-HCP cryogenic probe at 298 K using 0.1 mM DNA in 5 mM K-P_i (phosphate buffer: 38.5% KH₂PO₄ + 61.5% K₂HPO₄), 0.05 mM DSS, 10% D₂O at pH 7.0 if not otherwise indicated. Water suppression was achieved using a jump-return-echo pulse scheme for time-resolved data or excitation sculpting for reference 1D spectra. For processing of data, Topspin 3.5pl7 (Bruker Biospin) was used. Time-resolved NMR experiments were performed as a pseudo-2D experiment. Laser irradiation within the NMR tube was triggered via a TLL connection to a laser set up (Paladin Advanced 355–8000) (Figure 32). Kinetics were recorded after 4 s of laser irradiation. Kinetic traces were subsequently averaged over 64 individual experiments. For K⁺-induced folding experiments, 0.1 mM salt-free DNA in Bis-Tris-buffer at pH = 7 was used. K⁺-induced folding was achieved utilizing a rapid-mixing device, as described previously.¹⁴⁴ Kinetic traces were subsequently averaged over 64 (8 for folding) individual scans. For light-induced folding experiments, kinetic traces were not averaged (1 scan per point). NOESY- and HMBC-spectra for assignment have been recorded using a jump-return-echo pulse scheme for water suppression. Spectrometers, Temperatures and sample conditions are indicated in the respective Figures.

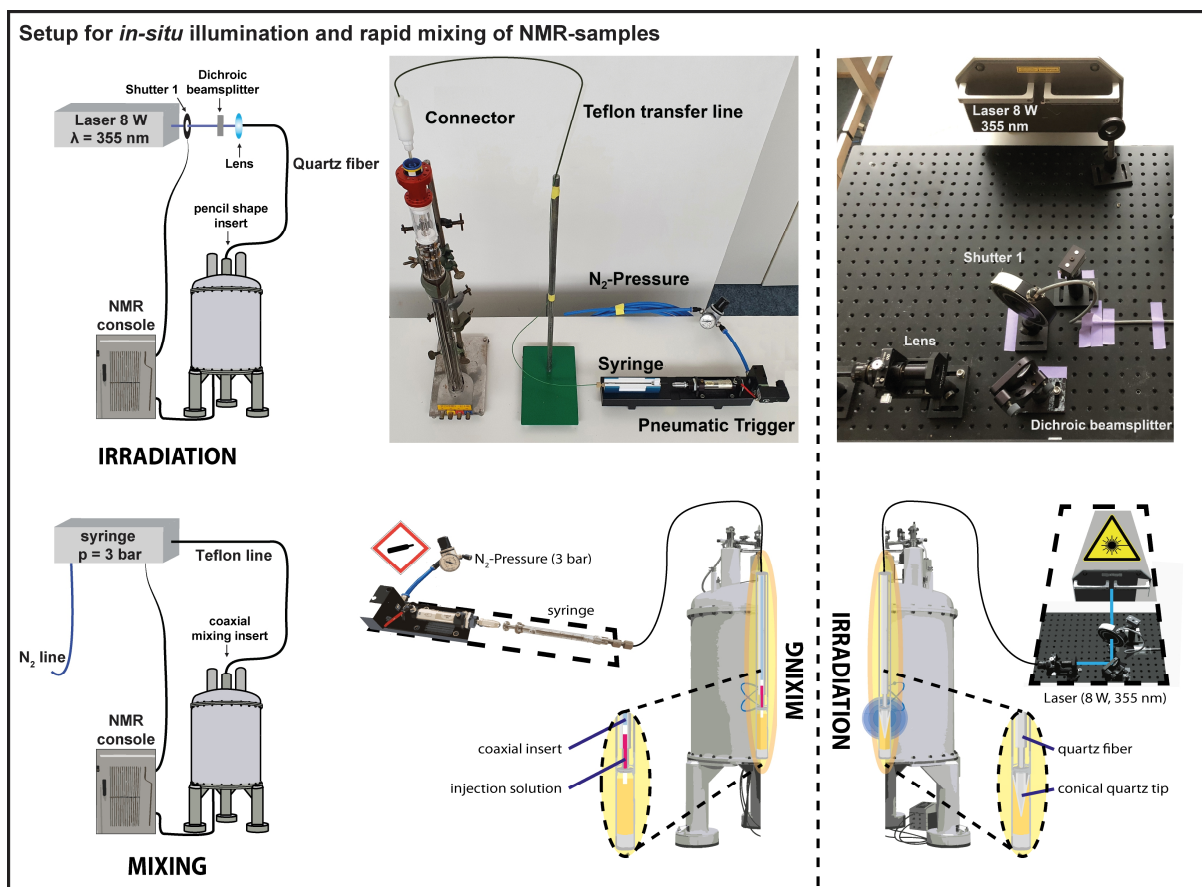


Figure 32: Setup for *in situ* illumination with a laser setup (8 W, 355 nm) connected via a glass fiber to a Shigemitsu tube with a conical quartz tip.⁶⁹⁶ Setup for *in situ* rapid mixing with a mixing device introduced by Hore *et al.*⁶⁸³. Photos from the setups are originally taken from J. Tassilo Grün.

Materials and Methods

4 G-Register Exchange Dynamics

4.1 Overview

This chapter examines the kinetics of folding from unfolded states and refolding from trapped, isolated folded states back to their native equilibrium for the major G4 conformation (2345) found in the *cMYC* promoter (2.4.2). The 18-mer oligonucleotide *cMYC* 2345-X3 (3.1.1) can coexist in two different G-register isomers (Figure 33 (left)); and chapters 2.1.3, 2.4.3). The G-register exchange dynamics for the interconversion of the two parallel *cMYC* G4 conformations have been compared to the interconversion of two conformations from the *hTERT* G4 that co-exists in a hybrid and a parallel conformation. The experimental data focus on the dynamics associated to the refolding between either of the two-state systems.

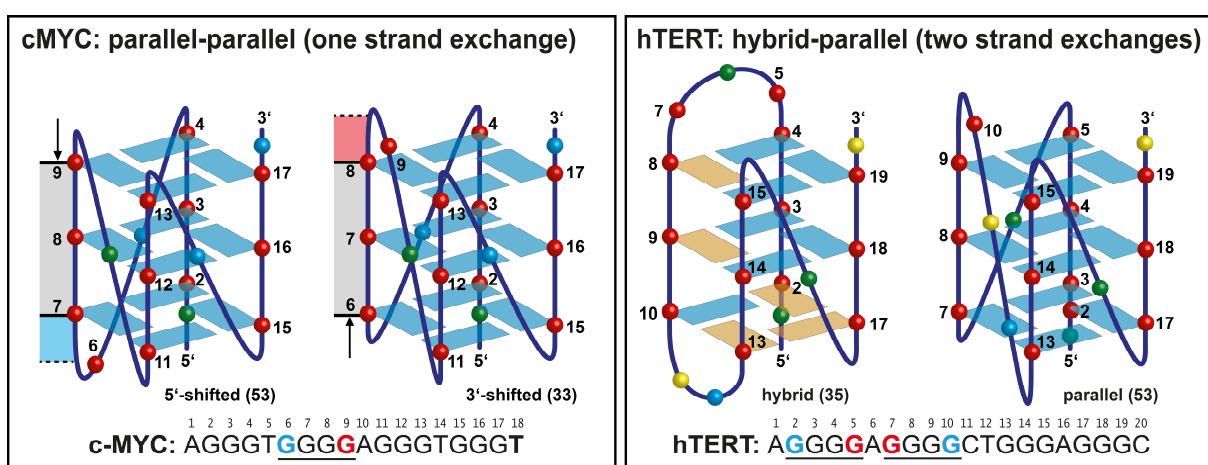


Figure 33: Schematic representation for two different pairs of G-register shift related G4 conformations. **(left)** Two parallel *cMYC* G-register isomers (53, 33) with G-tract III shifted either towards the 5'-terminal tetrad (5) or the 3'-terminal tetrad (3). G-tract V can form two additional isomeric states (3'- and 5'-shifted) in the wildtype sequence, but due to a G-to-T modification on residue 18 the 3'-shifted state is locked. **(right)** Two G-register isomers from an *hTERT* G4 in hybrid (35) or parallel (53) conformation.

The main results presented here have been published in Grün *et al.* 2020.⁷³² Experiments in chapter 4.6.1 have been conducted by Christopher Hennecker and Robert W. Harkness in the group of Anthony K. Mittermaier within a joint collaboration. All shown figures have been created originally by me (the author of this thesis) and are presented modified or unmodified as compared to the published version: Reprinted (adapted) with permission from *J. Am. Chem. Soc.* 2020, 142, 1, 264–273. Copyright 2020 American Chemical Society.

4.2 K⁺-Induced Folding into G-register Isomers

Extensive desalting (3.1) of the native 18-mer *cMYC* 2345-X3 oligonucleotides yields ¹H-NMR spectra that show no imino proton signals in the ppm-range typical for Hoogsteen base pair interactions in K⁺-bound G-quadruplexes (Figure 34). This state has been considered unfolded after the careful analysis of the NMR-spectra, even though the formation of pre-folded states cannot completely be precluded. After addition of K⁺ to a final concentration of 5 mM, the native folded ensemble can be recaptured. In thermodynamic equilibrium the folded state of *cMYC* 2345-X3 is consisting of two coexisting G-quadruplex conformations, as will be shown and discussed further in this chapter.

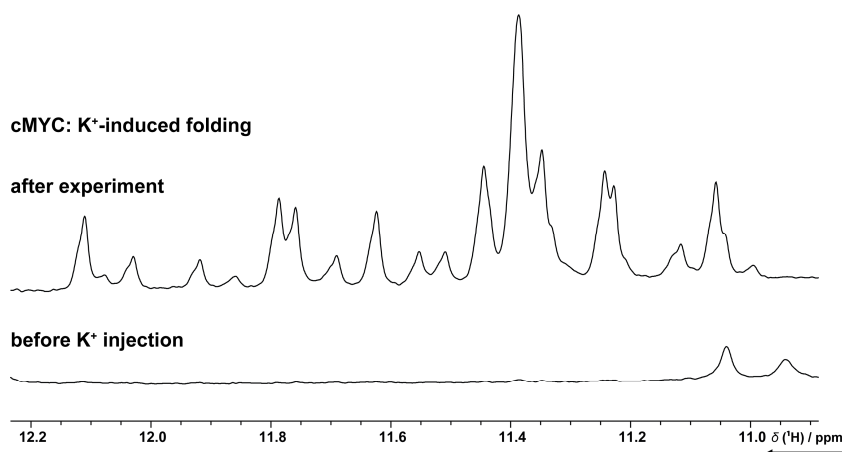


Figure 34: ¹H-NMR spectra of *cMYC* 2345-X3 in K⁺-free (**lower**) Bis-Tris buffered solution and several hours after addition of 5 mM KCl (**upper**). Comparably broad signals in the downfield (>12.1 ppm) and upfield (<11.1 ppm) shifted region indicate some pre-associated state that are unrelated to the K⁺-bound state. (0.1 mM DNA, 5 mM Bis-Tris buffer (pH=7.0), 90/10:H₂O/D₂O, 700 MHz, 298 K, jr-echo water suppression)

The folding reaction was initiated by a rapid injection of K⁺ using the rapid-mixing device outlined in section 2.5.3 and 3.4. Time correlated single scan 1D spectra were recorded as pseudo-2D dataset. Kinetic traces for the folding reaction were then extracted by plotting the respective signal intensities as function of time. The spectral resolution (~1.16 s) of the imino proton signals allows tracking of the kinetics of individual signals that represent either of the two folded conformations (Figure 35).

The folding kinetics exhibit a clear biphasic behaviour: (i) rapid folding into the two individual states and (ii) slow reequilibration to thermodynamic distribution. The kinetics also reveal that the minor conformation is slightly kinetically favoured and overshoots in the first folding phase, compared to thermodynamic equilibrium. The kinetics were analysed both with bi-exponential fitting and with global fitting to a kinetic model (Table 4).

The easiest model to describe the folding mechanism is a direct refolding from the minor to the major state (Figure 35). The global fitting procedure (3.3) yields numerical solutions for the individual rate constants.

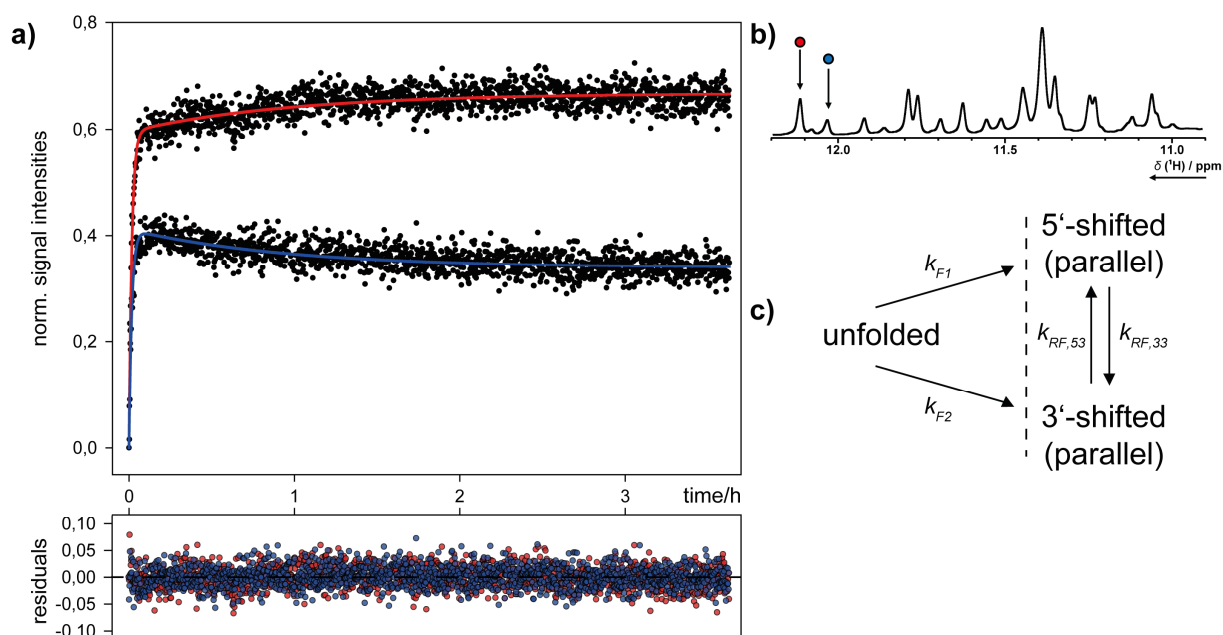


Figure 35: Folding transition of *cMYC* 2345-X3 G-register isomers. (a) Kinetic traces of individual imino proton signals (b), representative for the two folded conformations, shown in red (*cMYC*-2345-33) and blue (*cMYC*-2345-53). Fitting curves and residuals result from global fitting to the depicted kinetic model (c).

The residual plot shown in Figure 35 a results from fitting to the depicted kinetic model (Figure 35 c). Errors for bi-exponential fitting (k_{obs}) are within 5%, but should be considered as estimates.

Table 4: Kinetic rate constants for the overall (re-)folding transitions (k_{obs}) derived from exponential fitting using non-linear regression and results from global fitting to the kinetic model depicted in Figure 35. For non-linear regression, the calculated standard error is within 5% as well as the error estimate from fitting residuals.

	$k_{\text{obs}} [h^{-1}]$		Conformation	folding rate [h^{-1}]	refolding rate [h^{-1}]
	Folding	Refolding			
<i>cMYC</i>	59	1.0	2345-53	$k_{F,53} = 35$	$k_{RF,53} = 0.35$
			2345-33	$k_{F,33} = 25$	$k_{RF,33} = 0.69$

A Bis-Tris buffer was used instead of a phosphate buffer to obtain the K^+ -free state, which is different from NMR spectra of most other shown samples. The spectra indicate no structural changes in different buffers. Further, no influence of phosphate vs. Bis-Tris buffer on folding kinetics was observed in control experiments.

4.3 Light-Induced Refolding of Caged G-Register Isomers

4.3.1 Conformational Selection of G-Register Isomers

The second, slow phase of the folding reaction indicates that refolding occurs, which gives a dynamic link between the two folded conformations. The expected timescale for this rearrangement is too slow for NMR methods in equilibrium to investigate, if the discussed direct refolding mechanism is correct. The folding trajectory however, could also be affected by underlying *on*- or *off*-pathway intermediates or pre- and misfolded structures that cause the respective kinetic behaviour. In this regard, the corresponding population change would less be a refolding reaction from stable folded structures, but more a feature of perturbed, retarded folding from these somewhat partially folded states. The refolding dynamics therefore should be separated from initial folding and investigated independently.

The *cMYC* 2345-X3 oligonucleotide was synthesized in a way to carry a (*R*)-NPE photolabile protecting group either at position G6 to yield *cMYC* 2345-53 or at position G9 to yield *cMYC* 2345-33 (Figure 36). The ¹H-NMR spectra show that this approach is suitable to select either one of the two conformations, as the native spectrum directly decomposes into the spectra of the individual, trapped single conformations.

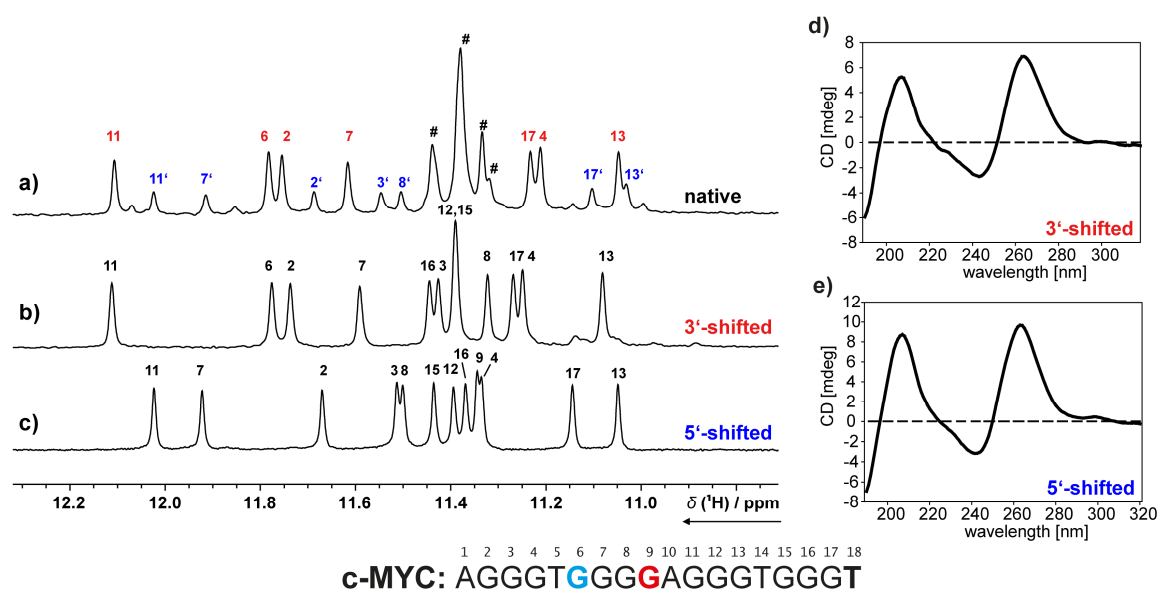


Figure 36: (a-c) 1D-¹H-NMR spectra of **a**) native *cMYC* 2345-X3, **b**) trapped (caged) 3'-shifted conformation (red) and **c**) caged 5'-shifted conformation (blue). (0.1 mM DNA, 5 mM K-P_i buffer (pH=7.0), 90/10:H₂O/D₂O, 700 MHz, 298 K, jr-echo water suppression) **(d-e)** CD-spectra of caged 3'-shifted **(d)** and 5'-shifted **(e)** conformations. The spectra show characteristic signatures for *all*-parallel G4 conformations. (10 μM DNA, 5 mM K-P_i buffer (pH=7.0), 298 K)

CD spectra of the two conformations show that the expected parallel conformation is maintained (Figure 36 d, 3'-shifted 33 and e, 5'-shifted 53). Spectral assignment (Figure 37, Table 5) of the respective conformations was achieved via NOESY (Figure 38, Figure 39) interactions, as well as long-range HMBC correlations and H-D-exchange experiments (Appendix 6.1, Figure 66). Taken together the expected overall *all*-parallel folding topology can be certainly supposed.

G-Register Exchange Dynamics

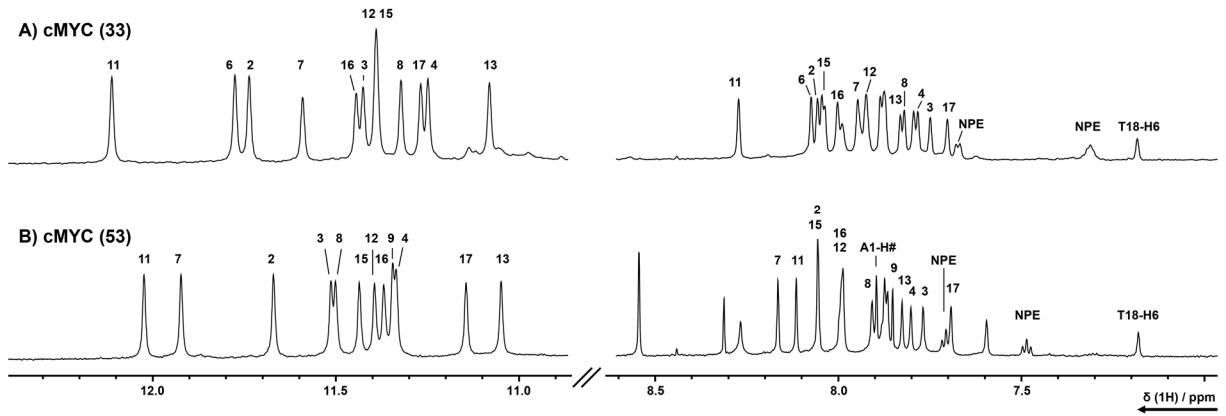


Figure 37: Assigned 1D ^1H -NMR spectra showing the imino and aromatic proton regions of photocaged **A)** *cMYC*-18 (**33**) and **B)** *cMYC*-18 (**53**) G-quadruplex, (700 MHz, 298 K). Referenced to DSS.

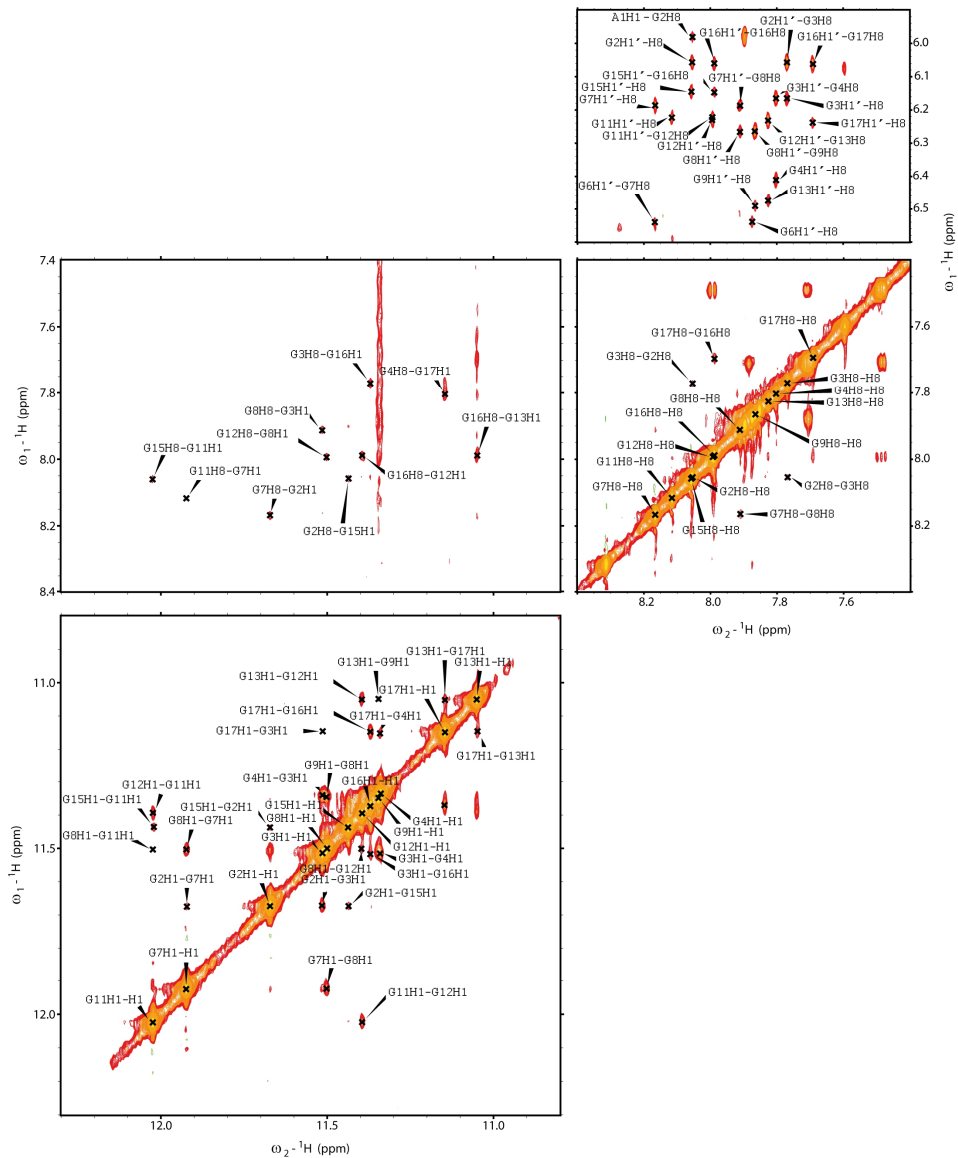


Figure 38: Assigned 2D-NOESY-spectra with characteristic H1-H1, H1-H8, H8-H8 and H1'-H8 regions of the photocaged 5'-shifted *cMYC*-2345-53 G-quadruplex. Recorded with 8192x1024 points (800 MHz, 298 K) using a pulse-sequence with jump-return-echo water-suppression. (5 mM K-P_i buffer (pH=7.0), 298 K)

G-Register Exchange Dynamics

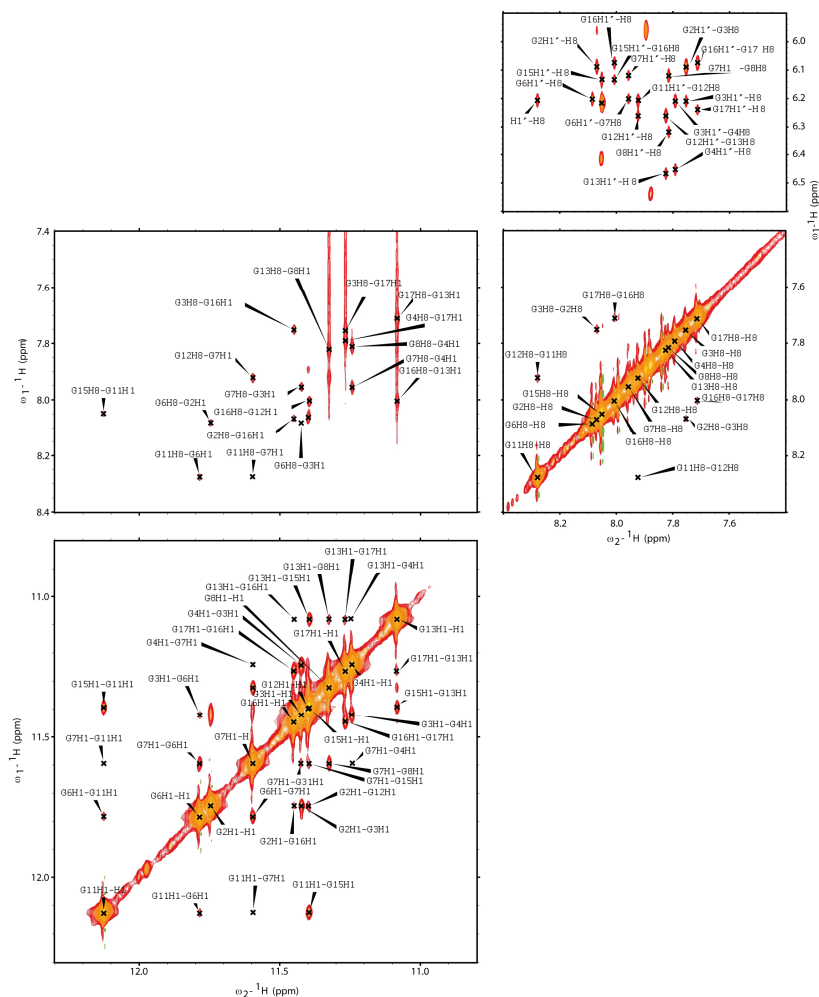


Figure 39: Assigned 2D-NOESY-spectra with characteristic H1-H1, H1-H8, H8-H8 and H1'-H8 regions of the photocaged 3'-shifted *cMYC*-2345-33 G-quadruplex. Recorded with 8192x1024 points (600 MHz, 298 K) using a pulse-sequence with jump-return-echo water-suppression. (5 mM K-P_i buffer (pH=7.0), 298 K)

Table 5: Chemical shifts of assigned H1 imino and H8 aromatic proton signals of photocaged *cMYC* conformations at 298 K in 5 mM K⁺ (pH = 7.0). Referenced to DSS.

Residue	<i>cMYC</i> (53) [ppm]		<i>cMYC</i> (33) [ppm]	
	H1	H8	H1	H8
G2	11.67	8.05	11.75	8.07
G3	11.52	7.77	11.42	7.75
G4	11.34	7.80	11.24	7.79
G6	-	-	11.78	8.09
G7	11.92	8.17	11.60	7.96
G8	11.50	7.91	11.32	7.82
G9	11.35	7.87	-	-
G11	12.02	8.12	12.13	8.28
G12	11.40	8.00	11.40	7.92
G13	11.05	7.83	11.08	7.83
G15	11.44	8.06	11.40	8.05
G16	11.37	7.99	11.45	8.00
G17	11.15	7.69	11.27	7.71

4.3.2 Refolding into a Two-State G-Register Ensemble

The refolding reaction was initiated by photolysis using the laser setup outlined in section 3.4. Time correlated single scan 1D spectra were recorded as pseudo-2D dataset and afterwards summed up to yield 64 scans per point in time. Kinetic traces for the refolding reaction were then extracted by plotting the respective signal intensities as function of time (Figure 40).

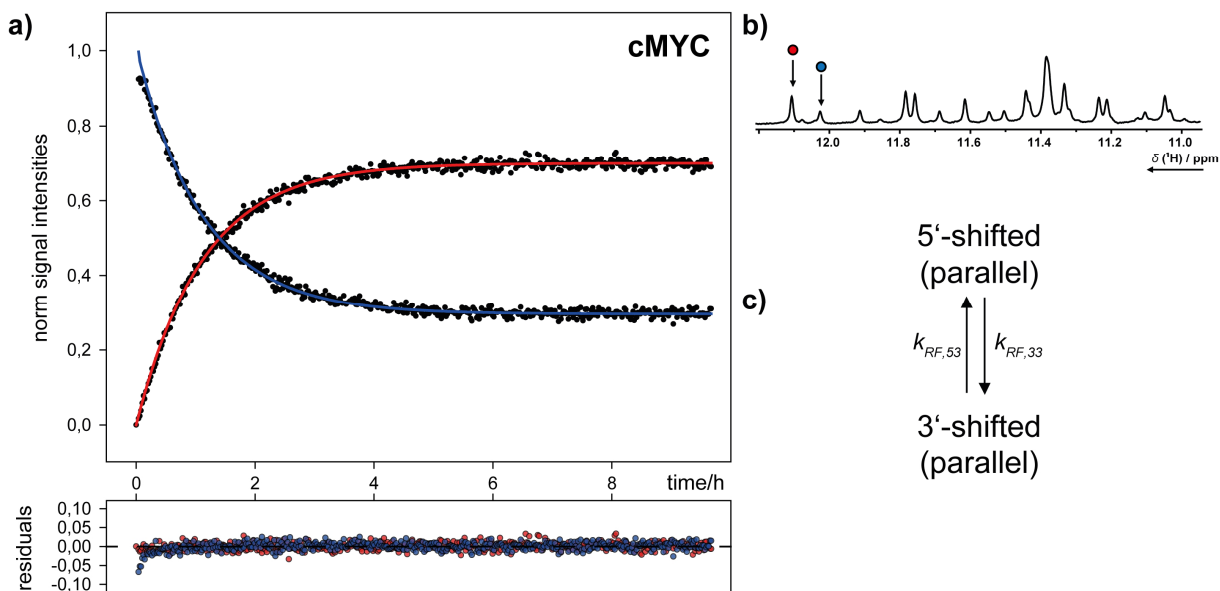


Figure 40: Refolding transition of *cMYC* 2345-X3 G-register isomers. (a) Kinetic traces of individual imino proton signals (b), representative for the two folded conformations, shown in red (*cMYC*-2345-33) and blue (*cMYC*-2345-53). Fitting curves and residuals result from global fitting to the depicted kinetic model (c).

The kinetic data reveal that the refolding process is slow, taking place for several hours at room temperature. The signals for the initially 100% populated single conformation decay, while signals for the second conformation build-up. These kinetics have been fitted globally according to a two-state kinetic model (Figure 40 c), yielding the rate constants $k_{RF,53}$ and $k_{RF,33}$ shown in Table 6. The observable rate constant k_{obs} for the refolding process was derived from mono-exponential fitting. The rates are in good agreement if $k_{obs} = k_{RF,53} + k_{RF,33}$ is assumed ($0.85 \text{ h}^{-1} \approx 0.9 \text{ h}^{-1}$).

Table 6: Kinetic rate constants for the overall refolding transition (k_{obs}) derived from exponential fitting using non-linear regression and results from global fitting to the kinetic model depicted in Figure 40. For non-linear regression, the calculated standard error is within 5% as well as the error estimate from fitting residuals.

	$k_{obs} [\text{h}^{-1}]$	refolding rate $[\text{h}^{-1}]$
<i>cMYC</i>	Refolding	$k_{RF,53} = 0.27$
	0.85	$k_{RF,33} = 0.63$

The two-state model is the simplest model that can be applied to this system of two conformations. Neither long-lived intermediate conformations can be detected, nor does the correlation between both traces indicate an accumulation of unfolded species.

The total signal intensity remains constant during the progression of the experiment. However, the overall kinetics are slow. The existence of any additional short-lived intermediate states (potentially unfolded states) that indicate more complex refolding cannot be postulated or precluded based on the experiments.

4.3.3 Refolding into a Full Set G-Register Ensemble

The refolding reported in 4.3.2, encompasses the two majorly populated *cMYC* G-register isomers 2345-53 and 2345-33. To investigate the kinetic effect of the full conformational space of the 18-mer *cMYC*-2345 wt-sequence, a trapped conformation was prepared using a second photocage at the 3'-terminal G-residue (G18). This position is replaced with a G-to-T mutation in all other presented experiments. Potentially, the incorporation of G18 into tetrad formation can yield two additional G-register isomers, resulting from a strand shift in the 3-terminal G-tract (namely the 35 and 55 conformation).

The incorporation of the second photocage was successful to trap the 53-conformation in a single folded state without further distortions to the structure (Figure 41 b). The refolding kinetics have been monitored after photolysis and analysed with mono-exponential fitting. The observable rate constant for refolding was obtained as a function of the build-up of the *cMYC*-33 conformation (Figure 41 c). The refolding (0.54 h^{-1}) is $\sim 36\%$ slower compared to the two state refolding process between the 53 and 33 conformations (0.85 h^{-1}). This shows that the refolding process in the *cMYC* G4 forming sequence that is able to adopt all four possible G-register isomers is clearly more complex. The NMR spectrum after the refolding process shows additional imino proton signals that indicate a polymorphic ensemble with more than two populated conformations (Figure 41 a).

The enlarged conformational space during refolding enables additional interactions in the transitory ensemble leading to kinetic traps that decelerate the process. These additional conformational states can only occur if the 3'-terminal G-tract (G-tract V) unfolds from the G4-core.

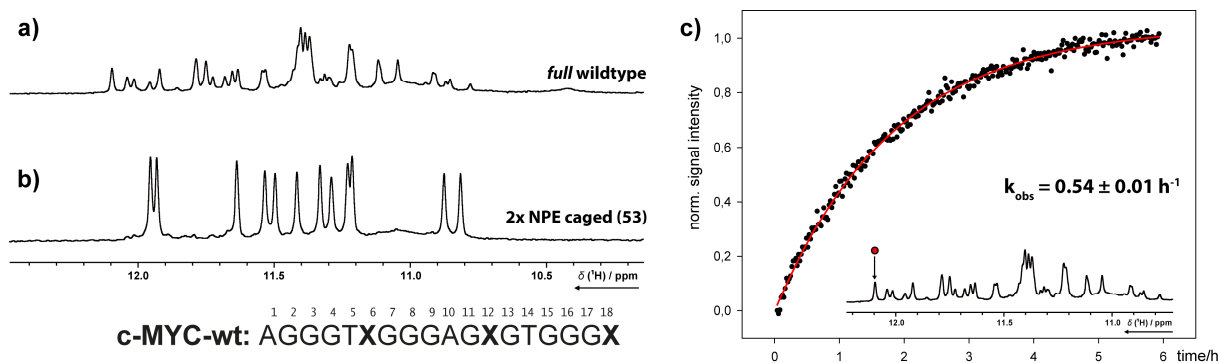


Figure 41: ^1H NMR spectrum of **a)** the 18-mer *cMYC* wt-sequence with all four possible G-register isomers in a native oligonucleotide; and **b)** with two (*R*)-NPE cages at position G6 and G18 (marked with X in the sequence, conformation-53). **c)** Kinetic trace for light-induced refolding and mono-exponential fitting (red) to yield k_{obs} .

4.4 G-Register Polymorphism in the *hTERT* Promoter G-Quadruplex

The human TElomerase Reverse-Transcriptase (*hTERT*) gene promoter region possesses a G4 forming element that has multiple G-rich strands, able to fold into a manifold of different G4 conformations.^{102,103} Two co-existing conformations have been published for a 20 nt long sequence that can be found 60 nt upstream of the TSS.

These conformations feature G-register isomerism as they incorporate different G-residues into the G4 tetrads. However, the reported conformations are markedly different, since they adopt either an *all-parallel* or a *hybrid* conformation. To investigate the G-register exchange dynamics in comparison to the reported kinetics of the *cMYC* G-register isomers, the same photocaging strategy was applied. For each oligonucleotide, two photocages have been incorporated to isolate a specific conformation, because the *hTERT* conformations distinguish for G-register shifts in G-tract I and II. The photocaging strategy was successful to trap and isolate the parallel and the hybrid conformations from the wt-sequence (Figure 42). Some minor additional imino proton signals are visible that vanish immediately after uncaging. This set of signals might arise from different stacked photocage orientations.

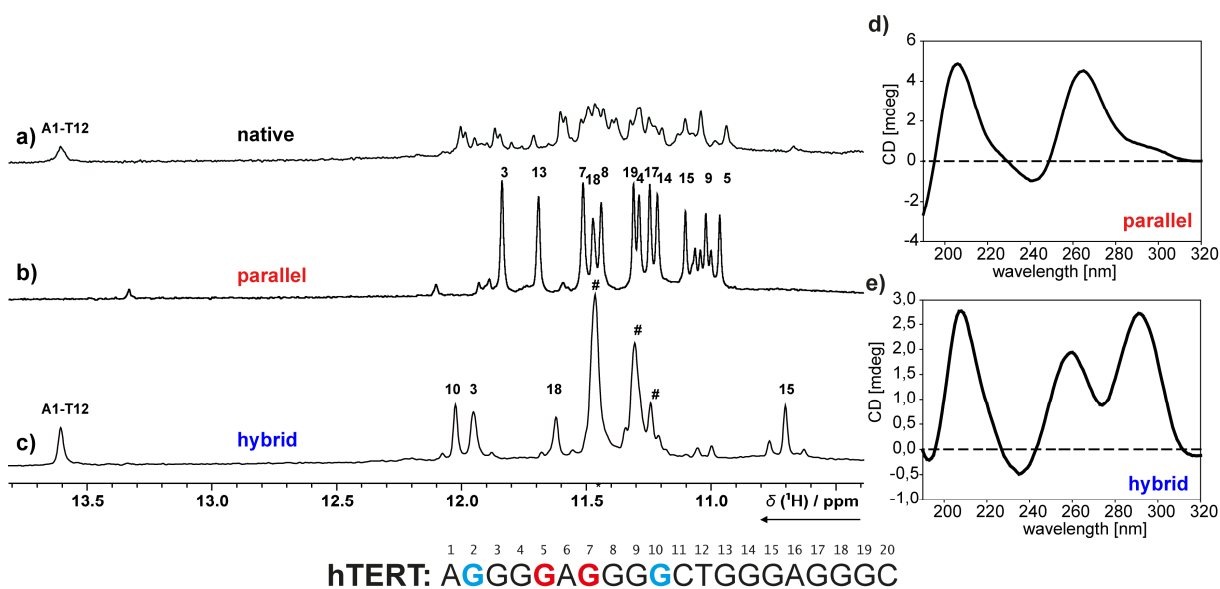


Figure 42: (a-c) 1D-¹H-NMR spectra of a) native *hTERT*, b) trapped (caged) parallel conformation (red) and c) caged hybrid conformation (blue). (0.1 mM DNA, 5 mM K-P_i buffer (pH=7.0), 90/10:H₂O/D₂O, 700 MHz, 298 K, jr-echo water suppression) (d-e) CD-spectra of caged parallel (d) and hybrid (e) conformations. The spectra show characteristic signatures for *all-parallel* and hybrid G4 conformations. Assignment taken from Lim *et al.*¹⁰², #-signals have not been assigned due to spectral overlap. (10 μM DNA, 5 mM K-P_i buffer (pH=7.0), 298 K)

The refolding kinetics have been measured and analyzed as described in 4.3.2. Figure 43 shows representative kinetic traces, starting from a caged parallel conformation. Fitted in blue is the decay of the parallel conformation, fitted in red the build-up of the hybrid conformation. The downfield-shifted signal (~13.6 ppm) can be assigned to an A1-T12 Watson-Crick base pair, which serves as a specific probe for the hybrid conformation.

G-Register Exchange Dynamics

The refolding reaction is slow with an observable rate constant (Table 7) that is comparable to the refolding of the *cMYC* G-register isomers.

This observation was unexpected; as the initial hypothesis was that the refolding should be slower, due to a drastically greater degree of unfolding needed for interconversion of a hybrid to a parallel conformation. The postulation of an unfolded ensemble in the kinetic model (Figure 43 c) is based both on analysis of fitting statistics and rational considerations, because refolding between the two states necessarily requires unfolding to a significant extent. The residual plot shows ~5% systematic deviations for the fitting in the initial 2 h of the experiment, which gives indications for a more complex underlying mechanism. The kinetic model is however adequate to describe the overall refolding progression.

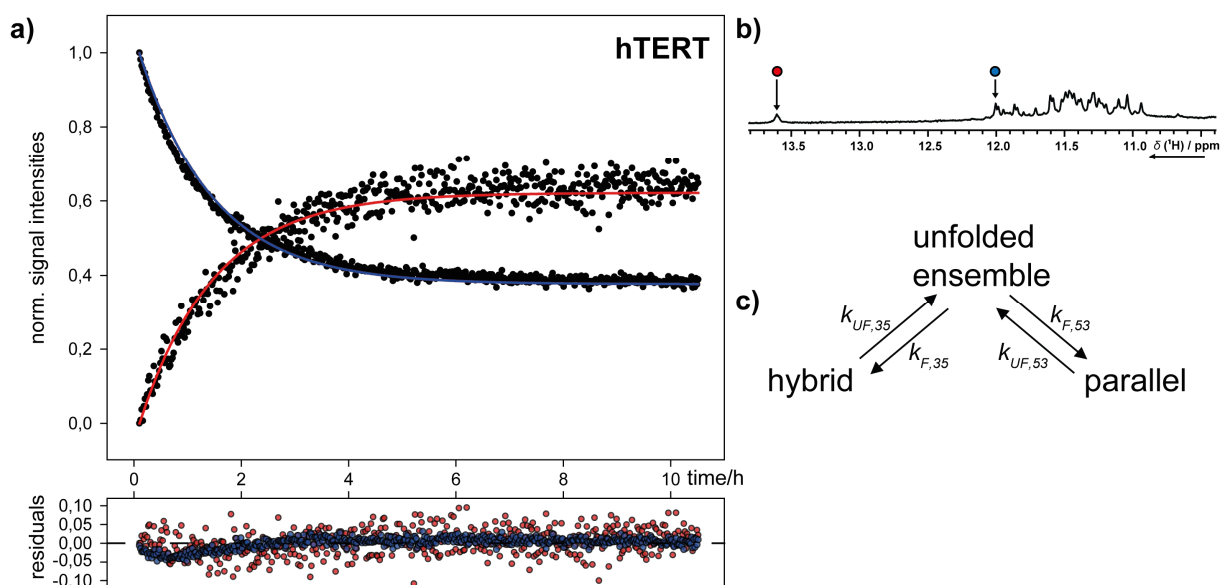


Figure 43: Refolding transition of *hTERT* G-register isomers. (a) Kinetic traces of individual imino proton signals (b), representative for the two folded conformations, shown in red (*hTERT* hybrid) and blue (*hTERT* parallel). Fitting curves and residuals result from global fitting to the depicted kinetic model (c).

Table 7: Kinetic rate constants for the overall refolding transition (k_{obs}) derived from exponential fitting using non-linear regression and results from global fitting to the kinetic model depicted in Figure 43. For non-linear regression, the calculated standard error is within 5% as well as the error estimate from fitting residuals.

	$k_{\text{obs}} [h^{-1}]$	refolding rate [h^{-1}]
<i>hTERT</i>	0.67	$k_{UF,35(\text{hybrid})} = 1.2$
		$k_{F,35(\text{hybrid})} = 491$
		$k_{UF,53(\text{parallel})} = 0.43$
		$k_{F,53(\text{parallel})} = 277$

4.5 Activation Energies of G-Register Exchange Transitions

The energy profiles of the G-register exchange transitions help to get mechanistic insight into the refolding processes of the inherently different G-register isomers in the *cMYC* (parallel to parallel) and in the *hTERT* G4s (parallel to hybrid).

The refolding kinetics have been measured in the range between 288 and 308 K. Arrhenius analysis (Figure 44) of the temperature-dependent rate constants yields an apparent activation energy for refolding (Table 8). The apparent activation energies for *cMYC* do not significantly differ, if the experiment is started from the caged *cMYC*-18 2345-53 or *cMYC*-18 2345-33 conformation. For *hTERT*, the apparent activation energies are larger compared to *cMYC*, and do differ for the experiments that start either in the hybrid or parallel conformation.

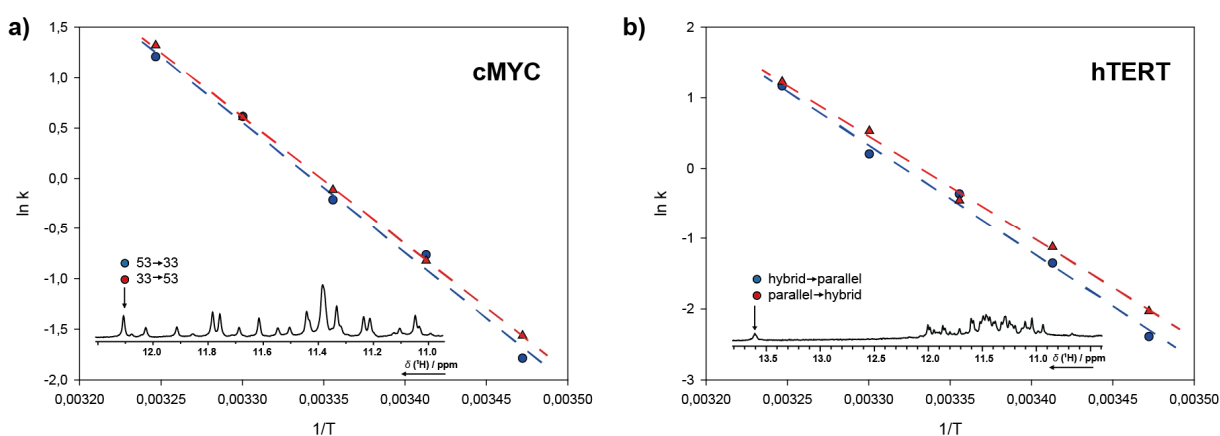


Figure 44: Arrhenius plots of activation energies for refolding of *cMYC* (a) and *hTERT* (b) G-register isomers. Apparent activation energies have been estimated using linear regression. Data points shown in red and blue represent individual experiments, starting from either of the two caged conformation at each temperature.

Table 8: Apparent activation energies for the refolding reactions between 5'-shifted vs. 3'-shifted parallel conformations (*cMYC*) and parallel vs. hybrid conformations (*hTERT*). Refolding for *cMYC* is irrespective of the experimental starting point (caged 53 or caged 33 conformation).

	<i>cMYC</i>	<i>hTERT</i>
apparent activation energies [kJ·mol ⁻¹]	107	hybrid: 128 parallel: 120

Melting temperatures (T_m) have been determined based on CD melting curves for all of the caged conformations (Figure 45). Melting temperatures are comparable to literature values (*cMYC*^{100,101} and *hTERT*¹⁰²) and reflect the relative population of the coexisting conformations. This is not a trivial finding, as it indicates that for the selected caged conformations no stabilizing or destabilizing effect from the photocages can be observed. The melting curve of the *hTERT* hybrid conformation is significantly broader than the melting curves from the other G4 conformations and has is not perfectly sigmoidal. This gives indications for a multiphasic, more complex unfolding behaviour of the hybrid

conformation. The given T_m is a superposition of those individual steps, which suggests that unfolding steps with higher energetic barriers exist.

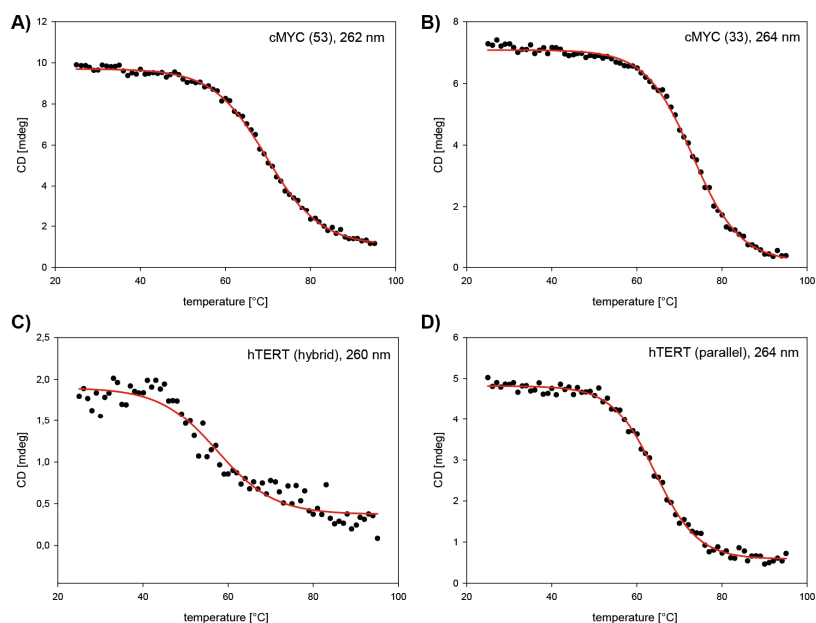


Figure 45: CD melting curves of all investigated caged G4 conformations in the range between 25-95 °C. T_m s have been calculated using sigmoidal fitting. The melting curve for *hTERT* hybrid (c) indicates multiphasic melting transition.

Table 9: Melting temperatures T_m of the caged G4 conformations from *cMYC* and *hTERT*, derived from sigmoidal fitting of CD melting curves.

DNA (caged conformation)	T_m [°C]
<i>cMYC</i> 2345-53	69.7 ± 0.2
<i>cMYC</i> 2345-33	73.2 ± 0.2
<i>hTERT</i> (hybrid)	57.4 ± 0.9
<i>hTERT</i> (parallel)	64.2 ± 0.2

The apparent activation energies do not allow evaluating directly the degree of unfolding during the refolding process. The calculated apparent activation energies are derived from observable rate constants (exponential fitting) and these relaxation rates cannot be directly assigned to a single rate constant in the proposed kinetic models. They further reflect the energy barriers of the rate-limiting steps during refolding, which is not directly linked to the overall thermal stability of the respective conformations. However, if the thermal stabilities for all conformations would be equal and the same refolding mechanism would be assumed (progressing through the same transitory ensemble), and then comparable activation energies must be expected. Here, opposing trends for these parameters have been observed: the ratio for the mean thermal stabilities for the *cMYC* G-register isomers compared to the *hTERT* G-register isomers is $\approx 1.0:0.84$, which means that *hTERT* has a 16% decreased thermal stability compared to *cMYC*. The ratio for the apparent activation energies for refolding is $\approx 1.0:1.2$, which means that *hTERT* requires a 20% higher activation energy for refolding (20% slower observable rate constant in refolding).

4.6 Folding Initiation in Pre-Equilibrated K⁺-Solution

4.6.1 Dynamics of Thermally-Induced (Un)-Folding of G-Register Isomers

All of the above-discussed investigations of G-register isomer dynamics have been conducted with oligonucleotides constituted of the wt-sequences. These oligonucleotides have been chemically modified to *transiently* disturb the native constitution and hence the native distribution of conformation populations. However, the nucleotide sequences of the oligonucleotides can also be modified with single nucleotide mutations. In this way, all relevant G-register isomers of the *cMYC* and *hTERT* G4 can also be isolated and separated *permanently* in four different oligonucleotides. These stable oligonucleotides also yield single conformations, but cannot refold from one conformation to another. Analysis from thermal melting profiles can yield kinetic information for unfolding and folding of distinct conformations. Rapid cooling or heating causes different cooling and heating profiles in denaturation experiments with shifted apparent melting temperatures. The observation of thermal hysteresis for the folding and unfolding of the G4 conformations provides quantitative information on the underlying rate constants at which thermal folding and unfolding occurs. These experiments based on thermal hysteresis (TH) and their kinetic analysis have been presented by A. K. Mittermaier and co-workers.^{101,142,737} The herein presented results from these experiments and their quantitative analysis have been done by Robert W. Harkness and Christopher Hennecker in the group of A. K. Mittermaier (McGill University, Canada). They are co-authors of the publication Grün *et al.* 2020.⁷³² Experimental details can be found there, the findings are briefly summarized in this chapter for a comprehensive and conclusive discussion.

Melting profiles have been monitored with UV-vis absorbance at 295 nm, with scan rates of 1, 2, 3 and 4 °C·min⁻¹. Based on the analysis of the thermal hysteresis (TH) experiments the refolding process was simulated according to the kinetic models depicted in Figure 46. These simulations yield apparent activation energies and rate constants shown in Table 10. Note, that the kinetic models must include unfolded states, because the experiments report on (complete) thermal unfolding (and folding). This can be in thermodynamic aspects fundamentally different from the isothermal refolding reported from the light-induced NMR experiments.

Table 10: Apparent activation energies and rate constants extracted from simulations for refolding between the G-register isomers in *cMYC* and *hTERT*. The simulations are based on rate constants for folding and unfolding that have been extracted from thermal hysteresis experiments. These experiments cannot account for direct conversion/refolding, but measure two-state transitions between folded and unfolded states.

	<i>conformation</i>	<i>app. activation energies</i> $E_{relax} [kJ \cdot mol^{-1}]$	<i>app. rate constants</i> $k_{relax} [h^{-1}]$
<i>cMYC</i>	5'-shifted (53)	129 ± 2	0.15 ± 0.01
	3'-shifted (33)	129 ± 2	0.15 ± 0.01
<i>hTERT</i>	parallel (53)	144 ± 4	0.93 ± 0.03
	hybrid (35)	146 ± 1	0.33 ± 0.01

G-Register Exchange Dynamics

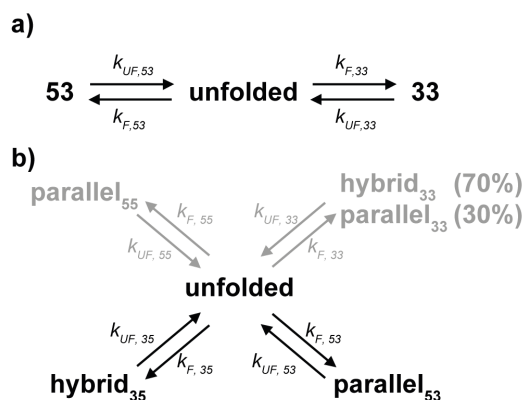


Figure 46: Kinetic models used for the simulation of refolding kinetics between the G-register isomers in *cMYC* (a) and *hTERT* (b). Shown in grey are additional states that have been proposed for *hTERT*, but are not significantly populated in equilibrium (Figure 42 and Phan *et al.*¹⁰²). Analysis and experimental data for these conformations are discussed in detail in Grün *et al.*⁷³²

For *hTERT*, the mean of the predicted apparent rate constants calculated for a refolding according to the kinetic model depicted in Figure 46 b, is in good agreement with the NMR derived rates (NMR: 0.67 h⁻¹; TH: 0.63 h⁻¹). For *cMYC* in contrast, the predicted apparent rates derived from TH according to the kinetic model depicted in Figure 46 a) are slower by a factor of ~6. The apparent activation energies calculated from TH are ~13% larger for *hTERT* and ~20% larger for *cMYC* compared to the NMR derived activation energies for isothermal refolding. Given the very different experimental approaches, the agreement for *hTERT* is still remarkable, even though it is significantly different. For *cMYC*, the apparent activation energies are somehow not comparable anymore for the two experimental results and their underlying models. Both the comparison of kinetic rates and the apparent activation energies are overall in good agreement for *hTERT*. This implies that the model, which has been proposed for the analysis in both kind of experiments, is suitable to describe the observed refolding process. For *cMYC*, the drastic differences in TH and isothermal NMR experiments imply that no complete unfolding occurs for the G-register exchange process.

The kinetic and energetic parameters from TH experiments have however not been determined directly, because in these experiments the two conformations are separated by mutations. The primary results from the TH experiments is folding (and unfolding, not shown) rates for the single conformations, given in Table 11. Most remarkably, the folding rates for both *cMYC* G-register isomers is approx. 10³-times faster as the K⁺-induced folding rates from the NMR experiments (~35 h⁻¹ for 53 and ~25 h⁻¹ for 33).

Table 11: Apparent activation energies and rate constants extracted from TH experiments for the thermal folding of G-register isomers in *cMYC* and *hTERT*.

	conformation	app. activation energies E_F [kJ·mol ⁻¹]	app. rate constants k_F [h ⁻¹]
<i>cMYC</i>	5'-shifted (53)	-75 ± 4	3800 ± 700
	3'-shifted (33)	-73 ± 3	6000 ± 1000

4.6.2 Light-Induced Folding into G-register Isomers

To scrutinize the drastic differences in the folding rates for thermally induced folding in the TH experiments and K^+ -induced folding in the NMR experiments, the photocaging strategy was expanded. An unfolded state in the presence of K^+ has been prepared, by introducing a total of three (*R*)-NPE photocages to the *cMYC*-2345-wt 18-mer. This strategy was successful to prevent completely the folding of the G4 forming oligonucleotide in the presence of K^+ and at ambient temperatures (Figure 47 b). The characteristic region for imino proton signals in slow exchange that shows Hoogsteen base pair interactions does not indicate any kind of pre-folded structures. The photolysis reaction is complete within 4 s of laser irradiation and the folded state can be recaptured after isothermal folding. The folding reaction here is accelerated by a factor of $\sim 10^3$ (mono-exponential fitting yields an approximation of $\sim 6300 \text{ h}^{-1}$), compared to K^+ -induced folding. Folding of the isolated 53-isomer is monophasic in good approximation. This folding relaxation time is in the range of the observed kinetics for thermal folding in the TH experiments (4.6.1).

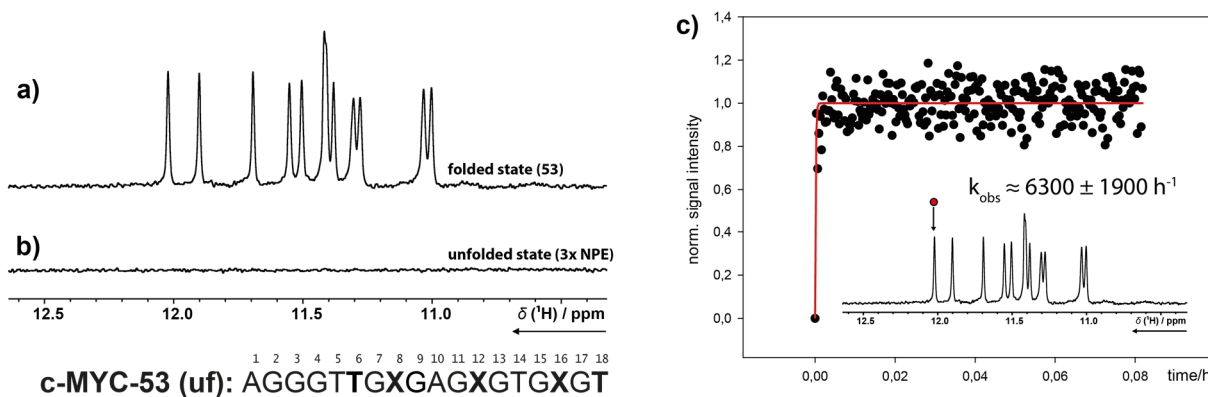


Figure 47: ^1H NMR spectrum of **a**) the 18-mer *cMYC* 53-sequence in a native oligonucleotide; and **b**) with three (*R*)-NPE cages at position G8, G12 and G16 (marked with X in the sequence, conformation-53). **c**) Kinetic trace for light-induced folding and mono-exponential fitting (red) to yield k_{obs} .

Two major effects can account for the differences in the observed kinetics. (i) While the initial folding experiment was induced by addition of K^+ , the folding experiment in Figure 47 and in TH was initiated in a pre-equilibrated K^+ solution. (ii) Folding in the TH experiments and the light-induced folding have been measured on *cMYC* sequences with a single isolated G-register isomer (53), while the initial folding experiment (4.2) reports on folding of a sequence that can fold into both G-register isomers 33 and 53.

To quantify these effects, the 18-mer *cMYC*-2345-wt (X3) that can fold in both G-register isomers has been trapped in an unfolded state following the same photocaging strategy. Figure 48 shows that this sequence can also be trapped in a mainly unfolded state. However, broad signals between 10.9 and 11.5 ppm indicate that in this case some pre-structured conformations can be formed. This nicely highlights the challenge to prevent base pair interactions in G4 forming oligonucleotides. Here, the addition (or better remaining with respect to the wt-sequence) of only a single G-residue enables at

least some extent of pre-formation. Therefore, a quantification of this experiment (in comparison to other light-induced folding experiments presented within this thesis) must be taken with caution. The fits to the kinetic traces shown in Figure 48 c, are meant to represent a trend in the folding kinetics. What can be qualitatively stated is that for the overall folding (integration between 10.8 – 12.2 ppm, black dots) is clearly biphasic.

The separated, conformation-specific traces follow again the same trends observed for the K^+ -induced folding, with a kinetic overshoot of the 33-isomer and subsequent re-equilibration. The initial folding phase (for overall folding) is finished within the dead time of the experiment, which is approximately $\sim 50\%$ faster compared to folding of the isolated 53-isomer (Figure 47). The subsequent refolding within the following ~ 5 minutes is $\sim 10^2$ -times faster than the refolding after K^+ -induced folding. The amplitude of the subsequent, slower refolding dynamics to reach the thermodynamic equilibrium is smaller than for the K^+ -induced folding experiment.

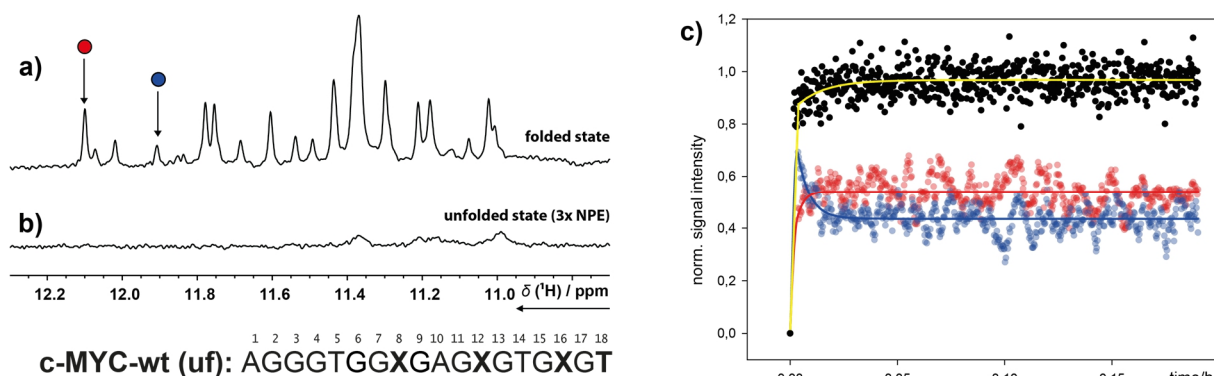


Figure 48: ^1H NMR spectrum of **a)** the 18-mer *cMYC* X3-sequence with two possible G-register isomers in a native oligonucleotide; and **b)** with three (*R*)-NPE cages at position G8, G12 and G16 (marked with X in the sequence, conformation-53). **c)** Kinetic trace for light-induced folding and bi-exponential fitting to show the trends in the folding kinetics (yellow, estimate for the relaxation time). Shown in black is the kinetic trace for the overall folding (normalized integration from 10.8 - 12.2 ppm). Shown in red (33) and blue (53) are the conformation-specific kinetic traces with exponential fits to show the trend in the refolding, subsequent to initial folding.

4.7 Conclusions

The major G-quadruplex conformation of the *cMYC* NHE-III₁ G4 forming sequence is able to adopt up to four different folded loop-isomers with *all*-parallel topology that are related by a G-register shift in the G-tract III or V (either 5'-shifted: 5; or 3'-shifted: 3, to yield 53, 33, 35 and 55). The folding and refolding kinetics of the two majorly populated^{101,142} G-register isomers 2345-53 and 2345-33 have been investigated in this chapter. Time-resolved NMR experiments of the folding process that were initiated either by rapid addition of K^+ (4.2) or by photolysis of multiple photocages (4.6.2) revealed a kinetic partitioning mechanism for the folding of the two co-existing G-register isomers. Folding is fast and proceeds on two parallel pathways that directly yield the isomeric folded states. Subsequent to this initial phase, a slower refolding takes places with an observable rate constant of $\sim 0.9 \text{ h}^{-1}$. The

refolding kinetics back to the conformational equilibrium between the two states have been delineated with light-induced experiments of photocaged isolated single conformations (4.3). The kinetics and activation energies (4.5) for two-state *cMYC* G-register exchange refolding (both parallel conformations) have been compared to two G-register shift related conformations in the *hTERT* promoter G4 (hybrid and parallel, 4.4). Finally, the isothermal NMR experiments have been cross-evaluated with an experimental approach based on thermal hysteresis for rapid thermal folding and unfolding (4.6.1). It was demonstrated that the presented photocaging strategy for conformational selection and conformational suppression of G4 states is a superior approach to unravel the kinetics of G-register exchange dynamics. It allows the observation of undisturbed relaxation dynamics for native, unmodified (after photolysis) oligonucleotides under isothermal and physiological relevant conditions.

The experimental data give striking evidence that the underlying refolding mechanisms for the different types of G-register exchange dynamics are fundamentally different (Figure 49). For *hTERT*, the evaluation of all presented experiments are in accordance with an unfolding-folding mechanism. The observed kinetics for isothermal refolding of the two G-register isomers in *cMYC* is 6x faster than expected for a complete unfolding-folding mechanism (estimated from thermal hysteresis experiments). A putative species of the transitory ensemble could be a triplex-like strand orientation that allows persisting tetrad scaffolds. The respective G-tract can be imagined as shifting or *sliding* along this remaining G4 core. However, the real physics of the transitory ensemble are probably better described by compacted coil-like structures.^{162,163}

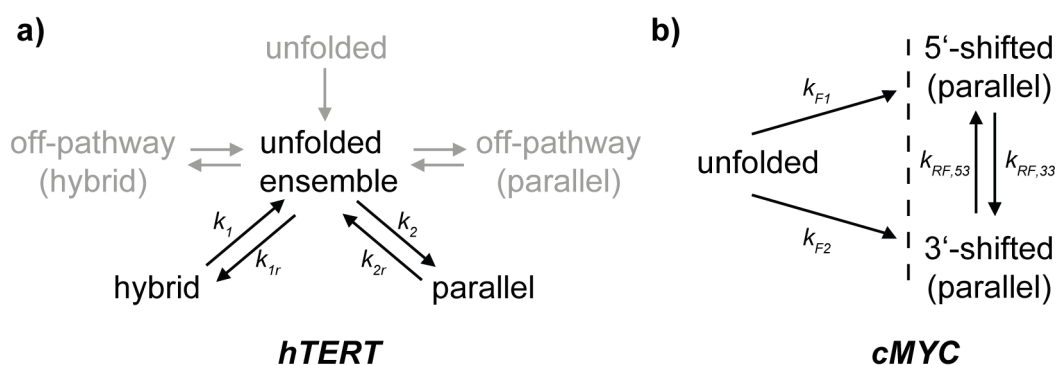


Figure 49: Kinetic models for the folding and refolding dynamics of G4-conformations that are related by a formal shift of G-registers with respect to the G4-core tetrads. (a) Model for the *hTERT* G4 ensemble with a co-existing hybrid and parallel conformation. (b) Model for the *cMYC*-2345 G4 ensemble with two co-existing parallel conformations.

The slow refolding kinetics for the G-register exchange dynamics exceed the timescales of biologically relevant processes by orders of magnitude (2.2.3). Therefore, the observed *off*-equilibrium distribution of G-register isomers directly after folding has to be taken into account for a structural evaluation of so-called “*major*” conformations in polymorphic G4 ensembles.

5 Spare-Tire Exchange Dynamics

5.1 Overview

Many G-quadruplexes have additional G-tracts (>4) that enables them to adopt different G-tract/loop isomers (2.1.3, 2.3.3). This non-canonical structural feature (2.1.3) is referred to as *spare-tire* isomerism. Accordingly, within this chapter the respective isomers will be referred to as “*spare-tire isomers*”.

As discussed and outlined above (4.6.2), folding kinetics following a K^+ -induced folding are remarkably slower than folding kinetics following folding initiation from uncaging or renaturation from a thermally denatured state. The successful approach of caging a completely unfolded state in the presence of K^+ was applied to investigate the folding kinetics of the *cMYC* G4 element in more detail.

To separate and deconvolute pure *spare-tire* exchange from G-register exchange dynamics, the sequence has been adapted to preclude the formation of G-register isomers (Figure 50). This sequence is called “*wildtype*”, with respect to the possibility to form all possible *spare-tire* isomers.

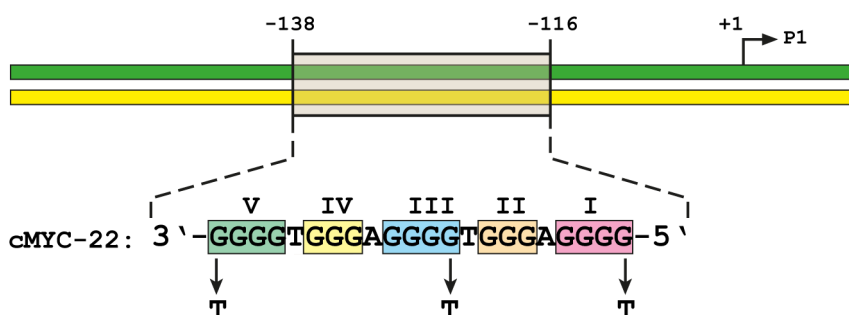


Figure 50: Depiction of the 22-mer sequence from the NHE-III₁ in the *cMYC* promoter region including G-tracts I-V (numbered from 5'-to-3' direction). Positions that were mutated (G-to-T) to preclude G-register isomers are indicated. Figure has been adapted according to the publications from the Hurley group.^{99,452}

The main results presented here have been published in Grün *et al.* 2021,⁷³³ *accepted manuscript (published online)* (doi: 10.1021/jacs.1c01089). All shown figures have been created originally by me (the author of this thesis) and are presented modified or unmodified as compared to the published version: Reprinted (adapted) with permission from *J. Am. Chem. Soc.* 2021, published online. (doi: 10.1021/jacs.1c01089) Copyright 2021 American Chemical Society.

5.2 Light-Induced Folding of Caged *Spare-Tire*-Isomers

5.2.1 Trapping Completely Unfolded *Spare-Tire* Isomers with Multiple Photocages

Here, the 22-nt long *cMYC* sequence was used to investigate the folding of the entire G4 forming sequence that spans all five (1-5) G-rich tracts. This sequence is able to fold into three distinct folded G4 conformations (Figure 51 b), namely the 5'-terminal conformation (5'-1234, abbr. **1234**), a long-looped conformation (5'-1245, abbr. **1245**) and the 3'-terminal conformation (5'-2345, abbr. **2345**). Figure 51 a shows the characteristic imino proton region of the ^1H 1D spectrum of the *cMYC*-22 wt-oligonucleotide. Wildtype here however should not be misleading, since in fact already mutations have been applied to lock a single G-register isomer for the 2345 conformation (G9T and G22T, resulting in *cMYC*-2345-53, see Figure 50). Following the distribution of populated conformations for the G4 isomers, 2345 will be called the **major conformation** and 1234 as well as 1245 will be called **minor conformations**.

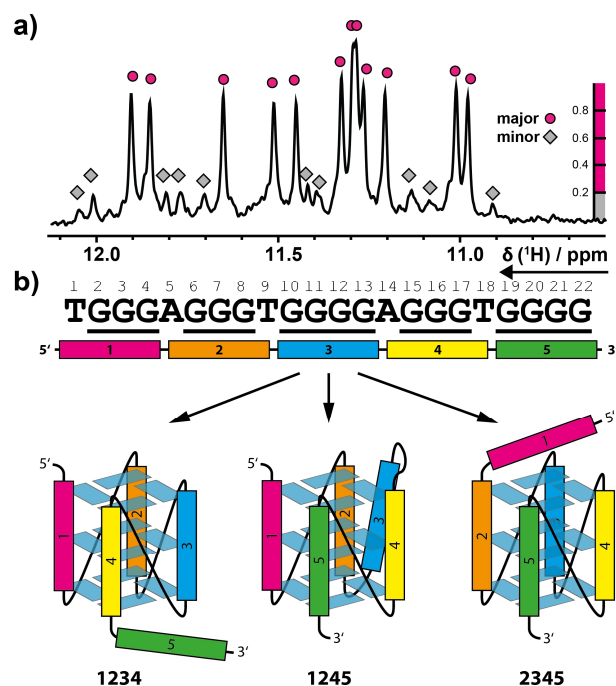


Figure 51: (a) N-H1 imino proton characteristic region of a ^1H -1D NMR spectrum (700 MHz, 298 K, excitation sculpting, 100 μM DNA, 5 mM K-P_i-buffer, pH=7.0) of wt-*cMYC*-22 and (b) simplified schematic representation of the *spare-tire* isomers.

To cage the *cMYC*-22-wt-12345 oligonucleotide in an unfolded state, three (*R*)-NPE-cages have been placed at the respective positions in G-tract III, IV and V (Table 12). The three different spare-tire isomers (1234, 1245 and 2345) have been separated and isolated in three different oligonucleotide sequences with G-to-T mutations in the respective G-tracts. In the folded state, these *stabilized* oligonucleotides can adopt only a single conformation, as the other *spare-tire* conformations cannot fold anymore. For each of these stabilized oligonucleotides then also up to three (*R*)-NPE-cages have been attached at different G-tracts to trap them in an unfolded state (Table 12). The photocaging strategy was successful to trap the oligonucleotides in an unfolded state.

Spare-Tire Exchange Dynamics

Table 12: Sequences and position of (*R*)-NPE-photocages of wildtype and stabilized spare-tire isomers. The wildtype sequence here denotes the 53-isomer only. (*X* = (*R*)-NPE-dG)

loop-isomer	oligonucleotide-sequence				
	I	II	III	IV	V
<i>wt</i> -12345	T <u>GGG</u> A <u>GGG</u> TT <u>GXG</u> A <u>GXG</u> T <u>GXG</u> T				
stab.-1234	T <u>GGG</u> A <u>GGG</u> TT <u>GXG</u> A <u>GXG</u> T <u>TTT</u> T				
stab.-1245	T <u>GGG</u> A <u>GXG</u> TT <u>TTT</u> A <u>GXG</u> T <u>GXG</u> T				
stab.-2345	T <u>TTT</u> A <u>GGG</u> TT <u>GXG</u> A <u>GXG</u> T <u>GXG</u> T				

The ¹H 1D spectra (Figure 52) show no signals in the characteristic region for N-H1 imino protons with Hoogsteen base pair interactions (10.4 – 12.5 ppm). From these spectra, the formation of any pre-folded states can be precluded. For partially folded or pre-folded states, at least broad signals would be expected, arising from interactions in the intermediate exchange regime.

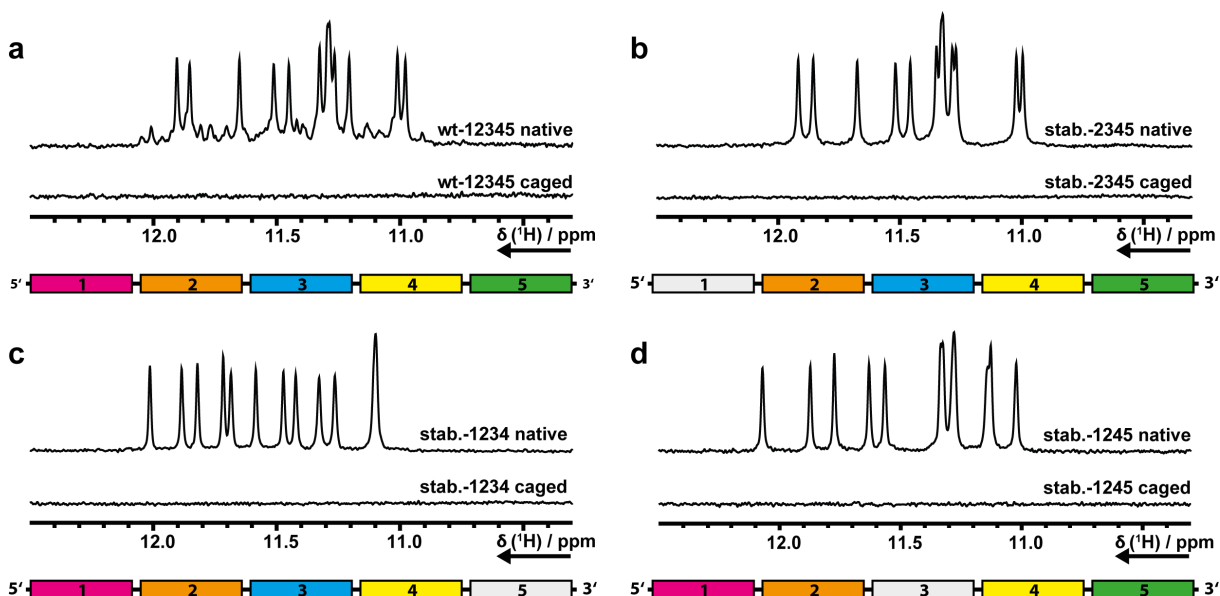


Figure 52: N-H1 region of ¹H 1D NMR-spectra of caged and native (uncaged) oligonucleotides for *cMYC*-22 wildtype (a) and *spare-tire* isomers (b-d), (700 MHz, 298 K, excitation sculpting, 100 μM DNA, 5 mM K-P_i-buffer, pH=7.0).

5.2.2 Light-Induced Folding Kinetics under Pre-Equilibrated K⁺ Conditions

Starting from these photocaged unfolded states, folding was triggered with 4 s of laser irradiation (3.4). The folding reactions were monitored using a pseudo-2D experiment that records a series of ¹H 1D spectra with jump-return-echo water suppression optimized for the detection of imino protons (700 MHz: d19 = 50 μs). The signals in the spectral region between 10.8-12.1 ppm (see also Appendix 6.2, Figure 67) have been integrated and the normalized signal intensity was plotted to yield kinetic traces for the build-up rate of the G4-specific imino protons (Figure 53).

Spare-Tire Exchange Dynamics

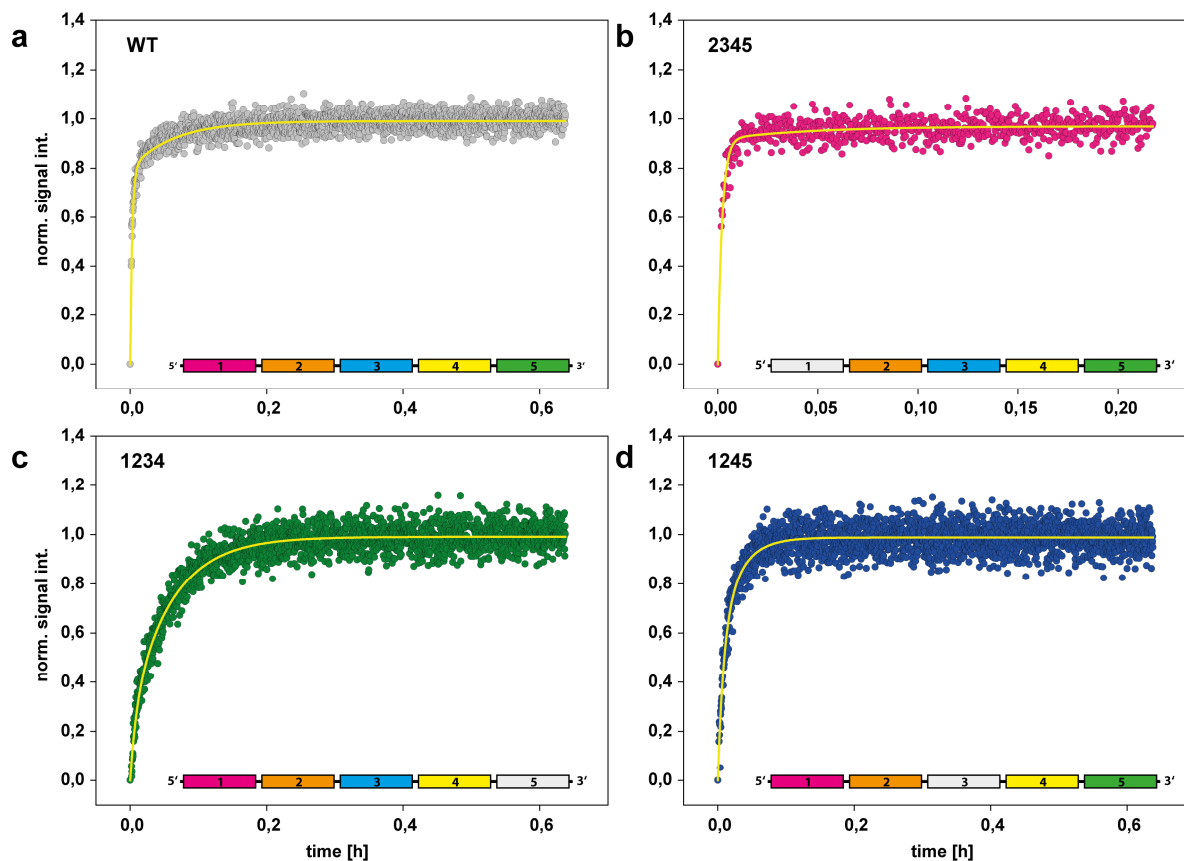


Figure 53: Plots of kinetic traces for light-induced folding of *cMYC-22*: wildtype (a, grey), major *spare-tire* isomer 2345 (b, magenta) and minor *spare-tire* isomers 1234 (c, green) and 1245 (d, blue). The plots show the normalized signal intensity for imino proton signals in the range between 10.8-12.1 ppm. Fits using bi-exponential non-linear regression are shown in yellow. (100 μ M DNA, 5 mM K-P_i-buffer, pH=7.0).

At room temperature the folding kinetics are sufficiently slow (s-min) to trace them, even after 4 s of laser irradiation, which is the initial dead time within the experiment. The time resolution between two points (1 scan per point) is ~ 1.155 s. All four folding kinetics (wt and three stabilized isomers) are clearly biphasic, with an initial faster folding regime and a subsequent slower folding regime. This statement is not meant to declare that the underlying folding process is only biphasic, but it can be multiphasic with convoluted kinetic steps. Based on the *observable* biphasic behaviour, the kinetics have been fitted with bi-exponential non-linear regression, which yields excellent representations for the folding reactions. The observable kinetic rate constants for folding as well as the amplitudes for the folding phases are given in Table 13.

Table 13: Observable kinetic rate constants (k_1 , k_2) for light-induced folding of *cMYC-22* *spare-tire* isomers obtained from bi-exponential fitting. The ratio $k_1:k_2$ describes the relative amplitudes of the folding phases.

isomer	k_1 [h^{-1}]	k_2 [h^{-1}]	$k_1 : k_2$
stab.-1234	110 ± 16	17 ± 13	0.26 : 0.74
stab.-1245	119 ± 18	35 ± 5	0.56 : 0.44
stab.-2345	461 ± 22	22 ± 8	0.94 : 0.06
wt-12345	348 ± 13	16 ± 1	0.80 : 0.20

The folding kinetics of *cMYC-22* have been compared to folding of shortened oligonucleotides, *cMYC-18*. Here no mutations are needed, because truncation of the 5'-terminal or 3'-terminal *spare-tire* preserves for alternative G-tract interactions. Table 14 shows the sequences for the 5'-truncated (short-1234) and 3'-truncated (short-2345) oligonucleotide. The *cMYC-18-short-2345* sequence is comparable to the *cMYC-18-X3* sequence used in (4.6.2), but with a G6T mutation, that locks the 53-G-register isomer.

Table 14: Sequences and position of (*R*)-NPE-photocages of shortened 5'-terminal and 3'-terminal conformations (*cMYC-18*). (X = (*R*)-NPE-dG)

loop-isomer	oligonucleotide-sequence									
	I	II	III	IV	V					
short-2345		A	<u>GGG</u>	TT	<u>GXG</u>	A	<u>GXG</u>	T	<u>GXG</u>	T
short-1234	T	<u>GGG</u>	A	<u>GXG</u>	TT	<u>GXG</u>	A	<u>GXG</u>	T	

Figure 54 shows the kinetic traces and bi-exponential fits of the light-induced folding. Here, again folding is clearly biphasic with the kinetic rate constants given in Table 15. Noteworthy, the kinetics are significantly accelerated compared to folding of the full-length stabilized *spare-tire* isomers. The fast folding phase for folding of 2345 cannot be quantified; hence folding is accelerated at least by a factor of ~ 6.5 (as estimated from the experimental dead time).

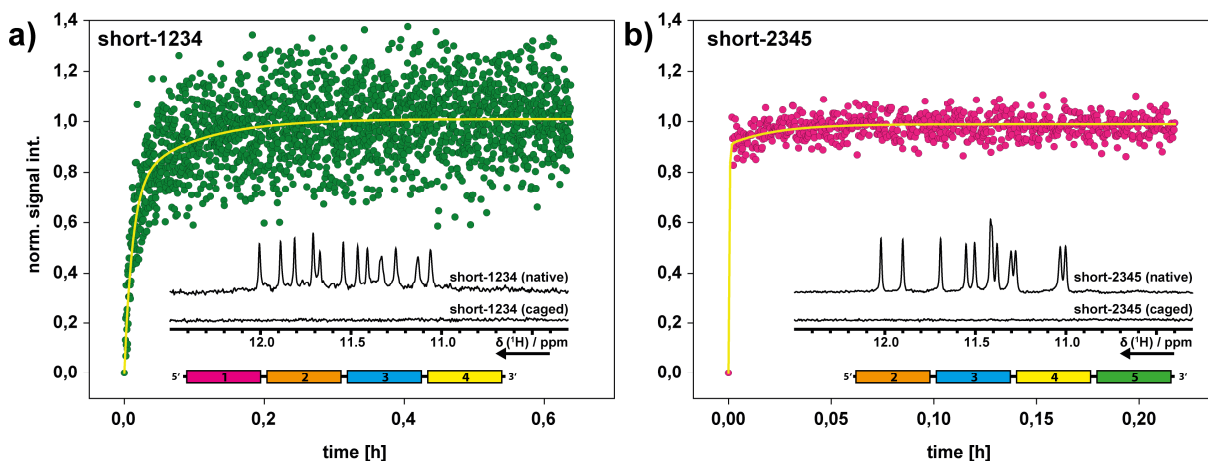


Figure 54: Plots of kinetic traces for light-induced folding of *cMYC-18* (shortened 5'-terminal (1234, **a**) and 3'-terminal (2345, **b**) conformations). The plots show the normalized signal intensity for imino proton signals in the range between 10.8-12.1 ppm. Fits using bi-exponential non-linear regression are shown in yellow. (100 μ M DNA, 5 mM $K-P_i$ -buffer, pH=7.0).

Table 15: Observable kinetic rate constants (k_1 , k_2) for light-induced folding of *cMYC-18* *spare-tire* isomers (1234 and 2345-53) obtained from bi-exponential fitting. The ratio $k_1:k_2$ describes the relative amplitudes of the folding phases.

loop-isomer	k_1 [h^{-1}]	k_2 [h^{-1}]	$k_1 : k_2$
short-2345 (53)	n.d.	45 ± 3	0.92 : 0.08
short-1234	86 ± 10	11 ± 2	0.74 : 0.26

5.2.3 Intermediate Formation and Folding Pathway for *cMYC*-1234

At lower temperatures (285 K), folding of *cMYC*-1234 revealed the formation of a long-lived intermediate state. In total eight additional imino proton signals build up after photolysis, mainly shifted to the downfield region (>12 ppm, Figure 55 a). The additional signal build up in parallel to the build-up of imino proton signals that are conformation specific for the 1234 conformation (see also Appendix 6.2, Figure 67).

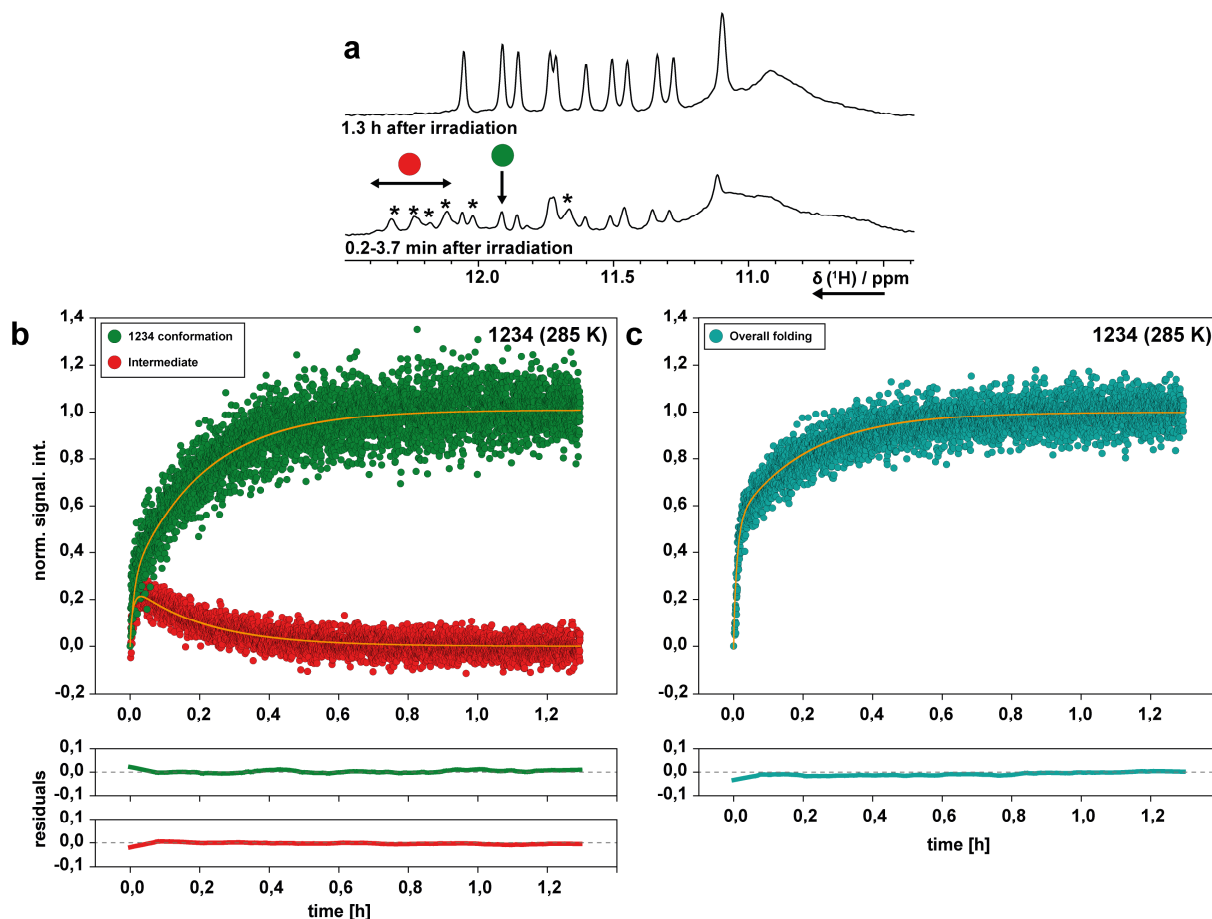


Figure 55: (a) Imino proton NMR spectra of stabilized 1234 at indicated times after light irradiation. (b) Kinetic traces for light-induced folding of stabilized 1234 *spare-tire* isomer. Shown in green is the trace of only a single conformation-specific signal and shown in red the conformation-specific signals for the intermediate conformation. (c) Shown in turquoise is the kinetic trace for the overall folding (10.8 – 12.4 ppm). Kinetic fits arise from fitting according to a kinetic model (Figure 56, 3.3) with the respective residual plots shown below.

Figure 55 b shows two kinetic traces from conformation specific (1234: green, intermediate: red, [12.1-12.4 ppm]) imino proton signals. After ~ 4 min, the kinetic trace for the intermediate conformation reaches a maximum at $\sim 20\%$ of the normalized total signal intensity in equilibrium. Afterwards it decays over a period of about 1 h until the peaks completely vanish. Figure 55 c (turquoise) instead shows the integrated peak intensity between 10.8-12.4 ppm, which reflects the overall, conformation independent folding progression. The overall folding is strictly separated in two regimes. The initial folding phase is significantly accelerated in comparison to the conformation specific folding of 1234 alone. The initial folding phase then abruptly passes into a slower phase, after

the maximum intensity for the intermediate conformation is reached. This kinetic behaviour is expected for parallel folding pathways, which defines the intermediate as *off*-pathway with respect to the exclusively populated 1234 conformation in equilibrium. The kinetic traces in Figure 55 b have been fitted to a kinetic model that accounts for parallel folding pathways (Figure 56). The fitting procedure (3.3) used global fitting and yields a prediction for the folding kinetics with the respective kinetic rate constants given in Table 16.

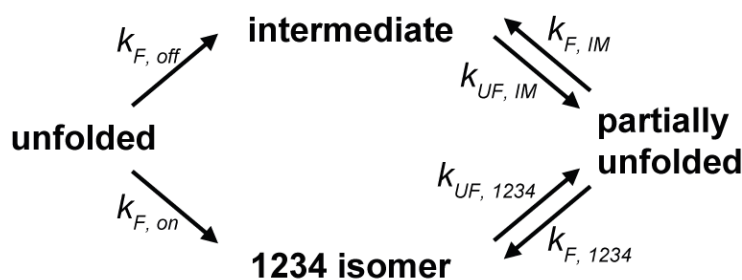


Figure S6: Kinetic model for parallel folding pathways of 1234. The initial folding is irreversible, as the completely (potentially random coil like) unfolded state with photocages cannot be recovered. The ensemble of partially unfolded states accounts for any “not-folded” species.

Table 16: Kinetic rate constants from global fitting for folding of 1234 at 285 K according to the kinetic model depicted in Figure S6. The calculated error is within 5% as well as the error estimate from fitting residuals.

fitted rates [h ⁻¹]			
<i>intermediate</i>		1234	
$k_{F,off}$	272	$k_{F,off}$	496
$k_{F,IM}$	7.44	$k_{F,1234}$	62.7
$k_{UF,IM}$	0.00136	$k_{UF,1234}$	29.3

To support the kinetic NMR data, time-resolved CD spectra have been measured. CD-spectroscopy can help to elucidate the secondary structure formation.^{569,571} The CD-cuvette was attached to the 355 nm laser-setup that has also been used for real-time NMR experiments in a self-designed experimental setup (3.2). After manually triggered laser-irradiation, CD-spectra in the range between 250-300 nm (Figure 57 a, 6 s interval) or single wavelength measurements at 290 nm (Figure 57 b, 1 s interval) have been acquired. Peaks in the CD-spectra give hints for the folding topology of G4s: positive bands at ~260 nm show *all*-parallel folds and positive bands at ~290 nm show *anti*-parallel (or hybrid) folds.^{170,571}

The time-resolved CD-spectra show that while the intensity of the band at 260 nm steadily increases, a small band at 290 nm again shows an intermediate build-up curve. The kinetic trace of this intermediate band at 290 nm then was normalized and plotted together with the kinetic trace obtained from time-resolved NMR-spectra. The traces obtained from both methods are completely superimposable, which allows concluding that the topology of the intermediate state is in an *anti*-parallel conformation.

Spare-Tire Exchange Dynamics

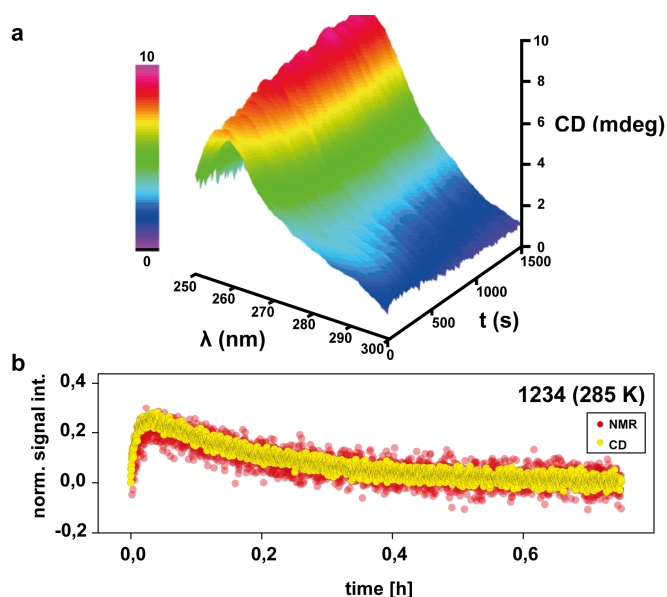


Figure 57: Time-resolved CD data for light-induced folding of 1234. (a) Series of time-resolved CD spectra in the range between 250 – 300 nm. (b) Normalized overlay of the kinetic trace for an intermediate conformation obtained from NMR (12.1 – 12.4 ppm red) and from CD (290 nm) spectroscopy.

5.3 Temperature Dependence of Folding

Light-induced folding has been investigated at variable temperatures in the range between 285 and 310 K. The kinetics for each isomer at every temperature have been fitted with bi-exponential non-linear regression to yield observable rate constants k_1 and k_2 . Arrhenius plots of the apparent kinetic rate constants are shown in Figure 58 and they show that the temperature difference for folding of each of the isolated *spare-tire* isomers is fundamentally different.

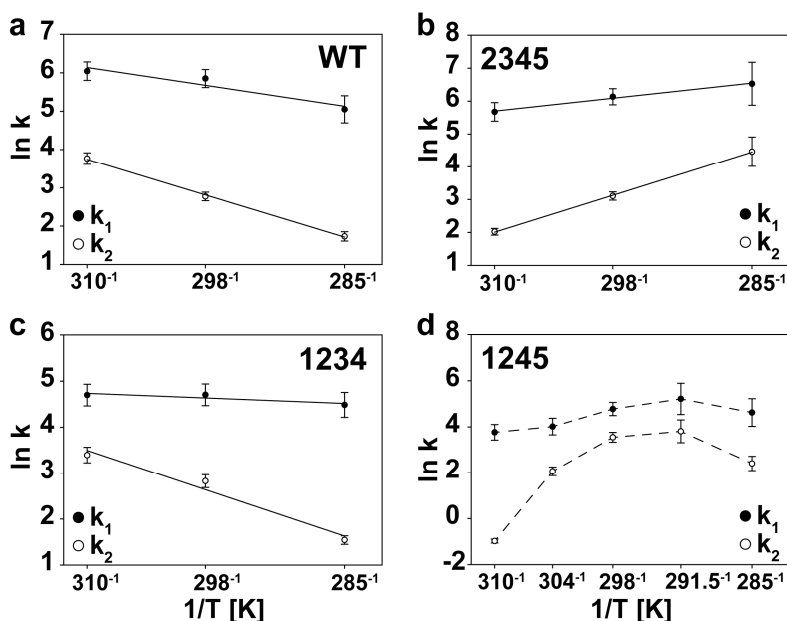


Figure 58: Arrhenius plots for temperature dependent light-induced folding of *cMYC-22 spare-tire* isomers. (a-c) have been fitted with linear regression, dashed lines in (d) connect the individual experimental points without underlying fitted model to follow the trend.

The observed apparent activation energy barriers are derived from observable rate constants and cannot be assigned directly to distinct kinetic steps during the complex folding process. They are a superposition of different energy barriers from individual rate constants that are involved in the folding mechanism. These barriers give strong indications for the rate-limiting barriers during folding.

Both the wildtype sequence and the isolated 1234 conformation have a linear Arrhenius correlation with a positive apparent activation energy barrier. The isolated 2345 conformation instead has a linear Arrhenius behaviour with a negative apparent activation energy barrier. The calculated apparent activation energies are given in Table 17. The 1245-conformation shows non-Arrhenius behaviour in the temperature range between 285 and 310 K.

Table 17: Calculated apparent activation energy barriers for *cMYC-22 spare-tire* isomers, derived from observable rate constants k_1 and k_2 . The energies are calculated from Arrhenius analysis shown in Figure S8.

isomer	ΔE_{A1} [kJ·mol ⁻¹]	ΔE_{A2} [kJ·mol ⁻¹]
stab.-1234	6.4 ± 3.5	54.9 ± 9.9
stab.-1245	-	-
stab.-2345	-25.0 ± 2.3	-71.8 ± 1.4
<i>wt</i> -12345	29.6 ± 8.9	59.7 ± 2.4

The positive activation energy barrier for folding of the stab.-1234 isomer can be explained with the formation of *off-pathway* intermediate conformations that need to be unfolded. The observed *anti-parallel* species presented presumably gives major contributions to the observed total energy barrier, but the activation energy does not necessarily reflect the unfolding energy of only this NMR-visible conformation.

The fast folding relaxation time of the isolated stab.-2345 isomer and the negative apparent activation energy barrier, indicate a mainly entropically driven, close to funnel-like folding process for this conformation. The conformational space during folding is drastically reduced for this sequence, since both the formation of G-register isomers and the competing *spare-tire* isomers is precluded. Nevertheless, the folding kinetics are remarkably slower than those obtained for the 18-mer *cMYC-2345* sequences (4.6.2 and Figure S4; both with and without possible G-register formation). This points to (i) an influence of the flexibility of the 5'-tail and (ii) possible transient base pair interactions with the 5'-elongated nucleotides. These interactions are precluded in the folding experiments of the short *cMYC-2345* oligonucleotides, which further guides an optimized folding process.

Folding of the 3'-terminal (2345) and 5'-terminal (1234) conformations mark limiting regimes for either (i) pronounced kinetic partitioning with intermediate formation or (ii) close to funnel-like folding. The non-Arrhenius temperature dependence for the isolated stab.-1245 isomer indicates that

here folding takes place in an interesting junction region on the folding energy landscape. At lower temperatures, folding is rate limited by unwinding kinetic traps with energy barriers, while at higher temperatures the process gets more entropy driven, with only a negative observed folding energy barrier.

Different kind of possible *off*-pathway interactions and transient kinetic traps might be present in the folding of the wt-sequence that shows a positive apparent activation energy barrier. This energy reflects refolding barriers from any possible microstates in the large conformational space of the wt-sequence that have to be overcome. To examine, if these *off*-states possibly involve fully folded minor conformations (*spare-tire* isomers 1234 and 1245), refolding from these states was investigated.

5.4 Light-Induced Refolding of Caged *Meta*-Stable Minor Conformations

5.4.1 Conformational Selection

In chapter 4.3.1 a strategy to trap and isolate single folded G4 conformations has been presented using site-specific suppression of Hoogsteen-interactions with (*R*)-NPE-photocages. This strategy has been applied to trap the minor conformations 1234 and 1245. Figure 59 shows the 1D ^1H NMR spectra in the relevant Hoogsteen base paired imino proton ppm-range for caged wt-1245 (G12 caged, b) and wt-1234 (G20 caged, c) and a spectrum of the native (uncaged) wt after complete relaxation (a). The spectra clearly show that the conformational selection was successful and that single conformations could be isolated. The caged wt-1245 however shows additional imino proton signals (marked with #) that might indicate the formation of a second (or partially folded) conformation.

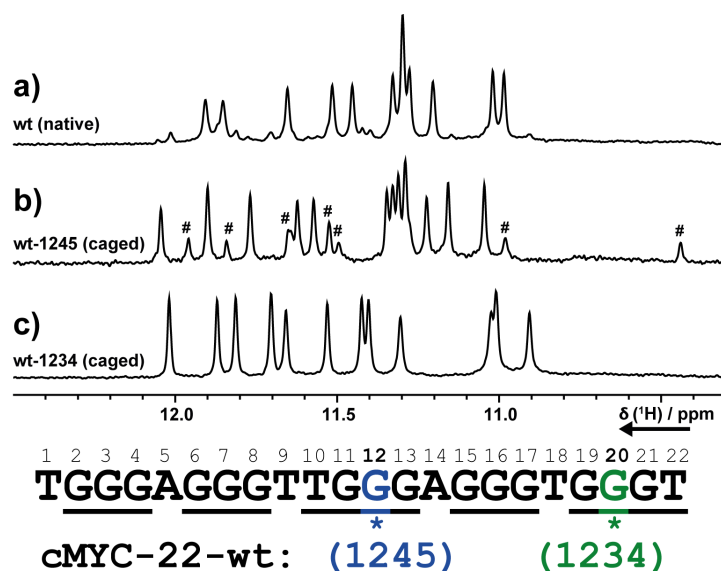


Figure 59: Trapped minor conformations as *meta*-stable states of *cMYC*-wt-1234 (G20 caged) and *cMYC*-wt-1245 (G12 caged). Before (b, c) and after uncaging (a, complete relaxation). (700 MHz, 298 K, excitation sculpting, 100 μM DNA, 5 mM K-P_i -buffer, pH=7.0). (*) indicates caging positions to yield 1245 (blue) or 1234 (green). (#) indicates an additional set of signals for the photocaged conformation.

Figure 60 shows a comparison of the caged spectrum (lower) and a spectrum directly after uncaging (upper, 4 s laser irradiation + 1.15 s acquisition time). The marked, additional signals are completely vanished, before any kind of spectral changes can be observed. Similar observations were made for the caged parallel-*hTERT* conformation (chapter 4.4, Figure 55). It cannot be completely ruled out that these kind of signals arise from partially unfolded or otherwise disturbed folded conformations. However, two findings strongly oppose that assumption: (i) the very sharp linewidth that indicate a defined state in the slow exchange regime; and (ii) the very fast decay of the signals in comparison to even the fastest folding, unfolding or refolding transitions for all of the investigated G4 dynamics. Furthermore, this would imply either that these “destabilized” conformations, after uncaging, convert directly to the completely folded conformation or that they completely unfold. If that were correct, a rapid folding of a stable G4 conformation must be expected. The NMR-spectra rule that out. Therefore, a likely explanation for these signals could be alternative arrangements (e.g. stacking) of the (R)-NPE-residues that cause perturbed chemical shifts. This explanation would be in line with a moderate shift of the imino proton signals after uncaging that has been observed for the trapped G4 conformations.

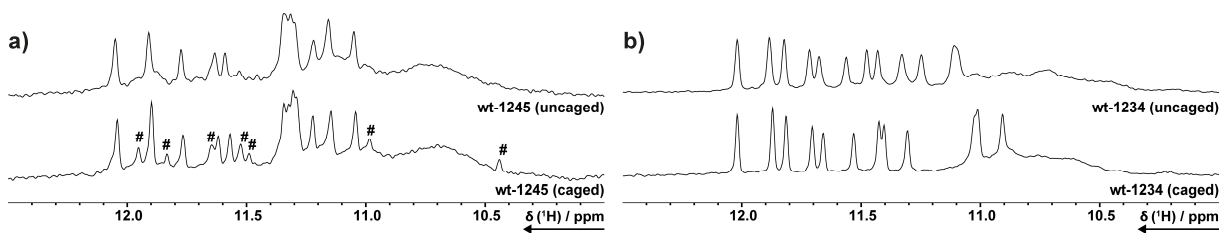


Figure 60: Trapped minor conformations as *meta*-stable states of *cMYC*-wt-1245 (a) and *cMYC*-wt-1234 (b). Before (lower spectra) and directly after uncaging (upper spectra) (700 MHz, 298 K, jump-return-echo, 100 μ M DNA, 5 mM K-P₁-buffer, pH=7.0) (#) indicates an additional set of signals for the photocaged conformation.

5.4.2 Refolding Kinetics

After quantitative uncaging, the native oligonucleotides are released, persisting in their initially folded G4 conformation. No additional imino proton signals were observed within a short time after photocleavage of the caging-groups, which would indicate the presence of additional conformations. Hence, the starting point of the experiments represents the unperturbed, native state of the single *spare-tire* G4 conformations. This exclusive population of the minor conformations could otherwise not be observed for the wildtype *cMYC*-sequence.

The initial conformations then decay over a period of several hours, while signals for the major conformation 2345 rise up (Figure 61). The kinetics for both the unfolding of the minor conformations and the subsequent folding of the major conformation are correlated and progress coherently. The observable rate constants have been obtained from a global fit of both traces with double mono-exponential non-linear regression (Table 18).

Spare-Tire Exchange Dynamics

At room temperature, a complete refolding of the minor conformations to the major conformation was observed. Hence, the minor conformations are completely *meta*-stable. Arrhenius analysis of temperature dependent refolding allows an estimate for the apparent activation energy of refolding (Table 18).

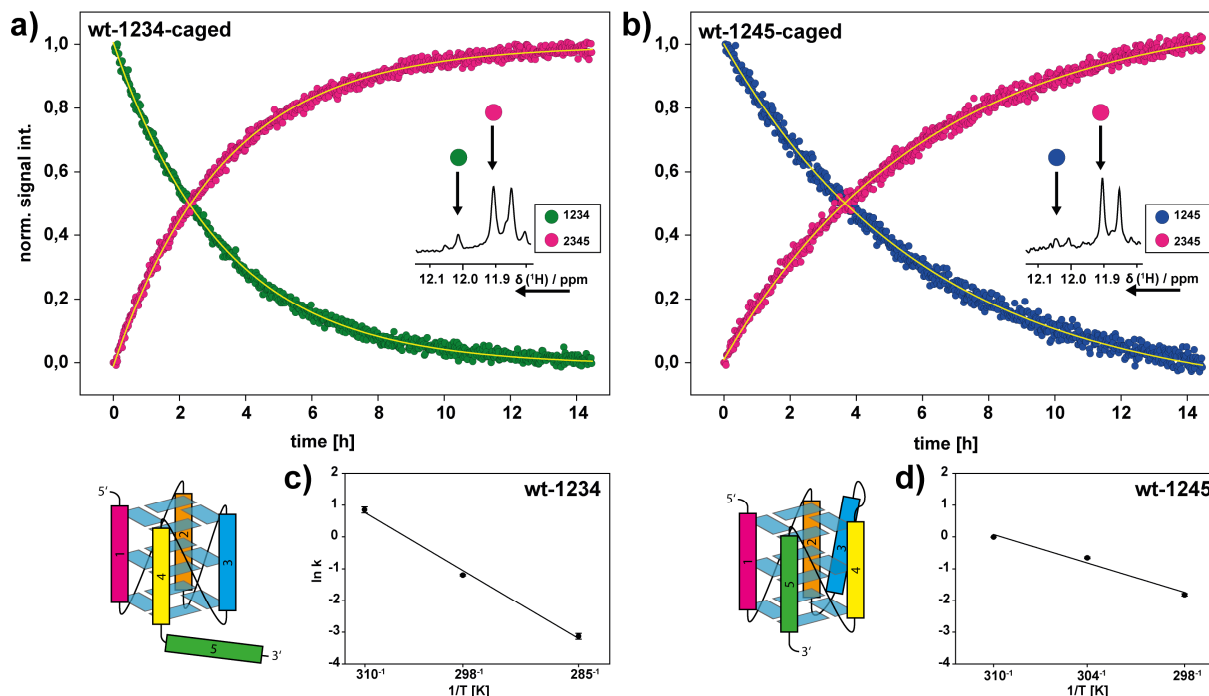


Figure 61: Refolding transition of trapped *meta*-stable *cMYC*-22 *spare-tire* isomers 1234 and 1245 at 298 K. (a, b) Kinetic traces of individual imino proton signals, representative for the two folded conformations (magenta: 2345, green: 1234, blue: 1245). Fitting curves from global mono-exponential regression are shown in yellow. (c, d) Arrhenius analysis with linear regression to calculate apparent activation energies for refolding (Table 18).

Table 18: Observable rate constants for refolding of trapped conformations 1234 and 1245 from the 22-mer *cMYC*-wt oligonucleotide. Apparent activation energies have been calculated from Arrhenius analysis (1234: 285 – 310 K; 1245: 298 – 310 K).

loop-isomer	k_{obs} [h^{-1}]	ΔE_{A} [$\text{kJ}\cdot\text{mol}^{-1}$]
caged-1234	0.302 ± 0.006	116 ± 8
caged-1245	0.163 ± 0.001	116 ± 19

For the 1234 conformation, a complete refolding was observed even at 285 K. For the 1245 conformation, at lower temperature (285 K), no refolding and no formation of conformation 2345 was observed (Figure 62). This is unexpected with regard to the calculated apparent activation energies of 1234 and 1245. The activation energy barriers however, reflect only the rate-limiting step for the unfolding of a certain conformation. The greater degree of required rearrangements for the 1245 conformation compared to 1234, might explain why 1245 has an increased lifetime. Irrespective of the same heights for the energy barriers that have to be overcome for unfolding to the transitory ensemble for both minor conformations, presumably 1245 is kinetically trapped.

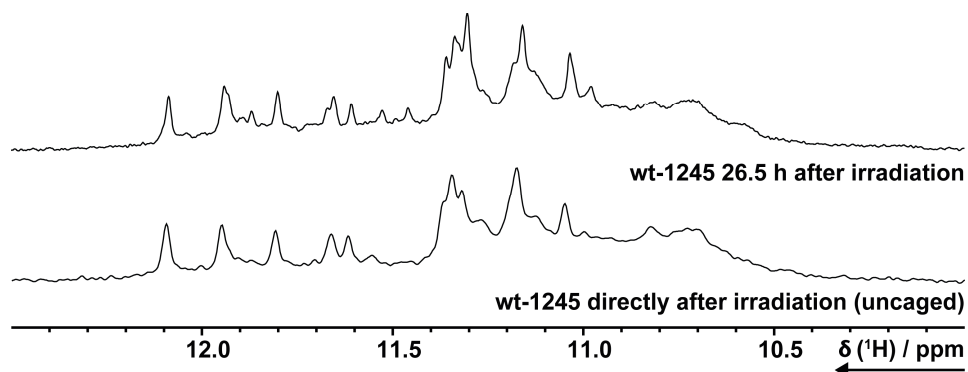


Figure 62: Trapped minor conformation of *cMYC*-wt-1245 at 285 K. Before (**lower spectrum**) and after directly after uncaging (**upper spectra**) (700 MHz, 298 K, jump-return-echo, 100 μ M DNA, 5 mM K-P_i-buffer, pH=7.0)

5.5 Conclusions

The G4 forming sequence in the *cMYC* NHE-III₁ is able to adopt up to three different folded loop-isomers with *all*-parallel topology that incorporate different G-tracts into tetrad-formation (either 5'-1234; 5'-1245 or 5'-2345). The folding and refolding kinetics of these *spare-tire* isomers have been investigated in this chapter. Time-resolved NMR experiments of the folding process that were initiated by photolysis of multiple photocages (5.2) revealed inherently different folding kinetics. Folding was investigated for separated and stabilized single conformations (1234, 1245 and 2345) and for the wildtype-sequence that is able to adopt all *spare-tire* conformations (but only a single G-register conformation). Folding of all *spare-tire* isomers is biphasic and proceeds on multiple parallel pathways that yield the isomeric folded states. Analysis of the temperature dependence of the folding reactions (5.3) reveals different manifestations of the underlying kinetic partitioning mechanisms. The minor conformations 1234 and 1245 have been isolated with a trapping strategy for conformational selection (5.4). The preparation of this single-conformation, *off*-equilibrium populations revealed that both minor conformations are *meta*-stable with respect to the major *cMYC*-G4 conformation 2345.

Folding kinetics of the wt-sequence could be disentangled as a superposition of the folding kinetics of the minor conformations (20% populated in the wt) and the major conformation (80% populated in the wt). The major conformation *cMYC*-2345 folds via a close to funnel-like pathway with optimized kinetics. For the minor conformation *cMYC*-1234, the formation of an *anti*-parallel off-pathway conformation was demonstrated both with time-resolved NMR and with CD experiments (5.2.3). The minor conformation *cMYC*-1245 showed an unprecedented non-Arrhenius temperature dependence in the folding kinetics, which is presumably linked to the increased flexibility of the long internal loop that separates the 5'-terminal and the 3'-terminal parts of the tetrad-constituting parts of the oligonucleotide. Thus, the observed inherently different folding dynamics for the formation of *spare-tire* G4 conformations demonstrate markedly different limiting regimes on the folding energy landscape for structurally very similar parallel G4s emerging from the same oligonucleotide sequence.

6 The Dynamic Behaviour of the cMYC Promoter G-Quadruplex

6.1 Folding Pathways for the 22 nt cMYC Full Length Sequence

6.1.1 K⁺-Recruitment During Folding

Folding has been investigated with several methods within this thesis, using time-resolved NMR-spectroscopy, time-resolved CD-spectroscopy and a thermal hysteresis approach with UV detection. To obtain kinetic information with these techniques, different approaches were used to prepare unfolded states. In chapter 4.2, cMYC G4 folding has been initiated by addition of K⁺-ions, while in chapters 4.6, and 5.2 and folding has been initiated under pre-equilibrated K⁺-conditions (thermal denaturation and photocaging). Strikingly, the folding kinetics for K⁺-induced folding deviate by several orders of magnitude, when compared to the other methods.

K⁺-induced folding has been used in numerous studies^{144,167,170} on G4 folding and the findings presented in these studies and within this thesis raise no doubt that this is a legitimate approach to study these processes. However, the recruitment and coordination of K⁺-ions has a significant effect on the overall folding dynamics, which is a direct conclusion from the presented folding kinetics. Experimental approaches can only describe macrostates (if any), but no microstates that in sum define the conformational entropy of G4 folding. Hence, the folding kinetics do not allow speculating about the reasons for the decelerated folding, after addition of K⁺. Although it seems reasonable to assume that the conformational space (conformational energy landscape) of the unfolded oligonucleotides is fundamentally different for (i) thermal denaturation, (ii) destabilization with photocages or (iii) K⁺-free conditions. The contributions of entropy and enthalpy for removing the hydrate shells from free vs. DNA-coordinated K⁺-ions may add important energetic aspects that alter the folding pathways. Insights for the binding properties of cations to G-quadruplexes have been obtained from theory⁷³⁸ and mass spectrometry^{167,739}. Especially ¹⁵NH₄⁺ has been extensively studied with NMR,^{93,740–747} but also G4-related experiments have been reported for a direct NMR detection^{748,749} of e.g. ²³Na⁺, ³⁹K⁺ or investigation of ion coordination sites with cross-correlated relaxation rates.⁷⁵⁰ Further NMR experimental data and insights from computational methods and simulations will be needed to elucidate the role of K⁺-binding during folding.

This is of particular importance if K⁺-coordination becomes rate limiting during folding in a way that e.g. K⁺-coordination does not happen concerted on both inter-tetrad “binding sites”.^{167,751} In this regard, the misfolded or partially folded states proposed by Marchand and Gabelica based on mass spectrometry¹⁶⁷ have to be evaluated carefully. Here, the authors describe, unlike in other folding studies, a step-by-step titration with substoichiometric amounts of K⁺. If the lifetimes (or better: *dwell times*) for this intermediate or transition states are longer (slower) than subsequent kinetic steps, then the associated misfolding pathways may not be relevant under native, K⁺-containing conditions.

Marchand and Gabelica describe a drastic acceleration of folding kinetics in the range of $<1 \text{ mM K}^+$ ($10 \text{ }\mu\text{M DNA}$). This effect seems to be mainly saturated in the range of $2 - 3 \text{ mM K}^+$ ($>100 \text{ }\mu\text{M DNA}$).¹⁷⁷

A rapid change of the ionic conditions is by no means physiological, and so is thermal denaturation. The approach presented in this thesis, using photocages to prepare unfolded states is a clear advancement in this regard, as it can be applied without further requirements to experimental and sample conditions. Folding can be studied under constant, isothermal experimental conditions with any biophysical detection method.

6.1.2 Parallel Reaction Pathways Accelerate Folding (G-Register Isomers)

This chapter relates to the major findings of the publication: “**Parallel reaction pathways accelerate folding of a guanine quadruplex**” (Harkness, Hennecker, Grün *et al.* NAR 2021).⁷⁵² This publication contains contributions from me (the author of this thesis) using the photocaging approach for light-induced folding.

In protein folding processes, the existence of parallel folding pathways reduces the entropy penalty for the folding, since the conformational entropy can be maximized. From these basic considerations⁷⁵³⁻⁷⁵⁹, the hypothesis follows that if different parallel folding pathways are occupied during G4 folding, the overall folding should be accelerated. The observable rate constant for the folding of a wildtype sequence into a structural ensemble that enfolds all conformations that can be possibly adopted (x, y, z) is given as:

$$k_{F, wt} = k_{F, w} + k_{F, x} + k_{F, y} + k_{F, z} \quad (1)$$

Given that the individual rate constants were equally fast and the related conformations were equally populated, folding of the wildtype would be 4x faster than the isolated conformations. Mittermaier and co-workers have outlined this hypothesis and investigated the predicted consequences on the four possible G-register isomer conformations from the 18-mer cMYC-2345 sequence (33, 35, 55, 53). Converting (1) into the following expression:

$$k_{F, wt} \cdot \langle k_{F, 33, 35, 53, 55} \rangle^{-1} = 4 \quad (2)$$

(with $\langle k_{F, \dots} \rangle$ as mean value of the rate constants) gives a folding acceleration that is equal to the number of folding pathways. They conducted a series of experiments based on rapid thermal (un)-folding with thermal hysteresis, comparable to those presented in chapter 4.6.1. These experiments have been supported with light-induced (isothermal) folding experiments observed with real-time NMR (from me, the author of this thesis). The trend in these experiments supports the general idea that G4 folding can be significantly accelerated by occupation of multiple parallel folding pathways.

6.1.3 Enlarged Conformational Space Decelerates Folding (*Spare-Tire* Isomers)

The influence of *spare-tire* isomers on the kinetics of the major folding pathways follows a different trend as observed for the G-register isomers. Here, the kinetics of the wildtype sequence with minimal restrictions to the conformational space are slower than those observed for the kinetics of the isolated major conformation. The wt-sequence is able to adopt all single isolated conformations and further all possible sub- and microstates, with unproductive interactions and misfolded species. This effect is expected for G4 folding with kinetic partitioning of competing basins of attraction on the folding energy landscape.

The fact that the competing minor conformations are also formed concurrently to the major conformation, despite their $\sim 4x$ slower folding kinetics, points towards irreversible folding steps that enable these slower pathways. These tipping points separate the folding pathways and lead to kinetic traps that collapse into the folding basins of the minor conformations. This results in an overall slower folding process (if described as a decay function of unfolded states towards *any* folded G4 structure).

6.2 Folding intermediates

In chapter 5.2.3 the experimental observation of a long-lived (several minutes at 285 K) intermediate formation for *cMYC*-1234 is presented. The kinetics for time-resolved NMR and CD spectra are superimposable and clearly indicate the formation of an *anti*-parallel conformation. Furthermore, the observation of eight well resolved imino proton resonances in the downfield-shifted region (>12.1 ppm) point to eight G-residues in slow-exchange (base paired). Marchand and Gabelica¹⁶⁷, as well as Gray *et al.*¹⁷⁰ (298 K, up to 25 mM K^+) have made strikingly similar observations on folding for *cMYC*-2345 (in **33** G-register conformation, PDB: 1XAV) with CD-spectroscopy. The *cMYC*-2345-**33** G4 features the same loop geometry as the *cMYC*-1234 G4 (lateral:proximal:lateral = 1:2:1 nts). They find *anti*-parallel intermediates persisting on a comparable time scale. Gray *et al.* have proposed a kinetic model that involves the formation of an *anti*-parallel *chair* G4 conformation that can evolve into the final *all*-parallel conformation. They suggest that this process does not require complete unfolding, because estimates for the unfolding rates (derived from complement trapping experiments) are significantly slower than the observed folding relaxation time. Finally, Marchand and Gabelica could show with mass spectrometry that the elusive *anti*-parallel intermediate binds only a single K^+ -ion.

The discussed experimental findings from the literature, in line with the experimental findings presented in this thesis, prompt proposing a 2-tetrad chair conformation as possible intermediate. Only the *anti*-parallel 2-tetrad G4 can account for all observed experimental data.

The formation of this intermediate species is further supported by predictions from molecular dynamics.^{157,161–163} The formation of a 3-tetrad *chair* G4 as proposed by Gray *et al.* does not account (i) for the single occupied K⁺-binding site, and (ii) the resulting 1 nt lateral loops are highly disfavoured for *anti*-parallel hairpins. The non-native *syn/anti* distribution in the *chair* G4 must be disentangled to enable a refolding transition to the *all-anti* sugar configuration in an *all*-parallel G4 conformation. Therefore, the observed intermediate formation is rather an *off*- than an *on*-pathway state.

The proposed structural model for the intermediate conformation is highly sensitive on the length of the proximal loop, which explains the differences in folding kinetics for the *spare-tire* isomers *cMYC* 2345 and 1245 (Figure 63). In the light-induced folding experiments for the *cMYC*-2345 in **53** G-register conformation (chapter 5.2.2), no stable intermediate formation was observed and folding was significantly faster compared to folding kinetics of other *cMYC* G4 conformations. The reason for this is likely the length of the 1 nt proximal loop resulting in a 2:1:1 loop geometry that prevents a tetrad formation of initially collapses terminal hairpins.

The unparalleled increased internal flexibility in the resulting 6 nt long proximal loop in the case of *cMYC*-1245, probably minimizes the chance for the formation of productive microstates. The separated hairpins are most likely not stable and will further collapse, but the coordination of K⁺ and subsequent tetrad formation will be less effective in this compacted ensemble. This gives a reasonable explanation for the non-Arrhenius temperature dependence of *cMYC*-1245 (5.3). This thermodynamic profile of folding kinetics is expected for hairpin folding.^{760–766} The propensity and stability of a stabilization in a G4-like arrangement, after the initial collapse into two intramolecular hairpins then is a rate-determining folding phase.

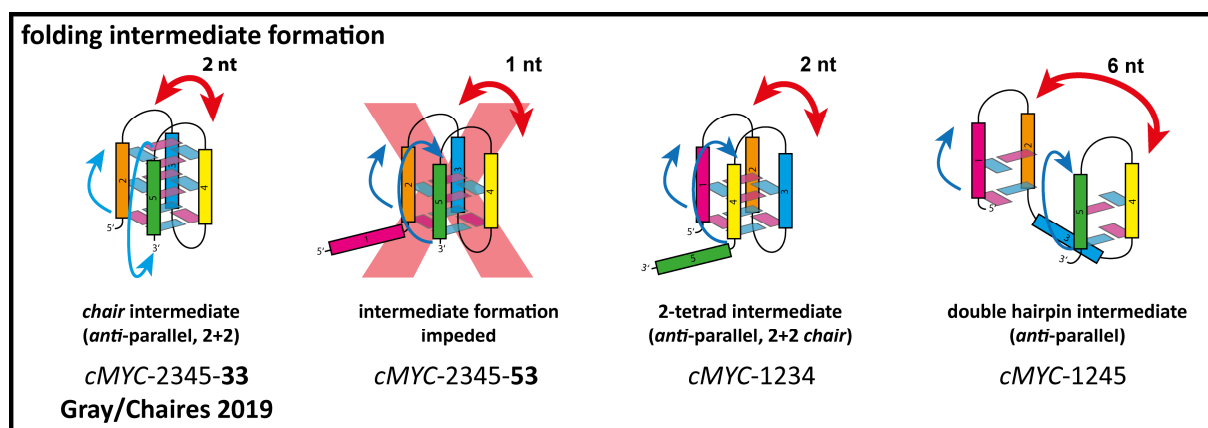


Figure 63: Different proposed folding intermediates and their stability in dependence to the length of the proximal loop. For *cMYC*-1234 and *cMYC*-2345-33, the 2 nt long proximal loop enables the intermediate formation of an *anti*-parallel chair G4. The resulting 1 nt long proximal loop for *cMYC*-2345-53 impedes the formation of defined tetrads in this strand arrangement. In case of *cMYC*-1245, the increased flexibility in the 6 nt long proximal loop separates the terminal hairpins (1-2 and 4-5) and thereby minimizes the probability of productive microstates and K⁺-coordination to form a 2-tetrad G4.

6.3 Refolding Across Different Transitory Ensembles

Refolding of folded G4 conformations has been investigated between *spare-tire* isomers (*cMYC* 1234 and 1245, transition to 2345) and different G-register isomers (between *cMYC*-2345-53 and *cMYC*-2345-33, both parallel; and between *hTERT* hybrid and parallel). The apparent activation energies and refolding kinetics allow proposing diverging mechanism that require different degrees of unfolding in the transitory ensemble. The experimental data can be explained following a simplified depiction of strand rearrangements as a reasonable model. As discussed above, the experimental data support the assumption that the observed apparent activation energies reflect the same rate-limiting step in all investigated refolding processes (chapters 4.5, 4.6.1 and 5.4.2). The subsequent refolding kinetics however are remarkably different, which links to different rearrangement steps.

Formally, the G-register exchange process for the *cMYC* isomers and the *spare-tire* exchange process for the transition from *cMYC* 1234 to 2345 requires only the removal of one G-tract, followed by the re-incorporation of the same, shifted G-tract; or the incorporation of another G-tract as substitute. The kinetic observations for both processes are strikingly similar, which prompts the proposal of a triplex-like strand orientation. For the refolding of the *hTERT* G-register isomers as well as for refolding of the *cMYC* 1245 *spare-tire* isomer, this degree of unfolding is not sufficient for the transition between these conformations. Both pairs of G4 isomers require at least a reorientation of two G-tracts, *hTERT* in addition requires redistribution of *syn/anti* configurations.

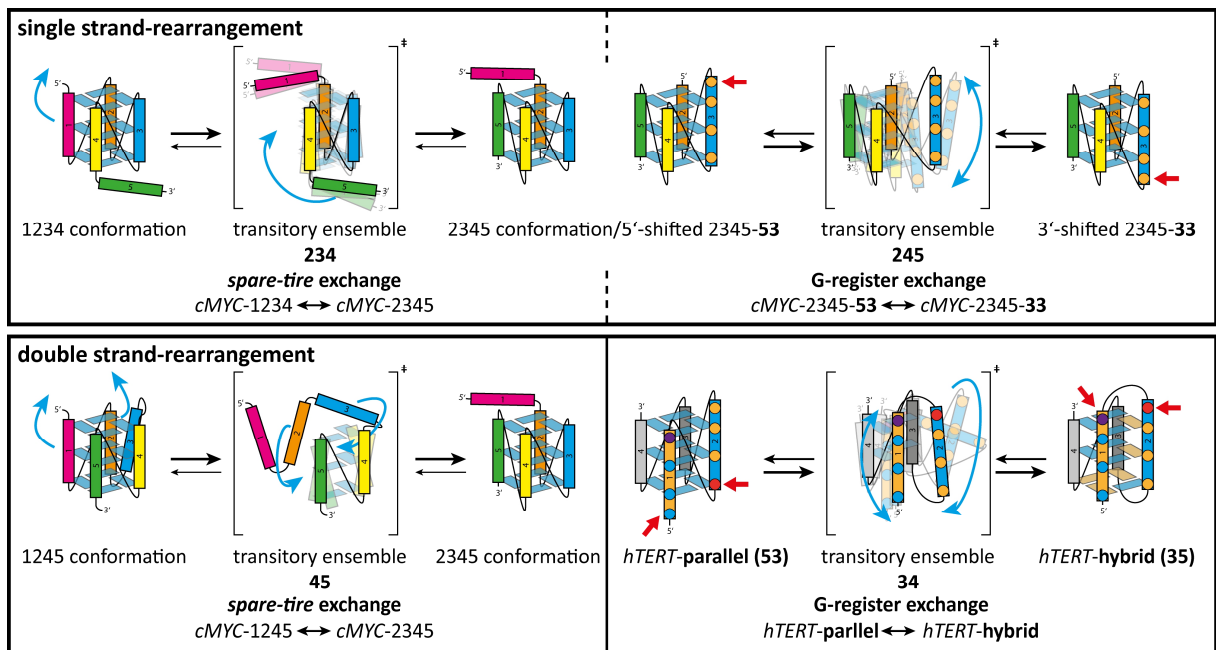


Figure 64: Different proposed species in the transitory ensembles for *spare-tire* exchange and G-register exchange dynamics. The measured refolding kinetics and analysis of the apparent activation energies support the hypothesis that the refolding mechanisms require different degrees of unfolding. A single strand-rearrangement does not necessarily require a complete unfolding-folding transition, but can evolve through compacted ensembles.

The proposed model for the different transitory ensembles is based on the experimental findings within this thesis. It is further supported by predictions from molecular dynamics simulations, on comparable systems.^{157,161–164} However, it cannot possibly account for putatively involved short-lived microstates that emerge from conformational diffusion on the sub- μ s timescale. The real polymer physics of the underlying mechanisms will be certainly more complex.

6.4 Non-Canonical Structural Polymorphism as Challenge

DNA G-quadruplexes are evasive and volatile structural species with an intrinsic dynamic nature. Within this thesis, the conformational dynamics related to the non-canonical G4-structural features of *spare-tire* and **G-register** isomerism have been described in the cMYC promoter G4 ensemble. The literature gives strong indications that these cMYC-G4 dynamics and the polymorphic character of the NHE-III₁ are vital and integral components for their functionality in gene regulation.^{99,340,341,440} The disentanglement of G4 ensembles into distinct folded conformations is however crucial to understand their structure-function relationship and the mechanisms that drive the dynamics and thermodynamic properties within such somehow blurred structural ensembles. Even small changes in the experimental conditions or the primary sequences of G4s have a drastic effect on the structural integrity.⁷⁶⁷

In this regard, the herein developed and presented strategy that utilizes photocages for conformational selection or suppression represents tremendous advantages over hitherto reported methods to investigate G4 folding (2.2.2). The approach was successful to trap and isolated different folded or unfolded states in all presented cases. However, Figure 48 in chapter 4.6.2 (triple caged cMYC-2345-X3 18-mer), shows that the effective inhibition of base pair interactions remains challenging and is not per se straightforward. The photocaging strategy with a (close to) minimal use of photocages has to be adjusted to a given oligonucleotide sequence to prevent pre- or partially folded species.

Chapter 4.2 as well as many examples from the literature^{144,159,160,172} show the challenge to accomplish a complete unfolding in the absence of K⁺, besides the additional considerations that might bias investigation of structural dynamics with altered ionic conditions (chapters 2.2.2, 6.1.1). The presence of pre-folded structures is a crucial aspect that has tremendous effects on the folding pathways and the resulting folded conformations.^{157–160,167} The thermodynamic effects that arise from sequence modifications to tailor single conformations are also not negligible. This can be seen e.g. for a direct comparison of the effects of G-to-T *vs.* G-to-A or G-to-I mutations in the same oligonucleotide sequence (cMYC-G4^{101,460,461}).

The successful application of transient chemical modification to G4 forming oligonucleotides with photocages is therefore remarkable. The time-resolved relaxation processes that have been presented

can be considered mostly unbiased by any of the discussed external or intrinsic factors (ionic conditions, temperature, sequence modifications, conformational entropy) that potentially disturb the folding energy landscape. Restraints to the investigated sequences (with respect to the respective wt) or e.g. buffer conditions ($[K^+]$ below physiological concentrations) have been deliberately introduced. The methodological approach however is universally applicable to diverse oligonucleotide sequences and experimental or sample conditions.

6.5 Model for cMYC G4 (Re)-Folding including Non-Canonical Polymorphism

The experimental data presented in this thesis allow drawing a comprehensive picture of the conformational landscape of the full-length G4 forming sequence found in the NHE-III₁ of the cMYC promoter. (Figure 65). The presented model includes the possible formation of both types of non-canonical polymorphism, which is **G-register** (chapter 4) and **spare-tire** (chapter 5) isomers. The major folding pathways have been delineated, important intermediate and *meta*-stable states have been characterized and key transitory ensembles are proposed.

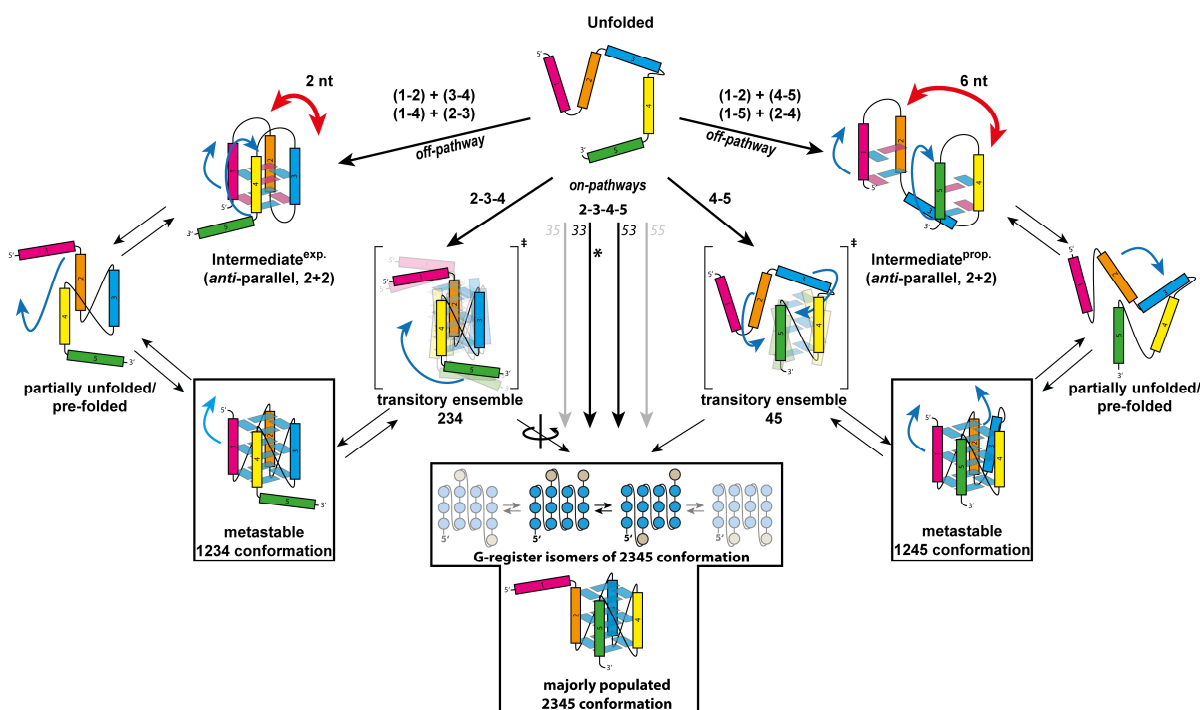


Figure 65: Overview and model for the major folding and refolding pathways in the cMYC NHE-III₁ promoter G4. The model is based on experimental findings presented in this thesis, experimental data from the literature and the above-discussed general considerations. The depicted model accounts for non-canonical structural polymorphism, including *spare-tire* and G-register isomers. (*see additional intermediate formation for cMYC-2345-33 in Figure 63. Folding pathway/mechanism reported by Gray *et al.*¹⁷⁰).

One of the key findings is that all different folding pathways can be occupied, starting from an unfolded state. The major *on*- and *off*-pathways here are defined with regard to the major populated 2345 conformations, including its G-register isomers. This definition might be misleading, since the major

off-pathways lead to the minor conformations 1234 and 1245, which themselves are no misfolded states, but rather alternative conformations.

However, the minor conformations are thermodynamically less stable, leading to refolding towards the major conformation at ambient temperatures. The *on*- and *off*-pathways have different impact on the overall folding kinetics: while a significant population of parallel *off*-pathways causes a deceleration, the parallel *on*-pathways cause an accelerated folding.

To go along these pathways, the early-stage, initial hydrogen bond interactions and orientation of the oligonucleotide backbone must be decisive. Otherwise, these pathways would not be productive by any means, since the kinetics of the pure *on*-pathways are significantly faster. This means that the initial, reversible dynamics (pre-equilibrium) of these pre-folded contacts must be kinetically in the same time range as subsequent steps or slightly slower. The *on*-pathways therefore lead to accelerated folding, since here more contacts between residues that are part of the major conformation may lead to productive macrostates. The irreversible decision (or branching) for any of the four possible G-register isomers might be on a late-stage along the folding pathway. This means that after initial pre-formation (including G-tracts 2-3-4-5) four different isomers share a common route along the folding energy landscape, leading to an acceleration of overall folding. Only in this scenario, the enlarged conformational entropy has this kinetic effect.

6.6 Conclusion and Outlook

Within the scope of this thesis important conformational dynamics related to newly discovered non-canonical structural features of DNA G4s have been described, as discussed and summarized in chapter 6. The presented experimental approach that exploits a photocaging approach for conformational selection and conformational suppression is a major improvement of hitherto reported experimental approaches to measure G4 kinetics. The experimental findings and their presented analysis have shed light on the dynamic features of the polymorphic G4 ensemble in the NHE-III₁ of the *cMYC* promoter. This insight will further deepen the understanding of the pivotal biophysical properties that govern G4 functionality.

In upcoming investigations, it will be highly desirable to elucidate the influence of *cMYC* G4 interacting proteins. Evidence is growing within the literature, that the binding affinities and processing efficiencies of binding proteins and helicases are sensitive to the described conformations with non-canonical polymorphism. Further, the rather slow interconversion dynamics harbour the speculation that there might exist proteins with a refolding chaperoning activity in the cell. Newly developed NMR methods (such as ultrafast 2D-NMR⁶³⁴ in combination with signal enhancement^{639,641,768}) will help to expand the experimental toolkit to tackle this challenge.

References

1. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
2. Vendruscolo, M., Zurdo, J., MacPhee, C. E. & Dobson, C. M. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* **361**, 1205–1222 (2003).
3. Dawkins, R. *River Out Of Eden: A Darwinian View of Life*. (page 133, Basic Books/Weidenfeld&Nicolson, ISBN: 0-297-81540-7, 1995).
4. Dahm, R. Friedrich Miescher and the discovery of DNA. *Developmental Biology* (2005). doi:10.1016/j.ydbio.2004.11.028
5. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human Genetics* **122**, 565–581 (2008).
6. Chargaff, E., Lipshitz, R. & Green, C. Composition of the desoxyribose nucleic acids of four genera of sea-urchin. *J. Biol. Chem.* **195**, 155–60 (1952).
7. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
8. Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxyribose Nucleic Acids. *Nature* **171**, 738–740 (1953).
9. Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740–741 (1953).
10. Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallogr.* (1953). doi:10.1107/s0365110x53001939
11. Marvin, D. A., Spencer, M., Wilkins, M. H. F. & Hamilton, L. D. The molecular configuration of deoxyribonucleic acid III. X-ray diffraction study of the C form of the lithium salt. *J. Mol. Biol.* **3**, 547–IN14 (1961).
12. Mitsui, Y. *et al.* Physical and Enzymatic Studies on Poly d(I–C).Poly d(I–C), an Unusual Double-helical DNA. *Nature* **228**, 1166–1169 (1970).
13. Wang, A. H. J. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686 (1979).
14. Wacker, A. *et al.* Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **48**, 12415–12435 (2020).
15. Wacker, A. *et al.* Structure and dynamics of the deoxyguanosine-sensing riboswitch studied by NMR-spectroscopy. *Nucleic Acids Res.* **39**, 6802–6812 (2011).
16. Niehrs, C. & Luke, B. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat. Rev. Mol. Cell Biol.* **21**, 167–178 (2020).
17. Nasiri, A. H., Wurm, J. P., Immer, C., Weickmann, A. K. & Wöhnert, J. An intermolecular G-quadruplex as the basis for GTP recognition in the class V–GTP aptamer. *RNA* **22**, 1750–1759 (2016).
18. Immer, C., Hacker, C. & Wöhnert, J. Solution structure and RNA-binding of a minimal ProQ-homolog from *Legionella pneumophila* (Lpp1663). *RNA* **26**, 2031–2043 (2020).
19. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
20. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
21. Ellington, A. D. & Szostak, J. W. Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **355**, 850–852 (1992).
22. Green, S. J., Lubrich, D. & Turberfield, A. J. DNA hairpins: Fuel for autonomous DNA devices. *Biophys. J.* (2006). doi:10.1529/biophysj.106.084681
23. Breaker, R. R. DNA enzymes. *Nat. Biotechnol.* **15**, 427–431 (1997).
24. Silverman, S. K. DNA as a Versatile Chemical Component for Catalysis, Encoding, and Stereocontrol. *Angew. Chem. Int. Ed.* **49**, 7180–7201 (2010).
25. Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Höbartner, C. & Pena, V. Crystal structure of a DNA catalyst. *Nature* **529**, 231–234 (2016).
26. Wirmner-Bartoschek, J. & Schwalbe, H. Understanding How DNA Enzymes Work. *Angew. Chem. Int. Ed.* **55**, 5376–5377 (2016).
27. Stojanovic, M. N., de Prada, P. & Landry, D. W. Aptamer-Based Folding Fluorescent Sensor for Cocaine. *J. Am. Chem. Soc.* **123**, 4928–4931 (2001).
28. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* (1992). doi:10.1038/355564a0
29. Heckel, A. & Mayer, G. Light Regulation of Aptamer Activity: An Anti-Thrombin Aptamer with Caged Thymidine Nucleobases. *J. Am. Chem. Soc.* **127**, 822–823 (2005).
30. Thiviyanathan, V. *et al.* Hybrid-hybrid matrix structural refinement of a DNA three-way junction from 3D NOESY-NOESY. *J. Biomol. NMR* **14**, 209–21 (1999).
31. van Buuren, B. N. ., Overmars, F. J. J., Ippel, J. H., Altona, C. & Wijmenga, S. S. Solution structure of a DNA three-way junction containing two unpaired thymidine bases. identification of sequence features that decide conformer selection. *J. Mol. Biol.* **304**, 371–383 (2000).
32. Duckett, D. R. & Lilley, D. M. The three-way DNA junction is a Y-shaped molecule in which there is no helix-helix stacking. *EMBO J.* **9**, 1659–1664 (1990).
33. Assenberg, R. Sequence-dependent folding of DNA three-way junctions. *Nucleic Acids Res.* **30**, 5142–5150 (2002).
34. Liu, Y. & West, S. C. Happy Hollidays: 40th anniversary of the Holliday junction. *Nat. Rev. Mol. Cell Biol.* **5**, 937–944 (2004).
35. Holliday, R. A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282–304 (1964).
36. Haider, S. *et al.* Holliday Junctions Formed from Human Telomeric DNA. *J. Am. Chem. Soc.* **140**, 15366–15374 (2018).
37. Shiba, T., Iwasaki, H., Nakata, A. & Shinagawa, H. SOS-inducible DNA repair proteins, RuvA and RuvB, of *Escherichia coli*: functional interactions between RuvA and RuvB for ATP hydrolysis and renaturation of the cruciform structure in supercoiled DNA. *Proc. Natl. Acad. Sci.* **88**, 8445–8449 (1991).
38. Shlyakhtenko, L. S., Potaman, V. N., Sinden, R. R. & Lyubchenko, Y. L. Structure and dynamics of supercoil-stabilized DNA cruciforms. *J. Mol. Biol.* **280**, 61–72 (1998).
39. Schroth, G. P. & Ho, P. S. Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res.* **23**, 1977–1983 (1995).
40. Brázda, V., Laister, R. C., Jagelská, E. B. & Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **12**, 33 (2011).
41. Zazopoulos, E., Lalli, E., Stocco, D. M. & Sassone-Corsi, P. DNA binding and transcriptional repression by DAX-1 blocks steroidogenesis. *Nature* **390**, 311–315 (1997).

References

42. Bikard, D., Loot, C., Baharoglu, Z. & Mazel, D. Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes. *Microbiol. Mol. Biol. Rev.* **74**, 570–588 (2010).
43. Greider, C. W. Telomeres Do D-Loop–T-Loop Minireview. *Cell* (1999).
44. Cornish-Bowden, A. Abbreviations and symbols for the description of conformations of polynucleotide chains (Recommendations 1982). *Pure Appl. Chem.* **55**, 1273–1280 (1983).
45. Markley, J. L. *et al.* Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J. Biomol. NMR* **12**, 1–23 (1998).
46. Mirkin, S. M. Discovery of alternative DNA structures: a heroic decade (1979–1989). *Front. Biosci.* **13**, 1064–71 (2008).
47. Shing, P. & Carter, M. DNA Structure: Alphabet Soup for the Cellular Soul. in *DNA Replication-Current Advances* (InTech, 2011). doi:10.5772/18536
48. Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **16**, 907–916 (1963).
49. Nikolova, E. N. *et al.* Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498–502 (2011).
50. Watson, J. D. & Crick, F. H. C. The Structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–131 (1953).
51. Yagil, G. Paranemic Structures of DNA and their Role in DNA Unwinding. *Crit. Rev. Biochem. Mol. Biol.* **26**, 475–559 (1991).
52. van Dongen, M. J. *et al.* Structure and mechanism of formation of the H-γ5 isomer of an intramolecular DNA triple helix. *Nat. Struct. Biol.* **6**, 854–9 (1999).
53. Pestov, D. G., Dayn, A., Siyanova EYu, George, D. L. & Mirkin, S. M. H-DNA and Z-DNA in the mouse c-Ki-ras promoter. *Nucleic Acids Res.* **19**, 6527–32 (1991).
54. Kinniburgh, A. J. A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. *Nucleic Acids Res.* **17**, 7771–8 (1989).
55. Radhakrishnan, I. & Patel, D. J. Solution structure of a purine-purine-pyrimidine DNA triplex containing G-GC and T-AT triples. *Structure* **1**, 135–152 (1993).
56. Radhakrishnan, I. & Patel, D. J. Solution structure of a pyrimidine-purine-pyrimidine DNA triplex containing T-AT, C+·GC and G-TA triples. *Structure* **2**, 17–32 (1994).
57. Pandey, S. *et al.* Transcription blockage by stable H-DNA analogs in vitro. *Nucleic Acids Res.* **43**, 6994–7004 (2015).
58. Christensen, J. J., Wrathall, D. P., Izatt, R. M. & Tolman, D. O. Thermodynamics of proton dissociation in dilute aqueous solution. IX. pK, changes in macroenthalpy, and changes in macroentropy values for proton ionization from o-, m-, and p-aminobenzoic acids and their methyl esters at 25 degrees. *J. Phys. Chem.* **71**, 3001–6 (1967).
59. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci.* **48**, 2013–2018 (1962).
60. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366 (1988).
61. Sen, D. & Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344**, 410–4 (1990).
62. Detellier, C. & Laszlo, P. Role of alkali metal and ammonium cations in the self-assembly of the 5'-guanosine monophosphate dianion. *J. Am. Chem. Soc.* **102**, 1135–1141 (1980).
63. Borzo, M., Detellier, C., Laszlo, P. & Paris, A. Proton, sodium-23, and phosphorus-31 NMR studies of the self-assembly of the 5'-guanosine monophosphate dianion in neutral aqueous solution in the presence of sodium cations. *J. Am. Chem. Soc.* **102**, 1124–1134 (1980).
64. Hud, N. V. & Plavec, J. The Role of Cations in Determining Quadruplex Structure and Stability. in *Quadruplex Nucleic Acids* (eds. Neidle, S. & Balasubramanian, S.) 100–130 (2007). doi:10.1039/9781847555298-00100
65. Largy, E., Mergny, J.-L. & Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. in *Metal Ions in Life Sciences* 203–258 (2016). doi:10.1007/978-3-319-21756-7_7
66. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).
67. Zhang, S. *et al.* Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule fluorescent probe. *Nucleic Acids Res.* **46**, 7522–7532 (2018).
68. Di Antonio, M. *et al.* Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **12**, 832–837 (2020).
69. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **6**, 75–80 (2014).
70. Laguerre, A. *et al.* Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.* **137**, 8521–8525 (2015).
71. Liu, H.-H. *et al.* RNA G-quadruplex formation in defined sequence in living cells detected by bimolecular fluorescence complementation. *Chem. Sci.* **7**, 4573–4581 (2016).
72. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371–aaf5371 (2016).
73. Gehring, K., Leroy, J. L. & Guéron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**, 561–565 (1993).
74. Guéron, M. & Leroy, J. L. The i-motif in nucleic acids. *Current Opinion in Structural Biology* **10**, 326–331 (2000).
75. Esmaili, N. & Leroy, J. L. i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats. *Nucleic Acids Res.* **33**, 213–24 (2005).
76. Lieblein, A. L., Buck, J., Schlepckow, K., Fürting, B. & Schwalbe, H. Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning. *Angew. Chem. Int. Ed.* **51**, 250–253 (2012).
77. Lieblein, A. L., Krämer, M., Dreuw, A., Fürting, B. & Schwalbe, H. The Nature of Hydrogen Bonds in Cytidine···H + ···Cytidine DNA Base Pairs. *Angew. Chem. Int. Ed.* **51**, 4067–4070 (2012).
78. Lieblein, A. L., Fürting, B. & Schwalbe, H. Optimizing the Kinetics and Thermodynamics of DNA i-Motif Folding. *ChemBioChem* **14**, 1226–1230 (2013).
79. Day, H. A., Pavlou, P. & Waller, Z. A. E. I-Motif DNA: Structure, stability and targeting with ligands. *Bioorganic Med. Chem.* **22**, 4407–4418 (2014).
80. Lannes, L., Halder, S., Krishnan, Y. & Schwalbe, H. Tuning the pH Response of i-Motif DNA Oligonucleotides. *ChemBioChem* **16**, 1647–1656 (2015).
81. Lannes, L., Young, P., Richter, C., Morgner, N. & Schwalbe, H. Interaction of the N-Terminal Tandem Domains of hnRNP LL with the BCL2 Promoter i-Motif DNA Sequence. *ChemBioChem* **18**, 2033–2044 (2017).

References

82. Wright, E. P., Huppert, J. L. & Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* **45**, 2951–2959 (2017).
83. Abou Assi, H., Garavís, M., González, C. & Damha, M. J. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **46**, 8038–8056 (2018).
84. Zeraati, M. *et al.* I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **10**, 631–637 (2018).
85. Agrawal, P., Hatzakis, E., Guo, K., Carver, M. & Yang, D. Solution structure of the major G-quadruplex formed in the human VEGF promoter in K⁺: Insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.* **41**, 10584–10592 (2013).
86. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
87. Dvorkin, S. A., Karsisiotis, A. I. & Webba da Silva, M. Encoding canonical DNA quadruplex structure. *Sci. Adv.* **4**, eaat3007 (2018).
88. Maity, A., Winnerdy, F. R., Chang, W. D., Chen, G. & Phan, A. T. Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res.* **48**, 3315–3327 (2020).
89. Cui, X. *et al.* Exploration of the Structure and Recognition of a G-quadruplex in the her2 Proto-oncogene Promoter and Its Transcriptional Regulation. *Sci. Rep.* **9**, 3966 (2019).
90. Palumbo, S. L. *et al.* A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res.* **36**, 1755–1769 (2008).
91. Adrian, M. *et al.* Structure and Conformational Dynamics of a Stacked Dimeric G-Quadruplex Formed by the Human CEB1 Minisatellite. *J. Am. Chem. Soc.* **136**, 6297–6305 (2014).
92. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A. & Feigon, J. Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci.* **90**, 3745–3749 (1993).
93. Podbevsek, P., Hud, N. V. & Plavec, J. NMR evaluation of ammonium ion movement within a unimolecular G-quadruplex in solution. *Nucleic Acids Res.* **35**, 2554–2563 (2007).
94. Brcic, J. & Plavec, J. NMR structure of a G-quadruplex formed by four d(G4C2) repeats: insights into structural polymorphism. *Nucleic Acids Res.* **46**, 11605–11617 (2018).
95. Lightfoot, H. L., Hagen, T., Tatum, N. J. & Hall, J. The diverse structural landscape of quadruplexes. *FEBS Lett.* **593**, 2083–2102 (2019).
96. Piazza, A. *et al.* Non-Canonical G-quadruplexes cause the hCEB1 minisatellite instability in *Saccharomyces cerevisiae*. *Elife* **6**, (2017).
97. Varizhuk, A. *et al.* The expanding repertoire of G4 DNA structures. *Biochimie* **135**, 54–62 (2017).
98. Patel, D. J., Phan, A. T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **35**, 7429–7455 (2007).
99. Seenisamy, J. *et al.* The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J. Am. Chem. Soc.* **126**, 8702–8709 (2004).
100. Hatzakis, E., Okamoto, K. & Yang, D. Thermodynamic Stability and Folding Kinetics of the Major G-Quadruplex and Its Loop Isomers Formed in the Nuclease Hypersensitive Element in the Human c-Myc Promoter: Effect of Loops and Flanking Segments on the Stability of Parallel-Stranded Intramolecul. *Biochemistry* **49**, 9152–9160 (2010).
101. Harkness, R. W. & Mittermaier, A. K. G-register exchange dynamics in guanine quadruplexes. *Nucleic Acids Res.* **44**, 3481–3494 (2016).
102. Lim, K. W. *et al.* Coexistence of Two Distinct G-Quadruplex Conformations in the hTERT Promoter. *J. Am. Chem. Soc.* **132**, 12331–12342 (2010).
103. Monsen, R. C. *et al.* The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res.* **48**, 5720–5734 (2020).
104. Cheung, I., Schertzer, M., Rose, A. & Lansdorp, P. M. Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat. Genet.* **31**, 405–409 (2002).
105. Zhao, Y., O'Neil, N. J. & Rose, A. M. Poly-G/poly-C tracts in the genomes of *Caenorhabditis*. *BMC Genomics* **8**, 403 (2007).
106. Sengar, A., Heddi, B. & Phan, A. T. Formation of G-Quadruplexes in Poly-G Sequences: Structure of a Propeller-Type Parallel-Stranded G-Quadruplex Formed by a G 15 Stretch. *Biochemistry* **53**, 7718–7723 (2014).
107. Li, X. *et al.* Guanine-vacancy-bearing G-quadruplexes responsive to guanine derivatives. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14581–6 (2015).
108. Heddi, B., Martín-Pintado, N., Serimbetov, Z., Mahfuzh Aufar Kari, T. & Tù an Phan, A. G-quadruplexes with (4n-1) guanines in the G-tetrad core: formation of a G-triad-water complex and implication for small-molecule binding. *Nucleic Acids Res.* **44**, 910–916 (2016).
109. Wang, K.-B., Dickerhoff, J., Wu, G. & Yang, D. PDGFR- β Promoter Forms a Vacancy G-Quadruplex that Can Be Filled in by dGMP: Solution Structure and Molecular Recognition of Guanine Metabolites and Drugs. *J. Am. Chem. Soc.* **142**, 5204–5211 (2020).
110. He, Y. *et al.* Selective Targeting of Guanine-Vacancy-Bearing G-Quadruplexes by G-Quartet Complementation and Stabilization with a Guanine–Peptide Conjugate. *J. Am. Chem. Soc.* **142**, 11394–11403 (2020).
111. Phan, A. T., Modi, Y. S. & Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J. Am. Chem. Soc.* **126**, 8710–8716 (2004).
112. Jana, J. & Weisz, K. A Thermodynamic Perspective on Potential G-Quadruplex Structures as Silencer Elements in the MYC Promoter. *Chem. – A Eur. J.* **26**, 17242–17251 (2020).
113. Fleming, A. M., Zhou, J., Wallace, S. S. & Burrows, C. J. A Role for the Fifth G-Track in G-Quadruplex Forming Oncogene Promoter Sequences during Oxidative Stress: Do These “Spare Tires” Have an Evolved Function? *ACS Cent. Sci.* **1**, 226–233 (2015).
114. Do, N. Q. & Phan, A. T. Monomer-Dimer Equilibrium for the 5'-5' Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study. *Chem. Eur. J.* **18**, 14752–14759 (2012).
115. Palumbo, S. L., Ebbinghaus, S. W. & Hurley, L. H. Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the hTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **131**, 10878–10891 (2009).
116. Chaires, J. B. *et al.* An Improved Model for the hTERT Promoter Quadruplex. *PLoS One* **9**, e115580 (2014).
117. Song, J. H. *et al.* Small-Molecule-Targeting Hairpin Loop of hTERT Promoter G-Quadruplex Induces Cancer Cell Death. *Cell Chem. Biol.* **26**, 1110–1121.e4 (2019).
118. Smargiasso, N. *et al.* G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.* **130**, 10208–16 (2008).
119. Marsh, T. C., Vesenka, J. & Henderson, E. A new DNA nanostructure, the G-wire, imaged by scanning probe microscopy. *Nucleic Acids Res.* **23**, 696–700 (1995).
120. Marsh, T. C. & Henderson, E. G-Wires: Self-Assembly of a Telomeric Oligonucleotide, d(GGGGTTGGGG), into Large Superstructures. *Biochemistry* **33**, 10718–10724 (1994).

References

121. Bose, K., Lech, C. J., Heddi, B. & Phan, A. T. High-resolution AFM structure of DNA G-wires in aqueous solution. *Nat. Commun.* **9**, 1959 (2018).
122. Mukundan, V. T. & Phan, A. T. Bulges in G-quadruplexes: Broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* **135**, 5017–5028 (2013).
123. Meier, M. *et al.* Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res.* **46**, 5319–5331 (2018).
124. Lenarčič Živković, M., Rozman, J. & Plavec, J. Structure of a DNA G-Quadruplex Related to Osteoporosis with a G-A Bulge Forming a Pseudoloop. *Molecules* **25**, 4867 (2020).
125. Das, P. *et al.* Bulges in left-handed G-quadruplexes. *Nucleic Acids Res.* **49**, 1724–1736 (2021).
126. Onel, B. *et al.* A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription. *J. Am. Chem. Soc.* **138**, 2563–2570 (2016).
127. Ngoc Nguyen, T. Q., Lim, K. W. & Phan, A. T. Duplex formation in a G-quadruplex bulge. *Nucleic Acids Res.* **48**, 10567–10575 (2020).
128. Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **1**, 167–173 (2005).
129. Schnarr, L., Jana, J., Preckwinkel, P. & Weisz, K. Impact of a Snap-Back Loop on Stability and Ligand Binding to a Parallel G-Quadruplex. *J. Phys. Chem. B* **124**, 2778–2787 (2020).
130. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D. J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter. *J. Am. Chem. Soc.* **129**, 4386–4392 (2007).
131. Chen, Y. *et al.* The major G-quadruplex formed in the human platelet-derived growth factor receptor β promoter adopts a novel broken-strand structure in K⁺ solution. *J. Am. Chem. Soc.* **134**, 13220–3 (2012).
132. Chung, W. J. *et al.* Structure of a left-handed DNA G-quadruplex. *Proc. Natl. Acad. Sci.* **112**, 2729–2733 (2015).
133. Fu, B. *et al.* Right-handed and left-handed G-quadruplexes have the same DNA sequence: distinct conformations induced by an organic small molecule and potassium. *Chem. Commun.* **52**, 10052–10055 (2016).
134. Bakalar, B., Heddi, B., Schmitt, E., Mechulam, Y. & Phan, A. T. A Minimal Sequence for Left-Handed G-Quadruplex Formation. *Angew. Chem. Int. Ed.* **58**, 2331–2335 (2019).
135. Winnerdy, F. R. *et al.* NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res.* **47**, 8272–8281 (2019).
136. Escaja, N., Gómez-Pinto, I., Pedroso, E. & González, C. Four-Stranded DNA Structures Can Be Stabilized by Two Different Types of Minor Groove G:C:G:C Tetrads. *J. Am. Chem. Soc.* **129**, 2004–2014 (2007).
137. Kocman, V. & Plavec, J. A tetrahelical DNA fold adopted by tandem repeats of alternating GGG and GCG tracts. *Nat. Commun.* **5**, 5831 (2014).
138. Kocman, V. & Plavec, J. Tetrahelical structural family adopted by AGCGA-rich regulatory DNA regions. *Nat. Commun.* **8**, 15355 (2017).
139. Hadži, S., Kocman, V., Oblak, D., Plavec, J. & Lah, J. Energetic Basis of AGCGA-Rich DNA Folding into a Tetrahelical Structure. *Angew. Chem. Int. Ed.* **58**, 2387–2391 (2019).
140. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
141. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200 (2014).
142. Harkness, R. W. & Mittermaier, A. K. G-quadruplex dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **1865**, 1544–1554 (2017).
143. Phan, A. T. & Patel, D. J. Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *J. Am. Chem. Soc.* **125**, 15021–15027 (2003).
144. Bessi, I., Jonker, H. R. A., Richter, C. & Schwalbe, H. Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex. *Angew. Chem. - Int. Ed.* **54**, 8444–8448 (2015).
145. Nufßbaumer, F. *et al.* Synthesis and incorporation of ¹³C-labeled DNA building blocks to probe structural dynamics of DNA by NMR. *Nucleic Acids Res.* **45**, 9178–9192 (2017).
146. Oh, K.-L., Kim, J., Park, C.-J. & Lee, J.-H. Dynamics Studies of DNA with Non-canonical Structure Using NMR Spectroscopy. *Int. J. Mol. Sci.* **21**, 2673 (2020).
147. Voter, A. F., Qiu, Y., Tippiana, R., Myong, S. & Keck, J. L. A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases. *Nat. Commun.* **9**, 4201 (2018).
148. Cahoon, L. A. & Seifert, H. S. An Alternative DNA Structure Is Necessary for Pilin Antigenic Variation in *Neisseria gonorrhoeae*. *Science* **325**, 764–767 (2009).
149. Portella, G., Orozco, M. & Vendruscolo, M. Determination of a Structural Ensemble Representing the Dynamics of a G-Quadruplex DNA. *Biochemistry* **59**, 379–388 (2020).
150. Carrino, S., Hennecker, C. D., Murrieta, A. C. & Mittermaier, A. Frustrated folding of guanine quadruplexes in telomeric DNA. *Nucleic Acids Res.* published online (2021). doi:10.1093/nar/gkab140
151. Levinthal, C. How to fold graciously. *Mössbauer Spectrosc. Biol. Syst. Proc.* **67**, 22–26 (1969).
152. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).
153. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Mol. Biol.* **4**, 10–19 (1997).
154. Englander, S. W. & Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci.* **111**, 15873–15880 (2014).
155. Callaway, E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* **588**, 203–204 (2020).
156. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. The Kinetics Of Formation Of Native Ribonuclease During Oxidation Of The Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci.* **47**, 1309–1314 (1961).
157. Šponer, J. *et al.* Folding of guanine quadruplex molecules-funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochim. Biophys. acta. Gen. Subj.* **1861**, 1246–1263 (2017).
158. Kuo, M. H.-J. *et al.* Conformational Transition of a Hairpin Structure to G-Quadruplex within the WNT1 Gene Promoter. *J. Am. Chem. Soc.* **137**, 210–218 (2015).
159. Freljih, T., Wang, B., Plavec, J. & Šket, P. Pre-folded structures govern folding pathways of human telomeric G-quadruplexes. *Nucleic Acids Res.* **48**, 2189–2197 (2020).
160. Kotar, A., Rigo, R., Sissi, C. & Plavec, J. Two-quartet kit* G-quadruplex is formed via double-stranded pre-folded structure. *Nucleic Acids Res.* **47**, 2641–2653 (2019).
161. Stadlbauer, P., Krepl, M., Cheatham, T. E., Koca, J. & Sponer, J. Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res.* **41**, 7128–7143 (2013).

References

162. Stadlbauer, P., Trantířek, L., Cheatham, T. E., Koča, J. & Šponer, J. Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie* **105**, 22–35 (2014).
163. Stadlbauer, P. *et al.* Parallel G-triplexes and G-hairpins as potential transitory ensembles in the folding of parallel-stranded DNA G-Quadruplexes. *Nucleic Acids Res.* **47**, 7276–7293 (2019).
164. Stadlbauer, P. *et al.* Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.* **43**, 9626–44 (2015).
165. Gajarský, M. *et al.* Structure of a Stable G-Hairpin. *J. Am. Chem. Soc.* **139**, 3591–3594 (2017).
166. Živković, M. L. *et al.* Insight into formation propensity of pseudocircular DNA G-hairpins. *Nucleic Acids Res.* **49**, 2317–2332 (2021).
167. Marchand, A. & Erie Gabelica, V. Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res.* **44**, 10999–11012 (2016).
168. Gabelica, V. A pilgrim's guide to G-quadruplex nucleic acid folding. *Biochimie* **105**, 1–3 (2014).
169. Gray, R. D., Trent, J. O. & Chaires, J. B. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *J. Mol. Biol.* **426**, 1629–1650 (2014).
170. Gray, R. D., Trent, J. O., Arumugam, S. & Chaires, J. B. Folding Landscape of a Parallel G-Quadruplex. *J. Phys. Chem. Lett.* **10**, 1146–1151 (2019).
171. Zhang, A. Y. Q. & Balasubramanian, S. The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *J. Am. Chem. Soc.* **134**, 19297–19308 (2012).
172. Müller, D., Bessi, I., Richter, C. & Schwalbe, H. The folding landscapes of human telomeric RNA and DNA G-quadruplexes are markedly different. *Angew. Chem. Int. Ed.* **ange.202100280** (2021). doi:10.1002/ange.202100280
173. Aznauryan, M., Søndergaard, S., Noer, S. L., Schiøtt, B. & Birkedal, V. A direct view of the complex multi-pathway folding of telomeric G-quadruplexes. *Nucleic Acids Res.* **44**, 11024–11032 (2016).
174. Noer, S. L. *et al.* Folding dynamics and conformational heterogeneity of human telomeric G-quadruplex structures in Na⁺ solutions by single molecule FRET microscopy. *Nucleic Acids Res.* **44**, 464–71 (2016).
175. Li, W., Hou, X.-M., Wang, P.-Y., Xi, X.-G. & Li, M. Direct Measurement of Sequential Folding Pathway and Energy Landscape of Human Telomeric G-quadruplex Structures. *J. Am. Chem. Soc.* **135**, 6423–6426 (2013).
176. Cheng, Y. *et al.* Folding/unfolding kinetics of G-quadruplexes upstream of the P1 promoter of the human BCL-2 oncogene. *J. Biol. Chem.* **294**, 5890–5895 (2019).
177. Gray, R. D. & Chaires, J. B. Isothermal folding of G-quadruplexes. *Methods* **57**, 47–55 (2012).
178. Mergny, J.-L. & Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **13**, 515–537 (2003).
179. Brown, N. M., Rachwal, P. A., Brown, T. & Fox, K. R. Exceptionally slow kinetics of the intramolecular quadruplex formed by the Oxytricha telomeric repeat. *Org. Biomol. Chem.* **3**, 4153–7 (2005).
180. Green, J. J., Ying, L., Klenerman, D. & Balasubramanian, S. Kinetics of unfolding the human telomeric DNA quadruplex using a PNA trap. *J. Am. Chem. Soc.* **125**, 3763–7 (2003).
181. Risitano, A. & Fox, K. R. Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry* **42**, 6507–13 (2003).
182. Lane, A. N. The stability of intramolecular DNA G-quadruplexes compared with other macromolecules. *Biochimie* **94**, 277–86 (2012).
183. Spoerner, M. *et al.* High pressure 31P NMR spectroscopy on guanine nucleotides. *J. Biomol. NMR* **67**, 1–13 (2017).
184. Munte, C. E. *et al.* High pressure response of 1H NMR chemical shifts of purine nucleotides. *Biophys. Chem.* **254**, 106261 (2019).
185. You, H., Wu, J., Shao, F. & Yan, J. Stability and Kinetics of c- MYC Promoter G-Quadruplexes Studied by Single-Molecule Manipulation. *J. Am. Chem. Soc.* **137**, 2424–2427 (2015).
186. Bugaut, A., Murat, P. & Balasubramanian, S. An RNA Hairpin to G-Quadruplex Conformational Transition. *J. Am. Chem. Soc.* **134**, 19953–19956 (2012).
187. Wang, Z.-F. *et al.* Cytosine epigenetic modification modulates the formation of an unprecedented G4 structure in the WNT1 promoter. *Nucleic Acids Res.* **48**, 1120–1130 (2020).
188. Lu, X.-M. *et al.* Folding Dynamics of Parallel and Antiparallel G-Triplexes under the Influence of Proximal DNA. *J. Phys. Chem. B* **122**, 9499–9506 (2018).
189. Søndergaard, S. *et al.* Dynamics of Fluorescent Dyes Attached to G-Quadruplex DNA and their Effect on FRET Experiments. *ChemPhysChem* **16**, 2562–2570 (2015).
190. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4**, (2016).
191. You, J. *et al.* Effects of monovalent cations on folding kinetics of G-quadruplexes. *Biosci. Rep.* **37**, (2017).
192. Nguyen, T. Q. N., Lim, K. W. & Phan, A. T. Folding Kinetics of G-Quadruplexes: Duplex Stem Loops Drive and Accelerate G-Quadruplex Folding. *J. Phys. Chem. B* **124**, 5122–5130 (2020).
193. Hou, X.-M. *et al.* Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* **45**, 11401–11412 (2017).
194. Cerofolini, L. *et al.* G-triplex structure and formation propensity. *Nucleic Acids Res.* **42**, 13393–13404 (2014).
195. Mashimo, T., Yagi, H., Sannohe, Y., Rajendran, A. & Sugiyama, H. Folding Pathways of Human Telomeric Type-1 and Type-2 G-Quadruplex Structures. *J. Am. Chem. Soc.* **132**, 14910–14918 (2010).
196. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **36**, 5482–5515 (2008).
197. Liu, W. *et al.* Kinetics and Mechanism of Conformational Changes in a G-Quadruplex of Thrombin-Binding Aptamer Induced by Pb²⁺. *J. Phys. Chem. B* **115**, 13051–13056 (2011).
198. Liu, W. *et al.* Kinetics and mechanism of G-quadruplex formation and conformational switch in a G-quadruplex of PS2.M induced by Pb²⁺. *Nucleic Acids Res.* **40**, 4229–4236 (2012).
199. Long, X. & Stone, M. D. Kinetic Partitioning Modulates Human Telomere DNA G-Quadruplex Structural Polymorphism. *PLoS One* **8**, e83420 (2013).
200. Gray, R. D. & Chaires, J. B. Kinetics and mechanism of K⁺- and Na⁺-induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res.* **36**, 4191–4203 (2008).
201. Xue, Y. *et al.* Kinetic and thermodynamic control of G-quadruplex folding. *Angew. Chem. Int. Ed.* **50**, 8046–50 (2011).
202. Bian, Y. *et al.* Atomistic Picture for the Folding Pathway of a Hybrid-1 Type Human Telomeric DNA G-quadruplex. *PLoS Comput. Biol.* **10**, e1003562 (2014).
203. Rocca, R. *et al.* Folding intermediate states of the parallel human telomeric G-quadruplex DNA explored using Well-Tempered Metadynamics. *Sci. Rep.* **10**, 3176 (2020).

References

204. Bian, Y. *et al.* Insights into the Kinetic Partitioning Folding Dynamics of the Human Telomeric G-Quadruplex from Molecular Simulations and Machine Learning. *J. Chem. Theory Comput.* **16**, 5936–5947 (2020).
205. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1**, 263–282 (1993).
206. Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876–880 (2002).
207. Dai, J., Carver, M., Punchihewa, C., Jones, R. A. & Yang, D. 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 (2007).
208. Dai, J. *et al.* Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.* **35**, 2440–2450 (2007).
209. Phan, A. T., Kuryavii, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res.* **35**, 6517–25 (2007).
210. Luu, K. N., Phan, A. T., Kuryavii, V., Lacroix, L. & Patel, D. J. Structure of the Human Telomere in K⁺ Solution: An Intramolecular (3+1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **128**, 9963–9970 (2006).
211. Isambert, H. The jerky and knotty dynamics of RNA. *Methods* **49**, 189–196 (2009).
212. Kolesnikova, S., Hubálek, M., Bednářová, L., Cvačka, J. & Curtis, E. A. Multimerization rules for G-quadruplexes. *Nucleic Acids Res.* **45**, 8684–8696 (2017).
213. Mergny, J.-L. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.* **33**, 81–94 (2005).
214. Bardin, C. & Leroy, J. L. The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.* **36**, 477–488 (2008).
215. Merkina, E. E. & Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophys. J.* **89**, 365–373 (2005).
216. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **126**, 16405–16415 (2004).
217. Agrawal, P., Lin, C., Mathad, R. I., Carver, M. & Yang, D. The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K⁺ solution. *J. Am. Chem. Soc.* **136**, 1750–3 (2014).
218. Greco, M. L. *et al.* Coexistence of two main folded G-quadruplexes within a single G-rich domain in the EGFR promoter. *Nucleic Acids Res.* **45**, 10132–10142 (2017).
219. Balkwill, G. D., Garner, T. P. & Searle, M. S. Folding of single-stranded DNA quadruplexes containing an autonomously stable mini-hairpin loop. *Mol. Biosyst.* **5**, 542–7 (2009).
220. Zhang, D.-H. *et al.* Monomorphous RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **49**, 4554–4563 (2010).
221. Tang, C.-F. & Shafer, R. H. Engineering the Quadruplex Fold: Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine Quadruplexes. *J. Am. Chem. Soc.* **128**, 5966–5973 (2006).
222. Małgowska, M. Overview of RNA G-quadruplex structures. *Acta Biochim. Pol.* **63**, (2017).
223. Mendoza, O., Porrini, M., Salgado, G. F., Gabelica, V. & Mergny, J.-L. Orienting Tetramolecular G-Quadruplex Formation: The Quest for the Elusive RNA Antiparallel Quadruplex. *Chem. Eur. J.* **21**, 6732–6739 (2015).
224. Xiao, C.-D., Ishizuka, T. & Xu, Y. Antiparallel RNA G-quadruplex Formed by Human Telomere RNA Containing 8-Bromoguanosine. *Sci. Rep.* **7**, 6695 (2017).
225. Karg, B., Haase, L., Funke, A., Dickerhoff, J. & Weisz, K. Observation of a Dynamic G-Tetrad Flip in Intramolecular G-Quadruplexes. *Biochemistry* **55**, 6949–6955 (2016).
226. Hirashima, K. & Seimiya, H. Telomeric repeat-containing RNA/G-quadruplex-forming sequences cause genome-wide alteration of gene expression in human cancer cells in vivo. *Nucleic Acids Res.* **43**, 2022–2032 (2015).
227. Martadinata, H. & Phan, A. T. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K⁺ solution. *J. Am. Chem. Soc.* (2009). doi:10.1021/ja806592z
228. Phan, A. T. Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **277**, 1107–17 (2010).
229. Collie, G. W. *et al.* Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.* **132**, 9328–34 (2010).
230. Martadinata, H., Heddi, B., Lim, K. W. & Phan, A. T. Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry* **50**, 6455–61 (2011).
231. Martadinata, H. & Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K⁺ Solution. *Biochemistry* **52**, 2176–2183 (2013).
232. Dhakal, S. *et al.* G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophys. J.* **102**, 2575–2584 (2012).
233. Cui, Y., Kong, D., Ghimire, C., Xu, C. & Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **55**, 2291–2299 (2016).
234. King, J. J. *et al.* DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J. Am. Chem. Soc.* **142**, 20600–20604 (2020).
235. Sengupta, P., Bose, D. & Chatterjee, S. The Molecular Tête-à-Tête between G-Quadruplexes and the i-motif in the Human Genome. *ChemBioChem* **cbic.202000703** (2021). doi:10.1002/cbic.202000703
236. Kang, H.-J., Kendrick, S., Hecht, S. M. & Hurley, L. H. The Transcriptional Complex Between the BCL2 i-Motif and hnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. *J. Am. Chem. Soc.* **136**, 4172–4185 (2014).
237. Kaiser, C. E. *et al.* Insight into the Complexity of the i-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* **139**, 8522–8536 (2017).
238. Kendrick, S. *et al.* The Dynamic Character of the BCL2 Promoter i-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J. Am. Chem. Soc.* **136**, 4161–4171 (2014).
239. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **277**, 3459–3469 (2010).
240. Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III 1. *J. Am. Chem. Soc.* **138**, 14138–14151 (2016).
241. Dai, J., Hatzakis, E., Hurley, L. H. & Yang, D. I-Motif Structures Formed in the Human c-MYC Promoter Are Highly Dynamic—Insights into Sequence Redundancy and I-Motif Stability. *PLoS One* **5**, e11647 (2010).
242. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–37 (2015).
243. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–16 (2005).
244. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406–413 (2007).

References

245. Du, Z., Zhao, Y. & Li, N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.* **18**, 233–241 (2008).
246. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18**, 279–284 (2017).
247. Maizels, N. & Gray, L. T. The G4 Genome. *PLoS Genet.* **9**, e1003468 (2013).
248. Todd, A. K., Johnston, M. & Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901–2907 (2005).
249. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–3896 (2006).
250. Vallur, A. C. & Maizels, N. Activities of human exonuclease 1 that promote cleavage of transcribed immunoglobulin switch regions. *Proc. Natl. Acad. Sci.* **105**, 16508–16512 (2008).
251. Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **13**, 1055–1059 (2006).
252. Moyzis, R. K. *et al.* A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci.* **85**, 6622–6626 (1988).
253. Ribeyre, C. *et al.* The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *PLoS Genet.* **5**, e1000475 (2009).
254. Fleming, A. M. *et al.* Human DNA Repair Genes Possess Potential G-Quadruplex Sequences in Their Promoters and 5'-Untranslated Regions. *Biochemistry* **57**, 991–1002 (2018).
255. Mao, S.-Q. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951–957 (2018).
256. Guillbaud, G. *et al.* Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat. Chem.* **9**, 1110–1117 (2017).
257. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
258. Reina, C. & Cavalieri, V. Epigenetic Modulation of Chromatin States and Gene Expression by G-Quadruplex Structures. *Int. J. Mol. Sci.* **21**, 4172 (2020).
259. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474 (2020).
260. Bedrat, A., Lacroix, L. & Mergny, J.-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
261. Brázda, V. *et al.* G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics* **35**, 3493–3495 (2019).
262. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
263. Kikin, O., D'Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682 (2006).
264. Piazza, A. *et al.* Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **38**, 4337–4348 (2010).
265. Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30**, 4033–4046 (2011).
266. Piazza, A. *et al.* Stimulation of Gross Chromosomal Rearrangements by the Human CEB1 and CEB25 Minisatellites in *Saccharomyces cerevisiae* Depends on G-Quadruplexes or Cdc13. *PLoS Genet.* **8**, e1003033 (2012).
267. Piazza, A. *et al.* Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.* **34**, 1718–1734 (2015).
268. Chai, W., Du, Q., Shay, J. W. & Wright, W. E. Human telomeres have different overhang sizes at leading versus lagging strands. *Mol. Cell* **21**, 427–435 (2006).
269. Lu, W., Zhang, Y., Liu, D., Songyang, Z. & Wan, M. Telomeres-structure, function, and regulation. *Experimental Cell Research* **319**, 133–141 (2013).
270. Morin, G. B. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**, 521–529 (1989).
271. Sandin, S. & Rhodes, D. Telomerase structure. *Current Opinion in Structural Biology* **25**, 104–110 (2014).
272. Sfeir, A. J., Chai, W., Shay, J. W. & Wright, W. E. Telomere-end processing the terminal nucleotides of human chromosomes. *Mol. Cell* **18**, 131–8 (2005).
273. De Lange, T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes and Development* **19**, 2100–2110 (2005).
274. De Cian, A. *et al.* Targeting telomeres and telomerase. *Biochimie* **90**, 131–55 (2008).
275. Hockemeyer, D. & Collins, K. Control of telomerase action at human telomeres. *Nat. Struct. Mol. Biol.* **22**, 848–52 (2015).
276. Wheelhouse, R. T., Sun, D., Han, H., Han, F. X. & Hurley, L. H. Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA. *J. Am. Chem. Soc.* **120**, 3261–3262 (1998).
277. Reed, J. E., Arnal, A. A., Neidle, S. & Vilar, R. Stabilization of G-quadruplex DNA and inhibition of telomerase activity by square-planar nickel(II) complexes. *J. Am. Chem. Soc.* **128**, 5992–5993 (2006).
278. Read, M. A. *et al.* Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: Structure-activity relationships. *J. Med. Chem.* **42**, 4538–4546 (1999).
279. Ou, T. *et al.* G-quadruplexes: targets in anticancer drug design. *ChemMedChem* **3**, 690–713 (2008).
280. Kim, M.-Y., Vankayalapati, H., Shin-Ya, K., Wierzbica, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular g-quadruplex. *J. Am. Chem. Soc.* **124**, 2098–9 (2002).
281. Bazzicalupi, C., Ferraroni, M., Bilia, A. R., Scheggi, F. & Gratteri, P. The crystal structure of human telomeric DNA complexed with berberine: An interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res.* **41**, 632–638 (2013).
282. Ohnmacht, S. A. & Neidle, S. Small-molecule quadruplex-targeted drug discovery. *Bioorganic Med. Chem. Lett.* **24**, 2602–2612 (2014).
283. McLuckie, K. I. E. *et al.* G-quadruplex DNA as a molecular target for induced synthetic lethality in cancer cells. *J. Am. Chem. Soc.* **135**, 9640–3 (2013).
284. Sun, D. *et al.* Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.* **40**, 2113–2116 (1997).
285. Oganessian, L. & Bryan, T. M. Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *BioEssays* **29**, 155–165 (2007).
286. Incles, C. M. *et al.* A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Mol. Cancer Ther.* **3**, 1201–6 (2004).
287. Tauchi, T. *et al.* Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene* **22**, 5338–5347 (2003).

References

288. Salvati, E. *et al.* Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Invest.* **117**, 3236–3247 (2007).
289. Harrison, R. J., Gowan, S. M., Kelland, L. R. & Neidle, S. Human telomerase inhibition by substituted acridine derivatives. *Bioorg. Med. Chem. Lett.* **9**, 2463–2468 (1999).
290. Bryan, T. M. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules* **25**, 3686 (2020).
291. Carvalho, J., Mergny, J.-L., Salgado, G. F., Queiroz, J. A. & Cruz, C. G-quadruplex, Friend or Foe: The Role of the G-quartet in Anticancer Strategies. *Trends Mol. Med.* **26**, 848–861 (2020).
292. Nakanishi, C. & Seimiya, H. G-quadruplex in cancer biology and drug discovery. *Biochem. Biophys. Res. Commun.* **531**, 45–50 (2020).
293. Zhao, J. & Zhai, Q. Recent advances in the development of ligands specifically targeting telomeric multimeric G-quadruplexes. *Bioorg. Chem.* **103**, 104229 (2020).
294. Hänsel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **13**, 551–564 (2018).
295. Hänsel-Hertsch, R. *et al.* Landscape of G-quadruplex DNA structural regions in breast cancer. *Nat. Genet.* **52**, 878–883 (2020).
296. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci.* **98**, 8572–8577 (2001).
297. Paeschke, K. *et al.* Telomerase recruitment by the telomere end binding protein- β facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.* **15**, 598–604 (2008).
298. Henderson, A. *et al.* Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* **42**, 860–869 (2014).
299. Liu, H.-Y. *et al.* Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.* **23**, 1261–1270 (2016).
300. Hänsel, R. *et al.* Evaluation of Parameters Critical for Observing Nucleic Acids Inside Living *Xenopus laevis* Oocytes by In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **131**, 15761–15768 (2009).
301. Hänsel, R. *et al.* The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. *Nucleic Acids Res.* **39**, 5768–5775 (2011).
302. Hänsel, R., Foldynová-Trantířková, S., Dötsch, V. & Trantířek, L. Investigation of quadruplex structure under physiological conditions using in-cell NMR. *Topics in Current Chemistry* **330**, 47–66 (2013).
303. Hänsel, R., Löhr, F., Trantířek, L. & Dötsch, V. High-resolution insight into G-overhang architecture. *J. Am. Chem. Soc.* **135**, 2816–24 (2013).
304. Hänsel, R., Luh, L. M., Corbeski, I., Trantířek, L. & Dötsch, V. In-cell NMR and EPR spectroscopy of biomacromolecules. *Angew. Chem. Int. Ed.* **53**, 10300–14 (2014).
305. Salgado, G. F., Cazenave, C., Kerkour, A. & Mergny, J.-L. G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* **6**, 3314–3320 (2015).
306. Yamaoki, Y., Nagata, T., Sakamoto, T. & Katahira, M. Recent progress of in-cell NMR of nucleic acids in living human cells. *Biophys. Rev.* **12**, 411–417 (2020).
307. Bao, H.-L. *et al.* Characterization of human telomere RNA G-quadruplex structures in vitro and in living cells using 19F NMR spectroscopy. *Nucleic Acids Res.* **45**, 5501–5511 (2017).
308. Bao, H.-L. & Xu, Y. Investigation of higher-order RNA G-quadruplex structures in vitro and in living cells by 19F NMR spectroscopy. *Nat. Protoc.* **13**, 652–665 (2018).
309. Bao, H.-L., Liu, H. & Xu, Y. Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res.* **47**, 4940–4947 (2019).
310. Adrian, M., Heddi, B. & Phan, A. T. NMR spectroscopy of G-quadruplexes. *Methods* **57**, 11–24 (2012).
311. Chilka, P., Desai, N. & Datta, B. Small Molecule Fluorescent Probes for G-Quadruplex Visualization as Potential Cancer Theranostic Agents. *Molecules* **24**, 752 (2019).
312. Huang, W.-C. *et al.* Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.* **43**, 10102–13 (2015).
313. Laguerre, A., Wong, J. M. Y. & Monchard, D. Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci. Rep.* **6**, 32141 (2016).
314. Summers, P. A. *et al.* Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* **12**, 162 (2021).
315. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry* **2**, 123–136 (2020).
316. Bryan, T. M. Mechanisms of DNA Replication and Repair: Insights from the Study of G-Quadruplexes. *Molecules* **24**, 3439 (2019).
317. Lerner, L. K. & Sale, J. E. Replication of G Quadruplex DNA. *Genes (Basel)*. **10**, 95 (2019).
318. Prorok, P. *et al.* Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.* **10**, 3274 (2019).
319. Valton, A.-L. & Prioleau, M.-N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* **32**, 697–706 (2016).
320. Tsukakoshi, K., Saito, S., Yoshida, W., Goto, S. & Ikebukuro, K. CpG Methylation Changes G-Quadruplex Structures Derived from Gene Promoters and Interaction with VEGF and SP1. *Molecules* **23**, 944 (2018).
321. Tu, A. T. T., Hoshi, K., Ikebukuro, K., Hanagata, N. & Yamazaki, T. Monomeric G-Quadruplex-Based CpG Oligodeoxynucleotides as Potent Toll-Like Receptor 9 Agonists. *Biomacromolecules* **21**, 3644–3657 (2020).
322. Jara-Espejo, M. & Line, S. R. DNA G-quadruplex stability, position and chromatin accessibility are associated with CpG island methylation. *FEBS J.* **287**, 483–495 (2020).
323. Yoshida, W., Terasaka, M., Laddachote, S. & Karube, I. Stabilization of G-quadruplex structure on vascular endothelial growth factor gene promoter depends on CpG methylation site and cation type. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 1933–1937 (2018).
324. Kim, N. The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.* **26**, 2898–2917 (2019).
325. Tong, X. *et al.* Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Res.* **39**, 6753–6763 (2011).
326. Agarwal, T., Roy, S., Kumar, S., Chakraborty, T. K. & Maiti, S. In the sense of transcription regulation by G-quadruplexes: Asymmetric effects in sense and antisense strands. *Biochemistry* **53**, 3711–3718 (2014).
327. Bourbon, H. M., Lapeyre, B. & Amalric, F. Structure of the mouse nucleolin gene. The complete sequence reveals that each RNA binding domain is encoded by two independent exons. *J. Mol. Biol.* **200**, 627–638 (1988).
328. Caizergues-Ferrer, M. *et al.* Nucleolin from *Xenopus laevis*: cDNA cloning and expression during development. *Genes Dev.* **3**, 324–333 (1989).
329. Maridor, G., Krek, W. & Nigg, E. A. Structure and developmental expression of chicken nucleolin and NO38: Coordinate expression of two abundant non-ribosomal nucleolar proteins. *BBA - Gene Struct. Expr.* **1049**, 126–133 (1990).

References

330. Peter, M., Nakagawa, J., Dorée, M., Labbé, J. C. & Nigg, E. A. Identification of major nucleolar proteins as candidate mitotic substrates of cdc2 kinase. *Cell* **60**, 791–801 (1990).
331. Tuteja, R. & Tuteja, N. Nucleolin: A Multifunctional Major Nucleolar Phosphoprotein. *Crit. Rev. Biochem. Mol. Biol.* **33**, 407–436 (1998).
332. Ginisty, H., Sicard, H., Roger, B. & Bouvet, P. Structure and functions of nucleolin. *J. Cell Sci.* **112** (Pt 6), 761–72 (1999).
333. Tajrish, M. M., Tuteja, R. & Tuteja, N. Nucleolin. *Commun. Integr. Biol.* **4**, 267–275 (2011).
334. Hanakahi, L. A., Sun, H. & Maizels, N. High Affinity Interactions of Nucleolin with G-G-paired rDNA. *J. Biol. Chem.* **274**, 15908–15912 (1999).
335. González, V., Guo, K., Hurley, L. & Sun, D. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* **284**, 23622–35 (2009).
336. Tosoni, E. *et al.* Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* **43**, 8884–8897 (2015).
337. De Nicola, B. *et al.* Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome. *Nucleic Acids Res.* **44**, 6442–6451 (2016).
338. González, V. & Hurley, L. H. The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry* **49**, 9706–14 (2010).
339. Lago, S., Tosoni, E., Nadai, M., Palumbo, M. & Richter, S. N. The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 1371–1381 (2017).
340. Saha, A. *et al.* Nucleolin Discriminates Drastically between Long-Loop and Short-Loop Quadruplexes. *Biochemistry* **59**, 1261–1272 (2020).
341. Sengupta, P., Bhattacharya, A., Sa, G., Das, T. & Chatterjee, S. Truncated G-Quadruplex Isomers Cross-Talk with the Transcription Factors To Maintain Homeostatic Equilibria in c-MYC Transcription. *Biochemistry* **58**, 1975–1991 (2019).
342. Dickerhoff, J., Onel, B., Chen, L., Chen, Y. & Yang, D. Solution Structure of a MYC Promoter G-Quadruplex with 1:6:1 Loop Length. *ACS Omega* **4**, 2533–2539 (2019).
343. Bian, W.-X. *et al.* Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res.* **47**, 56–68 (2019).
344. Santos, T. *et al.* Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* 114208 (2020). doi:10.1016/j.bcp.2020.114208
345. Figueiredo, J. *et al.* Targeting nucleolin by RNA G-quadruplex-forming motif. *Biochem. Pharmacol.* 114418 (2021). doi:10.1016/j.bcp.2021.114418
346. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Research* **44**, 1989–2006 (2016).
347. Wu, G., Xing, Z., Tran, E. J. & Yang, D. DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc. Natl. Acad. Sci.* **116**, 20453–20461 (2019).
348. De Magis, A. *et al.* Zuo1 supports G4 structure formation and directs repair toward nucleotide excision repair. *Nat. Commun.* **11**, 3907 (2020).
349. Hároniková, L. *et al.* IFI16 Preferentially Binds to DNA with Quadruplex Structure and Enhances DNA Quadruplex Formation. *PLoS One* **11**, e0157156 (2016).
350. Carle, C. M., Zaher, H. S. & Chalker, D. L. A Parallel G Quadruplex-Binding Protein Regulates the Boundaries of DNA Elimination Events of *Tetrahymena thermophila*. *PLOS Genet.* **12**, e1005842 (2016).
351. Niu, K. *et al.* Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res.* **47**, 7306–7320 (2019).
352. Williams, P., Li, L., Dong, X. & Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. *J. Am. Chem. Soc.* **139**, 12426–12429 (2017).
353. Chen, S. *et al.* Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 4769–4777 (2013).
354. David, A. P. *et al.* CNBP controls transcription by unfolding DNA G-quadruplex structures. *Nucleic Acids Res.* **47**, 7901–7913 (2019).
355. Fry, M. Tetraplex DNA and its interacting proteins. *Front. Biosci.* **12**, 4336–51 (2007).
356. Brázda, V., Hároniková, L., Liao, J. & Fojta, M. DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.* **15**, 17493–17517 (2014).
357. Brázda, V. *et al.* The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Molecules* **23**, 2341 (2018).
358. Huang, Z.-L. *et al.* Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J. Am. Chem. Soc.* **140**, 17945–17955 (2018).
359. Armas, P., Agüero, T. H., Borgognone, M., Aybar, M. J. & Calcaterra, N. B. Dissecting CNBP, a Zinc-Finger Protein Required for Neural Crest Development, in Its Structural and Functional Domains. *J. Mol. Biol.* **382**, 1043–1056 (2008).
360. Masuzawa, T. & Oyoshi, T. Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Omega* **5**, 5202–5208 (2020).
361. Mishra, S. K., Tawani, A., Mishra, A. & Kumar, A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **6**, 38144 (2016).
362. Heddi, B., Cheong, V. V., Martadinata, H. & Phan, A. T. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–quadruplex complex. *Proc. Natl. Acad. Sci.* **112**, 9608–9613 (2015).
363. Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465–469 (2018).
364. Cogo, S., Shchekotikhin, A. E. & Xodo, L. E. HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. *Nucleic Acids Res.* **42**, 8379–8388 (2014).
365. Cogo, S., Ferino, A., Miglietta, G., Pedersen, E. B. & Xodo, L. E. The regulatory G4 motif of the Kirsten ras (KRAS) gene is sensitive to guanine oxidation: implications on transcription. *Nucleic Acids Res.* **46**, 661–676 (2018).
366. Postel, E. H. Multiple biochemical activities of NM23/NDP kinase in gene regulation. *J. Bioenerg. Biomembr.* **35**, 31–40 (2003).
367. Dexheimer, T. S. *et al.* NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III 1. *Mol. Cancer Ther.* **8**, 1363–1377 (2009).
368. Ma, D. *et al.* NM23-H1 and NM23-H2 Repress Transcriptional Activities of Nuclease-hypersensitive Elements in the Platelet-derived Growth Factor-A Promoter. *J. Biol. Chem.* **277**, 1560–1567 (2002).
369. González, V. & Hurley, L. H. The c-MYC NHE III 1 : Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **50**, 111–129 (2010).

References

370. Sharma, S., Sengupta, A. & Chowdhury, S. NM23/NDPK proteins in transcription regulatory functions and chromatin modulation: emerging trends. *Lab. Investig.* **98**, 175–181 (2018).
371. Saha, D. *et al.* Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex-dependent fashion. *J. Biol. Chem.* **292**, 15205–15215 (2017).
372. Wu, Y., Shin-ya, K. & Brosh, R. M. FANCI Helicase Defective in Fanconi Anemia and Breast Cancer Unwinds G-Quadruplex DNA To Defend Genomic Stability. *Mol. Cell. Biol.* **28**, 4116–4128 (2008).
373. Sarkies, P. *et al.* FANCI coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res.* **40**, 1485–1498 (2012).
374. Wu, C. G. & Spies, M. G-quadruplex recognition and remodeling by the FANCI helicase. *Nucleic Acids Res.* **44**, 8742–8753 (2016).
375. Paeschke, K. *et al.* Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458–462 (2013).
376. Byrd, A. K. & Raney, K. D. Structure and function of Pif1 helicase. *Biochem. Soc. Trans.* **45**, 1159–1171 (2017).
377. Li, J.-L. *et al.* Inhibition of the Bloom's and Werner's Syndrome Helicases by G-Quadruplex Interacting Ligands. *Biochemistry* **40**, 15194–15202 (2001).
378. Wu, W. *et al.* HEC2 Facilitates BLM and WRN Helicase Complex Interaction with RPA to Suppress G-Quadruplex DNA. *Cancer Res.* **78**, 6371–6385 (2018).
379. Smestad, J. A. & Maher, L. J. Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. *BMC Med. Genet.* **16**, 91 (2015).
380. Wu, W.-Q., Hou, X.-M., Li, M., Dou, S.-X. & Xi, X.-G. BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res.* **43**, 4614–4626 (2015).
381. Budhathoki, J. B., Stafford, E. J., Yodh, J. G. & Balci, H. ATP-dependent G-quadruplex unfolding by Bloom helicase exhibits low processivity. *Nucleic Acids Res.* **43**, 5961–5970 (2015).
382. Chatterjee, S. *et al.* Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures. *Nat. Commun.* **5**, 5556 (2014).
383. van Wietmarschen, N. *et al.* BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat. Commun.* **9**, 271 (2018).
384. Lee, S., Kim, J., Han, S. & Park, C.-J. Recognition and Unfolding of c-MYC and Telomeric G-Quadruplex DNAs by the RecQ C-Terminal Domain of Human Bloom Syndrome Helicase. *ACS Omega* **5**, 14513–14522 (2020).
385. Aggarwal, M., Sommers, J. a., Shoemaker, R. H. & Brosh, R. M. Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. *Proc. Natl. Acad. Sci.* **108**, 1525–1530 (2011).
386. Johnson, J. E., Cao, K., Ryvkin, P., Wang, L.-S. & Johnson, F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* **38**, 1114–1122 (2010).
387. Budhathoki, J. B. *et al.* A Comparative Study of G-Quadruplex Unfolding and DNA Reeling Activities of Human RECQ5 Helicase. *Biophys. J.* **110**, 2585–2596 (2016).
388. Keller, H. *et al.* The intrinsically disordered amino-terminal region of human RECQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. *Nucleic Acids Res.* **42**, 12614–12627 (2014).
389. Ghosh, A. K. *et al.* RECQL4, the protein mutated in Rothmund-Thomson syndrome, functions in telomere maintenance. *J. Biol. Chem.* **287**, 196–209 (2012).
390. Castillo-Tandazo, W. *et al.* ATP-dependent helicase activity is dispensable for the physiological functions of Recq4. *PLOS Genet.* **15**, e1008266 (2019).
391. Sauer, M. & Paeschke, K. G-quadruplex unwinding helicases and their function in vivo. *Biochem. Soc. Trans.* **45**, 1173–1182 (2017).
392. Tippana, R., Hwang, H., Opreko, P. L., Bohr, V. A. & Myong, S. Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8448–8453 (2016).
393. Popuri, V. *et al.* The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *J. Biol. Chem.* **283**, 17766–76 (2008).
394. Ellis, N. A. *et al.* The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**, 655–666 (1995).
395. Gray, M. D. *et al.* The Werner syndrome protein is a DNA helicase. *Nat. Genet.* **17**, 100–103 (1997).
396. Larizza, L., Magnani, I. & Roversi, G. Rothmund-Thomson syndrome and RECQL4 defect: Splitting and lumping. *Cancer Lett.* **232**, 107–120 (2006).
397. Majello, B., De Luca, P., Suske, G. & Lania, L. Differential transcriptional regulation of c-myc promoter through the same DNA binding sites targeted by Sp1-like proteins. *Oncogene* **10**, 1841–8 (1995).
398. Suske, G. The Sp-family of transcription factors. *Gene* **238**, 291–300 (1999).
399. Han, H., Langley, D. R., Rangan, A. & Hurley, L. H. Selective interactions of cationic porphyrins with G-quadruplex structures. *J. Am. Chem. Soc.* **123**, 8902–13 (2001).
400. Haiyong, H., Cliff, C. L. & Hurley, L. H. Accelerated assembly of G-quadruplex structures by a small molecule. *Biochemistry* **38**, 6981–6986 (1999).
401. Izbicka, E. *et al.* Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.* **59**, 639–44 (1999).
402. Tian, T., Chen, Y.-Q., Wang, S.-R. & Zhou, X. G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem* **4**, 1314–1344 (2018).
403. Bessi, I., Wirmer-Bartoschek, J., Dash, J. & Schwalbe, H. Targeting G-quadruplex with Small Molecules: An NMR View. in *Modern Magnetic Resonance* 2189–2210 (Springer International Publishing, 2018). doi:10.1007/978-3-319-28388-3_126
404. Franceschin, M. *et al.* Natural and synthetic G-quadruplex interactive berberine derivatives. *Bioorganic Med. Chem. Lett.* **16**, 1707–1711 (2006).
405. Shirude, P. S. *et al.* Macrocyclic and Helical Oligoamides as a New Class of G-Quadruplex Ligands. *J. Am. Chem. Soc.* **129**, 11890–11891 (2007).
406. Monchaud, D. & Teulade-Fichou, M.-P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **6**, 627–36 (2008).
407. Burger, A. M. *et al.* The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* **65**, 1489–1496 (2005).
408. Sun, D., Guo, K., Rusche, J. J. & Hurley, L. H. 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 Res.* **33**, 6070–6080 (2005).
409. Shin-ya, K. *et al.* Telomestatin, a Novel Telomerase Inhibitor from *Streptomyces anulatus*. *J. Am. Chem. Soc.* **123**, 1262–1263 (2001).
410. Bessi, I. *et al.* Spectroscopic, molecular modeling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA. *ACS Chem. Biol.* **7**, 1109–1119 (2012).

References

411. Ruggiero, E. & Richter, S. N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.* **46**, 3270–3283 (2018).
412. Duarte, A. R., Cadoni, E., Ressurreição, A. S., Moreira, R. & Paulo, A. Design of Modular G-quadruplex Ligands. *ChemMedChem* **13**, 869–893 (2018).
413. O'Hagan, M. P., Morales, J. C. & Galan, M. C. Binding and Beyond: What Else Can G-Quadruplex Ligands Do? *European J. Org. Chem.* **2019**, 4995–5017 (2019).
414. De Magis, A. *et al.* DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci.* **116**, 816–825 (2019).
415. Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, D1115–D1123 (2013).
416. Wirmer-Bartoschek, J. *et al.* Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements that Affect Ligand Binding. *Angew. Chem. Int. Ed.* **56**, 7102–7106 (2017).
417. Lejault, P., Mitteaux, J., Sperti, F. R. & Monchaud, D. How to untie G-quadruplex knots and why? *Cell Chem. Biol.* in press (2021). doi:10.1016/j.chembiol.2021.01.015
418. Grand, C. L. *et al.* The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* **1**, 565–73 (2002).
419. Kim, M.-Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D. & Hurley, L. H. The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Res.* **63**, 3247–56 (2003).
420. Parkinson, G. N., Ghosh, R. & Neidle, S. Structural basis for binding of porphyrin to human telomeres. *Biochemistry* **46**, 2390–7 (2007).
421. Read, M. *et al.* Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4844–9 (2001).
422. Campbell, N. H., Parkinson, G. N., Reszka, A. P. & Neidle, S. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *J. Am. Chem. Soc.* **130**, 6722–6724 (2008).
423. Monchaud, D. *et al.* Ligands playing musical chairs with G-quadruplex DNA: A rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* **90**, 1207–1223 (2008).
424. De Cian, A., DeLemos, E., Mergny, J.-L., Teulade-Fichou, M.-P. & Monchaud, D. Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J. Am. Chem. Soc.* **129**, 1856–1857 (2007).
425. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound Phen-DC(3). *Angew. Chem. Int. Ed.* **53**, 999–1002 (2014).
426. Chung, W. J. *et al.* Solution Structure of an Intramolecular (3+1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative. *J. Am. Chem. Soc.* **135**, 13495–13501 (2013).
427. Casini, A. *et al.* Structural and solution chemistry, antiproliferative effects, and DNA and protein binding properties of a series of dinuclear gold(III) compounds with bipyridyl ligands. *J. Med. Chem.* **49**, 5524–5531 (2006).
428. Gabbiani, C., Guerri, A., Cinellu, M. A. & Messori, L. Dinuclear Gold(III) Complexes as Potential Anticancer Agents: Structure, Reactivity and Biological Profile of a Series of Gold(III) Oxo-Bridged Derivatives. *Open Crystallogr. J.* **3**, 29–40 (2010).
429. Gratteri, P. *et al.* Interactions of selected gold(III) complexes with DNA G quadruplexes. *Dalt. Trans.* **44**, 3633–9 (2015).
430. Heller, A. Spiers Memorial Lecture: On the hypothesis of cathodic protection of genes. *Faraday Discuss.* **116**, 1–13 (2000).
431. Fleming, A. M. & Burrows, C. J. 8-Oxo-7,8-dihydroguanine, friend and foe: Epigenetic-like regulator versus initiator of mutagenesis. *DNA Repair (Amst)*. **56**, 75–83 (2017).
432. Redstone, S. C. J., Fleming, A. M. & Burrows, C. J. Oxidative Modification of the Potential G-Quadruplex Sequence in the PCNA Gene Promoter Can Turn on Transcription. *Chem. Res. Toxicol.* **32**, 437–446 (2019).
433. Fleming, A. M. & Burrows, C. J. G-Quadruplex Folds of the Human Telomere Sequence Alter the Site Reactivity and Reaction Pathway of Guanine Oxidation Compared to Duplex DNA. *Chem. Res. Toxicol.* **26**, 593–607 (2013).
434. An, N., Fleming, A. M. & Burrows, C. J. Human Telomere G-Quadruplexes with Five Repeats Accommodate 8-Oxo-7,8-dihydroguanine by Looping out the DNA Damage. *ACS Chem. Biol.* **11**, 500–507 (2016).
435. Fleming, A. M., Ding, Y. & Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2604–2609 (2017).
436. Fleming, A. M., Zhu, J., Ding, Y. & Burrows, C. J. 8-Oxo-7,8-dihydroguanine in the Context of a Gene Promoter G-Quadruplex Is an On–Off Switch for Transcription. *ACS Chem. Biol.* **12**, 2417–2426 (2017).
437. Zhu, J., Fleming, A. M. & Burrows, C. J. The RAD17 Promoter Sequence Contains a Potential Tail-Dependent G-Quadruplex That Downregulates Gene Expression upon Oxidative Modification. *ACS Chem. Biol.* **13**, 2577–2584 (2018).
438. Omega, C. A., Fleming, A. M. & Burrows, C. J. The Fifth Domain in the G-Quadruplex-Forming Sequence of the Human NEIL3 Promoter Locks DNA Folding in Response to Oxidative Damage. *Biochemistry* **57**, 2958–2970 (2018).
439. Fleming, A. M., Zhu, J., Howpay Manage, S. A. & Burrows, C. J. Human NEIL3 Gene Expression Regulated by Epigenetic-Like Oxidative DNA Modification. *J. Am. Chem. Soc.* **141**, 11036–11049 (2019).
440. Fleming, A. M. & Burrows, C. J. Interplay of Guanine Oxidation and G-Quadruplex Folding in Gene Promoters. *Journal of the American Chemical Society* **142**, 1115–1136 (2020).
441. Zhou, J., Fleming, A. M., Averill, A. M., Burrows, C. J. & Wallace, S. S. The NEIL glycosylases remove oxidized guanine lesions from telomeric and promoter quadruplex DNA structures. *Nucleic Acids Res.* **43**, 4039–4054 (2015).
442. Konvalinová, H. *et al.* Diverse effects of naturally occurring base lesions on the structure and stability of the human telomere DNA quadruplex. *Biochimie* **118**, 15–25 (2015).
443. Bielskutė, S., Plavec, J. & Podbevšek, P. Impact of Oxidative Lesions on the Human Telomeric G-Quadruplex. *J. Am. Chem. Soc.* **141**, 2594–2603 (2019).
444. Bielskutė, S., Plavec, J. & Podbevšek, P. Oxidative lesions modulate G-quadruplex stability and structure in the human BCL2 promoter. *Nucleic Acids Res.* **49**, 2346–2356 (2021).
445. Takahashi, S. *et al.* Recovery of the Formation and Function of Oxidized G-Quadruplexes by a Pyrene-Modified Guanine Tract. *J. Am. Chem. Soc.* **140**, 5774–5783 (2018).
446. Cogo, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536–2549 (2006).
447. Cogo, S., Paramasivam, M., Spolaore, B. & Xodo, L. E. Structural polymorphism within a regulatory element of the human KRAS promoter: Formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res.* **36**, 3765–3780 (2008).

References

448. Paramasivam, M. *et al.* Protein hnRNP A1 and its derivative Up1 unfold quadruplex DNA in the human KRAS promoter: Implications for transcription. *Nucleic Acids Res.* **37**, 2841–2853 (2009).
449. Balasubramanian, S. & Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Current Opinion in Chemical Biology* **13**, 345–353 (2009).
450. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10**, 261–275 (2011).
451. Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev. Chem.* **1**, 0041 (2017).
452. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci.* **99**, 11593–11598 (2002).
453. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
454. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
455. Fouad, Y. A. & Aanei, C. Revisiting the hallmarks of cancer. *Am. J. Cancer Res.* **7**, 1016–1036 (2017).
456. Brooks, T. A. & Hurley, L. H. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **9**, 849–61 (2009).
457. Pirota, V., Nadai, M., Doria, F. & Richter, S. N. Naphthalene Diimides as Multimodal G-Quadruplex-Selective Ligands. *Molecules* **24**, 426 (2019).
458. Sanchez-Martin, V., Lopez-Pujante, C., Soriano-Rodriguez, M. & Garcia-Salcedo, J. A. An Updated Focus on Quadruplex Structures as Potential Therapeutic Targets in Cancer. *Int. J. Mol. Sci.* **21**, 8900 (2020).
459. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Mol. Cancer* **20**, 40 (2021).
460. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048–58 (2005).
461. Mathad, R. I., Hatzakis, E., Dai, J. & Yang, D. C-MYC promoter G-quadruplex formed at the 5'-end of NHE III 1 element: Insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res.* **39**, 9023–9033 (2011).
462. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34**, 5133–5144 (2006).
463. Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **128**, 5404–5415 (2006).
464. Sengupta, P., Chattopadhyay, S. & Chatterjee, S. G-Quadruplex surveillance in BCL-2 gene: a promising therapeutic intervention in cancer treatment. *Drug Discov. Today* **22**, 1165–1186 (2017).
465. Ducani, C., Bernardinelli, G., Högberg, B., Keppler, B. K. & Terenzi, A. Interplay of Three G-Quadruplex Units in the KIT Promoter. *J. Am. Chem. Soc.* **141**, 10205–10213 (2019).
466. Morgan, R. K., Batra, H., Gaerig, V. C., Hockings, J. & Brooks, T. A. Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 235–245 (2016).
467. Qin, Y., Rezler, E. M., Gokhale, V., Sun, D. & Hurley, L. H. Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.* **35**, 7698–7713 (2007).
468. Qin, Y. *et al.* Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. *Biochemistry* **49**, 4208–19 (2010).
469. De Armond, R., Wood, S., Sun, D., Hurley, L. H. & Ebbinghaus, S. W. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* **44**, 16341–50 (2005).
470. Xu, Y. & Sugiyama, H. Structural and functional characterizations of the G-quartet and i-motif elements in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Symp. Ser. (Oxf)*. 177–178 (2005).
471. Xu, Y. & Sugiyama, H. Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **34**, 949–954 (2006).
472. Garman, E. F. Rosalind Franklin 1920-1958. *Acta Crystallogr. Sect. D, Struct. Biol.* **76**, 698–701 (2020).
473. Wierstra, I. & Alves, J. The c-myc promoter: still MysterY and challenge. *Adv. Cancer Res.* **99**, 113–333 (2008).
474. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* **2**, 764–776 (2002).
475. Adhikary, S. & Eilers, M. Transcriptional regulation and transformation by Myc proteins. *Nat. Rev. Mol. Cell Biol.* **6**, 635–645 (2005).
476. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nat. Rev. Cancer* **8**, 976–990 (2008).
477. Dang, C. V. MYC on the Path to Cancer. *Cell* **149**, 22–35 (2012).
478. Chen, H., Liu, H. & Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct. Target. Ther.* **3**, 5 (2018).
479. Duffy, M. J. & Crown, J. Drugging “undruggable” genes for cancer treatment: Are we making progress? *Int. J. Cancer* **148**, 8–17 (2021).
480. Nair, S. K. & Burley, S. K. X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA. *Cell* **112**, 193–205 (2003).
481. Dang, C. V., Reddy, E. P., Shokat, K. M. & Soucek, L. Drugging the ‘undruggable’ cancer targets. *Nat. Rev. Cancer* **17**, 502–508 (2017).
482. Molyneux, E. M. *et al.* Burkitt’s lymphoma. *Lancet* **379**, 1234–1244 (2012).
483. Eisenman, R. N. Deconstructing Myc. *Genes Dev.* **15**, 2023–2030 (2001).
484. Levens, D. ‘You Don’t Muck with MYC’. *Genes Cancer* **1**, 547–554 (2010).
485. Gazin, C. *et al.* Nucleotide sequence of the human c-myc locus: provocative open reading frame within the first exon. *EMBO J.* **3**, 383–387 (1984).
486. Boles, T. C. & Hogan, M. E. DNA structure equilibria in the human c-myc gene. *Biochemistry* **26**, 367–376 (1987).
487. Siebenlist, U., Hennighausen, L., Battey, J. & Leder, P. Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in burkitt lymphoma. *Cell* **37**, 381–391 (1984).
488. Mirkin, S. M. *et al.* DNA H form requires a homopurine–homopyrimidine mirror repeat. *Nature* **330**, 495–497 (1987).
489. Cooney, M., Czernuszewicz, G., Postel, E., Flint, S. & Hogan, M. Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Science* **241**, 456–459 (1988).
490. Postel, E. H., Mango, S. E. & Flint, S. J. A nuclease-hypersensitive element of the human c-myc promoter interacts with a transcription initiation factor. *Mol. Cell. Biol.* **9**, 5123–5133 (1989).
491. Kinniburgh, A. J., Firulli, A. B. & Kolluri, R. DNA triplexes and regulation of the c-myc gene. *Gene* **149**, 93–100 (1994).
492. Olivas, W. M. & Maher, L. J. I. Competitive Triplex/Quadruplex Equilibria Involving Guanine-Rich Oligonucleotides. *Biochemistry* **34**, 278–284 (1995).
493. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **26**, 1167–1172 (1998).

References

494. Simonsson, T., Pribylova, M. & Vorlickova, M. A Nuclease Hypersensitive Element in the Human c-myc Promoter Adopts Several Distinct i-Tetraplex Structures. *Biochem. Biophys. Res. Commun.* **278**, 158–166 (2000).
495. Wang, G. & Vasquez, K. M. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc. Natl. Acad. Sci.* **101**, 13448–13453 (2004).
496. Belotserkovskii, B. P. *et al.* A Triplex-forming Sequence from the Human c-MYC Promoter Interferes with DNA Transcription. *J. Biol. Chem.* **282**, 32433–32441 (2007).
497. del Mundo, I. M. A., Zewail-Foote, M., Kerwin, S. M. & Vasquez, K. M. Alternative DNA structure formation in the mutagenic human c-MYC promoter. *Nucleic Acids Res.* **45**, 4929–4943 (2017).
498. Kim, H. G. *et al.* Inhibition of transcription of the human c-myc protooncogene by intermolecular triplex. *Biochemistry* **37**, 2299–304 (1998).
499. Catapano, C. V., McGuffie, E. M., Pacheco, D. & Carbone, G. M. Inhibition of gene expression and cell proliferation by triple helix-forming oligonucleotides directed to the c-myc gene. *Biochemistry* **39**, 5126–38 (2000).
500. McGuffie, E. M., Pacheco, D., Carbone, G. M. R. & Catapano, C. V. Antigenic and antiproliferative effects of a c-myc-targeting phosphorothioate triple helix-forming oligonucleotide in human leukemia cells. *Cancer Res.* **60**, 3790–9 (2000).
501. Christensen, L. A., Finch, R. A., Booker, A. J. & Vasquez, K. M. Targeting Oncogenes to Improve Breast Cancer Chemotherapy. *Cancer Res.* **66**, 4089–4094 (2006).
502. Sedoris, K. C. *et al.* Genomic c-Myc Quadruplex DNA Selectively Kills Leukemia. *Mol. Cancer Ther.* **11**, 66–76 (2012).
503. Boulware, S. B. *et al.* Triplex-forming oligonucleotides targeting c-MYC potentiate the anti-tumor activity of gemcitabine in a mouse model of human cancer. *Mol. Carcinog.* **53**, 744–752 (2014).
504. Muench, D. *et al.* Quadruplex-forming oligonucleotide targeted to the VEGF promoter inhibits growth of non-small cell lung cancer cells. *PLoS One* **14**, (2019).
505. Ponzielli, R., Katz, S., Barsyte-Lovejoy, D. & Penn, L. Z. Cancer therapeutics: Targeting the dark side of Myc. *Eur. J. Cancer* **41**, 2485–2501 (2005).
506. Kalkat, M. *et al.* MYC Deregulation in Primary Human Cancers. *Genes (Basel)*. **8**, 151 (2017).
507. Xu-Monette, Z. Y. *et al.* Clinical and Biologic Significance of MYC Genetic Mutations in De Novo Diffuse Large B-cell Lymphoma. *Clin. Cancer Res.* **22**, 3593–3605 (2016).
508. Rangan, A., Fedoroff, O. Y. & Hurley, L. H. Induction of Duplex to G-quadruplex Transition in the c-myc Promoter Region by a Small Molecule. *J. Biol. Chem.* **276**, 4640–4646 (2001).
509. Dembska, A., Rzepecka, P. & Juskowiak, B. Spectroscopic Characterization of i-motif Forming c-myc Derived Sequences Double-Labeled with Pyrene. *J. Fluoresc.* **23**, 807–812 (2013).
510. Reilly, S. M. *et al.* Folding and Hydrodynamics of a DNA i-Motif from the c-MYC Promoter Determined by Fluorescent Cytidine Analogs. *Biophys. J.* **107**, 1703–1711 (2014).
511. Paul, S., Hossain, S. S. & Samanta, A. Insights into the Folding Pathway of a c-MYC-Promoter-Based i-Motif DNA in Crowded Environments at the Single-Molecule Level. *J. Phys. Chem. B* **124**, 763–770 (2020).
512. Brooks, T. A. & Hurley, L. H. Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **1**, 641–649 (2010).
513. Liu, J. & Levens, D. Making myc. *Curr. Top. Microbiol. Immunol.* **302**, 1–32 (2006).
514. Michelotti, E. F., Tomonaga, T., Krutzsch, H. & Levens, D. Cellular Nucleic Acid Binding Protein Regulates the CT Element of the Human c-myc Protooncogene. *J. Biol. Chem.* **270**, 9494–9499 (1995).
515. Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell Biol.* **16**, 2350–2360 (1996).
516. Tomonaga, T. & Levens, D. Activating transcription from single stranded DNA. *Proc. Natl. Acad. Sci.* **93**, 5830–5835 (1996).
517. Postel, E., Berberich, S., Flint, S. & Ferrone, C. Human c-myc transcription factor PuF identified as nm23-H2 nucleoside diphosphate kinase, a candidate suppressor of tumor metastasis. *Science* **261**, 478–480 (1993).
518. Berberich, S. J. & Postel, E. H. PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene* **10**, 2343–7 (1995).
519. Ji, L., Arcinas, M. & Boxer, L. M. The Transcription Factor, Nm23H2, Binds to and Activates the Translocated c-myc Allele in Burkitt's Lymphoma. *J. Biol. Chem.* **270**, 13392–13398 (1995).
520. Rigo, R., Palumbo, M. & Sissi, C. G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 1399–1413 (2017).
521. Sun, D. & Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **52**, 2863–2874 (2009).
522. Stump, S., Mou, T.-C., Sprang, S. R., Natale, N. R. & Beall, H. D. Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene. *PLoS One* **13**, e0205584 (2018).
523. Dettler, J. M., Buscaglia, R., Le, V. H. & Lewis, E. A. DSC Deconvolution of the Structural Complexity of c-MYC P1 Promoter G-Quadruplexes. *Biophys. J.* **100**, 1517–1525 (2011).
524. Marchand, A., Rosu, F., Zenobi, R. & Gabelica, V. Thermal Denaturation of DNA G-Quadruplexes and Their Complexes with Ligands: Thermodynamic Analysis of the Multiple States Revealed by Mass Spectrometry. *J. Am. Chem. Soc.* **140**, 12553–12565 (2018).
525. Prochownik, E. V. & Vogt, P. K. Therapeutic Targeting of Myc. *Genes Cancer* **1**, 650–659 (2010).
526. Koh, C. M., Sabò, A. & Guccione, E. Targeting MYC in cancer therapy: RNA processing offers new opportunities. *BioEssays* **38**, 266–275 (2016).
527. Allen-Petersen, B. L. & Sears, R. C. Mission Possible: Advances in MYC Therapeutic Targeting in Cancer. *BioDrugs* **33**, 539–553 (2019).
528. Wolf, E. & Eilers, M. Targeting MYC Proteins for Tumor Therapy. *Annu. Rev. Cancer Biol.* **4**, 61–75 (2020).
529. Ross, J. *et al.* Targeting MYC: From understanding its biology to drug discovery. *Eur. J. Med. Chem.* **213**, 113137 (2021).
530. Wang, W. *et al.* Human MYC G-quadruplex: From discovery to a cancer therapeutic target. *Biochimica et Biophysica Acta - Reviews on Cancer* **1874**, 188410 (2020).
531. Liu, W. *et al.* Structures of 1:1 and 2:1 complexes of BMVC and MYC promoter G-quadruplex reveal a mechanism of ligand conformation adjustment for G4-recognition. *Nucleic Acids Res.* **47**, 11931–11942 (2019).
532. Asamitsu, S., Obata, S., Yu, Z., Bando, T. & Sugiyama, H. Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy. *Molecules* **24**, 429 (2019).
533. Freyer, M. W. *et al.* Biophysical Studies of the c-MYC NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin. *Biophys. J.* **92**, 2007–2015 (2007).

References

534. Gluszyńska, A., Juskowiak, B., Kuta-Siejkowska, M., Hoffmann, M. & Haider, S. Carbazole ligands as c-myc G-quadruplex binders. *Int. J. Biol. Macromol.* **114**, 479–490 (2018).
535. Felsenstein, K. M. *et al.* Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression. *ACS Chem. Biol.* **11**, 138–148 (2016).
536. Pany, S. P. P., Bommiseti, P., Diveshkumar, K. V. & Pradeepkumar, P. I. Benzothiazole hydrazones of furylbenzamides preferentially stabilize c-MYC and c-KIT1 promoter G-quadruplex DNAs. *Org. Biomol. Chem.* **14**, 5779–5793 (2016).
537. Wu, T.-Y., Huang, Q., Huang, Z.-S., Hu, M.-H. & Tan, J.-H. A drug-like imidazole-benzothiazole conjugate inhibits malignant melanoma by stabilizing the c-MYC G-quadruplex. *Bioorg. Chem.* **99**, 103866 (2020).
538. Dutta, D. *et al.* Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC G-quadruplex. *Nucleic Acids Res.* **46**, 5355–5365 (2018).
539. Panda, D., Saha, P., Das, T. & Dash, J. Target guided synthesis using DNA nano-templates for selectively assembling a G-quadruplex binding c-MYC inhibitor. *Nat. Commun.* **8**, 16103 (2017).
540. Nimbarte, V. D. *et al.* Synthesis and In vitro Evaluation of novel 5-nitroIndole Derivatives as c-Myc G-Quadruplex Binders with Anticancer Activities. *ChemMedChem* cmdc.202000835 (2021). doi:10.1002/cmdc.202000835
541. Chauhan, A. *et al.* Synthesis of Fluorescent Binaphthyl Amines That Bind c-MYC G-Quadruplex DNA and Repress c-MYC Expression. *J. Med. Chem.* **59**, 7275–7281 (2016).
542. Pavan Kumar, Y. *et al.* Fluorescent Dansyl-Guanosine Conjugates that Bind c-MYC Promoter G-Quadruplex and Downregulate c-MYC Expression. *ChemBioChem* **17**, 388–393 (2016).
543. Hu, M. H. *et al.* Discovery of a New Four-Leaf Clover-Like Ligand as a Potent c-MYC Transcription Inhibitor Specifically Targeting the Promoter G-Quadruplex. *J. Med. Chem.* **61**, 2447–2459 (2018).
544. Asamitsu, S. *et al.* Simultaneous Binding of Hybrid Molecules Constructed with Dual DNA-Binding Components to a G-Quadruplex and Its Proximal Duplex. *Chem. Eur. J.* **24**, 4428–4435 (2018).
545. Asamitsu, S., Li, Y., Bando, T. & Sugiyama, H. Ligand-Mediated G-Quadruplex Induction in a Double-Stranded DNA Context by Cyclic Imidazole/Lysine Polyamide. *ChemBioChem* **17**, 1317–1322 (2016).
546. Binas, O. *et al.* 19 F NMR-Based Fragment Screening for 14 Different Biologically Active RNAs and 10 DNA and Protein Counter-Screens. *ChemBioChem* **22**, 423–433 (2021).
547. Drygin, D. *et al.* Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
548. Xu, H. *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8**, 14432 (2017).
549. Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H. & Brooks, T. A. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* **286**, 41018–27 (2011).
550. Das, T., Panda, D., Saha, P. & Dash, J. Small Molecule Driven Stabilization of Promoter G-Quadruplexes and Transcriptional Regulation of c-MYC. *Bioconjug. Chem.* **29**, 2636–2645 (2018).
551. Calabrese, D. R. *et al.* Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat. Commun.* **9**, 4229 (2018).
552. Bouvard, C. *et al.* Small molecule selectively suppresses MYC transcription in cancer cells. *Proc. Natl. Acad. Sci.* **114**, 3497–3502 (2017).
553. Musso, L. *et al.* c-MYC G-quadruplex binding by the RNA polymerase I inhibitor BMH-21 and analogues revealed by a combined NMR and biochemical Approach. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 615–629 (2018).
554. Cercek, A., Wheler, J., Murray, P. E., Zhou, S. & Saltz, L. Phase 1 study of APTO-253 HCl, an inducer of KLF4, in patients with advanced or metastatic solid tumors. *Invest. New Drugs* **33**, 1086–1092 (2015).
555. Local, A. *et al.* APTO-253 stabilizes G-quadruplex DNA, inhibits MYC expression, and induces DNA damage in acute myeloid leukemia cells. *Mol. Cancer Ther.* **17**, 1177–1186 (2018).
556. Tsai, C. Y. *et al.* APTO-253 is a new addition to the repertoire of drugs that can exploit DNA BRCA1/2 Deficiency. *Mol. Cancer Ther.* **17**, 1167–1176 (2018).
557. Pal, S. & Paul, S. An in silico investigation of the binding modes and pathway of APTO-253 on c-KIT G-quadruplex DNA. *Phys. Chem. Chem. Phys.* **23**, 3361–3376 (2021).
558. Haddach, M. *et al.* Discovery of CX-5461, the first direct and selective inhibitor of RNA polymerase I, for cancer therapeutics. *ACS Med. Chem. Lett.* **3**, 602–606 (2012).
559. Lee, H. C. *et al.* RNA Polymerase I Inhibition with CX-5461 as a Novel Therapeutic Strategy to Target MYC in Multiple Myeloma. *Br. J. Haematol.* **177**, 80–94 (2017).
560. Mars, J.-C. *et al.* The chemotherapeutic agent CX-5461 irreversibly blocks RNA polymerase I initiation and promoter release to cause nucleolar disruption, DNA damage and cell inviability. *NAR Cancer* **2**, (2020).
561. Yan, S. *et al.* The RNA polymerase I transcription inhibitor CX-5461 cooperates with topoisomerase 1 inhibition by enhancing the DNA damage response in homologous recombination-proficient high-grade serous ovarian cancer. *Br. J. Cancer* **124**, 616–627 (2021).
562. Bruno, P. M. *et al.* The primary mechanism of cytotoxicity of the chemotherapeutic agent CX-5461 is topoisomerase II poisoning. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4053–4060 (2020).
563. Sanij, E. *et al.* CX-5461 activates the DNA damage response and demonstrates therapeutic efficacy in high-grade serous ovarian cancer. *Nat. Commun.* **11**, 1–18 (2020).
564. Sullivan, H. J., Chen, B. & Wu, C. Molecular dynamics study on the binding of an anticancer DNA G-quadruplex stabilizer, CX-5461, to human telomeric, C-Kit1, and c-Myc G-quadruplexes and a DNA duplex. *J. Chem. Inf. Model.* **60**, 5203–5224 (2020).
565. Khot, A. *et al.* First-in-human RNA polymerase I transcription inhibitor CX-5461 in patients with advanced hematologic cancers: Results of a phase I dose-escalation study. *Cancer Discov.* **9**, 1036–1049 (2019).
566. Jana, J. *et al.* Human cathelicidin peptide LL37 binds telomeric G-quadruplex. *Mol. Biosyst.* **9**, 1833–1836 (2013).
567. Sengupta, P. *et al.* Site-specific amino acid substitution in dodecameric peptides determines the stability and unfolding of c-MYC quadruplex promoting apoptosis in cancer cells. *Nucleic Acids Res.* **46**, 9932–9950 (2018).
568. Purcell, E. M. Research in nuclear magnetism. *Nobel Lecture* December 11 (1952). Available at: <https://www.nobelprize.org/uploads/2018/06/purcell-lecture.pdf>. (Accessed: 19th January 2021)
569. Binas, O., Bessi, I. & Schwalbe, H. Structure Validation of G-Rich RNAs in Noncoding Regions of the Human Genome. *ChemBioChem* **21**, 1656–1663 (2020).
570. Tunis-Schneider, M. J. B. & Maestre, M. F. Circular dichroism spectra of oriented and unoriented deoxyribonucleic acid films-A preliminary study. *J. Mol. Biol.* **52**, 521–541 (1970).

References

571. Del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed.* **57**, 7171–7175 (2018).
572. Kypr, J., Kejnovská, I., Renciuk, D. & Vorlicková, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713–1725 (2009).
573. Gray, D. M. *et al.* Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* **20**, 431–440 (2008).
574. Vorlíčková, M., Kejnovská, I., Bednářová, K., Renciuk, D. & Kypr, J. Circular Dichroism Spectroscopy of DNA: From Duplexes to Quadruplexes. *Chirality* **24**, 691–698 (2012).
575. Daly, S., Rosu, F. & Gabelica, V. Mass-resolved electronic circular dichroism ion spectroscopy. *Science* **368**, 1465–1468 (2020).
576. Overhauser, A. W. Polarization of Nuclei in Metals. *Phys. Rev.* **92**, 411–415 (1953).
577. Carver, T. R. & Slichter, C. P. Polarization of Nuclear Spins in Metals. *Phys. Rev.* **92**, 212–213 (1953).
578. Solomon, I. Relaxation Processes in a System of Two Spins. *Phys. Rev.* **99**, 559–565 (1955).
579. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553 (1979).
580. Kumar, A., Ernst, R. R. & Wüthrich, K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6 (1980).
581. Vögeli, B. The nuclear Overhauser effect from a quantitative perspective. *Prog. Nucl. Magn. Reson. Spectrosc.* **78**, 1–46 (2014).
582. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *ChemBioChem* **4**, 936–962 (2003).
583. Schnieders, R., Keyhani, S., Schwalbe, H. & Fürtig, B. More than Proton Detection—New Avenues for NMR Spectroscopy of RNA. *Chem. – A Eur. J.* **26**, 102–113 (2020).
584. Schnieders, R. *et al.* NMR Spectroscopy of Large Functional RNAs: From Sample Preparation to Low-Gamma Detection. *Curr. Protoc. nucleic acid Chem.* **82**, e116 (2020).
585. Webba da Silva, M. NMR methods for studying quadruplex nucleic acids. *Methods* **43**, 264–277 (2007).
586. Feigon, J., Koshlap, K. M. & Smith, F. W. 1H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* **261**, 225–55 (1995).
587. Zimmer, D. P. & Crothers, D. M. NMR of enzymatically synthesized uniformly 13C15N-labeled DNA oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3091–5 (1995).
588. Louis, J. M., Martin, R. G., Clore, G. M. & Gronenborn, A. M. Preparation of Uniformly Isotope-labeled DNA Oligonucleotides for NMR Spectroscopy. *J. Biol. Chem.* **273**, 2374–2378 (1998).
589. Lohmann, J. S., Stougaard, M. & Koch, J. A new enzymatic route for production of long 5'-phosphorylated oligonucleotides using suicide cassettes and rolling circle DNA synthesis. *BMC Biotechnol.* **7**, 49 (2007).
590. Nelissen, F. H. T., Girard, F. C., Tessari, M., Heus, H. A. & Wijmenga, S. S. Preparation of selective and segmentally labeled single-stranded DNA for NMR by self-primed PCR and asymmetrical endonuclease double digestion. *Nucleic Acids Res.* **37**, e114 (2009).
591. Nelissen, F. H. T., Goossens, E. P. M., Tessari, M. & Heus, H. A. Enzymatic preparation of multimilligram amounts of pure single-stranded DNA samples for material and analytical sciences. *Anal. Biochem.* **475**, 68–73 (2015).
592. Nelissen, F. H. T., Tessari, M., Wijmenga, S. S. & Heus, H. A. Stable isotope labeling methods for DNA. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 89–108 (2016).
593. Nunes, A. R. D., Chavante, S. F., Rocha, H. A. O. & Lanza, D. C. F. 'in-House' Production of Single Stranded Oligodeoxyribonucleotides. *Nucleic Acid Ther.* **27**, 115–120 (2017).
594. Meija, J. *et al.* Isotopic compositions of the elements 2013 (IUPAC Technical Report). *Pure Appl. Chem.* **88**, 293–306 (2016).
595. Phan, A. T. Long-range imino proton-13C J-couplings and the through-bond correlation of imino and non-exchangeable protons in unlabeled DNA. *J. Biomol. NMR* **16**, 175–178 (2000).
596. Phan, A. T. Through-bond correlation of sugar and base protons in unlabeled nucleic acids. *J. Magn. Reson.* **153**, 223–6 (2001).
597. Adrian, M., Winnerdy, F. R., Heddi, B. & Phan, A. T. Rotation of Guanine Amino Groups in G-Quadruplexes: A Probe for Local Structure and Ligand Binding. *Biophys. J.* **113**, 775–784 (2017).
598. Schnieders, R. *et al.* Novel 13C-detected NMR Experiments for the Precise Detection of RNA Structure. *Angew. Chem. Int. Ed.* **58**, 9140–9144 (2019).
599. Bothe, J. R. *et al.* Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat. Methods* **8**, 919–931 (2011).
600. Al-Hashimi, H. M. NMR studies of nucleic acid dynamics. *J. Magn. Reson.* **237**, 191–204 (2013).
601. Zhao, B. & Zhang, Q. Characterizing excited conformational states of RNA by NMR spectroscopy. *Curr. Opin. Struct. Biol.* **30**, 134–146 (2015).
602. Rinnenthal, J. *et al.* Mapping the Landscape of RNA Dynamics with NMR Spectroscopy. *Acc. Chem. Res.* **44**, 1292–1301 (2011).
603. Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).
604. Markwick, P. R. L., Malliavin, T. & Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput. Biol.* **4**, e1000168 (2008).
605. Marušič, M., Schlagnitweit, J. & Petzold, K. RNA Dynamics by NMR Spectroscopy. *ChemBioChem* **20**, 2685–2710 (2019).
606. Gáspári, Z. & Perczel, A. Protein Dynamics as Reported by NMR. in *Annual Reports on NMR Spectroscopy* 35–75 (2010). doi:10.1016/B978-0-08-089054-8.00002-2
607. Reich, H. NMR spectroscopy. *Chem 60S, University of Wisconsin - Madison* chapter 8, Relaxation in NMR spectroscopy (2020). Available at: <https://organicchemistrydata.org/hansreich/resources/nmr/>. (Accessed: 4th March 2021)
608. Bloembergen, N., Purcell, E. M. & Pound, R. V. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Phys. Rev.* **73**, 679–712 (1948).
609. Levitt, M. H. Spin dynamics: basics of nuclear magnetic resonance. *John Wiley Sons, LTD* **1. edition**, chapter 16 Relaxation, pp. 513–537, ISBN: 04714892 (2001).
610. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).
611. Akke, M., Fiala, R., Jiang, F., Patel, D. & Palmer, A. G. Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. *RNA* **3**, 702–9 (1997).
612. Shajani, Z. & Varani, G. 13C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.* **349**, 699–715 (2005).
613. Ban, D., Sabo, T., Griesinger, C. & Lee, D. Measuring Dynamic and Kinetic Information in the Previously Inaccessible Supra-tc Window of Nanoseconds to Microseconds by Solution NMR Spectroscopy. *Molecules* **18**, 11904–11937 (2013).
614. Tolman, J. R. & Ruan, K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chem. Rev.* **106**, 1720–36 (2006).

References

615. Zhang, Q. & Al-Hashimi, H. M. Domain-elongation NMR spectroscopy yields new insights into RNA dynamics and adaptive recognition. *RNA* **15**, 1941–1948 (2009).
616. Chiarparin, E., Rüdiger, S. & Bodenhausen, G. Hydrogen Bonds in RNA Base Pairs Investigated by Cross-Correlated Relaxation of Multiple-Quantum Coherence in NMR. *ChemPhysChem* **2**, 41–45 (2001).
617. Vögeli, B. Comprehensive description of NMR cross-correlated relaxation under anisotropic molecular tumbling and correlated local dynamics on all time scales. *J. Chem. Phys.* **133**, 014501 (2010).
618. Brutscher, B. Principles and applications of cross-correlated relaxation in biomolecules. *Concepts Magn. Reson.* **12**, 207–229 (2000).
619. Carr, H. Y. & Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **94**, 630–638 (1954).
620. Meiboom, S. & Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **29**, 688–691 (1958).
621. Juen, M. A. *et al.* Excited States of Nucleic Acids Probed by Proton Relaxation Dispersion NMR Spectroscopy. *Angew. Chem. Int. Ed.* **55**, 12008–12 (2016).
622. Forsén, S. & Hoffman, R. A. Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance. *J. Chem. Phys.* **39**, 2892–2901 (1963).
623. Vallurupalli, P., Sekhar, A., Yuwen, T. & Kay, L. E. Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. *J. Biomol. NMR* **67**, 243–271 (2017).
624. Wenter, P., Bodenhausen, G., Dittmer, J. & Pitsch, S. Kinetics of RNA refolding in dynamic equilibrium by ¹H-detected ¹⁵N exchange NMR spectroscopy. *J. Am. Chem. Soc.* **128**, 7579–87 (2006).
625. Kloiber, K., Spitzer, R., Tollinger, M., Konrat, R. & Kreutz, C. Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive ¹³C-methyl label. *Nucleic Acids Res.* **39**, 4340–51 (2011).
626. Van Nuland, N. A. J., Forge, V., Balbach, J. & Dobson, C. M. Real-Time NMR Studies of Protein Folding. *Acc. Chem. Res.* **31**, 773–780 (1998).
627. Fürtig, B. *et al.* Time-resolved NMR studies of RNA folding. *Biopolymers* **86**, 360–383 (2007).
628. Pintér, G. *et al.* Real-time NMR spectroscopy in the study of biomolecular kinetics and dynamics. *Magn. Reson.* preprint (2021). doi:10.5194/mr-2021-16
629. Rennella, E. & Brutscher, B. Fast Real-Time NMR Methods for Characterizing Short-Lived Molecular States. *ChemPhysChem* **14**, 3059–3070 (2013).
630. Favier, A. & Brutscher, B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR* **49**, 9–15 (2011).
631. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* (2005). doi:10.1021/ja051306e
632. Schanda, P., Kupče, Ě. & Brutscher, B. SOFAST-HMQC Experiments for Recording Two-dimensional Heteronuclear Correlation Spectra of Proteins within a Few Seconds. *J. Biomol. NMR* **33**, 199–211 (2005).
633. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.* **128**, 9042–9043 (2006).
634. Gouilleux, B., Rouger, L. & Giraudeau, P. Ultrafast 2D NMR: Methods and Applications. in *Annual Reports on NMR Spectroscopy* 75–144 (2018). doi:10.1016/bs.arnmr.2017.08.003
635. Giraudeau, P. & Frydman, L. Ultrafast 2D NMR: An Emerging Tool in Analytical Spectroscopy. *Annu. Rev. Anal. Chem.* **7**, 129–161 (2014).
636. Harris, T., Szekely, O. & Frydman, L. On the Potential of Hyperpolarized Water in Biomolecular NMR Studies. *J. Phys. Chem. B* **118**, 3281–3290 (2014).
637. Olsen, G., Markhasin, E., Szekely, O., Bretschneider, C. & Frydman, L. Optimizing water hyperpolarization and dissolution for sensitivity-enhanced 2D biomolecular NMR. *J. Magn. Reson.* **264**, 49–58 (2016).
638. Szekely, O., Olsen, G. L., Felli, I. C. & Frydman, L. High-Resolution 2D NMR of Disordered Proteins Enhanced by Hyperpolarized Water. *Anal. Chem.* **90**, 6169–6177 (2018).
639. Szekely, O., Olsen, G. L., Novakovic, M., Rosenzweig, R. & Frydman, L. Assessing Site-Specific Enhancements Imparted by Hyperpolarized Water in Folded and Unfolded Proteins by 2D HMQC NMR. *J. Am. Chem. Soc.* **142**, 9267–9284 (2020).
640. Olsen, G. L. *et al.* Sensitivity-enhanced three-dimensional and carbon-detected two-dimensional NMR of proteins using hyperpolarized water. *J. Biomol. NMR* **74**, 161–171 (2020).
641. Novakovic, M. *et al.* A 300-fold enhancement of imino nucleic acid resonances by hyperpolarized water provides a new window for probing RNA refolding by 1D and 2D NMR. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 2449–2455 (2020).
642. Mishkovsky, M. & Frydman, L. Progress in Hyperpolarized Ultrafast 2D NMR Spectroscopy. *ChemPhysChem* **9**, 2340–2348 (2008).
643. Lee, M.-K., Gal, M., Frydman, L. & Varani, G. Real-time multidimensional NMR follows RNA folding with second resolution. *Proc. Natl. Acad. Sci.* **107**, 9192–9197 (2010).
644. Gal, M., Mishkovsky, M. & Frydman, L. Real-Time Monitoring of Chemical Transformations by Ultrafast 2D NMR Spectroscopy. *J. Am. Chem. Soc.* **128**, 951–956 (2006).
645. Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negroni, A. & Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* (2012). doi:10.1038/nature11498
646. Kimsey, I. J. *et al.* Dynamic basis for dG•dT misincorporation via tautomerization and ionization. *Nature* **554**, 195–201 (2018).
647. Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).
648. Yoo, J., Louis, J. M., Gopich, I. V. & Chung, H. S. Three-Color Single-Molecule FRET and Fluorescence Lifetime Analysis of Fast Protein Folding. *J. Phys. Chem. B* **122**, 11702–11720 (2018).
649. Wirth, A. J., Liu, Y., Prigozhin, M. B., Schulten, K. & Gruebele, M. Comparing Fast Pressure Jump and Temperature Jump Protein Folding Experiments and Simulations. *J. Am. Chem. Soc.* **137**, 7152–7159 (2015).
650. Jung, A. High-temperature solution NMR structure of TmCsp. *Protein Sci.* **13**, 342–350 (2004).
651. Kachel, N., Kremer, W., Zahn, R. & Kalbitzer, H. R. Observation of intermediate states of the human prion protein by high pressure NMR spectroscopy. *BMC Struct. Biol.* **6**, 16 (2006).
652. Dreydoppel, M. *et al.* Equilibrium and Kinetic Unfolding of GB1: Stabilization of the Native State by Pressure. *J. Phys. Chem. B* **122**, 8846–8852 (2018).
653. Cavini, I. A. *et al.* Inhibition of amyloid A β aggregation by high pressures or specific d-enantiomeric peptides. *Chem. Commun. (Camb.)* **54**, 3294–3297 (2018).
654. Kamatari, Y. O., Kitahara, R., Yamada, H., Yokoyama, S. & Akasaka, K. High-pressure NMR spectroscopy for characterizing folding intermediates and denatured states of proteins. *Methods* **34**, 133–43 (2004).
655. Kalbitzer, H. R. High pressure NMR methods for characterizing functional substates of proteins. *Subcell. Biochem.* **72**, 179–197 (2015).

References

656. Akasaka, K. High pressure NMR spectroscopy. *Subcell. Biochem.* **72**, 707–721 (2015).
657. Caro, J. A. & Wand, A. J. Practical aspects of high-pressure NMR spectroscopy and its applications in protein biophysics and structural biology. *Methods* **148**, 67–80 (2018).
658. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814 (2009).
659. Roche, J., Ying, J., Maltsev, A. S. & Bax, A. Impact of Hydrostatic Pressure on an Intrinsically Disordered Protein: A High-Pressure NMR Study of α -Synuclein. *ChemBioChem* **14**, 1754–1761 (2013).
660. Kalbitzer, H. R. *et al.* Intrinsic allosteric inhibition of signaling proteins by targeting rare interaction states detected by high-pressure NMR spectroscopy. *Angew. Chem. Int. Ed.* **52**, 14242–6 (2013).
661. Naito, A., Nakatani, H., Imanari, M. & Akasaka, K. State-correlated two-dimensional NMR spectroscopy. *J. Magn. Reson.* **87**, 429–432 (1990).
662. Kawakami, M. & Akasaka, K. Microwave temperature-jump nuclear magnetic resonance system for aqueous solutions. *Rev. Sci. Instrum.* **69**, 3365–3369 (1998).
663. Ferguson, D. B., Krawietz, T. R. & Haw, J. F. Temperature-Jump MAS NMR with a Laser Heater. *J. Magn. Reson. Ser. A* **109**, 273–275 (1994).
664. Mildner, T., Ernst, H., Freude, D. & Hölderich, W. F. Temperature-Jump Relaxation Magic-Angle Spinning NMR Study of the Methyl tert-Butyl Ether Reaction in a Boron Pentasil Zeolite. *J. Am. Chem. Soc.* **119**, 4258–4262 (1997).
665. Yushmanov, P. V. & Furó, I. A temperature-jump design for conventional NMR probes. *J. Magn. Reson.* **181**, 148–153 (2006).
666. Rinnenthal, J. *et al.* A temperature-jump NMR probe setup using rf heating optimized for the analysis of temperature-induced biomacromolecular kinetic processes. *J. Magn. Reson.* **251**, 84–93 (2015).
667. Gal, M., Zibzener, K. & Frydman, L. A capacitively coupled temperature-jump arrangement for high-resolution biomolecular NMR. *Magn. Reson. Chem.* **48**, 842–847 (2010).
668. Reinhard Arnold, M., Robert Kalbitzer, H. & Kremer, W. High-sensitivity sapphire cells for high pressure NMR spectroscopy on proteins. *J. Magn. Reson.* **161**, 127–131 (2003).
669. Charlier, C. *et al.* Study of protein folding under native conditions by rapidly switching the hydrostatic pressure inside an NMR sample cell. *Proc. Natl. Acad. Sci.* **115**, E4169–E4178 (2018).
670. Heuert, U., Krumova, M., Hempel, G., Schiewek, M. & Blume, A. NMR probe for pressure-jump experiments up to 250 bars and 3 ms jump time. *Rev. Sci. Instrum.* **81**, 105102 (2010).
671. Kremer, W. *et al.* Pulsed Pressure Perturbations, an Extra Dimension in NMR Spectroscopy of Proteins. *J. Am. Chem. Soc.* **133**, 13646–13651 (2011).
672. Wu, W.-J., Vidugiris, G., Mooberry, E. S., Westler, W. M. & Markley, J. L. Mixing apparatus for preparing NMR samples under pressure. *J. Magn. Reson.* **164**, 84–91 (2003).
673. Peterson, R. W. & Wand, A. J. Self-contained high-pressure cell, apparatus, and procedure for the preparation of encapsulated proteins dissolved in low viscosity fluids for nuclear magnetic resonance spectroscopy. *Rev. Sci. Instrum.* **76**, 094101 (2005).
674. Akasaka, K., Naito, A. & Nakatani, H. Temperature-jump NMR study of protein folding: Ribonuclease A at low pH. *J. Biomol. NMR* **1**, 65–70 (1991).
675. Pintér, G. & Schwalbe, H. Refolding of Cold-Denatured Barstar Induced by Radio-Frequency Heating: A New Method to Study Protein Folding by Real-Time NMR Spectroscopy. *Angew. Chem. Int. Ed.* **59**, 22086–22091 (2020).
676. Alderson, T. R., Charlier, C., Torchia, D. A., Anfinrud, P. & Bax, A. Monitoring Hydrogen Exchange During Protein Folding by Fast Pressure Jump NMR Spectroscopy. *J. Am. Chem. Soc.* **139**, (2017).
677. Charlier, C., Courtney, J. M., Anfinrud, P. & Bax, A. Interrupted Pressure-Jump NMR Experiments Reveal Resonances of On-Pathway Protein Folding Intermediate. *J. Phys. Chem. B* **122**, 11792–11799 (2018).
678. Barnes, C. A., Robertson, A. J., Louis, J. M., Anfinrud, P. & Bax, A. Observation of β -Amyloid Peptide Oligomerization by Pressure-Jump NMR Spectroscopy. *J. Am. Chem. Soc.* **141**, 13762–13766 (2019).
679. Kamatari, Y. O., Yokoyama, S., Tachibana, H. & Akasaka, K. Pressure-jump NMR Study of Dissociation and Association of Amyloid Protofibrils. *J. Mol. Biol.* **349**, 916–921 (2005).
680. Munte, C. E., Beck Erlach, M., Kremer, W., Koehler, J. & Kalbitzer, H. R. Distinct conformational states of the Alzheimer β -amyloid peptide can be detected by high-pressure NMR spectroscopy. *Angew. Chem. Int. Ed.* **52**, 8943–7 (2013).
681. RÖDER, H. Rapid mixing methods for exploring the kinetics of protein folding. *Methods* **34**, 15–27 (2004).
682. Yushmanov, P. V. & Furó, I. A rapid-mixing design for conventional NMR probes. *J. Magn. Reson.* **175**, 264–270 (2005).
683. Mok, K. H. *et al.* Rapid sample-mixing technique for transient NMR and photo-CIDNP spectroscopy: Applications to real-time protein folding. *J. Am. Chem. Soc.* **125**, 12484–12492 (2003).
684. Franco, R., Favier, A., Schanda, P. & Brutscher, B. Optimized fast mixing device for real-time NMR applications. *J. Magn. Reson.* **281**, 125–129 (2017).
685. Balbach, J. *et al.* Following protein folding in real time using NMR spectroscopy. *Nat. Struct. Biol.* **2**, 865–870 (1995).
686. Kiefhaber, T., Labhardt, A. M. & Baldwin, R. L. Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* **375**, 513–515 (1995).
687. Balbach, J. *et al.* Protein Folding Monitored at Individual Residues During a Two-Dimensional NMR Experiment. *Science* **274**, 1161–1163 (1996).
688. Corazza, A. *et al.* Native-unlike long-lived intermediates along the folding pathway of the amyloidogenic protein β 2-microglobulin revealed by real-time two-dimensional NMR. *J. Biol. Chem.* **285**, 5827–5835 (2010).
689. ZEEB, M. Protein folding studied by real-time NMR spectroscopy. *Methods* **34**, 65–74 (2004).
690. Dobson, C. M. & Hore, P. J. Kinetic studies of protein folding using NMR spectroscopy. *Nat. Struct. Biol.* **5**, 504–507 (1998).
691. Schanda, P., Forge, V. & Brutscher, B. Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc. Natl. Acad. Sci.* **104**, 11257–11262 (2007).
692. Reining, A. *et al.* Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499**, 355–9 (2013).
693. Manoharan, V., Fürtig, B., Jäschke, A. & Schwalbe, H. Metal-induced folding of diels-alderase ribozymes studied by static and time-resolved NMR spectroscopy. *J. Am. Chem. Soc.* **131**, 6261–6270 (2009).
694. Barraud, P. *et al.* Time-resolved NMR monitoring of tRNA maturation. *Nat. Commun.* **10**, 3373 (2019).
695. Buck, J., Fürtig, B., Noeske, J., Wohnert, J. & Schwalbe, H. Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution. *Proc. Natl. Acad. Sci.* **104**, 15699–15704 (2007).
696. Kühn, T. & Schwalbe, H. Monitoring the kinetics of ion-dependent protein folding by time-resolved NMR spectroscopy at atomic resolution. *J. Am. Chem. Soc.* **122**, 6169–6174 (2000).

References

697. Wirmer, J., Kühn, T. & Schwalbe, H. Millisecond Time Resolved Photo-CIDNP NMR Reveals a Non-Native Folding Intermediate on the Ion-Induced Refolding Pathway of Bovine α -Lactalbumin. *Angew. Chem. Int. Ed.* **40**, 4248–4251 (2001).
698. Mayer, G. & Heckel, A. Biologically active molecules with a 'light switch'. *Angew. Chem. Int. Ed.* **45**, 4900–21 (2006).
699. Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G. & Heckel, A. Light-Controlled Tools. *Angew. Chem. Int. Ed.* **51**, 8446–8476 (2012).
700. Becker, Y. *et al.* A red-shifted two-photon-only caging group for three-dimensional photorelease. *Chem. Sci.* **9**, 2797–2802 (2018).
701. Becker, Y. *et al.* Selective Modification for Red-Shifted Excitability: A Small Change in Structure, a Huge Change in Photochemistry. *Chem. – A Eur. J.* **27**, 2212–2218 (2021).
702. Griffiths, J. II. Photochemistry of azobenzene and its derivatives. *Chem. Soc. Rev.* **1**, 481 (1972).
703. Yamana, K., Yoshikawa, A. & Nakano, H. Synthesis of a new photoisomerizable linker for connecting two oligonucleotide segments. *Tetrahedron Lett.* **37**, 637–640 (1996).
704. Asanuma, H., Ito, T. & Komiyama, M. Photo-responsive oligonucleotides carrying azobenzene in the side-chains. *Tetrahedron Lett.* **39**, 9015–9018 (1998).
705. Asanuma, Ito, Yoshida, Liang & Komiyama. Photoregulation of the Formation and Dissociation of a DNA Duplex by Using the cis-trans Isomerization of Azobenzene. *Angew. Chem. Int. Ed.* **38**, 2393–2395 (1999).
706. Bartrop, J. A. & Schofield, P. Photosensitive Protecting Groups. *Tetrahedron Lett.* **3**, 697–699 (1962).
707. Engels, J. & Schlaeger, E. J. Synthesis, Structure, and Reactivity of Adenosine Cyclic 3',5'-Phosphate-Benzyl Triesters. *J. Med. Chem.* **20**, 907–911 (1977).
708. Monroe, W. T., McQuain, M. M., Chang, M. S., Alexander, J. S. & Haselton, F. R. Targeting Expression with Light Using Caged DNA. *J. Biol. Chem.* **274**, 20895–20900 (1999).
709. Ando, H., Furuta, T., Tsieng, R. Y. & Okamoto, H. Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat. Genet.* **28**, 317–325 (2001).
710. Shah, S., Rangarajan, S. & Friedman, S. H. Light-activated RNA interference. *Angew. Chem. Int. Ed.* **44**, 1328–32 (2005).
711. Ting, R., Lermer, L. & Perrin, D. M. Triggering DNazymes with Light: A Photoactive C8 Thioether-Linked Adenosine. *J. Am. Chem. Soc.* **126**, 12720–12721 (2004).
712. Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H. & Pitsch, S. Kinetics of photoinduced RNA refolding by real-time NMR spectroscopy. *Angew. Chem. Int. Ed.* **44**, 2600–3 (2005).
713. Höbartner, C. & Silverman, S. K. Modulation of RNA tertiary folding by incorporation of caged nucleotides. *Angew. Chem. Int. Ed.* **44**, 7305–9 (2005).
714. Heckel, A. *et al.* An Anticoagulant with Light-Triggered Antidote Activity. *Angew. Chem. Int. Ed.* **45**, 6748–6750 (2006).
715. Kröck, L. & Heckel, A. Photoinduced transcription by using temporarily mismatched caged oligonucleotides. *Angew. Chem. Int. Ed.* **44**, 471–3 (2005).
716. Mayer, G., Kröck, L., Mikat, V., Engeser, M. & Heckel, A. Light-Induced Formation of G-Quadruplex DNA Secondary Structures. *ChemBioChem* **6**, 1966–1970 (2005).
717. Chaulk, S. Caged RNA: photo-control of a ribozyme reaction. *Nucleic Acids Res.* **26**, 3173–3178 (1998).
718. Chaulk, S. G. & MacMillan, A. M. Synthesis of oligo-RNAs with photocaged adenosine 2'-hydroxyls. *Nat. Protoc.* **2**, 1052–1058 (2007).
719. Zheng, G., Cochella, L., Liu, J., Hobert, O. & Li, W. Temporal and Spatial Regulation of MicroRNA Activity with Photoactivatable Cantimirs. *ACS Chem. Biol.* **6**, 1332–1338 (2011).
720. Griepenburg, J. C., Ruble, B. K. & Dmochowski, I. J. Caged oligonucleotides for bidirectional photomodulation of let-7 miRNA in zebrafish embryos. *Bioorg. Med. Chem.* **21**, 6198–6204 (2013).
721. Seyfried, P., Eiden, L., Grebenovsky, N., Mayer, G. & Heckel, A. Photo-Tethers for the (Multi-)Cyclic, Conformational Caging of Long Oligonucleotides. *Angew. Chem. Int. Ed.* **56**, 359–363 (2017).
722. Kaplan, J. H., Forbush, B. & Hoffman, J. F. Rapid photolytic release of adenosine 5'-triphosphate from a protected analog: utilization by the sodium:potassium pump of human red blood cell ghosts. *Biochemistry* **17**, 1929–1935 (1978).
723. Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H. & Pitsch, S. A Caged Uridine for the Selective Preparation of an RNA Fold and Determination of its Refolding Kinetics by Real-Time NMR. *ChemBioChem* **7**, 417–420 (2006).
724. Rodrigues-Correia, A. *et al.* Comparison of the duplex-destabilizing effects of nucleobase-caged oligonucleotides. *Anal. Bioanal. Chem.* **399**, 441–447 (2011).
725. Steinert, H. S., Schäfer, F., Jonker, H. R. A., Heckel, A. & Schwalbe, H. Influence of the absolute configuration of npe-caged cytosine on DNA single base pair stability. *Angew. Chem. Int. Ed.* **53**, 1072–5 (2014).
726. Seyfried, P. *et al.* Optimal Destabilization of DNA Double Strands by Single-Nucleobase Caging. *Chem. Eur. J.* **24**, 17568–17576 (2018).
727. Fürtig, B. *et al.* Conformational dynamics of bistable RNAs studied by time-resolved NMR spectroscopy. *J. Am. Chem. Soc.* **129**, 16222–16229 (2007).
728. Fürtig, B., Wenter, P., Pitsch, S. & Schwalbe, H. Probing Mechanism and Transition State of RNA Refolding. *ACS Chem. Biol.* **5**, 753–765 (2010).
729. Cao, S., Fürtig, B., Schwalbe, H. & Chen, S.-J. Folding Kinetics for the Conformational Switch between Alternative RNA Structures. *J. Phys. Chem. B* **114**, 13609–13615 (2010).
730. Höbartner, C., Ebert, M.-O., Jaun, B. & Micura, R. RNA Two-State Conformation Equilibria and the Effect of Nucleobase Methylation. *Angew. Chem. Int. Ed.* **41**, 605–609 (2002).
731. Höbartner, C., Mittendorfer, H., Breuker, K. & Micura, R. Triggering of RNA secondary structures by a functionalized nucleobase. *Angew. Chem. Int. Ed.* **43**, 3922–5 (2004).
732. Grün, J. T. *et al.* Conformational Dynamics of Strand Register Shifts in DNA G-Quadruplexes. *J. Am. Chem. Soc.* **142**, 264–273 (2020).
733. Grün, J. T. *et al.* Unraveling the Kinetics of Spare-Tire DNA G-Quadruplex Folding. *J. Am. Chem. Soc.* accepted manuscript, (published online) (2021). doi:https://doi.org/10.1021/jacs.1c01089
734. Georg, I. *et al.* Exhaustively Trichlorosilylated C 1 and C 2 Building Blocks: Beyond the Müller–Rochow Direct Process. *J. Am. Chem. Soc.* **140**, 9696–9708 (2018).
735. Georg, I. *et al.* Building up Strain in One Step: Synthesis of an Edge-Fused Double Silacyclobutene from an Extensively Trichlorosilylated Butadiene Dianion. *Angew. Chem. Int. Ed.* **59**, 16181–16187 (2020).
736. Helmling, C. *et al.* Life times of metastable states guide regulatory signaling in transcriptional riboswitches. *Nat. Commun.* **9**, 944 (2018).
737. Harkness V, R. W., Avakyan, N., Sleiman, H. F. & Mittermaier, A. K. Mapping the energy landscapes of supramolecular assembly by thermal hysteresis. *Nat. Commun.* **9**, 3152 (2018).

References

738. Wang, Z., Chen, R., Hou, L., Li, J. & Liu, J.-P. Molecular dynamics and principal components of potassium binding with human telomeric intramolecular G-quadruplex. *Protein Cell* **6**, 423–433 (2015).
739. Balthasar, F., Plavec, J. & Gabelica, V. Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? *J. Am. Soc. Mass Spectrom.* **24**, 1–8 (2013).
740. Hud, N. V., Schultze, P. & Feigon, J. Ammonium Ion as an NMR Probe for Monovalent Cation Coordination Sites of DNA Quadruplexes. *J. Am. Chem. Soc.* **120**, 6403–6404 (1998).
741. Hud, N. V., Schultze, P., Sklenár, V. & Feigon, J. Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. *J. Mol. Biol.* **285**, 233–43 (1999).
742. Šket, P., Črnugelj, M. & Plavec, J. d(G3T4G4) forms unusual dimeric G-quadruplex structure with the same general fold in the presence of K⁺, Na⁺ or NH₄⁺ ions. *Bioorg. Med. Chem.* **12**, 5735–5744 (2004).
743. Sket, P., Črnugelj, M., Kozminski, W. & Plavec, J. 15NH₄⁺ ion movement inside d(G4T4G4)₂ G-quadruplex is accelerated in the presence of smaller Na⁺ ions. *Org. Biomol. Chem.* **2**, 1970–3 (2004).
744. Sket, P. & Plavec, J. Not all G-quadruplexes exhibit ion-channel-like properties: NMR study of ammonium ion (non)movement within the d(G(3)T(4)G(4))(2) quadruplex. *J. Am. Chem. Soc.* **129**, 8794–800 (2007).
745. Podbevsek, P., Sket, P. & Plavec, J. Stacking and not solely topology of T3 loops controls rigidity and ammonium ion movement within d(G4T3G4)₂ G-quadruplex. *J. Am. Chem. Soc.* **130**, 14287–93 (2008).
746. Trajkovski, M., Šket, P. & Plavec, J. Cation localization and movement within DNA thrombin binding aptamer in solution. *Org. Biomol. Chem.* **7**, 4677 (2009).
747. Trajkovski, M. & Plavec, J. Assessing Roles of Cations in G-Quadruplex-Based Nanowires by NMR. *J. Phys. Chem. C* **116**, 23821–23825 (2012).
748. Wong, A., Ida, R. & Wu, G. Direct NMR detection of the “invisible” alkali metal cations tightly bound to G-quadruplex structures. *Biochem. Biophys. Res. Commun.* **337**, 363–366 (2005).
749. Ida, R. & Wu, G. Direct NMR Detection of Alkali Metal Ions Bound to G-Quadruplex DNA. *J. Am. Chem. Soc.* **130**, 3590–3602 (2008).
750. Fiala, R. *et al.* NMR Cross-Correlated Relaxation Rates Reveal Ion Coordination Sites in DNA. *J. Am. Chem. Soc.* **133**, 13790–13793 (2011).
751. Sket, P. Identification of mixed di-cation forms of G-quadruplex in solution. *Nucleic Acids Res.* **33**, 3691–3697 (2005).
752. Harkness, R. W. *et al.* Parallel reaction pathways accelerate folding of a guanine quadruplex. *Nucleic Acids Res.* **49**, 1247–1262 (2021).
753. Alberty, R. A. Principle of Detailed Balance in Kinetics. *J. Chem. Educ.* **81**, 1206 (2004).
754. Wright, C. F., Lindorff-Larsen, K., Randles, L. G. & Clarke, J. Parallel protein-unfolding pathways revealed and mapped. *Nat. Struct. Mol. Biol.* **10**, 658–662 (2003).
755. Dobson, C. M., Šali, A. & Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem. Int. Ed.* **37**, 868–893 (1998).
756. Baldwin, R. L. The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR* **5**, 103–9 (1995).
757. Wildegger, G. & Kiefhaber, T. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **270**, 294–304 (1997).
758. Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M. & Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339 (2000).
759. Zaidi, F. N., Nath, U. & Udgaonkar, J. B. Multiple intermediates and transition states during protein unfolding. *Nat. Struct. Biol.* **4**, 1016–1024 (1997).
760. Bonnet, G., Krichevsky, O. & Libchaber, A. Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl. Acad. Sci.* **95**, 8602–8606 (1998).
761. Ansari, A., Kuznetsov, S. V. & Shen, Y. Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc. Natl. Acad. Sci.* **98**, 7771–7776 (2001).
762. Wallace, M. I., Ying, L., Balasubramanian, S. & Klenerman, D. Non-Arrhenius kinetics for the loop closure of a DNA hairpin. *Proc. Natl. Acad. Sci.* **98**, 5584–5589 (2001).
763. Zhang, W. & Chen, S.-J. Exploring the Complex Folding Kinetics of RNA Hairpins: I. General Folding Kinetics Analysis. *Biophys. J.* **90**, 765–777 (2006).
764. Zhang, W. & Chen, S.-J. Exploring the Complex Folding Kinetics of RNA Hairpins: II. Effect of Sequence, Length, and Misfolded States. *Biophys. J.* **90**, 778–787 (2006).
765. Kuznetsov, S. V., Ren, C.-C., Woodson, S. A. & Ansari, A. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* **36**, 1098–1112 (2007).
766. Chen, S.-J. RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. *Annu. Rev. Biophys.* **37**, 197–214 (2008).
767. Črnugelj, M., Šket, P. & Plavec, J. Small Change in a G-Rich Sequence, a Dramatic Change in Topology: New Dimeric G-Quadruplex Folding Motif with Unique Loop Orientations. *J. Am. Chem. Soc.* **125**, 7866–7871 (2003).
768. Novakovic, M. *et al.* Sensitivity enhancement of homonuclear multidimensional NMR correlations for labile sites in proteins, polysaccharides, and nucleic acids. *Nat. Commun.* **11**, 5317 (2020).

Appendix

6.1 HMBC and Deuterium Exchange NMR Spectrum

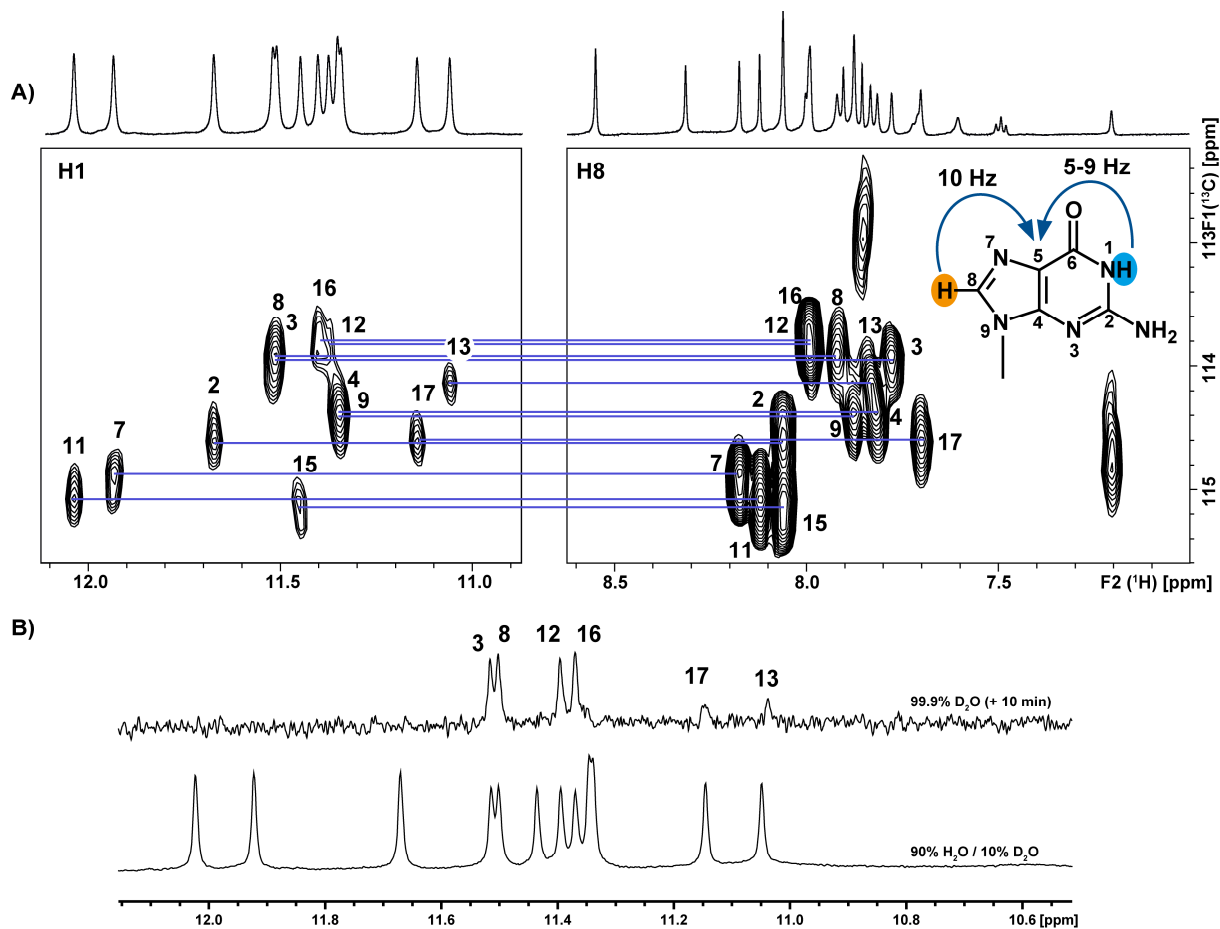


Figure 66: Representative NMR spectra of photocaged cMYC-2345-53 18-mer for H1-H8 correlations and hydrogen-deuterium exchange. **A)** 500 μM DNA, 90/10 H₂O/D₂O (298 K), 600 MHz (WW), HMBC (hmbcetgpl3nd.ric) with jump-return-echo water suppression, $d19 = 55 \mu\text{s}$, $J(\text{XH}) = 17 \text{ Hz}$; **B)** 100 μM DNA (298 K), 700 MHz, jump-return echo water suppression (hs11echo). Referenced to DSS.

6.2 Plots of Time-Resolved 1D ^1H NMR Spectra for *Spare-Tire* G4 Folding

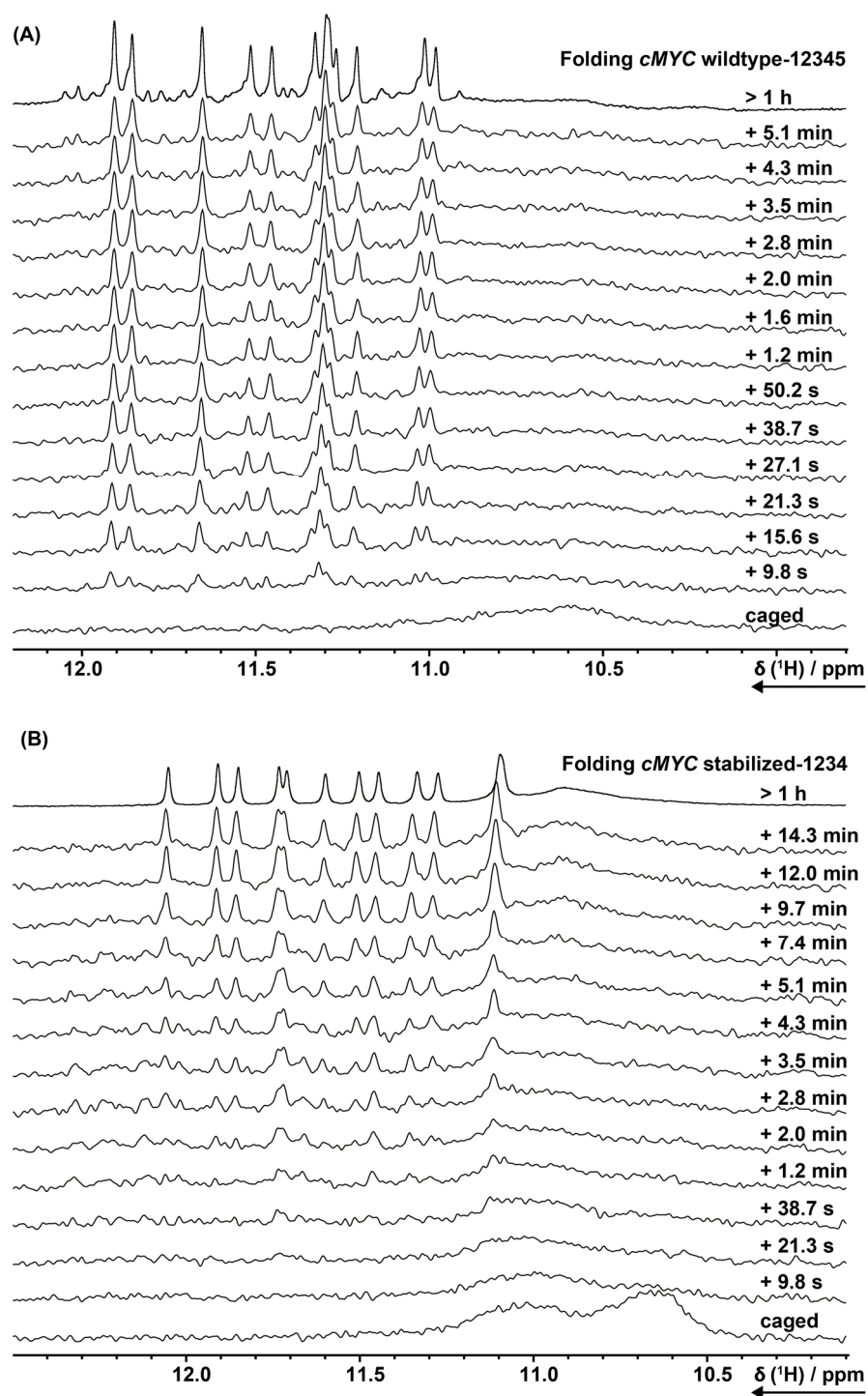


Figure 67: Plots of ^1H imino proton regions of time-resolved NMR spectra (700 MHz, jump-return-echo water suppression (hs1echo, d19 = 50 μs) at indicated times after light-induced folding initiation. (A) Folding of wt-12345 at 298 K. (B) Folding of stab.-1234 at 285 K.

6.3 Python Script for Kinetic Fits

```

1  __author__ = 'Dominik Brey'
2  # Used for plotting
3  from matplotlib.pyplot import plot, show, subplot, savefig, title, clf
4  # Package for all stuff mathematical
5  import numpy as np
6  # numerical solving of differential equations
7  from scipy.integrate import odeint
8  # optimization modules
9  from scipy.optimize import differential_evolution, minimize
10 # used for system specific stuff (commandline arguments, de leting offiles)
11 import sys
12
13 # read text file as commandline argument
14 # textFile = sys.argv[1]
15
16 # alternatively use name of textfile to be read; has to be in the same
17 # directory or full path
18 # must be specified
19 # the textfile has to be in the following form:
20 # x tab y1 tab y2
21 textFile = "1234.txt"
22 # use True if error -estimation should be done, otherwise False
23 doError = False
24 # use True if progress of optimization should be printed on screen
25 # use with caution as windows 10 seems to have a bug with unicode characters
26 # (march 2018) and the
27 # script might just die without warning
28 # should be fine on UNIX -systems
29 displayProgress = True
30
31 # opens the specified textfile in reading -mode
32 with open(textFile, "r") as file:
33     # file is read completely and all , are replaced with .decimal separator
34     # should be . for
35     # python
36     txt = file.read().replace(", ", ".")
37     # lines are split by linebreak-character; essentially separates data points
38     txtSplit = txt.split("\n")
39     # removes empty lines because they cause problems
40     if "" in txtSplit:
41         txtSplit.remove("")
42     # lists for x and y data
43     freq = []
44     freq2 = []
45     int1 = []
46     int2 = []
47
48     # each line in the textfile is split by tabstops and the first entry put
49     # into x-data,
50     # the second to y-data and the third to ydata2
51     # assignment can be changed arbitrarily, e.g. if the first and second
52     # entry in the file are
53     # both x-data with different scaling
54     for i in txtSplit:
55         tmp = i.split("\t")
56         if not '#' in tmp[1]:
57             freq.append(tmp[0])
58             int1.append(tmp[1])
59         if not '#' in tmp[2]:
60             freq2.append(tmp[0])
61             int2.append(tmp[2])
62
63     # lists are converted to numpy-arrays; makes calculations easier
64     freq = np.array(freq, dtype= float )

```

```

65     freq2 = np.array(freq2, dtype= float )
66     int1 = np.array(int1, dtype= float )
67     int2 = np.array(int2, dtype= float )
68
69     # start value (maximum concentration) is set as the mean of the sum of
70     # the (corresponding
71     # last 200 data points in ydata
72     # u0 = np.mean(int1[-1:100:-1]+int2[-1:100:-1])
73     u0_tmp = []
74     if len(int1) < len(int2):
75         for i in range(len(int1) - 50, len(int1)):
76             a = int1[i]
77             b = int2[np.where(freq2 == freq[i])[0]]
78             iflen (b) > 0:
79                 u0_tmp.append(a + b)
80     else:
81         for i in range(len(int2) - 50, len(int2)):
82             a = int2[i]
83             b = int1[np.where(freq == freq2[i])[0]]
84             iflen (b) > 0:
85                 u0_tmp.append(a + b)
86
87     u0 = np.mean(u0_tmp)
88
89     # normalization by dividing by u0 (normalized data is easier to optimize)
90     int1 = int1 / u0
91     int2 = int2 / u0
92     # u0 = 1
93
94     # sets of differential equations for each system
95     def ode_aub(x, t, k):
96         return (-k[1] * x[0] + k[0] * x[1], # A
97                k[1] * x[0] - (k[0] + k[2]) * x[1] + k[3] * x[2], # U
98                k[2] * x[1] - k[3] * x[2]) # B
99
100     def ode_aub_kreis(x, t, k):
101         return (-k[1] + k[4]) * x[0] + k[0] * x[1] + k[5] * x[2], # A
102                k[1] * x[0] - (k[0] + k[2]) * x[1] + k[3] * x[2], # U
103                k[4] * x[0] + k[2] * x[1] - (k[3] + k[5]) * x[2]) # B
104
105     def ode_u_to_aib_kreis(x, t, k):
106         return (-k[1] + k[4]) * x[0] + k[0] * x[1] + k[5] * x[2], # A
107                k[1] * x[0] - (k[0] + k[2]) * x[1] + k[3] * x[2], # U
108                k[4] * x[0] + k[2] * x[1] - (k[3] + k[5]) * x[2], # B
109                -k[6] * x[3]) # U
110
111     def ode_u_to_aib(x, t, k):
112         return (-k[1] * x[0] + k[0] * x[1], # A
113                k[1] * x[0] - (k[0] + k[2]) * x[1] + k[3] * x[2], # U
114                k[2] * x[1] - k[3] * x[2], # B
115                -k[4] * x[3]) # U
116
117     def ode_u_to_ab_ggw_i(x, t, k):
118         return (-k[1] * x[0] + k[0] * x[1] + k[4] * x[3], # A
119                k[1] * x[0] - (k[0] + k[2]) * x[1] + k[3] * x[2], # U
120                k[2] * x[1] - k[3] * x[2] + k[5] * x[3], # B
121                -(k[4] + k[5]) * x[3]) # U
122
123     def ode_u_to_ab(x, t, k):
124
125     def obj_u_to_ab_ggw_i(p, t1, t2, yData1, yData2):
126         sum = 0
127         for i in 0.5 * (yData1 - model_u_to_ab_ggw_i(p, t1)[: , 0]) ** 2:
128             sum += i
129         for i in (1 - 0.5) * (yData2 - model_u_to_ab_ggw_i(p, t2)[: , 2]) ** 2:
130             sum += i
131         return sum
132
133     def obj_u_to_ab(p, t1, t2, yData1, yData2):
134         sum = 0
135         for i in 0.5 * (yData1 - model_u_to_ab(p, t1)[: , 0]) ** 2:
136             sum += i
137         for i in (1 - 0.5) * (yData2 - model_u_to_ab(p, t2)[: , 1]) ** 2:
138             sum += i
139         return sum
140
141     # optimization through differential evolution
142     print ('Start Optimization of system aub')
143     bounds_aub = (
144         (u0 * 0.99, u0 * 1.01), (5, 250), (5, 250), (5, 250), (0.0, 2.5))
145     res_aub = differential_evolution(obj_aub, args=(freq, freq2, int1, int2),
146                                   bounds=bounds_aub,
147                                   popsize=5, disp=displayProgress)
148     print ('Success in global optimization of system aub: ', res_aub.success,
149           '\n')
150     k_opt_aub_diff = res_aub.x
151
152     print ('Start Optimization of system aub_kreis')
153     bounds_aub_kreis = (
154         (u0 * 0.99, u0 * 1.01), (5, 250), (5, 250), (5, 250), (0.0, 2.5),
155         (0.01, 10), (0.01, 10))
156     res_aub_kreis = differential_evolution(obj_aub_kreis,
157                                         args=(freq, freq2, int1, int2),
158                                         bounds=bounds_aub_kreis, popsize=2,
159                                         disp=displayProgress)
160     print ('Success in global optimization of system aub_kreis: ',
161           res_aub_kreis.success, '\n')
162     k_opt_aub_kreis_diff = res_aub_kreis.x
163
164     print ('Start Optimization of system u_to_aib')
165     bounds_u_to_aib = (
166         (u0 * 0.95, u0 * 1.05), (5, 100), (5, 100), (5, 100), (0.01, 10),
167         (50, 1000))
168     res_u_to_aib = differential_evolution(obj_u_to_aib,
169                                         args=(freq, freq2, int1, int2),
170                                         bounds=bounds_u_to_aib, popsize=2,
171                                         disp=displayProgress)
172     print ('Success in global optimization of system u_to_aib: ',
173           res_u_to_aib.success, '\n')
174     k_opt_u_to_aib_diff = res_u_to_aib.x
175
176     print ('Start Optimization of system u_to_aib_kreis')
177     bounds_u_to_aib_kreis = (
178         (u0 * 0.95, u0 * 1.05), (5, 500), (5, 500), (0.01, 50), (0.01, 50),
179         (50, 5000))
180     res_u_to_aib_kreis = differential_evolution(obj_u_to_aib_kreis,
181                                               args=(freq, freq2, int1, int2),
182                                               bounds=bounds_u_to_aib_kreis,
183                                               popsize=2,
184                                               disp=displayProgress)

```

```

260 print ('Success in global optimization of system u_to_aib_kreis: ',
261       res_u_to_aib_kreis.success,
262       '\n')
263 k_opt_u_to_aib_kreis_diff = res_u_to_aib_kreis.x
264
265 print ('Start Optimization of system u_to_ab_ggw_i' )
266 bounds_u_to_ab_ggw_i = (
267     u0 0.99, u0 1.01), (0.01, 1000), (0.01, 1000), (0.01, 1000), (0.01, 1000),
268     (0.01, 1000), (0.01, 1000),
269     (0.01, 1000))
270 res_u_to_ab_ggw_i = differential_evolution(obj_u_to_a b_ggw_i,
271     args=(freq, freq2, int1, int2.),
272     bounds=bounds_u_to_ab_ggw_i,
273     popsize=2,
274     disp=displayProgress)
275 print ('Success in global optimization of system u_to_ab_ggw_i: ',
276       res_u_to_ab_ggw_i.success,
277       '\n')
278 k_opt_u_to_ab_ggw_i_diff = res_u_to_ab_ggw_i.x
279
280 print ('Start Optimization of system u_to_ab' )
281 bounds_u_to_ab = (
282     u0 0.99, u0 1.01), (5, 100), (0, 1), (5, 50), (5, 50))
283 res_u_to_ab = differential_evolution(obj_u_to_ab,
284     args=(freq, freq2, int1, int2.),
285     bounds=bounds_u_to_ab,
286     popsize=2,
287     disp=displayProgress)
288 print ('Success in global optimization of system u_to_ab: ',
289       res_u_to_ab.success, '\n')
290 k_opt_u_to_ab_diff = res_u_to_ab.x
291
292 =====STATISTIC=====
293 # error approximation by bootstrap algorithm
294 if doError:
295     =====STANDARDERROR=====
296     print ('Start approximating standard error\n')
297
298     =====u_b=====
299     newDataFits_aub = []
300     print ('Standard error for system aub: ' )
301     for i in range (0, 100):
302         newXData1 = []
303         newYData1 = []
304         newXData2 = []
305         newYData2 = []
306         # smallest value in one list of y data; initially set to be the
307         # highest number in y data
308         smallest1 = max(int1)
309         smallest2 = max(int2)
310         # make subsamples the same size as the original sample
311         for index in range (0, len(int1)):
312             # choose a random number between 0 (inclusive) and the number
313             # of elemnts in
314             # ydata1 (exclusive)
315             rand = np.random.randint(0, len(int1))
316             # if value of y data at position rand is smaller than current
317             # smallest value,
318             # set accordingly
319             if (int1[rand] < smallest1):
320                 smallest1 = int1[rand]
321             # there is a problem with numerical solving differential
322             # equations if there are
323             # more than 4 entries with
324             # the same value at the beginning of the sample...
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

520 # choose a random number between 0 (inclusive) and the number
521 # of elemnts in
522 # ydata2 (exclusive)
523 rand = np.random.randint(0, len(int2))
524 # if value of y data at position rand is smaller than current
525 # smallest value,
526 # set accordingly
527 if (int2[rand] < smallest2):
528     smallest2 = int2[rand]
529 # there is a problem with numerical solving differential
530 # equations if there are
531 # more than 4 entries with
532 # the same value at the beginning of the sample...
533 # so of there is already 4 times the same smallest number in
534 # the sample but the
535 # next one would be the same
536 # choose a different one
537 while (newYData2.count(smallest2) >= 4 and int2[
538     rand] == smallest2):
539     rand = np.random.randint(0, len(int2))
540 # corresponding values of x2 and y2 are added to the new lists
541 newXData2.append(freq2[rand])
542 newYData2.append(int2[rand])
543 # the subsample is ordered by x values (smallest to highest)
544 # because numerical solving
545 # of differential equations needs it that way
546 newXData2, newYData2 = zip(sorted(zip(newXData2, newYData2)))
547
548 if (i % 5 == 0):
549     print('.', end=' ', flush=True)
550
551 randomFit = minimize(obj_u_to_aib, k_opt_u_to_aib_diff,
552                     args=(
553                         newXData1, newXData2, newYData1, newYData2),
554                     method='L-BFGS-B', bounds=bounds_u_to_aib)
555 newYDataFits_u_to_aib.append(randomFit.x)
556
557 newYDataFits_u_to_aib = np.array(newYDataFits_u_to_aib)
558 k_opt_u_to_aib = np.mean(newYDataFits_u_to_aib, 0)
559
560 sigmaInterval = 1.0
561
562 err_u_to_aib = sigmaInterval * np.std(newYDataFits_u_to_aib, 0)
563
564 print('\nu_to_aib:\nBootstrap\nMean:', k_opt_u_to_aib, '\nError: ',
565       err_u_to_aib, '\n')
566
567 #####u_to_aib_kreis#####
568 print('Standard error for system u_to_aib_kreis')
569 newYDataFits_u_to_aib_kreis = []
570 for i in range(0, 100):
571     newXData1 = []
572     newXData2 = []
573     newYData1 = []
574     newYData2 = []
575     # smallest value in one list of y data; initially set to be the
576     # highest number in y data
577     smallest1 = max(int1)
578     smallest2 = max(int2)
579     # make subsamples the same size as the original sample
580     for index in range(0, len(int1)):
581         # choose a random number between 0 (inclusive) and the number
582         # of elemnts in
583         # ydata1 (exclusive)
584         rand = np.random.randint(0, len(int1))

```

```

585 # if value of y data at position rand is smaller than current
586 # smallest value
587 # set accordingly
588 if (int1[rand] < smallest1):
589     smallest1 = int1[rand]
590 # there is a problem with numerical solving differential
591 # equations if there are
592 # more than 4 entries with
593 # the same value at the beginning of the sample...
594 # so of there is already 4 times the same smallest number in
595 # the sample but the
596 # next one would be the same
597 # choose a different one
598 while (newYData1.count(smallest1) >= 4 and int1[
599     rand] == smallest1):
600     rand = np.random.randint(0, len(int1))
601 # corresponding values of x and y1 are added to the new lists
602 newXData1.append(freq[rand])
603 newYData1.append(int1[rand])
604 # the subsample is ordered by x values (smallest to highest)
605 # because numerical solving
606 # of differential equations needs it that way
607 newXData1, newYData1 = zip(sorted(zip(newXData1, newYData1)))
608 # same for second data
609 for index in range(0, len(int2)):
610     # choose a random number between 0 (inclusive) and the number
611     # of elements in
612     # ydata2 (exclusive)
613     rand = np.random.randint(0, len(int2))
614     # if value of y data at position rand is smaller than current
615     # smallest value, set
616     # accordingly
617     if (int2[rand] < smallest2):
618         smallest2 = int2[rand]
619     # there is a problem with numerical solving differential
620     # equations if there are
621     # more than 4 entries with
622     # the same value at the beginning of the sample...
623     # so of there is already 4 times the same smallest number in
624     # the sample but the
625     # next one would be the same
626     # choose a different one
627     while (newYData2.count(smallest2) >= 4 and int2[
628         rand] == smallest2):
629         rand = np.random.randint(0, len(int2))
630     # corresponding values of x2 and y2 are added to the new lists
631     newXData2.append(freq2[rand])
632     newYData2.append(int2[rand])
633     # the subsample is ordered by x values (smallest to highest)
634     # because numerical solving
635     # of differential equations needs it that way
636     newXData2, newYData2 = zip(sorted(zip(newXData2, newYData2)))
637
638 if (i % 5 == 0):
639     print('.', end=' ', flush=True)
640
641 randomFit = minimize(obj_u_to_aib_kreis, k_opt_u_to_aib_diff,
642                     args=(
643                         newXData1, newXData2, newYData1, newYData2),
644                     method='L-BFGS-B', bounds=bounds_u_to_aib_kreis)
645 newYDataFits_u_to_aib_kreis.append(randomFit.x)
646
647 newYDataFits_u_to_aib_kreis = np.array(newYDataFits_u_to_aib_kreis)
648 k_opt_u_to_aib_kreis = np.mean(newYDataFits_u_to_aib_kreis, 0)

```

```

590 sigmaInterval = 1.0
591
592 err_u_to_aib_kreis = sigmaInterval * np.std(newYDataFits_u_to_aib_kreis,
593                                             0)
594
595 print('\nu_to_aib_kreis:\nBootstrap\nMean:', k_opt_u_to_aib_kreis,
596       '\nError: ',
597       err_u_to_aib_kreis, '\n')
598
599 #####u_to_ab_ggw#####
600 print('Standard error for system u_to_ab_ggw_i')
601 newYDataFits_u_to_ab_ggw_i = []
602 for i in range(0, 100):
603     newXData1 = []
604     newXData2 = []
605     newYData1 = []
606     newYData2 = []
607     # smallest value in one list of y data; initially set to be the
608     # highest number in y data
609     smallest1 = max(int1)
610     smallest2 = max(int2)
611     # make subsamples the same size as the original sample
612     for index in range(0, len(int1)):
613         # choose a random number between 0 (inclusive) and the number
614         # of elements in
615         # ydata1 (exclusive)
616         rand = np.random.randint(0, len(int1))
617         # if value of y data at position rand is smaller than current
618         # smallest value,
619         # set accordingly
620         if (int1[rand] < smallest1):
621             smallest1 = int1[rand]
622         # there is a problem with numerical solving differential
623         # equations if there are
624         # more than 4 entries with
625         # the same value at the beginning of the sample...
626         # so of there is already 4 times the same smallest number in
627         # the sample but the
628         # next one would be the same
629         # choose a different one
630         while (newYData1.count(smallest1) >= 4 and int1[
631             rand] == smallest1):
632             rand = np.random.randint(0, len(int1))
633         # corresponding values of x and y1 are added to the new lists
634         newXData1.append(freq[rand])
635         newYData1.append(int1[rand])
636         # the subsample is ordered by x values (smallest to highest)
637         # because numerical solving
638         # of differential equations needs it that way
639         newXData1, newYData1 = zip(sorted(zip(newXData1, newYData1)))
640         # same for second data
641         for index in range(0, len(int2)):
642             # choose a random number between 0 (inclusive) and the number
643             # of elements in
644             # ydata2 (exclusive)
645             rand = np.random.randint(0, len(int2))
646             # if value of y data at position rand is smaller than current
647             # smallest value,
648             # set accordingly
649             if (int2[rand] < smallest2):
650                 smallest2 = int2[rand]
651             # there is a problem with numerical solving differential
652             # equations if there are
653             # more than 4 entries with

```

```

715 # the same value at the beginning of the sample...
716 # so of there is already 4 times the same smallest number in
717 # the sample but the
718 # next one would be the same
719 # choose a different one
720 while (newYData2.count(smallest2) >= 4 and int2[
721     rand] == smallest2):
722     rand = np.random.randint(0, len(int2))
723 # corresponding values of x2 and y2 are added to the new lists
724 newXData2.append(freq2[rand])
725 newYData2.append(int2[rand])
726 # the subsample is ordered by x values (smallest to highest)
727 # because numerical solving
728 # of differential equations needs it that way
729 newXData2, newYData2 = zip(sorted(zip(newXData2, newYData2)))
730
731 if (i % 5 == 0):
732     print('.', end=' ', flush=True)
733
734 randomFit = minimize(obj_u_to_ab_ggw_i, k_opt_u_to_ab_ggw_i_diff,
735                     args=(
736                         newXData1, newXData2, newYData1, newYData2),
737                     method='L-BFGS-B', bounds=bounds_u_to_ab_ggw_i)
738 newYDataFits_u_to_ab_ggw_i.append(randomFit.x)
739
740 newYDataFits_u_to_ab_ggw_i = np.array(newYDataFits_u_to_ab_ggw_i)
741 k_opt_u_to_ab_ggw_i = np.mean(newYDataFits_u_to_ab_ggw_i, 0)
742
743 sigmaInterval = 1.0
744
745 err_u_to_ab_ggw_i = sigmaInterval * np.std(newYDataFits_u_to_ab_ggw_i, 0)
746
747 print('\nu_to_ab_ggw_i:\nBootstrap\nMean:', k_opt_u_to_ab_ggw_i,
748       '\nError: ',
749       err_u_to_ab_ggw_i, '\n')
750
751 #####u_to_ab#####
752 print('Standard error for system u_to_ab')
753 newYDataFits_u_to_ab = []
754 for i in range(0, 100):
755     newXData1 = []
756     newXData2 = []
757     newYData1 = []
758     newYData2 = []
759     # smallest value in one list of y data; initially set to be the
760     # highest number in y data
761     smallest1 = max(int1)
762     smallest2 = max(int2)
763     # make subsamples the same size as the original sample
764     for index in range(0, len(int1)):
765         # choose a random number between 0 (inclusive) and the number
766         # of elements in
767         # ydata1 (exclusive)
768         rand = np.random.randint(0, len(int1))
769         # if value of y data at position rand is smaller than current
770         # smallest value,
771         # set accordingly
772         if (int1[rand] < smallest1):
773             smallest1 = int1[rand]
774         # there is a problem with numerical solving differential
775         # equations if there are
776         # more than 4 entries with
777         # the same value at the beginning of the sample...
778         # so of there is already 4 times the same smallest number in
779         # the sample but the

```



```

1300 tmpX = np.linspace(start=0, stop=(
1301 freq[-1] if freq[-1] > freq2[-1] else freq2[-1]),
1302 num=len(freq) if len(freq) > len(freq2) else len(
1303 freq2))
1304 for i in range(0, len(tmpX)):
1305     string = str(tmpX[i]) + '\t' + str(
1306         model_aub_kreis(k_opt_aub_kreis_diff, tmpX[:, 1][i]) + '\n'
1307     )
1308     tmpFile.write(string)
1309 with open("aub_kreis_residuals.txt", "w") as tmpFile:
1310     tmpFile.write("#\n")
1311     for i in range(0, len(freq)):
1312         string = str(freq[i]) + '\t' + str(
1313             int1[i] - model_aub_kreis(k_opt_aub_kreis_diff, freq)[:, 0][
1314                 i]) + '\n'
1315         tmpFile.write(string)
1316     tmpFile.write("#\n")
1317     for i in range(0, len(freq2)):
1318         string = str(freq2[i]) + '\t' + str(
1319             int2[i] - model_aub_kreis(k_opt_aub_kreis_diff, freq2)[:, 2][
1320                 i]) + '\n'
1321         tmpFile.write(string)
1322 with open("u_to_aib.txt", "w") as tmpFile:
1323     tmpFile.write("#\n")
1324     for i in range(0, len(freq)):
1325         string = str(freq[i]) + '\t' + str(
1326             model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 0][i]) +
1327             '\n'
1328         tmpFile.write(string)
1329     tmpFile.write("#\n")
1330     tmpFile.write("#\n")
1331     for i in range(0, len(freq2)):
1332         string = str(freq2[i]) + '\t' + str(
1333             model_u_to_aib(k_opt_u_to_aib_diff, freq2)[:, 2][i]) +
1334             '\n'
1335         tmpFile.write(string)
1336     tmpFile.write("#\n")
1337 tmpX = np.linspace(start=0, stop=(
1338 freq[-1] if freq[-1] > freq2[-1] else freq2[-1]),
1339 num=len(freq) if len(freq) > len(freq2) else len(
1340 freq2))
1341 for i in range(0, len(tmpX)):
1342     string = str(tmpX[i]) + '\t' + str(
1343         model_u_to_aib(k_opt_u_to_aib_diff, tmpX[:, 1][i]) + '\n'
1344     )
1345     tmpFile.write(string)
1346     tmpFile.write("#\n")
1347 for i in range(0, len(tmpX)):
1348     string = str(tmpX[i]) + '\t' + str(
1349         model_u_to_aib(k_opt_u_to_aib_diff, tmpX[:, 3][i]) + '\n'
1350     )
1351     tmpFile.write(string)
1352 with open("u_to_aib_residuals.txt", "w") as tmpFile:
1353     tmpFile.write("#\n")
1354     for i in range(0, len(freq)):
1355         string = str(freq[i]) + '\t' + str(
1356             int1[i] - model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 0][
1357                 i]) + '\n'
1358         tmpFile.write(string)
1359     tmpFile.write("#\n")
1360     tmpFile.write("#\n")
1361     for i in range(0, len(freq2)):
1362         string = str(freq2[i]) + '\t' + str(
1363             int2[i] - model_u_to_aib(k_opt_u_to_aib_diff, freq2)[:, 2][
1364                 i]) + '\n'
1365         tmpFile.write(string)
1366 
```

```

1400 tmpFile.write("#\n")
1401 tmpX = np.linspace(start=0, stop=(
1402 freq[-1] if freq[-1] > freq2[-1] else freq2[-1]),
1403 num=len(freq) if len(freq) > len(freq2) else len(
1404 freq2))
1405 for i in range(0, len(tmpX)):
1406     string = str(tmpX[i]) + '\t' + str(
1407         model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, tmpX[
1408             :, 1][
1409                 i]) + '\n'
1410     )
1411     tmpFile.write(string)
1412     tmpFile.write("#\n")
1413     tmpFile.write("#\n")
1414     for i in range(0, len(tmpX)):
1415         string = str(tmpX[i]) + '\t' + str(
1416             model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, tmpX[
1417                 :, 3][
1418                     i]) + '\n'
1419         )
1420         tmpFile.write(string)
1421 with open("u_to_ab_ggw_i_residuals.txt", "w") as tmpFile:
1422     tmpFile.write("#\n")
1423     for i in range(0, len(freq)):
1424         string = str(freq[i]) + '\t' + str(
1425             int1[i] - model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[
1426                 :, 0][
1427                     i]) + '\n'
1428         tmpFile.write(string)
1429     tmpFile.write("#\n")
1430     tmpFile.write("#\n")
1431     for i in range(0, len(freq2)):
1432         string = str(freq2[i]) + '\t' + str(
1433             int2[i] - model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq2)[
1434                 :, 2][
1435                     i]) + '\n'
1436         tmpFile.write(string)
1437 with open("u_to_ab.txt", "w") as tmpFile:
1438     tmpFile.write("#\n")
1439     for i in range(0, len(freq)):
1440         string = str(freq[i]) + '\t' + str(
1441             model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 0][i]) + '\n'
1442         tmpFile.write(string)
1443     tmpFile.write("#\n")
1444     tmpFile.write("#\n")
1445     for i in range(0, len(freq2)):
1446         string = str(freq2[i]) + '\t' + str(
1447             model_u_to_ab(k_opt_u_to_ab_diff, freq2)[:, 2][i]) + '\n'
1448         tmpFile.write(string)
1449 tmpX = np.linspace(start=0, stop=(
1450 freq[-1] if freq[-1] > freq2[-1] else freq2[-1]),
1451 num=len(freq) if len(freq) > len(freq2) else len(
1452 freq2))
1453 for i in range(0, len(tmpX)):
1454     string = str(tmpX[i]) + '\t' + str(
1455         model_u_to_ab(k_opt_u_to_ab_diff, tmpX[:, 2][i]) + '\n'
1456     )
1457     tmpFile.write(string)
1458 with open("u_to_ab_residuals.txt", "w") as tmpFile:
1459     tmpFile.write("#\n")
1460     for i in range(0, len(freq)):
1461         string = str(freq[i]) + '\t' + str(
1462             int1[i] - model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 0][
1463                 i]) + '\n'
1464         tmpFile.write(string)
1465     tmpFile.write("#\n")
1466     tmpFile.write("#\n")
1467 
```

```

1366 with open("u_to_aib_kreis.txt", "w") as tmpFile:
1367     tmpFile.write("#\n")
1368     for i in range(0, len(freq)):
1369         string = str(freq[i]) + '\t' + str(
1370             model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq
1371                 )[:, 0][
1372                     i]) + '\n'
1373         tmpFile.write(string)
1374     tmpFile.write("#\n")
1375     tmpFile.write("#\n")
1376     for i in range(0, len(freq2)):
1377         string = str(freq2[i]) + '\t' + str(
1378             model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq
1379                 )[:, 2][
1380                     i]) + '\n'
1381         tmpFile.write(string)
1382     tmpFile.write("#\n")
1383 tmpX = np.linspace(start=0, stop=(
1384 freq[-1] if freq[-1] > freq2[-1] else freq2[-1]),
1385 num=len(freq) if len(freq) > len(freq2) else len(
1386 freq2))
1387 for i in range(0, len(tmpX)):
1388     string = str(tmpX[i]) + '\t' + str(
1389         model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, tmpX
1390             )[:, 1][
1391                 i]) + '\n'
1392         tmpFile.write(string)
1393     tmpFile.write("#\n")
1394     tmpFile.write("#\n")
1395     for i in range(0, len(tmpX)):
1396         string = str(tmpX[i]) + '\t' + str(
1397             model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, tmpX
1398                 )[:, 3][
1399                     i]) + '\n'
1400         tmpFile.write(string)
1401 with open("u_to_aib_kreis_residuals.txt", "w") as tmpFile:
1402     tmpFile.write("#\n")
1403     for i in range(0, len(freq)):
1404         string = str(freq[i]) + '\t' + str(
1405             model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq
1406                 )[:, 0][
1407                     i]) + '\n'
1408         tmpFile.write(string)
1409     tmpFile.write("#\n")
1410     tmpFile.write("#\n")
1411     for i in range(0, len(freq2)):
1412         string = str(freq2[i]) + '\t' + str(
1413             int2[i] -
1414             model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq
1415                 )[:, 2][
1416                     i]) + '\n'
1417         tmpFile.write(string)
1418 with open("u_to_ab_ggw_i.txt", "w") as tmpFile:
1419     tmpFile.write("#\n")
1420     for i in range(0, len(freq)):
1421         string = str(freq[i]) + '\t' + str(
1422             model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[
1423                 :, 0][
1424                     i]) + '\n'
1425         tmpFile.write(string)
1426     tmpFile.write("#\n")
1427     tmpFile.write("#\n")
1428     for i in range(0, len(freq2)):
1429         string = str(freq2[i]) + '\t' + str(
1430             model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq2)[
1431                 :, 2][
1432                     i]) + '\n'
1433         tmpFile.write(string)
1434     tmpFile.write("#\n")
1435 
```

```

1495 for i in range(0, len(freq2)):
1496     string = str(freq2[i]) + '\t' + str(
1497         int2[i] - model_u_to_ab(k_opt_u_to_ab_diff, freq2)[:, 2][
1498             i]) + '\n'
1499     tmpFile.write(string)
1500 #####PLOTING#####
1501 subplot(231)
1502 title('aub')
1503 plot(freq2, int2, 'k')
1504 plot(freq, int1, 'k')
1505 plot(freq, model_aub(k_opt_aub_diff, freq)[:, 0], 'r', linewidth=3.0)
1506 # plot(freq, model_aub(k_opt_aub_diff, freq)[:, 1], 'g', linewidth=3.0)
1507 plot(freq, model_aub(k_opt_aub_diff, freq)[:, 2], 'b', linewidth=3.0)
1508 subplot(232)
1509 title('aub_kreis')
1510 plot(freq2, int2, 'k')
1511 plot(freq, int1, 'k')
1512 plot(freq, model_aub_kreis(k_opt_aub_kreis_diff, freq
1513     )[:, 0], 'r',
1514     linewidth=3.0)
1515 # plot(freq, model_aub_kreis(k_opt_aub_kreis_diff, fre
1516     q)[:, 1], 'g',
1517     # linewidth=3.0)
1518 plot(freq, model_aub_kreis(k_opt_aub_kreis_diff, freq
1519     )[:, 2], 'b',
1520     linewidth=3.0)
1521 subplot(233)
1522 title('u_to_aib')
1523 plot(freq2, int2, 'k')
1524 plot(freq, int1, 'k')
1525 plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[
1526     :, 0], 'r',
1527     linewidth=3.0)
1528 plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[
1529     :, 1], 'g',
1530     linewidth=3.0)
1531 plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[
1532     :, 2], 'b',
1533     linewidth=3.0)
1534 # plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)
1535     [:, 3], 'yellow',
1536     # linewidth=3.0)
1537 subplot(234)
1538 title('u_to_aib_kreis')
1539 plot(freq2, int2, 'k')
1540 plot(freq, int1, 'k')
1541 plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis
1542     _diff, freq)[:, 0], 'r',
1543     linewidth=3.0)
1544 plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis
1545     _diff, freq)[:, 1], 'g',
1546     linewidth=3.0)
1547 plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis
1548     _diff, freq)[:, 2], 'b',
1549     linewidth=3.0)
1550 # plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_krei
1551     s_diff, freq)[:, 3],
1552     # 'yellow',
1553     # linewidth=3.0)
1554 subplot(235)
1555 title('u_to_ab_ggw_i')
1556 plot(freq2, int2, 'k')
1557 plot(freq, int1, 'k')
1558 plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_d
1559     iff, freq)[:, 0], 'r',
1560     linewidth=3.0)
1561 plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_d
1562     iff, freq)[:, 1], 'g',
1563     linewidth=3.0)
1564 plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_d
1565     iff, freq)[:, 2], 'b',
1566     linewidth=3.0)
1567 # plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i
1568     _diff, freq)[:, 3],
1569     # 'yellow', linewidth=3.0)
1570 
```

```

1550 subplot(236)
1551 title('u_to_ab')
1552 plot(freq2, int2, 'k')
1553 plot(freq, int1, 'k')
1554 plot(freq, model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 0], 'r',
1555      linewidth=3.0)
1556 plot(freq, model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 1], 'b',
1557      linewidth=3.0)
1558 # plot(freq, model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 2], 'g',
1559 #      linewidth=3.0)
1560 savefig('folding_all.pdf', format='pdf', transparent=True)
1561 show()
1562
1563 clf()
1564 title('aub')
1565 plot(freq2, int2, 'k')
1566 plot(freq, int1, 'k')
1567 plot(freq, model_aub(k_opt_aub_diff, freq)[:, 0], 'r', linewidth=3.0)
1568 # plot(freq, model_aub(k_opt_aub_diff, freq)[:, 1], 'g',
1569 #      linewidth=3.0)
1570 plot(freq, model_aub(k_opt_aub_diff, freq)[:, 2], 'b', linewidth=3.0)
1571 savefig('folding_aub.pdf', format='pdf', transparent=True)
1572
1573 clf()
1574 title('aub_kreis')
1575 plot(freq2, int2, 'k')
1576 plot(freq, int1, 'k')
1577 plot(freq, model_aub_kreis(k_opt_aub_kreis_diff, freq)[:, 0], 'r',
1578      linewidth=3.0)
1579 # plot(freq, model_aub_kreis(k_opt_aub_kreis_diff, freq)[:, 1], 'g',
1580 #      linewidth=3.0)
1581 plot(freq, model_aub_kreis(k_opt_aub_kreis_diff, freq)[:, 2], 'b',
1582      linewidth=3.0)
1583 savefig('folding_aub_kreis.pdf', format='pdf', transparent=True)
1584
1585 clf()
1586 title('u_to_aib')
1587 plot(freq2, int2, 'k')
1588 plot(freq, int1, 'k')
1589 plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 0], 'r',
1590      linewidth=3.0)
1591 plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 1], 'g',
1592      linewidth=3.0)
1593 plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 2], 'b',
1594      linewidth=3.0)
1595 # plot(freq, model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 3], 'yellow',
1596 #      linewidth=3.0)
1597 savefig('folding_u_to_aib.pdf', format='pdf', transparent=True)
1598
1599 clf()
1600 title('u_to_aib_kreis')
1601 plot(freq2, int2, 'k')
1602 plot(freq, int1, 'k')
1603 plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 0], 'r',
1604      linewidth=3.0)
1605 plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 1], 'g',
1606      linewidth=3.0)
1607 plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 2], 'b',
1608      linewidth=3.0)
1609 # plot(freq, model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 3],
1610 #      'yellow',
1611 #      linewidth=3.0)
1612 savefig('folding_u_to_aib_kreis.pdf', format='pdf', transparent=True)
1613
1614 clf()
1615
1616 savefig('folding_u_to_aib_kreis_residuals.pdf', format='pdf',
1617        transparent=True)
1618
1619 clf()
1620 title('Residuals u_to_ab_ggw_i')
1621 plot(freq, int1 - model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[:, 0],
1622      'o',
1623      markerfacecolor=(1, 0, 0, 0.5),
1624      markeredgewidth=0.5)
1625 plot(freq2,
1626      int2 - model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq2)[:, 2], 'o',
1627      markerfacecolor=(0, 0, 1, 0.5),
1628      markeredgewidth=0.5)
1629 savefig('folding_u_to_ab_ggw_i_residuals.pdf', format='pdf',
1630        transparent=True)
1631
1632 clf()
1633 title('Residuals u_to_ab')
1634 plot(freq, int1 - model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 0], 'o',
1635      markerfacecolor=(1, 0, 0, 0.5),
1636      markeredgewidth=0.5)
1637 plot(freq2, int2 - model_u_to_ab(k_opt_u_to_ab_diff, freq2)[:, 1], 'o',
1638      markerfacecolor=(0, 0, 1, 0.5),
1639      markeredgewidth=0.5)
1640 savefig('folding_u_to_ab_residuals.pdf', format='pdf', transparent=True)
1641
1642 clf()
1643 subplot(231)
1644 title('Residuals aub')
1645 plot(freq, int1 - model_aub(k_opt_aub_diff, freq)[:, 0], 'o',
1646      markerfacecolor=(1, 0, 0, 0.5),
1647      markeredgewidth=0.5)
1648 plot(freq2, int2 - model_aub(k_opt_aub_diff, freq2)[:, 2], 'o',
1649      markerfacecolor=(0, 0, 1, 0.5),
1650      markeredgewidth=0.5)
1651
1652 subplot(232)
1653 title('Residuals aub_kreis')
1654 plot(freq, int1 - model_aub_kreis(k_opt_aub_kreis_diff, freq)[:, 0], 'o',
1655      markerfacecolor=(1, 0, 0, 0.5),
1656      markeredgewidth=0.5)
1657 plot(freq2, int2 - model_aub_kreis(k_opt_aub_kreis_diff, freq2)[:, 2], 'o',
1658      markerfacecolor=(0, 0, 1, 0.5),
1659      markeredgewidth=0.5)
1660
1661 subplot(233)
1662 title('Residuals u_to_aib')
1663 plot(freq, int1 - model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 0], 'o',
1664      markerfacecolor=(1, 0, 0, 0.5),
1665      markeredgewidth=0.5)
1666 plot(freq2, int2 - model_u_to_aib(k_opt_u_to_aib_diff, freq2)[:, 2], 'o',
1667      markerfacecolor=(0, 0, 1, 0.5),
1668      markeredgewidth=0.5)
1669
1670 subplot(234)
1671 title('Residuals u_to_aib_kreis')
1672 plot(freq,
1673      int1 - model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 0],
1674      'o',
1675      markerfacecolor=(1, 0, 0, 0.5),
1676      markeredgewidth=0.5)
1677 plot(freq2,
1678      int2 - model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq2)[:, 2],
1679      'o',
1680      markerfacecolor=(0, 0, 1, 0.5),

```

```

1625 title('u_to_ab_ggw_i')
1626 plot(freq2, int2, 'k')
1627 plot(freq, int1, 'k')
1628 plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[:, 0], 'r',
1629      linewidth=3.0)
1630 plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[:, 1], 'g',
1631      linewidth=3.0)
1632 plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[:, 2], 'b',
1633      linewidth=3.0)
1634 # plot(freq, model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[:, 3],
1635 #      'yellow', linewidth=3.0)
1636 savefig('folding_u_to_ab_ggw_i.pdf', format='pdf', transparent=True)
1637
1638 clf()
1639 title('u_to_ab')
1640 plot(freq2, int2, 'k')
1641 plot(freq, int1, 'k')
1642 plot(freq, model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 0], 'r',
1643      linewidth=3.0)
1644 plot(freq, model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 1], 'b',
1645      linewidth=3.0)
1646 # plot(freq, model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 2], 'g',
1647 #      linewidth=3.0)
1648 savefig('folding_u_to_ab.pdf', format='pdf', transparent=True)
1649
1650 clf()
1651 title('Residuals aub')
1652 plot(freq, int1 - model_aub(k_opt_aub_diff, freq)[:, 0], 'o',
1653      markerfacecolor=(1, 0, 0, 0.5),
1654      markeredgewidth=0.5)
1655 plot(freq2, int2 - model_aub(k_opt_aub_diff, freq2)[:, 2], 'o',
1656      markerfacecolor=(0, 0, 1, 0.5),
1657      markeredgewidth=0.5)
1658 savefig('folding_aub_residuals.pdf', format='pdf', transparent=True)
1659
1660 clf()
1661 title('Residuals aub_kreis')
1662 plot(freq, int1 - model_aub_kreis(k_opt_aub_kreis_diff, freq)[:, 0], 'o',
1663      markerfacecolor=(1, 0, 0, 0.5),
1664      markeredgewidth=0.5)
1665 plot(freq2, int2 - model_aub_kreis(k_opt_aub_kreis_diff, freq2)[:, 2], 'o',
1666      markerfacecolor=(0, 0, 1, 0.5),
1667      markeredgewidth=0.5)
1668 savefig('folding_aub_kreis_residuals.pdf', format='pdf', transparent=True)
1669
1670 clf()
1671 title('Residuals u_to_aib')
1672 plot(freq, int1 - model_u_to_aib(k_opt_u_to_aib_diff, freq)[:, 0], 'o',
1673      markerfacecolor=(1, 0, 0, 0.5),
1674      markeredgewidth=0.5)
1675 plot(freq2, int2 - model_u_to_aib(k_opt_u_to_aib_diff, freq2)[:, 2], 'o',
1676      markerfacecolor=(0, 0, 1, 0.5),
1677      markeredgewidth=0.5)
1678 savefig('folding_u_to_aib_residuals.pdf', format='pdf', transparent=True)
1679
1680 clf()
1681 title('Residuals u_to_aib_kreis')
1682 plot(freq,
1683      int1 - model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 0],
1684      'o',
1685      markerfacecolor=(1, 0, 0, 0.5),
1686      markeredgewidth=0.5)
1687 plot(freq2,
1688      int2 - model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq2)[:, 2],
1689      'o',
1690      markerfacecolor=(0, 0, 1, 0.5),
1691      markeredgewidth=0.5)
1692
1693 plot(freq,
1694      int1 - model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq)[:, 0],
1695      'o',
1696      markerfacecolor=(1, 0, 0, 0.5),
1697      markeredgewidth=0.5)
1698 plot(freq2,
1699      int2 - model_u_to_aib_kreis(k_opt_u_to_aib_kreis_diff, freq2)[:, 2],
1700      'o',
1701      markerfacecolor=(0, 0, 1, 0.5),
1702      markeredgewidth=0.5)
1703
1704 subplot(235)
1705 title('Residuals u_to_ab_ggw_i')
1706 plot(freq, int1 - model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq)[:, 0],
1707      'o',
1708      markerfacecolor=(1, 0, 0, 0.5),
1709      markeredgewidth=0.5)
1710 plot(freq2,
1711      int2 - model_u_to_ab_ggw_i(k_opt_u_to_ab_ggw_i_diff, freq2)[:, 2], 'o',
1712      markerfacecolor=(0, 0, 1, 0.5),
1713      markeredgewidth=0.5)
1714
1715 subplot(236)
1716 title('Residuals u_to_ab')
1717 plot(freq, int1 - model_u_to_ab(k_opt_u_to_ab_diff, freq)[:, 0], 'o',
1718      markerfacecolor=(1, 0, 0, 0.5),
1719      markeredgewidth=0.5)
1720 plot(freq2, int2 - model_u_to_ab(k_opt_u_to_ab_diff, freq2)[:, 1], 'o',
1721      markerfacecolor=(0, 0, 1, 0.5),
1722      markeredgewidth=0.5)
1723 savefig('folding_all_residuals.pdf', format='pdf', transparent=True)

```


Acknowledgements - Danksagung

“... und vergiss nicht, Danke zu sagen!”

Meine liebe **Mama**, immer

Ich danke meinem Doktorvater und Mentor Harald Schwalbe;

[Redacted text block containing multiple lines of blacked-out content]

Curriculum Vitae

Personal

Name: Jens Tassilo Grün
Date of Birth: 31.01.1991
Place of Birth: Wetzlar, Germany
Nationality: German

Education

01/2017-Date of print **Doctorate** (PhD) in Chemistry with Prof. Harald Schwalbe
Goethe University, Frankfurt am Main
Fachbereich 14 – Biochemie, Chemie, Pharmazie

11/2017-03/2021 **Member and representative** of the graduate research programme
GRK 1986: “Complex Light Control”, (CLiC)

10/2014-09/2016 **Master of Science** (M.Sc.) in Chemistry, *with distinction*
Goethe University, Frankfurt am Main

04/2016-09/2016 **Master Thesis** with Prof. Harald Schwalbe: “NMR-Spectroscopic
Investigations of G-Quadruplex-Ligand Complexes”

10/2015-12/2015 **Study abroad and Research stay** with Prof. Robert G. Griffin
Massachusetts Institute of Technology, Cambridge (MA, USA)

10/2011-12/2014 **Bachelor of Science** (B.Sc.) in Chemistry
Goethe University, Frankfurt am Main

07/2014-09/2014 **Bachelor Thesis** with Prof. Harald Schwalbe: “Investigation of
G-Quadruplex-Ligand Complexes”

06/2010 **Abitur (allgemeine Hochschulreife)**
Johanneum Gymnasium, Herborn

Publications

- (9) G. Pinter, K.F. Hohmann, **J.T. Grün**, J. Wirmer-Bartoschek, C. Glaubitz, B. Fürtig, H. Schwalbe, Real-time NMR spectroscopy in the study of biomolecular kinetics and dynamics, *Magnetic Resonance* **2021**, *Robert Kaptein Festschrift, under revision*, as preprint available: doi.org/10.5194/mr-2021-1
- (8) **J.T. Grün**, A. Blümmler, I. Burkhart, J. Wirmer-Bartoschek, A. Heckel, H. Schwalbe, Unraveling the Kinetics of Spare-Tire DNA G-Quadruplex Folding, *Journal of the American Chemical Society* **2021**, *accepted manuscript (published online)*
- (7) M. Novakovic, E. Kupče, T. Scherf, A. Oxenfarth, R. Schnieders, **J.T. Grün**, J. Wirmer-Bartoschek, C. Richter, H. Schwalbe, L. Frydman, Magnetization transfer to enhance NOE cross-peaks among labile protons: Applications to imino-imino sequential walks in SARS-CoV-2-derived RNAs, *Angewandte Chemie (Int. Ed.)* **2021**, *accepted article* (doi: 10.1002/anie.202015948)
- (6) R. Schnieders, S.A. Peter *et al.* (**J.T. Grün** as co-author in alphabetical order), ¹H, ¹³C and ¹⁵N chemical shift assignment of the stem-loop 5a from the 5'-UTR of SARS-CoV-2, *Biomolecular NMR assignments* 2021, (doi: 10.1007/s12104-021-10007-w)
- (5) R.W. Harkness, C. Hennecker, **J.T. Grün**, A. Blümmler, A. Heckel, H. Schwalbe, A.K. Mittermaier, Parallel reaction pathways accelerate folding of a guanine quadruplex, *Nucleic Acids Research* **2021**, 49 (3), 1247-1262
- (4) A. Wacker, J. Weigand *et al.* (COVID19-NMR team, **J.T. Grün** as co-author in alphabetical order), Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy, *Nucleic Acids Research* **2020**, 48 (22), 12415-12435* (*Breakthrough article)
- (3) **J.T. Grün**, C. Hennecker, D.-P. Klötzner, R.W. Harkness, I. Bessi, A. Heckel, A.K. Mittermaier, H. Schwalbe, Conformational Dynamics of Strand Register Shifts in DNA G-Quadruplexes, *Journal of the American Chemical Society* **2020**, 142 (1), 264-273*
*Spotlight in: D. Lockwood, *Journal of the American Chemical Society* **2020**, 142 (6), 2697-2698
- (2) M. Reese, C. George, C. Yang, S. Jawa, **J.T. Grün**, H. Schwalbe, C. Redfield, R. J. Temkin, R.G. Griffin, Modular, triple-resonance, transmission line DNP MAS probe for 500 MHz/330 GHz, *Journal of Magnetic Resonance* **2019**, 307, 106573
- (1) J. Wirmer-Bartoschek, L.E. Bendel, H.R.A. Jonker, **J.T. Grün**, F. Papi, C. Bazzicalupi, L. Messori, P. Gratteri, H. Schwalbe, Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements that Affect Ligand Binding, *Angewandte Chemie (Int. Ed.)* **2017**, 56, 7102-7106

Conference Contributions and invited Seminar talks

- (7) **J.T. Grün**, A. Blümmler, A. Heckel, H. Schwalbe, Conformational Dynamics of unusual DNA structures: Insights into G-Quadruplex Polymorphism by Real-Time NMR-Spectroscopy
Poster Presentation, 62nd ENC (Virtual Conference), March 2021
- (6) **J.T. Grün**, Conformational Dynamics of unusual DNA structures: Insights into G-Quadruplex Polymorphism by Real-Time NMR-Spectroscopy
Seminar Talk (Zoom), Magnetic Resonance Seminar at Weizmann Institute of Science, October 2020
- (5) **J.T. Grün**, A. Blümmler, A. Heckel, H. Schwalbe, Conformational Dynamics of DNA G-Quadruplexes - Insights by Real-Time-NMR
Oral Presentation, V. Doktorandenseminar Nucleinsäurechemie (DNG), Bad Herrenalb (Germany), October 2020
- (4) **J.T. Grün**, D.-P. Klötzner, R.W. Harkness, C. Hennecker, I. Bessi, A. Heckel, A.K. Mittermaier and H. Schwalbe, Mechanistic Insights to G-Quadruplex Conformational Dynamics by Real-Time NMR-spectroscopy
Poster Presentation, 7th Int. Meeting on Quadruplex Nucleic Acids Changchun (China), September 2019
- (3) **J.T. Grün**, D.-P. Klötzner, R.W. Harkness, I. Bessi, A. Heckel, A.K. Mittermaier and H. Schwalbe, Investigation of DNA G-Quadruplex Structural Dynamics using Real-Time NMR-spectroscopy
Poster Presentation, XXIX. ICMRBS Dublin (Ireland), August 2018
- (2) **J.T. Grün**, D.-P. Klötzner, R.W. Harkness, I. Bessi, A. Heckel, A.K. Mittermaier and H. Schwalbe, Investigation of DNA G-Quadruplex Structural Dynamics using Real-Time NMR-spectroscopy
Poster Presentation, 20th JCF-Frühjahrssymposium Konstanz (Germany), March 2018
- (1) **J.T. Grün**, D.-P. Klötzner, A. Heckel, I. Bessi and H. Schwalbe, Investigation of Promoter G-quadruplex Folding Dynamics using Real-Time NMR-spectroscopy
Poster Presentation, 6th Int. Meeting on Quadruplex Nucleic Acids Prague (Czech Republic), June 2017

Honors and Awards

- 2019 Royal Society of Chemistry Poster Award, 7th Int. Meeting on Quadruplex Nucleic Acids, Changchun (China)
- 2018 Suraj Manrao Poster Prize, XXIX. ICMRBS, Dublin (Ireland)
- 2017 Valedictorian, Fachbereich Biochemie, Chemie und Pharmazie
- 2015 Dr. Albrecht Magen Fellowship, Steuben-Schurz Gesellschaft, Frankfurt am Main
Elisabeth-Meurer Fellowship, Elisabeth-Meurer Privatstiftung, Pulheim

Academic Teachers (in alphabetical order)

Prof. Dr. Norbert Auner, PD Dr. Markus Braun, PD Dr. George Bruls, Prof. Dr. Irene Burghardt, Prof. Dr. Björn Corzilius, Prof. Dr. Gunter Eckert, Prof. Dr. Ernst Egert, Prof. Dr. Joachim Engels, Prof. Dr. Michael Göbel, Prof. Clemens Glaubitz (D.Phil Oxon), Prof. Robert G. Griffin (PhD), Prof. Dr. Martin Grininger, Prof. Dr. Alexander Heckel, AOR Dr. Rainer Hegger, Prof. Dr. Mike Heilemann, Prof. Dr. Max C. Holthausen, PD Dr. Matthias Köck, Prof. Dr. Georg Manolikakes, Prof. Dr. Thomas Prisner, Prof. Dr. Matthias Rehahn, Prof. Dr. Martin U. Schmidt, Apl. Prof. Dr. Wolfgang Schubert, Prof. Dr. Harald Schwalbe, Prof. Dr. Andreas Terfort, Prof. Dr. Josef Wachtveitl, Prof. Dr. Matthias Wagner