Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXVI

Settore Concorsuale di afferenza: 05/E2 BIOLOGIA MOLECOLARE

Settore Scientifico disciplinare: BIO/11 BIOLOGIA MOLECOLARE

Interactions of G-quadruplex binders and Topoisomerase I inhibitors in cancer cells

Presentata da: **Tiang Yee Peng**

Coordinatore Dottorato

Prof. Davide Zannoni

Relatore

Prof. Giovanni Capranico

Co-relatore

Dott.ssa Jessica Marinello

Esame finale anno 2015

To

Mom & Dad

Acknowledgement

"Time flies but not memories" Ika Natassa.

So said a writer, hence, this thesis marks the end of a chapter in my life which began 4 years back. My adventures in Italy, is clear and concise; to do a PhD in quenching my thirst for knowledge and challenges to test the limit of capability.

I am greatly indebted to the Italian Government, through the Embassy of Italy in Kuala Lumpur, Malaysia, for awarding a fellowship for my study in the University of Bologna. It is the generosity of the Italian taxpayers that kickstarted this exciting journey, despite the unexpected reduction of fellowship to 2 years from a promised 3-year of fellowship.

Despite this great financial obstacle, I was awarded a lodging fellowship by the University of Bologna through the Institute of Advanced Studies (ISA). To which, I would like express my great gratitude as the fellowship help to reduce significantly my financial burden, without which, I might have to end my study in my second year. A sincere thank you to Dr. Barbara Cimatti, for her great coordination in ensuring a great experience with the ISA.

To Professor Giovanni Capranico, thank you for your patience and guidance plus constant critics in ensuring the highest quality work obtained. To Dr. Jessica Marinello, thank you for your technical guidance throughout the whole duration of my thesis. Thank you also to CIRB for the usage of microscopy facilities and Prof Natalia Calonghi for her technical help in microscopy. Also not forgetting members and ex-members from the laboratory, Davide, Stefano, Christian, Fabrizio and Iris for their ever willingness to help.

To my family, in particular my mother, who has been constantly giving her relentless support during these years. Her open-mindedness in giving me all the freedom to choose the path I desired for life is deeply appreciated.

To Gianfranco and Evelyn in Milan, to whom I have shared closely all my ups and downs, provided me a shelter when I needed them. These will be remembered for life and a word of thank you is insufficient to express my gratitude.

To Jia, my closest friend in Italy and also one of the greatest people, I have ever met. I am so lucky to have you as my friend. And to you, Kit, my best confidant and friend, despite the distance that keeps us apart, your encouragements, advises and sometimes weird theories are ever precious. Through the laughters and tears we shared, your friendship is greatly appreciated and may it flourishes till world ends. To Fabien, my apologies for putting your name in the end, but my gratitude to you is different from the rests. It is a special kind of gratitude that will follow you for a very long time, if not forever. Thank you for being there for me and become my punch bag when things go awry. For being gratitude, I will wear dresses for you, occasionally.

To all my friends and the people who had crossed my path, both in Italy and during my 800 km walk in the Camino de Santiago, Aravin, Kapila, Esther, Sophia, Hyeyoung and Patrick you have not been forgotten and you have had given me all the motivations and strength to go on. I have also learnt a great deal from you guys. Thank you.

The journey in Italy was indeed not a rosy path, particularly with the complications from the different cultures, language barrier in addition to the unexpected financial struggles. However, with my head held high, I am proud that all these had been overcome with a great deal of will, hard work, patience and professionalism.

Finally, with great humility, I would like to convey my most sincere apologies to anyone that I might have hurted unintentionally along the way. Thank you all for your tolerance.

Tiang Yee Peng

March 2015

Abstract

Coordination of gene expression in eukaryotes involves several hierarchical regulations in response to internal and environmental changes. DNA topology, the fundamental control for DNA functions, is regulated through different DNA topoisomerases such as Topoisomerase 1. These enzymes, having the ability to cut one or two strands of DNA, relief torsional stresses that are caused by processes such as replication, transcription, recombination and repair. Recently, DNA negative supercoiling is associated with the formation of non B-DNA structures such as G-quadruplexes and R-loops that may have functional roles in gene regulations.

This thesis is, hence, a continuation of an on-going research project which has been investigating the role of human Top1 during transcriptional consequences associated with Top1 inhibition by CPT (a specific Top1 poison) in human cell lines. We investigate the interaction of Top1 with G-quadruplex in cancer cells using pharmacological tools, ie. CPT inhibitors and G4 binders (binds and stabilizes G4 structures).

Previous findings demonstrate that Top1-DNA cleavage complexes (Top1ccs) trigger an accumulation of antisense RNAPII transcripts specifically at active divergent CpG-island promoters in a replication independent and Top1 dependent manner. A burst of Top1ccs, parallels the transient increase of R-loops is reported in these promoters and transcribed regions; indicating a response pathway leading to transcription-dependent genome instability and altered transcription regulation.

Using different cancer cell lines of colon and osteo origins, we show that they display different sensitivity to CPT that is independent from Top1 level. The cell lines also show different response to G4 binders. To look at the interactions between Top1 and G4, we show that co-treatment with G4 binders potentiate the cell cytotoxicity of CPT regardless of the treatment sequences. Potentiation is indicated by a reduced inhibition concentration (IC_{50}) with a more profound cytotoxicity in CPT-resistant cell lines, HCT15 and U2OS. We hence, show the interactions between Top1inhibitor and G4 binders. This is further supported by the presence of G4 motifs as determined by computational analysis on 225 genes with CPT-

induced antisense transcription. G4 motifs are present mostly 5000 bp upstream from transcription start site and notably lower in genes. Comparisons between genes with no antisense transcription and genes with antisense transcription show that G4 motifs in this region are notably lower in the genes with antisense transcripts.

Since CPT increases negative supercoils at promoters of intermediate activity, we then demonstrate that the formation of G4 is also increased in CPT-treated cells. Suprisingly, formation of G4 is regulated in parallel to the transient stabilization of R-loops, indicating a role in response to stress caused by CPT. Moreover, G4 formation is highly elevated in Pyridostatin treated cells, which previous study shows increased formation of γH2Ax foci. This effect is also seen in the CPT-resistant cell lines, HCT15, indicating that the formation is a general event in response to CPT. We also show that R-loop formation is greatly increased in Pyridostatin treated cells. In order to study the role of R-loops and G4 structures in Top1cc-dependant repair pathway, we inhibited tyrosyl-phosphodiestrase 1 (TDP-1) using a TDP-1 inhibitor. Although, we have not obtained a similar kinetics as seen in the formation of R-loops and G4, we show that prolonged TDP-1 inhibition show a small changes in their formation, if any. It is however, further experiments need to be performed to establish whether or not TDP1 has any role in G4 and R loop formation, in particular we need to assess the effects of TDP1 inhibitor on G4 and R- loop levels induced by CPT and pyridostatin.

Contents

Chapter 1

Introduction

A typical DNA structure consists of two complementary polynucleotide chains that are multiply interwound, forming a diameter of 20A double helix as coined by Watson and Crick in 1953. This form, also known as the B-DNA, is a right-handed helix with a period of approximately 10.5 base-pairs per turn at physiological conditions. Being in relaxed form, the B-DNA is kept at its lowest energy state. This underwound state of DNA, is referred as negatively supercoiled. Although in the beginning the high level of DNA supercoiling was seen as physicochemical tricks that were not directly clear to biology, it has later become clear that negatively supercoiled DNA is homeostatically maintained to provide means for the genome to be compacted and also allowing a mechanism where the access of genetic information is tightly regulated. When a DNA helicase separates the two strands of a (−) supercoiled DNA, the remaining DNA initially becomes relaxed, however further strand separation causes the accumulation of positive (+) supercoils.

The topological properties of DNA are defined by the linking number (Lk) which refers to the number of times the two helical strands are interwound. The Lk for a relaxed molecule, termed Lk0, is equal to the number of base pairs divided by the period of the DNA helix. Most theoretical and computational modeling of DNA assumes that the double helix behaves as an isotropic elastic rod. Due to its compacted structure, local melting of the supercoiled structure is required to allow access to DNA for transcription and replication. This led to the DNA to be overwound or having a positively supercoiled DNA helix. of the helix changes the twist (Tw), a parameter describing the number of times the individual strands coil around the helical axis. If the DNA behaves as an isotropic elastic rod, then as the value of Tw increases, the associated torque should increase linearly. When Tw reaches a critical density, the molecule bends to form plectonemic structures in which the double helix coils about itself, a property known as writhe (Wr). The coiling of the double helix

about itself is more commonly known as supercoiling. The sum of Tw and Wr is equal to the linking number such that:

$$Lk = Tw + Wr$$

Any change in the linking number must result in a change in the twist and/or writhe such that:

$$\Delta Lk = \Delta Tw + \Delta Wr$$

As the writhe of a relaxed molecule is equal to zero, hence:

$$\Delta Wr = Wr - 0 = Wr$$

To better understand the model DNA supercoiling, one can use a piece of rubber tubing. DNA is a right-handed helix, i.e. the helix spirals in a clockwise direction. Therefore, to simulate the effects of overwinding (positive supercoiling) one can introduce clockwise twist into the tubing. When sufficient twist is added, the tubing coils about itself analogously to DNA writhe. Positively supercoiled DNA coils about itself in a left-handed direction. In contrast negatively supercoiled DNA assumes a right-handed superhelical structure. Mathematically, this means that the value of Lk is lower than that of Lk0. The negative linking difference, ΔLk, is defined by:

$$\Delta Lk = Lk - Lk0$$

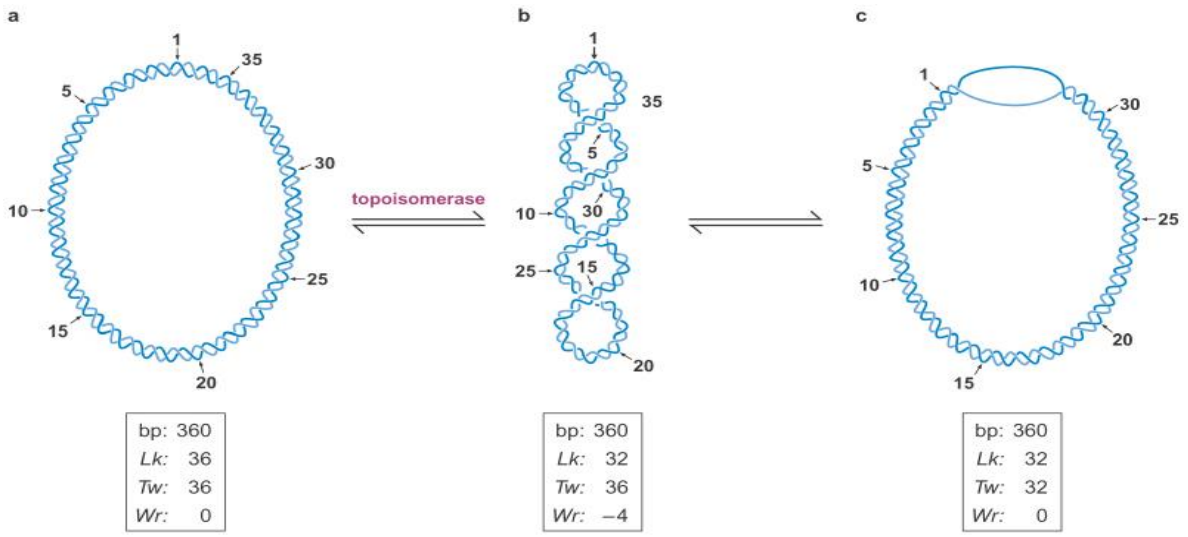


Figure 1. Diagrammatic depiction for DNA supercoiling. a) DNA supercoiling at relax state; b) DNA supercoiling with stress induced by topoisomerase; c) DNA supercoiling at a relax state with partial melting and producing a negatively coiled structure [Human Biology, Benjamin Cummings, Pearson 2004].

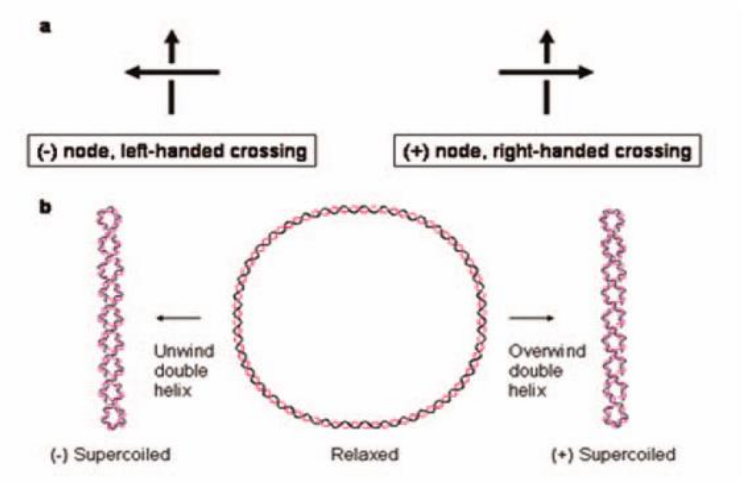


Figure 2. Topology of closed circular DNA. a) Sign convention for DNA crossings in closed-circular DNA. The convention corresponds to the normal right-hand rule in chemistry and physics: a left-handed crossing is counted as negative whereas a right-handed crossing is counted a positive. b) Conversion of relaxed DNA into negatively and positively supercoiled DNA. The description of supercoiling, the linking number, can be computed from one-half the sum of signed crossings of the red and black strands. In the case of relaxed DNA there is no writhe and the linking number, Lk is equal to the twist number, Tw. In negatively supercoiled DNA, reduction of Lk below Tw gives rise to right-handed interwound supercoils, or negatively writhe. Conversely, incrementing Lk above Tw generates left-handed interwound supercoils and positive writhe [Human Biology, Benjamin Cummings, Pearson 2004].

1.1 Biological importance of being underwound

Negative supercoiling serves as a store of free energy and provides the energy for localized, controlled melting of the DNA duplex to allow access of DNA polymerases, RNA polymerases, repair factors and recombinases to the internal nucleotide sequence. In addition to accessing these sequences, many DNA metabolic processes have additional specific DNA conformational requirements. For example, transcriptional regulation, through enhancers, and synapsis during site-specific recombination both require that distant DNA sites come in close physical proximity. All these require topological regulations to underwind (to facilitate strand separation) and overwind (to inhibit strand separation). Topological regulation in the bacterial genome is enzymatically regulated by a group of Topoismerases, namely DNA gyrase, Topoisomerase 1 and Topoisomerase IV. DNA gyrase introduces negative supercoils into DNA and Topoisomerase I relaxes the highly superhelical tension. Topoisomerase IV resolves DNA knots and catenanes, an important feature for genome duplication and segregation. In the eukaryotic where the regulation is much less clear, regulation of DNA topology has been based on the twin supercoiling domain model proposed by Liu and Wang in 1987 to explain how transcription by RNA polymerase can be used to stimulate DNA supercoiling. The model postulates that rotation of the RNA polymerase-RNA complex around the DNA helical axis during transcription becomes increasingly difficult as the size of the growing RNA chain increases. The rotation reaches a critical point when it is more feasible energetically to rotate the DNA on its axis rather than rotate the transcription complex and any associated proteins around the DNA. Further translocation of the polymerase generates transient DNA supercoils, positive supercoils in front of the polymerase and negative supercoils behind it.

1.2 DNA topoisomerases

DNA cleavage by all topoisomerases is accompanied by the formation of a transient phosphodiester bond between a tyrosine residue in the protein and one of the ends of the broken strand. In this way they regulate DNA superhelicity and solve topological problems arising during DNA metabolism. DNA superhelicity is influenced by topoisomerases I and

II (Top1p and Top2p), encoded by the *TOP1* and *TOP2* genes. Topoisomerases are classified into two main subfamilies in function of structural and mechanistical differences: type I cleaves only one strand of the double helix and type II cleaves both strands to generate a double-strand break. Type I topoisomerases are further classified in type IA subfamily members if the protein link is to a 5' phosphate and type IB subfamily members if the protein is attached to a 3' phosphate. Further division of the type II subfamily in IIA and IIB is based on structural considerations. Members of the same subfamily are similar in structure and mechanism of action [Wang, *et. al.*, 2002; Champoux, *et. al.*, 2001].

1.2.1 Type IA

The topoisomerases in the type IA subfamily share several properties. They are all monomeric (except *Methanopyrus kandleri* reverse gyrase) and require Mg^{2+} for the DNA relaxation activity. The type IA enzymes catalyse DNA strand passage by an 'enzyme-bridging' mechanism, in which the DNA ends, which are created in the DNA breakage reaction, are bridged by the topoisomerase [Lima *et. al.*, 1994]. Movements of the enzyme-bound DNA ends relative to each other mediate the opening and closing of the DNA gate [Lima *et. al.*, 1994]. During the cleavage of a DNA strand, a covalent attachment of one of the DNA ends to the enzyme is formed through a 5' phosphodiester bond to the active amino acid tyrosine. Negative supercoils are substrates for the relaxation reaction but relaxation requires an exposed single-stranded region within the substrate DNA to complete the reaction [Wang *et. al.*, 1996]. Bacterial DNA topoisomerases I and III, Yeast DNA topoisomerase III, *Drosophila melanogaster* DNA topoisomerases IIIα and IIIβ, Mammalian DNA topoisomerases IIIα and IIIβ are some of the representative from this subfamily [Wang *et. al.*, 2002].

The *E. coli* Top1, for example, can be divided into three domains.

- N-terminal domain, composed of the first 582 amino acids, corresponds to a core "cleavage/strand passage" domain containing the active site tyrosine at position 319;
- Zn^{2+}-binding domain, consisting of 162 amino acids that contain three tetracysteine motifs. This region of the protein is required for the strand passage reaction;

- C-terminal domain, enriched with basic amino acids and contributes to substrate binding preferentially for single-stranded DNA [Zhang *et. al.*, 2001].

While the N-terminal domain, containing the catalytic tyrosine, is highly conserved, the Zn^{2+}-binding domain is not. Notably, both yeast and *E. coli* topoisomerases III lack a Zn^{2+}-binding domain, while the human Top IIIβ has an incomplete Zn^{2+}-binding domain.

Further, all the Tops IIIα, even from phylogenetically different groups, possess a highly conserved set of eight CXXC motifs that could correspond to four zinc fingers [Wang *et. al.*, 2002].

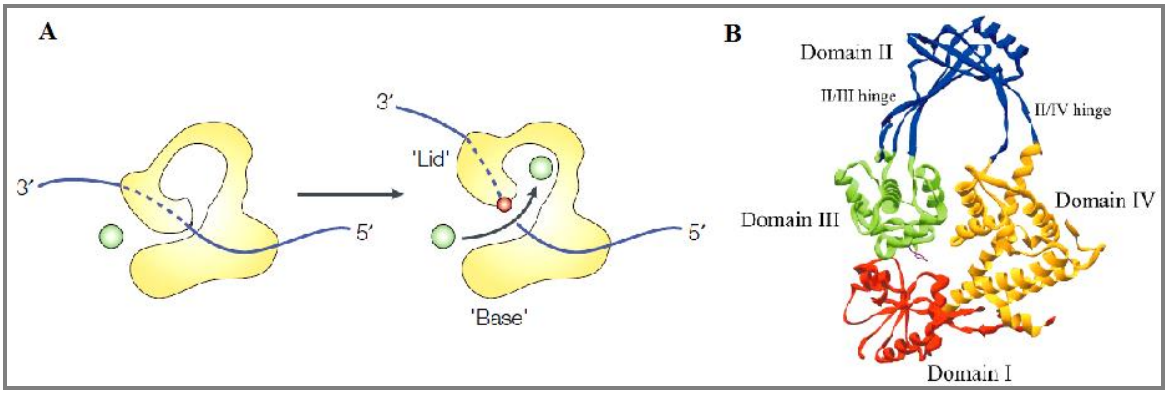


Figure 3. a) Mechanism of action of type IA topoisomerases. On transient breakage of a DNA strand, the 5'end is covalently attached to the active-site tyrosyl group (red circle) in the 'lid' of the enzyme, while the 3'end is noncovalently bound to the enzyme. The opening of the gate allows the passage of another DNA strand (green circle) [Wang *et. al.*, 2002]. b) Crystal structures of *E. coli* Top 1 [Champoux *et. al.*, 2001].

1.2.2 Type IB

This subfamily of topoisomerases is composed of three classes of enzymes: the topoisomerases I found in all eukaryotic cells, the poxvirus topoisomerases (vaccinia enzyme), and the prokaryotic Top V from *Methanopyrus kandleri* [Wang *et. al.*, 2002]. They share no sequence or structural homology with other known topoisomerases. The type IB subfamily members can relax both positive and negative supercoils, and relaxation goes to completion, hence, negating the need for partially single-stranded DNA as substrate. The type IB topoisomerases form a covalent intermediate in which the active site tyrosine becomes attached to the 3' phosphate end of the cleaved strand. The enzymes contain no bound metal ions, and DNA relaxation does not require Mg^{2+} [Champoux *et. al.*, 2001].

The type IB enzymes act by a 'DNA rotation', rather than by an 'enzyme-bridging' mechanism. When a DNA-bound type IB enzyme transiently cleaves one of the DNA strands, only the side of the DNA double helix that is upstream of the nick, the side containing the protein-linked 3' end of the broken strand, is tightly bound to the enzyme. Interaction between the downstream side of the dsDNA and the enzyme is mostly ionic physiologically, presenting a low barrier to a 75° rotation between the DNA and protein. The DNA segments that flank a transient nick can therefore rotate relative to each other by turning around one of the single bonds that opposes the nick [Champoux *et. al.*, 2001; Wang *et. al.*, 2002].

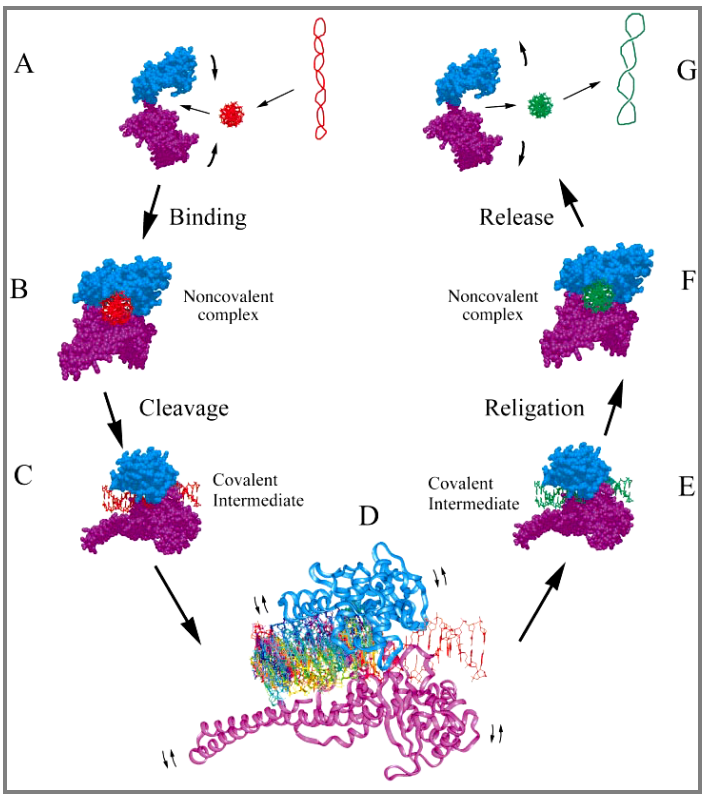


Figure 4. The human Top1 controlled rotation mechanism. A highly negatively superhelical DNA is converted to a less supercoiled state. Human Top1 is rendered as a bilobed structure with core subdomains I and II forming the "cap" lobe, cyan. The "catalytic" lobe is magenta. 30° intervaling DNA rotation states are different-colored in step (D). Small movements of the protein (small arrows) may be allowed during the events of controlled rotation [Stewart *et. al.*, 1998].

Human DNA Top I, a 91-kDa protein, is subdivided into four distinct domains:

- The N-terminal domain is a 214 amino acids sequence and is a hydrophilic, unstructured, and highly protease-sensitive region of the protein. It is dispensable for relaxation activity in vitro and presents four nuclear localization signals and sites for interaction with other cellular proteins, including nucleolin, SV40 T-antigen, certain transcription factors, p53, and the WRN protein.

- The core is a 421 amino acid domain that binds DNA and contains all of the catalytic residues except the active site tyrosine suggesting to be highly conserved. It has further subdivided into core subdomains I, II, and III.
- A protease-sensitive and poorly conserved linker domain comprising 77 amino acids connects the core domain to the 53 amino acid C-terminal domain. It is dispensable and assumes a coiled-coil structure in the 3D protein.
- The C-terminal domain that contains the active site such as the catalytic Tyr723.

In the 3D structure, the protein assumes a "lobulated" conformation. One of these lobes comprises core subdomains I and II and forms the "cap" of the protein. The front end of the cap consists of a pair of long α-helices in a V-shaped configuration. The other lobe forms a base that gets in touch with the DNA and consists of core subdomain III and the C-terminal domain. This second lobe is connected to the cap by a long α-helix labeled the "connector". On the side opposite to this α-helix is pair of opposing loops called the "lips". Opening and closing of the protein clamp during DNA binding and release must involve the breaking of this interaction between the lips and the lifting of the cap away from the base [Stewart *et. al.*, 1998].

The enzyme has a strong preference for binding to supercoiled DNA over relaxed DNA and nicks the DNA with a preference for a combination of nucleotides that extends from positions -4 to -1. The preferred nucleotides in the scissile strand are 5'-(A/T)(G/C)(A/T)T-3' with the enzyme covalently attached to the -1 T residue [Jaxel *et. al.*, 1991]. Other protein-DNA interactions, in addition to the Tyr723, play an important role in cleavage site selection, as residues Arg488, Arg590, His632 of human enzyme [Interthal *et. al.*,2001].

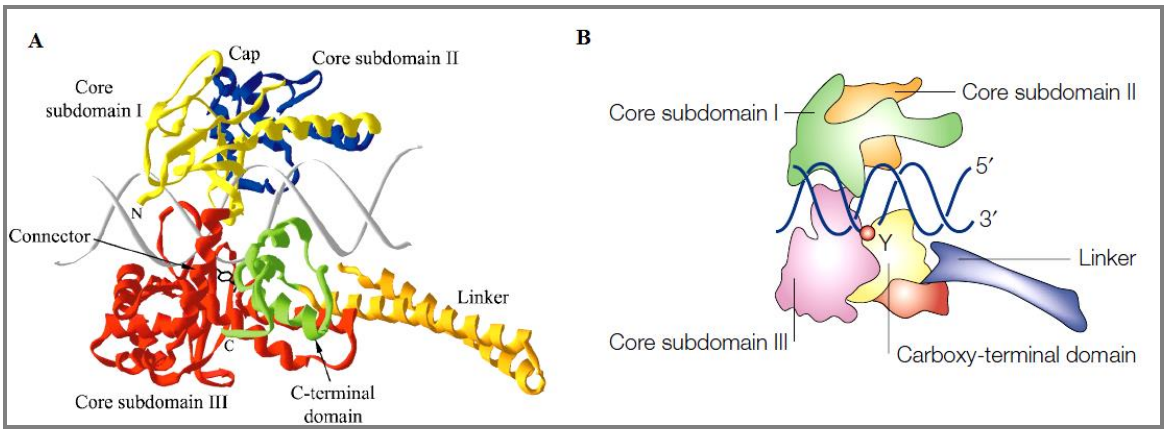


Figure 5. a) Structure of human Top I. Core subdomains I, II and III are colored yellow, blue and red, respectively. The linker and C-terminal domain are shown in orange and green, respectively. The active site tyrosine is shown in black ball and stick. The long α-helix that connects the cap to the base of the core is labeled "Connector" [Champoux *et. al.*, 2001]. b) Schematic of Type IB enzymes where an intermediate between a DNA fragment and a type IB DNA topoisomerase is shown. The 3' end of the broken DNA scissile strand is covalently linked to the active-site tyrosyl group (Y) of the enzyme (red circle) [Wang *et. al.*, 2002].

1.2.3 Type II

In contrast to the type I enzymes, the type II DNA topoisomerases are ATP-dependent dimeric enzymes. The mechanism of action of these enzymes involves covalent attachment of each subunit of the dimer to the 5' end of the DNA through a phosphotyrosine bond. The dimeric enzymes bind duplex DNA and cleave the opposing strands, while a conformational change pulls the two ends of the cleaved duplex DNA apart to create an opening in what is referred to as the gated or G-segment DNA. A second region of duplex DNA from either the same molecule or a different molecule, the transported or T-segment, is passed through this open DNA gate. This feature of the reaction explains why the linking number is changed in steps of two when the supercoiling of a circular DNA is changed. Besides ATP hydrolysis, the reactions also require Mg^{2+} [Champoux *et. al.*, 2001]. The crystal structures of several members reveal that the active site tyrosines are situated in a

helix-turn-helix (HTH). As with the type IB enzymes, a highly conserved arginine residue is implicated in catalysis by its close proximity to the active site tyrosine [Berger *et. al.*, 1996].

Within type IIA are Bacterial gyrase (DNA Top IV), Phage T4 DNA topoisomerase, Yeast DNA Top II, *Drosophila* DNA Top II and Mammalian DNA Tops IIα and IIβ.

All prokaryotic type II topoisomerases contain two different subunits and are therefore heterotetrameric in structure, whereas the eukaryotic enzymes are homodimers. Among all of the type II enzymes, DNA gyrase stands alone as the only enzyme capable of using the energy from ATP hydrolysis to introduce negative supercoils into the DNA. These enzymes use ATP hydrolysis to transport one DNA double helix through another, passage that reflects in several topological transformations, including catenation and decatenation of dsDNA rings, and the relaxation of positively or negatively supercoiled DNA [Berger *et. al.*, 1996]. Different members of the type II family can be distinguished by their relative proficiency at DNA relaxation versus decatenation/catenation, likely to reflect their specialized roles in the cell [Wang *et. al.*, 2002; Champoux *et. al.*, 2001].

1.3 Modulation of Topoisomerase

1.3.1 DNA replication

The initiation of replication begins with the opening a DNA region by unpairing the DNA strands. In *E. coli* plasmid-replication systems, a negatively supercoiled template is usually required for initiation, hence, requiring the ability of bacterial gyrase to introduce negatively supercoil for the initiation of DNA replication. In yeast, either TopI or TopII, both of which can relax positive and negative supercoils, are required for their functions in the movement of growing forks [Wang *et. al.*,1996].

The topological consequences of an advancing replication fork, and the roles of different DNA topoisomerases, depend on whether the replication machinery is allowed to rotate in the cellular milieu. If the replication machinery cannot rotate around the helical axis of the unreplicated DNA, it can force the helical intertwines of the DNA and the DNA becomes

positively supercoiled ahead, while the replicated bubble becomes progressively larger behind. In contrast, if replication machinery is allowed to rotate, the positive supercoils ahead of it might be redistributed to the region behind it, leading to intertwining of the pair of replicated DNA segments and positive supercoiling of the unreplicated DNA behind the fork [Wang *et. al.*, 2002].

The mechanisms of the various subfamilies of DNA topoisomerases predict that the positive supercoils that are generated by replication can be removed by a type IB or a type II enzyme [Wang *et. al.*, 2002; Champoux *et. al.*, 2001]. Because of the inefficiency of type IA enzymes at removing positive supercoils that do not have a pre-existing single-stranded region in the DNA, they are expected to be less suitable than type IB or type II enzymes for solving the topological problems that are associated with DNA chain elongation in replication [Wang *et. al.*, 2002].

1.3.2 Transcription

The best example of supercoiling generating process is transcription. Due to the huge molecular mass of the RNA polymerase, the DNA template is forced to rotate around its axis as the double helix passes through the transcriptional machinery [Liu *et. al.*, 1987]. The upstream DNA becomes untwisted, while the downstream DNA becomes positively supercoiled. This enormous torsional stress might inhibit an efficient transcription [Capranico *et. al.*, 2010], hence, playing its role in modulating gene expression. DNA torsional tension may coordinate also local or regional transcription by modulating the stability of protein-DNA interactions, e.g. interactions between transcription machinery factors and promoters or coding regions [Capranico *et. al.*, 2010].

DNA topoisomerases are required to restore the topological conformation of the DNA. It is however, the requirement for DNA topoisomerases in transcription depends on the ability of the transcription apparatus (which includes the RNA polymerase, proteins associated with the polymerase, and the nascent transcript and its associated proteins) to rotate. In transcription, the elongation of a nascent transcript resembles those of the

elongation step in replication except that it does not involve a continuous separation of the parental DNA strands, hence, modulating the local supercoiled state of the DNA, rather than fulfilling a topological necessity [Wang *et. al.*, 2002]. Another important factor which is the distribution of promoters in divergent orientation can reinforce DNA supercoiling upstream transcription start sites by untwisting the double helix and directly inducing plectonemes [Seila *et. al.*, 2008]. The type IIA enzymes in prokaryotes, especially DNA gyrase, remove positive supercoils, whereas the type IA enzyme DNA Top I is important in the removal of negative supercoils [Wang *et. al.*, 1996]. In *E. coli topA* mutants, the lack of DNA Top I induces an excessive negative supercoiling, a condition favours base-pairing between the nascent RNA and its template strand ('R-looping'), a condition which is often implicated in genome instability [Drolet *et. al.*, 1995].

In the eukaryotes, DNA Top I is present in actively transcribed regions [Wang *et. al.*, 2001], mainly functions to relax transcription-generated DNA supercoils [Wang *et. al.*, 2002]. Top I is also able induce hyperphosphorylation of Rpb1 subunit of the RNAPII to facilitate their escape from pausing sites. This escape requires the formation of an active TFIID–TFIIA protein complex on the promoter and could thus act as co-activator [Capranico *et. al.*, 2010; Baranello *et. al.*, 2010; Wang *et. al.*, 2002]. Silencing the Top1 gene causes genome instability in human cells as chromosomal translocations are increased, as a consequence of unresolved topological problems arising when the replication forks encounter translocating transcriptional machineries [Capranico *et. al.*, 2010].

Physiologically, in the eukaryotes, the DNA is compacted into nucleosomes as a means of packaging into the nucleus. The compacted DNA supramolecules are inaccessible to enzymes and proteins (polymerases) for transcriptopn initiation but also restricts transcription elongation along the DNA. Because of the strong binding energy between nucleosomes and DNA, chromatin remodelers are required to disrupt or to slide nucleosomes, hence, providing a mean for transcription regulation. Special ATP deoendent protein complexes called chromatin remodelers are able to remove or slide nucleosomes. Notably, *in vitro* experiments have shown that these chromatin remodeling activities directly generate torsional stress of DNA in the presence of nucleosomes which has to be solved by DNA topoisomerase [Baranello *et. al.*, 2010; Wang *et. al.*, 2002]. Additionally, in cases

where a stably base-paired region between a nascent RNA and its template strand has formed, a topoisomerase-mediated untwining of the transcript from the template strand, rather than the nucleolytic removal of the DNA-bound RNA by RNase H, could also be used to salvage the transcript [Pommier *et. al.*, 2006; Drolet *et. al.*, 1995].

1.3.3 DNA recombination

Resolution of an intermediate of recombinational repair is characterized by topological tension, can be solved by DNA topoisomerases [Wang *et. al.*, 2002]. The role of these enzymes in the modulation of recombination has been widely investigated pointed to an important role by the type IA enzymes. *Escherichia coli* cells that lack both type IA DNA topoisomerases are non-viable. Additionally, investigations have shown links between this lethal genotype to a gene involved in the process of recombination, the *recA* gene. There are indications that type IA enzymes may be involved in RecA-mediated recombination and that they can specifically resolve recombination intermediates before chromosome segregation [Zhu *et. al.*, 2001]. Also Top IIB might be involved in recombination, and more exactly, in meiotic recombination. It presents different homologues in various organisms, of which the *SPO11* gene product of *S. cerevisiae*, which binds in a covalent manner to the 5'ends of broken DNA, found in meiotic recombination hot-spots [Keeney *et. al.*, 2008].

1.3.4 Role of Topoisomerases in the chromosomal topological organization

Chromatin compaction, chromosome segregation, and DNA topology are intricately interrelated both in prokaryotes and eukaryotes [Nordstrom *et. al.*, 2001].

As the unreplicated segment of parental DNA becomes very short, a type II DNA topoisomerase is required to convert the residual intertwines between the parental strands into intertwines between the newly replicated daughter molecules so that the segregation of the newly replicated molecules can be established. Evidences show that in *E. coli* and yeast, DNA topoisomerases II are indispensable in chromosome segregation [Wang *et. al.*, 1996].

The eukaryotic Top II is involved in chromatin and chromosome condensation during mitosis and in apoptotic chromatin condensation [Durrieu *et. al.*, 2000]. Further, mammalian Top II has a structural role in chromosome. Its α-isoform, for instance, is a main non-histone protein in the axial core or scaffold of metaphase chromosomes [Wang *et. al.*, 2002]. However, this role remained open for debate as neither of the Top II isomers are not immobile structural components of the chromosomal scaffold [Christensen *et. al.*, 2002].

Chromosome condensation and decondensation is a key event not only during chromosome segregation but also in the fine tuning of gene expression in higher eukaryotic beings. In general, whenever a long chromatin fiber undergoes a structural change, any accompanying changes in its twist could require the catalytic action of one or more DNA topoisomerases and in some processes, chromatin remodeling is necessary [Capranico *et. al.*, 2010]. The association of *Drosophila* DNA Top II and human DNA Top IIβ with ATP-dependent chromatin remodeling complexes has been reported. A main Top1 function may be the regulation of nucleosome remodeling by modulating the torsional tension generated by the assembly and/or disassembly of nucleosomes [Wang *et. al.*, 2002]. In support to this, deletion and mutation in the Top1 gene sees an up-regulation in gene expression as well as increased acetylation of core histones at telomeric and sub-telomeric regions in S. cerevisiae [Lotito *et. al.*, 2008; Wang *et. al.*, 2002].

1.4 Camptothecin: a specific Topoisomerase inhibitor

The immense interest in topoisomerases in recent years derives not only from the recognition of their crucial role in managing DNA topology, but also because a wide variety of topoisomerase-targeted drugs have been identified, many of which generate cytotoxic lesions by binding to the interface between Top1 and DNA, thus, trapping the enzymes in covalent complexes on the DNA which inhibit the function of Top1 to relegate DNA after cleavage reaction [Liu *et. al.*, 2000]. These features enable topoisomerase an interesting therapeutic target in clinical use.

Camptothecin (CPT) is a natural product that was isolated from the Chinese tree *Camptotheca acuminata* by Wall and Wani and was shown to inhibit the growth of cancer cells in cell culture [Wall *et. al.*, 1966]. It penetrates vertebrate cells readily and targets Top1 within minutes of exposure to low or even sub-micromolar drug concentrations.

Camptothecin is a 5-ring heterocyclic alkaloid that contains an hydroxylactone within its E-ring that is unstable at physiological pH. The active lactone form is in equilibrium with its inactive carboxylate derivative, characterized by an open E-ring (Figure 7) [Pommier *et. al.*, 2006].

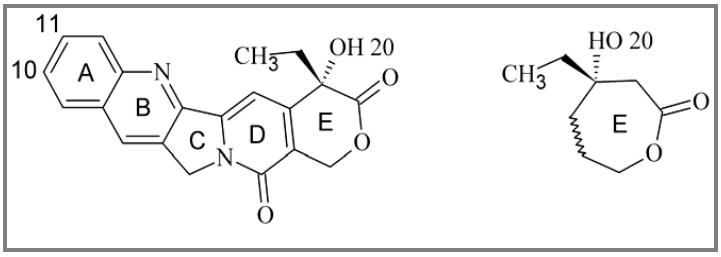


Figure 6. Structure of CPT and CPT E-ring [Pommier *et. al.*, 2009]

CPT specifically targets Top1. Structural studies established that camptothecin interacts with active site amino acid residues and DNA base pairs at the cleavage site, preventing strand religation and therefore increasing the half-life of the Top1–DNA cleavage complex (Top1cc). Formation of the cleavage complex is a critical event during the cell cycle since cell vitality is seriously compromised by poisoning this complex [Pommier *et. al.*, 2009]. This effect is highly reversible both in vitro and in vivo [Capranico *et. al.*, 2010]. Although reversible, it becomes lethal when collisions occur between Top1cc and an advancing replication fork, converting the single-strand breaks into irreversible double-stranded DNA (dsDNA) breaks and imminent cell deaths through the activation of DNA damage checkpoints [Capranico *et. al.*, 2010].

In the less studied aspect of CPT effect, the early effect inhibition of Top1 by CPT has been seen to introduce topological stress and hence, a major reorganization of chromatin is in place in response to Top1 mediated DNA damage.

Duann *et. al.* (1999) observed an increased in the linking number of the episomal DNA in a Top1-dependent manner and within 3 min of CPT treatment, indicating that Top1 is actively involved in maintaining the negatively supercoiled state of episomal DNA. Since Top1 must mainly remove positive supercoils, and as a consequence of CPT effect, inhibition of Top1 by camptothecin would result in more positively supercoiled DNA. Further support provided by single molecule manipulation which monitored the dynamics of human Top1 in the presence of Topotecan, an analog to CPT. The uncoiling activity of Top1 due to topotecan effect sees a 20-fold slower and hinders the removal of positive supercoiling. It is shown; however, the camptothecin-induced chromatin reorganization only alters nucleosome conformation and/or position along the studied DNA regions but not being removed from DNA template [Capranico *et. al.*, 2010; Wang *et. al.*, 2002; Duann *et. al.*, 1999]. One model suggested that Top1 inhibition *in vivo* lead to the transient wave of positive torsion downstream of Pol II destabilizes genic nucleosomes to allow progression, and the transient negative torsion stabilizes nucleosome formation behind Pol II to maintain chromatin structure. This supports a balance between destabilization of nucleosomes for Pol II passage and maintenance of chromatin structure for chromosomal integrity.

Since CPT readily penetrates into cells, its effects is seen as early as the first 2 minutes of cell treatment with 10uM, activating the initiation step of transcription [Marinello *et. al.*, 2013] and the expression of certain genes in human cells [Collins *et. al.*, 2001]. CPT triggers hyperphosphorylation of the largest subunit (Rpb1) of RNA polymerase II (Pol II) which apparently occurs on Ser-5 residues of heptapeptide repeats of the carboxy-terminal domain (CTD) possibly mediated by Cdk7, component of TFIIH [Sordet *et. al.*, 2008; Khobta *et. al.*, 2006]. In addition to this CPT-induced effect is reported to contribute to the disruption of the large inactive positive elongation factor b (P-TEFb) complex, which is kept an inactive state in the multisubunit 7SK snRNP. Disruption of P-TEFb releases free active P-TEFb complex (containing the Cdk9 subunit) that correlates with a concomitant hyper-phosphorylation of RNAPII, which in turn alters the levels and distribution of the RNAPII along transcribed genes [Amente *et. al.*, 2009].

Consequently, chromatin-bound Pol II is redistributed along transcribe genes in human cancer cells by enhancing the escape of Pol II from promoter proximal pausing sites [Khobta

et. al., 2006]. In HIF-1 α gene CPT stimulates antisense transcription in a Cdk 9/7 dependant manner. Consequently, this may lead to the de-repression of antisense transcription and reduced synthesis of mRNAs [Baranello *et. al.*, 2010].

It has been proposed that a sustained camptothecin interference with Pol II regulation that increases its pausing efficiency, may eventually lead to a general transcriptional stress which involves a more accessible chromatin conformation through histone modifications, specifically at transcribed loci in human cells. CPT causes a decrease of histone H1, and acetylation of H3 and H4 on the gene cluster of human chromosome 1 but not at repressed α-satellite DNAs [Khobta *et. al.*, 2006]. Unlimited to HIF-1α, the effect of CPT is further demonstrated in a genome-wide scale. CPT-stabilized Top1ccs is shown to have a specific and dynamic impact at divergent CpG-island (CGI) promoters in human cells. CPT enhances the levels of antisense RNAPII transcripts upstream of transcriptional start site (TSS) at divergent CGI promoters [Marinello *et. al.*, 2013]. This effect depends on cellular Top1 levels while is independent from DNA replication. Divergent promoters have been reported in several eukaryotic cells raising the question of how this process contributes to regulation of gene activity and chromatin structure. Divergent CGI promoters may constitute a set of gene control regions highly sensitive to exogenous and endogenous perturbation of Top1 activity.

Moreover, Top1 can undergo protein degradation within 1 h of camptothecin treatments in certain cancer cells. The removal of Top1 cleavage complexes and DNA break processing are transcription-dependent, and coupled to ubiquitination and degradation of Top1 and Pol II through the 26 S proteasome pathway [Sordet *et. al.*, 2008; Desai *et. al.*, 2003]. Further, blocking of Top1 activity by camptothecin promotes an increase of local negative supercoiling behind the transcriptional apparatus thus stabilizing R-loop structures, especially at nucleoli and mitochondria, and the kinetic closely parallels Top1cc formation. These non-B structures may increase genome instability. Notably, in bi-directional promoters, the rate of transcription-generated negative torsional tension is necessarily higher as divergent RNAPIIs elongate in different directions and this event could further ease R-loop formation [Marinello *et. al.*, 2013; Sordet *et. al.*, 2009]. Top1 inhibition by

camptothecin has also been reported to block the SR-kinase activity of Top1; in this case splicing might be inhibited because of ASF hypophosphorylation [Pommier *et. al.*, 2006].

In this scenario, CPT not only makes clear its antitumor effect by way of DNA replication and DNA damage checkpoints but can also interfere with transcription regulation, that can lead to alterations of gene expression patterns that may be relevant for cancer therapy, e.g. HIF-1α [Marinello *et. al.*, 2013; Capranico *et. al.*, 2010].

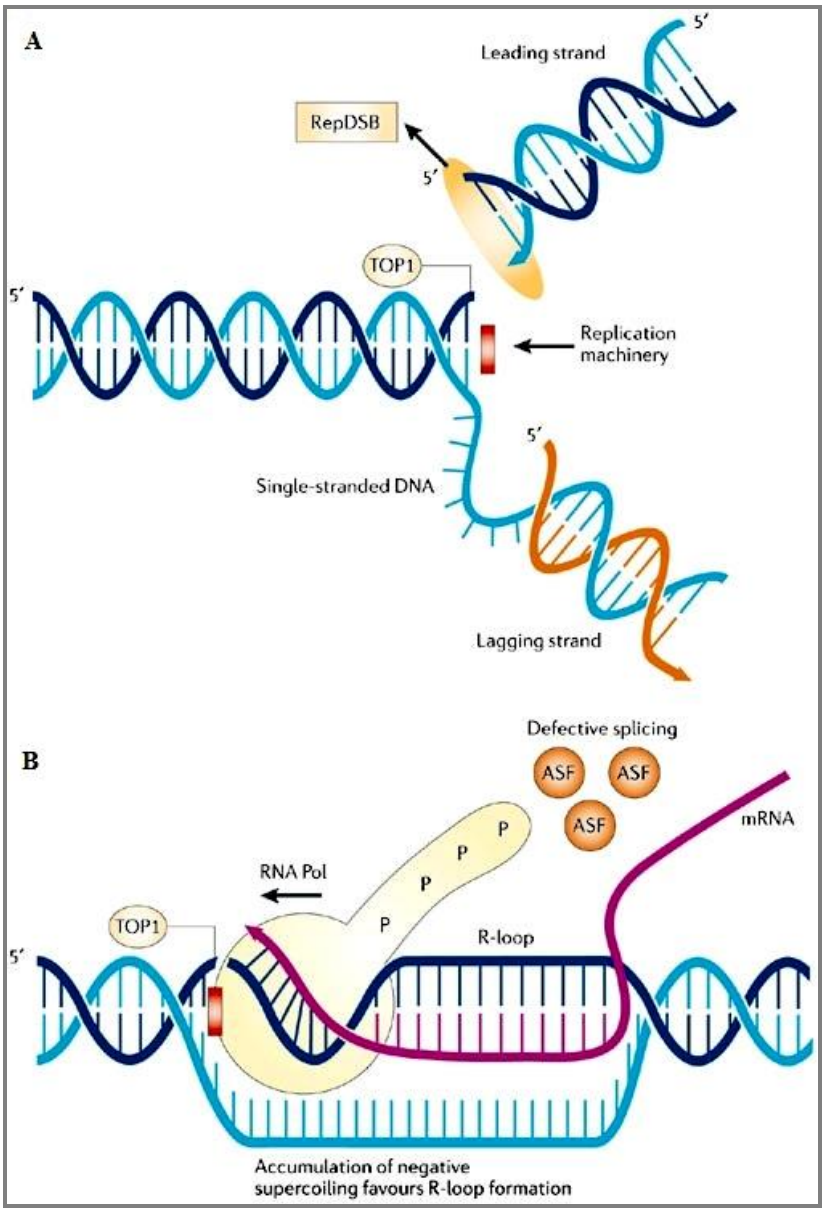


Figure 7. a) Conversion of Top1cc into DNA damage by replication-fork collision. The consequence is a replication double-strand break (RepDSB). b) Conversion of Top1 cleavage complexes into DNA damage by transcription: the RNA–DNA duplex prevents the religation of the Top1cc, and Top1 inhibition leads to an accumulation of negative supercoiling that could promote the formation of an R-loop. Inhibition of Top1 SR-kinase activity would also inactivate splicing because of ASF hypophosphorylation [Pommier *et. al.*, 2006].

1.5 Non B-form DNA structures

Non cell techniques such as circular dichroism and x-ray crystallgraphy has enabled the discovery secondary structures in DNA, which since its revelation in 1953 has been predominantly B-form (B-DNA) and right-handed double helical structure. Secondary DNA structures containing regions of single-stranded DNA that can adopt a variety of alternative conformations based on particular sequence motifs and interactions with different proteins. Cruciforms, Z-DNA, triplexes and G-quadruplexes are some of the non B-form secondary structures. Their non-random presence in the DNA regions involved in regulations has sparked increasing interests on their functional roles *in vivo*.

1.5.1 G-quadruplex motifs and structures

G-quadruplex structures are stacked nucleic acid structures that can form within specific repetitive G-rich DNA or RNA that have sequence motif G≥3NxG≥3NxG≥3NxG≥3. There are >375,000 predicted G4 motifs in the human genome and >1,400 G4 motifs have been predicted in *Saccharomyces cerevisiae*. The motif confers the ability to form a four stranded structure with two adjacent guanine are bonded by the Hogsteen hydrogen forming a square planar (G-quartet). Stacked G-quartets form a G4 structure, and the intervening sequences are extruded as single-strand loops except in tetramolecular G4 structures which may lack loops. The sequence and size of the loop regions varies, usually small (1–7 nucleotides (nt)). Smaller loops and longer G-tracks result in more stable G4 structures. This structure is stabilized by monovalent cations that occupy the central cavities between the stacks, neutralizing the electrostatic repulsion of inwardly pointing guanine oxygens. G-quadruplexes are stabilized by K+ or Na+ ions at physiological temperature and pH *in vitro*.

G4 motifs are abundant in specific chromosomal domains, genomic regions, and genes. In human cells, the telomeres, rDNA, immunoglobulin switch regions (S regions), some variable number tandem repeats (VNTRs), and some single copy genes are all enriched for

G4 motifs, meiotic double-strand break (DSB) sites, and transcriptional start sites (TSSs; often near promoters), hence, providing a considerable potential repertoire for formation of diverse structures that may correlate with specific functions.

1.5.2 Chromosomal location of G4 motifs.

G4 motifs found on the chromosomal locations forms very stable G4 structures of different topologies and are classified into various groups based on their orientation in the DNA strand. Their orientation can be parallel, antiparallel or hybrids and within one strand (intramolecular) or from multiple strands (intermolecular), and various loop structures. These features however, depended on several factors including the length and sequence composition of the total G4 motif, the size of the loops between the guanines, strand stoichiometry and alignment 11–13, and the nature of the binding cations.

Zhang et al (2013) using *in vitro* transcription method, reported that the formation of G4 was observed in the upstream but not downstream of TSS, which can be few thousands base pairs away from the promoters. G4 forms disregard to its distance of a travelling polymerase but rather the speed at which it is traveling [Zhang *et. al.*, 2013]. This has provided evidence to support computational studies in various organisms have revealed that G4 motifs are not randomly located within genomes, but rather they tend to cluster in particular genomic regions. Moreover, the G4 motifs are seen enriched at the 5’ end of the DNA among warm-blooded animals (Figure 8).

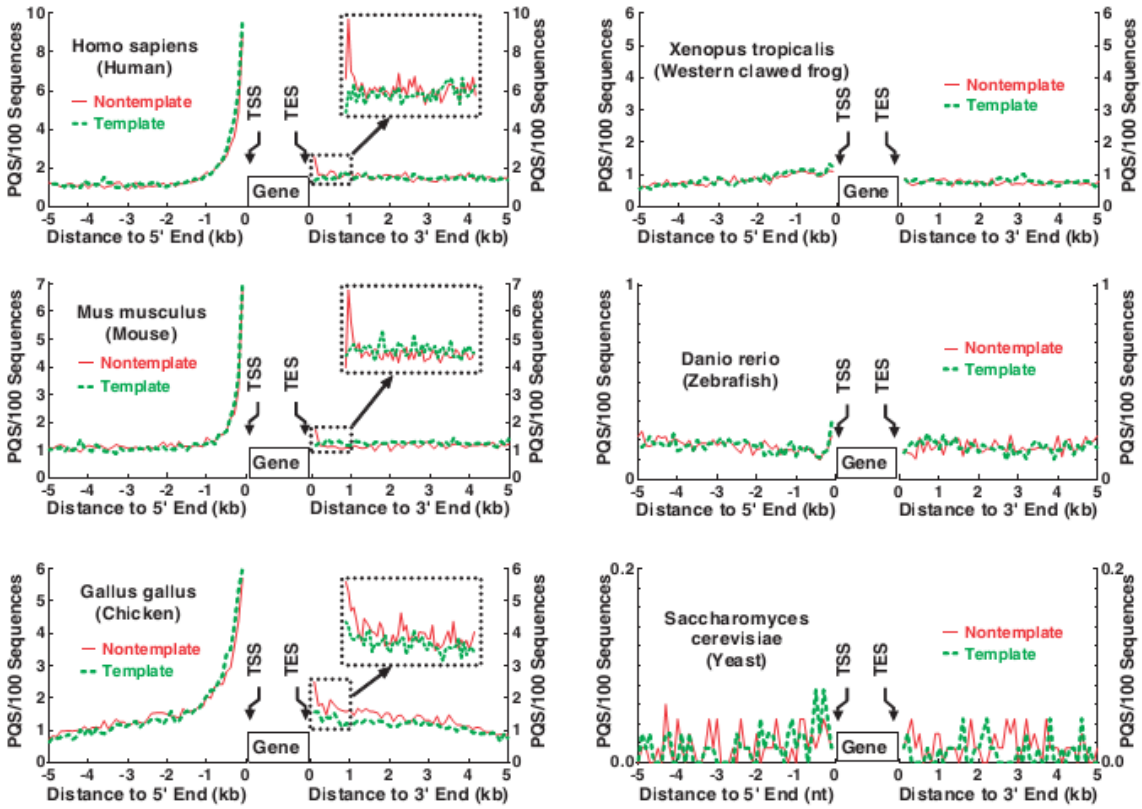


Figure 8. Distribution of PQS occurrence in the 5000bp region flanking the 5' and 3' end genes [Zhang *et. al.*, 2013]. Enrichment of G4 motifs at the 5' end of chromosomes in higher organisms but not in lower organisms.

1.5.3 Telomeric G4 structures

Telomeres are special structure at the extreme ends of eukaryotic chromosomes, composing of a DNA component and multiple protein components. They provide protections from enzymatic end- degradation. The DNA component of human telomere structure is a long stretch of double-stranded DNA composed of a hexanucleotide DNA repeat sequence (5'-T2AG3-3' in veterbrate and 5'-T4G4-3' in certain ciliated protozoans such as *Stylonychia lemnae*). In addition, all telomeres terminate with a 3' single-stranded G-rich DNA tail, which has the capacity to fold into a unique secondary structure called a G-quadruplex (G4). Evidence for possible G4 formation *in vivo* came from *in vitro* experiments on telomere structural proteins, such as TEBPα and TEBPβ in ciliates and Rap1

in *S. cerevisiae* which promotes the formation of G4 DNA and on the contrary to the human telomeric G-strand binding protein protection of telomeres protein 1 (POT1) promotes the unfolding of G4 structures in vitro [Wang *et. al.*, 2011]. The G4 existence in ciliates was made possible with the production of G4 antibody raised from ribosome display against parallel and anti-parallel telomeric G4T4 structures. In the *Stylonychia lemnae* only antibodies raised against antiparallel G4 structures bind to *S. lemnae* telomeres, providing important indications that antiparallel, and not parallel are present *in vivo*. Experimental controls also showed that the G4 formation is not induced by anti-G4 antibodies. Instead analysis from RNAi experiments showed that the formation of G4 depended on TEBPα to bind to the telomeric overhang and recruits TEBPβ with its highly charged carboxyl terminus. These telomeric G4 structures although are present during most of *S. lemnae* cell cycle, are resolved during replication. Removal of G4 requires phosphorylation TEBPβ and its removal from telomeres [Paeschke *et. al.*, 2005; Fang *et. al.*, 1993]. In support to these findings, the existence of G4 at the human telomere has been recently reported [Lam *et. al.*, 2013; Biffi *et. al.*, 2013]. The ability of ssTEL to form G4 could influence the competition between Replication protein A (RPA) and protection of telomeres (POT1)/telomere protection protein (TPP1) in protecting the telomeric termini. The telomeric termini is protected against DNA damage signals through the binding of POT1/TPP1 to single-stranded telomeric DNA (ssTEL). Association of POT1/TPP1 with shelterin or other telomere-associated proteins maybe required to compete against RPA binding since RPA is abundantly found in eukaryotes and having similar affinity towards ssTEL. (RPA binding to ssDNA, including telomeric overhangs, activates the ataxia telangiectasia and Rad3-related checkpoint). The formation of thermodynamically stable G4 could probably be regarded as obstacles for recruitment of telomerase and translocation of the DNA replication machinery, and their unfolding requires helicase activity or ssDNA binding proteins.

Several others ssDNA and helicases have also been shown to unwind G4 structures such as WRN, Pif1, FANCJ and BLM. Their mutations result in genomic instability, suggesting a role in processing of G4 structures is central for maintaining genome integrity. Moreover, the Human Bloom helicase (BLM) from RecQ family helicase is shown to bind with high specificity to G4 at the telomere.

1.5.4 G-quadruplex in Replication

DNA replication allows the faithful duplication of the whole genome before each cell division. During DNA replication, the double-stranded DNA double helix is separated in which one serves as the template for leading strand synthesis and the other for lagging strand synthesis. Due to the antiparallel nature of the DNA molecule, DNA polymerases moves from 3' → 5' direction on the template strand thus synthesizing new chains continous chain in the 5' → 3' direction. The lagging strand is however discontinuous as the new strand is formed in a direction opposite that of the movement of the replication fork, creating short fragments of transient single stranded DNA called the Okazaki fragments. This transient event provides a favorable condition for G4 to form, although the Okazaki fragment is protected by replication protein A (RPA). The protection is overcome as RPA since has low affinity for DNA that is made of guanines and adenines. The entire replication fork is stalled when DNA polymerase suddenly faces a four-stranded obstacle in the template which it cannot disentangle under normal conditions [Woodford *et. al.*, 1994]. These predictions were corroborated by Biffi *et. al.* (2013) that demonstrated the accumulation of G-quadruplexes during the S-phase of the cell cycle, the phase which replication occurs.

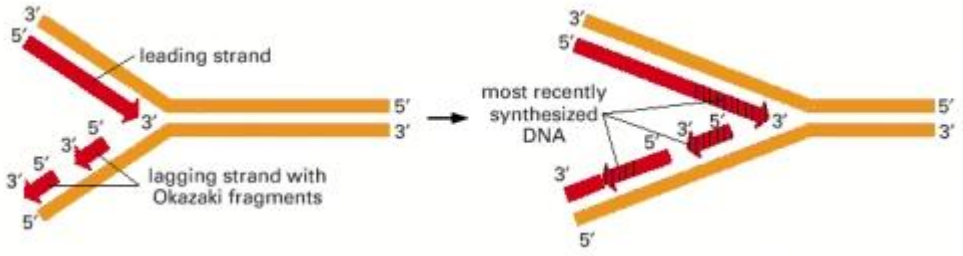


Figure 9. Replication in the leading and lagging strands [Molecular Biology of the Cell. 4th edition].

G-quadruplexes formed during replication or pre-formed before have to be resolved for the completion of DNA replication. Hence, helicases are likely to be recruited to unwind the G4 structures. Helicases such as the RecQ, Werner syndrome (WRN), Fanconi Anemia Group J (FANCJ), Bloom's syndrome (BLM) and PIF1 are some of the examples known to unwind G4 structures in vitro. Lost unwinding function often is associated to detrimental

effects on genome homeostasis which lead to genome instability. The association of these helicases with genome instability has renewed interests that G4 unwinding could be involved to suppress premature ageing and cancer. The FANCJ-depleted human cells are sensitive to a G4- specific binding compound and show elevated DNA damage and apoptosis upon exposure to the drug [Wu *et. al.*, 2008]. Moreover, FANCJ-deficient cells accumulate deletions at genomic sequences with a G4 DNA signature [London *et. al.*, 2008], suggesting that FANCJ prevents replication-associated DNA damage by removing G4 structures. Similarly, a distantly related FANCJ helicase, the *Caenorhabditis elegans* DOG-1 helicase, which its mutation causes genome-wide deletions in the G-rich regions containing G4 motifs. It is also been postulated that the human RTEL helicase, together with BLM, protect telomere from instability.

Another class of helicase, Pif1, acts at G4 motifs. Pif1 is a multi-functional DNA helicase that binds >1,000 sites in the genome of mitotic cells, of which ~10% overlap G4 motifs, representing approximately ~25% of the G4 motifs of the yeast. Absence of Pif1 slows DNA replication and DSBs to occur at G4 motifs. The yeast Pif1 is an efficient G4 unwinder of parallel intramolecular G4 and is involved in the maintenance of nuclear and mitochondrial genome stability [Boule *et. al.*, 2006]. In a study, Pif1 was shown to prevent genomic instability of a G4 forming human minisatellite sequence inserted into the S. cerevisiae nuclear genome [Ribeyre *et. al.*, 2009] and involved in the coordination of checkpoint activation following telomere uncapping [Dewar *et. al.*, 2010. The Dna2 helicase-nuclease implicated in Okazaki fragment processing, has dual functions to unwind G-quadruplex substrates with a 5’ ssDNA tail and as well as degrading G-quadruplexes in the presence of the single-stranded DNA binding protein RPA [Masuda-Sasa *et. al.*, 2008]. Lam *et. al.* (2014) in a study using chromatin immunoprecipitation followed by sequencing (ChIP-seq) reported that pyridostatin that preferentially bind to G4 motifs, causes replication and transcription-dependent damage, seen by its high γH2Ax content. These γH2Ax foci were seen overlapped with GFP–PIF1 foci in the pyridostatin-treated human cells. The current hypothesis is that G4 formation or stabilization blocks transcription and/or replication, resulting in DNA damage.

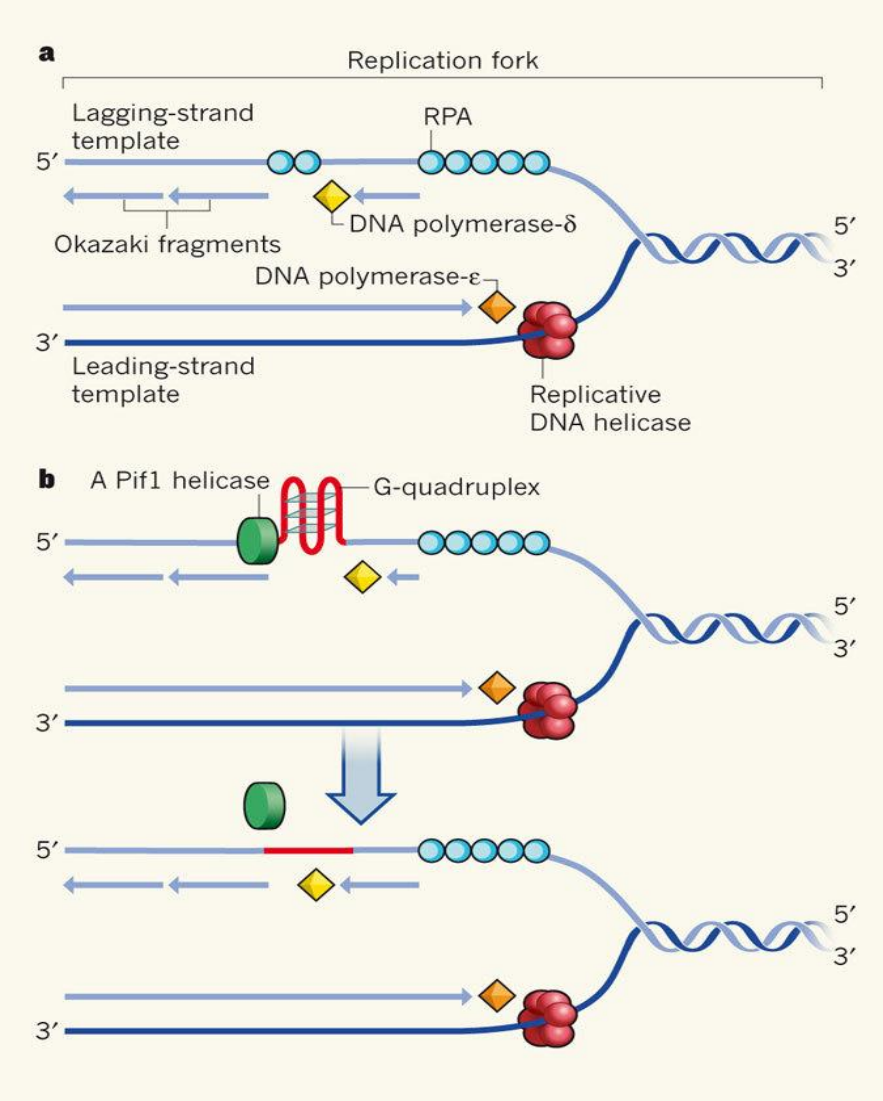


Figure 10. Diagrammatic depiction on the role of Pif1 helicase in resolving G-quadruplex during replication. a) A normal replication machinery without G4 formed. b) G-quadruplex formed in the lagging posed a blockage to an advancing replication machinery. Pif1 helicase resolves G4 to remove the block, enabling DNA polymerase-δ transcribing the Okazaki fragments. [Mirkin, 2013].

1.5.5 G-quadruplex in transcription

The high concentration and non-random distribution of G4 motifs particularly near promoter regions suggests a potential function of G4 structures in gene regulation. At least one or more G4 motifs are found within 1,000 nt upstream of the TSS of 50% of human genes. Intriguingly, bioinformatics show that the promoters of human oncogenes and regulatory genes (for example, transcription factors) are more likely than the average gene to contain G4 motifs, whereas G4 motifs are under-represented in the promoters of housekeeping and tumour suppressor genes. A similar enrichment of G4 motifs in promoter regions is found in other organisms, including yeast, plants and bacteria. Additionally, in humans, G4 motifs are less often found in the template strand than in the non-template strand. Those that are on the template strand tend to cluster at the 5′ end of the 5′UTR.

1.5.6 Possible consequences of G4 structures formed during transcription.

DNA topology as reviewed in Chapter 1.1 pointed out some of its important roles in transcriptions, both positive and negative. This includes non B-DNA such as G4. G4 structures have been postulated to form during transcription as a result from supercoiling-induced stress during transcription. First and foremost, their positions near the promoter regions especially on which strand they are formed. They are said to inhibit transcription when these structures are found on the template strand but enhancing transcription when G4 structures are formed on the non-template strands.

The proteins bound to the G4 structures could also affect transcription. The G4 structures formed in the mammalian MYC locus is one of the well-studied genes. G4 structures formed in G-rich region of NHE III_1 in the c-MYC have been reported to regulate transcription [Siddiqui-Jain *et. al.*, 2002]. MYC is a transcription factor whose expression is associated with cell proliferation. Increased levels of MYC expression are observed in 80% of human cancer cells, and this increase promotes tumorigenesis. Nuclease hypersensitive element (NHE III_1) is a highly conserved sequence located 142–115 bp upstream from the P1 promoter has been shown to be required for 80–95% of c-MYC transcription, controls >80% of the MYC transcription. NHE III_1 highly sensitive to DNase I and S1 nucleases, hence, a feature to enable it to equilibrate between transcriptionally active forms (duplex and single-

stranded DNA) and a silenced four-stranded structure under physiological conditions *in vitro.* Footprinting studies and luciferase reporter assays comparing the expression of a gene with a wild-type NHE III1 versus one with a mutated NHE III 1 that cannot form a G4 structure demonstrate that the G4 motif in NHE III represses transcription. In another study, TMPyP4, a compound that binds to and stabilizes G4 structures in the NHE III_1 reduced *MYC* transcription in lymphoma cell lines and showed antitumour activity in mice. TMPyP4 catalyzes the oxidation of DNA upon exposure to light, which results in DNA strand breakage in proximity to the binding sites.

In a study by Gonzales *et. al.*(2009), the G4 structures which are formed at the NHE III_1 act as signaling elements, through the binding of nucleolin to G4 structures as a c-myc G-quadruplex-binding protein. Nucleolin is a nucleolar phosphoprotein that is highly expressed in proliferating cells, known mainly for its role in ribosome biogenesis [Ginisty *et. al.*, 1998]; however, nucleolin also functions in chromatin remodeling [Angelov *et. al.*, 2006], transcription [Yang *et. al.*, 1994; Grinstein *et. al.*, 2007], G-quadruplex binding [Dempsey *et. al.*, 1999], and apoptosis. Nucleolin is a modular protein that can be structurally divided into three different domains as follows: the N-terminal, the central domain that includes the four RBDs, and the C-terminal domain [Ginisty *et. al.*, 1998]. It has been show that nucleolin binds with higher affinity to the c-*myc* G-quadruplex structure over its consensus NRE-RNA substrate. In addition, the ability of the c-*myc* G-quadruplex to displace the NRE-RNA from binding to nucleolin suggests that RBD1 and RBD2 preferentially bind to the c-*myc* G-quadruplex structure [González *et. al.*, 2009]. Other hypothesis on nucleolin-G4 binding associated transcription regulation is that nucleolin-mediated G4 formation in NHE III 1 inhibits *MYC* transcription by masking binding sites for *MYC* transcriptional activators, such as the transcription factor SP1 [Bochman *et. al.*, 2012] and cellular nucleic acid-binding protein (CNBP) [Borgognone *et. al.*, 2010].

1.5.7 Regulation through proteins binding to G4 structures.

Transcription may also be altered by G4 binding proteins that affect the formation and unfolding of G4 structures. The myosin D (MyoD) family proteins are transcription factors that bind to E-boxes in the promoters of several muscle-specific genes to regulate muscle development. *In vitro*, MyoD homodimers bind preferentially to G4 structures from the promoter sequences of muscle specific genes. It is hypothesized that MyoD homodimers preferentially bind to the G4 structure when G4 structures form in the promoters of E-box driven gene. Consequently, MyoD–MyoE heterodimers, which cannot bind G4 structures, bind to the E-box instead and enhance gene transcription. Genome-wide studies analysing the effects of drugs that stabilize and/or induce G4 formation have shown that the expression levels of many genes are affected by treating cells with G4 ligands. In support to this, the effects of mutations in helicases known to unwind G4 DNA on transcription genome wide were studied. For instance, in human fibroblasts deficient for the WRN or BLM RecQ helicases, the transcription of genes that are predicted to form intramolecular G4 structures is significantly upregulated which correlates with the G4 motifs but not G-richness. The genes associated with G4 motifs account for 20–30% of all transcripts that are upregulated in WRN and BML mutant cells. Despite these, the high stability and thermodynamic of G4 and the fact that G4 structures are slow formed that is resolved with the existence of chaperones (for example, TEBPβ and Rap1) that promote the formation of G4 DNA to overcome this slow formation [Bochman *et. al.*, 2012]. Thermodynamic and kinetic measurement of G4 structure formation indicates that G4 structures can form cooperatively and it is possible that other intramolecular G4 structures form as readily. Unwinding of G4 structures in a timely manner can also no longer be considered a problem given the discovery of helicases that bind and unwind G4 motifs with high efficiency.

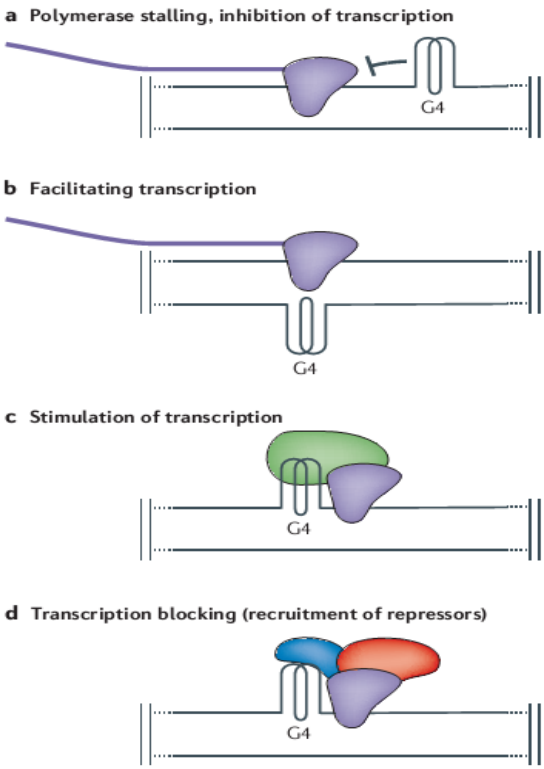


Figure 11. Putative functional roles of G-quadruplex during transcription [Bochman *et. al.*, 2012].

1.5.8 G-quadruplex binders

Numerous studies on synthetic molecules that interact with G-quadruplexes have helped demonstrate the existence and elucidate putative biological roles of these nucleic acid structures. The G4 stabilizers can be broadly classified into small molecules, non nucleoside compounds such as telomestatin, Braco-19, TMPyP4 are predicted to bind within the grooves or DNA intercalators such as porphyrins and cisplastin are compounds which tend to have large flat aromatic surfaces and are cationically charged to allow for π – stacking interactions. These platinum-bridged compounds are reported to inhibit telomerase activity in vitro, with distinct covalent linkage that could lock G4 irreversibly. Although reported to target the G4 structures at telomeres, some these compounds are also shown to bind to non-telomeric G-rich regions of the DNA promoters, preventing the access of transcription factors.

Despite the numerous G4 binders being found, only 3 of them are reviewed here as they have been used in this work. Pyridostatin has been the most recent to be studied for its ability to bind G4 structures and visualized in vitro. Pyridostatin, described by Müller *et.al.*, (2012), is a highly selective G-quadruplex-binding small molecule which alters transcription and replication of particular human genomic *loci* containing high G-quadruplex clustering within the coding region, which encompasses telomeres[57] and selected genes such as the proto-oncogene *SRC.* Downregulation of oncogenes has been shown in SRC and c-MYB in glioma cells. In the telomeres, Pyridostatin induces telomere dysfunction by competing for binding with telomere associated proteins such as human POT [Rodrigues *et. al.*, 2008]. Its biotinylated Py analogue is able to mediate the selective pull-down of telomeric fragments from genomic DNA by means of affinity matrix isolation. Pyridostatin has been shown to demonstrated high selectivity towards G-quadruplex nucleic acids, regardless of sequence variability and structure polymorphism, compared to double-stranded DNA [Müller *et. al.*, 2010].

Unlike other G4 binders, Pyridostatin is designed with the capability to adopt a flat but flexible conformation, facilitated by an internal hydrogen bonding network, prone to adapt to the dynamic and polymorphic nature of diverse G-quadruplex structures. It has an optimal electronic density of the aromatic surface to enable π–π interactions with the G-tetrad tuned by substituents (for instance alkoxy or halogens capable of altering the electron density) and the presence of free nitrogen lone pairs able to coordinate with a molecule of water or alternatively to sequester a monovalent potassium cation in the centre, thus locking the flat surface of the molecule and facilitating the interaction with G-quartets (Figure 12).

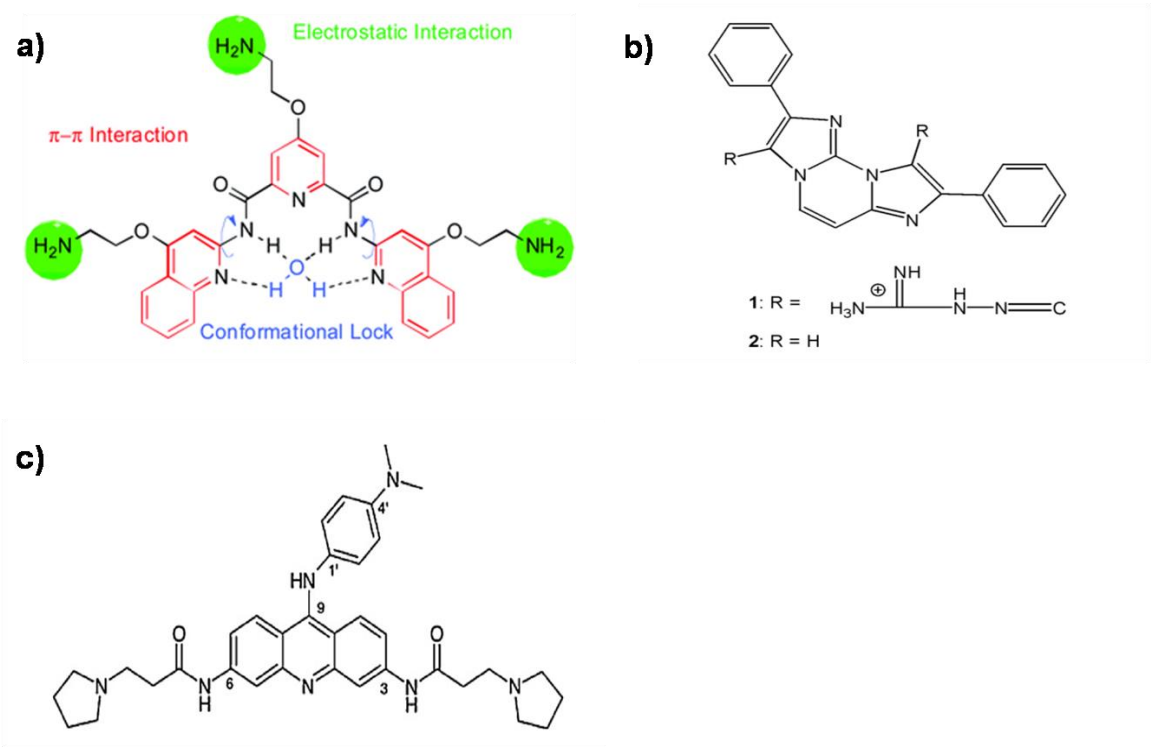


Figure 12. Chemical structures of G-quadruplex binders. a) Pyridostatin; b) Diimidazole [1,2-a:1,2-c]pyrimidine derivatives; Structure (1), termed as FG, has guanylhydrazone groups in their side chains; Structure (2), termed as FA, which lack charged side-chains, is devoid of quadruplex- or duplex- binding activity. c) BRACO-19 ((9-[4-(*N*,*N*-dimethylamino)phenylamino]-3,6-bis(3-pyrrolodino-propionamido) acridine) [Müller *et. al.*, 2012; Sparapani, 2005; Burger *et. al.*, 2005].

Other examples of G-quadruplex binders of alkylamidoacridines-based telomere-targeting agents, Braco-19 (Figure 12c) and diimidazole [1,2-a:1,2-c]pyrimidine derivatives (Figure 12b). Braco-19 interacts with G4 structures via π-π stacking. Positively charged side chains of the inhibitor interact with the negatively charged phosphate DNA backbone and thereby stabilize the G-quadruplex–inhibitor complex. It was reported that a treatment with BRACO19 not only resulted in telomerase inhibition but also in general telomere dysfunction that led to atypical mitosis and consequently to apoptosis (Burger, 2005). The diimidazole [1,2-a:1,2-c]pyrimidine derivative, bis-guanylhydrazone derivative of diimidazo[1,2- a:1,2- c]-pyrimidine is a potent in vitro inhibitor of telomerase. Molecular modeling suggests that the guanylhydrazone groups play an active role in quadruplex

binding [Sparapani *et. al.*, 2005; Andreani *et. al.*, 2004]. Biophysical assays by FRET melting assays showed outstandingly high ΔTm with the human telomeric quadruplex, quadruplexes c-kit1 and c-kit2 from the promoter region of the c-kit gene. Moreover, the compound was previously examined for their anti-proliferative activity in the NCI 60 cancer cell line panel with significant activity in ovarian line IGROV1 [Sparapani *et. al.*, 2005].

1.6 R-loops

R-loops are three stranded structure which form when RNA hybridizes to a complementary DNA strand of a DNA duplex, leaving the opposite DNA strand single-stranded. The R-loop structure has been first characterized *in vitro* over 35 years ago. As observed under the electron microscope, these thermodynamically stable R-loop structures appeared similar to the D-loop structures reported by Robberson et al. during mtDNA replication [Thomas *et. al.*, 1976].

The role of R-loops was firstly established in bacteria by Tomizawa et al in the 1980s. During the replication of ColE1 family of plasmids, where DNA synthesis was found to be initiated by an R-loop with a cleaved 3'-OH end [Itoh *et. al.*, 1980; Selzer *et. al.*, 1982]. Kogoma *et. al.* (1997) reported that the RNase H1in *E.coli* hydrolyses R-loops. The enzyme that targets RNA exclusively in RNA-DNA hybrids belong to the RNase H family. There are two types of RNase H, H1 and H2, characterized by their biochemical properties and substrate preference, are potentially capable of removing RNA-DNA hybrids. Rnase H1 being the most conserved type is present in retroviruses, bacteria and humans. Its specificity in target recognition is due to a ~50 aa N-terminal RNA/DNA hybrid-binding domain that is connected to the C-terminal catalytic domain (~150 aa) by a flexible linker, within the residues D145, E186, D210, and D274 form the active site. Any mutation of at least one of these residues inhibits enzymatic activity [Nowotny *et. al.*, 2007].

As a consequence of transcription process, R-loops formation resulted in hybridization between nascent RNA transcript and DNA template, called ‘co-transcriptional R-loop’ formation [Drolet *et. al.*, 1995]. These R-loops are reported to mediate the establishment of

replication forks for chromosomal DNA replication in a phenomenon described as constitutive stable DNA replication (cSDR) in *E.coli* cells deficient either for RNase HI or RecG but not both as it causes lethality.

The R-loops formation is well-characterized in the mammalian mitochondrial origin of replication and immunoglobulin (Ig) class switch regions of activated B lymphocytes [Yu *et. al.*, 2003; Lee *et. al.*, 1996]. Their formation depends on a number of features, such as G•C content, DNA supercoiling, and DNA cleavage. In bacteria, the Ig class-switch S regions consist of G clusters on the non-transcribed strand and are followed by regions with high G density. G-clustering or GC skew, a similar term coined for G-clustering on the non-transcribed strand are important for R-loop initiation and G-density involves in the stabilization and elongation of the RNA:DNA hybrid [Ginno *et. al.*,2012; Vertino *et. al.*, 2012; Roy *et. al.*, 2009].

The key element for R-loop formation *in vivo* being the negative superhelicity that negative supercoiling increases the length of the RNA:DNA hybrid and reduces the G dependency [Roy *et. al.*, 2009]. Both features facilitate the opening of the bubble of DNA duplex [Aguilera *et. al.*, 2012] and this can be seen in Top1-deficient mutants of *E. coli* that inability of Top1 to relax the negative superhelicity of DNA warrants lethality.

A R-loop model consisted of three distinct parts: R-loop initiation zone (RIZ), linker and R-loop elongation zone (REZ) [Roy *et. al.*, 2009]. G-clusters in RIZ are extremely important for the initiation of R-loop formation while the linker between RIZ and REZ consisted of any nucleotide composition. The final part of R-loop, REZ sequence, is required to be of high G density but does not necessarily have to be a G-cluster. This model can be applied for in vivo R-loop detection and facilitate the search of potential R-loop forming sequences (RLFS) in the genome [Roy *et. al.*, 2008]. Wongsurawat *et. al.* using predictive algorithm proposed a model based on Roy and Michael Lieber's model. RLFS can be partitioned into three segments: RIZ; linker and REZ or

RLFS=RIZ+linker+REZ

RIZ. The DNA regions of initiation of R-loops are considered as clusters of a few Gs (3–4 nt) in the region. Segment sequence initiates and terminates with G-cluster that contains at least three contiguous.

Linker. The DNA sequence region between RIZ and REZ regions is called linker. The nucleotides in this region are not specified and also the length is variable from 0 to 50 nt.

REZ. Downstream of RIZ and Linker, REZ can support the extension of R-loop with a high G density. REZ has to be G-rich but does not require G-cluster like RIZ. At least 40% of G is required for R-loop formation. In Wongsurawat's model, nucleotide number of REZ can vary from 100 to 2000 nt.

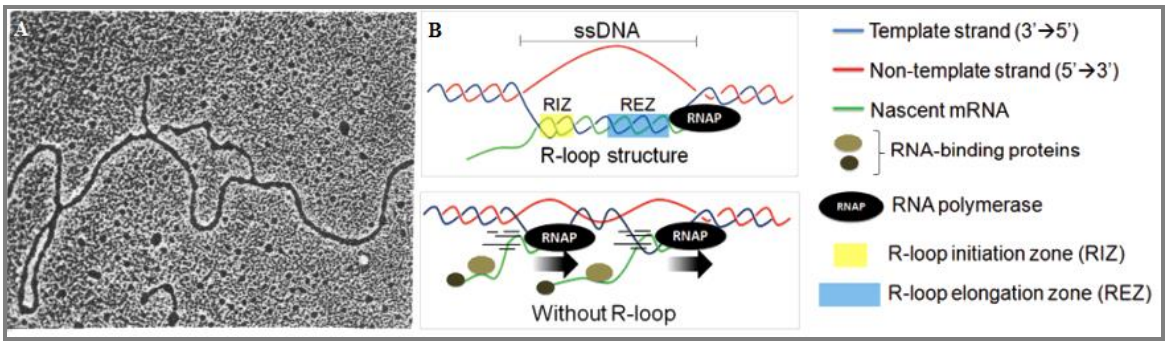


Figure 13. a) Identification of *in vitro* R-loop using electron microscopy (Thomas et al 1976). b) Transcription with and without R-loop forming structure. R-loop initiation zone (RIZ) and R-loop elongation zone (REZ) are highlighted in yellow blue, respectively [Wongsurawat *et. al.*, 2011].

Two mechanisms are proposed for the formation of R-loop at a transcribed sequence [Roy *et. al.*, , 2010; 2009; 2008). In the 'thread-back' model, the nascent transcript is ejected from the RNA polymerase at the site of transcription but threads back to bind to the DNA template strand to form the hybrid, as in the case of linear templates of the murine immunoglobin Sϒ3 class switch recombination region [Roy *et. al.*, 2008]. In the 'extended hybrid' model, the nascent transcript fails to denature from the template in the transcription bubble, due to the high thermodynamic stability between RNA–DNA hybrids. This mechanism is involved during a transcriptionally induced R-loop in immunoglobin sequence

on supercoiled templates but also at expanded trinucleotide repeat tracts [Reddy *et. al.*, 2011; Duquette, 2004].

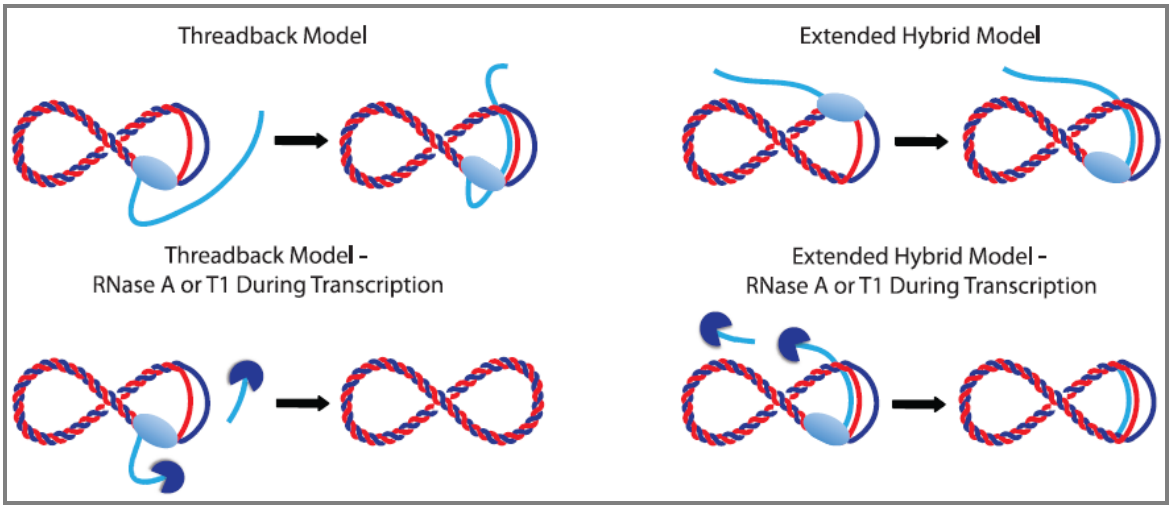


Figure 14. Schematic representation of the two possible mechanisms for R-loop formation. The "thread-back" model on the left and the "extended-hybrid model" on the right. The nascent transcript is depicted in light blue, free DNA template strand in red, RNAse A in dark blue and the RNA polymerase as an light blue oval [Reddy *et. al.*, 2011].

R-loops sequences in length from 150 to 650 bp in Ig switch region, from 110 to 1280 bp in Bcl6 and from 120 to 770 bp in RhoH [Wongsurawat *et. al.*, 2011]. Their stability, depending on the oligomeric length, the content of deoxypyrimidines/deoxypurines, and the A•T/U proportion, is typically higher than DNA:DNA associations and the relative stability of these hybrids [Shaw *et. al.*, 2008]. Hence, their formation can be a costly energy-consuming process. NMR and X-ray diffraction studies indicate that RNA:DNA hybrids adopt a conformation that is intermediate between those of dsDNA (B form) and dsRNA (A form) [Shaw *et. al.*, 2008]. This special structure might be important as a recognition element, since hybrids have to be distinguished in vivo from normal dsDNA for removal. [Aguilera *et. al.*, 2012].

1.6.1 Hypernegative Supercoiling and Cotranscriptional R-loops

One consequence of TopI depletion in cells is the accumulation of hypernegatively supercoiled DNA behind the transcribing RNAP. Negative supercoiling behind the transcription bubble can lead to opening of the DNA. When this happens, the nascent RNA may hybridize to the transcribed strand, creating RNA R-loops. Negative supercoiling is linked to the formation of R-loops in *E. coli*. RNase H abolishes transcription-dependent supercoil accumulation in vitro [Drolet *et. al.*, 1994]. In addition, R-loops have a negative effect on cell metabolism and growth as suggested by the lethality of topA rnhA mutants [Drolet *et. al.*, 1995]. Finally, *E. coli* topA-null mutants fail to accumulate full mRNAs and rRNAs, a phenotype suppressed by RNase H overexpression and consistent with a negative effect of R-loops in transcription efficiency. A connection between hypernegative supercoiled DNA and R-loops in vivo also exists in *S. cerevisiae*. RNA:DNA hybrids form in wild-type yeast cells at the rDNA region and are significantly increased in top1Δ top2Δ strains and further enhanced in the absence of RNase H1. Further analyses have shown that truncated fragments of pre-rRNA accumulate in top1Δ mutants [El Hage *et. al.*, 2010].

A link between R-loops and supercoiling also seems to occur in mRNA transcription as shown in human cells, in which Top1-depletion causes replication impairment at regions rich in protein-encoding genes. This impairment is suppressed by RNase. The observation that a nick in the DNA potentiates RNA:DNA hybrid formation in vitro strongly suggests that a free end either in the DNA or RNA strand would facilitate the intertwining of RNA with DNA. It may be possible thereby that cleavage of the nascent RNA molecule facilitates R-loop formation in negatively supercoiled DNA upstream of the elongating RNAP [Roy *et. al.*, 2010].

All these notions make reasonable that the R-loop formation in vivo is a dynamic process involving protein–DNA–RNA interactions. Top1 may prevent an accumulation of negative supercoiling downstream of transcription block and can prevent R-loop formation [Pommier *et. al.*, 2006].

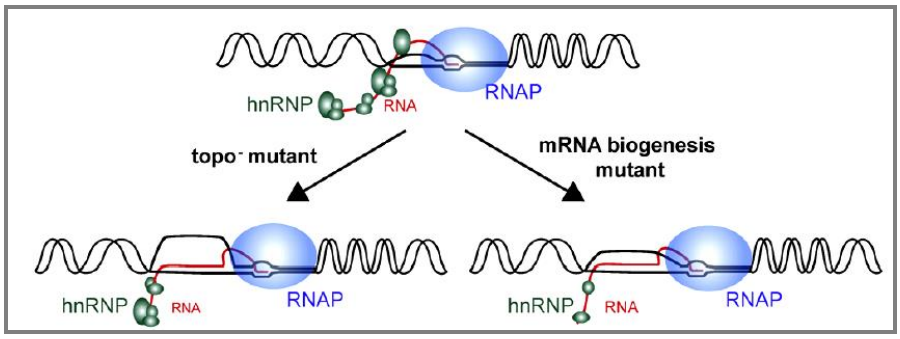


Figure 15. During transcription the nascent pre-mRNA is fully packed by the hnRNPs. Hypernegative supercoiling and mRNP biogenesis defect contribute to aberrant formation of R loops formed by the nascent RNA. [Aguilera *et. al.*, 2012].

1.6.2 R-loops formation as a natural event

1.6.2.1 Replication

R-loops are obligatory intermediates in specific cellular processes. They are most studied in the replication of bacteriophage T4, *E. coli* ColE1 plasmid and mitochondrial DNA.

The bacteriophage T4 initiates DNA replication by two mechanisms, one of which, the recombination-dependent replication (RDR) forms R-loop. Immediately after bacterial infection, replication is initiated at several origins including oriG, which harbors a promoter. Transcripts initiated at such promoter form persistent R-loops that are likely processed by T4 RNase H to generate the free 3'-end required for lagging-strand DNA synthesis. In bubble-migration synthesis model, lagging strand synthesis does not occur, and the newly synthesized single strand is extruded from the back of the D-loop as new DNA is synthesized at the front of the D-loop [Kreuzer *et. al.*, 2010].

In *E. coli*, initiation of replication ColE1 is RNAP-dependent RNA synthesis. ColE1 initiation of replication relies on the formation of an RNAP-driven 550 bp sequence, termed RNAII, which forms a stable RNA:DNA hybrid with the leading-strand DNA template. This RNAII is processed by RNase H1 to generate a 3' end and is extended by DNA Pol I [Itoh

et. al., 1980]. Interestingly, *E. coli* RNase H1 mutants (rnhA) replicate the chromosome in the absence of dnaA and other canonical replication initiation factors normally required at oriC [Kogoma *et. al.*, 1997]. It is believed that persistent R-loops in rnhA mutants can initiate replication independent of oriC, mimicking initiation of ColE1 [Lee *et. al.*, 1996].

Mitochondrial DNA (mtDNA) replication is similar to ColE1 replication mechanism, with DNA synthesis being primed by an RNA molecule produced by the mitochondrial RNAP [Xu *et. al.*, 1996]. mtDNA is a double-stranded circular molecule that encodes essential subunits of the mitochondrial respiratory chain, as well as the tRNAs and rRNAs required for their synthesis. mtDNA consists of two major promoters: the light strand promoter (LSP) and the heavy strand promoter (HSP). Transcription by RNA polymerase (POLRMT), starts at both LSP and HSP and is polycistronic, producing near genome-length transcripts that are later processed to give rise to the individual mRNA molecules [Falkenberg *et. al.*, 2007]. In mitochondria, transcription and DNA replication are closely linked, since POLRMT is responsible for the synthesis of a RNA primer required for initiation of DNA synthesis by the mtDNA polymeraseϒ (POLϒ) from the mitochondrial origins of replication, OriH and OriL [Fuste *et. al.*, 2010]. In addition to POLϒ, the mitochondrial replication machinery also consists of a replicative helicase, TWINKLE, and a single-stranded DNA binding protein, mtSSB. In combination, these three factors form replication machinery that can synthetize ssDNA molecules longer than 16.5 kb, the size of the human mtDNA genome [Falkenberg *et. al.*, 2007]. In mammals the RNA molecule is transcribed from the light-strand promoter (LSP), using this strand as a template, and terminates at OriH. Once processed by RNase H the RNA molecule is used as primer by DNA Polϒ and the light chain starts being replicated. The heavy chain is displaced, forming a loop (called "D-loop") and only in a second moment undergoes itself replication. Sequence comparison in vertebrates reveals three conserved sequence blocks (termed CSB I, CSB II, and CSB III), downstream of LSP. The RNA-DNA hybrid, is stable and persistent and its formation is dependent on the CSBII element, in yeast and in human [Wanrooij *et. al.*, 2012]. On the contrary, human mtRNA sequence upstream of the hybrid region is also important for efficient RNA-DNA hybrid formation, especially the conserved CSBIII element which is absent from the putative yeast mitochondrial origins. However, the exact role of CSBIII, in the form of DNA and RNA, is

not known. Its complete removal permits RNA-DNA hybrid formation in a manner similar to the natural situation in yeast [Xu *et. al.*, 1996]. The extraordinary stability of the mitochondrial R-loop was explained by the discovery of G-quadruplex DNA structure in the R-loop forming sequence, by preventing reannealing of the template and non-template DNA strands [Wanrooij *et. al.*, 2012; Lee *et. al.*, 1996].

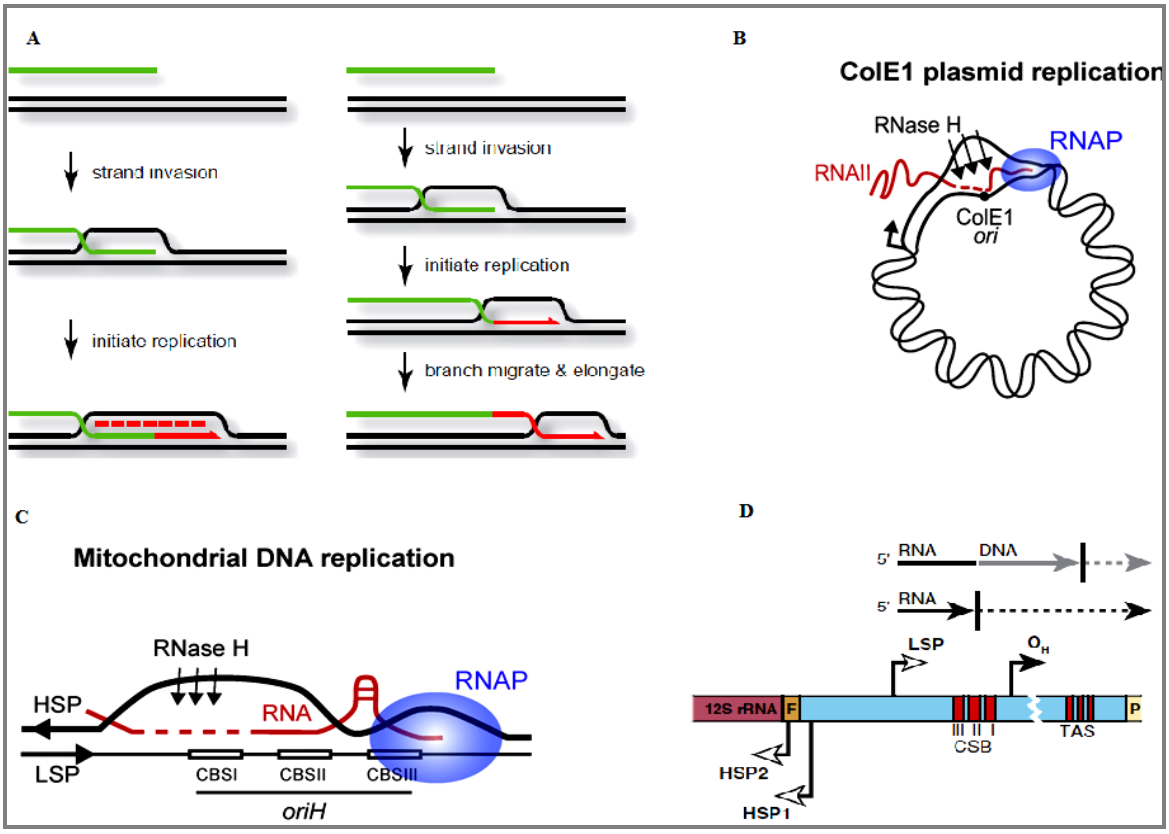


Figure 16. R-loops formation as natural event with a role in Replication. A) Recombination-dependent replication (RDR) of Bacteriophage T4. Left: semi-conservative RDR model, right: bubble-migration synthesis model. New leading strand replication is in solid red and new lagging strand replication is in dashed red (Kreuzer *et. al.*,2010). B) R loops in ColE1-type plasmids replication C) R loops in mammalian mitochondrial light DNA strand replication (Aguilera *et. al.*, 2012). D) Schematic representation of the D-loop regulatory region. The three conserved sequence blocks (CSB I, CSB II, and CSB III) are located just downstream of light-strand promoter (LSP). The conserved termination-associated sequence (TAS) elements are also represented Falkenberg *et. al.* (2007).

1.6.2.2 Recombination

The Ig class switching in vertebrate B cell plays an essential process for the generation of specific and high affinity immunoglobulins.

Various isotypes of immunoglobulins have different effector functions for optimal immune responses to pathogens. Prior to pathogen exposure, a highly diverse repertoire of low-affinity IgM antibodies are generated through V(D)J recombination. When B cells encounter antigens, more effective isotypes of immunoglobulins are produced through two additional DNA modifying mechanisms: somatic hypermutation (SHM) and class switch recombination (CSR) [Yu *et. al.*, 2003; Dunnick *et. al.*, 1993]. CSR is responsible for changing the heavy chain isotype from IgM to IgA, IgE or IgG.

CSR occurs only at the repetitive switch (S) regions, located downstream of a promoter. These sequences are GC-rich, located particularly on the non-template strand DNA. In addition, they consist of 25–80 bp repeat units stretching a total of 1–10 kb in length [Dunnick *et. al.*, 1993]. All switch regions have promoters that respond to B cell activation and cytokines. Transcription is required for CSR. Activation-induced cytidine deaminase (AID) is the key enzyme in both CSR and SHM that deaminates cytidines of the single stranded DNA. The action of AID at the displaced G-rich ssDNA is the first step for the generation of the double strand break responsible for CSR. The DNA in switch regions becomes single-stranded during SHM and CSR is by transcription. It is however, not all AID is seen to be recruited deaminates cytidines in ssDNA [Roy *et. al.*, 2008]. Therefore, both CSR and SHM possess special mechanisms to recruit AID to switch regions and the VDJ region, respectively. R-loops are known to form at switch regions but their role was unclear until it was shown that these structures could accentuate AID targeting in murine primary B cells. The R-loops of few hundred base pairs provide sufficient numbers of single-stranded cytidines for AID to act on with greater efficiency than would otherwise occur [Roy *et. al.*, 2008].

An example of R-loops, TERRA, form naturally by the telomeric transcripts. TERRA involves RNA:DNA hybrids is implicated in telomerase activity inhibition [Luke *et. al.*, 2008]. There are also noncoding RNAs (ncRNAs) reported to form R-loops, of unknown

physiologically meaning, but has a role in mediating RNAi-directed heterochromatin formation in fission yeast [Aguillera *et. al.*, 2012].

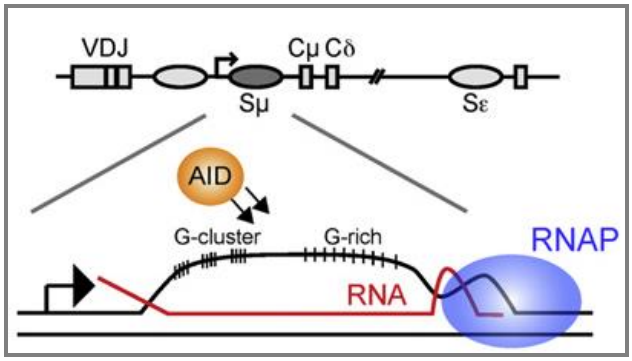


Figure 17. R loops in Ig class-switch recombination. Gs are indicated as thin vertical lines (Aguillera *et. al.*, 2002).

1.6.2.3 Gene expression

R-loop structures affect gene expression, whether by blocking gene elongation or by repressing gene expression itself. In *E.coli*, R-loop formation in the Top1 mutants leads to growth defects, impaired transcription elongation on the rDNA and extensive RNA degradation by RNase H [Drolet *et. al.*, 2005].

Hage and coworkers reported that pre-rRNA transcription is affected by the absence of Top1 in yeast and that these pre-rRNA fragments not only accumulated in the absence of Top1 but also hybridized to the template strand, forming an R-loop. rRNA synthesis was also reduced in these strains, supporting the conclusion from in vitro data that stable R-loops block transcription elongation [Aguilera *et. al.*, 2012]. These truncated pre-rRNA fragments were stabilized in top1Δ strains defective in TRAMP (trf4Δ) or the exosome (rrp6Δ), indicating that these complexes degrade the pre-rRNA fragments released by RNase H cleavage. Depletion of Top1 increases both the frequency of pileups and the numbers of contiguous polymerases, presumably reflecting increased duration of stalling of the leading polymerase. In this strand Top2 is supposed to resolve positive and negative supercoiling. Both activities should lead to the release of transcriptional blocks, but Top2 is not predicted to resolve

strand separation induced by negative torsion. In strains lacking both Top1 and RNase H activity, pileup frequency increased further. The presence of persistent R-loops might slow down local rotation of the rDNA, reducing the speed of elongation of Pol I in the same pileup, which would increase the residency times of pileups and impede their resolution. [Hage *et. al.*, 2010].

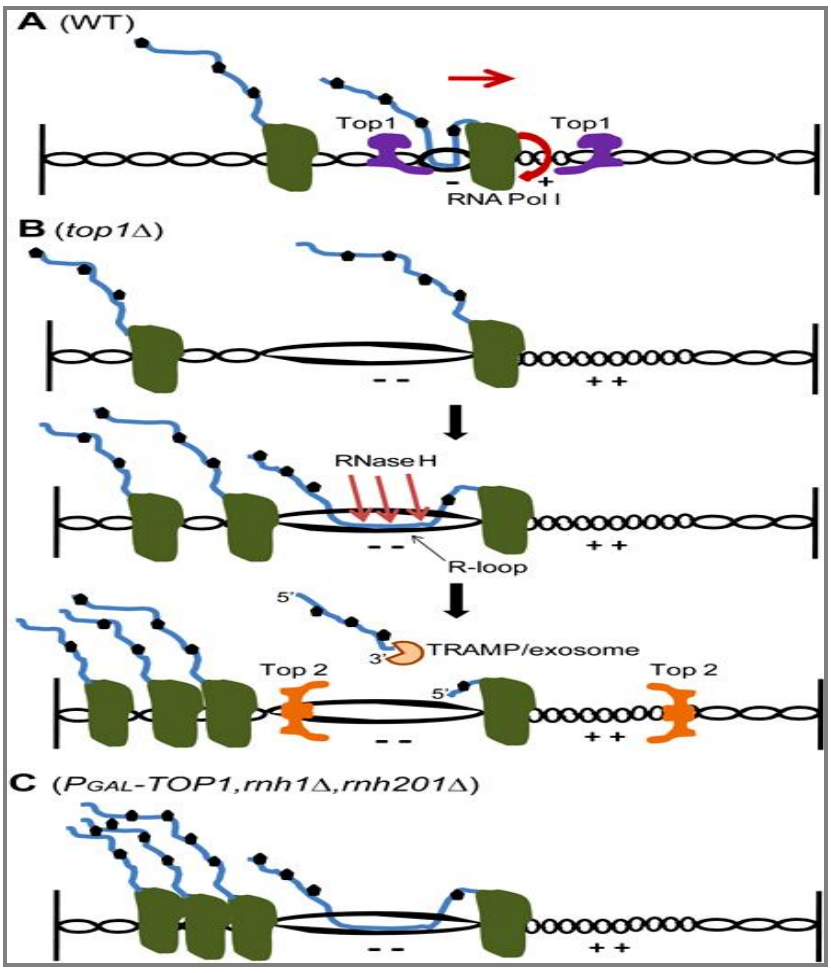


Figure 18. R-loops' role in blocking pre-rRNA transcription. a) Polymerase movement during transcription forces the rDNA to rotate, building up positive and negative torsion, resolved by Top1. b) In strains lacking Top1, more torsion is accumulated and R-loops occur more frequently, leading to an increase in pileup formation. RNase H1 and H2 cleave the RNA–DNA hybrids, releasing truncated pre-rRNA fragments that are targeted and degraded by the TRAMP and exosome complexes. c) In the absence of both Top1 and RNase H1 and H2, persistent R-loops block rotation of the rDNA and cause severe polymerase arrests and pileups [Hage *et. al.*, 2010].

R-loop formation in bacteria is reported with termination failures. The absence of functional Rho-dependent termination has been inferred to lead to R-loop accumulation, as these mutants require RNase H activity for survival.

Different studies suggest that transcription termination in RNAPII-driven genes may represent an additional and relevant case in which R-loops could form naturally in eukaryotic cells. Transcription pause sites located downstream of the poly(A) signal facilitates termination. Specific nucleases and helicases participate in the degradation of the downstream-cleaved RNA as key factors for template release, including yeast Rat1 exonuclease. In yeast the Sen1 RNA:DNA helicase cooperates with Rat1 to promote efficient transcription termination. Loss of Sen1 results in RNA:DNA hybrid accumulation, suggesting that R-loops may form as natural transcription intermediates that are removed by Sen1 [Mischo *et. al.*, 2011].

Senataxin (SETX), the human ortholog of Sen1, is required to avoid transcript read through. This is due to the incapacity of SETX-depleted cells to properly terminate transcription and correlates with RNA:DNA hybrid accumulation downstream of the poly(A) signal. It has been proposed that R-loops may be critical for RNAPII to pause downstream of the poly(A) site and that Senataxin may unwind the RNA:DNA hybrids. Therefore, termination could use short RNA:DNA hybrids to potentiate pause sites and/or as a step required for exonuclease degradation of the nascent RNA necessary for template release [Skourti-Stathaki *et. al.*, 2011]. G-rich sequences immediately downstream of the poly(A) signal are common in mammalian genes and potential G4-forming sequences are enriched at the 3'-UTR regions of genes. Given the potential of G-rich sequences to stabilize R-loops, it might be possible that R-loops are intrinsic elements of termination pause sites. [Belotserkovskii *et. al.*, 2010].

Interestingly, approximately 45% of all human genes exhibited GC skew at their core promoter region, the vast majority of which (95%) co-occur with promoter CpG islands [Marinello *et. al.*, 2013; Ginno *et. al.*, 2012; Vertino *et. al.*, 2012]. R-loops accumulate naturally at these regions, immediately downstream of the CpG-non-methylated promoters. It has been proposed that the displaced ssDNA in the R-loop acts as a signal to recruit either

the protective H3K4 trimethyl mark or the DNA demethylases complex [Ginno *et. al.*, 2012; Vertino *et. al.*, 2012]. Moreover, ssDNA in the R-loop structure is also able to recruit the AID enzyme, which is capable of editing nucleic acid through deamination of cytosines to uracils. They are most often act in immunoglobulins production and epigenetic reprogramming in mammals. They are in active state in primordial germ cells (PGCs) and in early embryos where demethylation occurs [Bhutani *et. al.*, 2010; Popp *et. al.*, 2010]. Hence, indicating that R-loop structure may be a potential target of AID-mediated epigenetic reprogramming [Wongsurawat *et. al.*, 2012].

R-loops regulate gene expression in an epigenetic manner. Topotecan-induced R-loop at Snord116 locus inhibits transcriptional elongation through Ube3a-antisense and promotes allele-specific chromatin decondensation [Powell *et. al.*, 2013]. Snord116 is a neuron-specific noncoding RNA that maps on the Prader–Willi syndrome (PWS)/Angelman syndrome (AS) locus, which is regulated by an imprinting control region that is maternally methylated and silenced. Epigenetic dysregulation of this locus leads to PWS or AS. The PWS imprinting control region is the promoter for a one megabase paternal transcript encoding, besides Snord116, the ubiquitous protein-coding Snrpn gene, other neuron-specific noncoding RNAs, and the antisense transcript to Ascausing ubiquitin ligase encoding Ube3a (Ube3a-ATS). All these results provide a deeper understanding of the relationship between epigenetic regulation, chromatin dynamics, transcription and R-loop formation [Powell *et. al.*, 2013].

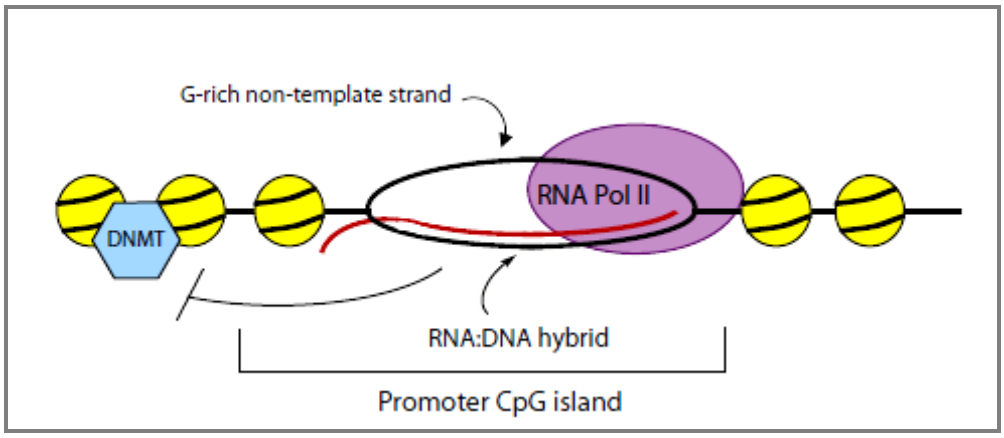


Figure 19. Persistent RNA:DNA hybrid influences DNMT activity [Vertino *et. al.*, 2012]

1.6.2.4 Source of genome instability

The computational analysis of human genomes has identified possible DNA hotspots for the formation of R-loops which are widespread throughout the human genome particularly located in genes involved in several diseases. In fact, it is well known that co-transcriptional R-loops are linked to different forms of genome instability including mutations, recombination, and chromosome rearrangements as well as chromosome loss. Evidence for R-loops as a source of genome instability is provided by the yeast THO and THSC/TREX-2 R-loop-forming mutants which show a transcription-associated hyper-recombination phenotype and elevated chromosome and plasmid loss [Wongsurawat *et. al.*, 2012]. R-loops can form in fragile sites *in vivo*. Common fragile sites (CFSs) are caused by collisions between the replication and transcription machineries and such CFSs are linked to the formation of RNA:DNA hybrids [Reddy *et. al.*, 2011; Lin *et. al.*, 2010].

In *S. cerevisiae* mutations in genes involved in transcription and mRNA processing/export can cause a significant increase in YAC minichromosome loss. This chromosome instability (CIN) is linked to mRNA biogenesis factors, in particular mRNA cleavage and polyadenylation factors. Cells depleted of such factors form RNA:DNA hybrids, and the CIN phenotype is suppressed by RNase H [Stirling *et. al.*, 2012].

The single-stranded DNA (ssDNA), a by-product from R-loops formation is prone for the mutagenic action of specific DNA-modifying enzymes such as AID. The ssDNA displaced by the RNA:DNA hybrid may be critical for such mutagenicity, since ssDNA is more susceptible to mutagenic DNA damage than dsDNA [Aguilera *et. al.*, 2012].

R-loops are also involved in the generation of hyper-recombination: translocations occurring between S regions and c-myc that are responsible for Burkytt's lymphoma are AID-dependent and occur at G-cluster-rich regions prone to form R-loops. Proto-oncogenes involved in translocations such as BCL6, RhoH, PIM1, or PAX5 also occur in primary B lymphocytes at sites of potential G-quadruplex-rich R-loops [Duquette *et. al.*, 2007]. Transcription- induced R loops can lead to genomic instability also by the creation of an impediment to replication fork progression, which constitutes a general and evolutionarily conserved mechanism underlying R-loop-induced genomic instability [Gan *et. al.*, 2011].

Recent evidence has linked R-loop formation at several trinucleotide repeat sequences, whose genetic instability, expansions, are the cause of numerous diseases. The genetic instability of gene-specific trinucleotide repeat sequences is the causative mutation for various neurological, neuromuscular as well as many neurodegenerative. Among these diseases there are spinocerebellar ataxia type 1 (SCA1), myotonic dystrophy (DM1) and fragile X type A (FRAXA). R-loop forming structures can be found in the Fmr1 and Fxn genes that are responsible for neurodegenerative disease. It was demonstrated that R-loops could co-localize with some classes of trinucleotide repeat tracks that occur in these genes [Groh *et. al.*, 2014]. R-loop structures are found when Fmr1 and Fxn genes are transcribed. The RNA–DNA hybridization via R-loop mechanism can generate genetic instability that may be associated with the expansion of the trinucleotide repeats within the disease related genes [Pereira *et. al.*, 2014; Groh *et. al.*, 2014].

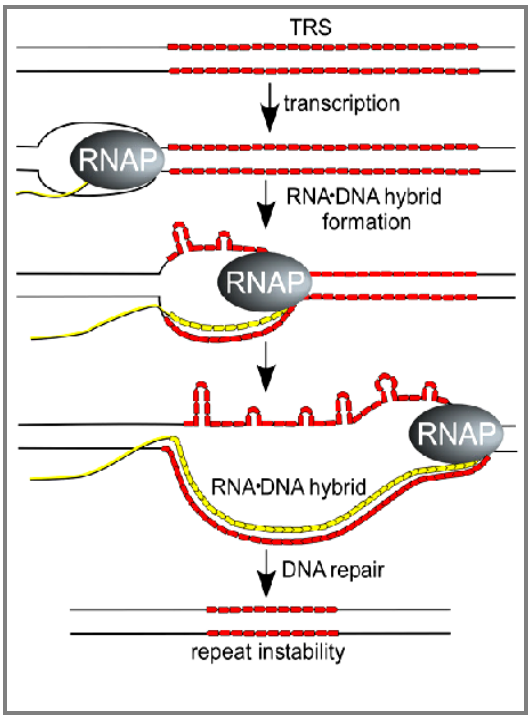


Figure 20. Proposed mechanism for stable RNA: DNA hybrids stimulating repeat instability [Lin *et. al.*, 2010].

A genome-wide siRNA screen in human cells to identify genes involved in genome stabilization through H2AX phosphorylation (γ-H2AX) (a cellular mark for double-strand breaks), identified a number of factors involved. Among these different proteins involved in mRNA processing (such as 3' end polyadenylation factors), the nuclear exosome involved in RNA surveillance, heterogeneous nuclear ribonucleoproteins (hnRNPs), splicing factors and nuclear pore complex (NPC) components. Overexpression of RNase H in depleted in any of these, reduces the γ-H2AX foci [Aguilera *et. al.*, 2012].

Acting as mutagenic intermediates, R-loops can cause either gene-specific or genome-wide instability. For example, mutations in the THO/TREX complex, which is required for proper coupling of transcription and mRNA export, cause wide-scale co-transcriptional R-loop formation triggering aberrant recombination leading to genome-wide instability [Gonzalez *et. al.,* 2011].

Diseases caused by gene-specific expansions of (CAG)·(CTG) repeats including Huntington's disease, myotonic dystrophy type 1 and a series of spinocerebellar ataxias are implicated with R-loop formation (Pereira, 2014). Others include fragile X mental retardation involving (CGG).(CCG) repeats and Friedreich's ataxia involving a (GAA)·(TTC) repeat [Groh *et. al.,* 2014].

Chapter 2

Materials and methods

2.1 Cell lines

HCT15 and COLO205 were grown in monolayer cultures in RPMI 1640 while HCT116 was grown in DMEM. All growth mediums were supplemented with 10% heat-inactivated foetal bovine serum and 2 mM glutamine. The cells were maintained by trypsinization and passed 2-3 times a week and incubated at 37 °C, 5 % CO_2.

2.2 Preparation of drug compounds

Camptothecin was initially dissolved in 100% DMSO to make up 10 mM stock solutions while Pyridostatin and Braco-19 in 5 mM stock solutions. The stock solutions were aliquoted in small vials of 30 µl to avoid repeating freeze-thaw. Stock solutions were diluted immediately prior to use.

2.3 Western blot

Cell lines HCT116, HCT15 and COLO205 were grown to 90 % of confluency. The medium was discarded and washed with PBS at room temperature. The cells were lysed with SDS buffer (Tris-HCl pH 6.8 50 mM; Sucrose 15 %; EDTA 12 mM; SDS 3 %; β-mercaptoethanol 10 %, bromophenol blue) and boiled for 20 mins. The protein lysate was put into ice and loaded an equal volume into the polyacrylamide gel.

Bis-acrylamide gel

Resolving (25ml)

- Bis acrylamide (19:1) 40 % 8 %
- Tris-HCl pH 8.8 405 mM
- SDS 10 % 0.1 %
- Temed 0.21 %
- APS 10 % 0.1 %

- H_2O to volume

Stacking (10ml)

- Bis acrylammide (19:1) 40 % 4 %
- Tris-HCl pH 6.8 125 mM
- SDS 10 % 0.1 %
- Temed 0.1 %
- APS 10 % 0.2 %
- H_2O to volume

The gel was electrophoresed at 20 mA constant for 2 hrs. Upon finishing, the protein was transferred to gel at 100V constant for 1 hour and 30 mins. The transfer buffer consisted of 25 mM Tris, 192 mM Glycine and 20 % methanol (v/v). After transfer, the membrane was colored for 5 minutes with Ponceau red. The membrane was washed with PBS for 2 times and once with 1 x TBS/Tween 0.1 % to de-colour red. The membrane was then blocked 5 % milk /1 x TBS/Tween 0.1% for 1 hr. The membrane was incubated with goat polyclonal IgG anti-Top1 (C-15; *Santa Cruz Biotechnology*) in dilution 1:200 for 2 hours at room temperature and anti β-actin (*Santa Cruz Biotechnology)* dilution 1:200. Then, membranes were washed in TBS 1 x- Tween 0.1% for 10 min. Secondary antibody, rabbit anti goat IgG conjugated with horse radish peroxidase (*Santa Cruz Biotechnology*) diluted in 1:2000 using 5 % milk /1 x TBS/Tween 0.1 % was added to the membranes and incubated for 1 hour RT. Membranes were wash with 1 x TBS/Tween 0.1 % for 10 mins. Fluorescence signal was detected using Pierce ECL plus western blotting substrate (Thermo Scientific).

2.4 Drug concentrations

2.4.1 For cell proliferative assay

Drug concentrations used were: Camptothecin (0.1- 10 µM) for all cell lines, Pyridostatin (10 µM for U2OS, 0.25 µM for HCT116, 10 µM for HCT15, 1.0 µM for COLO205) and Braco-19 (1.5 µM for HCT116).

2.4.2 For immunofluorescence

U2OS and HCT15 were treated with 10 µM of CPT and 10 µM Py in single drug treatment as well as co-treatment.

2.5 *In vitro* drug treatments

Two types of assay were carried out to evaluate their antiproliferative effects: 1) Single drug treatment with only CPT for 2 hrs, Py or Braco-19 for 24 hours in different cell lines 2) Co-treatments of CPT for 2 hrs followed by Py or Braco-19 for further 24 hrs. The effect of treatment sequence was investigated. Cells were either first treated with CPT followed by Py or Braco-19 (Figure 21a) or Py treatment followed by addition of CPT (Figure 21b).

A sufficient number of exponentially growing cells were used to avoid confluence of the culture during the treatments. Cells were seeded at 20,000 cells/well in Falcon 48-well plates. All treatments started 24 hrs after seeding. Cells were either treated with single drugs or co-treatments. After treatment ended, the drugs were replaced with fresh medium, and the cultures were maintained for another 72 hrs before the cell viability quantification. Experimental control was always treated with the same amount of DMSO as used in the corresponding experiments.

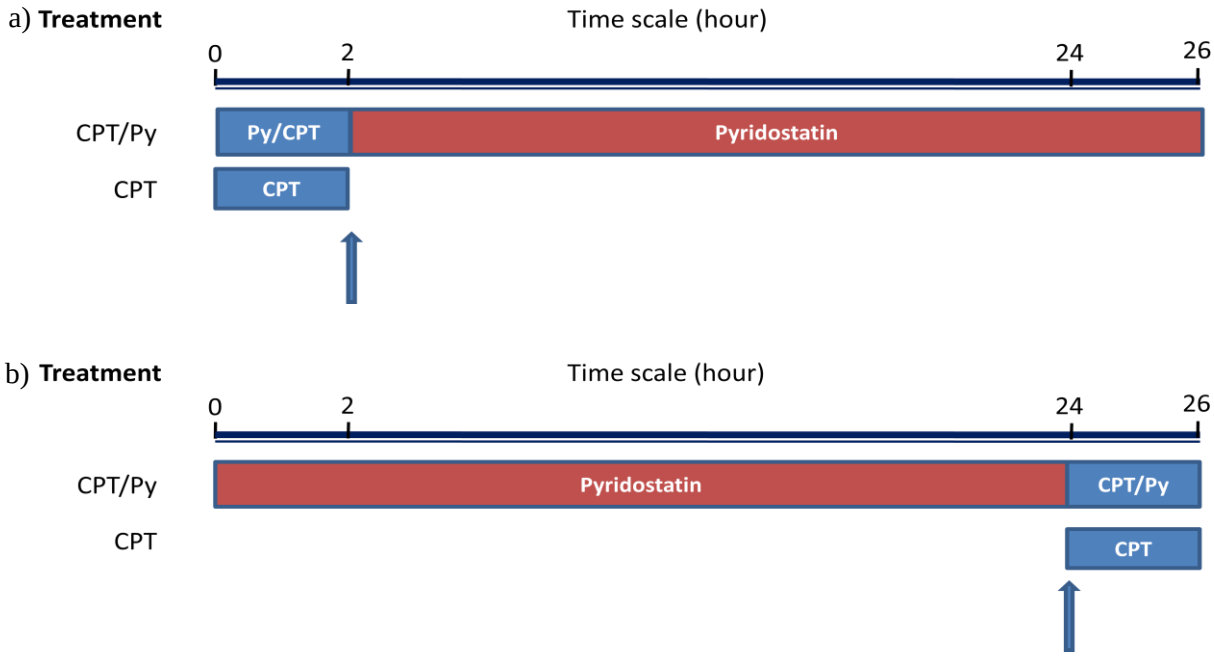


Figure 21. Diagrammatic depiction of co-treatment combinations. a) Py/CPT. Cells were treated with a drug mixture of Pyridostatin (10 µM) and CPT (10 µM) for 2 hrs after 24 hrs of seeding. Drugs were removed (indicated by blue arrow) and cells were further treated with Pyridostatin alone for 24 hrs. b) CPT/Py. Cells were treated with Pyridostatin (10 µM) for 24 hrs and further 2 hrs of treatment with the addition of CPT 10 µM. All drugs were removed after treatments in (a) and (b) and cells were left to grow for further 3 days. The MTT assay was performed to obtain cell survival data.

2.6 MTT assay

The MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide) colorimetric assay is commonly used to determine mitochondrial reductive function and hence it is used as an indicator of cell death or inhibition of growth. MTT was performed on cells which were left to grow for another 72 hrs after drug exposure. This was to provide a more accurate indication of the inhibition of growth caused by drug cytotoxicity rather than other situations, such as the cell quiescence, metabolic stopping, or induction of apoptosis. All assays were performed in duplicate. 200 µl of MTT solution (1 mg/ml in PBS, Sigma) was added to each well and incubated for 2 hrs. Medium was subsequently removed from wells and resulting formazan crystals solubilised in 300 µl of DMSO. Culture plates were rocked

gently for 30 mins to solubilise before measuring the optical density using a microplate reader at 595 nm.

2.6.1 Interpretation of MTT assay

The cell survival for each treatment condition was obtained from the absorbance values. Each absorbance value from treated cells is normalized against a negative control (cells with no drug treatment). Normalized values depicting the survival rate of cells is then used to compute the IC_{50} values.

$$\text{Survival rate (\%)} = \frac{A_{sample} - A_b}{A_c - A_b} \times 100$$

A_b = absorbance of blank

A_c = absorbance of negative control.

2.7 Half-maximal inhibitory concentration, IC_{50}

IC_{50} is the molar concentration of an inhibitor that produces 50 % of the maximal possible inhibitory effect of that agonist. The action of the agonist may be stimulatory or inhibitory. IC_{50} of all cell lines were obtained using GraphPad Prism 5. Data from proliferation assay were analysed using nonlinear regression of log dose-response. The program finds the best-fit values of the parameters from the model to the obtained data from MTT assay.

2.8 Putative G4 motifs

Two set of genes were chosen for this analysis. These genes were obtained from the previous genome-wide study on the effects of CPT. One of the gene sets consisted of list of 225 promoters which were found to display CPT-induced antisense transcription. Another, consisted of 253 genes which showed no antisense transcription. Sequences were downloaded from UCSC Genome browser. For each gene sequence, the transcription start site (TSS) was identified. The regulatory regions flanking different parts upstream and downstream of TSSs were downloaded and analysed in the QGRS for putative G4 structures. The analysed sequences were from 5000bp flanking upstream of TSS (-5000 to -

1), intergenic regions flanking 500bp and 1000bp downstream of TSS (+1 to +500 and +1 to +1000, respectively) and finally 5000bp downstream of 3' end. The presence of putative G4 motifs in genes was investigated using the QGRS mapper [Kikin *et. al.*, 2006]. G4 motifs $G_{3+}\ N_{1-7}\ G_{3+}\ N_{1-7}\ G_{3+}\ N_{1-7}\ G_{3+}$ were searched in designated regions of a gene namely within -5000bp upstream of TSS, +500bp, +1000bp downstream the transcription start-site and 5000bp downstream from transcription termination. The QGRS was set to find putative G4 structures with at least 3 guanines and constraining loop lengths of the G-quadruplex to a maximum of 7 bases.

2.9 BG4 plasmid

BG4 plasmid containing the sequence specific for G4 structure was given as a gift from Professor Balasubramanian, University of Cambridge [Biffi *et. al.*, 2013]. BG4 was then transfected in *E.coli* bacterial expression system to express antibody specific for the detection of G4.

2.9.1 Preparation of competent cells

BL21(DE3) cells, incorporated with T7 promoter expression system, is capable of producing more protein than any other bacterial expression system. The strain is a specially modified BL21 that will express genes from the T7 promoter. It is deficient in 2 types of proteases, *Ion* and *omp*T. The Ion protease is an intracellular protease that *E. coli* makes to degrade extracellular proteins. It degrades the protein after cells are lysed. *E. coli* is a living system and needs to be able to turn over protein and feed on peptides to stay healthy and productive.

Escherichia coli strains, BL2(DE3) was inoculated in SOC medium at 1:100 dilution and was let to grow for 2-3 hrs at 30 ^{o}C. This was to synchronize the cells and to improve its competent efficiency. The cells were grown at 37 ^{o}C until OD_{600} is between 0.37 - 0.40. This OD value corresponded to the log phase growth of the bacteria, when the T7 promoter is repressed to prevent expression of endogenous protein. The bacteria were then pelleted for 10 mins at 4000 rpm and resuspended in Transformation buffer (in 40 ml). Bacteria suspension was left in the cold room for 45-60 mins with continuous agitation. Pellet bacteria for 10 min at 4000 rpm. Cells were resuspended in 10ml of transformation buffer and 7 % DMSO (700 µl). Freeze cells in aliquots of 250 µl at -80 ^{o}C.

2.9.2 Transformation of BG4 plasmid

Competent cells kept in –70 ˚C was placed on ice for 5 mins or until just thawed. A DNA concentration of 1-50 ng (in a volume not greater than 5µl) was added to the competent cells by moving the pipette tip was moved through the cells while dispensing and quickly flicked the tube several times without vortexing. The tubes were immediately returned to ice for 5–30 mins. Cells were heat-shock for 15 seconds in a water bath at exactly 42 ˚C without shaking. The tubes were immediately placed on ice for 2 mins after which 450 µl of room-temperature SOC medium was added to each transformation reaction, and incubated for 60 mins at 37 ˚C with shaking (approximately 225 rpm). The tubes were laid on their sides and taped to the platform for best transformation efficiency. 100µl of undiluted cells and 1:10 and 1:100 cell dilutions were plated on antibiotic plates and incubated overnight at 37˚C.

2.9.3 Preparation of BG4 antibody

The induction of BG4 protein expression in *E.coli* was performed using the autoinduction method as described by Studier (2005), which requires a high bacterial density. For initiation culture, transformed-BL21 cells were inoculated in 2 ml of 2 x TY media + 2 % glucose + 50 mg/ml kanamycin. Cells were grown overnight at 200 rpm at 30 ºC.

The initiation culture was inoculated in 100 ml auto induction media and let to grow at 37 ºC at 250 rpm for 6 hours. The cell culture was pelleted for 30 min at 4 ºC at 4000 *g*. The pellet was resuspended in 8ml TES and left on ice for 10 mins. A further 12 ml TES diluted 1:5 was added into mix and left for 15 mins on ice. Cells debris was discarded by spinning down for 10 mins at 8000 *g* at 4 ºC. At this point, the solution contained a protein mixture of endogenous protein and BG4 antibody. BG4 was purified by using silica-based resin (Protino® Ni-IDA) pre-charged with Ni2+ ions. Since BG4 is tagged with polyhistidine, the protein would be bound by the immobilized Ni2+ on the resins. One gram of resin was weighed and packed into a column. The resin was then equilibrated in PBS, pH 8.0. The protein solution was added to the pre-equilibrated column and was allowed to drain by gravity. To ensure higher BG4 binding, the flowthrough was collected and was re-added to the column. The column was washed twice with cold PBS/100 mM NaCl/10 mM imidazole, pH 8.0 and drained by gravity. Elution was done in a new collecting tube by adding

PBS/250 mM imidazole, pH 8.0. The eluted antibody solution was dialysed overnight in PBS, pH 8.0 in cold room. The BG4 antibody is stored at 4 ^{o}C for few weeks.

2.10 Immunofluorescence

2.10.1 Detection of G-quadruplex

The U2OS and HCT15 cells were seeded at density of 100,000 and 80,000 cells, respectively. After 24 hrs of seeding, cells were subjected to CPT, Pyridostatin or Braco-19 treatment. The cells were then fixed in methanol/acetic acid (3:1) and permeabilized with 0.1 % Triton-X100/PBS. After blocking in 2 % milk/PBS, immunofluorescence was performed using standard methods with 1:30 of BG4, 1:800 anti-FLAG (No. 2368, Cell Signaling Technology) and 1:1000 anti-rabbit Alexa 594-conjugated (A11037, Invitrogen) antibodies. Slides are counterstained with DAPI (5 mg/ml) for 30 mins. After briefly washed with ddH_2O, the slides were mounted by Molwiol (Sigma Aldrich) and left to dry at room temperature before transferring to a 4 ^{o}C fridge for storage. The images were acquired by Nikon Ti-E microscopy equipped with A1R confocal system. Fluorescence quantitation analysis was performed using ImageJ software with the following formula: Corrected Total Cell Fluorescence=Integrated Density - (Area of selected cell X Mean Fluorescence of Background Readings)

2.10.2 Detection of R-loops

The U2OS cells were seeded at density 100,000 cells on glass coverslips. After 24 hrs of seeding, cells were treated with CPT and Pyridostatin. The cells were washed with ice cold PBS and fixed with ice cold methanol for 10 mins. After washing 3 x 5 mins with PBS, cells were blocked with 1 % BSA/PBS for 1 hr and followed by incubation with 1:50 of DNA/RNA hybrid antibody, S9.6 and detected by 1:800 of goat anti-mouse Alexa 594-conjugated (A11005, Invitrogen) antibodies. Slides were counterstained with DAPI (5mg/ml) for 30 mins. After briefly washed with ddH_2O, the slides are mounted by Molwiol (Sigma Aldrich) and left to dry at room temperature before transferring to a 4 ^{o}C fridge for storage. The images were acquired by Nikon Ti-E microscopy equipped with A1R confocal

system. Fluorescence quantitation analysis was performed using ImageJ software with the following formula: Corrected Total Cell Fluorescence=Integrated Density - (Area of selected cell X Mean Fluorescence of Background Readings).

Chapter 3

Results

3.1 Differential Sensitivity of the Human Cancer Cell Lines to Camptothecin, a DNA topoisomerase I poison

The cellular growth rate was measured for all the cell lines studied to evaluate their growth characteristic for establishing optimal experimental conditions. A typical cellular growth starts with a lag phase after reseeding. This phase could take a few hours to 48 hrs to recover from the trypsinization and activation of cellular regulation to rebuild its cytoskeleton and secretion of extracellular matrix to establish a linkage between the cells and the propagation substrate. Upon completion, the cells enter growth cycle and eventually reaching the log phase. It is this phase that the effect of drugs treatment is studied. As an example, a cellular growth curve of colon cancer HCT116 cell line is reported in Figure 22. The HCT116 cells were harvested and seeded in a 24-multiwells plate at different concentrations per well, the lowest concentration seeded was 10,000 and highest was 120,000 cells per well. At a fixed growth area of 2 cm^2 each well, cell concentration of 120,000 displayed a short lag-phase in the first 24 hrs of re-seeding and entered into logarithmic growth phase. Cells reached a short plateau at time 50 hrs and were likely to enter a cell death phase. For lower seeded cell numbers, the curve was similar but the initial gap (lag phase) increases before the logarithmic phase. Thus, a cell concentration of 20,000 was chosen for cell killing assay as it allows enough time to measure cell growth inhibition following drug treatments.

Similar experiments were performed with HCT15, U2OS and COLO205. They have displayed similar cellular growth characteristics for the different cell concentrations. This corresponded well to the reported cell doubling times for these cells line [Goldwasser *et. al.*, 1995; Fallica *et. al.*, 2012] that is between 20 - 24 hrs. Hence, a concentration of 20,000 was chosen and used for all cell lines.

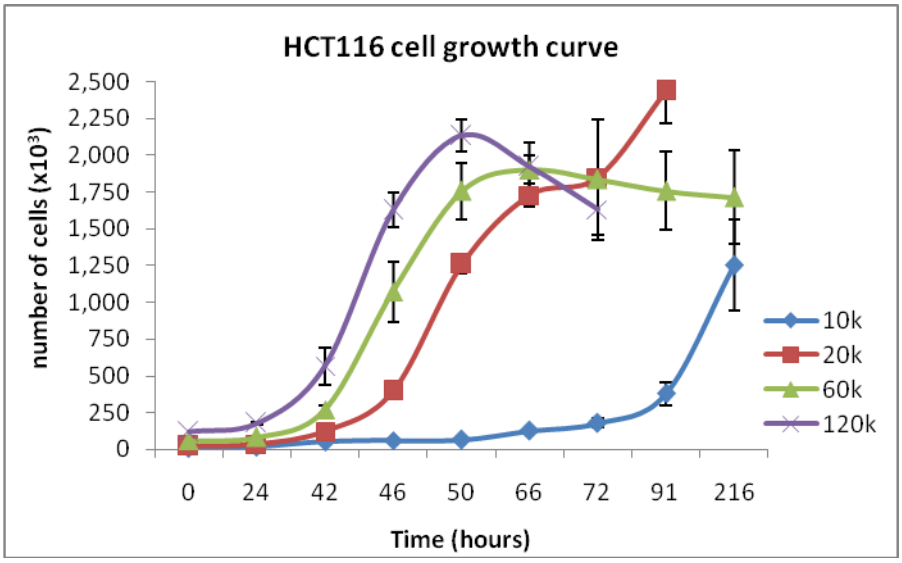


Figure 22. HCT116 cell growth curve. Graph showing a cell growth curve of different cell concentrations seeded in 24-multiwell plates. Cells were trypsinized and left for 24 hrs before counted every 12 hrs. The cell growth is indicated by the numbers of cells. Values are means ± S.D. of at least 3 wells for each time of calculation.

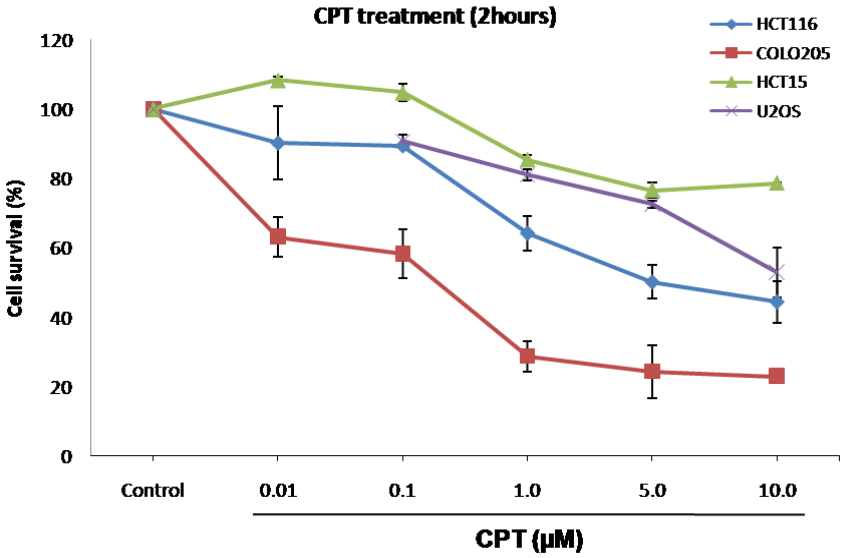


Figure 23. Cell survival after 2 hrs of CPT exposure. Three different cell lines; HCT116, COL205 and HCT15 were exposed to CPT after 24 hrs of seeding. Camptothecin concentrations between 0.01 to 10 µM were treated on cells for the indicated times (a) and (b). A control was use where cells were untreated. The cells were left to grow for 3 days after removal of drugs and the MTT assay was performed. Percentage of cell survival was obtained from absorbance values that were normalized to the control untreated cells. Values are mean ±S.D. of 2 - 4 wells of 2 independent experiments.

Upon understanding the cellular growth characteristics, we have investigated the sensitivities of COLO205, HCT116 and HCT15 to CPT treatment. These cell lines, despite originated from the colon cancers, showed different chromosomal ploidy mutations and exhibit natural differences in CPT sensitivity. Both HCT15 and HCT116 derived from a colon carcinoma while COLO205 was originated from a metastastic ascite fluid. Treatments of different CPT dosages were carried out for 2 hrs in all cell types. Cells with no treatment are used as experimental control. The survival of cells after CPT treatment was determined using the MTT assay. CPT exhibited marked cytotoxic effects in different colon cancer cell lines. COLO205 and HCT116 were most sensitive to CPT with IC50 less than 0.1 μM and 5 μM CPT, respectively (Figure 23). HCT15 was found to show the least sensitivity towards CPT with minimal drug activity at 2 hrs of treatment. The U2OS cells have sensitivity between HCT116 and HCT15. Cell exposure to CPT for 1 hr has yielded similar responses. These cell lines have been reported for their differences in sensitivity towards CPT and the obtained observations were in agreement with the previous publication [Goldwasser *et. al.*, 1995].

As Top1 is the sole target of CPT, I have then determined Top1 contents in these cells to understand if the differential sensitivity of cells towards CPT could be due to the different amount of cellular Top1. Hence, western blot experiments were performed on a whole cell lysate to evaluate the cellular Top1 levels in all colon cancer cell lines. Total protein was extracted from these cell lines was electrophoresed using a polyacrylamide gel. The presence of Top1 in these cell lines were detected using a Top1 antibody while an Actin antibody was used as loading control. The Top1 signal from each cell line was measured using ImageJ. Results showed that Top1 is detected in all 3 cell lines (~ 100 kDa). Additional band was detected at 70 kDa, indicating proteolysis of Top1 protein (Figure 24a). The amount of Top1 and its proteolysed form were quantified based on its signal intensity. Using ImageJ, Top1 signals were quantitated 3 times and the averaged mean values of Top1 were normalized against the actin signal to even-out protein loading error. Normalised Top1 signal is represented in Figure 24b. HCT116 has slightly lower Top1 signal comparing to HCT15 and COLO205 (which have almost the same amount of Top1). However, the differences in Top1 among these cell lines are not statistically significant (*t*-test, $p>0.05$), showing that the amount of Top1 in the cells is not the determinant factor of drug

sensitivities of the studied cancer cells, in agreement with previous studies [Perego *et. al.*, 1996; Goldwasser *et. al.*, 1995]. The U2OS cells, having CPT sensitivity between HCT116 and HCT15, were not evaluated in this comparison as it has a different histological origin (ovary), and therefore it is not homogenous with the other studied lines. Our findings and published data thus suggest that additional mechanisms may modulate the enzyme catalytic activity, and hence the response to the drug and eventually the cell sensitivity to CPT [Roy *et. al.*, 2014].

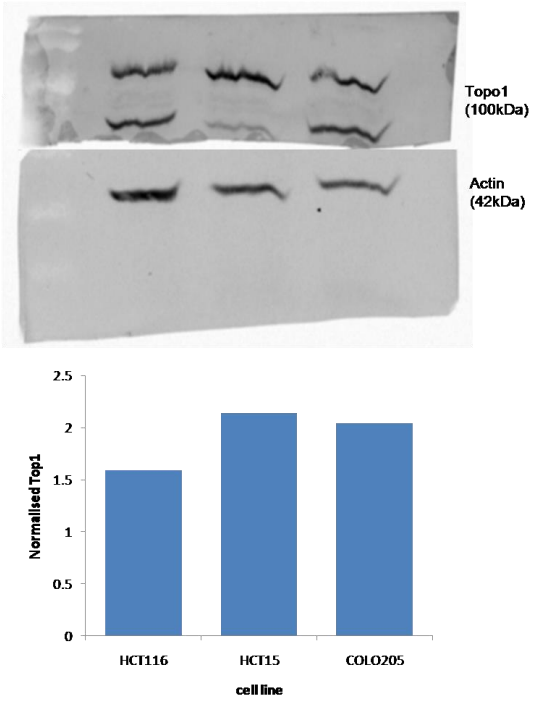


Figure 24. **Western blot for Topoisomerase 1 detection in HCT116, HCT15 and COLO205.** a) A polyacrylamide gel picture showing Top1 at 100 kDa and a loading control, Actin (42 kDa). Additional bands were observed at 70 kDa, indicating of Top1 proteolysis. b) Graph showing the amount of Top1 in each cell line. Signals were quantified using Image J (NIH). Each value was normalized against the loading control, Actin.

3.2 Differential Sensitivity of Human Cancer Cell Lines to Pyridostatin, Braco-19 and FG, agents that bind to G-quadruplex

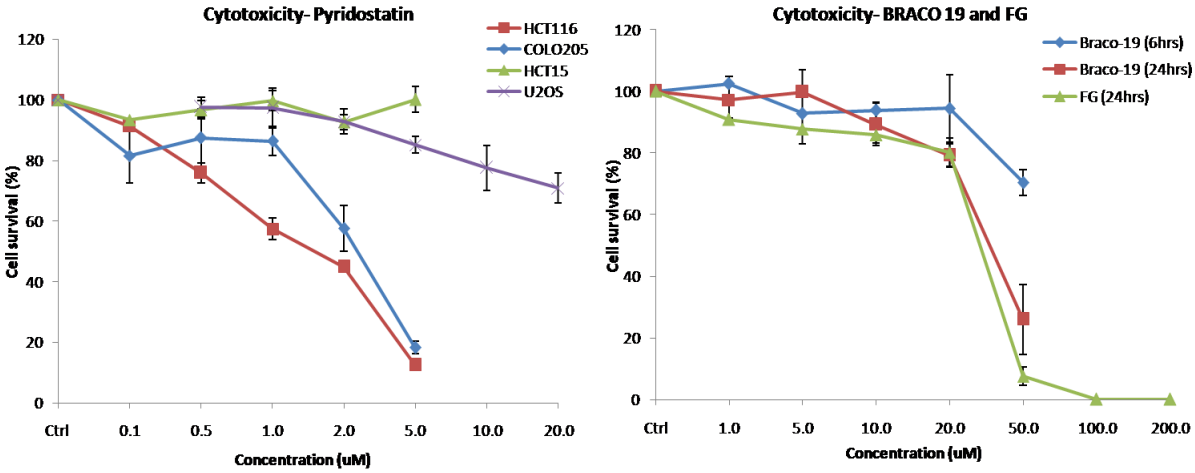


Figure 25. Cytotoxicity of Pyridostatin, Braco-19 and FG. a) Cytotoxicity of Pyridostatin in HCT116, COLO205, HCT15 and U2OS. A 20,000 cells were seeded and treated with 0.1 µM to at least 10 µM of Pyridostatin (except U2OS, a maximum of 20 µM was tested) for 24 hrs. b) Cytotoxicity of Braco-19 and FG in HCT116. Twenty-thousand cells were seeded and treated with 1.0 µM to 50 µM of Braco-19 for 6 and 24 hours while for FG 1.0 µM to 200 µM. Cells in (a) and (b) were left to grow for further 96 hrs after treatment. The MTT assay was performed. Untreated cells were used as controls. Cell survival was obtained from absorbance values at 595 nm (A_{595}) of treated cells that were normalized against the values of untreated cells. Values are mean absorbance±S.D. of at least 3 - 4 wells in 2 independent experiments.

Owing to the previous findings on CPT early effects in stimulating the formation of R-loops and associations that suggest G4 structures are modulated during transcriptional stress, we have hence use G4 binders as pharmacological tool to study the modulation of G4 in transcription. I have first started to investigate the cell sensitivity of G4 binders in all the cell lines, used in the previous sections. The cells were tested for their sensitivity to G-quadruplex binders. Cell proliferation assays were performed on different cell lines with G4 binders, ie: Pyridostatin, Braco-19 and FG. The cells lines, HCT116, COLO205, HCT15 and U2OS, were treated with Pyridostatin for 24 hrs at a concentration between 0.1 – 20.0 µM. Results showed that among the different colon cancer cell lines, HCT116 is most sensitive cell lines to Py followed by U2OS and COLO205. The least sensitive cell line, HCT15 has almost no detectable cytotoxic activity (Figure 25a). Albeit displaying resistance to

Pyridostatin and CPT, the use of both HCT15 and U2OS for synergistic studies between G4 and Top1 can be interesting. The use of them would display a clear interaction, if any, between Top1 and G4.

These observations were preceded using Braco-19 at different time exposures. Cytotoxicity of Braco-19 was investigated on HCT116 for 6 and 24 hrs of drug exposures. Both exposures showed similar reactivity with a higher cytotoxicity at 24 hrs of treatment, IC_{50} of approximately 35 µM (Figure 25b). The IC_{50}, however, could not be determinable at 6 hours of treatment in the tested range of concentrations. Exposure of HCT116 to FG for 24 hours showed similar response as in Braco-19.

3.3 Synergistic interaction between Pyridostatin and CPT sensitizes cells towards apoptosis

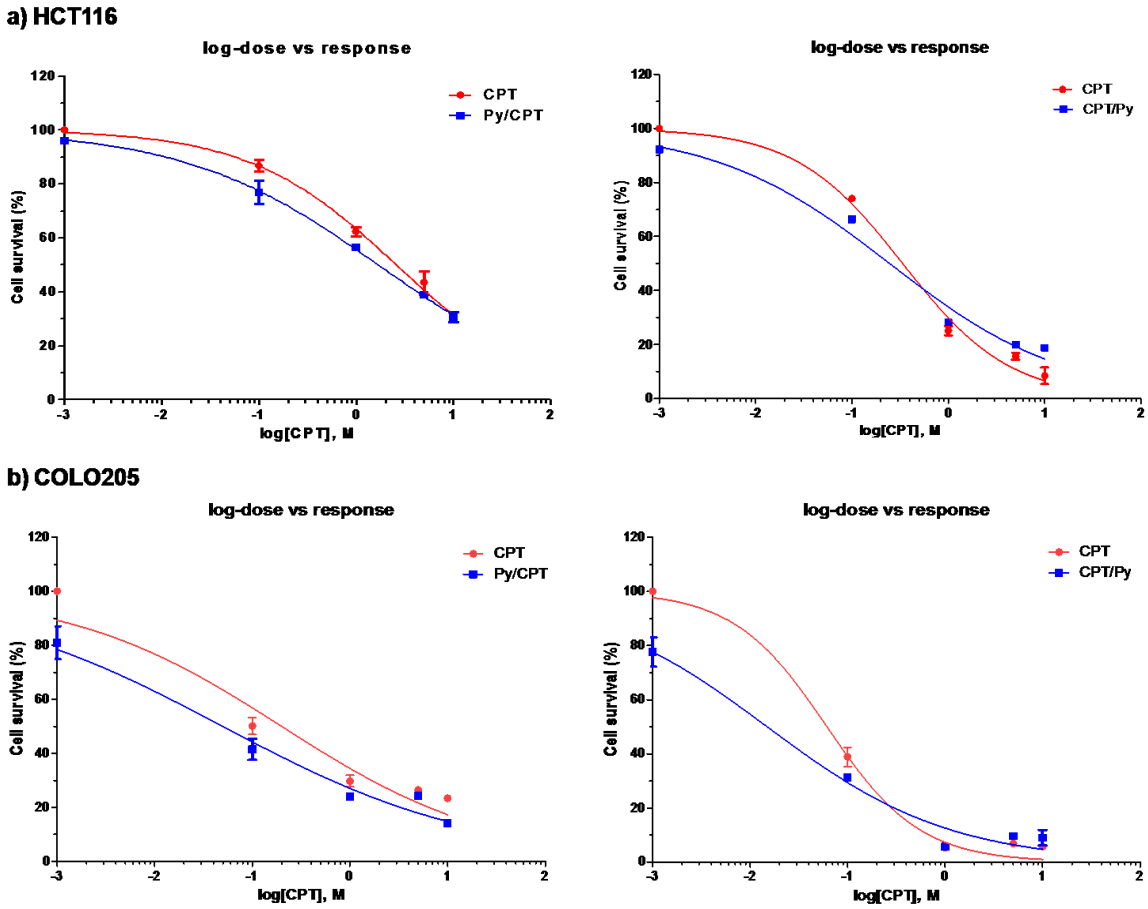


Figure 26 (*continue next page*)

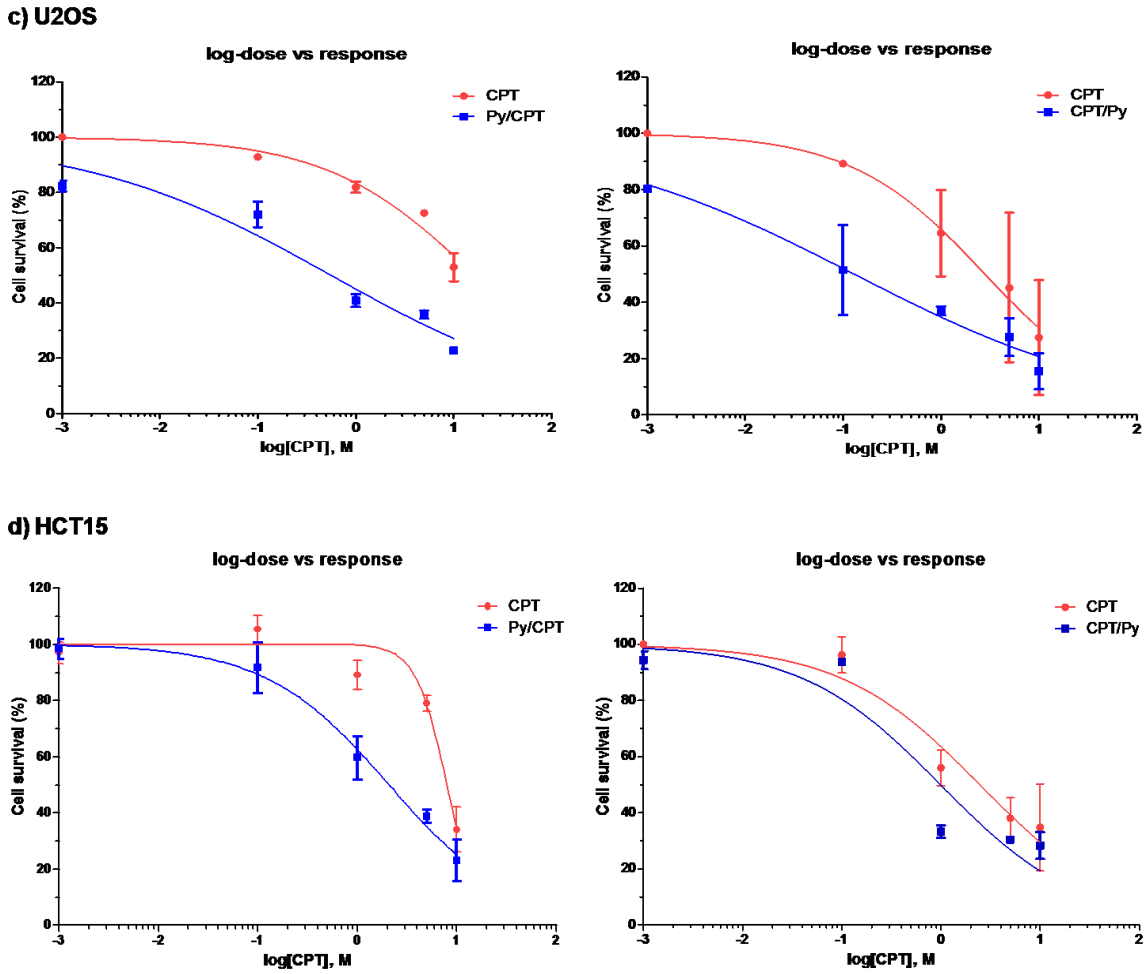


Figure 26. Log-dose versus response graphs for different cell lines. a) HCT116; b) COLO205; c) U2OS; and d) HCT15. Log-dose versus response graphs were computed from normalized absorbance values obtained from MTT assays using GraphPad Prism5. Nonlinear regression model was used to find the best fit curve in the log concentration and response data. The IC_{50} values were computed from these curves.

		HCT116		COLO205		HCT15		U2OS	
CPT (uM)	Py (uM)	Py/CPT	CPT/Py	Py/CPT	CPT/Py	Py/CPT	CPT/Py	Py/CPT	CPT/Py
0.1-10.0	0	2.584	0.3362			7.936	2.454		
0.1-10.0	0.25	1.691	0.2465			2.094	0.986		
0.1-10.0	0			0.1987	0.0614				
0.1-10.0	1.0			0.0495	0.0150				
0.1-10.0	0							16.50	2.813
0.1-10.0	10.0							0.559	0.1296

Table 1. Half-maximal inhibition concentration (IC_{50}) for each cell line. Table showed IC_{50} of CPT before and after co-treatment with Pyridostatin. The values were computed from cell survival rates in different cell lines using GraphPad Prism5.

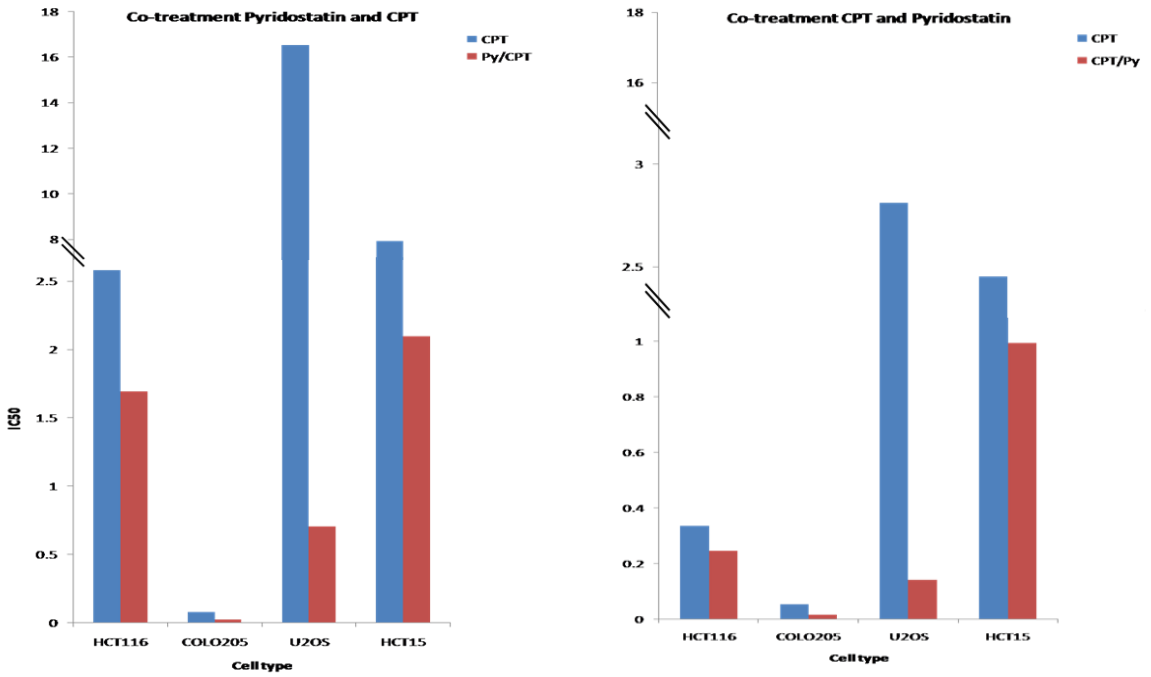
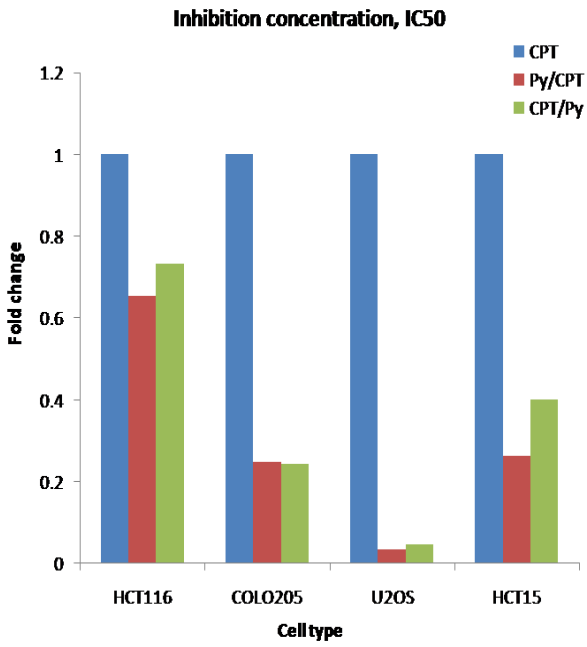


Figure 27. Graphs showing IC_{50} values for CPT and Pyridostatin co-treatments in different cell lines. a) Co-treatment Pyridostation/CPT; b) Co-treatment CPT/Pyridostatin. The IC_{50} values for CPT computed from log-dose versus response graphs were used to show their reductions when cells were co-treated with Pyridostatin, regardless of its sequences of treatment.



Fold change	Py ->CPT	CPT -> Py
HCT116	1.5 X	1.4 X
COLO205	4.0 X	4.1 X
U2OS	29.5 X	21.7 X
HCT15	3.8 X	2.5 X

Figure 28. Fold change of IC_{50} for CPT in co-treatments of different cell lines. a) Normalized IC_{50} of co-treatments to CPT treatment alone were obtained for each cell lines. The reduction of IC_{50} can be seen across tall cell lines and regardless of treatment sequences. b) Fold reductions of IC_{50} for CPT after co-treatment were then calculated as depicted in the table for each cell lines.

Extensive reviews on G-quadruplex on their roles in gene regulation based on in vitro studies [Broxson *et. al.*, 2011; Muller *et. al.*, 2010; Cogoi *et. al.*, 2006; Siddiqui-Jain *et. al.*, 2002] and their non-random putative distribution in the human genome [Zhang *et. al.*, 2013; Huppert *et. al.*, 2008], have sparked tremendous interests in this area. One of the recent findings have highlighted the interactions between Top1 and G4 in solid tumors. G4 binder on the telomeric region, RHPS4, in combination with camptothecin was shown to display synergistic effect on the antitumoral activity, hence, putting forward a hypothesis that both CPT and RHPS4 target G-strand of telomeric DNA especially during replication. Topologic aberrations during telomere replication might require more Top1 to be resolved and the stabilized G4 due to RHPS4, must be disrupted for replication to proceed [Biroccio *et. al.*, 2011; Leonetti *et. al.*, 2008]. These studies have given important clues on the interactions between Top1 and G4 structures. As an initial attempt to look into this, co-treatments

experiments were designed to look at drug synergism in cancer cell lines from the colon origin (HCT116, COL0205 and HCT15) and osteosarcoma (U2OS).

From previous cytotoxicity profiles, these cell lines posed different sensitivities towards CPT and G4 binders (Table 1). Interactions were investigated for their effect on different sequence of treatment since Leonetti *et. al.* (2008) reported a treatment sequence-dependent synergism. Treatments were performed as described in Methods. The cell survival obtained from the combination treatments were used to generate the log-dose versus response curves. Log-dose versus response curves showed a myriad magnitude of cell proliferation for co-treatments across different cell lines (Figure 26). Magnitude of reduction is greatest in U2OS comparing to HCT15, COLO205 and HCT116 showing the least reduction (Figures 27 and 28). The IC_{50} of CPT was reduced after co-treatments in all cell lines (Table 1). The IC_{50} for HCT116 with CPT treatment alone is 2.581 µM, reduced to 1.691 µM when co-treatment was performed with Py/CPT. Cell proliferation was also reduced in treatment condition CPT/Py (from 0.3362 µM to 0.2465 µM), reduced 1.5x and 1.4x respectively for the treatment sequences. For COLO205, IC_{50} for CPT alone is 0.07774 µM was reduced to 0.02331 µM in Py/CPT. For CPT/Py, the IC_{50} is 0.05475 µM and reduced to 0.01636 µM in co-treatment. In term of fold reductions, both treatment sequences showed a similar fold reduction across cell lines (Figure 28).

The U2OS showed greatest magnitude of IC_{50} reduction, 29.5x and 21.7x, although the cell line has the least sensitivity towards CPT and Py. IC_{50} for CPT alone was 16.54 µM and reduced to 0.7051 µM in Py/CPT while for CPT/Py, IC_{50} was reduced from 2.811 µM to 0.1422 µM (Table 1). Similarly like U2OS which showed least sensitivity to CPT and Py, co-treatments have enhanced reduction of IC_{50} by 3.8 and 2.5 times for both treatment sequences.

It is also noted that IC_{50} for CPT alone was different schedule experiments (Figure 27, left vs right panels). Depending on the exact combination treatment, CPT treatment was only started 48 hrs after cell seeding in the case of CPT only which started simultaneously when CPT was added into the 24 hrs Py-treated cells for Py/CPT combination. However, for CPT/Py sequence of treatment, CPT treatment started simultaneously after 24 hrs of seeding for both treatment conditions, ie; CPT only and CPT/Py combination. This difference is

most likely due to different cell-cycle distribution of cells after 48 hours of seeding as compared with 24 hrs. Moreover, Figure 3 also shows that cells are entering into its logarithmic phase at the 42nd hr. Cells are still in the lag phase in the first 24 hrs after seeding. Since CPT is most effect at S-phase cells, it is likely that cells entering the logarithmic phase of the growth are mostly cells in S-phase, hence the higher sensitivity to CPT treatment.

The potentiation of CPT cytotoxicity by Pyridostatin across cell lines is indeed a general event and the observation agree to the earlier reports. There is, however, no significant different between sequence of treatment. These results suggest that combination of Top1 inhibitor with G4 binder make cell unable to recover from drug-induced cytotoxicity. Stabilization of G4 by Pyridostatin decreases cell proliferation and causes cell to accumulate in G2 phase of the cell cycle, elicited DNA damage response activation such as phosphorylation of histone H2AX on Ser-139 [McLuckie *et. al.* 2013; Rodriguez *et. al.* 2012].

3.4 Synergistic interaction between other G4 binders and CPT

Using two other G4 binders, similar cell killing assay were carried n U2OS to confirm observations seen in Pyridostatin and CPT co-treatments. The potentiation of CPT effect by Pyridostatin was tested using Braco-19 and Bis-guanylhydrazone diimidazo[1,2-a:1,2-c]pyrimidine (FG) in U2OS cells. Similarly like the previous co-treatment experiments with Pyridostatin, U2OS cells were exposed to G4 binders after and prior to CPT exposure. Co-treatment of CPT with both G4 binders showed that once again, IC_{50} of CPT is reduced in co-treatmen, indication an elevation of CPT cytotoxicity in co-treatments. For Braco-19, IC_{50} of CPT was reduced 9.3 and 2.5 times with Braco-19 exposure before and after CPT treatment. The IC_{50} for CPT was reduced 3.6 and 9.9 times in co-treatment with FG (Table 2). These observations confirmed those of Pyridostatin co-treatments, hence showing that the potentiation effect in co-treatments is not limited to Pyridostatin but all G4 binders. Similarly, the sequence of co-treatments was not seen to be significant. Potentiation could also indicate interactions between G4 and Top1 that the stabilization of G4 and Top1 inhibition disrupt cell regulations which lead to apoptosis.

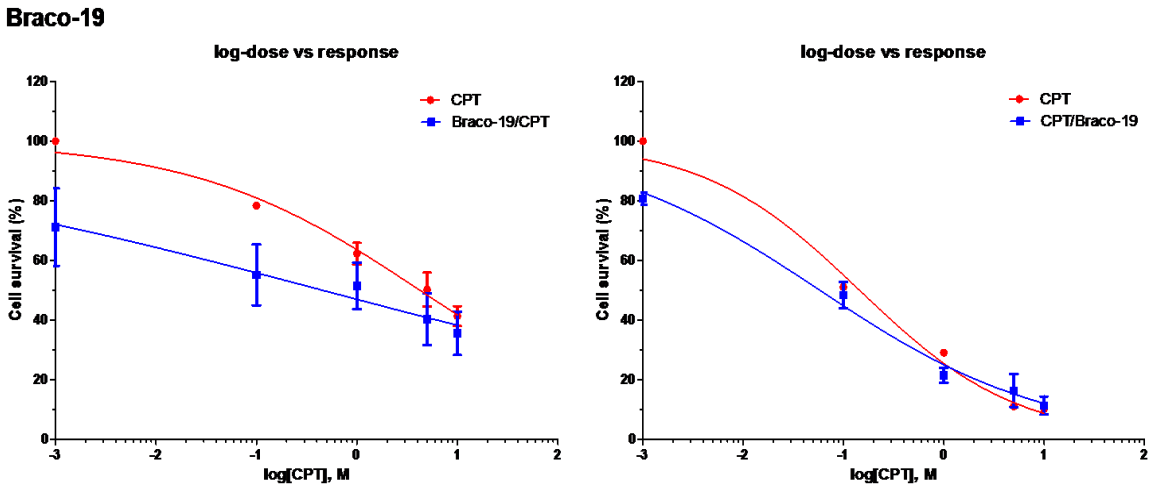


Figure 29. **Log-dose versus response graph for Braco-19 in HCT116.** a) Braco-19/CPT co-treatment; b) CPT/Braco-19 co-treatment. Log-dose versus response graphs were computed from normalized absorbance values obtained from MTT assays using GraphPad Prism5. Nonlinear regression model was used to find the best fit curve in the log concentration and response data. The IC_{50} values were computed from these curves.

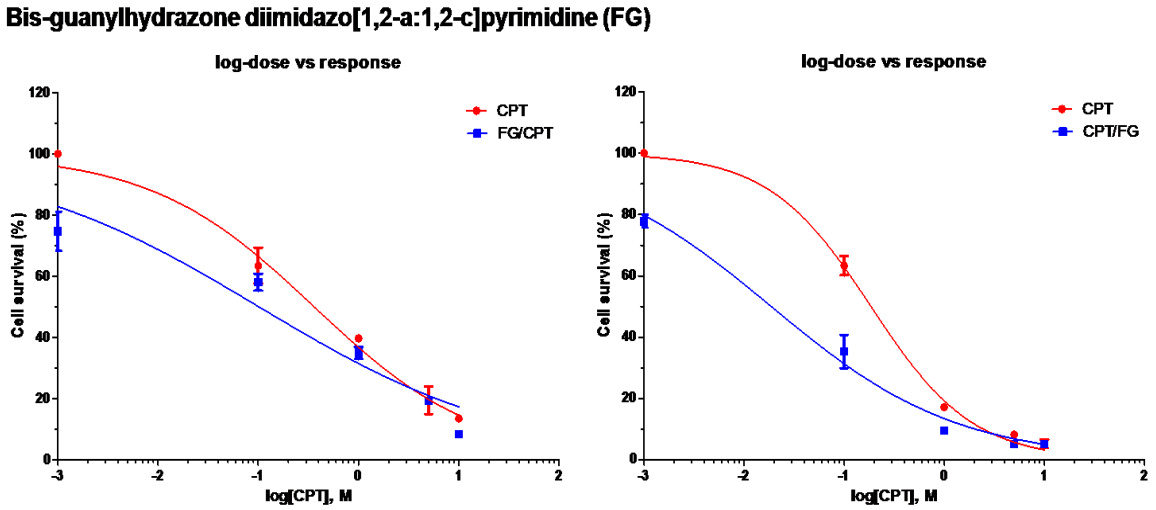


Figure 30. **Log-dose versus response graph for Bis-guanylhydrazone diimidazo[1,2-a:1,2-c]pyrimidine (FG) in HCT116.** a) FG/CPT co-treatment; b) CPT/FG co-treatment. Log-dose versus response graphs were computed from normalized absorbance values obtained from MTT assays using GraphPad Prism5. Nonlinear regression model was used to find the best fit curve in the log concentration and response data. The IC_{50} values were computed from these curves.

Fold change	Py →CPT	CPT → Py
Braco-19	**9.3x**	**2.5x**
FG	**3.6x**	**9.9x**

Table 2. Fold reductions of IC_{50} for CPT in co-treatments. IC_{50} obtained from log-dose versus response graphs for Braco-19 and FG were normalized to IC_{50} of CPT alone. The fold reductions of IC_{50} in co-treatments were calculated for each compound in HCT116 cells.

3.5 Putative G-quadruplex motifs on genes with CPT-induced antisense transcripts

Putative G-quadruplex (G4) motifs have been reported across the G-rich region of the genomic DNA amounting to 376,000 [Rodes *et. al.,* 2009]. The frequent G4 occurrence in telomeric regions, gene body and gene control regions suggest a functional correlation with gene activity [Zhang *et. al.,* 2013; Huppert *et. al.,* 2007]. Further to look into the interactions between Top1 and G4 structures, I have sought to predict formation of G4 structures putatively by searching for G4 motifs in the genes which has been shown to display antisense transcription, activated by CPT-inhibited Top1 [Marinello *et. al.,* 2013]. The distribution of the predicted G4 structures was then analysed. The information would enable us to ascertain whether putative presence of G4 motifs and its distribution in the gene could indicate possible correlation between G4 presence and dynamic effects of Top1 on divergent transcription at CpG-islands promoters. Putative G4 was predicted using QGRS mapper [Kikin *et. al.,* 2006] with high stringency as reported by Huppert *et. al.* (2006 and 2008) to search for $G_{3+}N_{1-7}G_{3+}\ N_{1-7}\ G_{3+}\ N_{1-7}\ G_{3+}$ motifs in the DNA sequences. The QGRS was set to find putative G4 structures with at least 3 guanines and constraining loop lengths of the G-quadruplex to a maximum of 7 bases.

All 225 genes that were found to have antisense transcription induced by CPT [Marinello *et. al.,* 2013] and a set of randomly selected genes, which showed no antisense transcription (253 genes), were analysed for the presence of putative G4 motifs. For each gene, the putative G4 motifs were categorized into pre-TSS (5000 bp upstream of transcription start

site), +500 bp downstream of TSS, +1000 bpdownstream of TSS and 5000 bp downstream of transcription termination site.

The occurrence of putative G4 motifs is found to be particularly higher in promoter genes with no antisense transcription. More than half of these genes have at least one G4 motifs at pre-TSS and +1000 bp TSS (Figure 31). Overall analysis of the putative G4 motifs looks to be enriched on the upstream region of the TSS. Interestingly, the occurrence of G4 motifs is found less prevalent in genes that displayed antisense transcription particularly on the template strand. These coincide with available genome wide data from Marinello *et. al.* (2013) that the accumulation of RNAPII antisense transcription occurs mainly in the upstream of TSS of divergent CpG island promoter, approximately 5000 bp upstream of TSS. Furthermore, analysis in the non-template versus the template strand (Figure 31b) reveals that approximately 55% of the genes contain at least one G4 motifs in the pre-TSS region of non-template strand compared to the same region of template strand. Our data corresponded well with those from a genome wide study by Zhang *et. al.* (2013). These observations implied a significant important in presence of G4 motifs which could affect the transcriptional activity of active promoters. The non-random enrichment of G4 motifs, particular on the upstream region of TSS could be a mode of downstream transcription activity sensing by G-quadruplex formation and its direct effect on DNA. The fast propagation of negative supercoiling generated by a proximal or distal downstream transcription or DNA tracking event induces a G-quadruplex formation at the PQS site and subsequently affects protein recognition and hinders protein translocation along the DNA.

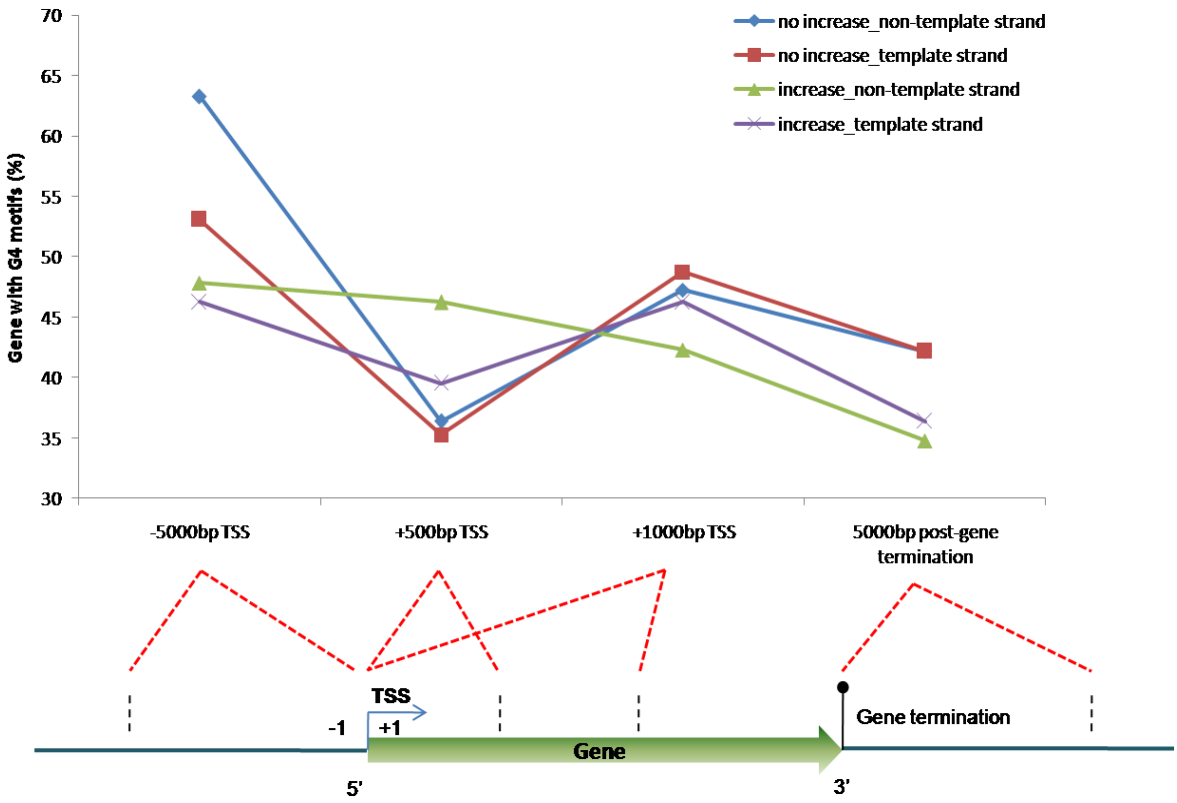


Figure 31. Putative G-quadruplex motifs. a) An overview of putative G4 motifs. Sequences downloaded from UCSC Genome Bioinformatics website were analysed for the presence of putative G4 motifs. A total of 225 promoters with antisense transcriptions and a set of 253 genes randomly chosen gene which showed no antisense transcription were analysed for putative G4 motifs using QGRS mapper. Regulatory sequences flanking 5000bp upstream of transcription start site (TSS), +500 bp and +1000 bp downstream of TSS plus 5000 bp downstream of 3'end were scanned for putative G4 motifs.; b) Diagram depicting the positions of the genome where G4 motifs were investigated in a gene.

3.6 Preparation of the antibody against G-quadruplex

The availability of G-Quadruplex specific antibody has provided valuable analytical tool to study the biological and therapeutical role of G4 structures [Handerson *et. al.*, 2014; Biffi *et. al.*, 2013]. In this study, the interplay between Top1 and G4 and the presence of putative G4 motifs have been established in cancer cells. As such, it is critical that these G4 can be confirmed through its visualization in vivo. We have, hence, proceeded to visualize G-quadruplexes in cells using a monoclonal antibody specifically directed against G4 motifs, named BG4, is engineered by Biffi and co-workers [Biffi *et. al.*, 2013]. The G4-specific

antibody is built as single-chained variable fragment fusion protein (scFv) that consists of the smallest functional antigen-binding domain of an antibody (~ 30 kDa) formed by engineering the heavy (VH) and light (VL) domains of the antibodies with a short linker polypeptide. Produced by phage display, BG4 was chosen from a library of scFv [Biffi *et. al.,* 2013]. The resultant scFV antibody populations were sub-cloned into an expression vector (pSANG10), which utilizes the high level expression promoter T7*lac*, driving expression of scFv to the bacterial periplasmic space via a *pelB*signal peptide sequence [Martin *et. al.,* 2006]. This resulted in the production of antibody that is fused with a six histidine tag for one step immobilised metal affinity chromatography (IMAC) and a tri-FLAG epitope tag for detection.

To prepare BG4 antibody, the recombinant plasmid is transfected into BL21(DE3) *E.coli* strain that possesses an inducible T7 RNA polymerase that transcribes the gene under the control of T7*lac* promoter. Together with a lac operator sequence is placed downstream to the start site of T7promoter to regulate the expression of basal protein expression. The presence of small percentage of both glucose and lactose in the auto-induction serves to reduce basal transcription and start expression when cells reached saturation. Glucose is favored over lactose when both are presence. Glucose reduces cAMP and catabolite activator protein (CAP). Under low concentration of cAMP, CAP is not able to bind to DNA, thus disabling the function of RNA polymerase. As glucose is depleted, catabolite repression is relieved, which leads to a shift in cellular metabolism toward the import and consumption of lactose and glycerol. Lactose import results in the production of allolactose from lactose by a promiscuous reaction of β-galactosidase. Allolactose then acts as the physiological inducer of the *lac* operon. Lactose is broken down by the lac operon to produce allolactose. Allolactose binds to lac repressor, thus removing it from the operator sequence and enabling the establishment of an elongation complex by T7 RNA polymerase at a T7*lac* promoter and substantially kick-starting the protein expression. It places the transition from the un-induced to induced state under metabolic control of the expression host. This protocol follows an autoinduction protocol for BG4 protein expression (as described by Studier, 2005) that is based on the ability of appropriate media to induce protein expression in *E. coli* when cells reach saturation.

Induction of protein expression in *E.coli* strain BL21(DE3) was monitored by taking a small amount of bacteria culture (5 ml) at times 0, 3, 6 and 24 hrs. SDS-PAGE of the bacterial crude protein lysate showed that, as induction time advanced, bacteria were seen to produce higher concentration of protein. Protein expression was visible as soon as 3 hrs of induction, which could probably be due to T7 promoter that has a very high activity. 24 hrs of induction gave a high concentration of proteins (Figure 32a).In comparison to induced cells, protein expression is visibly lower in control cells (not-induced).

Flowthrough represents the crude lysate that has been passed through the beads. The crude protein lysate were let to pass through the bead for 2 times. Electrophoresis on the flowthrough showed that both flowthroughs has almost equal intensity of protein. A second flowthrough did not seem to increase the binding capacity of beads. BG4 protein bound to the resin was then eluted with cold PBS/250 mM imidazole 3 times and a final elution with PBS/500 mM imidazole to ensure all proteins was totally dissociated from the resins. Purified antibody is seen mostly in the second and third elutions; these are combined and concentrated using Amicon Ultra to obtain higher concentration of antibody. High purity of antibody was obtained as no other visible bands are detected except an intense band at the molecular size of approximately 37 kDa (Figure 32b). BG4 was totally eluted after the third elution and no visible of BG4 was detected in the washing step.

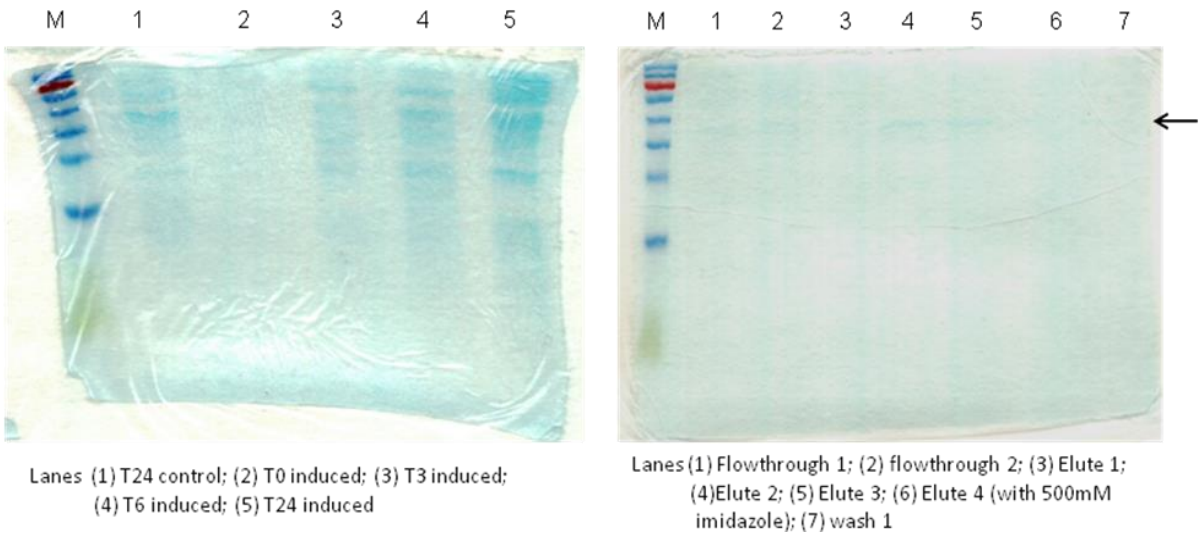


Figure 32. a) An SDS-PAGE of crude protein lysate from induced BL21(DE3) for different times. T indicates time (hour) of induction. Five mililitres of cell culture was collected from control which were uninduced for BG4 protein expression, Induced at T=0, Induced at T=3, induced at T=6 and T=24 were lysed and checked for overexpression of proteins using the SDS-PAGE. b) An SDS-PAGE showing purified protein lysate. Black arrow indicates the molecular weight where BG4 protein is seen. Eluted BG4 protein is visible at lanes 4 and 5. Protein from these lanes is accumulated and concentrated.

3.7 Visualization of G-Quadruplex in living cancer cells

As a continuation to our previous objective to produce BG4 for visualization of G4 in vivo, we have succeeded in expressing BG4 in bacterial expression and obtained a purified monoclonal antibody. We then attempted immunofluorescence staining to validate its functionality. The U2OS cells were seeded on glass slides and treated with Pyridostatin (Py) for 24 hrs at 10 µM. This concentration was used in agreement with published data [Biffi *et. al.*, 2013] and the 10 µM of Py was found to display low drug-induced cytotoxic effect on U2OS with a 80% of cell survival (Figure 6). Incubation of Py-treated U2OS with BG4 was carried out as described in Materials and Methods Section, and was detected by a fluorescein conjugated secondary antibody. Results showed that small punctuates of green fluorescence signal was seen in both controls and treated cells and they were localized in the nucleus of the cells, indicated by a nuclear staining with DAPI. Occasionally, fluorescence punctuates could be seen at areas in adjacent to the nuclear membrane, which is probably endogenous RNA G4 structures or G4 on RNA which was being transported out to the cytoplasm [Biffi

et. al., 2014]. Pyridostatin-treated cells displayed an increase of fluorescence staining as compared with control cells when viewed on a confocal microscope (Figure 33). Further to confirm these observations, we treated U2OS with Braco-19. Images from confocal microscope showed similar images as was seen from Pyridostatin treatment. The increased of fluorescence signal is further confirmed by Braco-19 treatment (Figure 34). Quantitation of the fluorescence signal showed an increased mean value for both experiment conditions compared to control (Figure 35). An increase of 1.5 and 1.8 of ratio fluorescence arbitrary unit in Py- and Braco-19 treated cells, respectively, when the fluorescence signal is normalized to the control cell (Figure 35b). Another G4 binder, FG was shown to produce similar effect on U2OS with an increase signal of 4x compared to normal cells. In addition, to control the specificity of FG towards G4 in vivo, we used another compound, named FA, which is a closed derivative of FG with few structural differences (Figure 12b, Structure 2), but cannot bind G4 motifs [Sparapani *et. al.,* 2010]. FA did not increase nuclear fluorescence of BG4 signals. The presence of fluorescence signal in control cells showed that stable physiological G4 are present in cells under normal conditions. As BG4 has previously been shown to display high specificity towards G4 structures, fluorescence signal indicates that the G4 structures formed were stabilised by Py, Braco-19 and FG. Treatment with G4 binders has enabled more G4 to be stabilized and hence a higher fluorescence signal. These observations indicated the functionality of our G4 antibody and are able to detect G4 in vivo.

Next, we used this G4 antibody to investigate the functional interaction between Top1 poisons and G4 binders. This might provide initial insights to assess whether transcriptional stress caused by CPT is associated with increased negative supercoiling, a condition favoring G4 formation. Hence, experiments were designed and carried out to investigate G4 formation under CPT effect. The U2OS cells were treated with 2 hrs of CPT alone and in combination treatments with Pyridostatin. Fluorescence signal for CPT alone show similar level with endogenous G4 and lower for co-treatments (Figures 34 and 35). A probable explanation to this observation is that the CPT was previously report to show an early effect on cells. An exposure of 2 hrs CPT would have caused G4 formed from CPT exposure to be resolved by helicases as RNAPII advances towards stalled Top1cc. It is also probably one of the homeostastical response towards a topologically balanced DNA structures. However,

Braco-19/CPT co-treatment did not show similar observations. Both Braco-19 only and Braco-19/CPT treatments showed elevated fluorescence signal compared to control. This could be due to the cytotoxic effect from high concentration used for Braco-19 as cytotoxicity assay was not tested on U2OS.

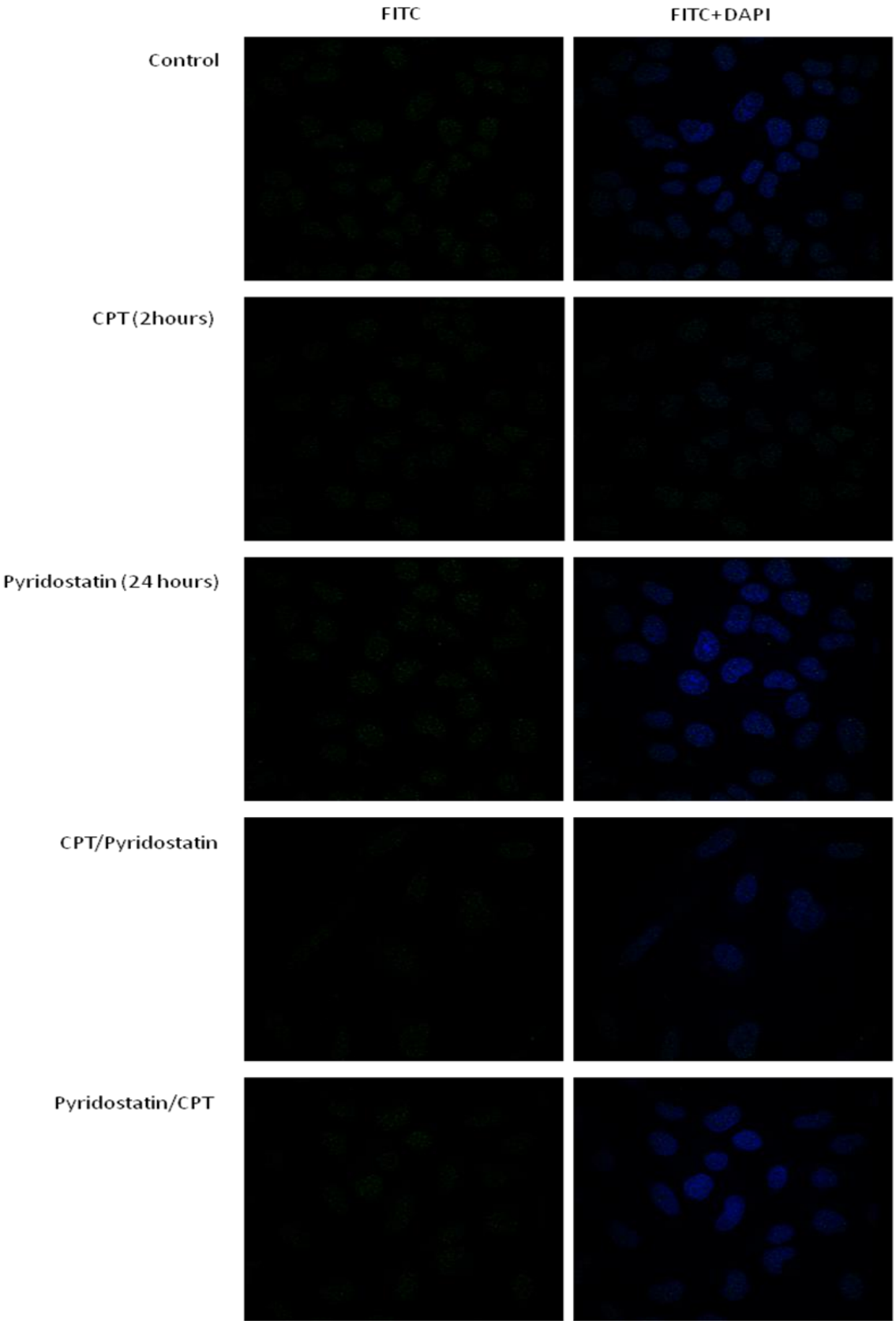

FITC
FITC+DAPI
Control
CPT (2hours)
Pyridostatin (24 hours)
CPT/Pyridostatin
Pyridostatin/CPT

Figure 33 (*Page 81*). Stabilization of G-quadruplex by Pyridostatin in U2OS. Images from confocal microscope showing G4 structures bound by Py. The U2OS cells were treated with Braco-19 and FG for 24 hrs at 10 μM, respectively. The cells were fixed with methanol/acetic acid and subjected to immunofluorescence staining. G-quadruplex structures were detected using the BG4 antibody and subsequently its detection using a fluorescein-conjugated secondary.

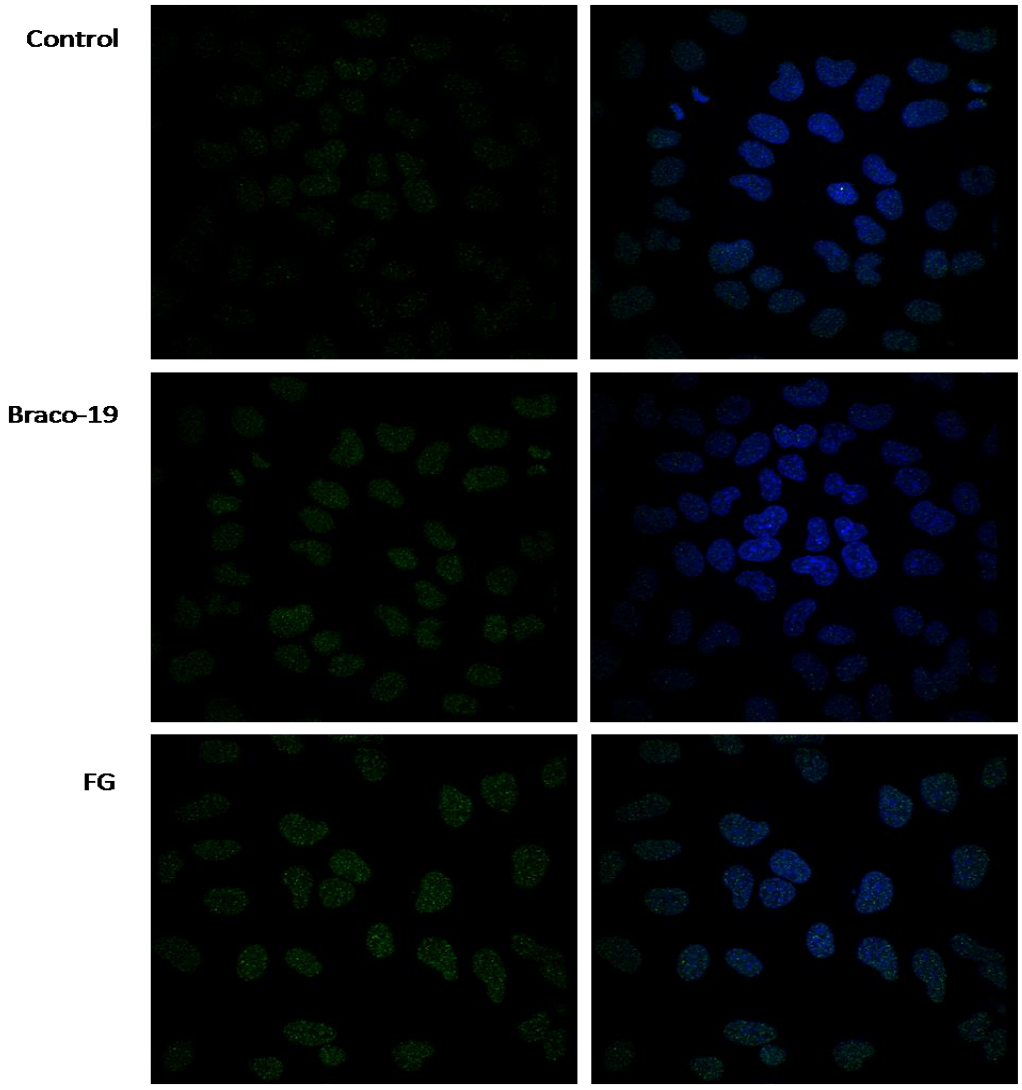


Figure 34. Stabilization of G4 by Braco-19 and FG in U2OS. Images from confocal microscope showing G4 structures bound by Braco-19 and FG. The U2OS cells were treated with Braco-19 and FG for 24 hrs at 10 μM, respectively. The cells were fixed with methanol/acetic acid and subjected to immunofluorescence staining. G-quadruplex structures were detected using the BG4 antibody and subsequently its detection using a fluorescein-conjugated secondary.

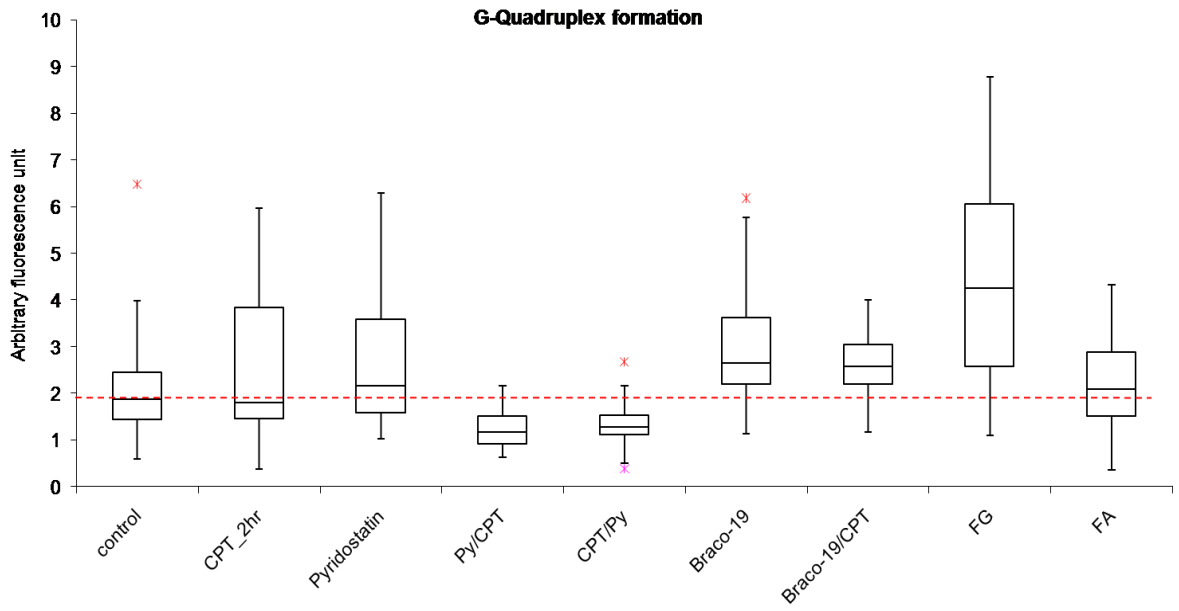


Figure 35. a) Box-plot of the G4 immunofluorescence staining. An * indicates maximum outlier fluorescence signal. b) Bar chart showing mean arbitrary fluorescence unit of G4 immunofluorescence staining.

3.8 Rapid formation of G-quadruplex by CPT

Previous data have shown that CPT-induced Top1ccs increased rapidly after 2 mins of treatment and followed by rapid removal at the promoters. The burst of Top1ccs forms a bell shaped curve with a peak at 2 mins and reduced thereafter. These, together with indications from my previous observations on the formation of G4 that is not increased upon treatment with 2 hrs of CPT have raised further questions if CPT effects in the same manner as Top1ccs. Hence, this prompted an investigation to look at the G4 formed by CPT in a time-dependant manner. In the experiment, shorter time exposures to CPT were planned. These times also correspond to those used for Top1ccs [Marinello *et. al.*, 2013]. The U2OS cells were exposed to CPT for 2 mins, 10 mins and 2 hrs at 10µM. Interestingly, with CPT alone, G4 formation peaked at 2 mins and reduced at 10 min and almost back to the level displayed by control cells at 2 hrs of CPT exposure (Figure 36). Moreover, G4 formation at 2 hrs of CPT agrees well with those from previous experiment (Figure 35). The formation of G4 was also investigated in co-treatment Py and CPT. Upon 24 hrs of Py exposure ended, the cells were exposed to CPT at different times as used in CPT treatment alone. Likely, G4 formation peaked at 2 mins and reduced at 10 mins and back to a level as in the control cells

for 2 hrs of CPT. These clearly showed a similar dynamic of G4 formation between CPT and Py/CPT treatments. All in all, the formation of G4 corresponds to the kinetics as seen in CPT-induced Top1ccs. CPT rapidly induced Top1ccs and so as G4 structure. As time of CPT is prolonged, both are also removed. G4 formation may be regulated when Top1ccs is induced by CPT which leads to the blocking of polymerase escape from promoter regions and their removal releases RNAPII allowing pausing escape and simultaneously maintaining a more balance DNA topology.

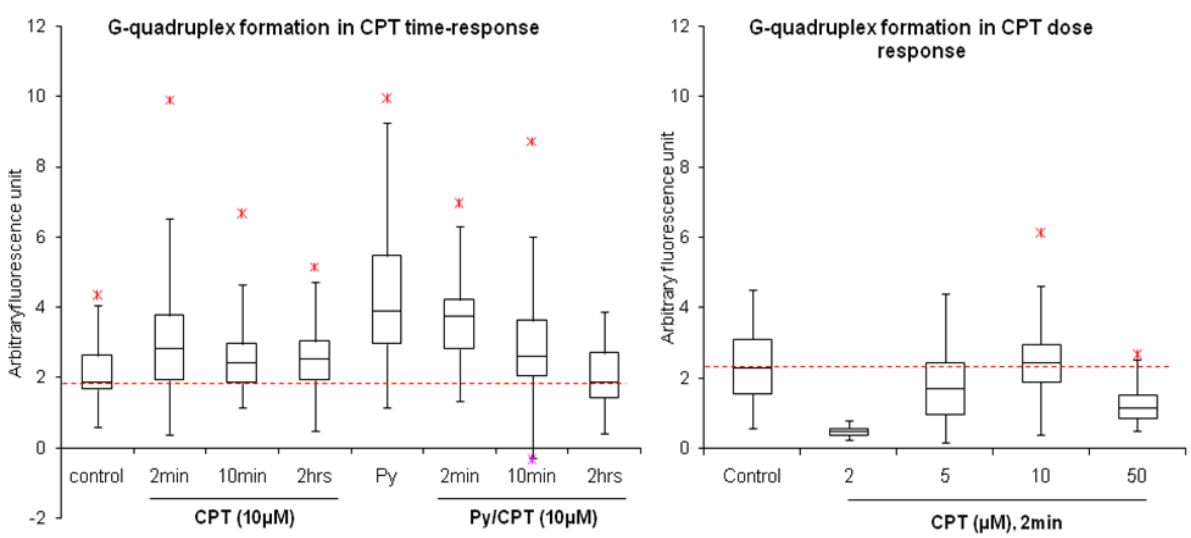


Figure 36. G-quadruplex formation in U2OS. a) CPT time-dependant experiment. An * indicates maximum outlier fluorescence signal. b) CPT dose-response experiment.

3.9 Rapid formation of G-quadruplex in CPT is dose dependent

Khobta *et. al.* (2006) and Marinello *et. al.* (2013) have previously reported that the effect of CPT is dose specific in HCT116 cell lines. They tested a range of CPT concentrations from as low as 0.08 μM to 10 μM and antisense transcriptions were seen at 10 μM of CPT treatment. Using this as a reference, the effect of CPT exposure on formation of G4 was investigated but in a larger concentration range. Experiment was carried on U2OS cells with different dose of CPT. Cells were exposed to 2 μM, 5 μM, 10 μM and 50 μM for 2 mins. Different doses of CPT showed an effect similar to the kinetics of Top1ccs, a bell-shaped

curve. G-quadruplex is least formed at 2 µM and increases at 5 µM of CPT. The fluorescence signal peaked at the exposure of 10 µM CPT but immediately reduced at 50 µM of CPT (Figure 36b). The results corresponded to those published earlier, hence, once again indicating an interesting regulations by Top1 on G4 formation when cells are exposed to CPT.

3.10 G-quadruplex formation in CPT-resistant HCT15 cells

HCT15 is one of the colon cancer cell lines used which has been previously established for its synergistic relationship between Topo1 inhibitor and G4 binder besides U2OS. The synergism relationship was indicated by a significant reduction in the IC_{50} of CPT when co-treatments were performed. It is, hence, of interest, to show that this kinetics is not limited to only U2OS. The relationship between CPT effects on the G4 formation was investigated in HCT15. Cells were first investigated for G4 formation in different CPT time exposures and in another, short exposure to different CPT dosage. In the CPT time exposure experiment, G4 formation in HCT15 displayed a peak fluorescence at 2 mins and reduced thereafter (10 mins and 2 hrs) (Figure 37a). In the CPT dose response experiment, G4 is least formed at 2 µM of CPT and steadily increase at 5 µM and peaks at 10 µM of CPT. The formation is however reduced at 50 µM of CPT (Figure 37b). In both cases, HCT15 was also treated with Pyridostatin for 24 hrs as control. G4 was shown to be formed and stabilized by Pyridostatin. These observations showed that CPT effect on the formation of G4 is not limited to type of cells used. Both experiments in HCT15 showed similar kinetics as in U2OS.

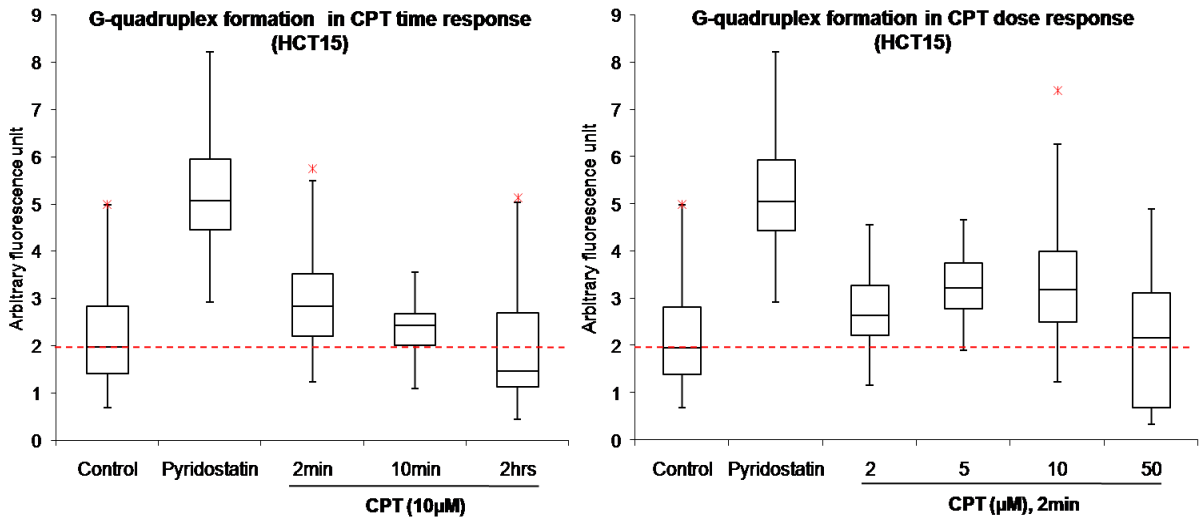


Figure 37. G-quadruplex formation in HCT15. a) CPT time-dependant experiment. The U2OS cells were treated with CPT in different times; b) CPT dose-response experiment. The U2OS cells were treated with different concentrations of CPT for 2 mins.

3.11 Elevation of R-loops formation in Pyridostatin-treated cells

R-loops are RNA:DNA hybrid structures, a by-product formed co-transcriptionally which are associated with genome instability. Their formation is one of CPT effects seen in cells. Previous study by Marinello *et. al.* (2013) reported that the burst of CPT-induced Top1cc paralleled with the formation of R-loops at transcriptionally active promoters. The R-loops formation is favored by the increasingly negative supercoiled DNA. Coincide where R-loops is present, G4 has been hypothesized to be present. Interestingly, it was shown in the previous section that G4 formation induced by CPT, displayed a similar kinetics in paralleled to the burst of CPT-induced Top1cc and R-loops formation. Hence, I sought to understand further modulation of R-loops formation in DNA damaged cells. The U2OS cells were treated with Pyridostatin at 10 µM for 24 hrs. Cells treated with 10 µM of CPT for 2 mins was used as experimental control for R-loops formation. After drug treatments, R-loops formation was detected using the DNA/RNA hybrid antibody, S9.6 (Figure 38). All R-loops fluorescence was shown to be localized to the nucleolus of the cells as visualized by the nucleus staining. A small amount of fluorescence was detected in the untreated U2OS cells and it increases 1.4 x when cells were treated with CPT for 2 mins, in agreement to previous experiments (Figure 39). Much to the excitement, fluorescence signal was significantly elevated by 4.5 x in Py-treated cells (Figure 39), leading us to believe that its formation could be modulated for DNA damage checkpoint activation. Treatment of cells with pyridostatin has been shown to induce DNA damage at specific genomic loci that contain putative quadruplex clusters sequences, such as the proto-oncogene *SRC*. Its G4-stalling nature by exerting mechanical forces stalls polymerases during transcription as indicated by an accumulation in G1 and G2 cells. This triggers transcriptional inhibition and elicited DNA damage as indicated by the production of γH2AX and cellular markers that lead to the ATM-mediated DNA damage response activation [Rodriguez *et. al.,* 2012]. The findings demonstrated that R-loops are formed in Pyridostatin–treated cells provide a strong indication that R-loops are modulated in response to DNA damage activation.

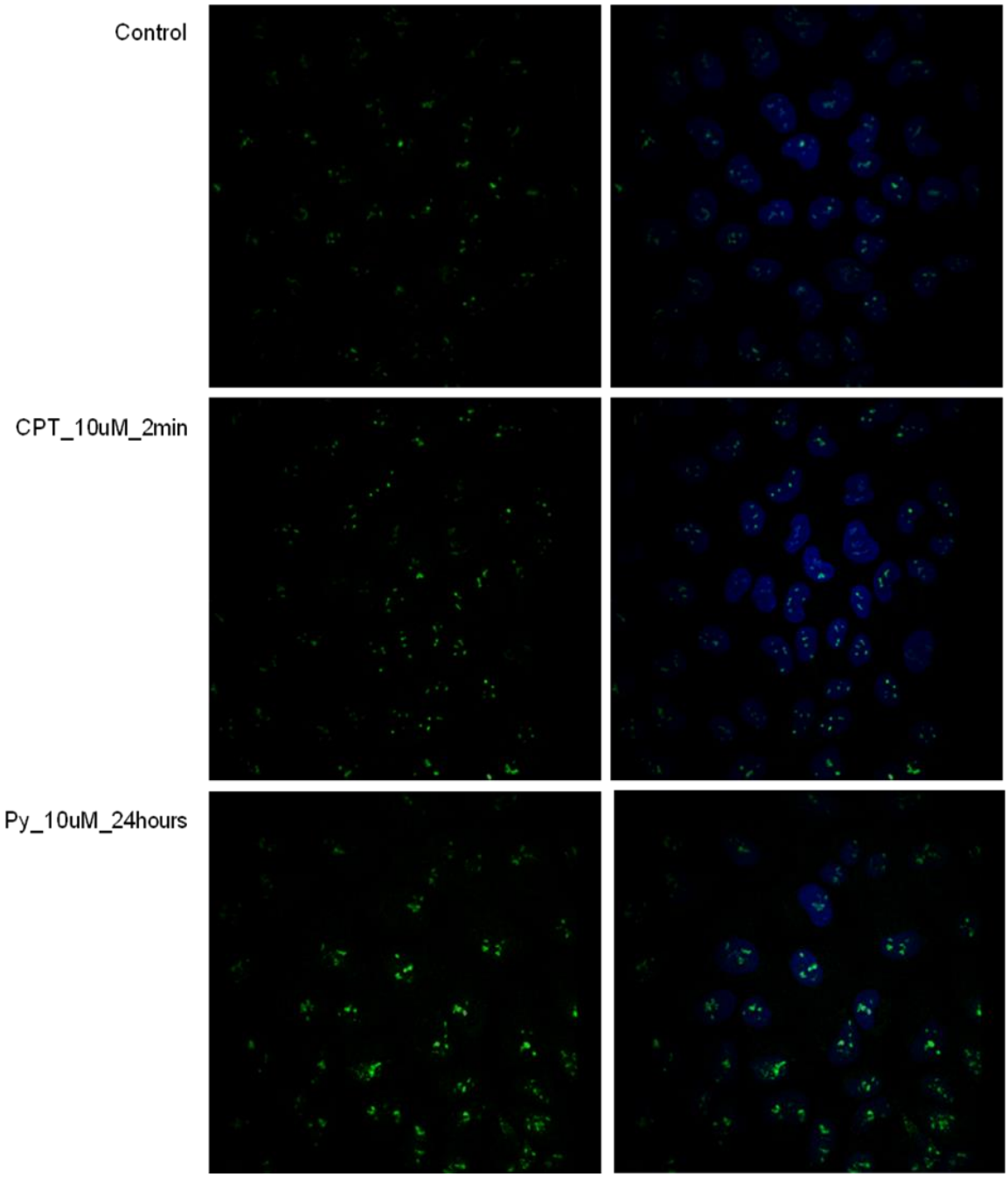


Figure 38. Images from confocal microscopy for R-loops formation. The U2OS cells were treated with 10 μM of CPT for 2 mins and 10 μM of Pyridostatin for 24 hrs. Control is cells without any drug treatment. The cells were fixed in ice cold methanol and subjected to standard immunofluorescence staining protocol. R-loops were detected using S9.6 antibody.

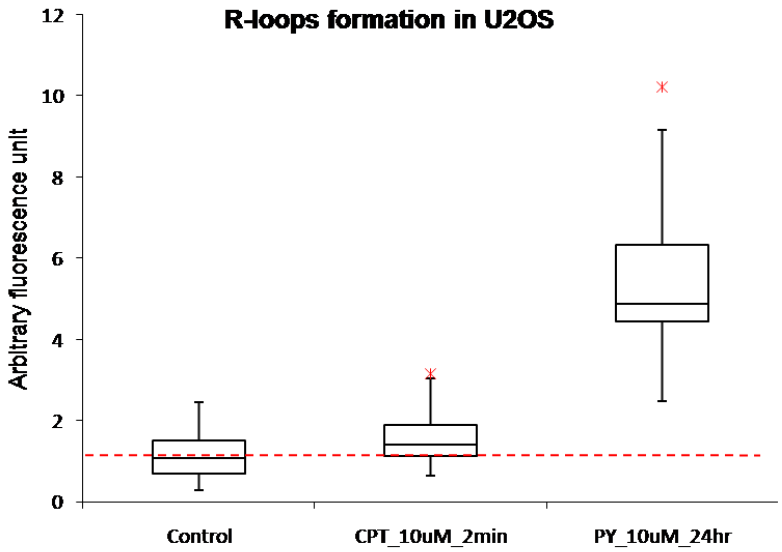


Figure 39. R-loops formation in U2OS. Quantification of images captured by confocal microscope (Figure 38) using ImageJ, NIH. Box-plot showed R-loops are formed in CPT-treated cells and is much elevated in Pyridostatin-treated cells.

3.12 Top1cc repair pathway resolves G-quadruplex structures and destabilizes the R-loops

In the previous sections, formation of R-loops and G4 were seen to be modulated in response to transcriptional stress by Top1 inhibition. To confirm these observations, we have sought to look at R-loops and G4 formations after inhibition of TDP1, an enzyme of the Top1cc repair pathway [Hubert *et. al.*, 2011; El-Khamisy *et. al.*, 2005].

To inhibit TDP-1, I have used the inhibitor, NSC88915, synthesized by Dexheimer *et. al.* (2009). The U2OS cells were incubated with TDP1 inhibitor over for 2 mins, 10 mins, 2 hrs and 24 hrs. The R-loops formation was then monitored using S9.6 antibody. R-loops formation seemed to be not affected by the short exposure of U2OS to TDP1 inhibitor. The level of R-loops is almost at physiological level and is not increase during all tested time of drug exposure. G4 formation after TDP-1 inhibition showed a slight increase in the first 2 hrs. This increment is only minor as compared with the dynamics seen in G4 binder time-course exposure, both in U2OS and HCT15 cell lines. The formation of G4 is reduced after a 24 hrs of drug exposure.

Results showed that the dynamics of R-loops (Figure 40a) and G4 (Figure 40b) formations which have been seen earlier in CPT treatment were not reproduced in TDP-1 inhibited cells. A small effect, if any, of drug is only seen after a prolonged treatment. Further experiments need to be performed to establish whether or not TDP1 has any role in G4 and R loop formation, in particular we need to assess the effects of TDP1 inhibitor on G4 and R loop levels induced by CPT and pyridostatin.

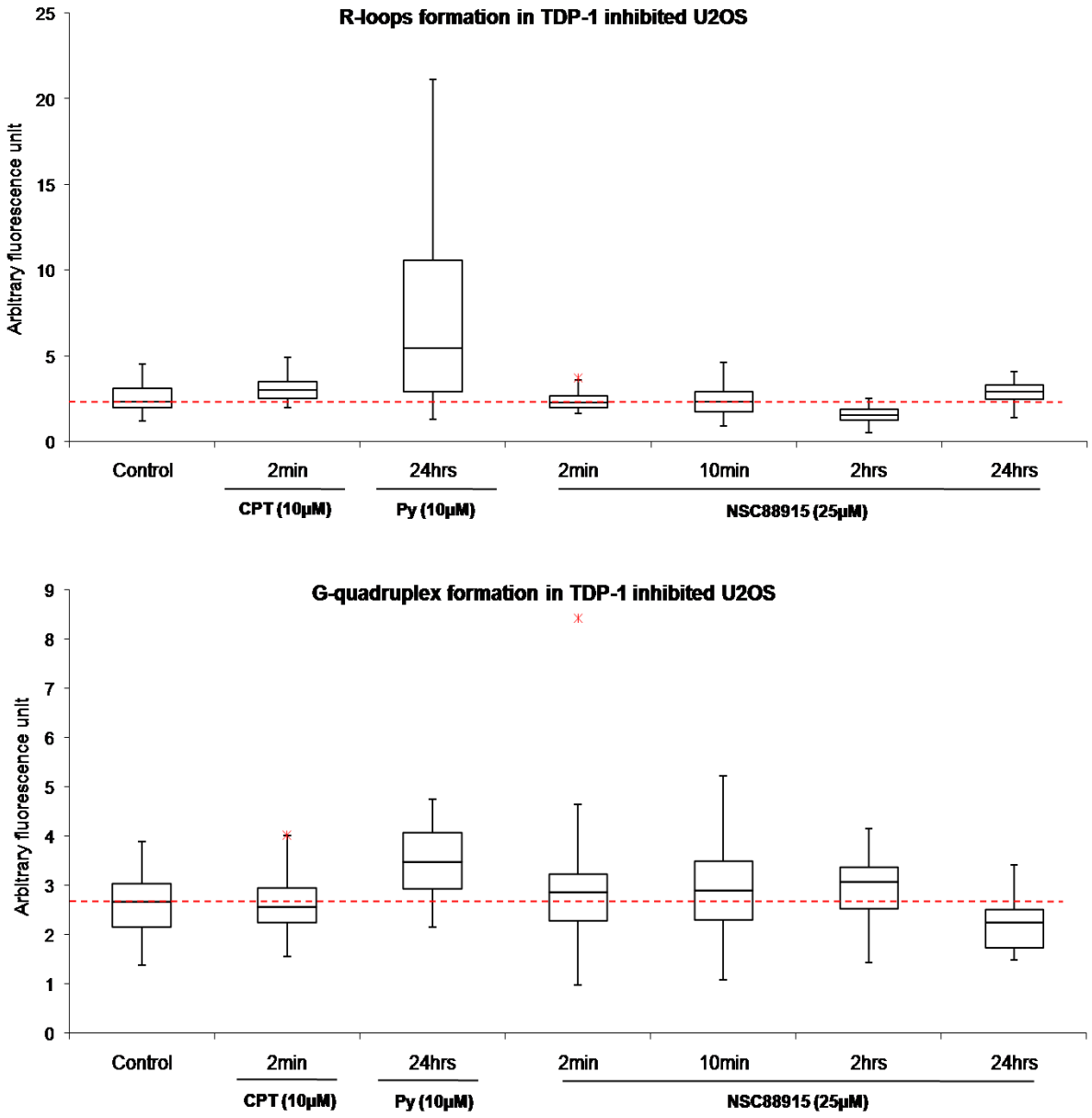


Figure 40. TDP-1 inhibition in U2OS. a) R-loops formation in TDP-1 inhibited U2OS. b) G-quadruplex formation in TDP-1 inhibited U2OS.

Chapter 4

Discussion

Previous studies from our laboratory have demonstrated intrigue findings on the role of Top1 during transcription and on the transcriptional consequences when Top1 is inhibited by CPT in human cell lines. Earlier work from the laboratory has unexpectedly observed an increase of RNAPII escape from promoter proximal pausing with marked alteration of HIF-1α co-transcriptional splicing, increased chromatin accessibility and activation/repression of antisense transcripts [Baranello *et. al.*, 2009].

In addition, CPT can impair the balance of cellular antisense and sense transcripts of cancer related HIF-1α. CPT increases transcription of a novel long RNA (5' aHIF-1α) which is an antisense to human HIF-1α mRNA and a known antisense RNA at the 3'-end of the gene. Eventually, this CPT effect is further investigated in a genome-wide scale. Consequently, the analysis shows that CPT-induced Top1ccs trigger an accumulation of antisense RNAPII transcripts specifically at active divergent CpG-island promoters. Likely to the previous effects shown by CPT, this phenomenon is replication independent and Top1-dependent manner. Moreover, time-course data showed a burst of Top1ccs increased by CPT at promoter sites and along transcribed regions, causing a transient block of RNAPII at the promoter. Similar time course also shows a transient increase in R-loops in highly transcribed regions.

It is, however, the function of R-loops is unknown although they are implicated in a molecular response pathway leading to transcription-dependent genome instability and alteration in transcription regulation. Our unpublished data also shows that CPT-induced R-loops is not limited to nucleoli but also in the mitochondria DNA, although the stability of R-loops is more stably formed compared to those formed in the genomic DNA. In both scenarios, R-loops are formed particularly in the hypernegatively coiled DNA and G-rich regions of DNA. Collectively, these data point to another interesting aspect of CPT effect on the regulation of DNA topology. R-loops are three stranded structure which are formed

when RNA hybridizes to a complementary DNA strand of a DNA duplex, leaving the opposite DNA strand single-stranded. DNA hypernegatively and G-richness of the DNA, conditions which favor the formation of R-loops [Aguilera *et. al.,* 2012; Ginno *et. al.,* 2013]; are also found to be favourable conditions for the formation of G-quadruplexes (G4) [Sun *et. al.,* 2009; Huppert *et. al.,* 2008]. G-quadruplex structures are stacked nucleic acid structures that can form within specific repetitive G-rich DNA or RNA that have sequence motif G≥3NxG≥3NxG≥3NxG≥3. Both R-loops and G4 have been shown to be involved in regulatory functions *in vivo* (reviewed in Chapter 1). Despite the numerous reports on the existence of G4, earliest reports on the interactions between G4 and Top1 only emerged in 2000 by Arimondo *et. al.* and followed by Marchand *et. al.* (2002) using *in vitro* studies. Both studies have reported that the Top1 is able to bind to intermolecular and intramolecular G4 and induces the formation of intermolecular ones. It is shown that Top1 and not TopII is able to displace G4-DNA [Arimondo *et. al.,* 2000]. This specificity has also seen that an inhibition of CPT-induced Top1 trapping of DNA by G-quartet-forming and G-rich Single-stranded oligodeoxynucleotides [Marchand *et. al.,* 2002]. In vitro interaction between G4 binders and CPT have shown to increase cell cytotoxicity in malignant cells [Leonetti *et. al.,* 2008; Birroccio *et. al.,* 2010], further establishing compelling evidences of an *in vivo* Top1-dependent mechanism of action in G4 formation.

Our work is focused on the role of DNA topoisomerase 1 (Top1) interactions with G4 during transcription and its involvement in the Top1-dependent repair pathway. Although, we have previously reported extensively on the role of Top1 in transcription in a single gene and genome-wide scale, the interactions of Top1 and G4 is less studied *in vitro*. We have hence sought to characterize their interactions including a possible role of Top1 in DNA repair mechanism.

Using different cancer cell lines of colon and osteo origins, we show that they display different sensitivity to CPT that is independent from Top1 level (Figure 23), consistent with available data [Goldwasser *et. al.,* 2005]. Sensitivity of CPT in cells is defined by a parameter, inhibitory concentration (IC_{50}), is an inhibitory concentration to achieve half maximal inhibition. Among the cell lines tested, we show that COLO205 has the least IC_{50} and HCT15 having the highest IC_{50} (Table 1). Since Top1 is the specific target for CPT, we

look that the level of Top1 in these cell lines. We show that the level of Top1 is not predictive for the differential sensitivity in these cell lines as indicated by an insignificant Student's t-test. The sensitivity is rather thought to be correlated directly with the intracellular hTop1 activity level since mutations in the Top1 can confer resistance to CPT [Gongora *et. al.*, 2011; Li *et. al.*, 1996]. The cell lines also show different response to G4 binders. The G4 binders are small molecules that are capable to stabilize G4 structures preformed in the cells and/or G4 that are formed due to stress. Most of the G4 binders, such as RHPS4 [Leonetti *et. al.*, 2008], are well tolerated and have good toxicology profiles. In our study, we show that cell exposures to different G4 binders (Pyridostatin, Braco-19 and FG) show absence of drug-toxicity except dose-related hypotension (Figure 25). U2OS and HCT15 are two CPT-resistant cell lines that shows low or no cytotoxic response to Pyridostatin.

To look at the interactions between Top1 and G4, we co-treat cells with CPT and G4 binders. Cell cytotoxicity confers by CPT through conversion of Top1 to Top1ccs, display a remarkable increase of cytotoxicity when cells are co-treated with G-quadruplex (G4) binders. We show that the co-treatment potentiate cell cytotoxicity of CPT regardless of the treatment sequences. Potentiation is indicated by a reduced inhibition concentration (IC_{50}) with a more profound cytotoxicity in CPT-resistant cell lines, HCT15 and U2OS. Hence, indicating a possible molecular mechanism that may be modulated through the interactions between Top1inhibitor and G4 binders, leading cell death. The interactions have, in fact, been shown in *in vivo* study that RHPS4 and CPT combination treatment inhibits and delays tumor growth in the colon cancer in mice, thus increases survival [Birroccio *et. al.*, 2010; Leonetti *et. al.*, 2008]. This cytotoxic consequence of combination drug treatment has open new avenues of deducing the molecular effect of CPT in the modulation of DNA topology that leads to genomic instability.

This is further supported by the presence of G4 motifs as determined by computational analysis on 225 genes with CPT-induced antisense transcription. G4 motifs are present mostly 5000 bp upstream from transcription start site and notably lower in genes body. Comparisons between genes with no antisense transcription and genes with antisense transcription show that G4 motifs in this region are notably lower in the genes with antisense

transcripts. It is however, unclear how this prevalence plays a role in the antisense transcription. These promoters have not been curated for their biological ontology as classified by Ginno *et. al.* (2013) which has identified four classes of promoters according to their GC skewness. Class IV genes, for instance, represented the likely distribution of PHQS which is seen from our analysis (Figure 31) contains gene sets enriched for developmental regulators. Nevertheless, our analysis corresponded well with recent bioinformatics analyses which show strong bias towards the non-template versus the template strand, in higher species [Lam *et. al.*, 2013; Xiao *et. al.*, 2013].

Collectively, the cell cytotoxicity and the computational analyses, we have, hence, hypothesized that the occurrence of PHQS in genes can be regulated to be functionally meaningful in cell regulations. Given the particular trends of G4 distribution in genes that display antisense transcription, it is likely that they are regulated by Top1 for transcription regulation. We can hence speculate that CPT causes fast propagation of negative supercoiling generated by a proximal or distal downstream transcription, which in turn induces G-quadruplex and R-loops at the potential G4 forming sites (PQS) sites. These may then posed a physical hindrance that affects protein recognition and their translocation along the DNA.

The biological importance of G-quadruplex formation such as transcriptional regulation and DNA replication in maintaining genomic stability has been reported extensively. G-quadruplex structures exist and are stable in human DNA [Lam *et. al.*, 2013]. Consistently, our study using an engineered antibody specific for G4 structures shows that G4 structures can be readily detected and visualized particularly localized in the nucleoli of cells, even without drug treatments (Figures 33, 34 and35). Although so, their molecular functions *in vivo* are not well characterized, *in vitro* analyses suggest that their formations may pose physical obstructions for transcriptional machinery. The KRAS gene, for example, contains a nuclease hypersensitive polypurine-polypyrimidine element that is essential for transcription. The G-rich strand of this element is able to form G4 as shown by circular dichroism (CD) and DMS footprinting experiments. Cogoi *et. al.* (2006) showed that the stabilization of G4 by TMPyP4 (a G4 binder) in this element increases the melting temperature of the oligonucleotides that competes with the nuclear protein, hence

postulating a transcriptional inhibition effect [Cogoi *et. al.*, 2006]. Similarly, in vitro experiment showed that G4 stabilization in the nuclease hypersensitivity element (NHE) III of the c-MYC promoter region results in suppression of transcription activation (Yang *et. al.*, 2006; Siddiqui-Jain *et. al.*, 2002].

Intriguingly, CPT exposures in the cells show an increase of G4 formation, noted by an increase compared to the controls in the first 10 minutes of treatment and reduced thereafter (Figure 35). It shows a rapid uptake by CPT that causes the accumulation of Top1ccs that in turn affect DNA hypernegativity. In relation to this, our previous data that shows CPT induces a hyperphosphorylation of the Rpb subunit of RNAPII and promote the escape of RNAPII from pausing sites after 1 hour of CPT exposure. In this experiment, we see that G4 is resolved further reduced at 2 hours of CPT, an effect which may probably occur simultaneuously with the escape of RNAPII from pausing sites. Consequently, the CPT-induced Top1ccs may regulate the binding of helicases to DNA to resolve G4 structures, although, it is unsure how this is achieved. Arimondo *et. al.* (2000) and Marchand *et. al.* (2002) were one of the firsts to observe the interactions between Top1 and G4. They reported that Top1 binds to intramolecular G4 and induces formation of intermolecular G4 structures. In certain cases, the formation of G4 is promoted by cleaving DNA duplexes, hence, suggesting a possible role of Top1 to prevent higher order of structures during transcription or the enzyme activity in RNA splicing. And further strengthening their findings, recent studies have reported several G4-specific helicases that are capable to resolve these G4 structures [Chen *et. al.*, 2015; Paeschke *et. al.*, 2013; Wu *et. al.*, 2008]. It is likely that Top1ccs regulate the recruitment of these helicases to G4 structures to counter-response the CPT-induced transcriptional stress.

Moreover, the rapid increase of CPT-induced Top1ccs affect the DNA supercoiling towards a more negatively form, induces the transient formation of R-loops. The kinetics of R-loops appears to be in parallel with the formation of CPT-induced Top1ccs. This prompted us to look at the formation of G4 in response to CPT-induced Top1ccs. To our delight, the kinetics of G4 formation appears to be rapidly formed, a kinetics similarly seen to that of Top1ccs and R-loops. This rapid response towards CPT exposure has led us to hypothesize that Top1ccs can in fact regulate the formation of G4 and this transcriptionally-dependent

CPT effect has not been reported elsewhere, hence, can provide a vital piece of information in deducing the molecular effect of CPT on transcription. The observation also shows that G4 are stable structures given the suitable conditions. It is shown that their formation and stabilization are dependent on monovalent cations, specifically K+ and Na+ [Sen *et. al.*, 1990]. The K+ which is present physiologically stabilizes G4 through the coordination of the positively charged cations with the electronegative O6 atoms in the center channel of the adjacent stacked G-tetrads [van Mourik *et. al.*, 2005]. Coupled with a favourable topological conditions (ie; DNA hypernegativity), the rapid formation of G4 can, hence, be supported. The rapid formation of G4 induced by CPT is dose dependent (Figure 36b) where we show that a bell curve pattern of G4 formation with a peak at CPT 10uM, a concentration that shows antisense transcription in HCT116 [Khobta *et. al.*, 2004; Marinello *et. al.*, 2013]. We then extend the investigations in HCT15, a colon cancer cell line that is being shown for elevated CPT cytotoxicity upon G4 stabilization (Figure 37). Similar kinetics of G4 formation is seen, showing that an interaction of Top1ccs with G4 structures is in fact, not cell specific but a universal event. We may hence speculate that Top1 modulate DNA superhelicity at transcribed active promoters in order to regulate gene activation.

Moreover, G4 formation is highly elevated in Pyridostatin treated cells, which previous study shows increased formation of γH2Ax foci. This effect is also seen in the CPT-resistant cell lines, HCT15, indicating that the formation is a general event in response to CPT. We also show that Although, G4 has been shown previously to be formed in Pyridostatin-treated cells [Biffi *et. al.*, 2013; Rodriguez *et. al.*, 2012], we are caught by surprise that our investigation shows that R-loop formation is also greatly increased in Pyridostatin-treated cells (Figures 17a and 18a), indicating that these structures maybe modulated in response to DNA damage. Collectively, the data prompted us to look into the roles of R-loops and G4 formations in Top1cc-dependent damage pathway.

Tyrosyl-DNA phosphodiesterase 1 (TDP1) having a role in the resolution of single strand breaks induced by Top1 by hydrolyzing the phosphodiester bond between the DNA 3'-end and the Top1 tyrosyl moiety. Studies have shown that TDP-1 together with Top1 is responsible in several neurodegenerative diseases such as Huntington disease and mytonic dystrophy type 1, implicated by expanded trinucleotide repeats. Small interfering RNA

(siRNA)-mediated knockdowns shows that the Top1-TDP1-SSBR (single strand break) pathway is required to modulate the contraction of these repeats [Hubert *et. al.*, 2011].

In order to study the role of R-loops and G4 structures in Top1cc-dependant repair pathway, we inhibited tyrosyl-phosphodiestrase 1 (TDP-1) using a TDP-1 inhibitor. Although, we have not obtained a similar kinetics as seen in the formation of R-loops and G4, we show that prolonged TDP-1 inhibition show a small changes in their formation, if any. Although further experiments need to be performed to establish whether or not TDP1 has any role in G4 and R loop formation, (particularly, assessment is needed to look at the effects of TDP1 inhibitor on G4 and R- loop levels induced by CPT and pyridostatin), existing studies have been shown independently for their roles in genome instability. In non-proliferating primary neurons, it has been recently shown that transcription arrest by stalled Top1ccs activates the DSB-ATM-DDR (DNA damage response) pathway which induces the formation of γH2AX foci [Sordet *et. al.*, 2009]. By blocking Top1 activity, CPT may promote an increase of local negative supercoiling behind the transcriptional machinery thus stabilizing R-loops, which may increase genome instability [Sordet *et. al.*, 2009]. Rodriguez *et. al.* (2012) on the other hand, shows that treatment with Pyridostatin stabilizes G4 structures, generates DNA damage at specific genomic loci (as seen in the formation of γH2AX foci), leading to cell cycle arrest and transcriptional downregulation of several genes that contains PQS clusters. Pyridostatin also decreases proliferation of simian virus (SV40)-transformed MRC-5 human fibroblasts (MRC-5–SV40 cells) and various cancer cell lines.

Bibliography

1. Aguilera, A., García-Muse, T., *R loops: from transcription byproducts to threats to genome stability*, Mol, Cell, (2012). **46**(2): **p**. 115-124.
2. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., Molecular Biology of the Cell 4th edition, (2002). Garland Science New York.
3. Amente, S., Gargano, B., Napolitano, G., Lania, L., and Majello, B., *Camptothecin releases P-TEFb from the inactive 7SK snRNP complex*, Cell Cycle, (2009). **8**: **p.** 1249–1255.
4. Amy, C., Seila, J., Calabrese, M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., *Sharp Divergent Transcription from Active Promoters*, Science, (2008). **322**(5909): **p.** 1849-1851.
5. Annamaria Biroccio, Porru, M., Rizzo, A., Salvati, E., D'Angelo, C., Orlandi, A., Passeri, D., Franceschin, M., Stevens, M. F. G., Gilson, E., Beretta, G., Zupi , G., Pisano, C., Zunino, F., Leonetti, C., *DNA Damage Persistence as Determinant of Tumor Sensitivity to the Combination of Topo I Inhibitors and Telomere- Targeting Agents*, Clin. Cancer Res., (2011). **17**(8): **p.** 2227–2236.
6. Andreani, A., Burnelli, S., Granaiola, M., Leoni, A., Locatelli, A., Morigi, R., Rambaldi, M., Varoli, L., Farruggia, G., Stefanelli, C., Masotti, L., Kunkel, M. W., *Synthesis and antitumor activity of guanylhydrazones from 6-(2,4-dichloro-5-nitrophenyl)imidazo[2,1-b]thiazoles and 6-pyridylimidazo[2,1-b]thiazoles(1),* J. Med. Chem., (2006). **49**(26): **p.** 7897-7901.
7. Angelov, D., Bondarenko, V. A., Almagro, S., Menoni, H., Mongélard, F., Hans, F., Mietton, F., Studitsky, V. M., Hamiche, A., Dimitrov, S., Bouvet, P., *Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes*, EMBO J., (2006). **25**(8): **p.** 1669-1979.
8. Arimondo, P. B., Riou, J.F., Mergny, J. L., Tazi, J., Sun, J.S., Garestier, T., Hélène, C., *Interaction of human DNA topoisomerase I with G-quartet structures*, Nucleic Acids Res., (2000). **28**(24): **p.** 4832–4838.
9. Baranello, L., Bertozzi, D., Fogli, M. V., Pommier, Y., and Capranico G., *DNA topoisomerase I inhibition by camptothecin induces escape of RNA polymerase II from promoter-proximal pause site, antisense transcription and histone acetylation at the human HIF-1 a gene locus*, Nucleic Acids Res., (2010). **38**(1): **p.** 159–171.
10. Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M. M., Mirkin, S. M., Hanawalt, P.C., *Mechanisms and implications of transcription blockage by guanine-rich DNA sequences*, Proc. Natl. Acad. Sci. USA, (2010). **107**(29): **p.** 12816–12821.
11. Berger, J. M., Gamblin, S. J., Harrison, S. C., Wang, J. C., Structure and mechanism of DNA topoisomerase II. Nature. (1996). **379**(6562): **p.** 225-232.

12. Bermejo, R., Doksani, Y., Capra, T., Katou, Y. M., Tanaka, H., Shirahige, K., Foiani, M., *Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation,* Genes Dev., (2007). **21**(15): **p.** 1921–1936.
13. Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., Blau, H. M., *Reprogramming towards pluripotency requires AID-dependent DNA demethylation*, Nature, (2010). **463**: **p.** 1042–1047.
14. Biffi, G., Tannahill, D., McCafferty, J., Balasubramanian, S., *Quantitative visualization of DNA G-quadruplex structures in human cells,* Nat Chem., (2013). **5**(3): **p.** 182-186.
15. 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., (2014). **6**(1): **p.** 75-80.
16. Bird, G., Zorio, D. A., Bentley, D. L., *RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation,* Mol. Cell. Biol., (2004). **24**(20): **p.** 8963-8969.
17. Bochman, M. L., Paeschke, K., Zakian, V. A., *DNA secondary structures: stability and function of G-quadruplex structures*, Nat. Rev. Genet., (2012). **13**(11): **p.** 770–780.
18. Borgognone, M., Armas, P., Calcaterra, N. B., *Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes*, Biochem. J., (2010). **428**(3): **p.** 491-498.
19. Bosch, P.C., S.-Bayona, S., Koole. W., van Heteren, J.T., Dewar, J.M., Tijsterman, M., Knipscheer, P., *FANCJ promotes DNA synthesis through G-quadruplex structures.* The EMBO Journal, (2014). **33**: **p.** 2521–2533.
20. Boule J.B., Zakian V.A., *Roles of Pif1-like helicases in the maintenance of genomic stability*. Nucleic Acids Res., (2006). **34**: **p.** 4147-4153.
21. Bronstein, I. B., Vorobyev, S., Timofeev, A., Jolles, C. J., Alder, S.L., Holden, J.A., *Elevations of DNA topoisomerase I catalytic activity and immunoprotein in human malignancies*, Oncol. Res., (1996). **8**(1): **p.**17-25.
22. Broxson, C., Beckett, J., Tornaletti, S., *Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene,* Biochemistry, (2011). **50**(19): **p.** 4162-4172.
23. Budhathoki, J.B., Ray, S., Urban, V., Janscak, P., Yodh, J.G., and Balci, H., *RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP*. Nucleic Acids Research, (2014). **42**(18): **p**.11528–11545.
24. Capra, J.A., Paeschke, K., Singh, M., Zakian, V.A., *G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae,* PLoS. Comput. Biol., (2010): **p.** e1000861.

25. Capranico, G., Marinello, J., Baranello, L., *Dissecting the transcriptional functions of human DNA topoisomerase I by selective inhibitors: implications for physiological and therapeutic modulation of enzyme activity*, Biochim. Biophys. Acta., (2010). **1806**(2): **p.** 240-250.
26. Chatterjee, S., Zagelbaum, J., Savitsky, P., Sturzenegger, A., Huttner, D., Janscak, P., Hickson, I. D., Gileadi, O., Rothenberg, E., *Mechanistic insight into the interaction of BLM helicase with intra-strand G-quadruplex structures.* Nature communications, (2014). **5**: doi:10.1038/ncomms6556.
27. Chen, M. C., Murat, P., Abecassis, K., Ferré-D'Amaré, A. R., Balasubramanian, S., *Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase*, Nucleic Acids Res., (2015). doi: 10.1093/nar/gkv051.
28. Champoux, J. J., *DNA topoisomerases: structure, function, and mechanism*, Annu. Rev. Biochem., (2001). **70**: **p.** 369-413.
29. Christman, M. F., Dietrich, F. S., Fink, G.R., *Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II*, Cell, (1988). **55**: **p.**413–425.
30. Christensen, M. O., Larsen , M. K., Barthelmes , H. U., Hock, R., Andersen, C. L., Kjeldsen, E., Knudsen, B. R., Westergaard, O., Boege, F., Mielke, C., *Dynamics of human DNA topoisomerases IIα and IIβ in living cells*, J. Cell Biol. (2002). **157**(1) **p.** 31-44.
31. Chun, Y.S., Choi, E., Kim, T. Y., Kim, M. S., Park, J. W., *A dominant negative isoform lacking exons 11 and 12 of the human hypoxia-inducible factor-1alpha gene*, Biochem. J., (2002). **362**(1): **p.** 71-79.
32. Collins, I., Weber, A., Levens, D. *Transcriptional consequences of topoisomerase inhibition.* Mol. Cell Biol., (2001). **21**: **p.** 8437–8451.
33. Cogoi, S., Xodo, L. E., *G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription*, Nucleic *Acids* Research, (2006). **34**(9), **p:** 2536–2549.
34. Dempsey, L. A., Sun, H., Hanakahi, L. A., Maizels, N., *G4 DNA Binding by LR1 and Its Subunits, Nucleolin and hnRNP D, A Role for G-G pairing in Immunoglobulin Switch Recombination*, Journal of Biological Chemistry, (1999). **274**: **p.** 1066-1071.
35. Desai, S. D., Zhang, H., Rodriguez-Bauman, A., Yang, J. M., Wu, X., Gounder, M. K., Rubin, E. H., Liu, L. F., *Transcription-Dependent Degradation of Topoisomerase I-DNA Covalent Complexes*, Mol. Cell Biol., (2003). **23**(7): **p.** 2341–2350.
36. Dewar, J.M., Lydall, D., *Pif1-and Exo1-dependent nucleases coordinate checkpoint activation following telomere uncapping*, EMBO J., (2010). **29**: **p.** 4020-4034.

37. Dexheimer, T.S., Gediya, L.K., Stephen, A.G., Weidlich, I., Antony, S., Marchand, C., Interthal, H., Nicklaus, M., Fisher, R.J., Njar, V.C., Pommier, Y, *4-Pregnen-21-ol-3,20-dione-21-(4-bromobenzenesulfonate) (NSC 88915) and related novel steroid derivatives as tyrosyl-DNA phosphodiesterase (Tdp1) inhibitors*, J. Med. Chem., (2009). **52**(22): **p.** 7122-7131.

38. Drolet, M., Phoenix, P., Menzel, R., Massé, E., Liu, L. F., Crouch, R. J., *Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I,* Proc. Natl. Acad. Sci. USA., (1995). **92**(8): **p.** 3526-3530.

39. Drolet, M., Bi, X., Liu, L.F., *Hypernegative supercoiling of the DNA template during transcription elongation in vitro,* J. Biol. Chem., (1994). **269**(3): **p.** 2068–2074.

40. Duann, P., Sun, M., Lin, C.T., Zhang, H., Liu, L.F., *Plasmid linking number change induced by topoisomerase I-mediated DNA damage*, Nucleic Acids Res., (1999). **27**: **p.** 2905–2911.

41. Dunnick, W., Hertz, G.Z., Scappino, L., and Gritzmacher, C. *DNA sequences at immunoglobulin switch region recombination sites*, Nucleic Acids Res., (1993). **21**: **p.** 365-372.

42. Duquette, M.L., Handa, P., Vincent, J. A., Taylor, A. F., Maizels, N., *Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA*, Genes Dev., (2004). **18**(13): **p.** 1618–1629.

43. Durrieu, F., Samejima, K., Fortune, J. M., Kandels-Lewis, S., Osheroff, N., Earnshaw, W. C., *DNA topoisomerase IIα interacts with CAD nuclease and is involved in chromatin condensation during apoptotic execution,* Current Biology, (2000). **10**(15): **p.** 923–926.

44. El-Khamisy, S. F., Saifi, G. M., Weinfeld, M., Johansson, F., Helleday, T., *Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1*, Nature (2005). **434**: **p.** 108-113.

45. El Hage, S. L., French, A. L., Beyer, D., Tollervey, *Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis*, Genes Dev., (2010). **24**: **p.** 1546–1558.

46. Falkenberg, M., Larsson, N. G., Gustafsson, C. M., *DNA replication and transcription in mammalian mitochondria,* Annu Rev Biochem., (2007). **76**: **p.** 679-699.

47. Fallica, B., Maffei, J. S., Villa, S., Makin, G., Zaman, M., *Alteration of Cellular Behavior and Response to PI3K Pathway Inhibition by Culture in 3D Collagen Gels*, PLoS ONE, (2012). **7**(10): **p.** e48024. doi:10.1371/journal.pone.0048024.

48. Fang, G., Cech, T. R., *The β subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA,* Cell, (1993). **74**: p. 875–885.

49. Fusté, J. M., Wanrooij, S., Jemt, E., Granycome, C. E., Cluett, T. J., Shi, Y., Atanassova, N., Holt, I. J., Gustafsson, C. M., Falkenberg, M., *Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication,* Mol. Cell, (2010). **37**(1): **p.** 67-78.

50. Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J. L., Li, X., *R-loop-mediated genomic instability is caused by impairment of replication fork progression*, Genes Dev., (2011). **25**(19): **p.** 2041–2056.

51. Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S. K., Collins, J. B., Li, L., Gilmour, D. S., Adelman, K., *NELF-mediated stalling of Pol II can enhance* 112 *gene expression by blocking promoter-proximal nucleosome assembly,* Genes Dev, (2008). **22**(14): **p.** 1921-1933.

52. Ginno, P.A., Lim, Y.W., Lott, P. L., Korf, I., Chédin, F., *GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination,* Genome Res., (2013). **23**(10): **p.** 1590–1600.

53. Ginisty, H., Amalric, F., Bouvet, P., *Nucleolin functions in the first step of ribosomal RNA processing*, EMBO J., (1998). **17**(5): **p.** 1476–1486.

54. Grinstein, E., Du, Y., Santourlidis, S., Christ, J., Uhrbergand, M., Wernet, P., *Nucleolin Regulates Gene Expression in CD34-positive Hematopoietic Cells,* The Journal of Biological Chemistry, (2007). **282**: **p.** 12439-12449.

55. Goldwasser, F., Bae, I., Valenti, M., Torres, K., Pommier, Y., *Topoisomerase I-related parameters and camptothecin activity in the colon carcinoma cell lines from the National Cancer Institute anticancer screen,* Cancer Res., (1995). **55**(10): **p.** 2116-2121.

56. Gómez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., Foiani, M., Aguilera, A., *Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles*, EMBO J. (2011). 30(15): p. 3106-3119.

57. Gómez-Herreros, F., Schuurs-Hoeijmakers, J. H. M., McCormack, M., Greally, M. T., Rulten, S., Romero-Granados, R., Counihan, T. J., Chaila, E., Conroy, J., Ennis, S., Delanty, N., Cortés-Ledesma, F., de Brouwer, A. P. M., Cavalleri, G. L., El-Khamisy, S. F., de Vries, B. B. A., Caldecott, K. W., *TDP2 protects transcription from abortive topoisomerase activity and is required for normal neural function,* Nature Genetics. (2014). **46**: **p.** 516–521.

58. Gomes-Pereira, M., Hilley, J. D., Morales, F., Adam, B., James, H. E., Monckton, D. G., *Disease-associated CAG·CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion,* Nucleic Acids Res., (2014). doi: 10.1093/nar/gku285.

59. Gongora, C., Vezzio-Vie, N., Tuduri, S., Denis, V., Causse, A., Auzanneau, C., Collod-Beroud, G., Coquelle, A., Pasero, P., Pourquier, P., Martineau, P., Del Rio, M., *New topoisomerase I mutations are associated with resistance to camptothecin,* Mol. Cancer, (2011). **10**: **p.** 64.

60. González, V., Guo, K., Hurley, L., Sun, D., *Identification and Characterization of Nucleolin as a c*-myc *G-quadruplex-binding Protein*, J. Biol. Chem., (2009). **284**(35): **p.** 23622–23635.

61. Groh, M., Lufino, M. M. P., Wade-Martins, R., Gromak, N., *R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome,* PLoS Genet., (2014). **10**(5): **p.** e1004318.

62. Henderson, A., Wu, Y., Huang, Y. C., Chavez. E. A., Platt, J., Johnson, F. B., Brosh, R. M. Jr., Sen. D., Lansdorp, P. M., *Detection of G-quadruplex DNA in mammalian cells,* Nucleic Acids Res., (2014). **42**(2): **p.** 860-869.

63. Hirose, Y., Tacke, R., Manley, J.L., *Phosphorylated RNA polymerase II stimulates pre-mRNA splicing.* Genes Dev., (1999). **13**(10): **p.** 1234-1239.

64. Hubert L. Jr., Lin, Y., Dion V., Wilson, J. H., *Topoisomerase 1 and Single-Strand Break Repair Modulate Transcription-Induced CAG Repeat Contraction in Human Cells,* Mol. Cell. Biol., (2011). **31**(15): **p.** 3105-3112.

65. Huppert, J, L., Bugaut, A ., Kumari, S., Balasubramanian, S., *G-quadruplexes: the beginning and end of UTRs*, Nucleic Acids Res., (2008). **36**(19): **p.** 6260–6268.

66. Huppert, J. L., Balasubramanian, S., *G-quadruplexes in promoters throughout the human genome,* Nucleic Acids Res., (2007), **35**(2): **p.** 406 -413.

67. Human Biology, Benjamin Cummings, Pearson (2004).

68. Hong, X., Cadwell, G.W., Kogoma, T., *Escherichia coli RecG and RecA proteins in R-loop formation,* EMBO J., (1995). **14**: **p.** 2385–2392.

69. Interthal, H., Chen H. J., Champoux J. J., *Human Tdp1 cleaves a broad spectrum of substrates including phosphoamide linkages,* J. Biol. Chem., (2005). **280**: **p.** 36518-36528.

70. Itoh, T., Tomizawa, J., *Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H,* Proc. Natl. Acad. Sci. USA, (1980). **77**(5): **p.** 2450–2454.

71. Jaxel, C., Capranico, G., Kerrigan, D., Kohn, K. W., Pommier, Y., *Effect of local DNA sequence on topoisomerase I cleavage in the presence or absence of camptothecin*, J Biol Chem., (1991). **266**(30): **p.** 20418–20423.

72. Keeney, S., *Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis: Genome dynamics and stability,* PMC, (2015). **2**(2008): **p.** 81–123.

73. Koster, D. A., Palle, K., Bot, E. S., Bjornsti, M. A., Dekker, N. H., *Antitumour drugs impede DNA uncoiling by topoisomerase,* Nature, (2007). **448**(7150): **p.** 213-217.

74. Khobta, A., Ferri, F., Lotito, L., Montecucco, A., Rossi, R., Capranico, G., *Early effects of topoisomerase I inhibition on RNA polymerase II along transcribed genes in human cells,* J. Mol. Biol. (2006). **357**(1): **p.** 127-138.

75. Kikin, O., D'Antonio, L., Bagga, P. S., *QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences*, Nucleic Acids Res. (2006). **34**(Web Server issue): **p.** W676-82.

76. Kogoma, T., *Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription,* Microbiol. Mol. Biol. Rev., (1997). **61**(2): **p.** 212–238.

77. Kreuzer, K. N., Brister, J. R., *Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives,* Virology Journal, (2010). **7**: **p.** 358-374.

78. Lam, E. Y. N., Beraldi, D., Tannahill, D., Balasubramanian, S., *G-Quadruplex structures are stable and detectable in human genomic DNA,* Nat Commun., (2013). **4**: **p.** 1796.

79. Lee, D. Y., Clayton, D. A., *Properties of a Primer RNA-DNA Hybrid at the Mouse Mitochondrial DNA Leading-strand Origin of Replication,* Journal of Biological Chemistry, (1996). **271**: **p.** 24262-24269.

80. Leonetti, C., Scarsella, M., Riggio, G., Rizzo, A., Salvati, E., D'Incalci, M., Staszewsky, L., Frapolli, R., Stevens, M. F., Stoppacciaro, A., Mottolese, M., Antoniani, B., Gilson, E., Zupi, G., Biroccio, A., *G-quadruplex ligand RHPS4 potentiates the antitumor activity of camptothecins in preclinical models of solid tumors,* Clin. Cancer Res., (2008). **14**(22): **p.** 7284-91.

81. Li, X. G., Haluska, Jr. P., Hsiang, Y.H., Bharti, A., Kufe, D.W., Rubin, E.H., *Identification of topoisomerase I mutations affecting both DNA cleavage and interaction with camptothecin,* Ann. N.Y. Acad. Sci., (1996). **803**: **p.** 111–127.

82. Lima, C. D., Wang, J.C., Mondragón, A., *Three-dimensional structure of the 67K N-terminal fragment of E. coli DNA topoisomerase I,* Nature, (1994). **367**(6459): **p.** 138-146.

83. Lin, Y., Dent, S. Y., Wilson, J. H., Wells, R. D., Napierala, M., *R loops stimulate genetic instability of CTG.CAG repeats,* Proc. Natl. Acad. Sci. USA.,(2010). **107**(2): **p.** 692-697.

84. Lipps, H. J., Rhodes, D., *G-quadruplex structures: in vivo evidence and function,* Trends Cell Biol., (2009). **19**(8): **p.** 414-422.

85. Liu, L. F., Desai, S. D., Li, T.K., Mao, Y., Sun, M., Sim, S.P., *Mechanism of Action of Camptothecin,* Ann. N. Y., Acad Sci., (2000). **922**: **p.** 1-10.

86. Liu, L. F., Wang, J. C., *Supercoiling of the DNA template during transcription*. Proc. Nat. Acad. Sci. USA, (1987). **84**: **p**. 7024-7027.

87. London, T.B., Barber, L.J., Mosedale, G., Kelly, G.P., Balasubramanian, S., Hickson, I.D., Boulton, S.J., Hiom, K., *FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts*, J. Biol. Chem., (2008). **283**: **p.** 36132-36139.
88. Lotito, L., Russo, A., Chillemi, G., Bueno, S., Cavalieri, D., Capranico, G., *Global transcription regulation by DNA topoisomerase I in exponentially growing Saccharomyces cerevisiae cells: activation of telomere-proximal genes by TOP1 deletion*, J. Mol. Biol., (2008). **377**(2): **p.** 311-322.
89. Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., Lingner, J., *The Rat1p 5′ to 3′ exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae*, Mol. Cell, (2008). **32**: **p.** 465–477.
90. Ljungman, M., Hanawalt, P.C., The anti-cancer drug camptothecin inhibits elongation but stimulates initiation of RNA polymerase II transcription. Carcinogenesis, (1996). **17**: **p**. 31–35.
91. Marchand, C., Pourquier, P., Laco, G.S., Jing, N.,and Pommier, Y., *Interaction of Human Nuclear Topoisomerase I with Guanosine Quartet-forming and Guanosine-rich Single-stranded DNA and RNA Oligonucleotides*, The Journal of Biological Chemistry, (2002). **277**: **p.** 8906-8911.
92. Marinello, J., Chillemi, G., Bueno, S., Manzo, S. G., Capranico, G., *Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation*, Nucleic Acids Res., (2013). **41**(22): **p.** 10110–10123.
93. Masuda-Sasa T., Polaczek, P., Peng, X.P., Chen, L., Campbell JL. *Processing of G4 DNA by Dna2 helicase/nuclease and replication protein A (RPA) provides insights into the mechanism of Dna2/RPA substrate recognition*, J. Biol. Chem., (2008). **283**: **p.** 24359- 24373.
94. 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., (2011). **39**(20): **p.** 9023-9033.
95. McLuckie, K. I. E., Di Antonio, M., Zecchini, H, Xian, J., Caldas, C., Krippendorff, B. F., Tannahill, D., Lowe, C., Balasubramanian, S., *G-Quadruplex DNA as a Molecular Target for Induced Synthetic Lethality in Cancer Cells*, Journal of the American Chemical Society, (2013). **135**(26): **p.** 9640-9643.
96. Miao, Z.H., Player, A., Shankavaram, U., Wang, Y. H., Zimonjic, D. B., Lorenzi, P. L., Liao, Z. Y., Liu, H., Shimura, T., Zhang, H. L., Meng, L. H., Zhang, Y. W., Kawasaki, E. S., Popescu, N. C., Aladjem, M. I., Goldstein, D. J., Weinstein, J. N., Pommier, Y., *Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses.* Cancer Res., (2007). **67**(18): **p.** 8752-8761.
97. Mirkin, S. M., *Driving past four-stranded snags*, Nature, (2013). **497**: **p.** 449-450.

98. Mischo, H. E., Gómez-González, B., Grzechnik, P., Rondón, A. G., Wei, W., Steinmetz, L., Aguilera, A., Proudfoot, N. J., *Yeast Sen1 helicase protects the genome from transcription-associated instability*, Mol Cell., (2011). **41**(1): **p.** 21-32.

99. Morten, O., Morten, C. K., Larsen, Barthelmes, H. U., Hock, R., Andersen, C. L., Kjeldsen, E., Knudsen, B. R., Westergaard, O., Boege, F., Mielke, C., *Dynamics of human DNA topoisomerases IIα and IIβ in living cells*, J. Cell Biol., (2002). **157**(1): **p.** 31–44.

100. Müller, S., Kumari, S., Rodriguez, R., Balasubramanian, S., *Small-molecule-mediated G-quadruplex isolation from human cells*, Nature Chemistry, (2010). **2**(12), **p:** 1095–1098.

101. Nasiri, H.R., Bell, N. M., McLuckie, K. I. E., Husby, J., Abell, C., Neidle, S., Balasubramanian, S., *Targeting a c-MYC G-quadruplex DNA with a fragment library*, Chem. Commun.**,** (2014).**50: p.** 1704-1707.

102. Nordström, K., Dasgupta, S., *Partitioning of the Escherichia coli chromosome: superhelicity and condensation*, Biochimie, (2001). **83**(1): **p.** 41-48.

103. Nowotny, M., Gaidamakov, S. A., Ghirlando, R., Cerritelli, S. M., Crouch, R. J., Yang, W., *Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription*, Mol. Cell, (2007). **28**(2): **p.** 264-276.

104. Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., Zakian, V. A., *Pif1 family helicases suppress genome instability at G-quadruplex motifs*, Nature, (2013). **497**: **p.** 458–462.

105. Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., Lipps, H. J., *Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo*, Nature Struct. Mol. Biol., (2005). **12: p.** 847–854.

106. Perego, P., Giarola, M., Righetti, S. C., Supino, R., Caserini, C., Delia, D,, Pierotti, M. A., Miyashita, T., Reed, J. C., Zunino, F., *Association between cisplatin resistance and mutation of p53 gene and reduced Bax expression in ovarian carcinoma cell systems*, Cancer Res., (1996). **56**: **p.** 556-562.

107. Perego, P., Cossa, G., Tinelli, S., Corna, E., Carenini, N., Gatti, L., De Cesare, M., Ciusani, E., Zunino, F., Luison, E., Canevari, S., Zaffaroni, N., Beretta, G. L., *Role of tyrosyl-DNA phosphodiesterase 1 and inter-players in regulation of tumor cell sensitivity to topoisomerase I inhibition.* Biochem. Pharmacol., (2011). doi:10.1016/j.bcp.2011.09.021.

108. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., Reik, W., Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency, Nature, (2010). **463**: **p.** 1101–1105.

109. Pommier, Y., *DNA topoisomerase I inhibitors: chemistry, biology, and interfacial inhibition,* Chem. Rev., (2009). **109**: **p.** 2894–2902.

110. Pommier Y., *Topoisomerase I inhibitors: camptothecins and beyond*, Nat. Rev. Cancer, (2006). **6**(10): **p.** 789-802.

111. Powell, W. T., Coulson, R. L., Gonzales, M. L., Crary, F. K., Wong, S. S., Adams, S., Ach, R. A., Tsang, P., Yamada, N. A., Yasui, D. H., Chédin, F., LaSalle, J. M., *R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation*, Proc. Natl. Acad. Sci. USA., (2013). **110**(34): **p.** 13938-13943.

112. Ray, S., Bandaria, J.N., Qureshi, M. H., Yildiz, A. and Balci, H., *G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding.* Proc. Nat. Acad. Sci. USA, (2013). **111**(8): **p.** 2990–2995.

113. Reddy, K., Tam, M., Bowater, R. P., Barber, M., Tomlinson, M., Edamura, K. N., Wang, Y.H., Pearson, C.E., *Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats*, Nucleic Acids Res., (2011). **39**(5): **p.** 1749–1762.

114. Ribeyre C, Lopes J, Boule JB, Piazza A, Guedin A, Zakian VA, Mergny JL, Nicolas A. *The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo*, PLoS.Genet. (2009). **5**: **p.** e1000475.

115. Rodriguez, R., Miller, K. M., Forment, J. V., Bradshaw, C. R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., Jackson, S. P., *Small-molecule-induced DNA damage identifies alternative DNA structures in human genes*, Nat. Chem. Biol., (2012). **8**(3): **p.** 301-310.

116. Roy, A., Tesauro, C., Frøhlich, R., Hede, M. S., Nielsen, M. J., Kjeldsen, E., Bonven, B., Stougaard, M., Gromova, I., Knudsen, B. R., *Decreased Camptothecin Sensitivity of the Stem-Cell-Like Fraction of Caco2 Cells Correlates with an Altered Phosphorylation Pattern of Topoisomerase I*, PLoS ONE, (2014). **9**(6): **p.** e99628.

117. Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L., Lieber, M. R., *Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site*, Mol. Cell. Biol., (2010). **30**(1): **p.** 146–159.

118. Roy, D., Lieber, M.R., *G Clustering Is Important for the Initiation of Transcription-Induced R-Loops In Vitro, whereas High G Density without Clustering Is Sufficient Thereafter*. Mol. Cell. Biol., (2009). **29**(11): **p.** 3124-3133.

119. Roy, D., Yu, K., Lieber, M. R., *Mechanism of R-Loop Formation at Immunoglobulin Class Switch Sequences*, Mol. Cell Biol., (2008). **28**(1): **p.** 50–60.

120. Selzer, G., *Specific cleavage of the p15A primer precursor by ribonuclease H at the origin of DNA replication*, Proc. Natl. Acad. Sci. USA., (1982). **79**(23): **p.** 7082-7086.

121. Sen, D., Gilbert, W., *Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis*, Nature, (1988). **334**: **p.** 364-366.

122. Shaw, N.N., Arya, D.P., *Recognition of the unique structure of DNA:RNA hybrids*, Biochimie, (2008). **90**(7): **p.** 1026–1039.

123. 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*, PNAS, (2002). **99** (18): **p.** 11593–11598.

124. Skourti-Stathaki, K., Proudfoot, N. J., Gromak, N., *Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination*, Mol. Cell, (2011). **42**(6): **p.** 794-805.

125. Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G., Champoux JJ. *A model for the mechanism of human topoisomerase I*, Science, (1998). **279**(5356): **p.** 1534-1541.

126. Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., Hieter, P., *R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants*, Genes Dev., (2012). **26**: **p.** 163–175.

127. Sparapani, S., Bellini, S., Gunaratnam, M., Haider, S. M., Andreani, A., Rambaldi, M., Locatelli, A., Morigi, R., Granaiola, M., Varoli, L., Burnelli, S., Leoni, A., Neidle, S., *Bis-guanylhydrazone diimidazo[1,2- a:1,2- c]pyrimidine as a novel and specific G-quadruplex binding motif*, Chem. Commun., (2010). **46**: **p.** 5680–5682.

128. Sordet, O., Redon, C. E., Guirouilh-Barbat, J., Smith, S., Solier, S., Douarre, C., Conti, C., Nakamura, A. J., Das, B. B., Nicolas, E., Kohn, K. W., Bonner, W. M., Pommier, Y., *Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks.* EMBO Rep., (2009). **10**(8): **p.** 887-893.

129. Sordet, O., Larochelle, S., Nicolas, E., Stevens, E. V., Zhang, C., Shokat, K. M., Fisher, R. P., Pommier, Y. *Hyperphosphorylation of RNA polymerase II in response to topoisomerase I cleavage complexes and its association with transcription- and BRCA1-dependent degradation of topoisomerase I*, J. Mol. Biol., (2008). **381**: **p.** 540–549.

130. 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., (2009). **52**(9): **p.** 2863–2874.

131. Thomas, M., White, R. L., Davis, R. W., *Hybridization of RNA to double-stranded DNA: formation of R-loops*, Proc. Natl. Acad. Sci. USA., (1976). **73**(7): **p.** 2294-2298.

132. van Mourik, T., Dingley, A.J., *Characterization of the monovalent ion position and hydrogen-bond network in guanine quartets by DFT calculations of NMR parameters*, Chemistry, (2005). **11**(20): **p.** 6064-6079.

133. Vertino, P. M., Wade, P. A., *R Loops: Lassoing DNA Methylation at CpGi*, Molecular Cell, (2012). **45**(6): **p.** 708–709.

134. Wang, H., Nora, G. J., Ghodke, H., Opresko, P. L., *Characterization of the monovalent ion position and hydrogen-bond network in guanine quartets by DFT calculations of NMR parameters, J. Biol. Chem.*, (2011). **286**: **p.** 7479–7489.

135. Wang, J. C., *Cellular roles of DNA topoisomerases: a molecular perspective*, Nat. Rev. Mol. Cell Biol., (2002). **3**(6): **p.** 430-40.

136. Wang, J. C., *DNA topoisomerases*, Annu. Rev. Biochem., (1996). **65**: **p.** 635-692.

137. Wang, J. C., Caron, P. R., Kim, R. A., *The role of DNA topoisomerases in recombination and genome stability: a double-edged sword?*, Cell, (1990). **62**(3): **p.** 403–406.

138. Wang, Q., Liu, J. Q., Chen, Z., Zheng, K. W., Chen, C. Y., Hao, Y. H., Tan, Z., *G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase*, Nucleic Acids Res., (2011). 39(14): p. 6229-6237.

139. Watson, J.D., Crick, F.H.C., Molecular structure of nucleic acids: *A structure for deoxyribose nucleic acid*, Nature, (1953). **171**(4356): **p.** 737–738.

140. Wongsurawat, T., Jenjaroenpun, P., Kwoh, C.K., Kuznetsov, V., *Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity*, Nucleic Acids Res., (2012). **40**: **p.** e16.

141. Woodford, K. J., Howell, R. M., Usdin, K., *A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes*, J. Biol. Chem., (1994). **269**(43): **p.** 27029–27035.

142. Wu, Y., Shin-ya, K., Brosh, R. M. Jr., *FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA To Defend Genomic Stability*, Mol. Cell. Biol., (2008). **28**(12): **p.** 4116-4128.

143. Xiao, S., Zhang, J.y., Zheng, K.w., Hao, Y.h., Tan, Z., *Bioinformatic analysis reveals an evolutional selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals*, Nucleic Acids Res., (2013). **41**(22): **p.** 10379–10390.

144. Xu, B., Clayton, D.A., *RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers*, EMBO J., (1996). **15**(12): **p.** 3135–3143.

145. Yang, D., Hurley, L.H., *Structure of the biologically relevant G-quadruplex in the c-MYC promoter*. Nucleosides Nucleotides Nucleic Acids, (2006). **25**(8): **p.** 951-968.

146. Yang, T. H., Tsai, W. H., Lee, Y. M., Lei, H. Y.,Lai, M. Y., Chen, D. S., Yeh, N. H., Lee, S. C., *Purification and characterization of nucleolin and its identification as a transcription repressor*, Mol. Cell. Biol., (1994). **14**(9): **p.** 6068-6074.

147. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., Lieber, M. R., *R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells*, Nat. Immunol., (2003). **4**(5): **p.** 442–451.

148. Zhang, C., Liu, H.h., Zheng, K.w., Hao, Y.h., Tan, Z., *DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix*, Nucleic Acids Research, (2013). **41**(14): **p.** 7144–7152.

149. Zhang, H., Barceló, J. M., Lee, B., Kohlhagen, G., Zimonjic, D.B., Popescu, N. C., Pommier, Y., *Human mitochondrial topoisomerase I*, Proc. Natl. Acad. Sci. USA, (2001). **98**: **p.** 10608–10613.

150. Zhu, Q., Pongpech, P., DiGate, R.J., *Type I topoisomerase activity is required for proper chromosomal segregation in Escherichia coli*, Proc. Natl. Acad. Sci. USA., (2001). **98**(17): **p.** 9766-9771.