UNIVERSITÀ DEGLI STUDI DI PADOVA

SEDE AMMINISTRATIVA: UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE CHIMICHE

CORSO DI DOTTORATO IN SCIENZE MOLECOLARI
INDIRIZZO FARMACEUTICO
XXXII CICLO

Non Canonical structures within MYC and BCL2 oncogenes: novel targets for gene expression modulation.

Coordinatore: Ch.mo Prof. Leonard J. Prins
Supervisore: Ch.mo Prof. Claudia Sissi

Dottorando: Cristofari Camilla

“Se vuoi aver successo nella vita,
fai della perseveranza il tuo migliore amico,
dell’esperienza il tuo saggio consigliere,
della cautela il tuo fratello maggiore
e della speranza il tuo angelo custode.”

Joseph Addison

TABLE OF CONTENTS

ABSTRACT

Cancer diseases are increasing worldwide and more than 20 million new cancer cases per year are expected by 2025. At these days the treatment of neoplastic forms took advantage of classic approaches, based on chemotherapeutics and radiotherapeutics agents. However they are characterized by numerous limitations as remarkable side effects, toxicity and selection of resistant phenotypes to such therapies. This prompted the development of so-called targeted therapies, where selective chemical entities (small molecules, monoclonal antibodies, miRNAs, siRNAs etc.) hit a single molecular target of the tumor phenotype. Despite, these therapies have proven to be efficient alternatives they also present several limitations that make them quite ineffective. In order to overcome these remarkable drawbacks, he modulation of the gene expression, that exploits the ability of nucleic acids to assume different conformations, defined as non-canonical, became extremely interesting. Among these non-canonical conformations, extremely fascinating are the tetrahelical conformations known as G-quadruplex (G4) and i-Motif (iM), that seem to be involved in the blockage of the cancer development.

G4 structures occur at DNA and RNA sequences presenting a high abundance of consecutive guanines that interact each other through Hoogstein hydrogen bonds to generate a planar structure called G-tetrads. Stacking interaction between two or more G tetrads create the overall structure. Bioinformatics studies revealed that prevalently these regions are contained along the telomeres and within the untranslated region (UTR) or within the promoter sites of several oncogenes (approximately 40%) directly implicated in the development of tumor phenotypes. The UTR domains, as the promoter regions, are double-stranded DNA sequences. Therefore the complementary strand results enriched in cytosine, that under specific environmental conditions can folds into a tetrahelical conformation, known as i-Motif. Unlike the G4s, the building block of the entire structure is a dimer of cytosine mainly stabilized by the presence of three Hoogstein hydrogen bonds. The in vivo formation of G4 and iM leads to a steric hindrance at the DNA level; this suggests an inhibition/activation effect on the elongation process of the telomere or on the gene expression process

Under the supervision of Dr. Laurence J. Hurley, the structural characterization of the cytosine rich sequence contained within the NHE(III)$_1$ region of *MYC* promoter was completed. In particular, was assessing the effect of the loop composition on the stability and folding process of the already characterized iM. Since it was proved that this conformation in vivo is involved in the transcriptional activation, the possibility to target it by using a selected compound (IMC-30) was considered. Furthermore, we took into consideration the possibility to use this compound (IMC-30) as an anticancer

drug by testing its ability to induce the apoptosis process in a cancer cell line in which the selected gene was overexpressed.

Besides the several evidence reported for the tetrahelical conformations assumed by the GC-rich promoter regions, more recently the efforts moved forward to the G-rich tract contained in the untranslated (UTR) domains, both the 5'- and the 3'-UTR, of the primary transcript. Since, they can act as modulators of the translation process. Based on this evidence, in this project, the guanine rich sequences contained in the 5'-UTR region, both at the DNA and RNA levels, of the *BCL2* gene were considered. In particular, the structural characterization study was initially carried out on the minimal sequences (dBcl2_G and rBcl2_G), then the effect exerts by the presence of additional nucleotides on the folding process towards the G-quadruplex was taken into consideration (dBcl2_G + 3 WC, rBcl2_G + 3 WC and rBcl2_48). Additionally, the cytosine rich tract contained on the DNA complementary strand was considered and characterized. Our data have shown that the dBcl2_G and rBcl2_G are able to assume multiple G4 conformations. While, the presence of additional nucleotides strongly modulates their ability to assume the non-canonical conformation. Indeed, we proved that the presence of 3 WC pairing partially prevents the formation of G4 both in the DNA and in the RNA, while the addition of a greater number of bases (rBcl2_48) leads to the formation of a different conformation that competes with the G4 structure. Regarding the cytosine rich string, its conformational equilibria have been taken into consideration both in a mildly acidic environment and in an environment that mimics the physiological condition. Finally, we implemented our work, by screening a library of compounds on each tested sequences in order to find a ligand that selectively recognizes and stabilizes one conformation. From the acquired data it emerged the feasibility to stabilize/induce the iM using the Bisanthrene compound and its derivative Bis 1-8. For the guanine rich sequences, Sanguinarine and Chelerythrine provide the best results on each tested tracts, therefore they cannot be considered selective compounds. Similarly, also the Bisanthrene derivatives recognize and interact with each tested guanine tracts, although with different selectivity.

RIASSUNTO

Oggigiorno una delle “piaghe” che affligge maggiormente la popolazione mondiale è il cancro. Il trattamento di queste forme neoplastiche sfrutta agenti chemioterapici e radioterapici, caratterizzati da numerose limitazioni legate ai notevoli effetti collaterali, alla tossicità e alla selezione di fenotipi resistenti a tali terapie. Ciò ha portato allo sviluppo delle targeted therapy, che sfruttano entità chimiche (small molecules, anticorpi monoclonali, miRNA, siRNA ecc.) selettive per un bersaglio molecolare caratteristico del fenotipo tumorale. Nonostante più mirati anche questi approcci presentano degli effetti collaterali Pertanto la modulazione dell’espressione genica che sfrutta la capacità degli acidi nucleici di assumere differenti conformazioni, definite non canoniche, ha destato sempre più interesse.

Tra le possibili strutture non canoniche di notevole interesse sono le conformazioni tetraelicoidali note come G-quadruplex (G4) e i-Motif (iM). La struttura G4 è propria di sequenze di DNA e RNA contenenti un’elevata abbondanza di guanine consecutive che, mediante legami a idrogeno di tipo Hoogstein, generano delle strutture planari chiamate tetradi. Dall’’impilamento di due o più tetradi si genera la struttura a tetraelica. Poiché il DNA è una doppia elica, il filamento complementare a queste regioni G ricche presenta un’elevata abbondanza di citosine. Anche questi domini in particolari condizioni ambientali, possono generare una conformazione tetraelicoidale, nota come i-Motif. A differenza del G4, il building block dell’intera struttura è un dimero di citosine stabilizzato dalla presenza di tre legami a idrogeno. In vivo l’esistenza di queste conformazioni, genera una sorta d’ingombro sterico a livello del DNA e ciò presuppone un effetto d’inibizione/attivazione del processo di elongazione del telomero o del processo trascrizionale.

Sotto la supervisione del Dott. Laurence J. Hurley, è stata implementata la caratterizzazione strutturale della stringa di citosine contenute nel promotore del gene *MYC*. In seguito un selezionato ligando è stato testato con l’idea di poter modulare il processo di folding/unfolding alla base dell’attivazione trascrizionale. Infine, l’effetto mediato da questo composto sul processo apoptotico è stato preso in considerazione lavorando su una selezionata linea cellulare.

Di notevole interesse sono le regioni GC-ricche contenute nella porzione non tradotta del trascritto primario (mRNA). Sulla base di ciò, in questo progetto, sono state prese in considerazioni, le stringhe di guanina e citosina contenute nella regione del 5’-UTR, sia a livello del DNA sia del RNA, del gene *BCL2*. Inizialmente è stato condotto uno studio di caratterizzazione sulle sequenze minimali dBcl2_G, dBcl2_C e rBcl2_G. In seguito è stato preso in considerazione l’effetto della presenza di nucleotidi adiacenti sul processo di folding verso il G-quadruplex (dBcl2_G + 3WC, rBcl2_G + 3WC e rBcl2_48).

I dati ottenuti dimostrano che le sequenze dBcl2_G e rBcl2_G sono in grado di assumere molteplici conformazioni G4. La presenza di nucleotidi addizionali modula la loro capacità di assumere queste conformazioni. In particolare, la presenza di tre appaiamenti WC impedisce parzialmente la formazione del G4 sia nel DNA, che nel RNA mentre, l'aggiunta di un maggior numero di basi (rBcl2_48) sposta l'equilibrio conformazionale verso una conformazione in forte competizione con il G4. Per la sequenza ricca di citosine, l'equilibrio conformazionale è stato valutato sia in ambiente blandamente acido, che in un ambiente che mima la condizione fisiologica. Infine, poiché negli ultimi anni è stata dimostrata la capacità di alcuni ligandi sintetici/naturali, di spostare gli equilibri conformazionali del DNA, dalla classica forma a doppio filamento, verso queste conformazioni tetraelicoidali, una selezionata libreria di composti è stata, scrinata allo scopo di individuare un ligando in grado di riconoscere e stabilizzare selettivamente una conformazione al pari di un'altra.

1 INTRODUCTION

Accordingly to the global demographic characteristic, more than 20 million new cancer cases per year are expected by 2025, meaning that the health problems due to this type of disease will be expected to increase worldwide. At this day several evidence support the idea that tumorigenesis is a multistep process caused by genetic alterations that can drive the transformation of a healthy cell into a malignant phenotype. Despite the high complexity of this disease in 2000 Hanahan and Weinberg outlined that all of them are characterized by six shared ability named as hallmark of cancer, acquired during the neoplastic development and including invasion, metastasis, cell death resistance angiogenesis replicative immortality and evasion of growth suppression. Each mentioned ability is associated with a particular mutation on a specific gene that results in overexpression or gain of function for oncogenes or a loss of function for tumor suppressor's genes. Among the 50.000-100000 genes that composed our entire genome, hundreds of them was been identified as an oncogenes such as *MYC*, *BCL2*, *RAS, BRAF, EGFR* and other, that overexpression has been associated to different of solid or liquid cancer. Alongside the proto-oncogenes, there are the tumor suppressors that physiologically act as a protecting agent from the cancer transformation supressing the cell proliferation, inducing the apoptosis or by inhibiting metastasis. As an example p53 is the well-studied tumor suppressor, that result misregulated in half of the human cancer forms[1,2].

To suppress the aberrant oncogenes activities anticancer therapy are required. Up today several approaches depending by the localization, the type of cancer and the progression step could be selected as example radiotherapy, chemotherapy surgery, and immunological approach. Despite, these classical approaches are essential for cancer eradication, they present several negative aspects and high number of side effects. Therefore, the development of a selective compound able to target specifically the cancer forms is required. As the cancer development is strictly controlled to a specific DNA mutation that leads the overexpression or the loss of function, targeting the DNA sequences involved in the cancer development may be a significant breakthrough in the anticancer fields. In this specific contest the gene therapy based on antisense-oligonucleotide, ribozyme or miRNA are already used. In this case, the main issue is the delivery of the nucleic acid and the controlled of the immunological response. For this reason, this specific approach is going to be replaced by the use of specific ligands and small molecules able to recognize and stabilize/destabilize specific DNA structure that could be assumed by the nucleic acids in vivo.

1.1. Nucleic acids polymorphism

The DNA, that carries all of the genetic information, is a polymorphic and flexible macromolecule. This peculiar characteristic was observed for the first time by Rosalind

Franklin, who set up a pioneering X-ray diffraction studies, through which she pointed out the existence of two allomorphic forms which she defined as shape A and shape B. Based on this preliminary knowledge in the fifties James Dewey Watson and Francis Crick focused their attention on a single conformation, thus they describe and define the predominant secondary structure assumed in solution, the B-DNA. In this contest, two nucleic acid strands are held together in an antiparallel orientation by hydrogen bonds between the four nitrogen bases following the Watson and Crick rules (AT and GC)[3]. However, over the years several searches highlighted the ability of the DNA to assumes other secondary structures depending by the solution condition i.e. the presence of salt, dehydration, binding with nuclear proteins or ligands, and finally, within peculiar nucleotide sequences[4]. These unusual secondary structures, although not preponderant, are implicated in several biological processes, as the transcription and the translation. Moreover, it has been reported that these structures can also be the triggering cause of human pathologies. Thus, prompted the comprehension of the structural features, as well, the physiological/pathological role exerted by these conformations.

Among them, as an example, we found the A form, likely adopted by the RNA and the Z-form which is a left-handed, unstable double strand, assumed by purine-pyrimidine or pyrimidine-purine dinucleotide repeated sequences (Ravichandran S. 2019). Beside these conformations, which are mainly stabilized by the canonical Watson and Crick bp, different conformations, involving different H-bond as the Hoogstein interactions, or characterized by peculiar building block, can exist. Thus, they are classified as non-canonical. Examples of these structures are the hairpins, the DNA cruciform, the triplex and tetraplex conformation (G-quadruplex and i-Motif) and the parallel double strand[6]. The cruciform DNA, as the stem-loop conformation, is capable of forming from inverted oligonucleotide repeats. This structure consists of a branch point, a stem and a loop, which length is dependent by the gap between the two inverted repeats. Despite it is thermodynamically unstable, several reports revealed its formation in vivo both in prokaryotes and eukaryotes. Moreover, the implication in the development of several diseases as cancer or the Werner's syndrome, was reported[7]. DNA inverted sequence or palindromic region lead also the formation of the hairpin structure (stem-loop) that might be in dynamic equilibrium with the double-strand form. Again this conformation is characterized by the presence of a base-paired stem connected to a small loop in which the bases are unpaired. Although hairpins formed in long palindromes are genetically unstable, is already known that they play a key role in several cellular processes as the gene expression, recombination, and transcription. Triplex conformation, firstly revealed in 1957, by Felsenfeld et coworker[8] are formed both in purine and pyrimidine motifs via Hoogsteen or reverse Hoogsteen hydrogen bonds. This conformation result to be involved in the inhibition of the transcription[9]. Finally, the DNA apart from the usual antiparallel duplex has the ability to form

parallel-stranded duplex (ps-duplex), which is stabilized by reverse Watson–Crick or Hoogsteen hydrogen bond. Despite less stabile it has been known that ps-duplex may play an important role in regulation of transcription, mutational processes and chromosomal folding.

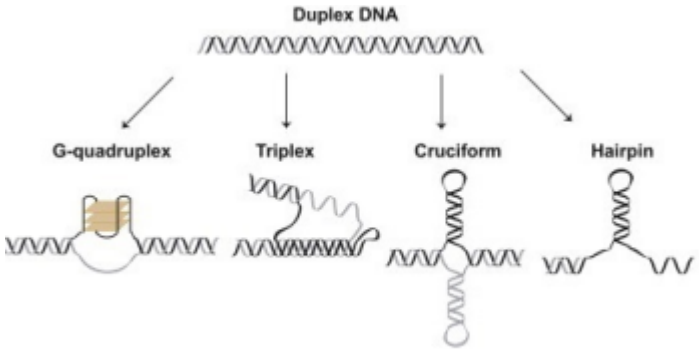


Figure 1.1 Several secondary structures adopted by the nucleic acid.

The RNA molecule is a linear polymer consisting of four nucleotides Adenine (A), Guanine (G), Cytosine (C) and Uracil (U) on a sugar-phosphate backbone. This RNA linear chain in general folds into a three-dimensional structure, both for molecular stability and to perform specific biological functions. Generally, they first fold onto itself forming a double-stranded regions characterized by additional hydrogen bonds. These double-stranded regions that follow the Watson-Crick rules lend energy stability to the molecule.

A hypothetical example of an RNA secondary structure is reported in figure 1.2. The picture illustrates the several basic motifs that arise in the RNA secondary structures. The double stranded regions formed by the stacking of two or consecutive base-pairs are referred to as stems. While the single stranded regions can form a variety of motifs corresponding all to loops. As examples, the hairpin loop is a single-stranded region that ends into a stem. An interior loop can interrupt a stem on either side. Whereas, a region into which more than two stems meet is known as multi-junction loop. All of this single-stranded region in the secondary structure are extremely important since they could serve as potential sites for protein/RNA binding or it may imply in the formation of RNA tertiary structure[10]. Additionally, over the years it has been proved the existence of supramolecular structures as the G4, the triplex and stem-loop, and many more, that can modulate the physiological processes involving the RNA, as well as pathological conditions[11,12]. Therefore the characterization and the understanding of their relevance are necessary.

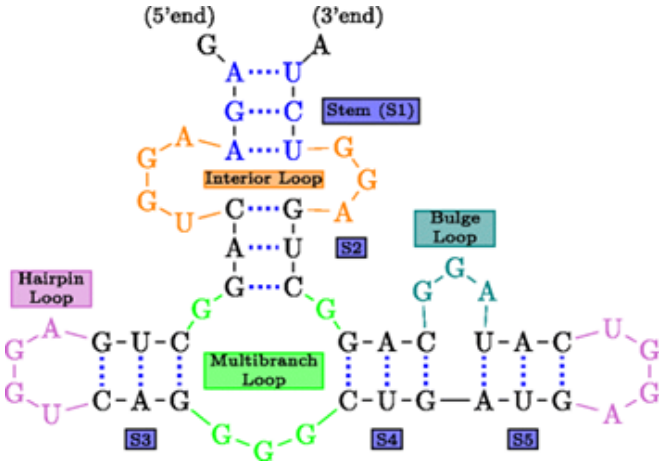


Figure 1.2 Basic Structural Motifs in RNA secondary structures

1.2. The G-quadruplex

The ability of the guanine base to self-assembly into planar square was firstly reported by Gellert et al., when in 1962 they proposed the ability of the guanylic acid to generate a supramolecular structure[13]. However, an increasing interest towards this supramolecular conformation adopted by the guanine bases, started in 1988, when Sen at colleagues proved the existence of a guanine rich tract, able to generate a tetrahelicoidal conformation along the telomeric region[14]. Over the years, multiple "G-rich" sequences were identified in other human genomic portions as the promoter region, along the 5'- or the 3'-UTR regions of several oncogenes as well as on the primary transcript[15]. This in-depth evaluation made possible to obtain a series of information regarding their structural characteristics but also their implication in many biological processes as: transcription and translation, aging, neurodegenerative or cancer progression emerged. Moreover, since their important location along with the nucleic acid, the interest on the G4 as an anticancer target raised.

The G-quadruplex (G4) is a very polymorphic non canonical secondary structure assumed by DNA and RNA guanine rich domain. Differently from the B-form, in which the building blocks are C-G and A-T base pairs, in this case, four guanines interact one each other via Hoogstein hydrogen bonds that involves the N1, N7, O6 and N2, forming a planar square named G-tetrad. Following stacking interaction two or more of these building blocks generate the overall structure (figure 1.3).

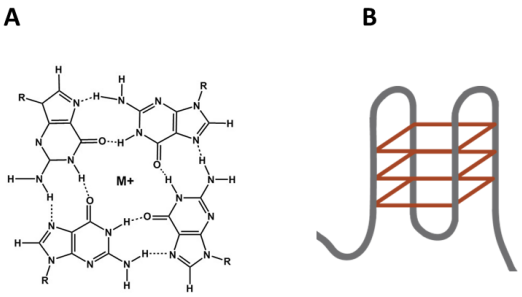


Figure 1.3 Schematic representation of a G-tetrad and G-quadruplex conformation upon tetrads stacking.

As in the case of the other conformations, several factors influence the folding process and the stability of this structure as the saline condition, the oligonucleotide sequence, the number of guanine in each run, the number of strand and their directionality, the angle of the glycosidic bond, the size, composition and type of the connecting loop and finally the presence/absence of dehydrating or crowding agents as well as the binding with nuclear proteins or ligands[16]. This conformation exhibits extensive structural polymorphism and diversity. In general, the polymorphism depends mostly from the nature of loop, strand polarity and stoichiometry, the nature of the glycosidic angle and the location of the loop. Meanwhile, also the presence of cationic species or ligands as well crowding agent may alter the conformation assumed by the guanine rich domain.

The first level of polymorphism is related to the strand stoichiometry, G-quadruplexes, in fact, can be folded from a single guanine rich tract intramolecularly or by the association of two or four separate strands. The polarity of the strand affects the parallel, propeller or antiparallel conformation assumed by the G4. Variation in the strand polarity affect also the location of the loop, parallel G4 requires a double chain reversal loop, while the antiparallel one can be linked by diagonal or latera loop. Moreover, it has been proved that the sequence of the loop highly contribute to the flexibility of the conformation as well as it can be considered a potential site for drug targeting. Regarding the glycosidic angles, this is an important parameter for the G4 topologies. Indeed, in parallel G4 the guanine adopt the *anti* conformation, whereas antiparallel one can adopt both *syn* and *anti* conformation. Another peculiar characteristic is the presence of four grooves, the dimensions of these cavities are variable depending by the topology and by the nature of the loop that connects the single tetrads. As an example for G4 with lateral o diagonal loop, the grooves are structurally simple, while the presence of a propeller loop complicates the structural features of the four grooves[17–20] Alongside these mentioned parameters, the folding and the stability of the G-quadruplex conformation result to be monovalent-cation dependent. This fact can be ascribed to the extremely negative electrostatic potential generated by the presence of eight carbonyl O6 atoms that create a central channel

between the two tetrads. Here the cationic species can be accommodated. The precise localization of the metal species depends by the nature of the cation. The Na^+ generally accommodates in a plane with the G-tetrads, while the K^+ result equidistant from each tetrad plane, thus symmetric tetragonal bipyramidal configuration is formed. Beside the Na^+ and K^+, that are the most relevant physiologically cations, other ions were considered, for example, Thallium (1^+), that presents a comparable ionic radius to the K^+ one, allows the formation of the identical topology promotes by the potassium. On the contrary, Li^+ was proved to have a destabilizing effect on the overall conformation.

By merging all possible factors able to modulate the G4 conformation it appears that different G4 structures occur, thus, this conformation can be compared to a library of multiple tridimensional structures in where each topology can be an interesting target to better understand the biological relevance of the G-quadruplex and, also to gain a new anticancer therapy extremely selective and with fewer side effects.

1.3. The i-Motif

Differently, from the telomeric portion, the guanine rich tract within the promoter region or along the untranslated domain of the oncogenes is paired to its complementary cytosine rich strand. Under specific environmental condition also this region may fold into a non-canonical tetrahelical conformation called i-Motif (iM) (figure 1.4). Particularly, this arrangement, firstly characterized by Gehring and coworkers in 1993[21] is the only known conformation composed by two parallel duplexes held together in an antiparallel manner. Contrarily from the G4 structure where the building block is a planar square of four G, the key interactions for the iM is a dimer of hemi-protonated cytosines ($C:C^+$), in which three Hoogsteen hydrogen bonds characterized by a higher base-pairing energy compared to the canonical Watson Crick GC, are present[22]. Additionally, the distance between two consecutive base pairs and the twist angle of the right-handed helical are both significantly smaller compared to the double-stranded form and finally, since the intercalation of the building block from the two parallel duplexes generates an antiparallel tetrahelical conformation, two wide grooves, and two minor grooves were observed[23].

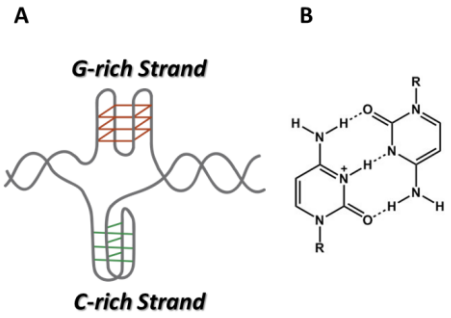


Figure 1.4 (A) Schematic representations of the guanine and relative G-quadruplex structure and cytosine rich tract and the iM conformation along the DNA. (B) Building block of the iM conformation.

The formation of the cytosine dimer, requires the protonation of the nitrogen in position 3, it derives that in general a specific environmental condition in which the pH value is quite close to the pKa of the cytosine (pKa ≈ 4.5), promotes and stabilizes the overall structure.

Based on this specific requirement, the iM firstly found several applications in analytical, diagnostic and material science. Modi and colleagues in 2009 developed and studied a nanomachine, based on i-Motif, to obtain a mapping of the variation in time and space of the cellular pH associated with the maturation of the endosome[24], in other cases, the i-tetraplex folding/unfolding process pH balanced is exploited in order to obtain drug delivery systems (nanoparticles and hydrogels) or biocompatible, biodegradable and non-toxic diagnostics[25–28]. Despite several cytosine rich sequences along the promoters region were already structurally characterized[29–34] until now, no clear indication regarding the existence of this conformation in vivo were known. For this reason, since 2017, the biological relevance of the iM structures has been largely questioned. Waller's and Burrows' groups by using bioinformatic tool highlighted the existence of C-rich sequences with higher potential to forming the iM[35,36]. In particular, they found that these tracts were not randomly located along the genome, in fact, most of them were concentrated along the 5'- and 3'-UTRs, introns, in the coding region and also in the promoters of certain genes directly involved in the gene expression. A confirmation regarding the in vivo existence of the iM was recently stated by an in-cell NMR study performed by Dzatko and colleagues on transfected Hela cells with four distinct i-Motif forming sequences[37]. Overall the NMR spectra acquired indicates that after the introduction of the pre-folded iM, these conformations persist under the complex nuclear environmental cell. In addition to this set of data, Christ's and Dinger's group in 2018 developed an antibody, named as iMab, that was find able to binds with a high affinity and specificity different well-known iM, whereas no binding ability towards double-strand, hairpin and G4 was showed. Also, by immunofluorescence staining of three different cell lines, they

revealed the presence of punctate foci attributed to the recognition of iM inside the nuclei that varied along the cell cycle. Altogether, a clear indication of the in vivo existence of the iM and its relevance in the biological processes is finally proved[38].

Like the other nucleic acid conformations, the stability of this tetrahelical conformation depends by several external factor including low temperature, crowding condition, presence of cations, ligands or proteins, as well the number of involved cytosine, the type of the bases adjacent to the cytosine core and finally the length of the connecting loop[39]. Moreover, as in the case of the quadruplex conformation, the iM can present as an intramolecular structure, when the cytosine bases involved belongs to a single nucleic acid strand, or it can folds into a multimeric conformation since the association of two or four separate strands.

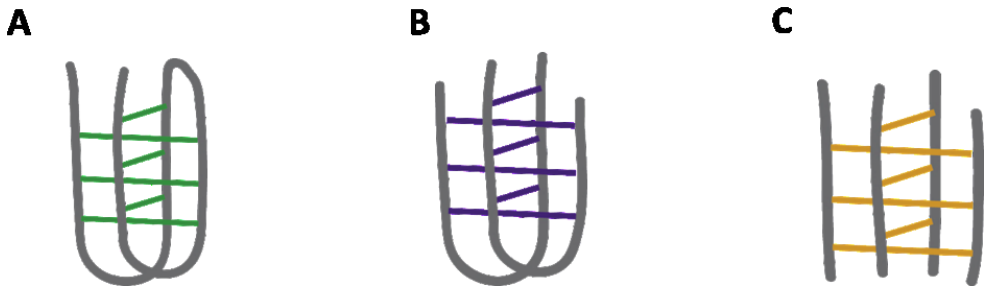


Figure 1.5 (A) Monomolecular (B) bimolecular and (C) tetramolecular iM conformation.

A fundamental parameter for the iM stabilization is the C-tract length. In 2001 it was reported that under the same experimental condition, i-Motif forming sequences possessing a high number of C would result more stable rather than sequences with a shorter C-tract[40]. In line with this preliminary data, Waller's group reported that increasing the C-tract was strictly related to the increment of the transitional pH, as well as the thermal stability. However, more recent studies show that also shorter C-tracts present the tendency to assume a stable iM under neutral condition[41]. Besides the number of cytosines in each tract, one other parameter affecting the stability of the overall structure is the nature of the bases contained into the loop region that connects the C-runs as well the presence of capping residues at the end of the iM[31]. In 2010 Tracy Brooks and coworkers classified the intramolecular iM structure depending by the length of the connecting loop. Class I comprises tetrahelical structures having short loops (2: 3-4: 2), whereas iM with longer loops (6-8: 2-5: 6-7) belong to class II. Generally, iM belonging to the class I favor the formation of dimeric or tetrameric conformation, while longer loops allow the formation of intramolecular structures. Moreover, since the longer loop might be characterized by the presence of additional stabilizing interaction it has been proposed that the class II are more stable compared the class I[42–44]. Nevertheless, recent reports partially disagree with this classification, since it has been observed that short loops allow the formation of stable iM at neutral

or nearly neutral pH[45]. Also the nature of the loops influences the stability of i-Motif. It has been observed that the presence of purine bases in the loops induces conformational changes that lead the destabilization or disruption of the i-motif core, whereas the presence of T or additional C bases confers a higher stability[46–48]. As for the G-quadruplex structures, also the iM can be affected by the nature of cationic species contained in solution. Therefore, several studies focused on the effects exerted by cationic species already extensively studied for their ability to modulate the G4 folding as Na^+, K^+, and Li^+. In particular, Saxena and colleagues pointed out that C-rich sequences in presence of Na^+ ions do not fold into the i-tetraplex structure, although in acidic conditions. Whereas, Mg^{2+} and K^+ in a system that contains the two complementary sequences lead the formation of three conformations in equilibrium each other (duplex, iM, and G4)[49]. Finally, the effect of LiCl was took into consideration, in 2014 Kim et al proved that the presence of the Li^+ leads to a decrease of the i-motif population in solution. Particularly, this effect can be related to the size of the cation, in fact, compared to the sodium or potassium, Li^+ is smaller, thus it presents a higher charge density and a very large hydration shell. This means that despite it is able to insert between a cytosines couple, it cannot mimic the mechanical effects hydrogen-mediated [50]. Alongside these studies, that referred to physiological cations, also the possibilities to obtain the iM foldings under neutral condition following the presence of non-physiological metal species as Ag^+ or Cu^+ was take into consideration[51–54]. As the metal species also the crowding condition and the negative superhelicity can favors the unwinding of the ds-form facilitating the formation the non canonical secondary structures[55,56].

Since the geometry and charge distribution of the iM makes it an attractive model for specific structural recognition of DNA by proteins, the interaction of C-rich DNA strands with proteins has been a matter of study for many years. From these studies, the existence of poly-C-binding proteins, able to specifically interact with the C-rich DNA tract and playing a fundamental role in the regulation of the gene expression was obtained. Among them, of great interest is the family of heterogeneous nuclear ribonucleoproteins (hnRNPs) that comprises proteins able to bind the G4 and/or iM structure assumed by DNA or RNA. One of the best-studied i-motif binding proteins is the heterogeneous nuclear ribonucleoprotein LL (hnRNP LL). This protein, which is a paralog of the hnRNPL a pre-mRNA splicing factor, was found able to interact and recognized the i-Motif conformation formed in the promoter region of *BCL2*. In particular, using two of the four RRM (RNA recognition motif) this protein recognizes and binds the two lateral loops of the *BCL2* iM leading the unfolding of the DNA secondary structure and the transcriptional activation of the gene [57–59]. Similarly to the hnRNP LL, one other RNA-binding protein that plays a critical role in several cellular processes is the hnRNPK. In particular, it has been reported that this protein is

characterized by the presence of three domains, named as KH domains, through which it is able to interact with the C-rich portion contained along the MYC promoter region. Especially, by using two of this three KH domains this protein recognizes and binds the 3' lateral and the central loops of the i-Motif (CCCT domain) assumed by the promoter of c-Myc, leading again the unfolding of this structure and the transcriptional activation. Finally, very recently Niu et al identified a BmILF protein, able to binds the iM structures adopted by the promoter region of BmPOUM2[60].

1.4. G-quadruplex and iM as a therapeutic target

Nowadays several experimental evidences prove the existence of G4 structures in cell, and its multiple roles physiological/pathological processes[15,61–63]. This fact prompted the idea regarding the G4 as a novel drug target that potentially leads to novel strategies having a wide therapeutic potential. In this contest, small molecules selectively able to recognized and bind the G4 scaffold might replace the nucleic acid-based therapy and also the classical anticancer approach based on the chemotherapy and radiotherapy.

Since the first assumption to develop a G4 binder is the ability of the scaffold to discriminate the G4 over the ds-form the investigators initially developed these scaffold by modifying the already known DNA-binding molecules[64]. The first compound identifies as a possible G4 binder was the cationic porphyrin, TMPyP4, that thanks to the planar skeleton and cationic moieties, in 1998 was proved able to bind the G4[65]. After this first report this chemical scaffold, due to its ability to discriminate the tetrahelix over the double helix became one of the best studied compounds. Unfortunately, solubility issue and lower penetration ability associated with the lacking of specific binding in vivo did not provide it to overcome the pre-clinical trials[66,67]. Following the TMPyP4, another compound was found to exhibit telomerase-inhibiting activity by binding the telomeric G4 structures, leading telomere dysfunction, length shortening and delocalization of telomere-related proteins; the Telomestatine, that is a natural derivative isolated from the Streptomyces anulatus[68,69]. Another extremely interesting compound was the Phen-DC(3), that result able to interact through π-stacking with guanine bases of the top G-tetrad with the intramolecular G4 within the promoter region of *MYC* gene[70]. These pioneering works accelerated the development and screened of G4-selective synthetic molecules that according to the structural features can be divided into several categories. Among them great interest received the Anthraquinones and naphthalene diimide scaffold that due the presence of a large a planar region associated to different substituents can interact selectively with the G4 structure both via stacking and electrostatic interaction[71–74]. Other interesting compounds are the phenanthroline derivatives that in 2010 were patented since they showed a telomerase inhibition and a strong selectivity for the G4 over the duplex

form. Also, very recently a 2,6-di-(4-carbamoyl-2-quinolyl)pyridine derivative was been evaluated as G4 binder. In particular, It has proved to be selective for the G4 over duplex with a preferential selection for the parallel conformation[75]. It is worthy of note that additionally to the stacking over the external tetrads, other binding modes have been proposed over the years including the grooves interaction[76], the electrostatic interaction with the loops and the binding to the negatively charged phosphate groups along the backbone. Among these additional binding sites, the best attractive strategy appears to be the grooves targeting. Indeed, as expected the Quadruplex grooves are extremely different from the double-strand ones, both in terms of shape and dimension. Moreover, the high structural polymorphism of the G4 implies the possibilities to obtained G4 binders that selectively recognize precise topologies compared to the other and also compared to the ds-DNA.

Compared to the numerous and well-documented examples of G4 ligands[77–81], the discovery of putative i-Motif ligands lags far behind. In this contest, the first study was reported in 2000, when Hurley and colleagues investigated the interaction of TMPyP4 to the tetramolecular i-motif structure assumed by the human telomeric sequence. Particularly, the compound was found able to bind and promote the iM conformation under acidic condition (pH 4.5). Modelling NMR experiment suggested a non-intercalative binding mode, further confirmed by Gargallo et colleagues, which investigated the selected compound also on the BCL2 iM revealing the interaction between the 2 components and the disruption of the iM at slightly acidic condition. (pH 6.1)[82–85]. After this first reports different G4 ligands were screened on this i-tetraplex as example bisacridine compound[86], phenanthroline derivatives[87,88] and metal complex[89,90]. Despite some of the recorded data were quite encouraging, all of these molecules lack structures specificity and somehow lead to a slight destabilization of the iM. The first selective iM ligand reported was the Carboxyl-modified single walled carbon nanotube (SWNT), since they increase the thermal stability of the telomeric C-rich sequence and inhibit the telomerase activity both in vitro and in vivo[91–94]. Interesting compounds are also the Graphene quantum dot (GQD) that were shown to be able to promote the iM under neutral condition (pH 8)[95]. Another successful study for the identification of iM binding ligands was set up by Hurley and colleagues. In particular, among 1990 screened compounds, they revealed the ability of IMC-48 to selectively bind and stabilize the iM assumed by the promoter of BCL2, while in parallel the IMC-76 scaffold was proved able to interact with the same iM conformation, leading the conformational switch towards the hairpin conformation. Since both of them are able to interact with the iM contained into the promoter region of BCL2, it appears that they can be used to modulate the dynamic equilibria of the promotorial features, leading the transcriptional activation or repression. An idea further confirmed by cellular studies[96]. The same group in 2017, identified the nitidine

as selective compound able to disrupt the hairpin in the hybrid hairpin/iM within the k-ras promoter, downregulating the gene expression, while benzothiophenecarboxamide compound was proved able to selectively binds the PDGFR-B i-Motif, leading again to the downregulation of the promoter activity [97,98].

1.5. c-*MYC* and its promoter region

c-Myc protooncogene encodes for a 65 kDa helix-loop-helix leucine zipper transcription factor that, after a dimerization step with MAX, activates the transcription of genes that contain the E-box (enhancer box)[99,100]. This protein plays a fundamental role in a multitude of physiological processes i.e. cell growth and proliferation, differentiation and programmed cell death. Also, it is widespread upregulated in multiple human malignancies, such as gastric, colon, breast, and cervical cancers, osteosarcomas, lymphomas, and leukemia as a result of gene amplification, chromosomal translocation, retroviral transduction or proviral insertion[101–103].

The mechanisms that govern the gene transcription are extremely complex and involve multiple elements known as cis-regulatory elements (CRE), Far Upstream Element (FUSE) and Nuclear hypersensitivity element (NHE)III_1 able to interact with several DNA binding protein,[104,105] as well as multiple promoters and transcriptional start sites[106]. Notably, the (NHE)III_1 element (CT element) accounts the 85-90 % of the total MYC transcription. Initially identified as a major site of the DNAse I hypersensitivity, this region located -142 to -115 base pair upstream the P1 promoter, and corresponding to bases 2186-2212 in the c-myc locus, presents an unusual strand asymmetry: one strand is almost a perfect homopurine filament (noncoding strand) whereas, the complementary one is a homopyrimidine tract (coding strand). This region appears able to engage a slow equilibrium between the ds-form and stabile paranemic structures as the G4 and the iM for which was proposed a role in the regulation of the entire expression pathway[107]. Additionally, many transcription factors as hnRNP A1, hnRNPK, Sp1, and NM23-H2 has proved able to recognized and bind selectively these purine or pyrimidine rich-tract as a single strand or as a non B-DNA form.

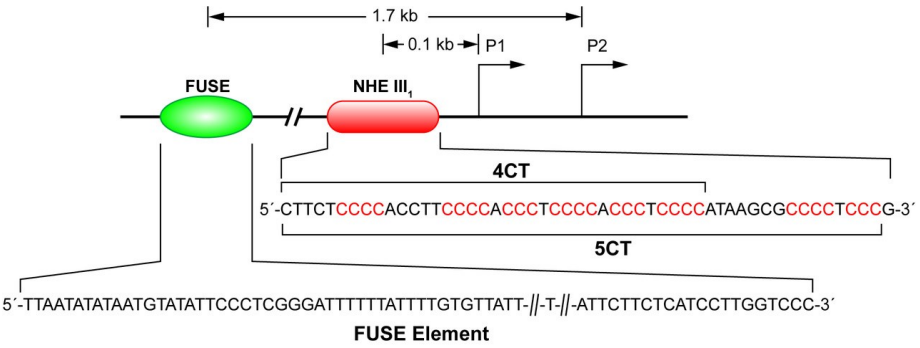


Figure 1.6 Representation of the Far Upstream Element (FUSE) and Nuclear hypersensitivity element (NHE)III_1.

The pioneer in the structural characterization of this domain was Simonsson, that in 1998 proved the ability of the guanine rich domain to assume a three tetrads intramolecular G4 structure[108]. Lately, several groups focused their attention on this peculiar guanine rich tract, in 2002 Siddiqui-Jain et al revealed an equilibrium between two different G4, a two tetrads chair G4 with 2-nts lateral loop and one orthogonal 3-nts bridging loop and the previously mentioned three tetrads basket G4 with two lateral two-base loops and an orthogonal three nucleotide bridging Despite, the tree tetrads result to be thermodynamically favored, mutational analysis proved that the chair form is the biological relevant G4. Indeed mutation or disruption of this conformation caused an increment of the transcriptional activity[83]. Additional mutational studies, as well as NMR assay, revealed the ability of this guanine rich tract to assume a parallel G4 conformation composed by three tetrad producing four isomers with a loop sizes of5'-(1:2:1)-3', 5'-(2:1:1)-3', 5'-(1:1:2)-3' and 5'-(2:1:2)-3. However, the predominant loop isomer adopted is the 5'-(1:2:1)-3', in which the four 5' guanine runs are utilized [42,109,110].

Since this conformation was proved to be an efficient repressor element for the transcription process, the identification of physiological proteins able to modulate this folding, as well as the possibility to target it specifically via small molecules received a lot of interest.

An interesting protein partner in the expression of this gene is Nucleolin, a 100 kDa phosphoprotein involved in several important processes as ribosomal biogenesis, transcriptional regulation, chromatin decondensation, cell proliferation, and differentiation. In 1994 the LR1 nucleolin-hnRNPD heterodimer was reported able to regulate the transcription of *MYC* in a B cell lymphoma by interacting with the $NHE(III)_1$ element. Whereas, in 1999 the specific interaction of Nucleolin with DNA G4 conformation, was reported. A further demonstration that Nucleolin act as a *MYC* G4 binding protein was stated in 2009 by Gonzales and colleagues. By applying in vitro an in vivo experiments they proved the binding of this protein on the G-quadruplex form adopt by this promoter, moreover, a transcriptional down regulation promoted by this interaction was observed[111–113].

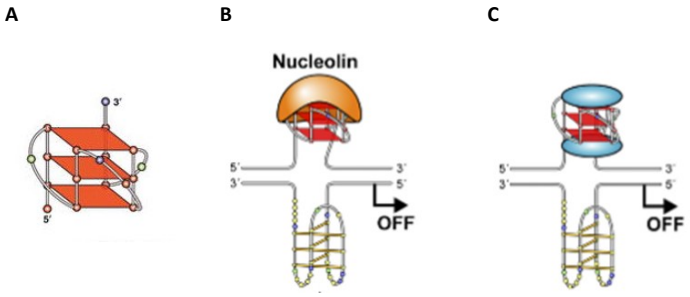


Figure 1.7 (A) Parallel G4 conformation adopts by the purine rich strand within MYC NHE(III)$_1$ element. (B) Proposed interaction of Nucleolin protein on the G4. (C) Proposed binding mode of GQC-05 compound. Regarding the possibility to target this G4 form, the first tested small molecule was the cationic porphyrin TMPyP4. This compound was proved able to lower the transcription and the protein expression of MYC. However, an implication in other promoter systems as *BCL2*, *hTERT*, and *VEGF* as well as, solubility issue, lower penetration ability, associated with the lacking of specific binding in vivo made it a non-specific compound for this gene. A more specific ligand was discovered in 2011 by Brown and colleagues. This ellipticine analog GQC-05 (NSC338258) was proved able to selectively recognize and binds the *MYC* G4 structure compared to dsDNA or to other tetrahelical conformation. Moreover, by applying ChiP assay the molecular mechanism of this compound was discovered. Especially, it was observed that the binding of this compound on the G4 structure alters the recruitment and the binding of specific protein actively involved in the transcriptional activation as CNPB and hnRNPK. Therefore, a downregulation of the gene expression was observed [Brown et al]. Alongside this compound, in 2016 a quinazoline derivative, able to affects the telomere by interacting with the G4, was investigated as a putative MYC G4 binder, both in vitro and vivo. Biophysical assay revealed the ability of this compound to selectively bind and stabilized the G4 conformation within *MYC* promoter promoting a decrement of the gene expression and antitumor effects[114].

Opposite to the guanine rich strand, the coding region contains eight runs of cytosine possibly involved in the formation of the iM. The existence of this conformation was firstly provided by Simonnson et al in 2000[115], later NMR study validated the folding towards the iM of the cytosine rich stretch suggesting also the ability of this cytosine rich tract to adopts at least two different intramolecular iM[116]. In 2009 the effect of the negative supercoiling on the iM folding was taken into consideration, in this specific condition Sun and Hurley proposed the ability of this sequence to adopts an intramolecular iM with a loop conformation of 2:6:2 where the first five 5' runs of cytosines are involved[55]. However, in 2016, Sutherland and colleagues by CD spectroscopy and footprinting assay revealed the formation of a unique iM with a loop conformation of 5:5:5. Furthermore, they revealed that as in the case of hINS and

hTERT[117,118], the formation of the two tetrahelical conformations results mutually exclusive. Confirming the importance of both conformations and their different implication on the transcriptional regulation.

As the G-rich counterpart, also the iM could be recognized by several proteins. One such protein is the heterogeneous ribonucleoprotein K (hnRNPK), an RNA-binding protein that plays a critical role in several cellular processes, including the regulation of cell cycle progression, cell growth, and differentiation, transformation, angiogenesis, and apoptosis. In early studies, hnRNPK was proved able to binds the pyrimidine-rich strand contained in *MYC* promoter, promoting the transcriptional activation[119]. Later studies revealed the presence of a conserved motif in all RNA-binding protein, repeated three times within the primary sequence of this protein, and in 1995 Takimoto and colleagues showed that the binding with the protein required at least two CT (CCCT) repeats separated by at least three nucleotides, suggesting that the interaction involves the KH1 and the KH2 domains. Furthermore, they proved that the binding of the KH1 domain, located at the N-terminus of the protein is necessary for the transactivation of the gene[120,121]. The involvement of the last KH domain on the DNA-protein complex was demonstrated in 2002 by Braddock et coworkers. This last domain located at C-terminal end was proved able to recognize and bind the CCCT element contained along the NHE(III)1 region. Moreover, they observed that the fifth amino acidic residue contained in each KH domain plays a crucial role in the discrimination of the first two bases of the CCCT tetrad[121].
In analogy to the observed data for the promoter region of Bcl2, the consensus sites for the hnRNPK are contained within the iM loop are inserted in a rigid conformation. This idea was further confirmed by Sutherland et al. Using both in vivo and in vitro techniques, they revealed that hnRNPK recognize the c-myc iM and bind the central loop and the 3'-lateral loop using the KH1 and the KH2 domains. They also revealed that the third KH domain interacts with the fifth CT element located at the 3' end of the C-rich sequence, leading the unfolding of the DNA secondary structure and the transcriptional activation of the gene[121].
In this contest, it emerged the possibility to target the iM via small molecules able to stabilize it and alter the recruitment of this protein. Based on this, very recently two papers were published regarding the possibility to stabilized selectively the iM and/or inhibit the interaction iM-hnRNPK. Shu and coworkers in 2018 proved the ability of an acridone derivative to induce and stabilize the iM as well as its ability to downregulate the transcriptional process. Moreover, by testing the compound on a SiHa cell line they revealed its ability to inhibit the cell proliferation by inducing the apoptosis[122]. Later this group focused their efforts on the possibility to inhibit the formation of the DNA-protein complex by sequestering the hnRNPK. In 2019 they reported the ability of a quindoline derivative to bind tightly the protein disrupting the interaction with the c-

myc iM and downregulate the gene expression. Furthermore, they observed the ability of this compound to inhibit the proliferation of various cancer cell line, to induce the apoptosis and inhibit the long-term proliferation, migration, and invasion of tumor cells[123].

1.6. *BCL2*: an important partner in the apoptosis process.

The apoptosis, also known as programmed cell death, is a fundamental process discovered in the early 1970s that allows the removal of aged, useless and damaged cells in order to maintain the homeostasis and the tissue functions. Moreover, it plays an important roles in controlling physiological pathways as development, immune response as well as diseases progress as neurodegenerative syndromes or cancer[124,125]. This process physiologically is tightly controlled by two major pathways: the extrinsic pathway (death receptor pathway), activated by a ligand-receptor interactions and the intrinsic one, promoted by intracellular signaling and controlled by a heterogeneous family of protein known as B cell lymphoma 2 protein family. This proteins depending by their pro- or anti-apoptotic activity can classified into two subfamilies[126,127]. The founder of this family is the Bcl2 protein, a 239 amino acid protein confined in the inner mitochondrial membrane, whose function is to inhibit the process. This structure was firstly identified in 1986 as an oncogene in human follicular B cell lymphomas, characterized by translocation t(14;18)[128] and overexpressed in several human tumors, i.e. prostate, breast or colorectal cancer. With such an important role, the comprehension of the mechanism that regulates its expression and its gene structure was demand.

The encoding gene for this protein is a large gene composed by the three exons (≈200 kb), a facultative 220 bp introns I and a 370 kb intron II[129]. The transcriptional process results under the control of two promoters (P1 and P2). The P1 promoter is located in the first exon near the nuclease-hypersensitive element (NHE) and commence the 95% of the transcription, whereas the P2 is located at ≈ 1.3 kb downstream the P1 and regulates only the 5% of transcript[130]. Since the discovery of the G4 structures, computational searches scanned the entire genome to found where these peculiar domains are localized. From this bioinformatic research the presence of multiple G-rich sequence along the promoter region of *BCL2* emerged (figure 1.5).

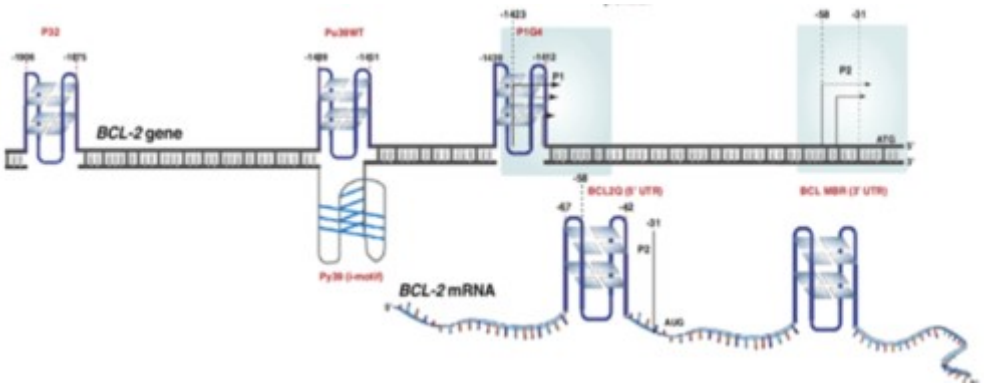


Figure 1.8 Representation of the guanine rich domain within BCL2

Pu39 sequence is located at 58-19 bp upstream the P1 promoter, it contains six runs of two or more contiguous guanines and present the propensity to form three individual intramolecular G-quadruplex structures. Trough systematic mutagenesis and NMR experiment each conformation were isolated and characterized and the pivotal conformer was identified as BCL2MidG4 that folds into a mixed parallel/antiparallel G-quadruplex with two lateral loops and a single nucleotide propeller loop. Since three tetrahelical conformations are present along this region it appears that selectively target one of them may alter the destiny of the transcription. Dexheimer and colleagues focused their effort on the possibility to target them via different small molecules (TMPyP4, Telomestatine and se2SAP). Surprisingly, their results revealed that two of these compounds (TMPyP4 and Se2SAP) showed a selectiveness for the different G-quadruplexes, whereas Telomestatin showed the ability to interact quite strongly with all sequences[131–133]. Similarly in 2010 Wang and colleagues took into consideration the effect of Quindoline derivatives as a turning off the transcription process. Particularly, after having proved that the disruption of this G-rich sequence lead to an increment of the transcriptional activation, the stabilizing effect of the quindoline compounds were explored, revealing their activity both as a repressor for the promoter activity and activators for the apoptosis process[134]. Also a pyridostatin analog was taken into account as a putative G-quadruplex binder. In 2016 it has been reported that PDF compound, as the quindoline derivatives, exhibits a high specificity and stabilizing effect on the G4, as well as interesting suppressor activity on the transcription process and a promoting effect on the apoptosis[135]. In 2014 Agrawal and coworkers focused their effort on a 30-nts sequence, involving four nonsuccessive guanine runs within Pu39 (1245G4). CD spectroscopy and NMR demonstrated that this sequence adopts a stable (Tm 71 °C) intramolecular parallel topology with three chain-reversal loops of 1, 13, and 1 nt. It is highly intriguing that the two stable intramolecular G4 structures formed in the BCL2 promoter, 1245G4 andMidG4, assume completely different folding structures moreover they can exist simultaneously[136]. The coexistence of this two interchangeable G-quadruplexes

envisage a new mechanistic of regulation since each G-quadruplex is likely to be recognized by different proteins leading to different gene modulation. Additionally, they may be act as different scaffold for small molecules. In 2019 the mechanical stability of these two G4s and the folding/unfolding kinetic was assessed by Cheng and coworkers. They found that although kinetically favored the hybrid topology (Pu39) result less stable compared to the parallel 1245G4 form. Therefore, this second topology may act as a kinetic barrier for the transcription process. Furthermore, they determined the existence of additional stable DNA secondary structure, the hybrid stranded Bcl2-1234 G4[137].

Another interesting tract is P1G4. In 2016 Onoel et colleagues reported the ability of this 28-nts sequence, immediately located upstream the P1 promoter, to assume an intramolecular G4 folding. This structure that consists in five runs of three guanines generates two three tetrads parallel topology in highly dynamic equilibrium each other and characterized by the presence of longer middle loop (12 nts or 11 nts) able to generate a hairpin structure involving the classical WC base pairs. This structure that was proved to be extremely stable, fold independently from the Pu39 G4. Moreover, it emerged that compared to the Pu39 G4, P1G4 has the major role in the repressing of transcriptional activity[138].

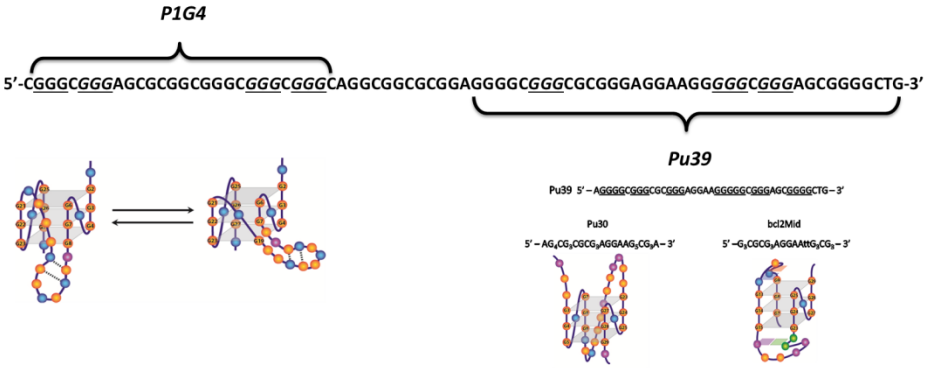


Figure 1.9 Schematic representation of the resolved G4 within the promoter region of Bcl2.

Another putative G4 forming sequence is P32 that resides 451 bp upstream the P1 promoter. P32 forms a high thermal stable intramolecular hybrid-type, composed by four G-tetrads. Differently from the previously G4s cited this folding appear involves in the transcriptional activation of the gene. However a detailed structure and its real relevance on the BCL2 expression are yet to be explored[139].

In physiological condition the guanine rich tract interacts stabily with its complimentary pyrimidine sequence (C-rich). This suggests that for each quadruplex forming sequence exist an i-M forming sequence. Up today, experimental data regarding this tetraplex structure are available only for the cytosine-rich tract opposite to Pu39 (Py39). In 2009 Hurley's group focused their effort on this cytosine rich tract; they revealed the ability of this sequence to generate a stable intramolecular i-Motif

with a 8:5:7 loop folding pattern and having a high transitional pH (pH 6.6) likely supported by the presence of a capping structure involving the three loops of the iM (figure 1.7)[30]. Lately, through NMR assay, the highly dynamics of this conformation, that coexists in equilibrium with a flexible hairpin form was observed. Moreover two distinct compounds were proved able to interact specifically with these conformations leading the redistribution of them in solution and two opposite effect on the gene expression. The cholestane derivative IMC-48 after a binding interaction on the central loop of the iM improve its thermal stability, shift the equilibrium towards it and lead the upregulation of the Bcl2. On the contrary the preganol derivative (IMC-76) well recognizes the hairpin conformation and shifts the dynamic equilibria towards this conformation inducing the repression of the *BCL2*. Since the antagonist effect of this two molecules was observed in vitro as well in cellular system, Hurley's group further assessed the presence of a transcription factor able to modulate this peculiar equilibrium[96]. The hnRNP LL transcription factor, belonging to the same family as hn RNPK, was identify able to recognize and stabilize the iM through specific domains known as RRM at the two lateral loop, leading the unfolding of the sequence and transcriptional activation. Since the two mentioned compound were able to modulate the iM/hairpin equilibria, appears clear that each considered partner may alter the conformational equilibria[57].

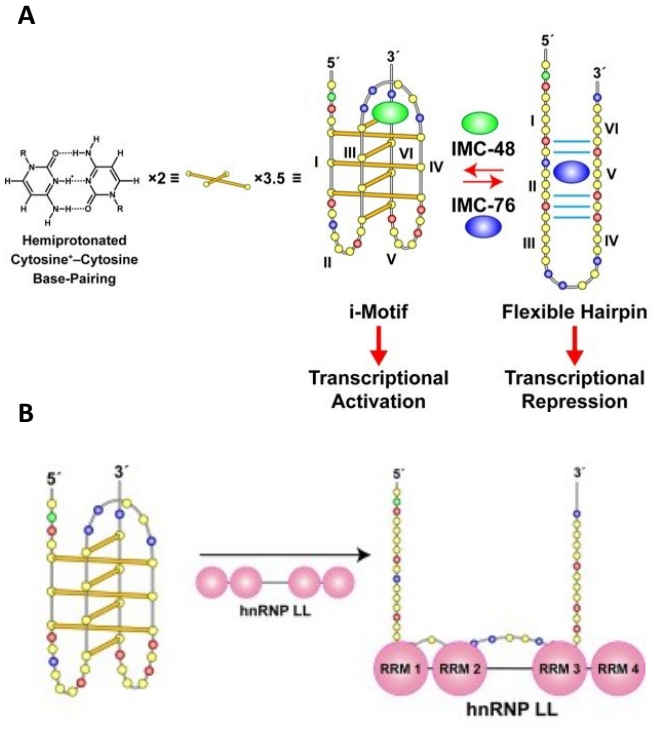


Figure 1.10 (A) Effect of the IMC-48 and IMC-76 on the dynamic equilibrium and on transcriptional process of *BCL2*. (B) Proposed interaction between BCL2 iM and the hnRNPLL protein.

Alongside the G4 and iM forming sequences contained along the promoter region another extremely interesting domain is the 5'-untranslated domain of *BCL2* transcript, since it a reservoir of multiple non canonical structures element. One of these elements is the highly conserved 25-nucleotides PQS, located at 42 bases upstream the translation start site (BCL2Q). Despite less is known compared to the promoter G4s, Zhang et colleagues, as well as, Balasubramanian and coworkers, proved the aptitude of this sequence to assume a parallel fairly stable G4 conformation in absence of stabilizing factor, whereas the thermal stability improves in presence of KCl. Additionally to the structural characterization, also the potential role in vivo was explored, Bugaut and coworker stated the reduction of the in vitro translation and the abrogation of the gene expression both in cell-free lysate and human cell lines [140–143]. Finally, also the 3'-UTR is populated by conserved non B-DNA which are involve both in the translational machinery process, as well as in the polyadenylation and mRNA stability[144].

2 AIM OF THE WORK

Over the years, in silico prediction disclosed the existence of several G4/iM forming sequences upstream the trascriptional start site or within the untranslated region of several oncogenes. Among them, *MYC* and *BCL2* are two of the most studied genes due to their involvement in cancer progression.

Based on the multiple physiological roles, i.e. cell growth and proliferation, differentiation as well as apoptosis, associated to its implication in a wide range of human malignancies, *MYC* gene became an interesting and important candidate for the anticancer therapy. Nowadays several approaches targeting the protein are available; however, they present several limitations. In the last decades the possibility to regulate directly the expression of this gene raised in interest, therefore several elements involves in the regulation of the gene expression were very well characterized. In particular one of the most interesting is the polyguanine/polycytosine region contained within the $NHE(III)_1$ element. Recently it has been proved the importance of the iM conformation assumed by the polycytosine tract, on the transcriptional activation of this gene, following the interaction with a precise transcription factor (hnRNPK). For this reason, small molecules able to stabilize this conformation, avoiding the unfolding of the iM promoted by the transcription factor, could be a new and selective approach for anticancer therapy. In this contest, we explored the impact of the loops composition on the stability of the iM structure previously characterized and published. Subsequently, we investigated the feasibility to target it with a small molecule, 5-[(benzylamino)methyl]quinolin-8-ol compound (IMC-30), selected by a FRET screening assay from NCI library compounds. In this contest were assessed the effect on the iM stability, and the localization of binding site for this chemical entity. Moreover, since *MYC* gene is involved in the regulation of the apoptosis, we explored the possibility to modulate this process in a B-cell lymphoma cell line (U2932) that simultaneously overexpress *MYC* and *BCL2*. Finally, a drug combination target was taken into account, with the purpose to evaluate the possibility to hit simultaneously both overexpressed genes.

In the last decades, the in silico prediction was also expanded on the untranslated domain of the oncogenes, and the existence of putative G-quadruplex forming sequences in the UTR domain of extremely important oncogenes was reported. Particularly, the presence of a highly conserved PQS sequence along the 5'-untranslated domain of BCL2 gene was observed (BCL2Q). Differently, from the well-studied guanine and cytosine rich sequences along the promoter of this gene, less information are reported for this 25-nts RNA sequence. In 2010 Zhang and coworkers stated the ability of a truncated version (22nts) of this sequence to assume a parallel G4 conformation fairly stable in absence of G4 promoting cation. Similarly, Bugaut and

colleagues proved that BCL2Q assume a stable parallel G4 conformation, also they demonstrated the inhibition of the in vitro translation following the stabilization of this G4 conformation[141,142]. The presence of this guanine rich tract on the primary transcripts (mRNA) implies the presence of the same sequence within the 5-'untranslated DNA region (figure 2.2).

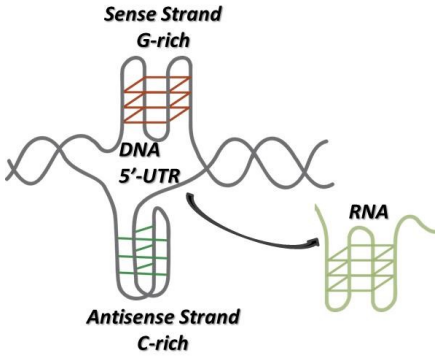


Figure 2.1 Schematic representation of the 5'-Untranslated region of *BCL2.*

Particularly, this G-rich tract will be localized in the Sense strand of DNA, therefore, the antisense strand will be enriched in cytosine. A 22nts sequence was taken into account by Zhang and colleagues. They revealed the ability of this DNA guanine rich sequence to generate different G4 conformation depending by the working condition[145]. However, no indication regarding the ability of the DNA sequence corresponding to BCL2Q to assume or not the G4 are yet reported.
Recent evidences report the possibility to modulate the recruitment of the Polymerase II enzyme, as well as, the scanning step of the ribosome through these non-canonical conformations. Therefore, the structural characterization of these tracts is demanded. Since the literature lack of information about the structural characterization of the 25-nts DNA guanine and cytosine rich sequences contained in the DNA 5'-UTR region, the first purpose of this project was to assess the conformational equilibrium and the in vitro ability to assume the G4 and iM conformation of these two sequences.. Moreover, we explored the balance of the G4 assumed by BCL2Q, with the aim to derive a comparison of the differences existing between the G4s assumed by the DNA and the RNA 5'-UTR region.

dBcl2_G: 5'-GGG GGC CGT **GGG GTG GG**A GCT **GGG G**-3'
dBcl2_C: 5'-**CCC C**AG CT**C CC**A **CCC C**AC GG**C CCC C**-3'
rBcl2_G: 5'-**GGG GG**C CGU **GGG GUG GG**A GCU **GGG G**-3'

Physiologically the 25-nts selected sequences are contained in a longer gene portion, and it is well known that the presence of additional bases can modulate the ability to assume or not the non canonical conformation of selected tract as well as the

possibility to shifts the conformational equilibrium towards a different structure or towards a single species. Therefore, we extended our study considering the conformational equilibrium in solution of three longer sequences containing the 25-nts G-rich sequence (dBcl2_G or rBcl2_G).

dBcl2_G+3WC: 5'-CTC T**GG GGG CCG TGG GGT GGG AGC TGG GG**C GAG-3'
rBcl2 + 3WC: 5'-CUC U**GG GGG CCG UGG GGU GGG AGC UGG GG**C GAG-3'
rBcl2_48: 5'-U**GG GGG CCG UGG GGU GGG AGC UGG GG**C GAG AGG UGC CGU UGG CCC CCG-3'

The extremely important role and the implication of this gene in several cancer forms made BCL2 a promising anticancer target. To date, chemotherapy, radiotherapy and interfering strategies that includes peptides and peptidomimetics, small molecule inhibitor and antisense oligonucleotide are the most used approaches. However all of these strategies are partially successful due to remarkable side effects that make them quite vain. Therefore, an alternative approach targeting the non canonical conformations assumed by this gene is highly desirable. From this point of view after the structural characterization step, we explored the feasibility to target selectively the iM and the G4 structures by performing a preliminary screening assay with selected compounds already known for their ability to recognize the G4 structure (figure 2.2)

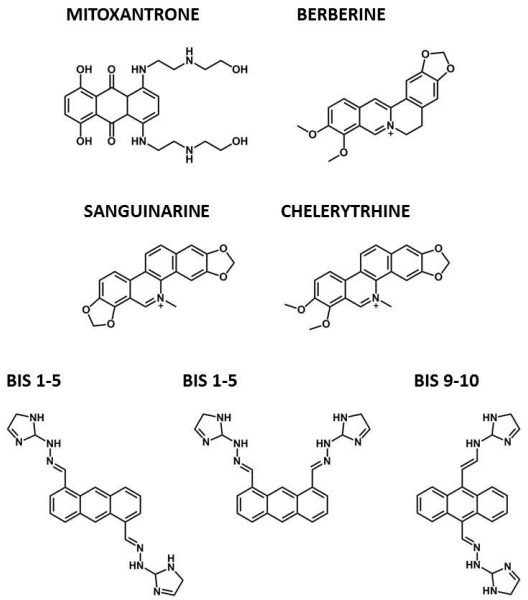


Figure 2.2 chemical structure of the ligand selected for the preliminary screening assay on Bcl2 sequences.

3 MATERIALS METHODS

3.1. Materials

Oligonucleotides

All deoxyriboligonucleotide sequences were purchased lyophilized from Eurogentec (Belgium) and stock solutions of 1 mM or 100 µM were prepared in MilliQ water. Before use, all DNA solutions were diluted in 10 mM Tris pH 7.4 buffer, heated for 10 mins at 95 °C and slowly cool down.

dBcl2_G: 5'-GGG GGC CGT GGG GTG GGA GCT GGG G-3'
dBcl2_G+3WC: 5'-CTC T**GG GGG CCG TGG GGT GGG AGC TGG GG**C GAG-3'
dBcl2_C: 5'-CCC CAG CTC CCA CCC CAC GGC CCC C-3'
EGFR-272_C: 5'-CCC AGC ACT GCC CCT CTG GAC CCG GTC CCC-3'
EGFR-37_C: 5'-CCC TCC TCC TCC CGC CCT GCC TCC CC -3'
Scramble: 5'-GGA TGT GAG TGT GAG TGT GAG G-3'
Coscramble: 5'-CCT CAC ACT CAC ACT CAC ATC C-3'
Coscramble-CC: 5'-TCA CAC TCA CAC TCA CAT-3'

All ribonucleotide sequences were purchased lyophilized from IDT () as standard desalting purified products. Stock solutions of 100 µM were prepared in DEPC water. Before use, all sequences were diluted in 10 mM Tris pH 7.4 buffer, heated at 95 °C for 10 mins and slowly cool down.

rBcl2_G: 5'-GGG GGC CGU GGG GUG GGA GCU GGG G-3'
rBcl2 + 3WC: 5'-CUC U**GG GGG CCG UGG GGU GGG AGC UGG GG**C GAG-3'
rBcl2_48: 5'-U**GG GGG CCG UGG GGU GGG AGC UGG GG**C GAG AGG UGC CGU UGG CCC CCG-3'

The cytosine rich sequences used for the fluorescence melting assay were labelled with FAM at 3' end and Dabcyl at 5' termini. All oligonucleotides were purchased lyophilized from Eurogentec (Belgium) and resuspended in MilliQ water as a stock solution of 100 µM.

For *MYC* sequences, all oligonucleotides were purchased lyophilized from Eurofins Scientific. Standard solutions were prepared at 100 µM concentration in MilliQ water. Before use, each sequence were diluted in the required buffer (10 mM Na-Cacodylate pH 6.0 or 6.7) heated for 60 s at 95 °C and the slowly cool down.

4CT: 5'-TCC CCA CCT TCC CCA CCC TCC CCA CCC TCC CCA-3
5CT: 5'-TCC CCA CCT TCC CCA CCC TCC CCA CCC TCC CCA TAA GCG CCC CTC CCG-3
4CT L1M1: 5'-TCC CCA TCT TCC CCA CCC TCC CCA CCC TCC CCA-3'
4CT L1M2: 5'-TCC CCA CTT TCC CCA CCC TCC CCA CCC TCC CCA-3'

4CT L2M1: 5’-TCC CCA CCT TCC CCA TCC TCC CCA CCC TCC CCA-3’
4CT L2M2: 5’-TCC CCA CCT TCC CCA CTC TCC CCA CCC TCC CCA-3’
4CT L3M1: 5’-TCC CCA CCT TCC CCA CCC TCC CCA TCC TCC CCA-3
4CT L3M2: 5’-TCC CCA CCT TCC CCA CCC TCC CCA CTC TCC CCA-3’
4CT A_6G: 5’-TCC CCG CCT TCC CCA CCC TCC CCA CCC TCC CCA-3’
4CT T_9C: 5’-TCC CCA CCC TCC CCA CCC TCC CCA CCC TCC CCA-3’
4CT T_{10}C: 5’-TCC CCA CCT CCC CCA CCC TCC CCA CCC TCC CCA-3’
4CT A_6T/T_9A: 5’-TCC CCT CCA TCC CCA CCC TCC CCA CCC TCC CCA-3’
4CT A_6T/T_{10}A: 5’-TCC CCT CCT ACC CCA CCC TCC CCA CCC TCC CCA-3’

Compounds

- Mitoxantrone and Anthracene derivatives were synthesized and purified in Prof. Giuseppe Zagotto’s laboratory the University of Padova.
- Chelerythrine, Sanguinarine and Berberine were purchased purified from Sigma Aldrich

All compounds were dissolved in 100 % DMSO to obtain a 5 mM stock concentration based on the molecular weight of each compound. Before screening assay each stock solution were diluted in MilliQ water.

3.2. Methods

3.2.1 Circular dichroism measurement (CD)

All CD analyses were conducted using a Jasco 810 spectropolarimeter equipped with a Peltier system using a quartz cell of 1 mm or 10 mm optical path length. CD Spectra were obtained using a scanning speed of 100 nm/min, with a response time of 1s, over a wavelength range of 230−330 nm in increasing concentration of HCl, cations (KCl, LiCl and AgNO3), ligand (IMC-30) or protein (hnRNPK). For BCL2 G4 forming sequences, DNA and RNA samples were diluted in 10 mM Tris-HCl, pH 7.4 at a 4 µM or 2 µM, while the iM forming sequence was prepared in 10 mM Na-Cacodylate pH 5.0 or pH 7.5 at a strand concentration of 4 µM. MYC cytosine rich sequences were prepared at a 5 µM concentration in 10 mM Na-Cacodylate pH 6.76.The final concentration of each sample was controlled using NanoDrop1000 Spectrophotometer (Thermo Scientific) and before data acquisition, each oligos solution were heated at 95°C and slowly cooled to room temperature. Each reported spectrum represents the average of two scans, baseline corrected and smoothed out. The acquired dichroic signal was finally converted in molar ellipticity per residue ([Θ] = deg x cm2 x dmol-1).

Experimental data were fitted according to a single binding event process according the following equation (Eq.1):

$$y= (Bmax*x)/ (Kd+x) \quad \text{(Eq. 1)}$$

Melting and annealing curves were recorded by monitoring the variation of the dichroic signal at the λ of the maximum molar ellipticity over a temperature range of 25−95°C with a heating rate of 50°C/h or 25°C/h for BCL2 sequences, whereas for MYC mutants the variation of the dichroic signal as a function of the temperature was monitored over a temperature range of 15−70°C with a heating rate of 60°C/h.
Kinetic data were recorded for dBcl2_C after the addition of HCl to the oligonucleotide solution. After a mixing time of 10 s, the spectra acquisition was initiated using an interval scan of 120 s. The temperature was maintained constant at 25°C.

Thermal differential spectrum (TDS)

The thermal difference spectra were calculated by subtracting the UV spectra at 25°C from the UV spectra recorded at 95°C. The experiments were performed in 10 mM Tris-HCl w/o KCl pH 7.5 for the G-rich strands and in 10 mM Na-Cacodylate pH 5.0 for the C-rich strand. All the resulting thermal difference spectra have been normalized to the value of one at the maximal intensity calculated.

3.2.2 Fluorescence melting assay

Fluorescence melting curves were acquired using a Roche LightCycler480 (λecc 488 nm, λem 520 nm). A volume reactions of 20 µL was loaded in a 96-well plate containing 0.25 µM of the tested sequence in 10 mM LiOH with H_3PO_4 pH 7.4 and increasing concentrations of KCl (0-200 mM) or AgCl (0-20 µM). For the compounds screening each well of a 96-well plate contained 0.25 µM of dBcl2_C in 10 mM LiOH with H_3PO_4 pH 7.4 or pH 5.0 and increasing concentrations of ligands (0-20 µM). During the experiment the oligonucleotide was first denatured by heating to 95 °C at a rate of 0.1 °C/s, maintained for 5 mins at 95 °C and then annealed by cooling to 30 °C at a rate of 0.1 °C/s and maintained for 5 mins at 30°C before being slowly heated and annealed at a rate of 1°C/min. The Tm values were determined from the first derivative of the melting profiles using a Roche LigthCycler software.

3.2.3 Electromobility shift assay

The electrophoretic mobility of Bcl2 sequences in absence/presence of increasing concentration of KCl, LiCl, AgCl or PEG200 was monitored by a 15% native polyacrylamide gel (acrylamide:bis-acrylamide 19:1) in 1x TBE (89 mM Tris, 89 mM Boric acid, 2 mM EDTA, pH 8.0) containing or not KCl.

Before loading, all samples were heated at 95 °C for 10 minutes and then cooled down to room temperature overnight. Solved products were visualized and quantified on a Geliance system (Roche) before and after staining with Sybr Green II.
For Myc sequences, the electrophoretic mobility was monitored for 4CT and 5CT oligonucleotides 5'-end-labeled with IrDye88. Oligonucleotides were prepared at 10 or 100 nM concentration in a binding buffer contained 20 mM HEPES (pH 6.5), 1 mM DTT, 2 mM MgCl2, 1 mM EDTA, 10% glycerol, 100 mM KCl, 1 µg/µL BSA, 0.1% Tween-20, and 0.01µg/µL poly [d (I−C), and incubated on ice for 20 mins with or without increasing concentration of hnRNPK (0-70-115-700-1400 ng). Protein−DNA complexes were resolved by electrophoresis on a 8% native acrylamide gel in 1x TBE (89 mM Tris, 89 mM Boric acid, 2 mM EDTA, pH 8.0). Solved products were visualized and quantified using a LICOR xS system.

3.2.5 Fluorescent Intercalator Displacement assay (FID)

The DNA binding by Thiazole Orange (TO) was monitored using a Victor^3 multilabel counter (PerkinElmer). A total volume reaction of 100 µL was loaded in a 96-well black plate containing constant concentration of a target DNA (0.5 µM) in 10 mM LiOH with H_3PO_4 pH 7.4, 100 mM KCl and increasing concentrations of TO, until the saturation of the signal was reached. Changes in fluorescence emission were recorded at 535 nm and then plotted against the TO/ [oligo] ratio.
For FID assay, a 10 mM LiOH with H_3PO_4 pH 7.4, and 100 mM KCl solution containing DNA and TO in a 1:2 ratio, was added of increasing concentrations of tested derivatives. Changes in fluorescence emission were recorded and the percentage of TO displacement was calculated (Eq.2).

% TO displacement = 100-[(F/F0) x 100] (Eq.2)

Where, F0 is the fluorescence before the addition of ligands. The TO displacement was plotted as a function of compounds concentration. Each titration was repeated in triplicate.

3.2.6 Cell culture and caspase-3 assay

Human B cell lymphoma (U2932) cells were cultured in 10% FBS, 1% penicillin/streptomycin-supplemented RPMI medium at 37°C in a humidified atmosphere of 5% CO_2. Cells were counted using a haemocytometer and the viability was assessed using trypan blue exclusion. Caspase-3 activity in presence of selected compounds was evaluated using the ApoAlert Caspase-3 Plate Assay as the manufacturer's specification (Clontech, Mountain View, CA).

3.2.7 Protein production and purification

pET28-hnRNP K-positive transformants were seeded in Miller's LB broth and grown overnight. Then 600 mL Miller's LB broth was inoculated with the overnight culture and incubated at room temperature for 2-3 h, and then protein expression was induced by 1 mM IPTG overnight at room temperature. Harvested cells were resuspended in a lysis buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 1% Triton X-100, 1 mg/mL lysozyme and 1× protease inhibitor cocktail) and underwent to 10 cycles of the following: incubation on ice, vortexing, and sonication. Cell debris was removed by centrifugation at 4°C, and the supernatant was incubated with a HisPur Cobalt Resin for 1 h at 4 °C. The resin was washed with a wash buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, and 20 mM imidazole) and an elution buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, and 250 mM imidazole) was used to separate hnRNPK from the resin. Purified hnRNPK was subjected to buffer exchange using a Centricon centrifugal filter. Purity of hnRNPK was confirmed by Coomassie blue staining and Bradford assay was performed to determine its concentration.

4 RESULT AND DISCUSSION

4.1. *MYC* iM: a potential anticancer target

Several data are reported in literature about the proximal promoter region of *MYC* gene and the secondary structures that it contains. In particular, the coding strand comprises a Cytosine-rich sequence composed by eight runs of three or four consecutive C able to assume an iM conformation with a loop configuration of 5:5:5, a transitional pH close the physiological one and acting as an important element in the entire transcription process. Based on these reported data an in-depth evaluation of this domain was assess, focusing the attention on the DNA-protein complex involved during the transcription and the feasibility to modulate the effect of this complex through small molecules.

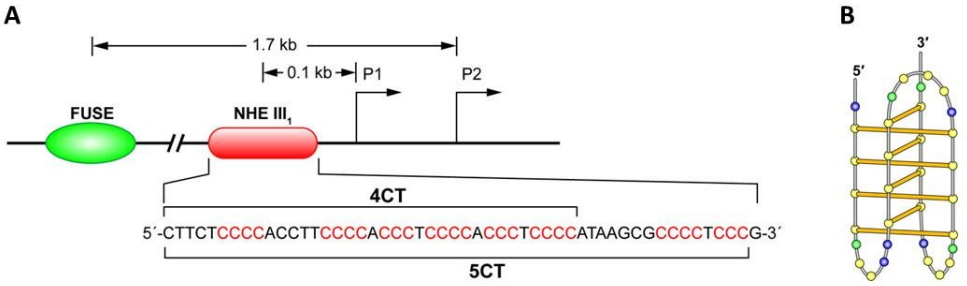
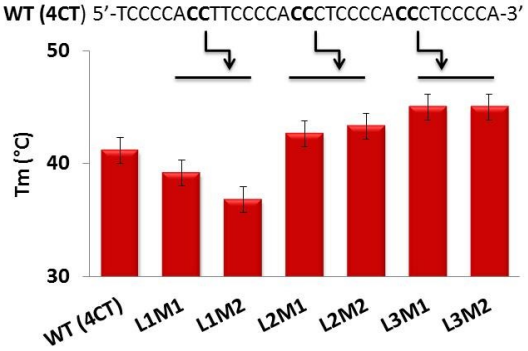


Figure 4.1 (A) Schematic representation of *MYC* proximal promoter region. (B)Proposed folding pattern of MYC iM composed by the 4CT sequence[146].

4.1.1. Loop composition affects the thermal stability of the i-Motif structure.

Lots of evidences is reported in literature showing that the stability of this secondary structure is strictly dependant both by the number of cytosines that composed each run[35,36,147]and by the oligonucleotide composition of the connecting loop By keeping that in mind, we decided to explore the impact of different loop on the stability of the iM conformation within the promoter region of *MYC*. In particular for each loop a single C-to-T mutation were considered. By using a spectroscopic technique as the CD assay, and working at the transitional pH of this structure (pH 6.7) were firstly confirmed the thermal stability of the wild type sequence, named as 4CT recording the variation of the CD signal at 286 nm, that is the most important positive contribution for the iM structure, as a function of the temperature, subsequently the same experimental approach was set up on each selected mutants in which one C was replaced by a T (Figure 4.2)



	Tm (°C)	ΔTm (°C)
WT (4CT)	41	0.0
L1M1	39.2	-1.8
L1M2	36	-5.0
L2M1	42.7	1.7
L2M2	43.3	2.3
L3M1	45.0	4.0
L3M2	45.0	4.0

Figure 4.2 Melting temperature and derived ΔTm recorded for wild type sequence and for each mutants at 286 nm in 10 mM Na-Cacodylate at pH 6.7.

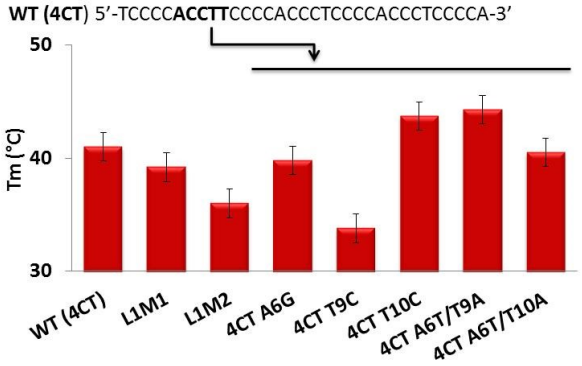
As expected the melting temperature reported in figure 4.2 A evidenced that the loop composition strongly affect the thermal stability of the non canonical structure assumed by the selected C-rich domain. More generally, all considered mutation does not alter the CD features recorded at the transitional pH, indeed the positive contribution at 286 nm was always maintained (data not shows). However, a detectable variation in the thermal stability of the iM assumed in solution was observed. In particular, for the oligonucleotides that contain the single C-to-T mutation in the first loop (5' end) a sensible decrease of the melting temperature was detected, in fact, by comparing the Tm derived for L1M1, in which the mutation involves the position 7, with the value derived for the wild type sequence, a decrease of about 2°C was observed, while a higher loss of stability (5 °C) was detected for the oligonucleotide in which the C8 was replaced by a T (L1M2). Suggesting that the first loop exert a strong impact on the stability of the overall structure.

Quite different is the behaviour observed for the second set of sequences, in which the mutation toward the T was localized in the second or central loop (ACCT). Particularly, compared to the iM adopted by the 4CT sequence, a small increment of the thermal stability was unexpectedly observed, that results to be 1.7 for the L2M1 (C_{16}-to –T) or 2.3 °C for L2M2 (C_{17}-to –T). Mostly the same was the behaviour observed also for the last set of oligonucleotide selected, where the single mutation was carried on the 3' later loop (third loop), in fact 4 °C stabilization was derived for both sequences (L3M1 and L3M2). Suggesting that in this case the replacement of a C to a T does no destabilize the tetraplex, but contrarily it exert a stabilization effect.

Considering together, this preliminary data exhibit that basically the single mutation exert different action based on its localization, in particular modifications in the second loop and in third one allows the formation of a more iM arrangement compared to the tetrahelical conformation assumed by the 4CT sequence. Probably, this stabilizing effect is due to the presence of an additional methyl group carried by the T able to stack over the other bases involved in the second and third loops. On the other hand,

the presence of this additional methyl group in the first loop may create a steric hindrance, thus a partial distortion of the loop is induced and a less stable conformation is assumed by both selected sequences (L1M1 and L1M2).

To fully understand the role exert by these three loops on the iM stability, for each of them we deeply analysed the base composition. By the comparison of the base contained into the first, second and third loop, was been immediately evident that the central loop (loop2) and the 3' lateral loop (loop3) are composed by the same five nucleotides while they differs from the first loop (5' lateral loop) for a T to C replacement (ACCCT instead of ACCTT). Moreover, considering only the first loop composition, it came out that the A6 may easily interact with the T9 or 10 through a Watson and Crick base pairing. This, may be in line with the thermal data collected for the first loop mutants, also suggests that the presence of an additional base pair could exert an important role for the stability of the overall iM structure. To further asses this idea a second set of mutant sequence for the first loop and two different double mutants were considered, and again the thermal properties at the transitional pH were evaluated (Figure 4.3).



	Tm (°C)	ΔTm (°C)
WT (4CT)	41	0.0
4CT A_6G	39.8	-1.2
4CT T_9C	33.8	-7.2
4CT $T_{10}C$	43.7	2.7
4CT A_6T/T_9A	44.3	3.3
4CT $A_6T/T_{10}A$	40.5	-0.5

Figure 4.3 Thermal evaluation for single and double loop one mutant sequences. Data collected at 286 nm in 10 mM of Ma-Cacodylate pH 6.7.

In line with the preliminary data previously describe, each selected mutations sensible alters the stability of the original system generate by our C-rich sequence. Particularly the data collected using the same experimental condition, confirmed our hypothesis regarding the existence of a WC base pairs, moreover, they provide us the exactly localization of this additional pairing. In fact, if we compared the melting temperature (Tm) and the ΔTm derived, a remarkable loss of the thermal stability was observed for the sequences carrying the single mutation in position 6 (A-to-G) and in position 9 (T-to-C), suggesting that the purine in position six and the pyrimidine in position 9 are extremely important for the iM. On the contrary, the behaviour observed when the position 10 was mutate was quite unexpected, in fact a higher melting temperature rather than the Tm derived for the wild type sequence was collected. In order to explain why a thermal stabilization was observed for the 4CT $T_{10}C$ sequence, we

considered once again the nucleotide composition. By merging all data collected we can assume that the A6 and T9 are involved in the formation of a canonical Watson and Crick base pairs that exert a fundamental role in the stabilization of the iM. While the replacement of the T into a C in position 10 leads the formation of a longer C-run, composed by 5 consecutive pyrimidines, that probably allows the formation of a different and more stable iM in which different C are involved.
In order to definitively confirm the presence of a WC base pair between the position 6 and 9, the thermal properties of two sequences carried a double mutation were evaluated. Both of them contained the same A-to-T mutation, while the T-to-A mutation involves firstly the position 9 and secondly the position 10.
Unfortunately, the collected data in for these two double mutants cannot allow us to totally confirm the position involving the WC base pairs indeed; the acquired data were quite unexpected. Specifically, following the effect exerted by the double mutation in positions 6 and 10 on the thermal stability of the iM, a comparable Tm value with the one derived for the wild type sequence was derived. Suggesting that, the switching of these two positions does not alter the iM stability. On the contrary, the mutations involving the positions 6 and 9 provided a higher Tm value. This data was quite surprising, in fact, in our idea, the inversion of these two positions does not change sensibly the thermal stability of the structure. Despite that, to understand why a thermal stabilization was observed, again we took into consideration the nucleotide composition. For the second mutant selected the position 6 was replaced with a T, a pyrimidine base, which differs from the C for the presence of an additional methyl group. Based on that, we suggest that thanks to the presence of this additional –CH3 group, additional interactions able to stabilize the last C-C pair involving the first C run may exist, thus, a more stable iM conformation can be assumed by the oligonucleotide.

Taking together all the collected data, it is reasonable to assume that for the iM within the c-Myc promoter, thanks to the presence of an additional WC base pair, involving the positions 6 and 9 and more generally the first loop plays the most important role for the stability of the overall structure. Also, it is reasonable to transfer this suggestion to the third loop in which probably the position 24 and 28, due to their spatial proximity, allows the formation of the same WC interaction. Moreover considering the importance of the first loop on the stability, the iM folding is driven by the first and the third loop. On the contrary, concerning the second loop, it covers a marginal role on the iM stability and that could be reasonably associated with its dimension indeed, being the wide loop the additional WC bp isn't allowed.

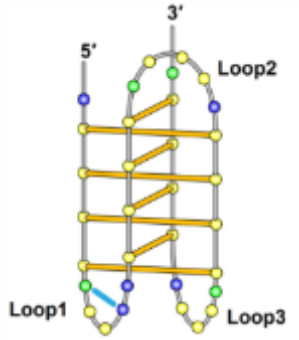


Figure 4.4 Proposed schematic representation for *MYC* iM. The light blue interaction is referred to the WC base pair between the position 6 and 9.

4.1.2. IMC-30: a potential iM stabilizer compound.

Based on its implication in a wide range of cancers, over the years MYC became one of the most attractive anticancer targets. Unfortunately, the lack of a targetable domain and of a specific enzymatic activity, associated with an extremely short half-life of the primary transcript make extremely difficult to gain a direct therapeutic approach. Additional strategies are obviously required and one of them is referred to an indirect inhibition of the MYC expression using a compound called JQ1, that prevent the transcriptional activation by a competitively binding towards the bromodomains region or to the acetyl-lysine recognition domain of several transcription factors involved in MYC expression pathway[148]. Despite, this compound is look as a promising anticancer drug, the direct modulation of the gene expression still remain the main focus for the scientific community and this may be obtained by targeting the non canonical secondary arrangements (G4 and iM) contained in the proximal promoter, that as already reported in literature, act as a molecular controls for the gene expression pathway.

An important goal for achieving that is the identification of small molecules able to modulate the formation/disruption of these non-canonical structures or able to modulate their recruitment of transcription factors. Recently, Hurley and co-workers have identified an ellipticine derivative (GQC-05) able to repress the transcription of MYC and kill tumor cells through a specific targeting to the G4 structure[149]. Even though the G-quadruplex has been demonstrated to be an effective target for repressing the MYC activation, the iM structure presents the potential to be the most successful target. Indeed, compared to the G-Quadruplex conformation, this folding present unique chemical properties that make it less polymorphic and easier to target. Thus, several scaffolds were screened with the aim to find a selective iM ligand. Particularly, using a fluorescence resonance energy transfer (FRET) assay, three compound libraries were previously screened in Hurley's laboratory and the collected data revealed the presence of several drug-like structures able to recognize the iM.

Among them, using the z-score method and based on the quenching effect, two small molecules were selected: IMC-16 and IMC-30. Concerning the IMC-16, a very well characterization is already be done and its ability to selectively bind and stabilize the iM leading the transcriptional activation was already published [146]. Regarding the second hit-compound selected (IMC-30) no details about its activity were known. Thus, by applying the CD technique, the effect of this small molecules on the conformational equilibria of 4CT sequence was evaluated firstly by working at slightly acidic pH (pH 6.1) and then at the transitional pH (pH 6.76). Moreover, its effect on the thermal properties was evaluated (Figure 4.4).

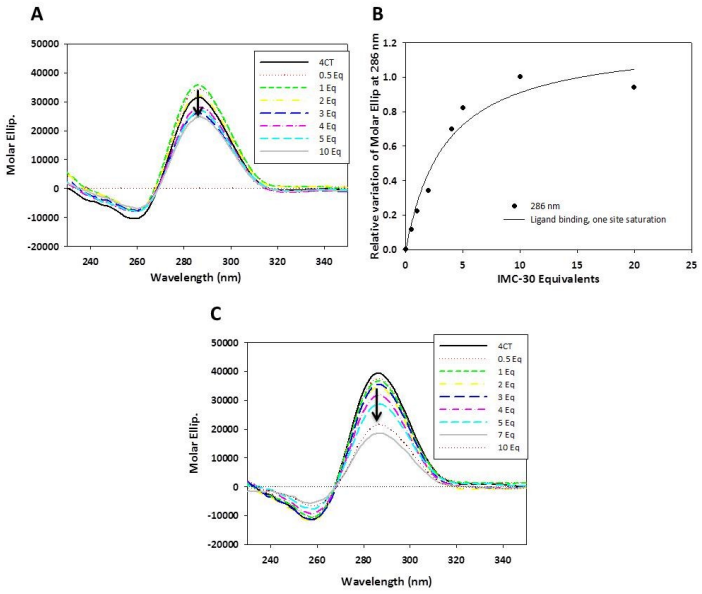


Figure 4.5 Effect of IMC-30 on MYC iM. (A) CD titration of 5 µMT 4CT in 10 mM Na-Cacodylate pH 6.7 with increasing concentration of compound. (B) Relative variation of the molar ellipticity at 286 nm. (C) CD titration in increasing concentration of IMC-30 in 10 mM Na-Cacodylate pH 6.0.

Unexpectedly, as figure 4.4 A shows, working at the transitional pH, the increasing concentration of compound was not associated with an increment of the CD signal. Indeed upon the addition of IMC-30 the shape recorded still maintained the positive contribution at 286 nm and the negative one at 261 nm; however a sensible decrease of the CD intensity was observed suggesting that the selected compound partially alters the tetraplex conformation. By plotting the relative variation of the molar ellipticity towards the compound concentration a higher Kd value of 17.4 ± 5.1 µM was derived towards the compound concentration compared to the already reported value obtained for the IMC-16. From this first set of data, we can suppose that despite the selected compound was able to recognize the already existing iM conformation in solution; increasing concentration of it partially promote the switching towards a different conformation. To assess if this compound is able to recognize the C-rich

domain, leading the destabilization of the tetraplex conformation and bring it towards to a different conformation, as the hairpin one, the same titration was repeats by working in iM promoting condition (pH 6.0). As the collected data show, (figure 4.4 C) increasing concentration of IMC-30 induced the behaviour observed at pH 6.7, thus only a decrement of the intensity of the signal recorded at 286 nm was detected. From the collected data during the two titrations, it appears that the IMC-30 compound, as the IMC-16, is able to recognize the cytosine rich sequence within MYC promoter and particularly, it seems able to bind the iM conformation contained in solution. Furthermore, considering that high concentration of compound leads only a decrement of the CD intensity probably this interaction partially distorted the tetrahelical structure. Thus, on the contrary of IMC-16, our hit-compound may be acting as an iM destabilizer compound.

In order to check this hypothesis, the thermal profile of the wild type sequence in presence of a constant concentration of IMC-30 was considered and the variation of the CD contribution at 286 nm was followed as a function of the temperature (figure 4.6 A and B).

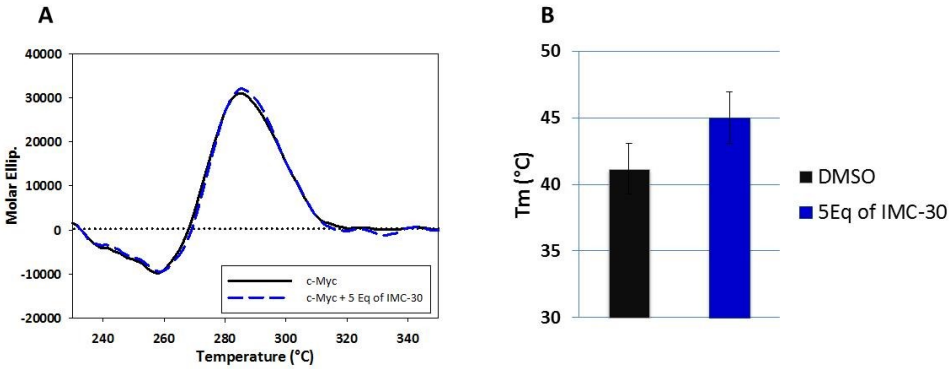
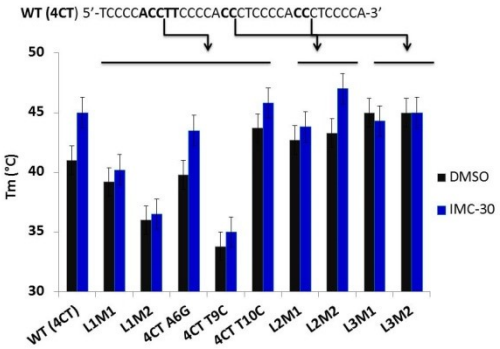


Figure 4.6. Evaluation of the thermal properties of c-*Myc* iM in presence of compound. (A) CD spectra recorded w/o 5 Equivalents of IMC-30 in 10 mM Na-Cacodylate pH 6.7. (B) Melting temperature derived in absence/presence of compound.

As reported in figure 4.6 the addition of 5 Eq of IMC-30 does not alter the CD features recorded either in term of intensity or shape. However, a remarkable variation in terms of thermal stability was observed. Indeed, by following the denaturation process at 286 nm as a function of the temperature a 4 °C increment of the melting temperature was detected upon the addition of compound. Thus, in contrast with the previously reported idea regarding the ability to destabilize the iM, this last set of data exhibit that in presence of our compound an increment of the thermal stability of the iM assumed in solution by the cytosine-rich sequence was observed.

After having established that our compound stabilizes the iM conformation assumed by the 4CT sequence, our issue became the identification of the binding site on the iM for the IM-30. Thus, a constant concentration of compound was added to each single

mutants selected then after an incubation time of 20 minutes, the denaturation profile at 286 nm was recorded as a function of the temperature. Collected data are reported in figure 4.7



	DMSO (Tm °C)	IMC-30 (Tm °C)	ΔTm (°C)
WT (4CT)	41	45.0	4.0
L1M1	39.2	40.2	1.0
L1M2	36	36.5	0.5
4CT A_6G	39.8	43.5	3.7
4CT T_9C	33.8	35.0	1.2
4CT $T_{10}C$	43.7	45.8	2.1
L2M1	42.7	43.8	1.1
L2M2	43.3	47.0	3.7
L3M1	45.0	44.3	n.a.
L3M2	45.0	45.0	0.0

Figure 4.7 Effect of IMC-30 on *MYC* mutant sequences. Data collected in 10 mM Na-Cacodylate pH 6.7 in presence/absence of 5Equivalent of IMC-30.

As expected, the collected data (Figure 4.7) showed that the behaviour of all sequences in presence of IMC-30 remarkably depends both by the type and the position of the single mutation. Generally, the addition of compound always induce an increment of the thermal stability, indeed, with the exception of loop 3 mutants, all Tm values derived upon the addition of IMC-30 were higher than the Tm derived for each mutants in 10 mM Na-Cacodylate pH 6.7. With the aim to identify the binding site, we deeply evaluate the melting temperatures and the derived ΔTm for each loop mutant. Considering the first loop, the value collected revealed that the addition of IMC-30 remarkably affects the thermal stability of the iM assumed in solution. Especially, a 3.7 °C increment for A_6G mutant, a 2.1 °C for $T_{10}C$ mutant and 2°C for L1M2 were detected, meaning that upon the addition of the small molecules, it recognize the first loop improving the stability of the iM. Moreover, based on the mutants presenting the higher ΔTm, we can also suppose the nucleotide involved in the binding process between the iM and the compound, which are the A in position 6, the T in position 10 and finally the T_{10}.

Regarding the two mutant sequences selected for the second loop, a sensible improved of the melting temperature was observed just for the sequence in where the C17 was mutated into a T (ΔTm 3.7 °C).On the contrary, when we tested the compound on the loop 3 mutants no stabilization effect was observed suggesting that this lateral loop is not the preferred binding site. It also should be noted that the mutation in this loop even in the absence of compound results in a 4°C increase in the Tm compared to WT.

Sadly, based on our evidence, we cannot unquestionably assume where the binding site for the selected compound is. However, we can propose that this chemical entity recognizes and improves the thermal stability of the iM following the interaction with

the first and the second loop. Particularly, we can also suggest the bases involved in this binding process, which are the A6, C8 and the T10 in the first loop, whereas for the second loop, probably the bases involved in the binding process should be the A15 C17 and T19.

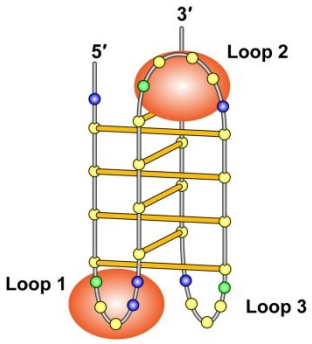


Figure 4.8. Schematic representation of the proposed site involves in the binding process between the IMC-30 and the iM adopted by *MYC* C-rich sequence.

4.1.3. IMC-30 and DNA-protein complex: new tool for the modulation of the transcription process.

Thanks to several contribute reported in the literature the secondary structure adopted by the DNA, both canonical and not, may act as a scaffold for the recruitment of protein involved in the activation/repression of the gene expression. Specifically, for the MYC gene and the iM structure several pieces of evidence report that the heterogeneous RNA binding protein K (hnRNPK) binds the promoter region of this gene as a single strand leading the transcriptional activation. Particularly, Boss and colleagues demonstrated that the K-homology (KH) domains contained within this protein, recognize and binds selectively the C-rich sequence located into the loop region of the MYC iM [150–153]. Thus the iM may represent the optimal scaffold to enhance the transcription. Sutherland and colleagues in 2016 also demonstrated that this protein preferentially binds the longer C-rich domain contained into the promoter region of MYC and in particular they evidenced that the protein was able to recognize both Cytosine domain, the shorter one (4CT), that is required for the completely folding towards the iM and the longer ones, that is proved to be able to assume exactly the same iM conformation of the 4CT despite the presence of more C. However, physiologically, to get the competent form required for the maximal transcription activation the additional runs of C contained into the 5CT sequence is required. Thus a more stable DNA-protein complex was observed.

After having established that our compound was able to recognize and stabilize the iM, the next step was to explore its ability to modulate the recruitment of the mentioned transcription factor. By using two different approaches (native PAGE assay and spectroscopic technique) firstly we confirmed the ability of this protein to recognize, binds and unfold the iM structure adopt both by the 4CT and the 5CT sequences. Then,

the effect exerts by the addition of IMC-30 on hnRNPK protein recruitment was explored.

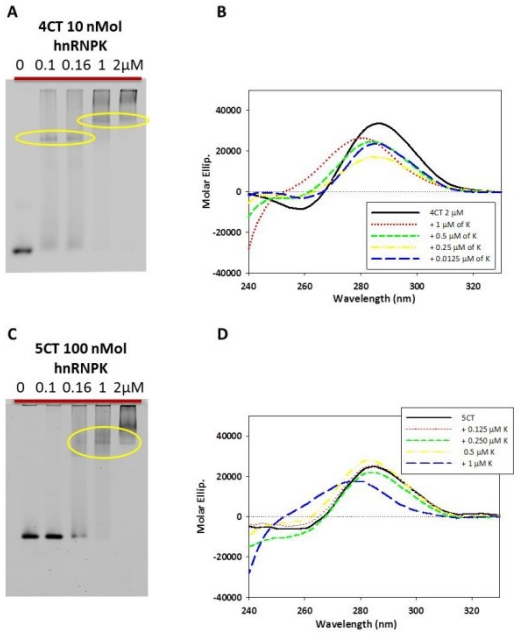


Figure 4.9 Unfolding process lead by hnRNPK. (A and C) 10 % native PAGE assay for 4CT and 5CT in increasing concentration of protein. (B and D) CD spectra of 2 µM of oligonucleotide sequence (4CT or 5CT) in increasing concentration of protein.

Briefly, the collected data confirmed the ability of the hnRNPK to bind and unfold the iM structure adopts both by the 4CT and the 5CT sequence. Indeed, as reported in figure 4.9 A and C, by working in a native condition, the addition of increasing concentrations of protein provide the formation of a new species characterized by a very low electrophoretic migration. Thus, the formation of a high molecular weight complex, in which the protein after the recognition step induced the unfolding of the secondary structure, is confirmed. One other technique that allows us to evaluate the conformational changes in solution is the CD spectroscopy, in order to confirm definitively our preliminary data; a titration with increasing concentration of protein was set up for both sequences. Data collected are reported in figure 4.8 B and C, working at the transitional pH, we firstly induced the iM formation, then increasing concentrations of hnRNPK were added to our samples and the CD spectra were recorded after 20 min incubation at room temperature. Again the unfolding of the iM was observed, in fact, a sensible variation of the CD features both in terms of intensity and shape was observed. Especially, at the end of the titration in where 1 µM of protein was added a decrease of the intensity of the signal associated to a 6 nm red shift of the positive contribution was observed.

Despite, for both C-rich sequences the presence of hnRNPK induced the unfolding of the non canonical secondary structure, the observed behaviours were quite different.

Concerning the 4CT element, two different DNA-protein complexes were resolved, whereas just only single DNA-protein complex was detected for the longer sequence. Similarly, during the CD titration, for the 4CT element, the addition of the protein immediately induced the switching towards a different DNA structure. On the contrary, to get the same variation on the CD spectra for the 5CT element, a higher concentration of the protein has been necessary.

This suggests that in contrast with the previously reported data, the unfolding activity of our protein is more prominent for the shorter (4CT) C-rich sequence in spite of the lacking of the third KH domain.

Based on our data, the IMC-30 acts as an iM stabilizer compound while the protein leads its unfolding. As for the G4, a compound able to stabilize/destabilize the secondary structure may be use as a potential anticancer drug. Thus, the following step was to explore the effect of our compound on the unfolding process mediated by hnRNPK. To check that, we set up the same protocols mentioned before (native PAGE assay and CD spectroscopy). As figure 4.10 A and B show, the PAGE assay did not allow us to defined the effect of the compound the iM-hnRNPK complex. Indeed, for both sequences (4CT and 5CT) after the addition of increasing concentration of the compound, a single electrophoretic band, having the same electrophoretic mobility of the control, was observed. This could be due to the small size of our compound, which is therefore not able to increase the hydrodynamic volume of our complex, otherwise, the interaction between the DNA-protein and the compound is extremely weak and under the electrophoretic migration, the chemical entity has been lost. For that reason, a different approach was needed and again, the CD spectroscopy was applied (figure 4.11).

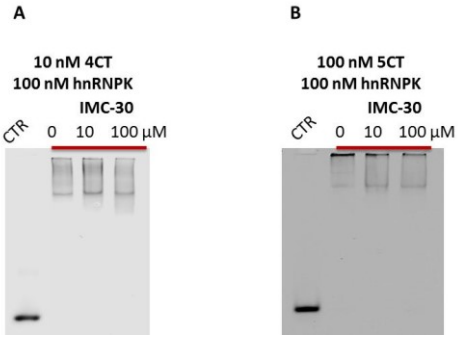


Figure 4.10 Effect of IMC-30 on protein-DNA complex. (A) 10 % native PAGE assay of 10 nM 4CT in presence of 100 nM of hnRNPK and increasing concentration of compound. (B) 10 % native PAGE assay of 10 nM 4CT in presence of 100 nM of hnRNPK and increasing concentration of compound. All samples were loaded after an incubation time of 20 minutes.

As reported in figure 4.10 B, the titrations with increasing concentration of hnRNPK exhibit that the addition of protein does not provoke the unfolding of the iM contained in solution. Indeed, by comparing the behaviour collected in the absence and in the

presence of compound (Figure 4.10 and 4.111) clearly neither the addition of one Equivalent of protein, promotes the unfolding of the iM assumed by both selected cytosine rich sequences. Thus, in line with the data collected for the thermal stability of the iM in presence of the compound, this experiment suggests that our compound stabilize the tetraplex structure preventing the unfolding activity of our protein.
Moreover, focusing the attention in the wavelength range between 240 to 260 nm, the same variation in terms of shape and intensity was observed after the addition of the protein provokes. Thus we can presume that during the titration the protein still able to recognize and bind the sequence. However, thanks to the presence of IMC-30, that has a stabilizing effect on the iM, the entire unfolding process is prevented.

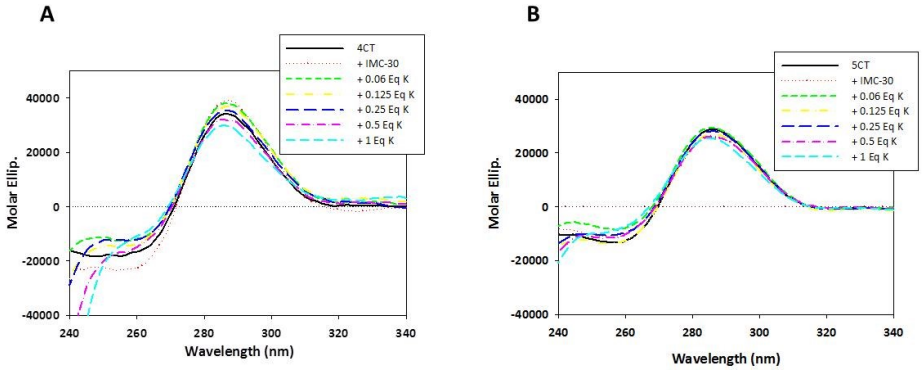


Figure 4.11 CD spectroscopy in increasing concentration of protein and constant concentration of compound on 4CT (A) or 5CT sequence (B).

4.1.4 IMC-30 and the apoptosis. A possible target to gain an innovative treatment.

MYC is one of the most-studied genes, however, at this day; a strategy able to directly modulate its expression is not available. Our previous data exhibit that IMC-30 compound recognizes stabilizes and binds the i-Motif structure within the promoter of this gene. Moreover, we observed that our compound by following the binding step, alter the recruitment of the transcription factor (hnRNPK) and the unfolding process of the iM.
Based on that evidence and with the aim to determine if this selected compound may be considered as a potential anticancer drug, we decided to evaluate if the stabilization of the iM lead by the presence of IMC-30 would result in activation or repression of the gene expression.
Using a human B lymphoma cell line that simultaneously presents the MYC translocation-positive and the Bcl2 amplification-positive, the cytotoxicity of IMC-30 was initially assessed, and then the effect on the gene expression at the mRNA production level was evaluated. Despite, the obtained results are not reported in this chapter, differently, from what was observed after the treatment with IMC-16, the presence of IMC-30 leads a sensible decrement in the mRNA production was observed,

which also results to be time- and dose-dependent. Thus, our compound seems to be able to negatively modulate the expression of MYC.
Based on the implication of MYC in different physiological processes, including apoptosis, we decided to evaluate the activity of this compound on the apoptotic pathway. In particular, the ability of this compound to activate or not the caspase-3, one of the most important caspases involved in the entire process, was assessed (figure 4.11).

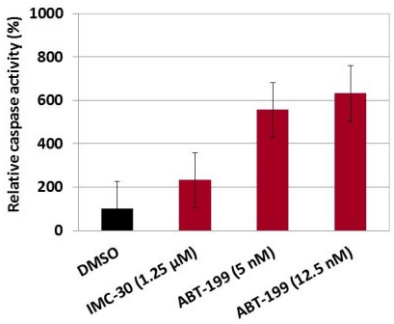


Figure 4.12 Caspase-3 test after a 24-hour treatment in the presence of 1.25 µM of compound (IMC-30). Compound ABT-199, a promising anticancer agent for the treatment of chronic lymphocytic leukaemic and estrogen receptor-positive breast cancer, was selected as the reference compound. While as a negative control was used the DMSO.

Following the protocol, after a 24 h treatment, the caspase-3 activity was assessed. As reported in figure 4.12 compared to the negative control (DMSO), when the cell line was treated with our compound, an increment of the enzymatic activity was observed. Thus, suggests that IMC-30 may affect and improved the apoptotic pathway of the U2932 cell line. To check the activity of IMC-30, we compared its activity with the one derived for the reference compound (ABT-199)[154]. As expected, after 24h treatment, ABT-199 induced a sensible variation of the enzymatic activity. This can be associated with the ability of this compound to recognized and interact with the BCL2 protein, that is implicated into the regulation of the apoptosis process.
Considering this first set of data, our results are extremely promising; in fact, they revealed that IMC-30 may be used as a potential drug thanks to its ability to hit the iM and also thanks to its ability to modulate the apoptosis.
It's reported that the non canonical conformations may be used as a scaffold for selective small molecules in order to modulate the transcription of the gene in a direct mode. Considering that, we decided to evaluate which was the best target among the non-canonical structures contained into the MYC promoter (G4 or iM), thus the same experimental protocol was set up using an ellipticine derivative (GQC-05) already known for its ability to binds and stabilize the G4. Collected data and the comparison with those reported in the presence of IMC-30, after a 24h treatment, are reported in

figure 4.13. Clearly, also in the presence of the same concentration of GQC-05, the induction of the caspase-3 activity was promoted. Indeed, an increment of the activity of this enzyme was observed if compared to the negative control. Moreover, as we already observed for IMC-30, again the activity of Venetoclax compound (ABT-199) results to be more prominent and a higher increment of the enzyme activity was detected. By comparing the experimental results obtained for both selected compounds it appears that they present a quite similar ability to modulate the apoptosis. Indeed, working at the same compound concentration, no sensible differences among the percentages of activation was detected. This, suggest that both compounds may be selected as a possible new anticancer treatment, thanks to their ability to reduce the mRNA production and also thanks to their ability to induce the apoptosis.

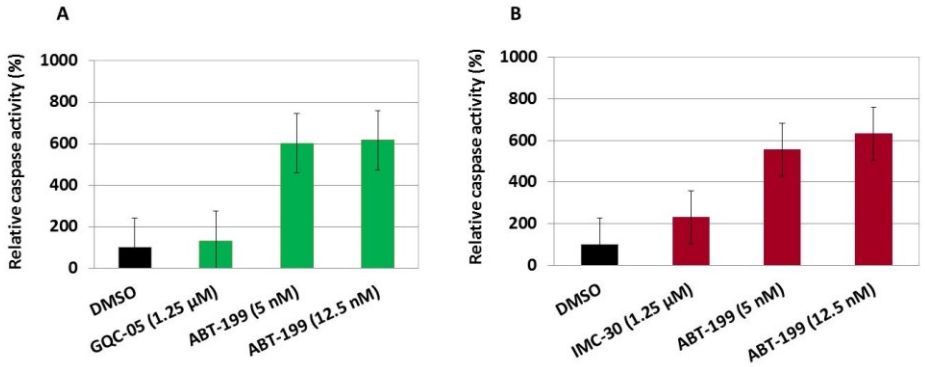


Figure 4.13. Comparison of the caspase-3 assay after a 24h treatment with GQC-05 (A) or IMC-30 (B).

After having considered the MYC gene alone and the possibility to directly modulate it with small molecules. The concomitant overexpression of MYC and BCL2 was taken into consideration. As already know the concurrent overexpression of these two oncogenes is a condition that occurs in at least 30% of patients affected by non-Hodgkin lymphoma, which is characterized by a rapid disease progression and drug resistance. Currently, no effective strategies are available to treat those patients, thus new approaches in which both critical oncogenes are simultaneously targeted are encouraged. In order to evaluate this possibility, we decided to evaluate the effect of a drug combination on the caspase-3 activity. Again, ABT-199 (Venetoclax) was selected for its ability to block the Bcl2 protein, whereas for Myc we selected IMC-30 and GQC-05 as compounds able to modulate the conformational equilibria at the promoter level and JQ1 a member of triazolo-diazepine already known for its ability to indirectly reduce the MYC expression through the inhibition of the BET bromodomain proteins [155]. Firstly the effect of this compound was evaluate in comparison to ABT-199, then the combination treatments in where the Venetoclax was combined with the G4 or iM binding compounds or with the JQ1 compound were considered (figure 4.14). As

reported in figure 4.14 A despite the JQ1 is known for its ability to indirectly modulate the MYC gene by targeting the bromodomain proteins, the observed activation of the caspase-3 was similar compared to the iM and G4 binders activities, suggesting that all selected compounds able to directly or indirectly modulate the expression of the MYC protein, efficiently induces the apoptosis. Regarding the combination assay, the behaviour observed was quite similar. Indeed, after a 24h of treatment, each considered combination induced the activation of the caspase-3. However, again the best results were detected in presence of GCQ-05 and IMC-30, suggesting that a connection between the possibility to directly modulate the protein expression and the activation of the caspase-3 and the apoptosis process may exist. Moreover, it appears that the combination in presence of a G-quadruplex binders give the best activation of the enzymes. Thus, in this case targeting the G4 may be the best choice.

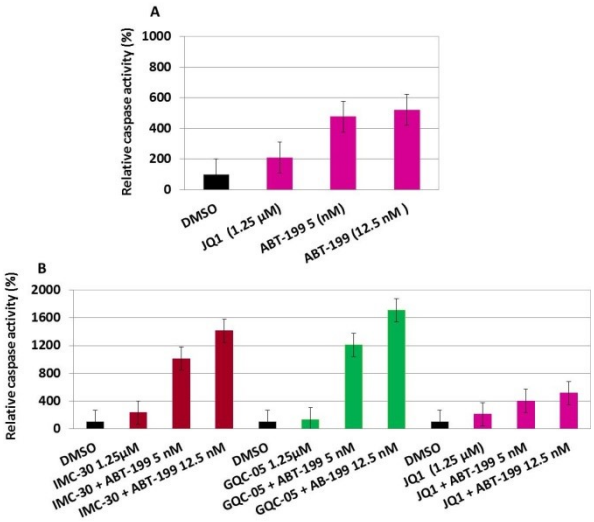


Figure 4.14. (A) caspase-3 assay in presence of JQ1 on U2932 cell line. (B) Comparison of the combination treatment in presence of 1.25 µM IMC-30, GQC-05 or JQ1 and 5 or 12.5 nM or ABT-199.

After having proved that in combination the G4 and iM binders present a higher aptitude to induce the apoptosis compared to the JQ1, different combinations were considered and especially, we focused on the possibility to combine the iM and the G4 binders. Then the possibility to obtain a significant activation of the caspase-3 by targeting the MYC gene both directly and indirectly (figure 4.15) was assessed.

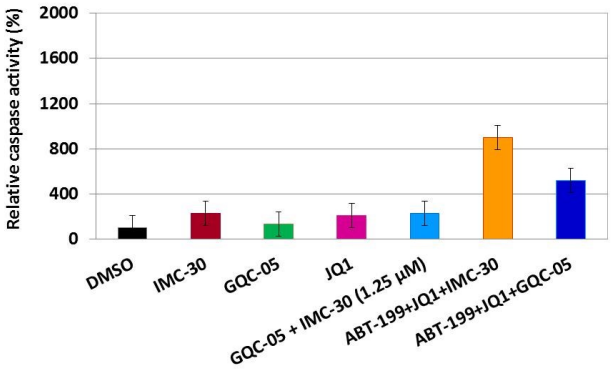


Figure 4.15 Comparison of drug combination on U2932 cell line.

As expected, since the formation of the G4 and the iM within *MYC* gene is mutually exclusive, the data collected and reported in figure 4.1 revealed that the combination of the GQC-05 and IMC-30 does not strongly affect the apoptosis process. Indeed, the percentage of activation derived after a 24 h treatment in the presence of both compounds is comparable to the percentage observed after the treatment with each single compound tested. On the contrary, the other two combinations induced an increment of the apoptosis. Especially, working in the presence of two compounds able to hit the MYC gene (IMC-30 and JQ1) and in the presence of a *BCL2* protein inhibitor the highest activity was observed.

By considering all collected data, it appears that each tested compounds and combinations produce the activation of the caspase-3 in a cell line presenting the overexpression of two important genes as *MYC* and *BCL2*.

Furthermore, by keeping in mind our aim, that was the identification of a new strategy for patients in which the classical treatment based on radiotherapy or chemotherapy is not responsive due to the overexpression of *MYC* and *BCL2*, it appears that a multidrugs approach that combined ABT-199, a compound able to inhibit the *BCl2* protein, with compound able to modulate the conformational features of *MYC* gene, may be used as an innovative anticancer treatment.

4.2. DNA G-quadruplex folding sequences: structural characterization

4.2.1. dBcl2_G

As already mentioned, in the sense strand of the DNA 5' untranslated domain a 25-nucleotide long putative quadruplex forming sequence (PQS) is contained.

dBcl2_G: 5'-**GGG GG**C CGT **GGG GT**G **GG**A GCT **GGG G**-3'

In order to assess the ability of this tract to assume a G4 conformation we studied it in the absence/presence of quadruplex promoting cations (KCl and NaCl). In this contest, the circular dichroism (CD) is a useful technique to obtain a general indication of the nucleic acid arrangements in solution. Indeed, DNA and RNA are able to interact with a polarized light beam and provide a specific signal in terms of shape and intensity, depending on the structure assumed in solution[156–158]. The spectroscopic evaluation of dBcl2_G was initiated by recording the CD signal in 10 mM Tris buffer pH 7.4. After dilution of the DNA stock solution to 4 µM in the required buffer, the recorded CD signal was remarkably intense and characterized by the presence of a positive peak at 263.5 nm and a positive contribution at 290 nm. However, an annealing cycle strongly alters the CD signal associated to the selected sequence. Indeed, as reported in figure 4.16 after the thermal treatment the intensity of the CD signal recorded dropped and a 4 nm red shift of the main positive peak was detected. Since the intensity of the CD signals recorded before the melting/annealing was remarkably high, our data suggest the existence of a pre-organized conformation possibly deriving from the stock solution. In order to understand the nature of this initial arrangement, two thermal different spectra were derived (TDS). For nucleic acid, the thermal difference spectrum is obtained by simply recording the ultraviolet absorbance spectra of the unfolded and folded states at temperatures above and below its melting temperature (Tm). The difference between these two spectra is the TDS that as the CD signal, presents a unique shape for each type of oligonucleotide rearrangements[159]. The comparison of the two TDS calculated before (solid line) and after (long dash line) the annealing cycle is reported in figure 4.1 B. The two CD shapes are remarkably different; moreover, due the presence of a negative contribution at 299 nm, the TDS profile recorded before the denaturation step suggests the presence of a G4 structure in solution that was lost after the annealing step. Since the aptitude of dBcl2_G to keep a G4 conformation even in the absence of KCl, possibly deriving by the presence of traces of cations in the stock buffer, was suggested, before each experiment, our DNA solution was subjected to a proper thermal treatment (10 mins at 95 °C followed by overnight annealing to the room temperature).

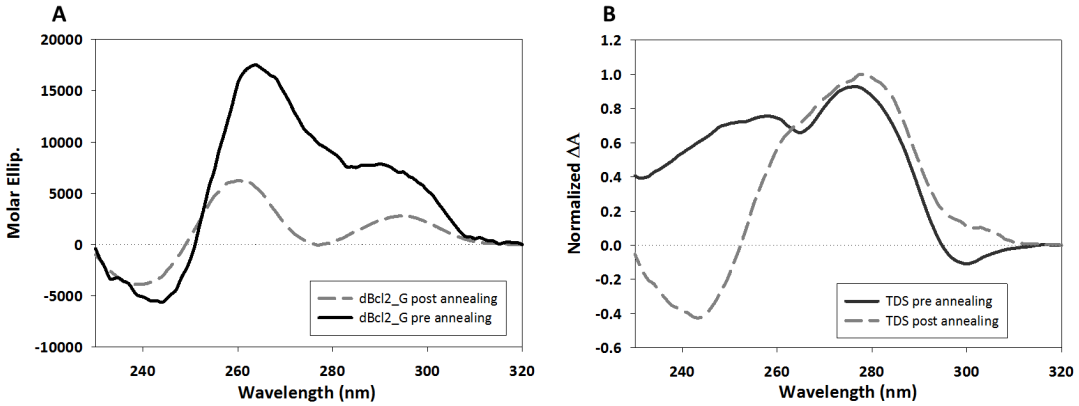


Figure 4.16. (A) Dichroic spectra for 4 µM dBcl2_G in 10 mM Tris pH 7.4 recorded before and after the annealing process. (B) Thermal difference spectra calculated in Tris 10 mM pH 7.4.

To evaluate the effect exerted by the potassium ion on the G4 folding process, a CD titration with increasing KCl concentrations was performed.

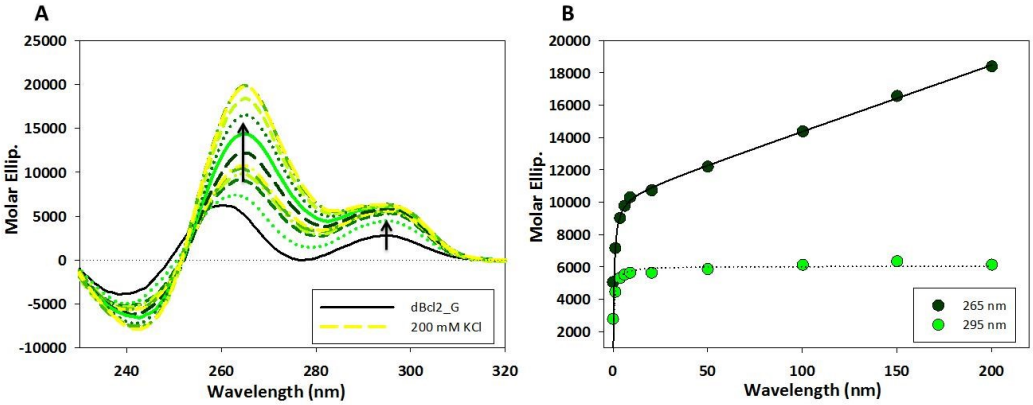


Figure 4.17. (A) CD titration of 4 µM dBcl2_G in 10 mM Tris pH 7.4 with increasing concentration of KCl. (B) Relative variation of the CD signal at 265 nm and 295 nm plotted against the cation concentration.

Collected data exhibit that the spectrum registered in absence of cations remarkably changes during the titration, both in terms of shape and intensity (Figure 4.17 A). Indeed, the intensity of the two main positive (+ve) peaks increased, and a 4 nm blue shift of the positive peak at 260 nm was observed. This suggests the ability of the KCl to induce a conformational change on DNA folding. The presence of the two +ve contributions at 265 nm and 295 nm might be referred to the presence of a G4 of mixed topology, in which the strands are oriented in opposite directions, or to the coexistence of multiple conformations (parallel and hybrid form) in solution that differently contribute to the CD signal. By plotting the variation of the molar ellipticity at these two wavelengths as a function of KCl concentration two different trends were observed (figure 4.17 B). In particular, at 295 nm a single process occurs. The

saturation is reached at about 20 mM of KCl and by fitting the data according to a one binding site saturation equation y= (Bmax*x)/(Kd+x) a dissociation constant of 1.4 ± 0.2 mM was derived. Extremely different is the behaviour observed at 265 nm. The plot of the molar ellipticity towards the KCl concentration revealed the presence of two processes in solution. The first one reminds what observed at 295 nm and correspond to a dissociation constant of 2.5 ± 1.0 mM for the KCl. However, higher concentrations of cation induced an event that reminds the occurrence of non-specific DNA-cation interactions. This suggests the coexistence of multiple species in solution that differently contribute to the overall CD shape.

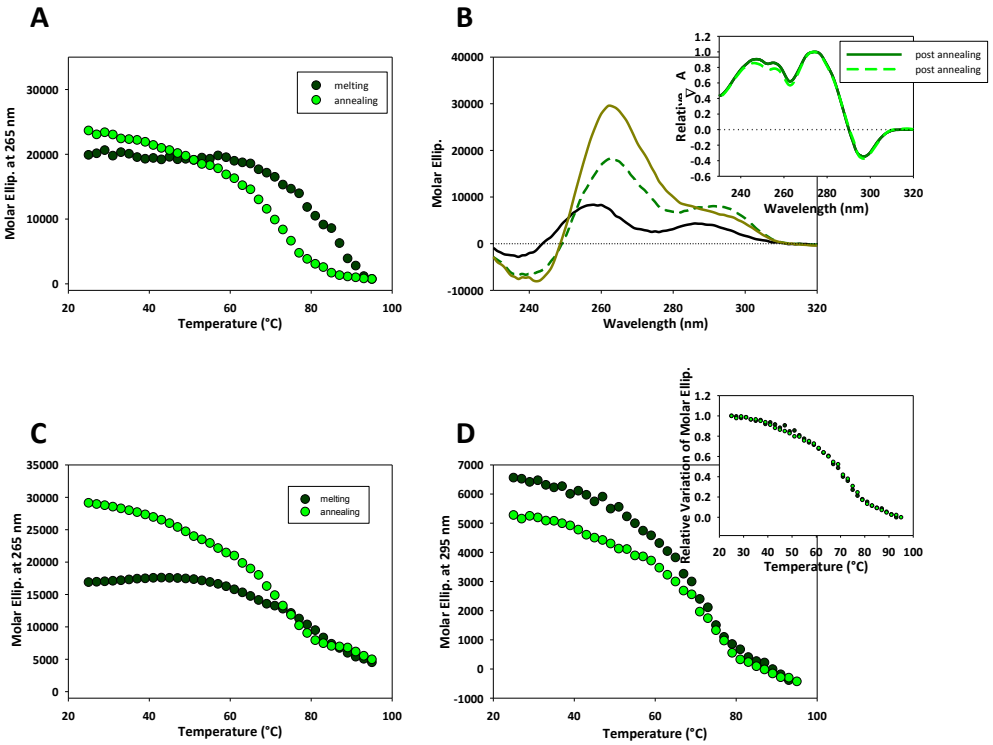


Figure 4.18. Melting experiment for dBcl2_G. (A) Melting (dark green) and annealing (light green) curves collected upon the addition of 150 mM KCl at 265 nm. (B) CD signal recorded before the addition of KCl (black line), in presence of 150 mM KCl before (medium dash line) and after annealing (solid dark green line). (C and D). Melting (dark green) and annealing (light green) curves recorded at 265 and at 295 nm in presence of 100 mM KCl at a heating rate of 50 °C/h.

To better understand the effect of potassium chloride on the conformational equilibria of dBcl2_G the thermal properties in the presence of a physiological concentration of KCl (150 mM) were investigated. First, we followed the variation of the molar ellipticity at 265 nm as a function of the temperature, by applying a heating ramp of 50 °C/h (figure 4.18 A). The collected profiles were not superimposable and intense hysteresis was observed. Multiple transitions on both curves were observed and a sensible increment of the molar ellipticity was observed at the end of the thermal cycle. For

reference the CD spectra in absence of KCl and then in presence of KCl before and after the thermal treatment are reported in figure 4.17 B. By comparing them, it emerged that the thermal treatment remarkably affects the G4 foldings. Indeed, after the annealing step in presence of KCl, the intensity at the positive contribution at 265 nm markedly increases, while the contribution at 295 nm slightly decreases. Nevertheless the TDS in presence of KCl before and after the annealing cycle (figure 4.18 B) overlap quite well and both remind to a G4 structure. Indeed a negative contribution at 299 associated to a two positive contributions at 274 and at 247 nm was observed. This fact supports the hypothesis of the coexistence of multiple G4s in solution before and after the melting/annealing. Since in presence of high concentration of potassium chloride a non-specific DNA-cation interaction was observed, we decided to reduce the KCl concentration (100 mM). By using the same heating rate two different behaviours were observed at the two major positive contributions (265 nm and 295 nm) (figure 4.18 C and D). At 295 nm, the melting and annealing curves overlap quite well in a temperature range between 69-95 °C, while a slight reduced CD intensity was observed during the annealing step. However, it is worth of note that the relative variation of the curves completely overlaps, and a Tm of 69.8 ± 0.6 °C was derived. This behaviour suggests that the species that contributes at 295 nm reversibly melts however its relative abundance in solution is reduced after the annealing. Conversely at 265 nm, the observed behaviour reminds the data acquired in presence of 150 mM of KCl. In fact, although less intense, the hysteresis between the curves was still observed. The intensity of the signal in terms of molar ellipticity was higher after the annealing cycle. Thus supporting the redistribution between the parallel and the hybrid contributions. To check if kinetic affects the parallel G4, melting and annealing at 100 mM KCl were acquired at a slower heating rate (25 °C/h). Again, the melting and annealing curves collected did not overlap, the hysteresis was observed and the intensity of the signal at the end of the process was higher compared to the value observed after the addition of KCl.

By comparing the curves at the two heating rates, it was clear that our system was not strongly affected by the selected heating ramp. While, the cation concentration affects the thermal stability of the system under investigation. Additionally, considering the increment of the recorded signal after the entire process, the redistribution towards different G4 conformations was confirmed (figure 4.19).

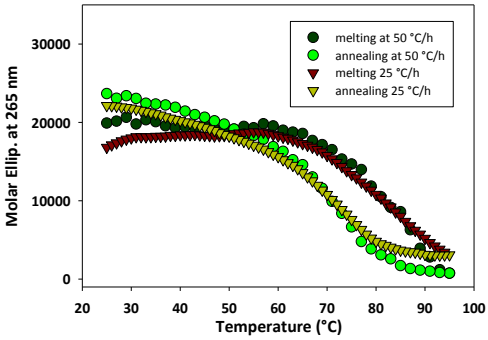


Figure 4.19. Comparison of melting and curves collected at 265 nm upon the addition of 100 mM using different heating rate (50 °C/h or 25 °C/h).

Table I Tm and Ta derived from the curves acquired in 100 mM KCl at 25 or 50 °C/h heating rate at 25 nm and 295 nm.

	265 nm	**265 nm**	**295 nm**
	25 °C/h	**50 °C/h**	**50 °C/h**
Tm (°C)	88.2 ± 1.1 °C	81.4 ± 0.6 °C/h	69.8 ± 0.6 °C/h
Ta (°C)	69.7 ± 0.7 °C	68.7 ± 0.6 °C	69.8 ± 0.6 °C/h

The G-quadruplex structure can exist as monomeric or multimeric forms that are expected to generate CD signal presenting differences in terms of intensity [160]. To check, the multiplicity of the species promoted by the presence of KCl and their redistribution upon the annealing, an electrophoretic mobility shift assay (EMSA) in native condition and increasing KCl concentration was set up.

Since the DNA molecules have a constant mass/charge ratio their migration into a native polyacrylamide gel is largely affected by their hydrodynamic volume. Particularly, species with a higher volume will migrate slower, it derives that monomolecular G4, that are more compact than the linear DNA, will run faster in the gel, while dimeric or multimeric forms will present a slower electrophoretic mobility. Samples were analysed immediately after the addition of metal ion, as well as, after an annealing step with variable KCl concentration (10 min at 95 °C and slow cool down to room temperature). As reported in figure 4.20 the addition of KCl provides the formation of different species. The presence of monovalent cation in the running buffer allows the DNA sequence to fold into an intramolecular arrangement even for sample prepared in the absence of KCl. Indeed, the oligonucleotide runs faster that a 22 bases linear oligonucleotide. Addition of KCl in the samples induces the formation of an additional form that we can attribute to the dimeric G4. Conversely, for the annealed sample in KCl, three different species were detected, the M form and the D form that disappear at the higher KCl concentrations, and the tetrameric ones (T) that

results mostly promoted by the annealing step (Figure 4.20). By merging all information acquired it is clear that the addition of KCl induces the formation of a fascinating system, in which multiple species, monomeric, dimeric and tetrameric may coexist depending on the working condition applied.

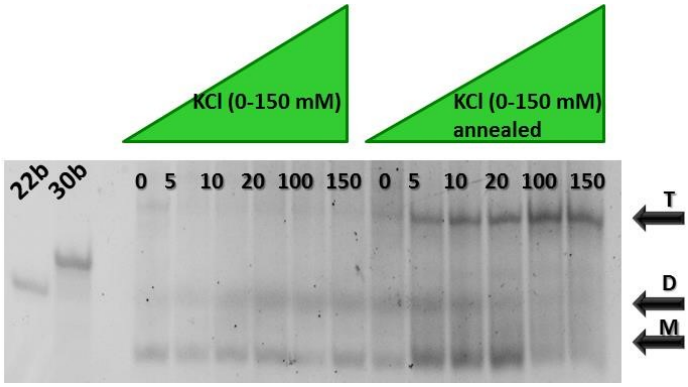


Figure 4.20. 15% native EMSA assay for dBcl2_G in increasing KCl concentration run in 1xTBE buffer containing 1 mM KCl. Samples were annealed in Tris 10 mM pH 7.4 and loaded immediately upon the addition of KCl or after and annealing step in KCl.

Since other monovalent cation can promote the folding of this sequence, we investigated the effect of Na^+.

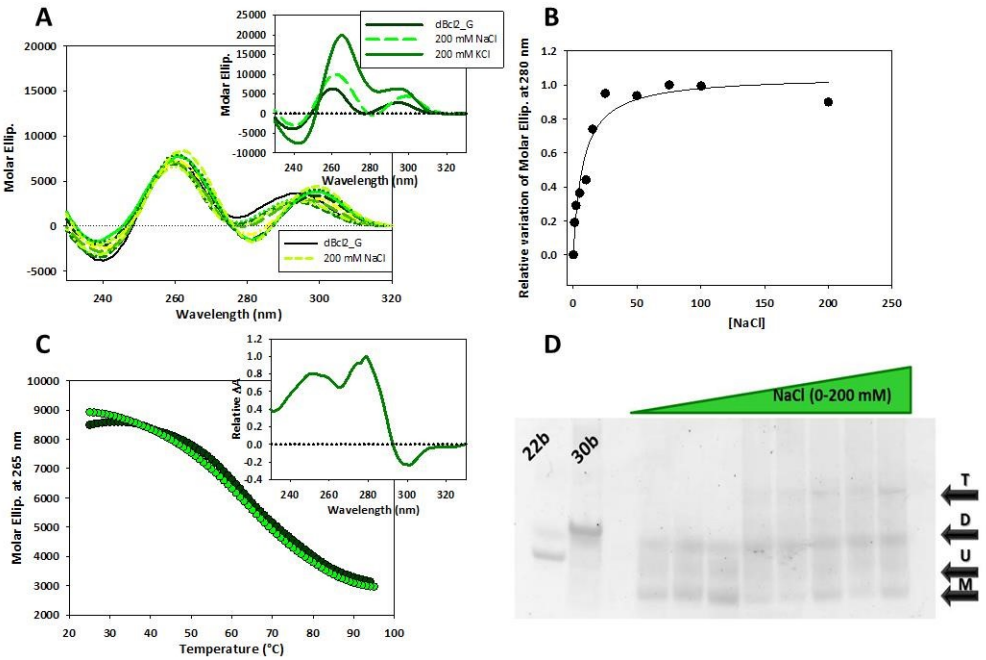


Figure 4.21. Effect of NaCl on dBcl2_G. (A) CD titration of 4 µM of dBcl2_G with NaCl in 10mM Tris pH 7.4. (B) Relative variation of the molar ellipticity at 280 nm as a function of NaCl concentration. (C) Melting and annealing curves recorded in presence of 200 mM NaCl, TDS calculated in presence of 200 mM NaCl. (D) Native EMSA assay in increasing concentration of cation. Before loading all samples were annealed in Tris buffer, then increasing concentration of cation was add. Buffer used 1x TBE.

CD titration of dBcl2_G with increasing Na+ concentration (figure 4.21 A) revealed that this cation only modestly modified the CD features of dBcl2_G. Nevertheless, a CD spectrum, having two main positive contributions, was observed and TDS confirmed that the addition of the sodium chloride promotes the switching towards a G4 conformation. Differently from the KCl, here the variation of the signal at 265 nm and 295 nm upon the addition of the cationic species was extremely low. Therefore we decided to consider the variation of the signal at 280 nm as a function of the cationic concentration. By fitting the data according to a one binding site model a Kd of 7.5 ± 1.8 μM was derived. Moreover, the saturation is reached at about 75 mM of the cation. Again, differently from what observed in presence of KCl, the melting and annealing profiles overlap quite well. Accordingly, a melting temperature of 66.2 ± 1.4 °C was derived. Finally, to evaluate the number of species in solution we performed an EMSA assay in increasing cations concentration. Since the annealing cycle slightly alter the conformational equilibria in solution, all samples were analysed immediately after the addition of metal species (figure 4.21 C). As above reported, in presence of increasing potassium chloride multiple species were detected confirming the ability of dBcl2_G to assume several G4 conformations. In particular at high ionic strength. Moreover, it appears evident that the Na+ cation is sufficient to drive the G4 folding.

So far acquired data evidenced that the selected guanine rich sequence, depending by the working conditions (annealing cycle, presence of KCl or NaCl), generates a very complex system in which multiple conformations (monomeric, dimeric and tetrameric), coexist in solution. It is worth of note that the herein presented data are not completely in line with those previously reported in literature by Zhang et colleagues [145]. Specifically, in 2010 they took into consideration the conformational equilibria in presence of KCl, NaCl and PEG_{200} of a truncated version of our sequence (22-nts). For this tract, they revealed the ability of NaCl to promote the formation of a single antiparallel G4, while a mixed structure was suggested in the presence of KCl. Also, no indication regarding the formation of a multimeric conformation was been derived. Despite our data results partially in contrast, it is worth to note that dBcl2_G contains 3 more guanines, this means that reasonably the conformational equilibria can be strongly modulated by their presence. Bearing this in mind, we decided to expand our characterization study considering the effect of flanking region on the conformational folding towards the G4 structures. Particularly, our efforts were focused on the possibility to get a conformational selection towards a single G4 structure.

4.2.2. dBcl2_G + 3WC

As anticipated, the sequence context around the minimal G4 forming sequence largely affects the conformational selection towards a single structure.

Therefore, we decided to scan the entire 5'-untranslated region with a bioinformatics tool. Then, we focused our attention on the adjacent portion to the already characterized guanine rich tract.

In this contest our sequence appears slightly symmetric; therefore, in a collaboration with Professor Toppo group, at the Department of Molecular Medicine at the University of Padova, we scanned the entire 5'-untranslated region looking for a longer sequence, where the symmetry of dBcl2_G is maintained and with a higher ability to assume into a G4 conformation. To do this a customizable imperfection-tolerant algorithm based on dynamic programming recently developed, and known as NeSSie (Nucleic acids elements of Sequence Symmetry identification) was applied[161].

From this in silico evaluation, several sequences were derived and among them, a 33-nts sequence, with the highest ability to folds into a G4 conformation was selected. It is characterized by the presence of 6 additional nucleobases potentially involved in the formation of 3 Watson-Crick base pairs...

dBcl2_G+3WC: 5'-CTC T**GG GGG CCG TGG GGT GGG AGC TGG GG**C GAG-3'

As for the dBcl2_G sequence, the conformational equilibria of the dBcl2_G + 3WC and its ability to assume or not a G4 arrangement were evaluated in specific experimental conditions (KCl or LiCl). Before use, again, each DNA sample was diluted from a stock solution to the required concentration in 10 mM Tris pH 7.4 and annealed.

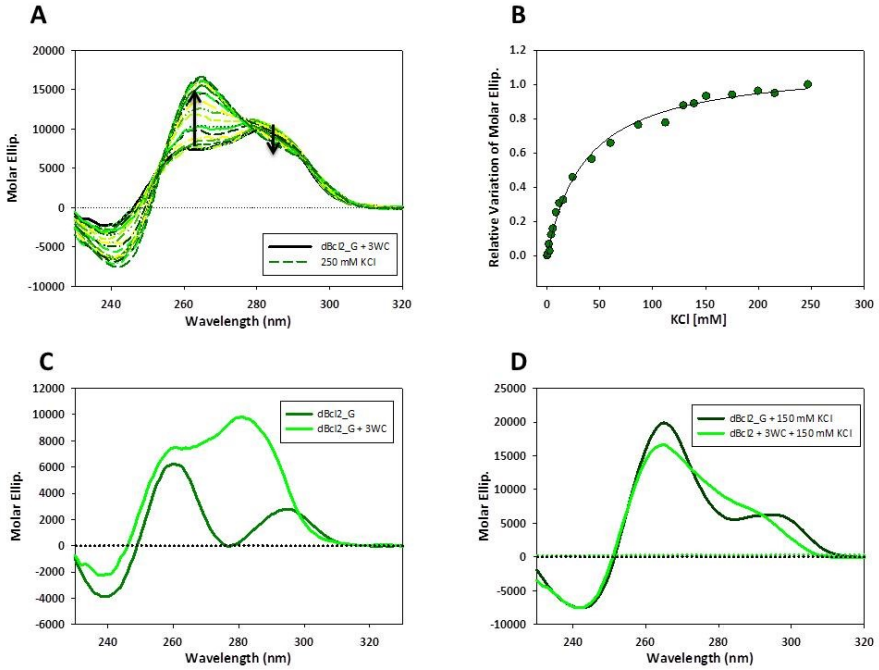


Figure 4.22 (A) CD titration of 4 µM dBcl2_G + 3 WC in 10 mM Tris pH 7.4 with increasing concentration of KCl. (B) Variation of the molar ellipticity at 265 nm plotted against the cation concentration.(C) Comparison of the spectra recorded in Tris 10 mM pH 7.4 for dBcl2_G and dBcl2_G + 3WC. (D) Comparison of the spectra recorded in presence of 250 mM KCl for dBcl2_G and dBcl2_G + 3WC.

As reported in figure 4.22 A and C the CD spectrum recorded in Tris 10 mM pH 7.4 (solid black line) was characterized by the presence of two positive bands centered at 260 nm and 282 nm. Also, the intensity of this signal compared to the one derived for the shorter sequence appears extremely high, suggesting that dBcl2_G + 3WC is already structured even in the absence of KCl. Upon the addition of KCl (dark green line), a sensible variation both in terms of shape and intensity of the signals was observed. In fact, a 5 nm blue shift associated with an increment of the CD intensity at 265 nm was detected, while the intensity of +ve contribution at 282 nm first slight increased but then it was reduced. In order to evaluate if KCl provide the same conformational shifting on the two guanine rich sequences under investigation, we compared the spectra obtained at the same KCl concentration (150 mM). As reported in figure 4.21C clearly the two obtained features are different (figure 4.22 D). Therefore, the KCl does not provide the shift toward the same conformation on the selected oligonucleotides.

By plotting the relative variation at the most relevant +ve contribution (265 nm) as a function of KCl concentration it emerged that the saturation of the process was reached at about 250 mM of cation, a value extremely higher compared to the data collected for dBcl2_G. By fitting the data according to a one site saturation binding model dissociation constant for the KCl of 39.3 ± 3.9 mM was derived (Figure 4.22 B).

In the previous chapter, the ability of dBcl2_G to assume multiple conformations in solution was proved. Since the longer sequence was selected with the aim to simplify the system in solution. A native EMSA assay, in increasing KCl concentrations, was performed to check the number of species in solution. (Figure 4.23).

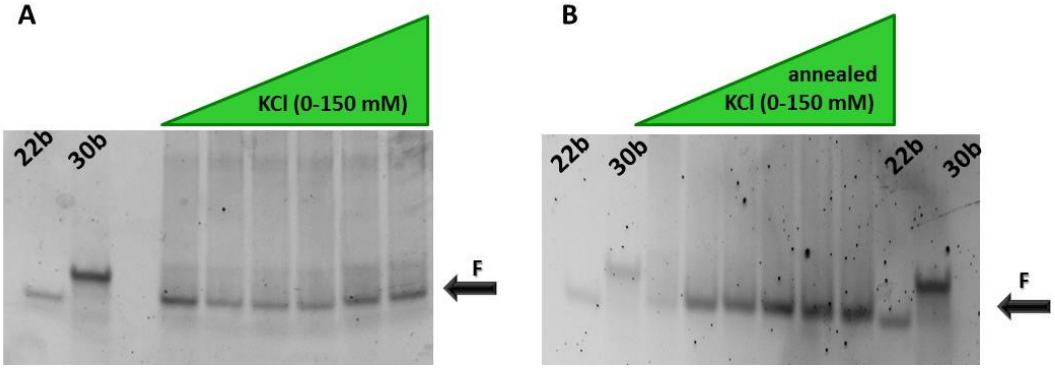


Figure 4.23. 15% native EMSA assay for dBcl2_G + 3WC with increasing KCl concentration. (A) Samples were loaded immediately upon the addition of KCl. (B) Samples loaded after an annealing step in KCl. Buffer used 1x TBE containing 20 mM of cation.

Again, the observed behaviour in presence of KCl of dBcl2_G + 3 WC was quite different from the one observed for the shorter sequence (dBcl2_G). Indeed, even for the annealed samples, a single electrophoretic band was detected, that runs comparable with the selected 22-nts marker. This indicated that dBcl2_G + 3WC folds into a more compact secondary structure. Furthermore, no multimeric species were ever observed. This likely suggests that the presence of the 3 canonical pairings that prevent the formation of multimeric structures.

To verify that the selected sequence generates in solution a less complex system its thermal melting profile was evaluated working at 150 mM of potassium chloride (figure 4.24). Melting and annealing curves were recorded using a 50 °C/h heating rate as previously made for dBcl2_G, and the CD spectra before and after the entire thermal treatment were acquired. As reported in figure 4.24 A, the CD spectrum collected at the end of the thermal cycle is extremely different from the one collected immediately after the addition of KCl as well from the CD signature acquired in Tris buffer. In contrast to the data for the shorter sequence, a less intense signal at 265 nm was detected associated to a small increment of the positive contribution at 282 nm. Also, considering the melting and annealing curves recorded at 265 nm a remarkable hysteresis was observed.

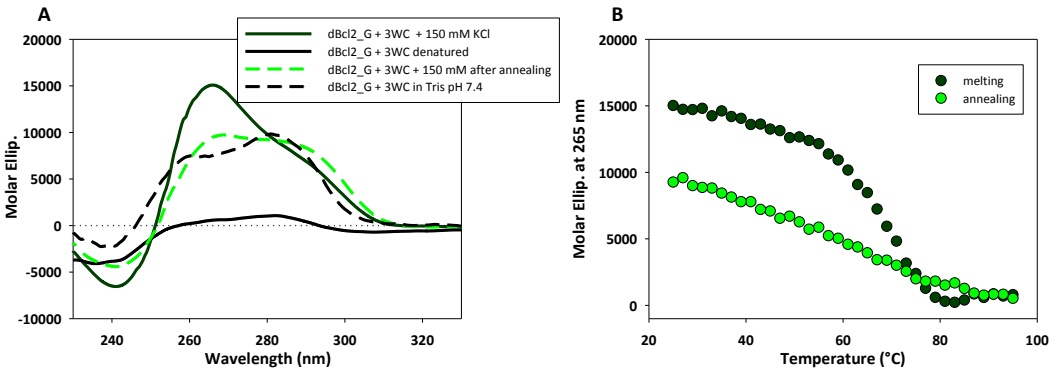


Figure 4.24. Evaluation of the thermal properties in presence of 150 mM of KCl. (A) CD spectra recorded before and after the thermal treatment. (B) Melting and annealing curves recorded at 265 nm.

In analogy to previous studies, the effect of lower potassium chloride concentration (100 mM KCl) on the thermal melting profile was considered. By applying a 50 °C/h heating ramp melting and annealing curves were recorded at 265 nm, which is the most relevant contribution on the CD features and also at 295 nm. Collected data are reported in figure 4.25

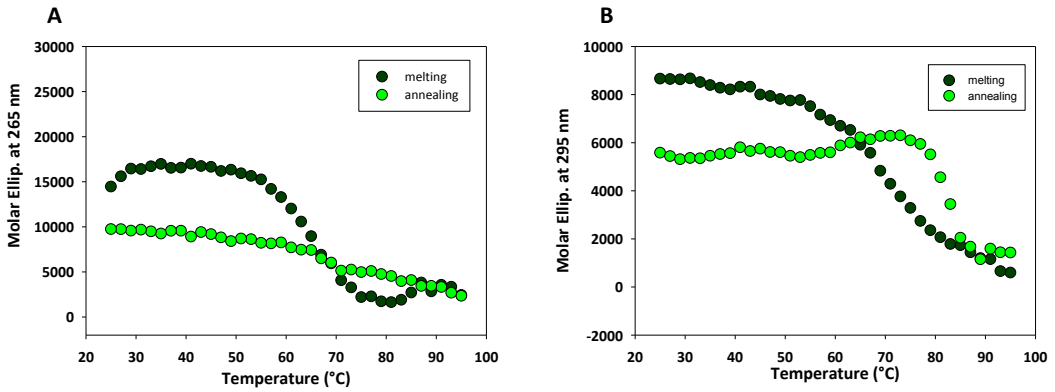


Figure 4.25. (A) Meting and annealing curves collected at 265 nm in presence of 100 mM KCl. (B) Melting and annealing curves acquired at the cation concentration at 295 nm.

Generally, the behaviour observed partially resembles the one reported in presence of 150 mM of KCl. Worth of note for this sequence the hysteresis was still present both at 265 nm and 295 nm, and multiple transitions can be observed.

To complete the evaluation of the thermal properties of the longer sequence in presence of KCl, as for the dBcl2_G, we set up the annealing experiment in presence of 100 mM KCl and by applying a heating ramp of 25 °C/h (Figure 4.25). Again the collected curves did not overlap and the hysteresis is observed.

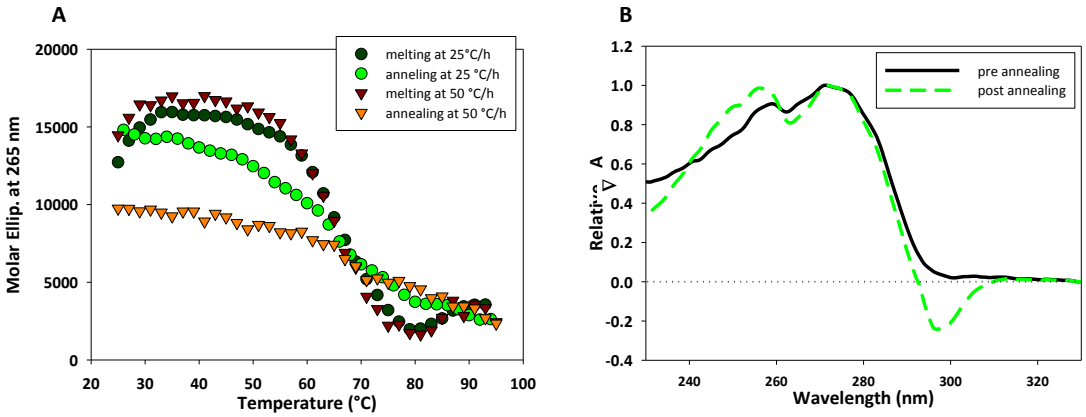


Figure 4.26. (A) Comparison of the melting and annealing curves recorded at the same cation concentration but at two different heating rates (50 °C/h or 25 °C/h). (B) TDSs calculated in presence of KCl before and after the annealing cycle performed at 25 °C/h.

By comparing the acquired curves at the two selected heating rates it emerged that the heating rate doesn't affected the melting profiles but remarkably change the corresponding annealing curves in the 65-25 °C range.
This indicates that the denaturation step is fast but during the renaturation step, based on the heating ramp a different structures are entrapped. In order to evaluate which type of secondary structures were involved during the thermal process; the TDSs were derived (figure 4.26). Clearly the two calculated result extremely different. In particular, before the melting/annealing two positive contributions at 258 nm and at 273 nm, while after the annealing, a spectrum that reminds the G4 structure was observed.

Considering all the collected data in presence of KCl, it appears that the potassium promotes on the tested nucleic acid a conformational rearrangement. However, the formation of the tetrahelical the G4 structure is promoted only by an annealing cycle in presence of the cation. This intriguing behaviour may related to the presence of the 3 WC base pairs, that contribute both to the CD signal and to the thermal profile of dBcl2_G + 3WC or it can be related to the ability of this sequence to adopted different conformation, sufficiently stable at room temperature, that partially prevents the G4 folding. To evaluate the existence of this non G4 structural rearrangement, the effect of lithium chloride, that is the best know cationic species able to stabilize the stem-loop conformation was assessed. First, a CD titration was performed, and then t the variation of the CD signal at 265 nm as a function of the temperature was followed. Finally, two TDSs were derived.

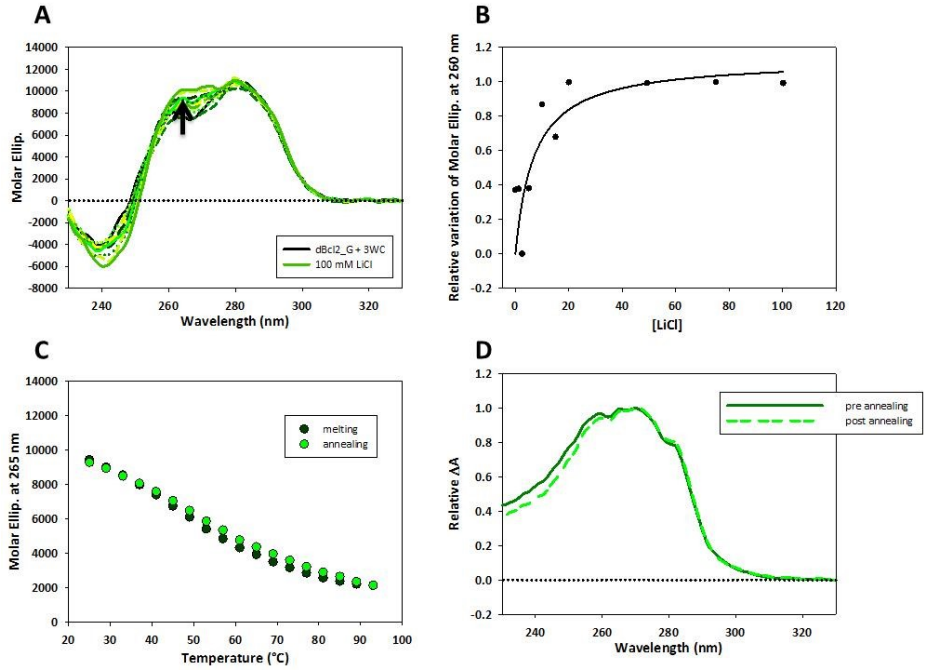


Figure 4.27 (A) CD spectra of 4 µM of rBcl2_G in 10 mM Tris pH 7.4 upon the addition of increasing LiCl concentration. (B) Relative variation of the molar ellipticity at 260 nm as a function of LiCl. (C) Melting and annealing curve collected in presence of 100 mM of LiCl. (C)TDS calculated before and after the thermal treatment in presence of metal species.

As reported in figure 4.27 A, the addition of LiCl induced only a small variation of the CD spectrum, in particular, with a modest increment of the intensity at 260 nm. The relative variation of the signal at 260 nm was considered, and by fitting the data, a Kd of 6.98 ± 3.9 µM was derived towards the cation. This suggests that the LiCl, as well the KCl, alters the conformational equilibria of dBcl2_G + 3WC. The thermal profile in presence of this cation and the TDS before and after the annealing cycle were considered. As reported in figure 4.27 C the melting and annealing curves completely overlap, suggesting the fully reversibility of the annealing cycle. Whereas from the shape of the two TDS derived it emerged that melting/annealing doesn't affect the structure assumed by dBcl2_G + 3WC in LiCl. Moreover the overall shapes did not remind to a G4 conformation.

The herein presented data revealed that the additional nucleotides allow the formation of 3WC base pairs, both in presence of KCl or LiCl. Furthermore, it appears that these additional pairings did not favour the G4 folding, indeed, the tetrahelical arrangement appears only after melting/annealing in presence of KCl. Finally, compared to the shorter sequence, no multimeric conformations (dimeric or tetrameric) were observed.

4.3. RNA G-quadruplex folding sequences: structural characterization

4.3.1. rBcl2_G

Since the DNA guanine rich sequence is contained into the sense strand, during the transcription process, a messenger RNA that contains the 25-nts PQS sequence within its 5'-UTR is generated. Again this stretch of guanine can easily lead the G4 folding. Its aptitude to assume this tetrahelical conformation was considered by applying the same experimental approach used for the DNA sequences previously mentioned.

rBcl2_G 5'-**GGG GGC** CGU **GGG G**UG **GGA G**CU **GGG G**-3'

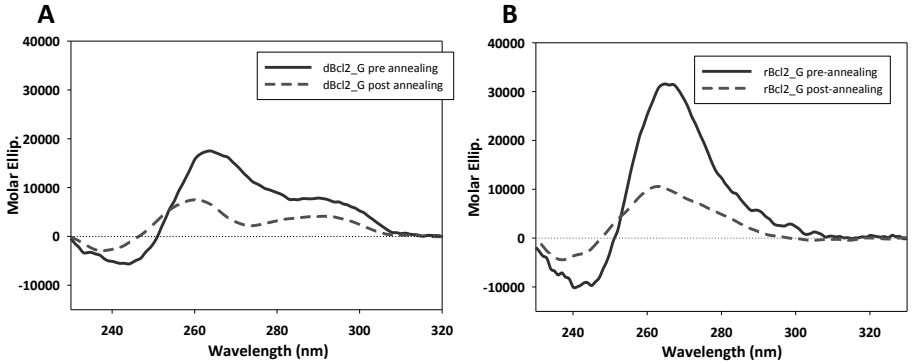


Figure 4.28 (A) CD spectra of 4 µM of rBcl2_G in 10 mM Tris pH 7.4 before and after the annealing cycle. (B)) CD spectra of 4 µM of dBcl2_G in 10 mM Tris pH 7.4 before and after the annealing cycle.

Based on the collected data for the DNA G-rich sequence, the existence of a pre-organized structure was taken into consideration. After dilution of the stock solution to the required sample concentration (4 µM) CD spectra were recorded before and after annealing. Figure 4.28 A reports the collected signals. Both spectra showed a single +ve band at 265 nm and a –ve one at 241 nm. However, a sensible drop of the intensity was observed. This suggests that also this RNA guanine rich sequence may adopt in solution a pre-organized structure that partially disappears after an annealing cycle
Moreover, both TDS shape derived before and after the annealing, showed two positive peaks at 250 nm and 277 nm associated to a negative contribution at 300 nm, thus suggesting that the pre-organized conformation in solution is likely a G4 structure, which was not completely removed after the thermal treatment (Figure 4.29).

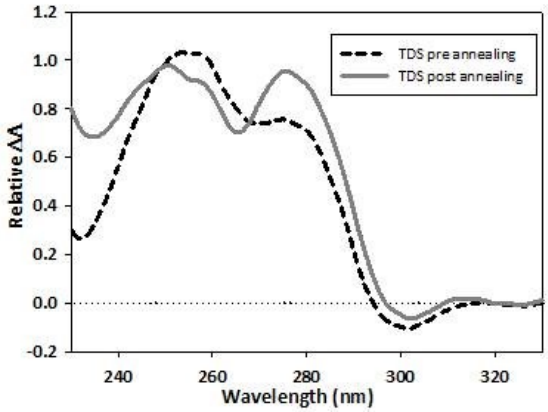


Figure 4.29. TDS derived for rBcl2_G in Tris 10 mM pH 7.4 before (black dotted line) and after the thermal treatment (grey solid line).

Considering the peaks position of the CD spectra of rBcl2_G a parallel G-quadruplex topology was suggested. This fits with the available literature data that indicate this as the only G4 topology for the RNA sequences.

Consistently to that, the evaluation of the effects exerts by KCl and LiCl on the conformational equilibrium involving this sequence was performed on annealed samples. Considering this set of data, in analogy with the corresponding DNA very low amount of cation might be sufficient to drive the conformational equilibrium towards the tetrahelical structure. Therefore, the evaluation of the effects exerts by KCl and LiCl on the conformational equilibrium involving this sequence was always performed on annealed samples. First, we evaluate the effect exerts by KCl on rBcl2_G (figure 4.30).

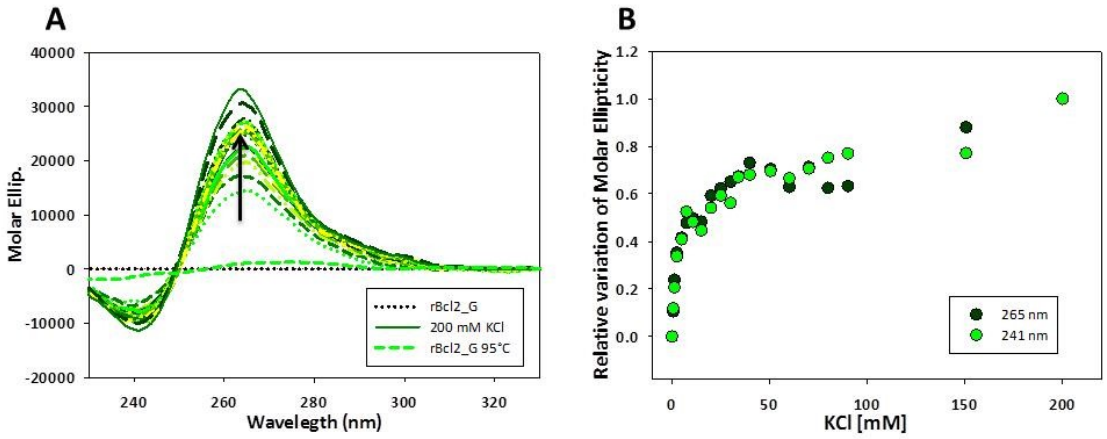


Figure 4.30 (A) CD titration of rBcl2_G in increasing KCl concentration. (B) Relative variation of the molar ellipticity at 265 and 241 nm plotted as a function of cation.

The collected data showed that potassium chloride markedly increased the intensity of the CD signal (Figure 4.30 A and B). The relative variation at the two wavelengths of

interest was plotted against the cation concentration. Differently from what observed for the DNA sequence, the detected behaviours were superimposable suggesting the simple conversion of the pre-organized arrangement towards a parallel G4 structure. Data fitted according to a one binding site saturation equation allow us to derive a dissociation constant of 3.99 ± 0.5 mM towards the KCl.

To fully describe our system we considered how this metal species affects the thermal stability.

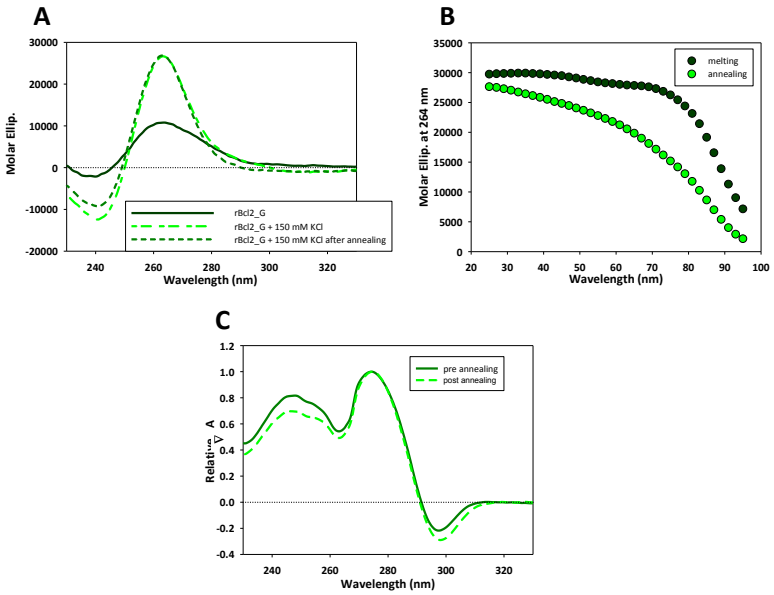


Figure 4.31 Thermal properties of rBcl2_G.in physiological concentration of KCl. (A) CD spectra of 4 µM of rBcl2_G recorded in 10 mM Tris pH 7.4 in absence/presence of 150 mM of KCl before and after the thermal treatment. (B) Melting and annealing curves recorded at 265 nm. (C) TDS derived before and after the melting/annealing.

The thermal profile recorded in 150 mM KCl using a scanning speed of 50 °C/h pointed out the remarkable stability of this parallel structure. In fact, clearly the denaturation process was not complete up to 95 °C. At the applied heating rate, melting and annealing curves were not superimposable, and a hysteresis phenomenon was observed (figure 4.31). This, unfortunately, do not allow us to derive a melting temperature. Nevertheless, the CD spectra acquired before and after the melting/annealing step nicely overlap as well as the two corresponding TDS. This might be indicate that for this sequence the hysteresis is not due the entrapping of different G4 topology but only to kinetic issues.

To clarify this point simply the experiment was repeated at lower cation concentration (100 mM) and at two different heating rates (50°C/h or 25 °C/h). As figure 4.32 A and B showed at both heating rate used, at a lower concentration of KCl the melting process appears as completed at 95 °C. However, the hysteresis between melting and annealing curves was reduced. A comparison of the collected melting curves, at the

same cation concentration but a different rating step was made (figure 4.32 C). Contrarily from the behaviour observed for the DNA G-rich sequence, here the two melting curves did not complete. Suggesting that the applied heating ramp was not sufficiently slow.

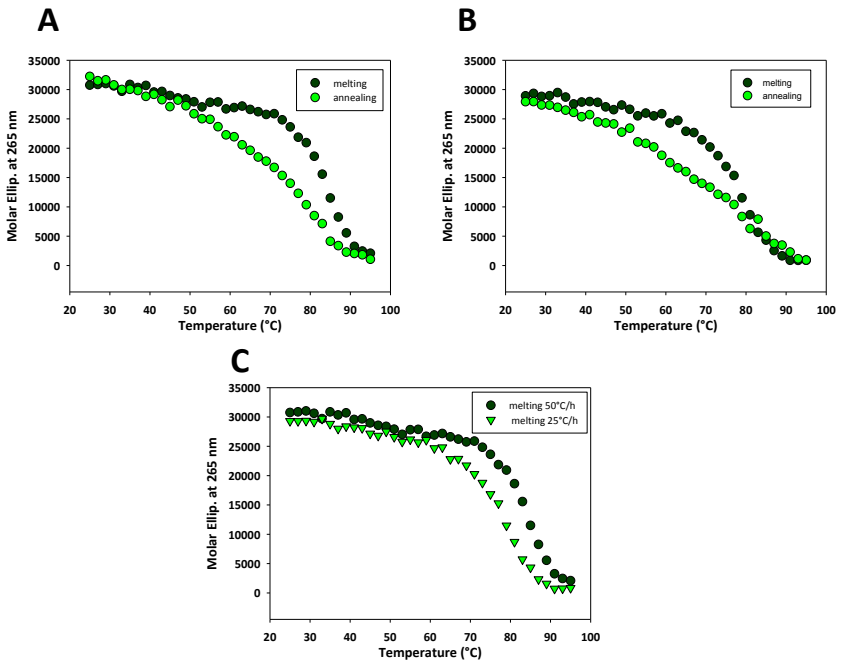


Figure 4.32 Thermal properties of rBcl2_G.in 100 mM of KCl recorded at different heating rate. (A) Melting and annealing curves recorded at 265 nm in 100 mM KCl at 50°C/h s heating rate. (B) Melting and annealing curves collected in the same cation concentration but at 25°C/h. (C) Comparison of the two melting curves collected at 265 nm.

To verify the number of species promoted by the presence of KCl a native EMSA assay was performed (figure 4.33). It emerged that sample preparation is crucial to gain the controlled modification of the conformational equilibria. Indeed, in absence of cation, as upon its addition, a single species with comparable electrophoretic mobility was observed. On the contrary, the annealing step after the addition of cation provides the formation of multiple species having different electrophoretic mobility. Therefore, as for dBcl2_G, our data revealed a complex system that contains multiple conformations depending on the working condition. Also, it is worth to note that these results are quite in contrast with the previous set of data reported by Balasubramanian et al in 2010. Although, they stated the ability of this domain to assume the parallel topology, they never observed multimeric conformation.

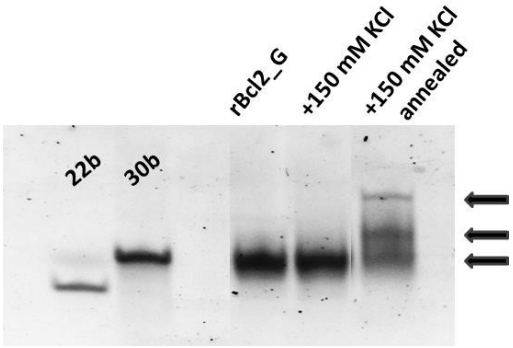


Figure 4.33 15 % native EMSA assay run in TBE 1x in presence of a physiological concentration of KCl. All samples were annealed in Tris 10 mM. After the addition of KCl, one sample was loaded immediately; the other was annealed in KCl presence.

To complete the structural characterization of the selected nucleic acid, the effect of LiCl was taken into consideration (figure 4.34).

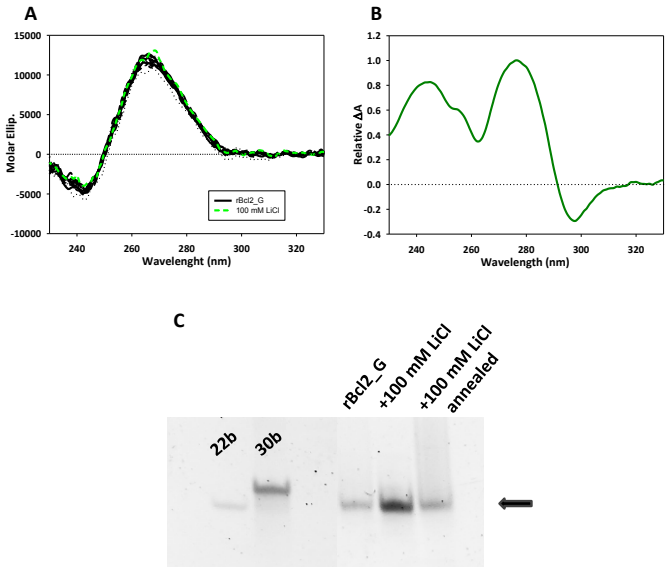


Figure 4.34 Effect of LiCl on RNA G-rich domain. (A) CD titration of 4 µM rBcl2_G in 10 mM Tris pH 7.4 in increasing concentration of LiCl. (B) TDS spectra calculated in presence of 100 mM of cation. (C) 15 % native EMSA assay in presence of LiCl. All samples were annealed in Tris 10 mM. Before the addition of LiCl, one sample was loaded immediately the other was annealed in LiCl presence.

CD titration revealed that the selected cation didn't promote any conformational change. Unexpectedly also the TDS did not show any variation thus meaning that the metal ion is not inducing the G4 but as well, is also not able to disrupt the residual G4 that is carried after annealing in Tris 10 mM. Consistently a native EMSA assay in absence/presence of lithium chloride revealed the presence of a single species having the same mobility of the 22-nts sequences selected as a marker for each tested

condition (figure 4.34 C). Therefore it is clear that the LiCl doesn't alter the hydrodynamic volume of the conformation assumed in solution by our sequence.

The data obtained so far revealed several similarities between the DNA and the RNA G-rich domains. Indeed, in an environmental condition comparable with those found in cell, both G-rich sequences are able to assume an extremely stable tetrahelical arrangement that can be converted into multimeric arrangements.

4.3.2 rBcl2_G +3WC

Since the minimal G4 forming sequence generates multiple species although with the same topology, by applying exactly the same experimental approaches followed for the DNA counterpart, we performed the structural characterization of the rBcl2_G + 3WC, characterized by the presence of 6 additional bases and a theoretical high propensity to assume the G4.

rBcl2 + 3WC 5'-CUC U**GG GGG CCG UGG GGU GGG AGC UGG GG**C GAG-3

The first step was to evaluate the presence/absence of a pre-organized arrangement. After a stock dilution the CD signal before and after the annealing cycle was recorded (figure 4.35). The recorded CD features shows a positive contribution at 265 nm, whereas after the renaturation step a blue shift, associated to a sensible reduction of the intensity of the signal was detected. This suggests the presence in solution of a pre-initial conformational possibly deriving from the stock solution and likely supported by traces of cations in the working buffer.

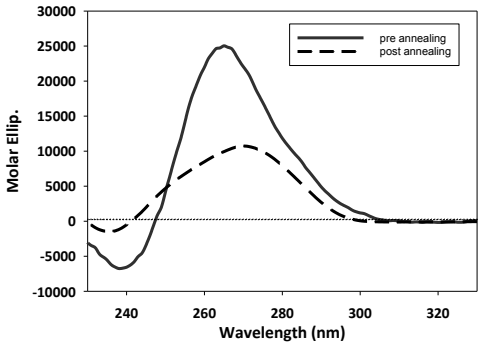


Figure 4.35. Melting and annealing curves, registered at 265 nm with a heating ramp of 50°C/h.

Each reported experiment was thus performed on pre-annealed samples (10 min at 95°C followed by a slow cool down to RT). First the effect of KCl was evaluated (figure 4.36).

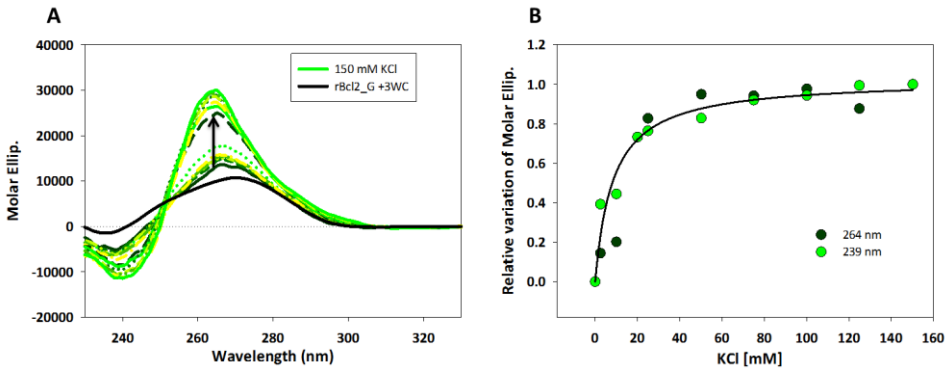


Figure 4.36. (A) CD titration in increasing KCl concentration. (B) Relative variation of the molar ellipticity recorded at 264 and 239 m as a function of the cation concentration.

As reported in figure 4.36 A the spectrum, recorded in 10 mM Tris shows a positive peak at 271 nm, while upon the addition of cation a 6 nm red shift associated to a sensible increment of the intensity of the overall CD shape was observed. This suggests that the selected metal species promotes conformational changes towards a parallel G4 topology. The relative variation of the molar ellipticity at 264 as well as at 241 nm were considered and plotted as a function of the KCl concentration. Similarly to the reported data for rBcl2_G sequence, the detected behaviour was superimposable By fitting the data using the one binding site model; a dissociation constant for the KCl of 12.19 ± 2.6 mM was derived. Thus the affinity for the metal ion was actually reduced. To complete the evaluation of the effects of KCl on rBcl2_G + 3WC, the thermal properties were evaluated in presence of a physiological concentration of cation (150 mM). Melting and annealing curves were recorded using a heating ramp of 50°C/h and following the denaturation process at 264 nm. In line with the shorter sequence, at 150 mM KCl the denaturation process was not complete up to 95 °C (data not shown). Surprisingly, even reducing the KCl concentration to 100 mM the unfolding was not complete at 95 °C as well. Therefore, we cannot derive a melting temperature. Considering the intensity of the signal at the end of the entire cycle a sensible increment of the molar ellipticity was observed, suggesting that the annealing step actually further promotes the G4 formation that, as reported in figure 4.37 C still preserve the parallel orientation. The same emerged from the TDS where the negative contribution at 298 nm appears as relevant only after the annealing cycle (figure 4.37 D).

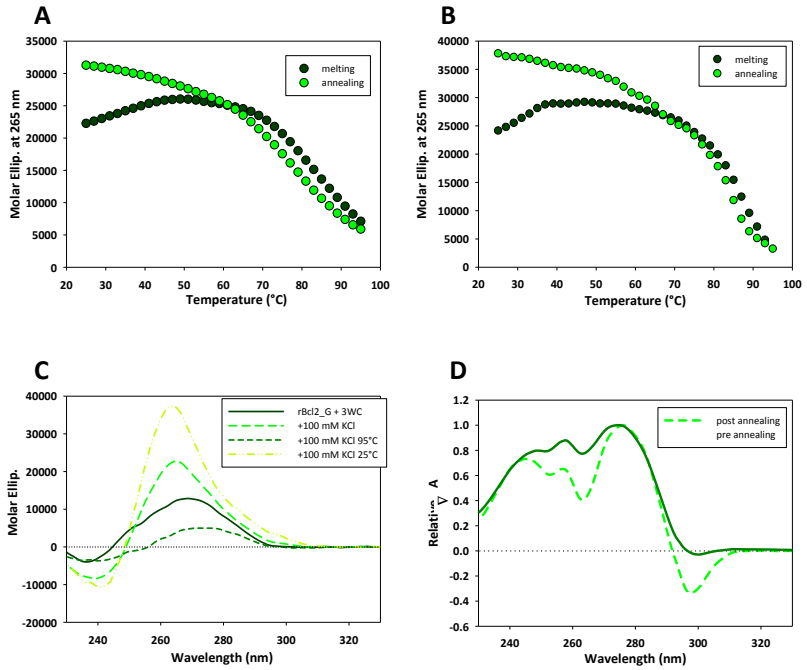


Figure 4.37. (A) Melting and annealing curves collected in 100 mM KCl with a heating ramp of 50°C/h. (B) Melting and annealing curves collected in 100 mM KCl with a heating ramp of 25°C/h. (C) Spectra recorded during the entire annealing cycle. (D) TDSs calculated before and after the thermal treatment upon the addition of 100 mM KCl.

This behaviour reminds the one previously observed for the dBcl2_G + 3WC and confirmed that the presence of 6 additional bases partially prevents the folding into G4. However, a native EMSA assay (figure 4.38) revealed the higher complexity of this system compared to the DNA counterpart (dBcl2_G + 3WC). In particular, it appears that the addition of KCl did not alter the number of species contained in solution, in fact, three main bands were identified both in the absence/presence of KCl and, it appears that their relative intensity change with the band having the slower mobility increasing after the thermal treatment.

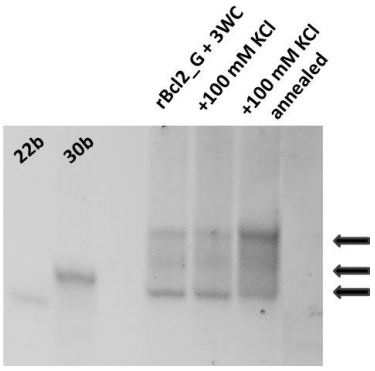


Figure 4.38. 15% native polyacrylamide gel in presence of 100 mM of KCl. One sample was loaded upon the addition of cation, the other one was submitted to an annealing cycle prior the loading.

To better understand our system, the effect of Lithium chloride on the conformational equilibria of the sequence was evaluated.

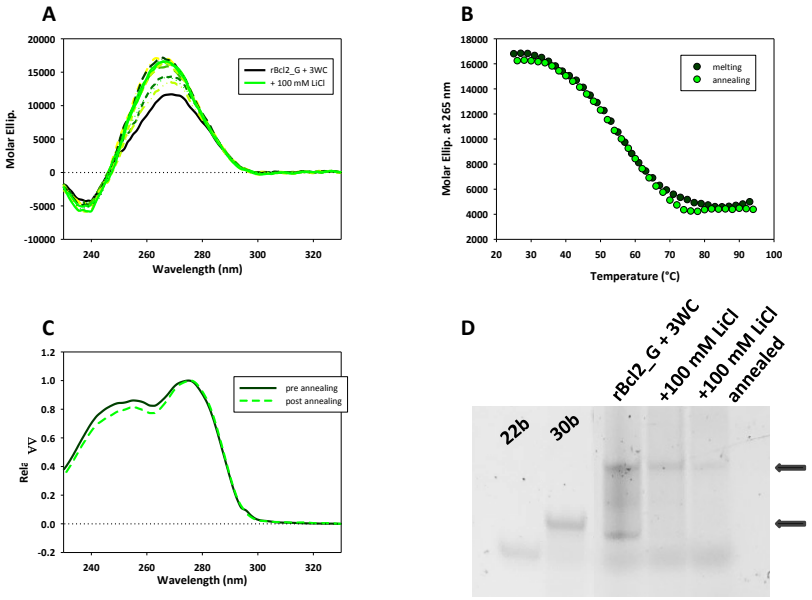


Figure 4.39. (A). CD titration in increasing LiCl concentration. (B) Melting and annealing curves collected for rBcl2_G + 3WC upon the addition of 100 mM LiCl. (C) TDSs calculated. (D) 15% native polyacrylamide gel in presence of 100 mM of LiCl. One sample was loaded upon the addition of cation, the other one was submitted to an annealing cycle prior the loading.

In contrast to what we observed on rBcl2_G, here the metal ion increases the CD signal at 266 nm (figure 4.39 A). This is quite similar to the one reported for dBcl2_G + 3WC and in line with the stabilization of the WC pairing. Indeed, the TDS derived did not remind to a G4 conformation (figure 4.39 C). Moreover, the melting and annealing curves acquired at a heating rate of 50 °C/h in 100 mM LiCl showed a perfectly reversible single transition with a Tm of 54.4 ± 0.2 °C (figure 4.39 B). Finally, as reported in figure 4.39 D, native EMSA assay showed that 100 mM Li+ shift the species distribution towards the form at lower electrophoretic mobility, irrespectively of the presence/absence of the annealing step.

By merging all collected data, we can suggest that this specie might derive from the WC pairing of two strands that is stabilized either by potassium and lithium.

4.3.3 rBcl2_48

As we already proved, the conformational equilibrium and the aptitude to folds into a G4 conformation of the PQS is strongly modulated by the presence of additional nucleotides at 3' and 5' ends. In particular, literature data evidenced the formation of secondary structures alternative to G4, as the stem-loop conformation (hairpin), that might competes with the tetrahelical one and having a physiological impact and relevance on the gene expression as well[162].

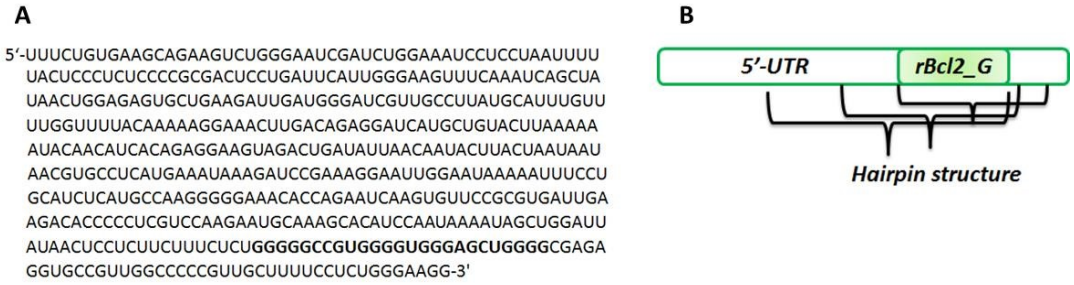


Figure 4.40 5'-UTR domain within *BCL2.* (A) UTR nucleotide composition. (B) Schematic representation of the 5'-UTR domain and possible secondary structure contained.

Bearing this in mind we considered the entire 5'-UTR portion and we searched for the presence of other longer domain comprising rBcl2_G that can possibly form a hairpin structure. By using mFold as a bioinformatics tool [163], we fractionized in a 50-nts long slots the entire 5'-untranslated domain and for each selected tract the ability to assume a stem-loop conformation was considered. Thanks to this approach several fragments able to assume an intramolecular stem-loop conformation were derived. Then, we restricted the area of interest to the portion that contains the selected PQS. The same bioinformatics tool was applied by changing the position in which the PQS is located within the 50-nts fragment. Again for each stretches we derived the theoretical stability of the intramolecular pairing and based on the score obtained a longer sequence, composed by 48-nt and characterized by the higher aptitude to assume a stable stem-loop conformation, was choose. (Figure 4.41).

rBcl2_48: 5'-U**GG GGG CCG UGG GGU GGG AGC UGG GG**C GAG AGG UGC CGU UGG CCC CCG-3'

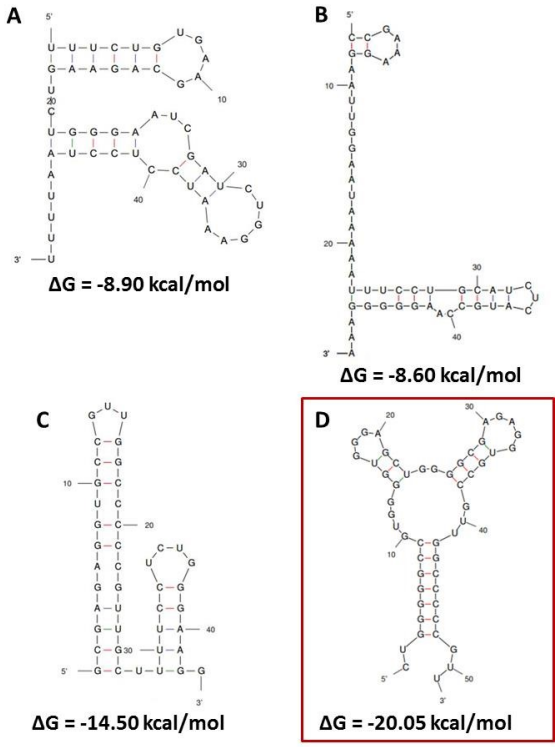


Figure 4.41 Example of non canonical secondary structures derived during the bioinformatics analysis of the entire UTR domain. (A) sequence 1-50 nts. (B) 76-126 nts. (C) 490-493 nts. (D) 425-475 nts.

The conformational equilibria of this sequence (rBcl2_48) and its propensity to balance between G4 or a hairpin were explored by spectroscopic and electrophoretic techniques in absence or presence of selected cations (KCl or LiCl).

The first step of our study was again evaluate the existence of a pre-initial structures, therefore a CD signals were recorded in 10 mM Tris pH 7.4 before and after an annealing cycle (figure 4.42)

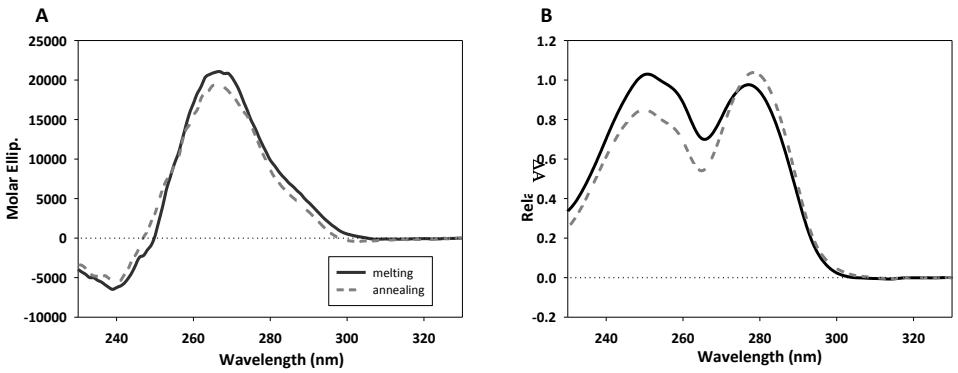


Figure 4.42. Effect of the Tris buffer on the rBcl2_48 (A) CD spectra recorded in Tris 10 mM before and after the annealing cycle. (B) TDs derived before (solid line) and after (dashed line) the thermal treatment.

For this sequence the two CD spectra nicely overlap. However, they were characterized by the presence of a strong positive peak at 265 nm and a negative one at 241nm that might refer to a G4 conformation. This was unfortunately ruled out by the TDS which shape did not correspond to the differential spectra generally associated to a G-quadruplex conformation, since the negative peak at 295 nm is missed. From this preliminary set of data a clear indication regarding the pre-initial conformation with no involvement of the G4 arrangement was obtained.

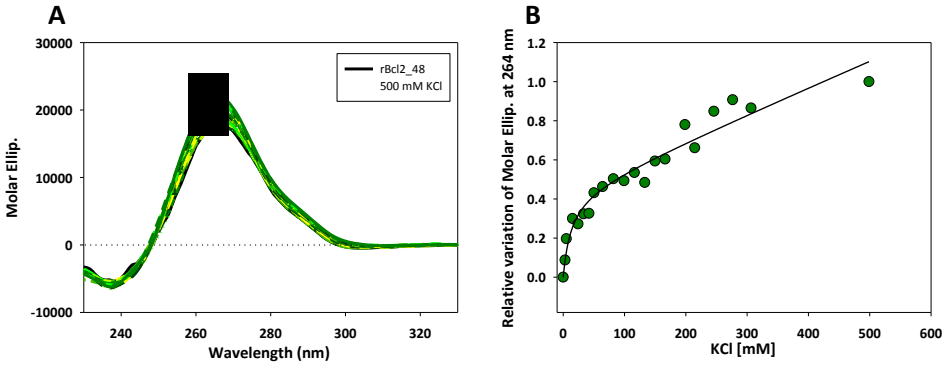


Figure 4.43. Effect of KCl on rBcl2_48 conformational equilibria. (A) CD titration in 10 mM Tris buffer pH 7.4 in increasing cation concentration. (B) Relative variation of the molar ellipticity at 264 nm.

Then we moved to evaluate the effect of KCl by using circular dichroism. As reported in figure 4.43 A the CD signal recorded upon the addition of KCl up to 500 mM revealed that the metal species did not extremely modulate the conformational balance of our oligonucleotide. Indeed, no sensible variation in the intensity of the CD feature recorded, as well as the shape. By plotting the molar ellipticity recorded at 264 nm as a function of the cation concentration a trend that remind of a non-specific interaction in presence of higher KCl concentrations was observed

To better understand this system we evaluated its thermal properties. In 150 mM KCl the melting and annealing curves modestly deviated (data not shown), whereas at lower cation concentration (100 mM KCl), the recorded data evidenced the completely reversibility of the annealing process and a Tm of 76.5 ± 0.5 °C was derived. Under this experimental condition the process resulted in equilibrium since a further reduction of the heating rate (25 °C/h) provides a similar thermal profile and melting temperature (77.4 ± 0.1 °C) (figure 4.44). Accordingly, TDS did not change and a native EMSA assay in TBE 1x buffer and in absence/presence of 100 mM of KCl. Data acquired did not highlights the formation of species at differ electrophoretic mobility in the sample annealed or not (data not shows).

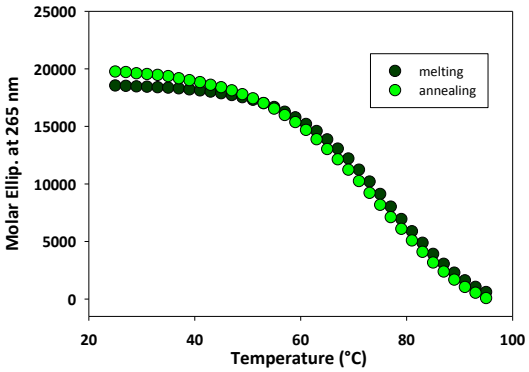


Figure 4.44. Effect of KCl on the rBcl2_48 thermal properties Melting and annealing curves collected at 265 nm in presence of 100 mM cation at a 25°C/h as heating rate.

Table II Tm derived from the curves acquired in Tris 10 mM, 100 mM or 150 mM KCl at 25 or 50 °C/h heating rate at 265 nm.

	Tris 10 mM 50 °C/h	**100 mM KCl 25 °C/h**	**100 mM KCl 50 °C/h**	**150 mM KCl 50 °C/h**
Tm (265 nm)	62.7 ± 0.6 °C	77.4 ± 0.1 °C	76.5 ± 0.5	84.8 ± 0.5

Considering the data acquired it emerged that the KCl doesn't promote the folding towards the G4 structures, despite a thermal stabilization effect was observed.
To understand the nature of the structure assumed by rBcl2_48 the same data were acquired in presence of lithium chloride (100 mM). Again, from the spectroscopic data (figure 4.45 A) no sensible variation in the CD signal, neither in terms of intensity or shape, was observed, and melting and annealing curves overlapped quite well, confirming the reversibility of the annealing process. By fitting the data as a function of the temperature a Tm close to 75 °C was derived that result comparable with the value derived in KCl. Also the TDS was reminiscent to the obtained trend in absence and in presence of KCl. To conclude the evaluation of the LiCl effect on rBcl2_48 a native EMSA assay in TBE 1x was performed. As reported in figure 4.45 B for each lane a single species, with the same electrophoretic mobility was observed, suggesting that the LiCl does not alter the hydrodynamic volume of the species contained in solution.

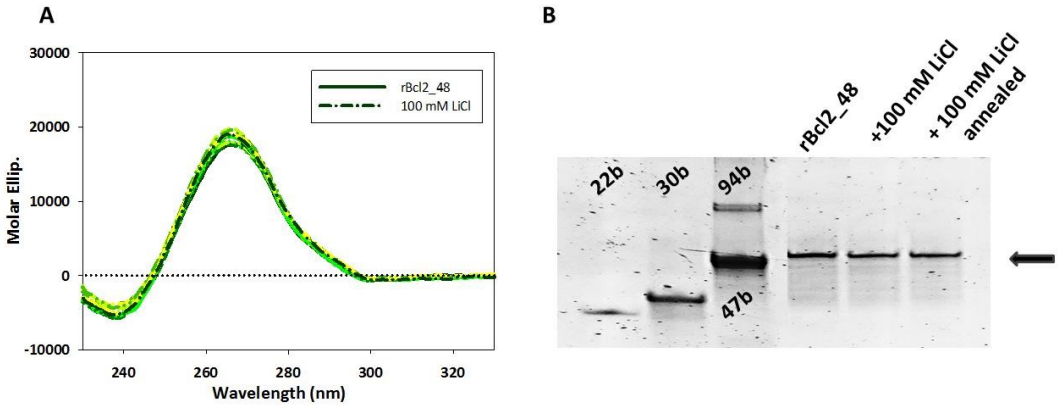


Figure 4.45 (A) CD titration of 4 µM rBcl2_48 in 10 mM Tris pH 7.4 in increasing concentration of LiCl. TDS spectra calculated at the cation concentration of 100 mM. (B) Native EMSA assay in presence of LiCl. All samples were annealed in Tris 10 mM. Before the addition of LiCl, one sample was loaded immediately after the addition of cation; the other one was annealed in presence of LiCl.

Considering all acquired data it emerged that the longer sequence is not able to assume the tetrahelical conformation neither in presence of potassium chloride and/or after an annealing cycle in presence of it. LiCl gave a comparable set of data. Indeed, no detectable variation in the CD shape and TDS were observed. Thus, both of them did not promote a conformational shift towards a different conformation. However, it is worth to note that both selected cations promote a thermal stabilization of the structure in solution. Therefore, all evidence supports and confirms the in silico analysis performed on our domain. This means that the insertion of the minimal G-rich sequence into a larger domain, prevent the formation of a G4 and promote the formation of an intramolecular stable hairpin structure.

4.4 DNA Cytosine-rich domain: structural characterization

4.4.1. dBcl2_C

The previous chapters (4.2 and 4.3) were focused on the structural characterization of the putative G-quadruplex forming sequences. However, physiologically the DNA is a double helix, thus a Cytosine rich domain is present on the complementary strand. For this gene, the C-rich domain is located in the antisense/template strand that contains the consensus site for the RNA Polymerase II. In 2014 Agarwal et colleagues[164] reported, that the presence of iM conformation in the antisense strand can creates a physical blockage for the enzyme, therefore the translation might be reduced. Keeping this in mind in this chapter the conformational equilibrium of dBcl2_C and its ability to assume the iM structure was explored.

dBcl2_C: 5'-**CCC C**AG CT**C CC**A **CCC C**AC GG**C CCC C**-3'

As reported in figure 4.46. A, in Tris 10 mM pH 7.4, the first recorded spectrum was characterized by the presence of a positive peak at 281 nm, a negative one at 236 nm with a cross-point at 260 nm, that reminds to a canonical DNA B-Helix. After the denaturation process, a 2 nm shift of the positive peak was observed (279 nm). However, the intensity of the CD signal recorded was maintained. Melting and annealing curves recorded at 281 nm overlap quite well. No sigmoidal profile was observed, suggesting that probably the conformation assumed in solution is not extremely stable. A native PAGE assay (figure 4.46 C) showed a single species with an electrophoretic migration comparable with the expected one for a 25-nts longer sequence. This might suggest that, differently from the guanine rich sequence, dBcl2_C did not assume a significantly stable pre-organized conformation.

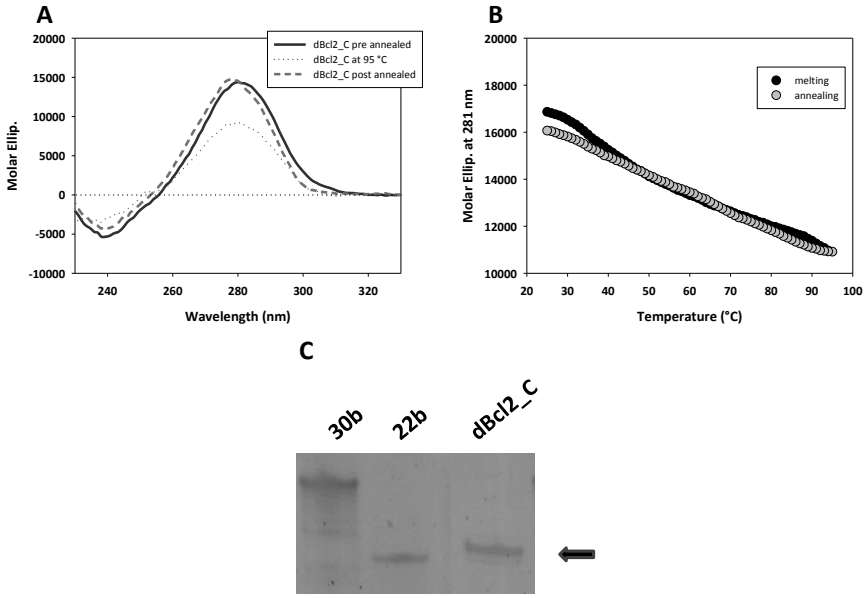


Figure 4.46 (A) Dichroic signal of 4 µM of dBcl2_C in 10 mM Na-Cacodylate at pH 7.4 before and after the thermal treatment. (B) Melting and annealing curves recorded for dBcl2_C. (C) 15% native PAGE assay in TBE 1x.

Even though this pre-folded conformation is not significantly stable, in silico evaluation using IDT OligoAnalyzer as a bioinformatics tool was set up to check the possible nature of this structure. As reported in figure 4.47, dBcl2_C might assume a poorly stable intramolecular pairing.

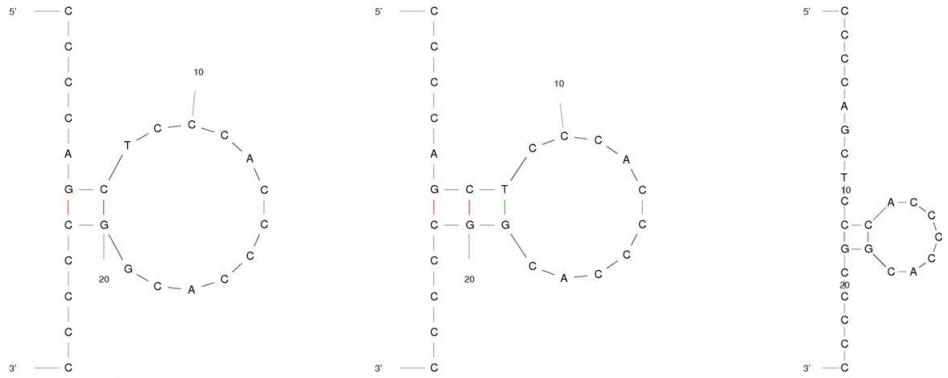


Figure 4.47 Possible intramolecular hairpin-like or conformation assumed by dBcl2_C calculated via IDT OligoAnalyzer tool with their relative theoretical stability.

The formation of the iM was subsequently followed working in a wide range of pH (8.3-2.8). The acidification of the environment strongly altered the recorded spectroscopic signal (figure 4.48). Specifically, by decreasing the pH from 8.3 to 5.5 a sensible increment of the intensity associated to a blue shift of the +ve peak at 288 nm and was observed. Comparing this trend with the CD dataset reported in literature[164,165], clearly, it reminds to an iM conformation. Moreover, according to the formation of the tetraplex, the addition of HCl up to pH 2.8 caused a sensible drop of

the signal as a consequence of the fully protonation of the cytosine bases[166]. Since the iM requires a specific pH condition to fold, from a pharmacological and physiological point of view is extremely important determines the transitional pH, which is the pH value where half of our sample is folded into this intercalated form. By plotting the relative variation of the molar ellipticity towards the pH value, a sigmoidal trend was observed, suggesting a cooperative folding process. By fitting the data according to y= B/ (1+ 10^ (n*(pH – pKa) equation, a transitional pH (pHt) value of about 6.6 was derived, a value quite close to the physiological one. As the cellular environment is extremely crowded, the pKa of the cytosines may change to allow the iM occurrence in vivo, as now confirmed by in-cell NMR spectroscopy and imaging with a selective antibody.

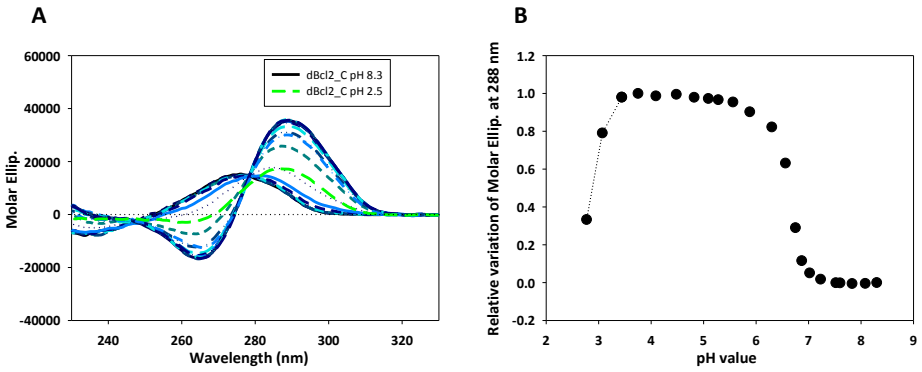


Figure 4.48 CD features recorded for dBcl2_C in 10 mM Na-Cacodylate during the pH titration (pH 8.3-2.5). (B) Relative variation of the molar ellipticity at 288 nm plotted as a function of the pH value

In addition we decided to follow the kinetic of the folding process. We acquired the CD spectrum at neutral condition (10 mM Na-Cacodylate pH 7.4), then HCl was added to get the transitional pH and the variation of the CD signal as a function of the time was evaluated (figure 4.49).

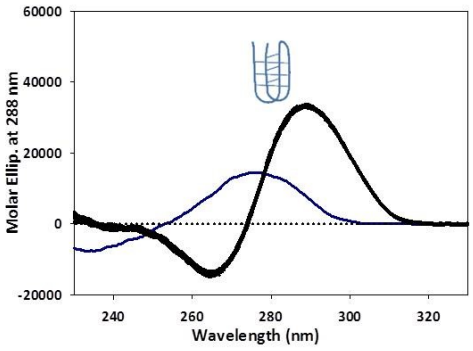


Figure 4.49.Kinetic experiment for dBcl2_C in 10 mM Na-Cacodylate upon the addition of HCl. The solid blue spectrum referred to the pre-folded conformation assumed at pH 7.4, the black solid line referred to the iM structure at the transitional pH (pH 6.6).

As reported in figure 4.48, by comparing the CD spectra recorded it appears that upon the addition of the cloridic acid immediately the pre-organized structure shifts towards the iM. Indeed, a blue shift associated with an increment of the intensity was observed. However, since the kinetic parameter was evaluated using a CD instrument, a lag time between the addition of the HCl and the first CD spectrum recorded exists. Thus, we cannot ensure that the iM folding process does not involve other intermediates.

The thermal profile at pH 5.5 is reported in figure 4.50 A. Melting and annealing curves recorded revealed the fully reversibility of the annealing process, moreover since a sigmoidal transition was observed a melting temperature of 53.0 ± 0.1 °C was derived.

Compared to the G4 conformation, the iM structure is considered less polymorphic, due to the spatial arrangement of the CC+ base pairs, the iM can assume just two different topologies: the 5'E when the outmost CC+ is located at the 5'end of the entire structure. While, if the terminal CC+ is at the 3' terminus of the structures the topology assumed is known as 3'E[39]. Also, as in the case of the guanine rich sequence, the iM can exist as a monomeric structure, dimeric or tetrameric forms. Considering this, the effect of strand concentration was taken into consideration. As reported in figure 4.50 B, the nature of the derived iM was asses by comparing the CD spectra recorded at different strand concentrations (3.60 and 20 µM). Nicely, the obtained trends overlap, the presence of a monomolecular tetraplex in solution.

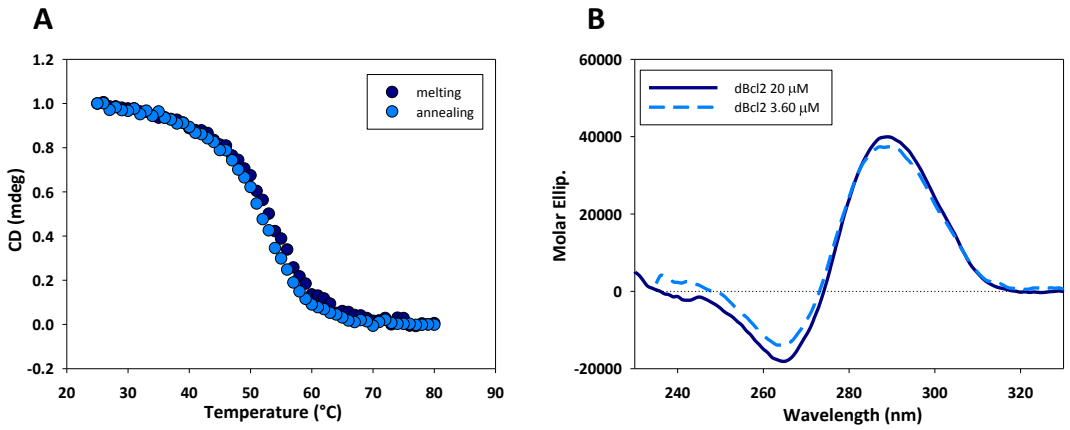


Figure 4.50 (A) Melting and annealing curve of dBcl2_C recorded in 10 mM Na-Cacodylate pH 5.5 at 288 nm (B) CD spectra recorded at pH 5.5 of 3.6 µM or 20 µM dBcl2_C.

The formation and the stability of the DNA secondary structures are largely affected by the ionic condition within the cell environment. Therefore, the conformational equilibrium of dBcl2_C was explored in presence of KCl.

To do this, working under neutral condition (pH 7.4) three different techniques were applied: Circular Dichroism, native PAGE assay, and fluorescence melting experiment.

The effect of KCl was initially assessed performing a Fluorescence melting assay at pH 7.4. This technique correlates the variation of fluorescence emitted by a suitably

sequence functionalized with a fluorophore and a quencher as a function of the temperature increment. In a condition that allows the folding into iM or other secondary structure, the fluorophore and quencher are in close proximity and the emission of the fluorophore will be quenched. On the contrary, when the sequence is completely unstructured, the maximum fluorescence emission is observed, since fluorophore and quencher are spatially distant. During this assay, the shift from the structured condition to the unstructured one is induced by temperature and the obtained out-put is a thermal profile where the variation of the fluorescence emitted is reported as a function of the temperature.

Considering the melting curves acquired none melting profile was detected even after the addition of 200 mM of KCl (data not showed). This suggests that the metal ion does not thermally stabilize the conformation assumed by our sequence. In agreement CD spectroscopy (figure 4.51 A) revealed that increasing KCl concentration doesn't alter the conformational equilibria in solution and native PAGE assay showed no variation in the electrophoretic mobility. Thus we can conclude that KCl doesn't induce the iM.

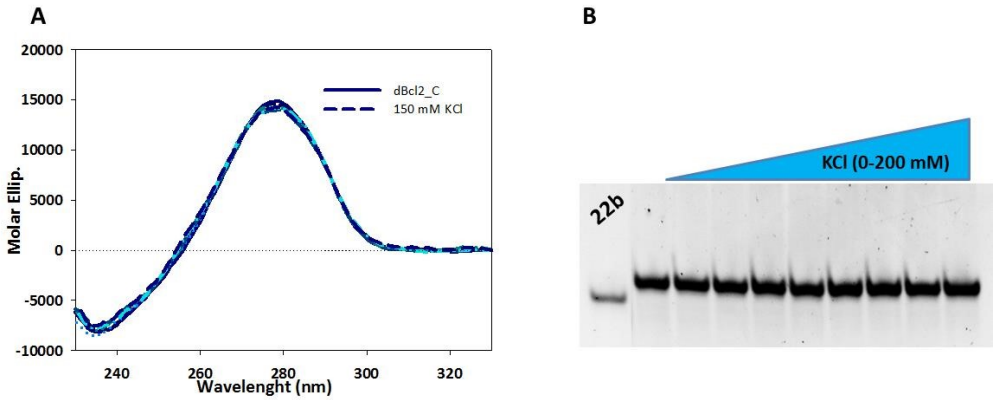


Figure 4.51 Effect of KCl on the iM forming sequence. (A) CD titration of 4 µM of dBcl2_C in 10 mM Na-Cacodylate pH 7.4 with increasing cation concentration. (B) 15% native PAGE assay runs in TBE 1x for the 25-nts iM forming sequence in presence of increasing KCl.

These collected data, unexpectedly are partially in contrasts with those published for EGFR-272_C, a 30-nts iM forming sequence previously characterized in our lab[88] where the iM was competing with a more defined hairpin that caused an impairment of the tetrahelical formation by the presence of monovalent cation. While in this context the monomeric stable iM assumed by dBcl2_C in mildly acidic condition is not in competition with other significantly folding.

4.5. Environmental factors affecting the G4 and iM conformational equilibria.

The great interest received by the non-canonical structures in not solely confined on a pharmacological field. In literature, several contributes focused on the possibility to use this arrangement as a tool for the evaluation and detection of a physiological and pathological condition or as a nanodevice loaded with protein or small molecule able to selectively release the cargo. Several contributes also reports that external factors, as small molecules, non-physiological cations, protein or dehydrating agent and co-solvent may alter these conformational equilibria. By keeping this in mind, after having established that all of the selected sequences along the 5'-UTR of *BCL2* are able to assume a non-canonical secondary structure (hairpin, G4, and iM), we decided to explore how the environment can modulate these conformational equilibria. In particular, the effect of selected ligands, non-physiological metal species (AgCl) or crowded agent, as the PEG_{200}, and the presence of the complementary strand, to modulate these foldings and their stability were assessed.

4.5.1 ds-DNA vs tetrahelical arrangements

The selected G- and C-rich domains and their ability to folds into peculiar non-canonical secondary structures were been confirmed using different experimental conditions where a single strand was present. Physiologically the DNA is a right-handed double helix stabilized by the Watson and Crick (WC) base pairs A-T and G-C. For the G-C base pairs, there are three hydrogen bonds (ΔG – 17.20 Kcal*mol-1) in contrast to the two present in the A-T base pairs. Considering the high content of Guanine and Cytosine of the DNA selected region, the formation of a stable double-stranded can be predicted. The base composition of the sequence exerts a fundamental role in the stability of the double-strand form, as well the cations physiologically present in cells. Indeed, positively charged species can neutralize the negatively charged backbone of the DNA leading to a differential stabilization of specific foldings. For this reason, here we assessed the ability of the dBcl2 25-nts long domain to form a ds-form instead of G4 and iM conformations in conditions that mimic the physiological environment.

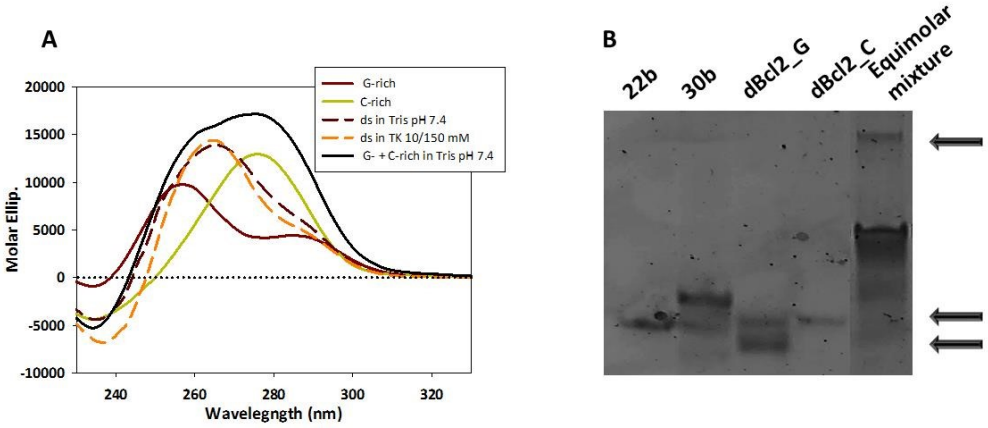


Figure 4.52 (A) CD spectra collected for dBcl2_G and dBcl2_C in 10 mM Tris pH 7.4. CD trend of an equimolar solution of both oligonucleotide in absence/presence of KCl and mathematical sum of the G- and C-rich spectra recorded. (B) 15% native EMSA assay runs in TBE 1x in presence of dBcl2_G-, dBcl2_ C and an equimolar mixture of selected strands.

A brief summary is reported in figure 4.52. As previously reported at pH 7.4 the single G and C-rich strands are differently organized. The guanine rich strand is folded into a G4 conformation even in the absence of cation, whereas for the C-rich sequence, a pre-organized conformation probably corresponding to a self-dimer was detected (+ve peak at 277 nm). When an equimolar solution of C- and G-rich sequences was considered, an unexpected behaviour was detected. Indeed, although the selected experimental conditions were set up to promote the formation of a double strand-form, the recorded CD features did not fit to a classical B-DNA (+ve peak at 280 nm and –ve ones at 245 nm). Indeed, it was characterized by a +ve contribution 265 nm, suggesting that the ability of the G-rich domain to assume the tetrahelical conformation might significantly prevent the formation of the ds-form. However, it is worth to remind that depending on the base composition different structures were adopted as well different CD spectra. In particular, compared to the canonical shape associated to the B-DNA, poly[d(G)] poly[d(C)] duplex generate a different CD spectrum, with a positive contribution at about 260 nm[167]. In order to understand the behaviour observed during the spectroscopic evaluation, a native EMSA assay was set up working in the same experimental condition (figure 4.52 B). Again, the collected data confirmed the propensity of the G- and C-rich sequence to assume a pre-organized conformation. Indeed electrophoretic bands having a comparable migration with the 22-nts marker were observed. The equimolar mixture of the two strands provides two species having different migration. The faster one runs comparable with the single strands, whereas the slower one might be related to the formation of a ds-form. Therefore, by merging all the presented data, we cannot fully exclude the formations of the double-stranded DNA in the selected condition.

4.5.2 AgCl: an interesting cation able to modulate the conformational features of DNA C-rich strand.

The DNA is a pivotal biopolymer comprises four distinct nucleobases (A, T, G, and C) and having the ability to self-assemble in intricate nanostructures in a reversible manner.

This peculiar ability and the excellent mechanical properties raised the interest regarding the nanotechnological application of the DNA. An interesting area of research is the metal-base interaction[168–170]. Indeed, each nucleobase offers several sites for the interaction with metal species. Over the years, many cations were found able to interact with the DNA. As example, in 1963 the ability of the Hg^{2+} to interact with the T residues, allowing the formation of T-HG^{2+}-T base pairs, was observed[171–173]. Similarly, in 1967 the interaction with the silver ion was proposed[174]. An hypothesis further confirmed lately[175–178]. In addition to that, the discovery of the non canonical secondary structures improves the research among the environmental condition able to stabilized/destabilize them. The presence of metal ions is one of the most intriguing partners for the iM folding process. In the last decade, the effect of monovalent or divalent cations such as calcium, magnesium and sodium was explored[50,179]. In 2013 Waller's group stated the ability of the silver ion to mediate the iM folding under neutral condition where the silver to acting as a bridge between the two N3 mimicking the structure of the hemi-protonated C-H^{+}-C dimer responsible for the stabilization of the i-motif structure in DNA[51,180]

Following these mentioned reports, we decided to evaluate the effect of AgCl on the conformational equilibria involving our cytosine rich sequence.

As in the case of the KCl, the ability of the silver ion to improve the thermal stability of our sequence was assessed through a fluorescence melting assay at neutral pH. As reported in figure 4.53, increasing concentrations of cation significantly alter the thermal profile associated to dBcl2_C.

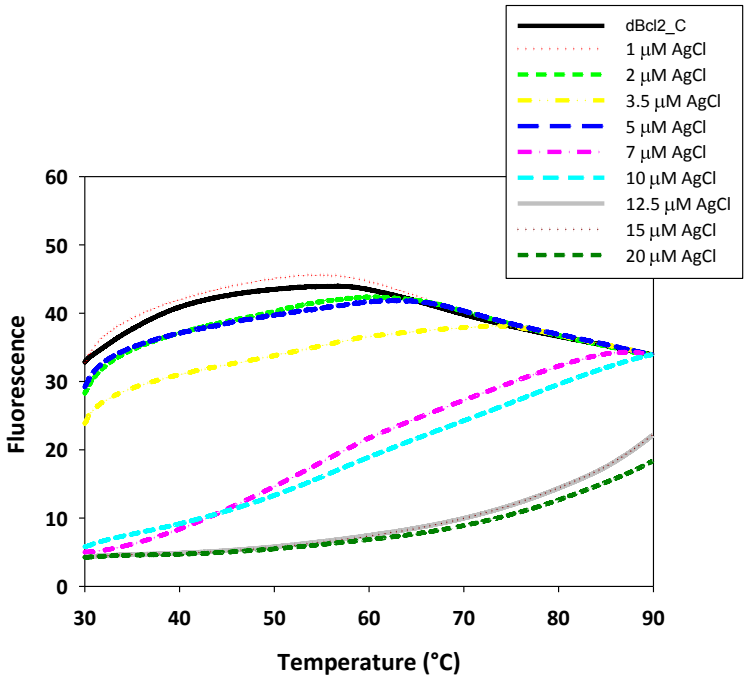


Figure 4.53 Melting curves recorded for 0.25 µM of labelled oligonucleotide in presence of increasing concentration of cation.

To evaluate the occurrence of a conformational rearrangement a native polyacrylamide gel was performed in the same experimental condition (10 mM Na-Cacodylate pH 7.4). Figure 4.54 A shows that the presence of cation did not alter the electrophoretic mobility of the tested nucleic acid. Indeed the electrophoretogram revealed the presence of a single band, having slower mobility compared to the selected 22-nts marker. Interesting results were derived by the CD titration that evidenced, as reported in figure 4.54 B, the aptitude of the silver ion to induce a conformational change. In particular, the positive contribution at 277 nm disappears and two different CD profiles were observed. The first one, characterized by a -ve contribution at 267 nm was observed up to the addition of 4 µM of cation. The second one, at higher AgCl concentrations, was characterized by a -ve contribution at 290 nm that was not completed up to 80 µM of AgCl. Each of these steps showed an isodicroic point. This particular behaviour might suggest that the conformational rearrangement promoted by the silver ion proceeds with a two-step process. To confirm our idea, the relative variations of the molar ellipticity at 267 nm and at 290 nm were plotted towards the cation concentration (figure 4.54 C). As expected, completely different trends were observed. At 267 nm, two distinct events were observed: in presence of lower cation concentrations a linear increment of the negative contribution was observed. However, by increasing the AgCl concentration, the –ve peak at 267 nm shifts towards 290 nm. Quite different is the behaviour at 290 nm, were only the

second step can be easily monitored. By fitting the data at 290 nm according to a one binding site equation the Kd for the second process was derived (Kd of 16.0 ± 1.6 µM).

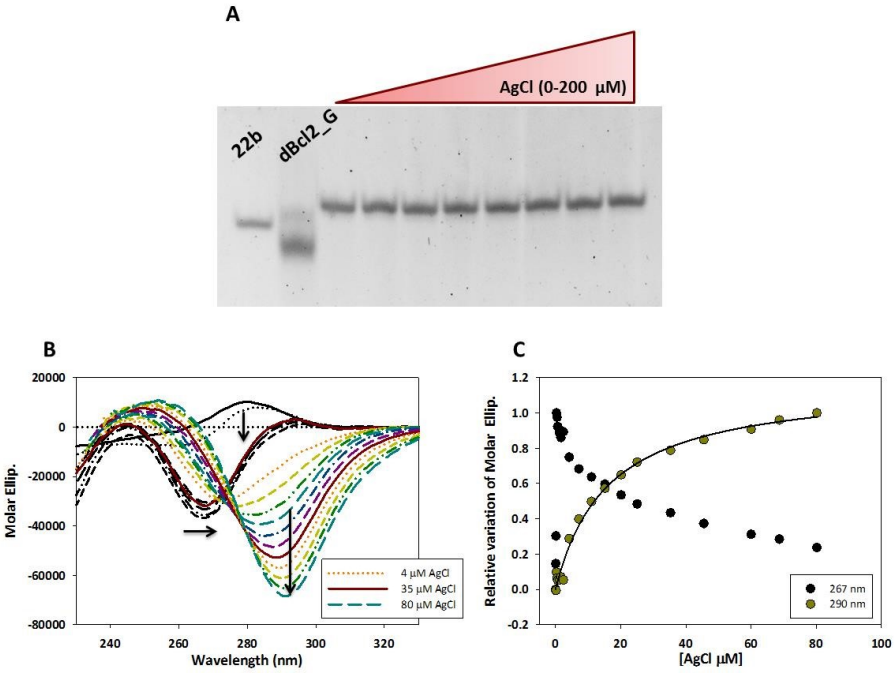


Figure 4.54. (A) 15% native PAGE assay runs in TBE 1x upon addition of increasing concentration of cation. (B) CD titration of 4 µM of dBcl2_C in 10 mM Na-Cacodylate with increasing AgCl concentration. (C) Relative variation of Molar Ellipticity at the two most relevant contributions (267 nm and 290 nm).

Together, the collected data revealed that AgCl promotes a conformational rearrangement towards a conformation that unfortunately not corresponds to an iM structure. To understand if the effect of this cation is specific for the iM forming sequences, CD titrations were performed on a different set of sequences. In particular, two C-rich sequences able to fold into an iM (cMYCpu24 and hTeloC) were selected. Coscramble-CC and Coscramble sequences were selected as cytosine-rich sequences not able to assume the iM, whereas the ctDNA as an example of a stable double strand.

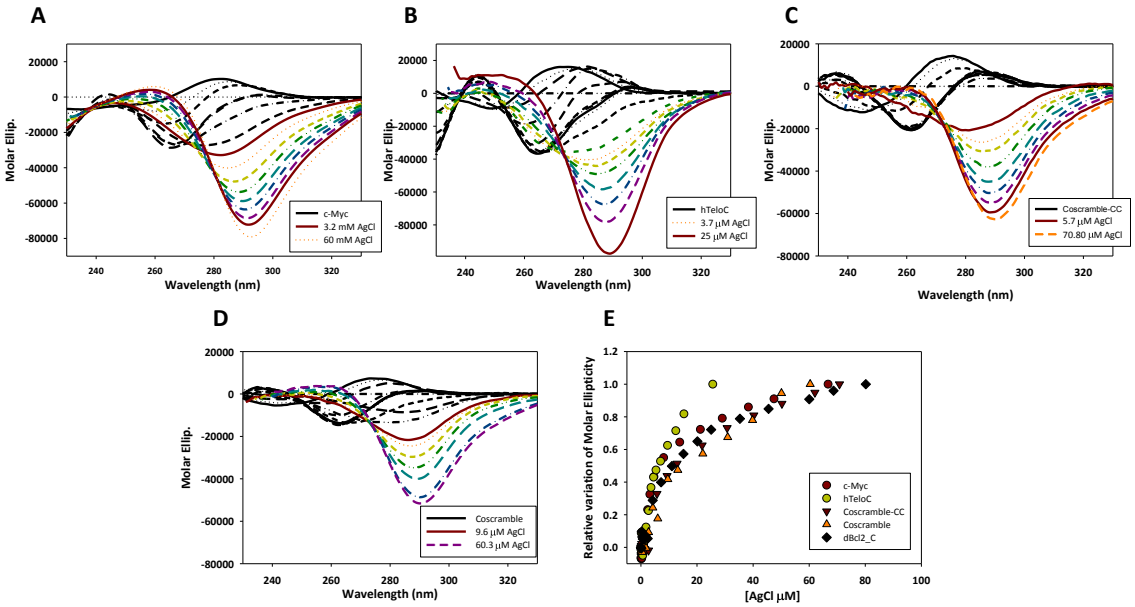


Figure 4.55 CD titrations on 4µM of (A) c-Myc (B) hTeloC, (C) Coscramble-CC with increasing concentrations of silver ions. (B) Relative variation of the signal plotted against the cation concentration.

Figure 4.55 revealed that the general behaviour observed reminds the one observed for dBcl2_C. Indeed, the main +ve contribution disappears and two different CD profiles, both characterized by isodicroic point, were observed. Again, we followed the second step of the entire process and by fitting the relative variation at the most relevant -ve contribution (293 nm for c-Myc, 289 nm for hTeloC and 290 nm for Coscramble-CC and Coscramble) according to a one binding site model a dissociation constant were derived (9.3 ± 1.6 µM and 14.9 ± 2.6 µM 23.5 ± 4.6 µM, 32.3 ± 5.8 µM and of 3.6 ± 0.8 µM). A comparison of this second step for each tested sequences is reported in figure 4.55 D. It emerged that unfortunately none of the selected sequences reached the saturation point, moreover, with the exception of the hTeloC, the behaviours observed for the other sequences nicely overlap. This suggests that the cation acts similarly on the iM forming sequences as well as on the Coscramble-CC and Coscramble that are enriched in cytosine but not designed to fold into an iM. Quite different was the scenario exhibit in figure 4.56 where the ctDNA was titrated with increasing AgCl. Indeed, again the silver ion induces a conformational change. The positive contribution at 270 nm disappears, two different CD profiles were observed and each of them showed an isodicroic point. However, by plotting the relative variations at 286 nm, which correspond to the second step, two different events occur, thus we cannot derive dissociation constant. Therefore, it appears that the selected metal ion acts differently on a double-strand form.

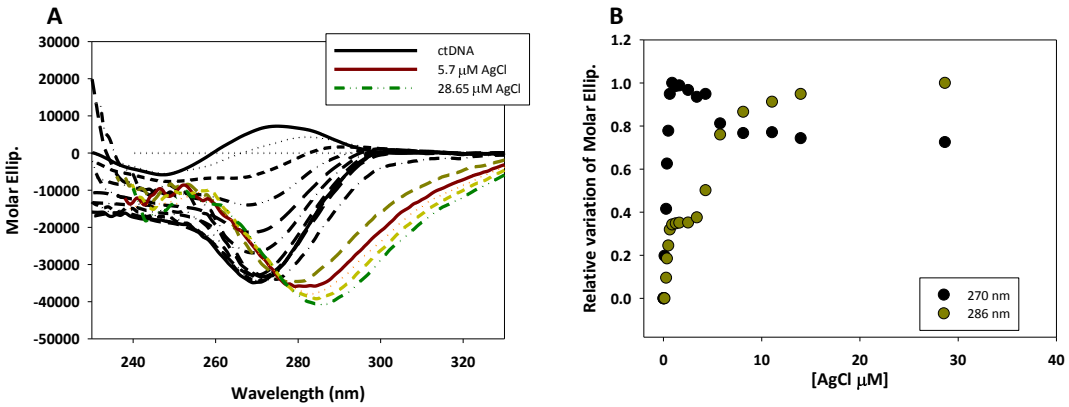


Figure 4.56 CD titration of ctDNA with increasing concentrations of AgCl. (B) Relative variation of Molar Ellipticity at the two most relevant contributions (270 nm and 286 nm).

The herein presented data are fully in contrast with the reported ability of the cation to induce the iM. Indeed, CD features recorded does not remind the signal associated to an iM conformation and more relevant, with the exception of the ctDNA, that is a double strand, the same trend was observed for each tested C-rich sequences.

Recently Swasey and colleagues[181] stated the ability of this cationic species to induce the formation of a parallel duplex by using ESI-MS and FRET assay. The formation of this peculiar pairings can fit our results. However, PAGE data excluded the formation of a dimeric species. This fully excludes the formation of a parallel dimer induce by the presence of AgCl.

Since unfortunately, our recorded CD features don't fit either with the CD spectra reported for the iM nor for the parallel duplex CD trend, in order to accurately define which conformation is assumed by each selected sequence further studies under these conditions are needed.

4.5.3 PEG_{200}: a crowded agent able to modulate the conformational equilibria of dBcl2 G- and C-rich sequences.

The presented data so far were acquired in a dilution condition. However, it is worthy to note that the intracellular environment is considerably different from the diluted aqueous solution. Indeed, in living cells the 20-40% of the total volume is occupied by various biomolecules, thus, the dynamics and the interaction between biomolecules, as well the efficiency of the gene transcription and translation and the cell growth is modulated by this crowded condition[182–184]. Based on this, in vitro studies, using buffered solution in presence of protein as albumin, chymotrypsin and lysozyme were performed. However, in their presence is not easy to analyse the reaction of interest, also, the added proteins might bind and/or degrade the nucleic acid. For this reason, to generate the intracellular crowding condition, synthetic cosolutes are considered more convenient reagents. Especially, among them, neutral or net neutral molecules as ethanol or small Poly(ethylene glycol) (PEG) reduces the water molecules activities, generates an osmotic stress and reduces the dielectric constant[185,186]. In the case of nucleic acids, it has been reported that the addition of cosolutes stabilizes the less hydrated structures, as the G4; moreover, they influence the folding topology. Also, the C-rich domain is affected by of these agents, in fact, several reports revealed the folding into iM structures at neutral pH lead by the addition of PEG_{200}[33,187,188]. An evidence completely in line with the reported data regarding the destabilization of the Watson-Crick duplex[189,190].

For this reason, the effect of PEG_{200} on the folding pathway of both DNA G- and C-rich tracts contained along the *BCL2* DNA untranslated domain, was assessed (figure 4.57).

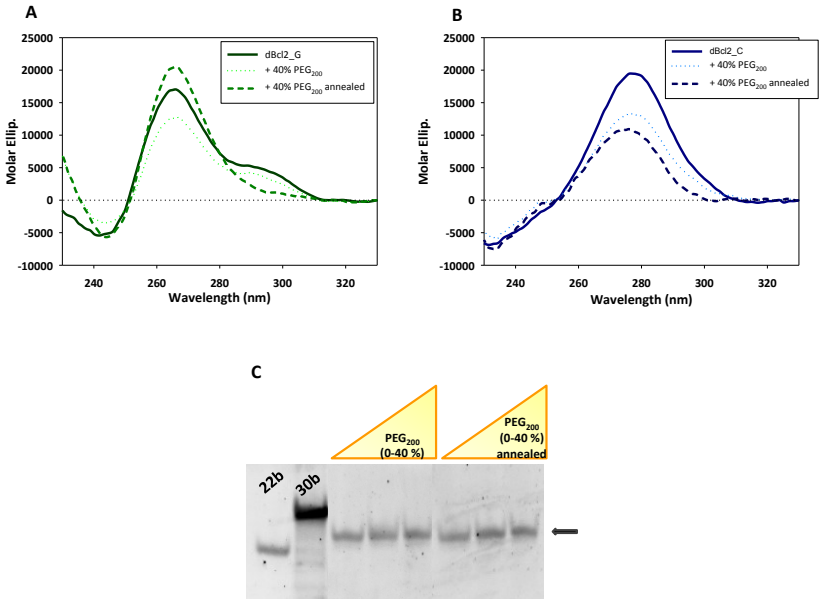


Figure 4.57 (A) CD spectra of the guanine rich sequence in absence/presence of 40% of cosolvent before and after an annealing cycle. (B) CD spectra of dBcl2_C in absence/presence of 40% of cosolvent before and after an annealing cycle. (C) 15% PAGE assay for the C-rich sequence in presence of increasing % of $PEG_{200.}$ One set of sample was loaded after an annealing cycle in presence of the polymer.

As reported in figure 4.57, the addition of the PEG_{200} caused different behaviours on the tested sequences. In particular, the addition of 40% of PEG_{200} on dBcl2_G immediately leads to the reduction of the CD signal. However, after an annealing cycle, the recorded CD spectrum exhibits an increment of the signal at 265 nm, while the +ve peak at 295 nm disappears. This behaviour suggests that the presence of the polymer partially alters the conformational equilibria in solution of dBcl2_G and, in line with the datasets reported in literature it appears that the parallel conformation is promoted (figure 4.57 A). Quite different was the behaviour observed for the cytosine rich fragment. As reported in panel B the molar ellipticity immediately dropped upon the addition of 40% of PEG_{200} that was even more evident after an annealing step. Additionally, considering the peak position, clearly, no induction of the iM conformation was provided. Despite, our set of data is totally in contrast with the reported behaviour of PEG_{200} on the iM folding process of several genes; these data are totally in line with those observed for the C-rich tract, along the EGFR promotorial region. Indeed, neither for the EGFR-272_C sequence, the addition of Poly (ethylene glycol) promotes the iM folding. To conclude our evaluation on dBcl2_C a 15% native PAGE assay was set up. In particular, the effect of increasing percentage of PEG_{200}, as well the annealing cycle after its addition, on the conformational equilibria of dBcl2_C was assessed. From the collected data, reported in panel C it appears that the addition of increasing % of PEG_{200} doesn't alter the electrophoretic mobility of the samples

loaded immediately after the addition of polymer, whereas a very small shift of the electrophoretic bands was observed for the samples loaded after an annealing cycle.

Considering all the acquired data we can propose that for the dBcl2_G crowding condition partially promotes a conformational selection, probably towards the parallel topology. Contrarily, the same percentage of polymer does not promote the iM folding of dBcl2_C under neutral condition. Indeed, no detectable changes towards a CD trend that reminds the iM were observed. Moreover, considering the sensible drop off of the CD signal and the small variation on the electrophoretic mobility after the annealing cycle, we might suggest that the selected polymer strongly interact with the 25-nts C-rich sequences, allowing the formation of a DNA-PEG complex that might precipitate in solution.

4.5.4 Targeting the iM and the G4 structures: screening of small molecules on DNA and RNA selected sequences.

The first goal of this project was the structural characterization of the guanine and cytosine rich sequences contained in the 5'-untranslated domain of *BCL2*. Since these conformations, may modulate the recruitment of the Polymerase II enzyme and the transcription process or the scanning of the ribosome along the primary transcript, we moved forward to a compounds screening, in order to identify small molecules able to selectively recognized, stabilized or induced them. At the first time the possibility to alter the recruitment of the RNA polymerase enzyme by inducing the iM was taken into account.

To get this aim the fluorescence melting assay was performed in presence of increasing concentration of selected derivatives already known for their ability to recognize the G4 form (Figure 4.58). As reference compounds Mitoxantrone and Berberine were selected since it was recently stated their ability to destabilized the iM under neutral condition (MX) or stabilized it at acidic pH[191–193]. All selected compounds were screened at slightly acidic pH (pH 5.5) to check their stabilization properties on the folded iM and at pH 7.4 to check the possible induction of the iM at physiological pH.

Differently, from the other techniques used so far, a labelled sequence, with Dabcyl (quencher) at the 5'-end and FAM (Fluorophore) at the 3'-end was titrated with increasing concentrations of small molecules. In conditions that allow the folding into iM, fluorophore and quencher result in close proximity, if the fluorophore will be excited its emission will be quenched by the Dabcyl. On the contrary, when the environmental conditions lead to a destabilization of the structure in solution, as during the denaturation step, the fluorophore and quencher move apart from each other and an increase of the fluorescence signal associated with the FAM will be observed[194]. This means that the thermal denaturation profile can be easily recorded and the melting temperature can be derived. Since this value, is related to the thermal stability of the structure contained in solution, ligands able to induced/stabilize the arrangements in solution provides an increment of the melting temperature, while the opposite effect will be observed in presence of inactive/destabilizer compounds.

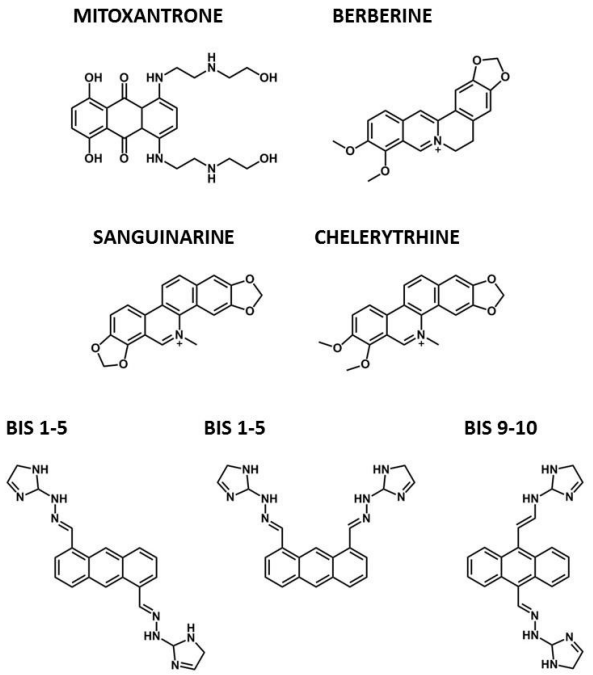


Figure 4.58 Chemical structures of the tested compounds.

The recorded data, summarized in figure 4.59 highlighted the different activity of the selected compounds on the cytosine rich sequence. With the exception of the Sanguinarine for which was observed a 7 °C stabilization of the iM structure, Berberine and Chelerythrine do not promote the stabilization or the induction of the iM. Considering the MX, the observed behaviour was quite unexpected, however in in 2016 Waller and coworkers stated the high stabilization effect of the MX on the hTeloC iM structures compared to the double-strand form or G4 conformation[195]. Since two different behaviours are reported in literature, we cannot give a clear explanation of the real effect exerts by this compound on the selected iM. However we can speculate that the MX interact with the unfolded form during the thermal denaturation step, thus the renaturation step might be affected by this peculiar complex, or it is able to selectively recognizes the pre-organized form observed at pH 7.4 and promotes the shifting towards this conformation at pH 5.5.

Quite encouraging was the data recorded for the anthracene derivatives. Indeed, an increment of the thermal stability was observed upon the addition of compounds at both tested pH, although this was more prominent in acidic condition.

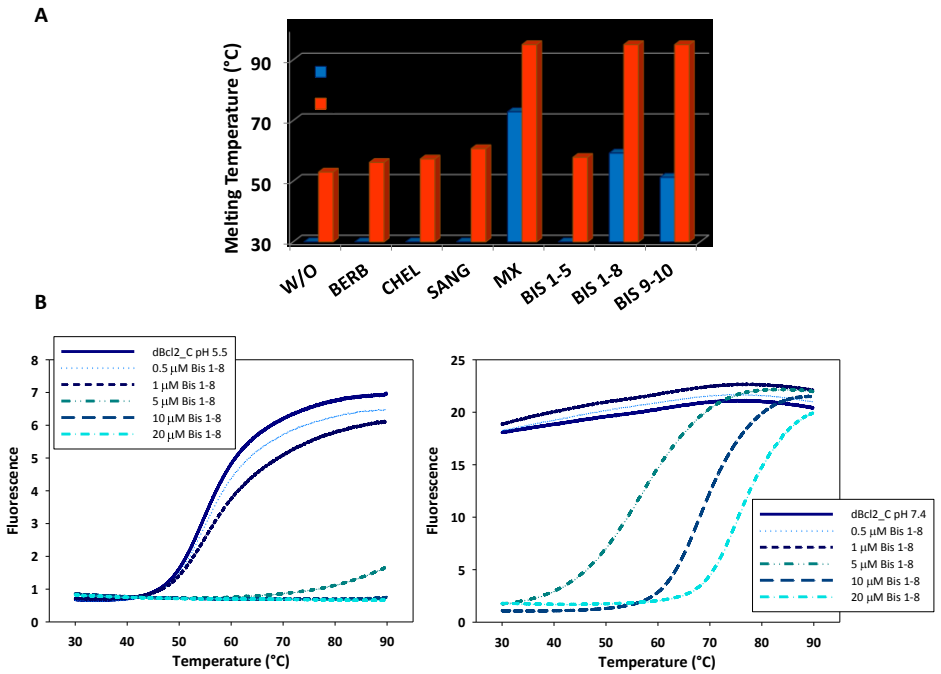


Figure 4.59 (A) Melting temperature derived from the denaturation profile of dBcl2_C at pH 7.4 or 5.5 in presence of 5 µM of selected compound. (B) Thermal profiles derived at pH 5.5 or 7.4 in increasing concentration of Bis 1-8 compound.

Since at these days, no information are reported for this class of compounds on the iM structure, our data might be revealed the possibility to stabilize the existing iM at acidic pH and the possibility to induce it in condition that mimics the physiological pH. In our lab, the iM forming sequence along the *BCL2* 5'-UTR region is not the solely C-rich tract under investigation, recently we stated the ability of a 30-nts C-rich sequence (EGFR-272_C) within the EGFR promoter region to assumes the iM, also, promising results have been collected for the C-rich tract located 37 bases upstream the TSS of the same gene. To corroborate the data previously described, the effect of these compounds on these aforementioned C-rich sequences was evaluated.

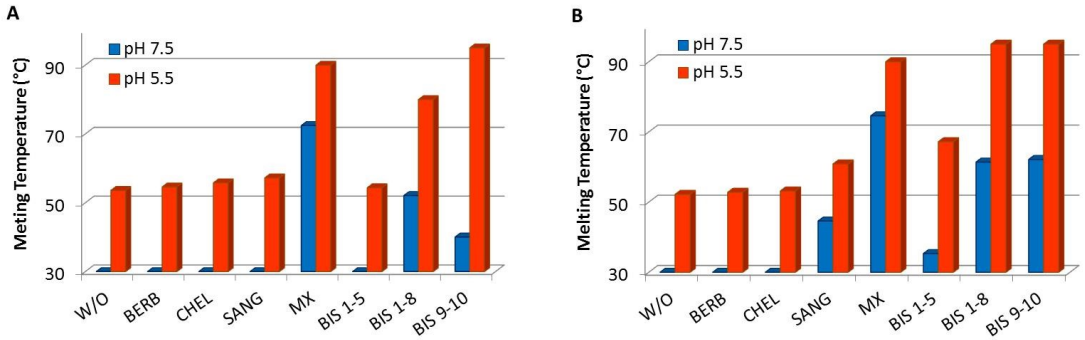


Figure 4.60 (A) Fluorescence melting assay of 0.25 µM of (A) EGFR-37_C and (B) EGFR-272_C at pH 5.5 or 7.4 in presence of 5 µM of selected compounds.

Data acquired and summarized in figure 4.60 revealed that again Berberine and Chelerythrine result quite inactive at both tested pHs. An exception is the Sanguinarine that improves the thermal stabilization at pH 7.4 for the EGFR-272_C sequence. As expected, MX and the anthracene derivatives increase the melting temperatures at both tested conditions. An in-depth evaluation of the collected data revealed that for EGFR-37 at physiological pH the compound 1-8 acts better compared to the 9-10, contrarily to acidic pH. On the contrary, at the two tested pH a comparable stabilization of the Tm was detected for EGFR-272_C upon the addition of Bis 1-8 and Bis 9-10, moreover, also the compound 1-5 results quite active in acidic condition.

Depending by the working condition, with the exception of the rBcl2_48, the guanine rich sequences assume different conformations in solution. In this contest, one possible drawback of the fluorescence melting assay is the obtained of not defined melting curves. To overcome this fact the fluorescent intercalator displacement assay (FID) was considered. This specific assay follows the loss of fluorescence of thiazole orange (TO) upon a competitive displacement from DNA or RNA by the addition of putative ligands. In fact, upon stacking interaction on the external tetrads or when it intercalates between the WC base-pair the TO emits a fluorescence signal, that decreases if the added compound efficiently competes with these interactions.
Before performing the assay in presence of the selected compound, we checked the binding of TO on the tested oligonucleotides performing a titration. As expected, data reported in figure 4.61 revealed the ability of the Thiazole Orange to recognize and interact with each guanine rich tracts. Indeed, an increment of the Fluorescence signal was observed towards the TO concentration.

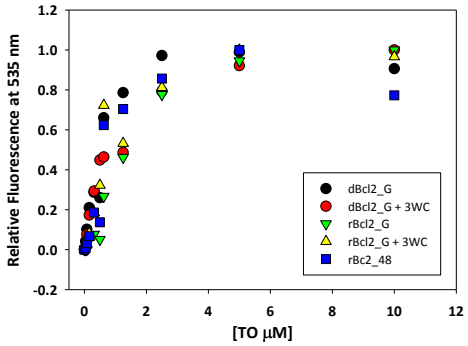


Figure 4.61 TO titration of the selected DNA and RNA guanine rich sequences.

Once proved the interaction between the TO and the G-rich sequences, a high throughput screening in 10 mM LIP pH 7.4, 100 mM KCl was set up in presence of increasing concentration of compound (Figure 4.62).

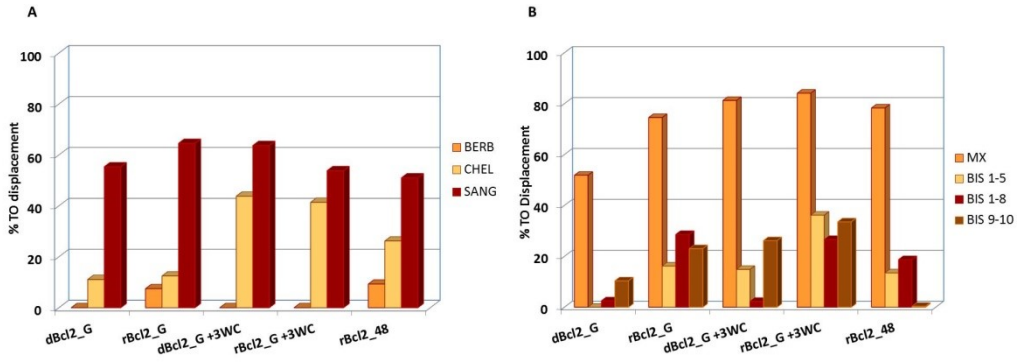


Figure 4.62 Fluorescence intercalator displacement assay in presence of 2 µM of selected ligand on the characterized guanine rich domains contained within the DNA or RNA 5'-UTR region of *BCL2*.

The selected compounds recognize and interact differently the tested sequences. Indeed, different percentage of displacement was observed.
Unexpected was the observed behaviour upon the addition of Berberine. Indeed, in literature, its ability to stabilized the telomeric G4 has been reported[196], however, a small displacement was observed only for rBcl2_G and rBcl2_48 sequences. This might suggest that the Berberine compound binds preferentially the parallel RNA G4 and the intramolecular hairpin structure. As expected, the Sanguinarine[197], despite with different efficient on the sequences, promotes the higher TO displacement, while an intermediate efficiency was detected for the Chelerythrine compound.
As expected, the MX gives us the highest displacement. Therefore, it cannot be considered a selective compound towards a specific DNA or RNA conformation. Regarding the anthracene derivatives, their activity results sequence-dependent and less intense compared to the MX. Especially, the addition of Bis 1-8 provides the highest displacement on rBcl2_G and rBcl2_48, Bis 1-5 provides the highest displacement for rBcl2_G + 3WC, while compound 9-10 give the best displacement on both DNA selected sequences.

Generally our data revealed that Berberine, Sanguinarine and Chelerythrine cannot be used to induce/stabilized the conformation assumed by dBcl2_C whereas quite encouraging were the data acquired in presence of compound 9-10 and 1-8. Concerning the guanine rich tracts, the tested compounds do not interact specifically with a single conformation, thus we cannot derive clear information regarding which compound can stabilizes or induces the G4 conformation or the intramolecular hairpin-like conformation

5 CONCLUSIONS

5.1 C-Myc i-Motif

The overexpression of MYC is a common feature across several types of cancer that often is associated with a poor prognosis and multidrug resistance, especially when in combination with the overexpression of BCL2. Despite a lot of work has been made to selectively target this protein, compounds able to inhibit or reduce the expression of this gene by targeting the secondary structures within the promoter region could be an innovative approach. One such compound is GQC-05, indeed is already reported its ability to recognize and stabilize the G4 structure contained in MYC promoter, inducing transcription repression. In this first part of the project, based on the already published structural characterization we explored one of the main important factors affecting the iM stability: the loop composition. Data acquired clearly demonstrated that the first loop, thanks to the presence of an additional WC base pair between positions 6 and 9 plays the most important role in the stability of the overall structure. This suggestion might be also transferred to the third loop, in fact being the narrowest grooves, as the first one, the spatial proximity may allow the formation of Watson-Crick base pairs between the position 24 and 28. On the contrary, our evidence revealed that the central loop, since it is the major groove, marginally acts on the stability of the tetrahelical structure. Moreover, no additional WC pairing was observed. Once having proved that the remarkable stability of this iM is affected by the presence of canonical Watson-Crick base pair along the loop region, a small molecule (IMC-30) was considered as a putative iM target. From the acquired data, the binding site of the compound was suggested. Particularly, we proposed that the recognition site is located on the first and second loop and involves position 6, 8 and 10 for the first one whereas the position 15, 17 and 19. Moreover, the addition of IMC-30 thermally stabilized the iM conformation. Our compound was also proved able to affect the binding of the hnRNPK protein, thus the unfolding process was prevented. Finally, with a cellular test, we demonstrated that this compound actively modulates the apoptosis leading an increment of this process. Thus, it can be considered an interesting compound for innovative anticancer treatment either as a single treatment or in drug combination with a well-known compound able to target the BCL protein.

5.2 *BCL2* and the non canonical secondary structures

A conclusion of this structural characterization study, we can say that all tested sequences are actually able to fold into a non-canonical secondary structure under the selected experimental condition. However, it appears that the features of all these folded forms are quite different from each other.

Regarding the 5 selected guanine rich strings, they demonstrated the ability to generate different conformations depending by the working condition (KCl, annealing

cycle and LiCl). In particular, considering the two 25-nts selected sequences (dBcl2_G and rBcl2_G), the collected data revealed their ability to assumed a pre-organized conformation, that partially disappears after an annealing cycle in 10 mM Tris buffer pH 7.4. As expected, for both of them, the presence of KCl promotes the conformational shifts towards several G4 conformations coexisting in solution, both in a monomeric and multimeric fashion form. Also, from the evaluation of the thermal properties, it was clear the remarkable stability of both systems.

Since the RNA in the cellular compartment is present as a single strand, the G4 conformation may be in strong competition with other secondary conformation. To confirm the tetrahelical nature of the conformation observed for rbcl2_G the effect of LiCl was considered. Acquired data nicely confirmed our idea regarding the presence of parallel G4 conformation. Indeed no significant variation was detected in a solution containing a high concentration of this cation.

Our data underline that both sequences in an environmental condition comparable with those found in the cell (150 mM KCl pH 7.4), assume multiple conformations. Since the sequence context around the minimal G4 forming sequence largely affects the conformational selection towards a single structure. Two longer sequences were taken into consideration (dBcl2_G + 3WC and rbcl2_G + 3WC) in order to get a conformational selection. In spite of the theoretical data derived through bioinformatics tool regarding their high aptitude to assume the G4 conformation, the collected data partially disagree. It appears evident that in each tested condition (KCl or LiCl) the addition of 6 nucleotides, allows the formation of 3 WC pairings that partially prevent the G4 folding. Moreover, in presence of KCl a fascinating system composed by at least three different species was observed for the rBcl2_G+ 3WC.

Since literature data reports the formation of alternative secondary structures that might have a physiological role, we explored the possibility to get a hairpin conformation able to compete with the G4 by implementing he number of nucleotide adjacent to the guanine rich core. All presented data for rBcl_48 confirmed the derived theoretical ability of this sequence to assume a stable intramolecular conformation even in G4 promoting condition despite, its abundance in guanine bases.

In addition to the guanine rich tracts, also the structural properties of the cytosine rich sequence contained in the antisense DNA strand were taken into account.

The collected data for dBcl2_C revealed its ability to assume a monomeric iM structure with a pHt quite close to the physiological pH. Since the iM folding requires a specific pH condition that does not fit the intracellular environment, its potential occurrence was asses also in a condition mimicking the nuclear environment (KCl and pH 7.4). From the acquired data it emerged that the KCl does not alter the conformational equilibria of the selected sequence; as well no stabilization effect was observed. It is well known that external condition, as the presence of complementary strand, crowding agent ligands or metal species, alters the conformational equilibria of the

guanine and cytosine rich tract. In this contest, we proved that despite the higher stability observed in vitro; the G4 conformation does not completely impair the formation of the ds-form. As crowding condition, the activity of PEG_{200} was tested. Contrarily to the expected data, the selected polymer does not promote either the folding or the stabilization of the i-Motif. While, the formations of non-specific DNA-PEG complexes was suggested. Interesting are the derived data for dBcl2_G. Indeed, molecular crowding condition in concert with the annealing cycle partially promotes a conformational selection towards a structure that reminds to a parallel topology. This fact may suggest that, among the several topologies assumed in vitro by the guanine rich tracts, the parallel one might exert a physiologically relevant role compared to the others. Extremely interesting is also the effects exerted by the silver ion. Our data pointed out the ability of this ion to modulate the conformational equilibria involving the C-rich tract. However, contrarily to the expected behaviour, this cation promotes a conformational shift towards a structure totally different from the i-Motif. The most recent data revealed the ability of this cation to induce a parallel duplex structure, unfortunately, our acquired CD feature is not reminiscent with this specific conformation as well our electrophoretic data. Based on this unclear set of data, we cannot define exactly which conformation is assumed by dBcl2_C at the tested experimental condition and further studies are required. However, our evidence nicely supports the possibility to get a DNA-metalled nanostructure that can be used as an innovative nanotechnological tool. To conclude our project the feasibility to target the G4, the iM and the hairpin-like conformation previously characterized was taken into consideration. Despite an improvement of the knowledge is required, the performed screening protocol pointed out that the natural compounds cannot be considered putative iM binders. While, compounds 1-8 and 9-10 were proved able to directly interact with the C-rich sequence at both selected pHs, promoting a thermal stabilization of the conformation contained in solution.

Regarding the guanine rich tracts, unfortunately, from the acquired data, despite a sequence-dependent activity was observed, it appears that all screened ligands do not interact specifically with a single conformation (G4 or hairpin). Therefore, we cannot define which compound can be used for inducing the G4 conformation and which one can induce the intramolecular hairpin structure.

6 REFERENCES

1. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* (2000). doi:10.1016/S0092-8674(00)81683-9
2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* (2011). doi:10.1016/j.cell.2011.02.013
3. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. J.D. Watson and F.H.C. Crick. Published in Nature, number 4356 April 25, 1953. *Nature* (1974).
4. Choi, J. & Majima, T. Conformational changes of non-B DNA. *Chemical Society Reviews* (2011). doi:10.1039/c1cs15153c
5. Subramaniyam Ravichandran, Vinod Kumar Subramani, and K. K. K. Z-DNA in the genome: from structure to disease. *Bhiophysical Rev.* **11**, 383–387 (2019).
6. Kaushik, M. *et al.* A bouquet of DNA structures: Emerging diversity. *Biochemistry and Biophysics Reports* (2016). doi:10.1016/j.bbrep.2016.01.013
7. Brázda, V., Laister, R. C., Jagelská, E. B. & Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Molecular Biology* (2011). doi:10.1186/1471-2199-12-33
8. Felsenfeld, G., Davies, D. R. & Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *Journal of the American Chemical Society* (1957). doi:10.1021/ja01565a074
9. Morgan, A. R. & Wells, R. D. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J. Mol. Biol.* (1968). doi:10.1016/0022-2836(68)90073-9
10. Achar, A. & Sætrom, P. RNA motif discovery: A computational overview. *Biology Direct* (2015). doi:10.1186/s13062-015-0090-5
11. Cammas, A. & Millevoi, S. RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkw1280
12. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *Journal of Molecular Biology* (2017). doi:10.1016/j.jmb.2017.05.017
13. GELLERT, M., LIPSETT, M. N. & DAVIES, D. R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* (1962).
14. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* (1988). doi:10.1038/334364a0
15. Lipps, H. J. & Rhodes, D. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology* (2009). doi:10.1016/j.tcb.2009.05.002
16. Bidzinska, J., Cimino-Reale, G., Zaffaroni, N. & Folini, M. G-quadruplex structures

in the human genome as novel therapeutic targets. *Molecules* (2013). doi:10.3390/molecules181012368
17. Simonsson, T. G-quadruplex DNA structures - Variations on a theme. *Biological Chemistry* (2001). doi:10.1515/BC.2001.073
18. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkl655
19. Patel, D. J., Phan, A. T. & Kuryavyi, V. Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Research* (2007). doi:10.1093/nar/gkm711
20. Ou TM, Lu YJ, Tan JH, Huang ZS, Wong KY, G. L. G-Quadruplexes: Targets in Anticancer Drug Design. *ChemMedChem* **3**, 690–713 (2008).
21. Gehring, K., Leroy, J. L. & Guéron, M. A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* (1993). doi:10.1038/363561a0
22. Yang, B. & Rodgers, M. T. Base-pairing energies of proton-bound heterodimers of cytosine and modified cytosines: Implications for the stability of DNA i-motif conformations. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja409515v
23. Berger, I., Egli, M. & Rich, A. Inter-strand C-H···O hydrogen bonds stabilizing four-stranded intercalated molecules: Stereoelectronic effects of O4' in cytosine-rich DNA. *Proc. Natl. Acad. Sci. U. S. A.* (1996). doi:10.1073/pnas.93.22.12116
24. Modi, S. *et al.* A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.* (2009). doi:10.1038/nnano.2009.83
25. Keum, J. W. & Bermudez, H. DNA-based delivery vehicles: PH-controlled disassembly and cargo release. *Chem. Commun.* (2012). doi:10.1039/c2cc37471d
26. Li, W., Feng, L., Ren, J., Wu, L. & Qu, X. Visual detection of glucose using conformational switch of i-motif DNA and non-crosslinking gold nanoparticles. *Chem. - A Eur. J.* (2012). doi:10.1002/chem.201201914
27. Dembska, A. The analytical and biomedical potential of cytosine-rich oligonucleotides: A review. *Analytica Chimica Acta* (2016). doi:10.1016/j.aca.2016.05.007
28. Sellner, S. *et al.* Dexamethasone-conjugated DNA nanotubes as anti-inflammatory agents in vivo. *Biomaterials* (2017). doi:10.1016/j.biomaterials.2017.04.031
29. Guo, K., Gokhale, V., Hurley, L. H. & Sun, D. Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn380
30. Kendrick, S., Akiyama, Y., Hecht, S. M. & Hurley, L. H. The i-motif in the bcl-2 P1

promoter forms an unexpectedly stable structure with a unique 8:5:7 loop folding pattern. *J. Am. Chem. Soc.* (2009). doi:10.1021/ja9076292

31. Dai, J., Hatzakis, E., Hurley, L. H. & Yang, D. I-motif structures formed in the human c-MYC promoter are highly dynamic-insights into sequence redundancy and I-motif stability. *PLoS One* (2010). doi:10.1371/journal.pone.0011647
32. Greco, M. L., Folini, M. & Sissi, C. Double stranded promoter region of BRAF undergoes to structural rearrangement in nearly physiological conditions. *FEBS Lett.* (2015). doi:10.1016/j.febslet.2015.06.025
33. Benabou, S. *et al.* i-motif structures in long cytosine-rich sequences found upstream of the promoter region of the SMARCA4 gene. *Biochimie* (2017). doi:10.1016/j.biochi.2017.06.005
34. Miglietta, G., Cogoi, S., Pedersen, E. B. & Xodo, L. E. GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1. *Sci. Rep.* (2015). doi:10.1038/srep18097
35. Wright, E. P., Huppert, J. L. & Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx090
36. Fleming, A. M. *et al.* 4n-1 Is a “Sweet Spot” in DNA i-Motif Folding of 2’-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.6b10117
37. Dzatko, S. *et al.* Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chemie - Int. Ed.* (2018). doi:10.1002/anie.201712284
38. Zeraati, M. *et al.* I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* (2018). doi:10.1038/s41557-018-0046-3
39. Assi, H. A., Garavís, M., González, C. & Damha, M. J. I-motif DNA: Structural features and significance to cell biology. *Nucleic Acids Research* (2018). doi:10.1093/nar/gky735
40. Fojtik, P. The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. *Nucleic Acids Res.* (2001). doi:10.1093/nar/29.22.4684
41. Mir, B. *et al.* Prevalent Sequences in the Human Genome Can Form Mini i-Motif Structures at Physiological pH. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.7b07383
42. Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS Journal* (2010). doi:10.1111/j.1742-4658.2010.07759.x
43. Fujii, T. & Sugimoto, N. Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys. Chem. Chem. Phys.* (2015). doi:10.1039/c5cp02794b
44. Gurung, S. P., Schwarz, C., Hall, J. P., Cardin, C. J. & Brazier, J. A. The importance of loop length on the stability of i-motif structures. *Chem. Commun.* (2015).

doi:10.1039/c4cc07279k
45. Lieblein, A. L., Fürtig, B. & Schwalbe, H. Optimizing the Kinetics and Thermodynamics of DNA i-Motif Folding. *ChemBioChem* (2013). doi:10.1002/cbic.201300284
46. Nonin-Lecomte, S. & Leroy, J. L. Structure of a C-rich strand fragment of the human centromeric satellite III: A pH-dependent intercalation topology. *J. Mol. Biol.* (2001). doi:10.1006/jmbi.2001.4679
47. Benabou, S. *et al.* Understanding the effect of the nature of the nucleobase in the loops on the stability of the i-motif structure. *Phys. Chem. Chem. Phys.* (2016). doi:10.1039/c5cp07428b
48. Školáková, P. *et al.* Systematic investigation of sequence requirements for DNA i-motif formation. *Nucleic Acids Res.* (2019). doi:10.1093/nar/gkz046
49. Saxena, S., Bansal, A. & Kukreti, S. Structural polymorphism exhibited by a homopurine·homopyrimidine sequence found at the right end of human c-jun protooncogene. *Arch. Biochem. Biophys.* (2008). doi:10.1016/j.abb.2008.01.015
50. Kim, S. E., Lee, I. B., Hyeon, C. & Hong, S. C. Destabilization of i-motif by submolar concentrations of a monovalent cation. *J. Phys. Chem. B* (2014). doi:10.1021/jp500120d
51. Day, H. A., Huguin, C. & Waller, Z. A. E. Silver cations fold i-motif at neutral pH. *Chem. Commun.* (2013). doi:10.1039/c3cc43495h
52. Day, H. A., Wright, E. P., MacDonald, C. J., Gates, A. J. & Waller, Z. A. E. Reversible DNA i-motif to hairpin switching induced by copper(ii) cations. *Chem. Commun.* (2015). doi:10.1039/c5cc05111h
53. Kang, B. H. *et al.* Thiazole orange as a fluorescent probe: Label-free and selective detection of silver ions based on the structural change of i-motif DNA at neutral pH. *Talanta* (2016). doi:10.1016/j.talanta.2016.05.006
54. Abdelhamid, M. A. S. *et al.* Redox-dependent control of i-Motif DNA structure using copper cations. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky390
55. 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). doi:10.1021/jm900055s
56. Jonchhe, S. *et al.* Decreased water activity in nanoconfinement contributes to the folding of G-quadruplex and i-motif structures. *Proc. Natl. Acad. Sci. U. S. A.* (2018). doi:10.1073/pnas.1805939115
57. Kang, H. J., Kendrick, S., Hecht, S. M. & Hurley, L. H. The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja4109352
58. Roy, B. *et al.* Interaction of Individual Structural Domains of hnRNP LL with the

BCL2 Promoter i-Motif DNA. *J. Am. Chem. Soc.* (2016). doi:10.1021/jacs.6b05036
59. Lannes, L., Young, P., Richter, C., Morgner, N. & Schwalbe, H. Interaction of the N-Terminal Tandem Domains of hnRNP LL with the BCL2 Promoter i-Motif DNA Sequence. *ChemBioChem* (2017). doi:10.1002/cbic.201700390
60. Niu, K. *et al.* BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in Bombyx mori. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1207
61. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* (2013). doi:10.1038/nchem.1548
62. 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). doi:10.1038/nchem.1805
63. Simone, R., Fratta, P., Neidle, S., Parkinson, G. N. & Isaacs, A. M. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Letters* (2015). doi:10.1016/j.febslet.2015.05.003
64. Monchaud, D. & Teulade-Fichou, M. P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* (2008). doi:10.1039/b714772b
65. Anantha, N. V., Azam, M. & Sheardy, R. D. Porphyrin binding to quadruplexed T4G4. *Biochemistry* (1998). doi:10.1021/bi973009v
66. Grand, C. L. *et al.* The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* (2002).
67. Parkinson, G. N., Ghosh, R. & Neidle, S. Structural basis for binding of porphyrin to human telomeres. *Biochemistry* (2007). doi:10.1021/bi062244n
68. Shin-ya, K. *et al.* Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus [17]. *Journal of the American Chemical Society* (2001). doi:10.1021/ja005780q
69. Gomez, D. *et al.* Interaction of telomestatin with the telomeric single-strand overhang. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M406123200
70. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. P. & Phan, A. T. Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew. Chemie - Int. Ed.* (2014). doi:10.1002/anie.201308063
71. Huang, H. S. *et al.* Synthesis and human telomerase inhibition of a series of regioisomeric disubstituted amidoanthraquinones. *Chem. Pharm. Bull.* (2007). doi:10.1248/cpb.55.284
72. Rahman, K. M. *et al.* Biaryl polyamides as a new class of DNA quadruplex-binding ligands. *Chem. Commun.* (2009). doi:10.1039/b902359c
73. Zagotto, G. *et al.* Tuning G-quadruplex vs double-stranded DNA recognition in regioisomeric lysyl-peptidyl-anthraquinone conjugates. *Bioconjug. Chem.* (2011).

doi:10.1021/bc200389w
74. Sissi, C. & Palumbo, M. Telomeric G-Quadruplex Architecture and Interactions with Potential Drugs. *Curr. Pharm. Des.* (2014). doi:10.2174/1381612820666140630094300
75. Das, R. N. *et al.* Design, synthesis and biological evaluation of new substituted diquinolinyl-pyridine ligands as anticancer agents by targeting G-Quadruplex. *Molecules* (2018). doi:10.3390/molecules23010081
76. Jain, A. K. & Bhattacharya, S. Groove binding ligands for the interaction with parallel-stranded ps-duplex DNA and triplex DNA. *Bioconjug. Chem.* (2010). doi:10.1021/bc900247s
77. Pagano, B. *et al.* Thermodynamic analysis of quadruplex DNA-drug interaction. in *Nucleosides, Nucleotides and Nucleic Acids* (2007). doi:10.1080/15257770701499069
78. Lin, C. & Yang, D. Human telomeric G-quadruplex structures and G-quadruplex-interactive compounds. in *Methods in Molecular Biology* (2017). doi:10.1007/978-1-4939-6892-3_17
79. Rocca, R. *et al.* Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization. *Biochim. Biophys. Acta - Gen. Subj.* (2017). doi:10.1016/j.bbagen.2016.12.023
80. Amato, J. *et al.* Targeting the BCL2 Gene Promoter G-Quadruplex with a New Class of Furopyridazinone-Based Molecules. *ChemMedChem* (2018). doi:10.1002/cmdc.201700749
81. Che, T. *et al.* Natural alkaloids and heterocycles as G-quadruplex ligands and potential anticancer agents. *Molecules* (2018). doi:10.3390/molecules23020493
82. Fedoroff, O. Y., Rangan, A., Chemeris, V. V. & Hurley, L. H. Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism. *Biochemistry* (2000). doi:10.1021/bi001528j
83. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* (2002). doi:10.1073/pnas.182256799
84. Fernández, S., Eritja, R., Aviñó, A., Jaumot, J. & Gargallo, R. Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the i-motif formed by the 5'-(C3TA2)4-3' sequence of the human telomere. *Int. J. Biol. Macromol.* (2011). doi:10.1016/j.ijbiomac.2011.07.004
85. Qin, T., Liu, K., Song, D., Yang, C. & Su, H. Porphyrin Bound to i-Motifs: Intercalation versus External Groove Binding. *Chem. - An Asian J.* (2017). doi:10.1002/asia.201700398
86. Alberti, P. *et al.* Interaction of an acridine dimer with dna quadruplex structures. *J. Biomol. Struct. Dyn.* (2001). doi:10.1080/07391102.2001.10506758

87. Wang, L., Wu, Y., Chen, T. & Wei, C. The interactions of phenanthroline compounds with DNAs: Preferential binding to telomeric quadruplex over duplex. *Int. J. Biol. Macromol.* (2013). doi:10.1016/j.ijbiomac.2012.08.015
88. Cristofari, C., Rigo, R., Greco, M. L., Ghezzo, M. & Sissi, C. pH-driven conformational switch between non-canonical DNA structures in a C-rich domain of EGFR promoter. *Sci. Rep.* (2019). doi:10.1038/s41598-018-37968-8
89. Shi S, Zhao J, Geng X, Yao T, Huang H, Liu T, Zheng L, Li Z, Yang D, J. L. Molecular 'light switch' for G-quadruplexes and i-motif of human telomeric DNA: [Ru(phen)2(dppz)]2+. *Dalt. Trans* 2490–3 (2010). doi:10.1039/b916094a [Indexed for MEDLINE]
90. Shi S, Geng X, Zhao J, Yao T, Wang C, Yang D, Zheng L, J. L. Interaction of [Ru(bpy)2(dppz)]2+ with human telomeric DNA: preferential binding to G-quadruplexes over i-motif. *Biochimie.* **92**, 370–7 (2010).
91. Li, X., Peng, Y., Ren, J. & Qu, X. Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. *Proc. Natl. Acad. Sci. U. S. A.* (2006). doi:10.1073/pnas.0607245103
92. Zhao, C., Ren, J. & Qu, X. Single-walled carbon nanotubes binding to human telomeric i-motif DNA under molecular-crowding conditions: More water molecules released. *Chem. - A Eur. J.* (2008). doi:10.1002/chem.200800280
93. Chen, Y. *et al.* Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat. Commun.* (2012). doi:10.1038/ncomms2091
94. Sun, H., Ren, J. & Qu, X. Carbon Nanomaterials and DNA: From Molecular Recognition to Applications. *Accounts of Chemical Research* (2016). doi:10.1021/acs.accounts.5b00515
95. Chen, X. *et al.* Stabilization and induction of oligonucleotide I-motif structure via graphene quantum dots. *ACS Nano* (2013). doi:10.1021/nn304673a
96. Kendrick, S. *et al.* The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja410934b
97. Kaiser, C. E. *et al.* Insight into the Complexity of the i-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.7b02046
98. Brown, R. V. *et al.* The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of. *J. Am. Chem. Soc.* (2017). doi:10.1021/jacs.6b10028
99. Dang, C. V. *et al.* Function of the c-Myc oncogenic transcription factor. *Exp. Cell*

Res. (1999). doi:10.1006/excr.1999.4686

100. Nair, S. K. & Burley, S. K. X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* (2003). doi:10.1016/S0092-8674(02)01284-9
101. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E. & Bishop, J. M. Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* (1983). doi:10.1073/pnas.80.6.1707
102. Collins, S. J., Robertson, K. A. & Mueller, L. Retinoic acid-induced granulocytic differentiation of HL-60 myeloid leukemia cells is mediated directly through the retinoic acid receptor (RAR-alpha). *Mol. Cell. Biol.* (1990). doi:10.1128/mcb.10.5.2154
103. Graham, M., Adams, J. M. & Cory, S. Murine T lymphomas with retroviral inserts in the chromosomal 15 locus for plasmacytoma variant translocations. *Nature* (1985). doi:10.1038/314740a0
104. Gazin, C. *et al.* Nucleotide sequence of the human c-myc locus: provocative open reading frame within the first exon. *EMBO J.* (1984). doi:10.1002/j.1460-2075.1984.tb01816.x
105. Spencer, C. A. & Groudine, M. Control of c-myc regulation in normal and neoplastic cells. *Adv. Cancer Res.* (1991).
106. Levens, D. *et al.* DNA conformation, topology, and the regulation of c-myc expression. in *Current Topics in Microbiology and Immunology* (1997).
107. Postel, E. H., Mango, S. E. & Flint, S. J. A nuclease-hypersensitive element of the human c-myc promoter interacts with a transcription initiation factor. *Mol. Cell. Biol.* (1989). doi:10.1128/mcb.9.11.5123
108. Simonsson, T., Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* (1998). doi:10.1093/nar/26.5.1167
109. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* (2005). doi:10.1021/bi048242p
110. Yang, D. & Hurley, L. Structure of the biologically relevant g-quadruplex in the c-MYC promoter. *Nucleosides, Nucleotides and Nucleic Acids* (2006). doi:10.1080/15257770600809913
111. Brys, A. & Maizels, N. LR1 regulates c-myc transcription in B-cell lymphomas. *Proc. Natl. Acad. Sci. U. S. A.* (1994). doi:10.1073/pnas.91.11.4915
112. 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. *J. Biol. Chem.* (1999). doi:10.1074/jbc.274.2.1066

113. 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). doi:10.1074/jbc.M109.018028
114. Li, Z. *et al.* Quinazoline derivative QPB-15e stabilizes the c-myc promoter G-quadruplex and inhibits tumor growth in vivo. *Oncotarget* (2016). doi:10.18632/oncotarget.9088
115. Simonsson, T., Pribylova, M. & Vorlickova, M. A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem. Biophys. Res. Commun.* (2000). doi:10.1006/bbrc.2000.3783
116. Dai, J., Ambrus, A., Hurley, L. H. & Yang, D. A direct and nondestructive approach to determine the folding structure of the I-motif DNA secondary structure by NMR. *J. Am. Chem. Soc.* (2009). doi:10.1021/ja900967r
117. Dhakal, S. *et al.* G-quadruplex and i-motif are mutually exclusive in ILPR double-stranded DNA. *Biophys. J.* (2012). doi:10.1016/j.bpj.2012.04.024
118. Cui, Y., Kong, D., Ghimire, C., Xu, C. & Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* (2016). doi:10.1021/acs.biochem.6b00016
119. Takimoto, M. *et al.* Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. *J. Biol. Chem.* (1993).
120. Tomonaga, T. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J. Biol. Chem.* (1995). doi:10.1074/jbc.270.9.4875
121. Tomonaga, T. & Levens, D. Activating transcription from single stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.* (1996). doi:10.1073/pnas.93.12.5830
122. Shu, B. *et al.* Syntheses and evaluation of new acridone derivatives for selective binding of oncogene c-: Myc promoter i-motifs in gene transcriptional regulation. *Chem. Commun.* (2018). doi:10.1039/c8cc00328a
123. Shu, B. *et al.* Syntheses and evaluation of new Quinoline derivatives for inhibition of hnRNP K in regulating oncogene c-myc transcription. *Bioorg. Chem.* (2019). doi:10.1016/j.bioorg.2018.12.020
124. Ashkenazi, A., Fairbrother, W. J., Leverson, J. D. & Souers, A. J. From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. *Nature Reviews Drug Discovery* (2017). doi:10.1038/nrd.2016.253
125. Cui, J. & Placzek, W. J. Post-transcriptional regulation of anti-apoptotic BCL2 family members. *International Journal of Molecular Sciences* (2018). doi:10.3390/ijms19010308
126. Siddiqui, W. A., Ahad, A. & Ahsan, H. The mystery of BCL2 family: Bcl-2 proteins and apoptosis: an update. *Archives of Toxicology* (2015). doi:10.1007/s00204-014-1448-7
127. Cory, S. & Adams, J. M. The BCL2 family: Regulators of the cellular life-or-death

switch. *Nature Reviews Cancer* (2002). doi:10.1038/nrc883
128. Tsujimoto, Y. & Croce, C. M. Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* (1986). doi:10.1073/pnas.83.14.5214
129. Lang, G., Gombert, W. M. & Gould, H. J. A transcriptional regulatory element in the coding sequence of the human Bcl-2 gene. *Immunology* (2005). doi:10.1111/j.1365-2567.2004.02073.x
130. Sengupta, P., Chattopadhyay, S. & Chatterjee, S. G-Quadruplex surveillance in BCL-2 gene: a promising therapeutic intervention in cancer treatment. *Drug Discovery Today* (2017). doi:10.1016/j.drudis.2017.05.001
131. Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* (2006). doi:10.1021/ja0563861
132. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkl610
133. Dai, J. *et al.* An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* (2006). doi:10.1021/ja055636a
134. Wang, X. D. *et al.* Turning off transcription of the bcl-2 gene by stabilizing the bcl-2 promoter quadruplex with quindoline derivatives. *J. Med. Chem.* (2010). doi:10.1021/jm100445e
135. Feng, Y. *et al.* Stabilization of G-quadruplex DNA and inhibition of Bcl-2 expression by a pyridostatin analog. *Bioorganic Med. Chem. Lett.* (2016). doi:10.1016/j.bmcl.2016.02.065
136. Agrawal, P., Lin, C., Mathad, R. I., Carver, M. & Yang, D. The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in k+ solution. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja4118945
137. Cheng Y, Tang Q, Li Y, Zhang Y, Zhao C, Yan J, Y. H. Folding/unfolding kinetics of G-quadruplexes upstream of the P1 promoter of the human BCL-2 oncogene. *J Biol Chem* 5890–5895 (2019). doi:10.1074/jbc.RA119.007516
138. Onel, B. *et al.* A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription. *J. Am. Chem. Soc.* (2016). doi:10.1021/jacs.5b08596
139. Sun, H. *et al.* A newly identified G-quadruplex as a potential target regulating Bcl-2 expression. *Biochim. Biophys. Acta - Gen. Subj.* (2014). doi:10.1016/j.bbagen.2014.07.014
140. Huppert, J. L., Bugaut, A., Kumari, S. & Balasubramanian, S. G-quadruplexes: The beginning and end of UTRs. *Nucleic Acids Res.* (2008). doi:10.1093/nar/gkn511
141. Zhang, D. H. & Zhi, G. Y. Structure monomorphism of RNA G-quadruplex that is

independent of surrounding condition. *J. Biotechnol.* (2010). doi:10.1016/j.jbiotec.2010.07.026

142. Shahid, R., Bugaut, A. & Balasubramanian, S. The BCL-2 5′ untranslated region contains an RNA G-quadruplex-forming motif that modulates protein expression. *Biochemistry* (2010). doi:10.1021/bi100957h
143. Bugaut, A. & Balasubramanian, S. 5′-UTR RNA G-quadruplexes: Translation regulation and targeting. *Nucleic Acids Research* (2012). doi:10.1093/nar/gks068
144. Beaudoin, J. D. & Perreault, J. P. Exploring mRNA 3′-UTR G-quadruplexes: Evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt265
145. Zhang, D. H. *et al.* Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry* (2010). doi:10.1021/bi1002822
146. Sutherland, C., Cui, Y., Mao, H. & Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* (2016). doi:10.1021/jacs.6b09196
147. Dembska, A., Kierzek, E. & Juskowiak, B. Studying the influence of stem composition in pH-sensitive molecular beacons onto their sensing properties. *Anal. Chim. Acta* (2017). doi:10.1016/j.aca.2017.07.040
148. Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* (2011). doi:10.1016/j.cell.2011.08.017
149. Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H. & Brooks, T. A. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.274720
150. Michelotti, G. A. *et al.* Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol. Cell. Biol.* (1996). doi:10.1128/mcb.16.6.2656
151. Baber, J. L., Libutti, D., Levens, D. & Tjandra, N. High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol.* (1999). doi:10.1006/jmbi.1999.2818
152. Braddock, D. T., Baber, J. L., Levens, D. & Clore, G. M. Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J.* (2002). doi:10.1093/emboj/cdf352
153. Backe, P. H., Messias, A. C., Ravelli, R. B. G., Sattler, M. & Cusack, S. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* (2005). doi:10.1016/j.str.2005.04.008
154. Korycka-Wolowiec, A., Wolowiec, D., Kubiak-Mlonka, A. & Robak, T. Venetoclax

in the treatment of chronic lymphocytic leukemia. *Expert Opin. Drug Metab. Toxicol.* (2019). doi:10.1080/17425255.2019.1606211
155. Ferri, E., Petosa, C. & McKenna, C. E. Bromodomains: Structure, function and pharmacology of inhibition. *Biochemical Pharmacology* (2016). doi:10.1016/j.bcp.2015.12.005
156. Bishop, G. R. & Chaires, J. B. Characterization of DNA Structures by Circular Dichroism. in *Current Protocols in Nucleic Acid Chemistry* (2002). doi:10.1002/0471142700.nc0711s11
157. Vorlíčková, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* (2012). doi:10.1016/j.ymeth.2012.03.011
158. Randazzo, A., Spada, G. P. & Da Silva, M. W. Circular dichroism of quadruplex structures. *Top. Curr. Chem.* (2013). doi:10.1007/128-2012-331
159. Mergny, J. L., Li, J., Lacroix, L., Amrane, S. & Chaires, J. B. Thermal difference spectra: A specific signature for nucleic acid structures. *Nucleic Acids Res.* (2005). doi:10.1093/nar/gni134
160. Tóthová, P., Krafčíková, P. & Víglaský, V. Formation of highly ordered multimers in G-quadruplexes. *Biochemistry* (2014). doi:10.1021/bi500773c
161. Berselli, M., Lavezzo, E. & Toppo, S. NeSSie: A tool for the identification of approximate DNA sequence symmetries. *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty142
162. Endoh, T. & Sugimoto, N. Conformational dynamics of the RNA G-quadruplex and its effect on translation efficiency. *Molecules* (2019). doi:10.3390/molecules24081613
163. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* (2003). doi:10.1093/nar/gkg595
164. Agarwal, T., Roy, S., Kumar, S., Chakraborty, T. K. & Maiti, S. In the sense of transcription regulation by G-quadruplexes: Asymmetric effects in sense and antisense strands. *Biochemistry* (2014). doi:10.1021/bi401451q
165. Manzini, G., Yathindra, N. & Xodo, L. E. Evidence for intramolecularly folded I-DNA structures in biologically relevant CCC-repeat sequences. *Nucleic Acids Res.* (1994). doi:10.1093/nar/22.22.4634
166. Bucek, P., Gargallo, R. & Kudrev, A. Spectrometric study of the folding process of i-motif-forming DNA sequences upstream of the c-kit transcription initiation site. *Anal. Chim. Acta* (2010). doi:10.1016/j.aca.2010.10.008
167. Vorlíčková, M., Kejnovská, I., Bednářová, K., Renčiuk, D. & Kypr, J. Circular dichroism spectroscopy of DNA: From duplexes to quadruplexes. in *Chirality* (2012). doi:10.1002/chir.22064
168. Müller, J. Metal-mediated base pairs in parallel-stranded DNA. *Beilstein J. Org. Chem.* (2017). doi:10.3762/bjoc.13.265
169. Jash, B. & Müller, J. Metal-Mediated Base Pairs: From Characterization to

Application. *Chemistry - A European Journal* (2017). doi:10.1002/chem.201703518
170. Jash, B. *et al.* A metal-mediated base pair that discriminates between the canonical pyrimidine nucleobases. *Chem. Sci.* (2017). doi:10.1039/c6sc03482a
171. Katz, S. Mechanism of the reaction of polynucleotides and HgII. *Nature* (1962). doi:10.1038/194569a0
172. Katz, S. The reversible reaction of Hg (II) and double-stranded polynucleotides a step-function theory and its significance. *BBA - Biochim. Biophys. Acta* (1963).
173. Miyake, Y. *et al.* MercuryII-mediated formation of thymine-HgII-thymine base pairs in DNA duplexes. *J. Am. Chem. Soc.* (2006). doi:10.1021/ja056354d
174. Eichhorn, G. L., Butzow, J. J., Clark, P. & Tarien, E. Interaction of metal ions with polynucleotides and related compounds. X. Studies on the reaction of silver(I) with the nucleosides and polynucleotides, and the effect of silver(I) on the zinc(II) degradation of polynucleotides. *Biopolymers* (1967). doi:10.1002/bip.1967.360050306
175. Ono, A. *et al.* Specific interactions between silver(i) ions and cytosine-cytosine pairs in DNA duplexes. *Chem. Commun.* (2008). doi:10.1039/b808686a
176. Megger, D. A., Fonseca Guerra, C., Bickelhaupt, F. M. & Müller, J. Silver(I)-mediated Hoogsteen-type base pairs. *J. Inorg. Biochem.* (2011). doi:10.1016/j.jinorgbio.2011.07.005
177. Torigoe, H., Miyakawa, Y., Ono, A. & Kozasa, T. Thermodynamic properties of the specific binding between Ag+ ions and C:C mismatched base pairs in duplex DNA. *Nucleosides, Nucleotides and Nucleic Acids* (2011). doi:10.1080/15257770.2011.553210
178. Fortino, M., Marino, T. & Russo, N. Theoretical study of silver-ion-mediated base pairs: The case of C-Ag-C and C-Ag-A systems. *J. Phys. Chem. A* (2015). doi:10.1021/jp5096739
179. Chaudhari, S. G. & Saxena, S. Effect of Divalent ions on the structure and stability of c-tetraplex. *Chem. Sci. Rev. Lett.* (2018).
180. Berdakin, M., Steinmetz, V., Maitre, P. & Pino, G. A. Gas phase structure of metal mediated (Cytosine)2Ag+ mimics the hemiprotonated (Cytosine)2H+ Dimer in i -motif folding. *J. Phys. Chem. A* (2014). doi:10.1021/jp5038969
181. Swasey, S. M., Rosu, F., Copp, S. M., Gabelica, V. & Gwinn, E. G. Parallel Guanine Duplex and Cytosine Duplex DNA with Uninterrupted Spines of Ag I -Mediated Base Pairs. *J. Phys. Chem. Lett.* (2018). doi:10.1021/acs.jpclett.8b02851
182 Morelli, M. J., Allen, R. J. & Rein Ten Wolde, P. Effects of macromolecular crowding on genetic networks. *Biophys. J.* (2011). doi:10.1016/j.bpj.2011.10.053
183. Theillet, F. X. *et al.* Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). *Chemical Reviews* (2014). doi:10.1021/cr400695p

184. Matsuda, H., Putzel, G. G., Backman, V. & Szleifer, I. Macromolecular crowding as a regulator of gene transcription. *Biophys. J.* (2014). doi:10.1016/j.bpj.2014.02.019
185. Nakano, S. I., Miyoshi, D. & Sugimoto, N. Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chemical Reviews* (2014). doi:10.1021/cr400113m
186. Nakano, S. ichi & Sugimoto, N. Model studies of the effects of intracellular crowding on nucleic acid interactions. *Molecular BioSystems* (2017). doi:10.1039/c6mb00654j
187. Cui, J., Waltman, P., Le, V. H. & Lewis, E. A. The effect of molecular crowding on the stability of human c-MYC promoter sequence I-motif at neutral pH. *Molecules* (2013). doi:10.3390/molecules181012751
188. Saxena, S., Joshi, S., Shankaraswamy, J., Tyagi, S. & Kukreti, S. Magnesium and molecular crowding of the cosolutes stabilize the i-motif structure at physiological pH. *Biopolymers* (2017). doi:10.1002/bip.23018
189. Miyoshi, D., Nakao, A. & Sugimoto, N. Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry* (2002). doi:10.1021/bi020412f
190. Miyoshi, D., Karimata, H. & Sugimoto, N. Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. *J. Am. Chem. Soc.* (2006). doi:10.1021/ja061267m
191. Xu, L. *et al.* Berberine as a novel light-up i-motif fluorescence ligand and its application in designing molecular logic systems. *Chem. Commun.* (2016). doi:10.1039/c5cc08242k
192. Pagano, A. *et al.* Common G-quadruplex binding agents found to interact with i-motif-forming DNA: Unexpected multi-target-directed compounds. *Front. Chem.* (2018). doi:10.3389/fchem.2018.00281
193. Abdelhamid, M. A. S., Gates, A. J. & Waller, Z. A. E. Destabilization of i-Motif DNA at Neutral pH by G-Quadruplex Ligands. *Biochemistry* (2019). doi:10.1021/acs.biochem.8b00968
194. De Cian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* (2007). doi:10.1016/j.ymeth.2006.10.004
195. Wright, E. P. *et al.* Mitoxantrone and analogues bind and stabilize i-motif forming DNA sequences. *Sci. Rep.* (2016). doi:10.1038/srep39456
196. Arora, A. *et al.* Binding of berberine to human telomeric quadruplex - Spectroscopic, calorimetric and molecular modeling studies. *FEBS J.* (2008). doi:10.1111/j.1742-4658.2008.06541.x
197. Ji, X. *et al.* The interaction of telomeric DNA and C-myc22 G-quadruplex with 11 natural alkaloids. *Nucleic Acid Ther.* (2012). doi:10.1089/nat.2012.034

7 APPENDIX