UNIVERSIDADE DA BEIRA INTERIOR
Ciências da Saúde

Molecular Modelling of a novel G-quadruplex structure and its interaction with ligands

Rui Pedro Abrantes Carrilho

Dissertação para obtenção do Grau de Mestre em
Ciências Biomédicas
(2º ciclo de estudos)

Orientador: Prof. Carla Cruz
Co-orientador: Prof. Fani Sousa

Covilhã, Outubro de 2018

Agradecimentos

Firstly, I must thank my supervisor, Dr. Carla Cruz, and my co-supervisor, Dr. Fani Sousa, for all the help they gave me. Without their patience, guidance, support, and advice provided throughout the year, this work would never have seen fruition.

To the great Tiago Santos and Josué Carvalho, who stuck with me throughout the year, lending me aid whenever I needed, and without whom this work would not be half as good as it has any right to be, I give my sincerest thanks. Gentlemen, it's been an honor.

To the many people who helped me see this year through. Whether it was the "Icons" Nádia, Ricardo and Elizabete, who put up with me down at the lab, or Zé, Duarte and Bruno over at the old Masters students room, or Vanessa, with our online chats, or my fantastic lab bro, Henrique, with a flawless taste in anime and strategy games: you made long hours fly, and I wouldn't have had it any other way.

To my family, especially my Mom, Dad, Filipe and grandparents. For their enduring support, guidance, and help, and giving me the chance I needed to finish this, I thank you.

To my grandpa. I wish you could have seen this work to the end. I hope you're looking down at this with a smile on your face. I did it.

Resumo alargado

O ADN pode existir sob a forma de diversas estruturas, contrariamente ao que a vasta maioria da população pensa, ao imaginar a dupla hélice de Watson e Crick. Uma das formas que tem sido mais investigada ultimamente consiste no G-quadruplex. Esta estrutura não canónica do DNA ocorre quando guaninas se emparelham e organizam em estruturas cíclicas através de pontes de hidrogénio *Hoogsteen*, chamadas G-quartetos. Estas estruturas formam-se por empilhamento π-π entre elas próprias, originando o G-quadruplex, desde que haja um catião (preferivelmente K^+) para assumir uma localização central entre todos os quartetos.

Estas estruturas desempenham funções importantes a nível de regulação da transcrição e replicação do DNA. Alguns estudos indicam também que podem ser relevantes a nível de manutenção do DNA, e que várias secções do DNA humano se encontram num estado de equilíbrio entre a forma de G-quadruplex e duplex. São também considerados alvos para certas abordagens terapêuticas a nível do cancro. Por exemplo, vários oncogenes como c-kit e c-myc têm a capacidade de formar G-quadruplexes nos seus promotores. Controlando a forma que estes genes assumem, seria possível controlar a sua transcrição, e possivelmente impedir a formação de cancro. Outra possibilidade cinge-se à inibição da telomerase, uma enzima responsável pela replicação celular, que está sobreexpressa em células cancerígenas. Se uma parte do telómero assumir uma estrutura em G-quadruplex, a ação desta enzima fica inibida, efetivamente parando a progressão do cancro.

Portanto, torna-se necessário induzir e estabilizar a formação de estruturas do G-quadruplex. A estratégia é utilizar ligandos que interajam por interações intermoleculares de forma a estabilizar a estrutura do G-quadruplex, e outra topologia que esteja em equilíbrio. No entanto, analisando a literatura, conclui-se que apenas alguns grupos de ligandos são efetivamente ligandos de G-quadruplex.

Este trabalho de investigação teve como objetivo comparar 7 ligandos promissores da estrutura de G-quadruplex designada por pre-miR-149 literatura. Os ligandos selecionados foram macrociclos derivados de fenantrolina ([16]phenN_2, [32]phen$_2N_4$, Phen-DC_3, e derivados de laranja de acridina C_8 e C_8-NH_2. Determinou-se a afinidade e a estabilização destes ligandos com a estrutura do RNA G-quadruplex, a pre-miR-149. Isso será feito em duas etapas principais. Primeiro, foram realizadas simulações computacionais para determinar quais os ligandos mais promissores e quais os seus métodos de interação com a estrutura G-quadruplex. Estas dividiram-se em três passos: primeiro, foram geradas as estruturas da sequência e de cada ligando em software adequado. Segundo, foram feitas simulações de *docking* de modo a averiguar os locais de ligação de cada ligando ao G-quadruplex, e a conformação e interações entre o ligando e o quadruplex, sendo também calculadas energias de ligação entre o ligando e o G-quadruplex. Finalmente, foram feitas simulações de dinâmica molecular sobre como essa

conformação evoluiria num ambiente fisiológico simulado e calculadas novas energias de ligação, que comparadas entre si, revelam diferenças de afinidades entre os ligandos. Após estas técnicas computacionais, foram executadas técnicas biofísicas, como espetroscopia de dicroísmo circular e estudos de desnaturação térmica, e espectroscopia de fluorescência para determinar experimentalmente as afinidades de cada ligando para com a estrutura escolhida. Foram também executadas experiências de cromatografia de afinidade para determinar o comportamento de um ligando para com sequência do RNA G-quadruplex, a pre-miR-149. O programa usado para avaliar as conformações iniciais gerou estruturas demasiado rígidas e pouco flexíveis com os ligandos macrocíclicos [16]phenN_2 e [32]phen$_2N_4$. As energias de ligação obtidas revelaram a nível de afinidade a seguinte ordem decrescente: piridostatina > [32]phen$_2N_4$ > [16]phenN_2 > PhenDC_3 > L-arginina > C_8 > C_8-NH_2. Esta tendência não foi a mesma verificada experimentalmente, e logo, foi descartada. A nível destas experiências, retiram-se maioritariamente apenas as conformações dos ligandos que não são macrociclos.

A nível das experiências de dicroísmo circular mencionadas, as variações de temperatura de desnaturação térmica ligando-quadruplex foram diferentes,verificando-se a seguinte ordem: C_8 > piridostatina > C_8-NH_2 > [16]phenN_2. Seguidamente, foram realizadas titulações por espectroscopia de fluorescência as quais revelaram a seguinte tendência: C_8 > C_8-NH_2 > [16]phenN_2. De notar que apenas quatro dos sete ligandos ([16]phenN_2, [32]phen$_2N_4$, C_8 and C_8-NH_2) possuíam fluorescência intrínseca, e que desses, apenas estes três puderam ser selecionados. Estes resultados mostraram que a piridostatina, e derivados de laranja de acridina C_8 e C_8-NH_2 apresentaram maior afinidade para esta estrutura de G-quadruplex.

Por último, os resultados de cromatografia de afinidade revelaram que o ligando C_8-NH_2 tem maior afinidade com o RNA G-quadruplex pre-miR-149 . Das seis sequências testadas, três delas (c-myc, c-kit e pre-miR-149) formam G-quadruplexes com topologia paralela, e tiveram tempos de retenção mais altos. Outras sequências (TBA e AG23) formam G-quadruplexes com topologia antiparalela, e mostram tempos de retenção mais baixos. A sequência ds26 (duplex) teve o tempo de retenção mais baixo. Conclui-se que este ligando tem maior especificidade para com G-quadruplexes com topologia paralela em detrimento do duplex. As simulações de *docking* corroboram esta conclusão.

Deste modo, conclui-se que os melhores ligandos a nível de afinidade para com a sequência pre-miR-149 são os derivados de laranja de acridina C_8 e C_8-NH_2 e a piridostatina, de modo que futura investigação nesta área deve considerar estes três como fortes candidatos a ligandos de RNA G-quadruplex.

Abstract

DNA can exist under many different forms. Lately, G-quadruplexes, which are one example of the non-canonical DNA forms, have been getting a lot of attention due to the role they play in certain biological processes and as potential targets for therapeutic interventions. For example, these structures can exist in certain parts of the telomeres, structures responsible for cell replication. In cancer cells, if the enzyme telomerase could be inhibited, by inducing the formation of a G-quadruplex structure in guanine-rich telomere sequences, the spread of cancer cells would cease. For this and other reasons, it becomes important to be able to induce the formation of G-quadruplex structures and/or stabilize them, and one of the ways of doing so consists of targeting these sequences with ligands that have good affinity to G-quadruplex structures. However, few G-quadruplex ligands demonstrated the needed properties to fulfill the clinical needs, and further efforts to determine which would be better suited to target any particular sequence are needed. This work aimed at comparing the affinity to the pre-miR-149 G-quadruplex structure of seven promising ligands found in the literature, through the latest techniques fit for that purpose. The seven ligands tested were: [16]phenN_2, [32]phen$_2N_4$, phen-DC3, pyridostatin, acridine orange derivatives C_8 and C_8-NH_2 and L-arginine. Firstly, they underwent computational tests, with the molecular structure of the quadruplex and the ligand being simulated, and their optimal binding site and conformation found. Their binding energies were compared, and they underwent molecular dynamics runs to simulate their behavior in an environment with solvent, followed by another binding energy comparison. The trend obtained in order of decreasing binding affinity was: pyridostatin > [32]phen$_2N_4$ > [16]phenN_2 > Phen-DC3 > L-arginine > C_8 > C_8-NH_2. Biophysical techniques were then performed, to determine the binding affinities experimentally. First, circular dichroism spectroscopy and melting studies (performed on four ligands) established the following trend: C_8 > pyridostatin > C_8-NH_2 > [16]phenN_2. Fluorescence spectroscopy titration (performed on three) revealed a similar trend: C_8 > C_8-NH_2 > [16]phenN_2. Lastly, affinity chromatography experiments were held to test how other DNA sequences would bind to C8-NH_2. The results revealed that the ligand has better binding affinity with parallel quadruplexes over antiparallel ones, and poor binding with a duplex sequence. Overall, the best ligands identified for binding to the G-quadruplex structure were the acridine orange derivatives C_8 and C_8-NH_2, and pyridostatin. These three ligands should be considered prime candidates for further research in this area.

Keywords:

G-quadruplex; pre-miR-149; Molecular modelling; Circular dichroism; Fluorescence spectroscopy; Affinity chromatography

Table of Contents

Figures List

Tables List

Acronym List

A	Adenine
A-DNA	A-form of DNA
B-DNA	B form of DNA
bp	Base pairs
C	Cytosine
C_8	10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium iodide
C_8-NH_2	3,6-diamino-10-(8-aminooctyl)acridin-10-ium
CD	Circular dichroism
c-Kit	V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
cm	centimeter
c-Myc	V-myc avian myelocytomatosis viral oncogene homolog
DNA	Deoxyribonucleic acid
Duplex	DNA double strand
EDTA	Ethylene diamine tetra-acetic acid
G	Guanine
g	gram
G4	G-quadruplex
h	hours
kbp	Kilo base pairs
KD	Affinity constant
KRAS	Kirsten rat sarcoma viral oncogene homolog
L	liter
ln	linear
M	Molar
min	Minutes
mL	milliliter
mM	Millimolar
mRNA	Messenger RNA
NMR	Nuclear magnetic resonance
ºC	Celsius
pre-miR-149	Pre micro RNA-149
RCF	Relative centrifugal force
RNA	Ribonucleic acid
rpm	Rotations per minute
Tris	Tris(hydroximethyl)methylamine
UV	Ultraviolet light
Z-DNA	Z-form of DNA, left handed-DNA

µg	microgram
µL	Microliter

Chapter I

Introduction

1.1 - DNA history

The history of DNA began in 1869, when Friedrich Miescher successfully isolated the hereditary material. Miescher, a Swiss Physician who did not have enough recognition, discovered DNA by accident upon isolating it from leucocytes, and termed it "nuclein", as it was obtained from the nucleus of cells. Further analysis on the precipitated "nuclein" revealed the molecular building blocks of DNA, with high amounts of phosphorous. He also believed that nuclein played an important part in fertilization, but due to the low information at the time, he was unable to determine the biological implications of his discovery. [1]

Other scientists followed up Miescher's work by attempting to discover the chemical nature of "nuclein". The most prominent was Phoebus Levene, a Russian physicist turned biochemist. A prolific researcher, Levene discovered many new things about nucleic acids, namely, the structure of a nucleotide (phosphate-sugar-base), the carbohydrate components of RNA and DNA (ribose and deoxyribose, respectively), and most importantly, a "polynucleotide" model of the structure of nucleic acids. [2]

Around 75 years later, in 1944, Oswald Avery, along with his colleagues Maclin McCarty and Colin MacLeod, delivered the next discovery in this area. Upon doing numerous experiments involving a "transforming principle" in pneumococcus bacteria, they concluded that this transforming principle consisted of desoxyribonucleic acid, or DNA. In other words, unlike what was currently believed at the time, genes consisted of DNA, not protein. [3]

Chargaff built upon this knowledge by inventing a new paper chromatography method, which allowed him to separate and identify small amounts of organic material. This allowed him to discover two important facts about DNA: the fact that different species had different nucleotide compositions, and that, regardless of species, the amounts of adenine and thymine was often equal, as well as the amounts of guanine and cytosine (first parity rule of Chargaff). [4]

Over the next years, genes began to undergo a paradigm shift in the minds of researchers the world over. Prominent researchers, such as Nobel-Prize-winning physicist Erwin Schrödinger, Kurt Stern, Erwin Chargaff began to argue the role of genes as a "code", a means of "transmitting information" – partly due to that being the main paradigm in the scientific world during and after the Second World War. [5]

In 1953, Watson and Crick delivered the next piece of the puzzle by revealing the structure of DNA. Based on the previously described works, as well as important X-ray crystallography studies performed by Maurice Wilkins and Rosalind Franklin, the authors discovered the famous double helical structure of DNA. [6]

This structure has been further studied by other scientists, who discovered different conformations of the double helix. The most common conformation discovered by Watson and Crick in living cells was the B-DNA. There are other conformations, namely A-DNA (a dehydrated conformation seldom found in normal biological circumstances) and Z-DNA (a left-handed conformation that often comes about temporarily as a response to some biological activity). [7]

In 1910, a German chemist named Ivar Bang, discovered that guanylic acid forms a gel at certain concentrations. More than 50 years later, in 1962, Gellert and colleagues determined the molecular structure that resulted of that gel. X-ray diffraction studies revealed it to consist of tetrameric units assembled into large helical structures, more specifically consisting of four guanine residues connected by cyclic hydrogen bonds at the center. The overall structure is helical, with 4 units per turn of helix. [8]

After this discovery, there were few others until the late 1980's, when Elizabeth Blackburn discovered that telomeric DNA could form non-Watson Crick structures, with guanine-guanine base pairs, with much relevance for telomere function. [9] After that, James Williamson and others raised considerable interest in this area, when they discovered that some telomeric sequences in *Oxytricha* and *Tetrahymena* could form G-quadruplexes (G4)s in certain ionic conditions. [10] Since then, the interest in this area has grown.

1.2 - Definition and structure of G-quadruplex (G4)

G-quadruplexes (G4)s are DNA/RNA single-stranded secondary structures, formed by guanine-rich sequences. These structures are a result of the self-stacking of G-quartets (4 guanines arranged in a planar quartet, where every base is connected to 2 others through Hoogsteen base pairing).

The G4 are stabilized by monovalent ions ($K^+ > Na^+ > NH_4^+ > Li^+$, in decreasing order of strength of stabilization), as long as the ionic radius fits inside the G4 core. The ions intercalate into the anionic core of a G-tetrad, as well as bind between quartets, in order to stabilize the structure. The larger the ion, more oxygens can coordinate, and enhance the stabilizing effect. [11]

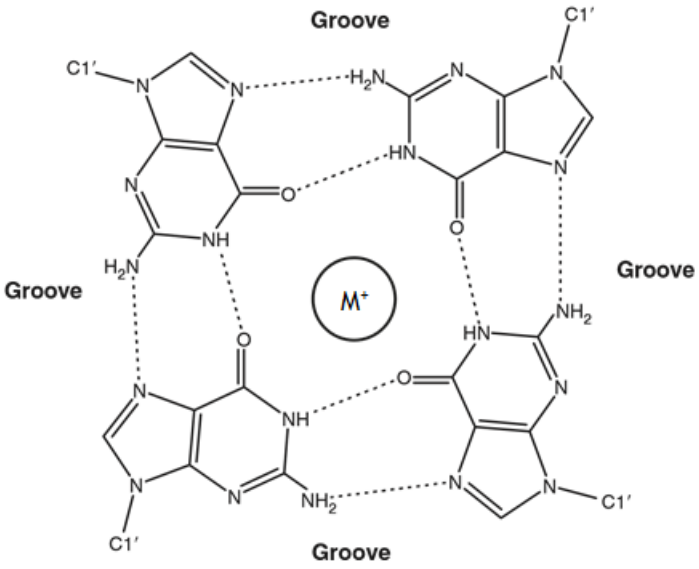

Figure 1 – The arrangement of guanine bases in a G-quartet, with a centrally placed metal ion. Adapted from [12].

The G4 structure can be defined according to numerous parameters, mostly being: number of strands involved, helical parameters of twist and rise, backbone strand polarity, connecting loop length and type, guanine glycosidic angles, groove widths and the nature of the ion channel. [12], [13]

A G4 is formed by one, two or four strands of DNA (or RNA), being respectively classified as monomeric, dimeric or tetrameric. Overall, a G4 consists of at least two stacked G-quartets, bound by loops consisting of mixed-sequence nucleotides not involved in the tetrads directly.

Unimolecular G4s usually can be described with the expression $G_m X_n G_m X_o G_m X_p G_m$, where m is the number of residues in each G tract, and X_n, X_o, X_p can be any combination of residues (including G) forming the loops. Di-molecular and tetramolecular G4s are less studied, but usually follow similar rules, except for the fact that the sequence of each strand does not necessarily have to be equal (though they often are).

The backbone polarity of any G4 refers to whether the strands are parallel or anti-parallel. If a single strand is anti-parallel to the others, the G4 has an antiparallel polarity, and otherwise, it would be parallel. Depending on the polarity, the G4 may end up adopting different loop types.

Parallel polarities exhibit propeller loops. These loops, also called “double chain reversal” have a shape, as the name indicates, like a propeller. Antiparallel polarities have two more loop type possibilities, in addition to propeller loops. They can also have lateral or edgewise loops (which join adjacent G4s strands). These loops can be located on the same or different faces of a G4, and the polarity of these strands may vary, as long as one of them is anti-parallel. The

other loop type is the diagonal loop, which joins opposing G4 strands. In this type of loop, there is always one anti-parallel strand and a parallel one.

Guanine glycosidic angles also tend to vary between different types of G4s, more specifically, due the strand polarity of the G4. In parallel G4s, every guanine glycosidic angle is in an *anti* conformation. Anti-parallel G4s can have both *syn* and *anti* glycosidic angles, although the arrangements of these angles depend on the topology and set of strand orientations of the G4, since the positions of the four strands in any topology differ between themselves.

Every G4 structure has four grooves, which are defined as the cavities bounded by the phosphodiester backbones. Their dimensions vary, according to the G4 topology and loop type. On G4s with only lateral or diagonal loops, the grooves are structurally simple, with walls bounded by monotonic sugar phosphodiester groups. G4s with propeller loops are more complex however, with more complex structural features, since the variable sequence loops are inserted into the grooves.

Regarding the ion channel, as was mentioned, G4s require monovalent cations in order to be formed and maintain stability. This is due to a strong negative electrostatic potential created by the guanine O6 oxygen atoms, forming a central channel which contains the ions in the center of the G-tetrad stack. Depending on the nature of the ion, their placement in the channel is different.

K^+ ions for instance are always equidistant from each tetrad plane, making the eight oxygen atoms assume a symmetrical tetragonal bipyramidal configuration. Na^+ ions on the other hand are on a plane with the G-tetrads. [12], [13]

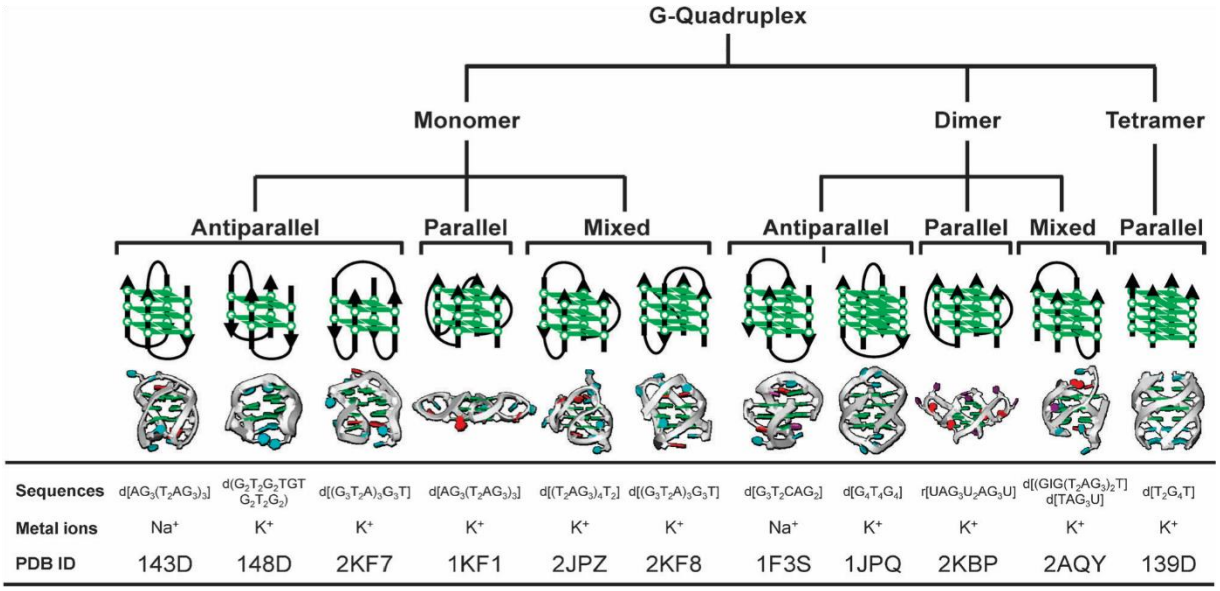

Figure 2 – Different G4 structures, according to number of strands, strand polarity, and ion channel nature. Adapted from [14].

1.2.1 - RNA G4s compared to DNA G4s

RNA G4s are different from DNA G4s in many ways. Firstly, the structure observes the same differences that exist between DNA and RNA in their natural forms, with the replacement of thymine for uracyl, and that of the deoxyribose sugar for a ribose sugar, respectively.

The presence of the extra 2 – OH' group in the ribose sugar results in more intramolecular interactions, as well as an added likeliness of bringing in more water molecules, making the structure more stable when compared to DNA G4s. The presence of uracyl instead of thymine in RNA G4s presumably also has an important stabilizing effect, as it improves the stability and decreases the hydration within the loop around the grooves. These factors make it so that RNA G4s have a greater thermodynamic stability, as well as being more compacted and less hydrated than DNA G4s.

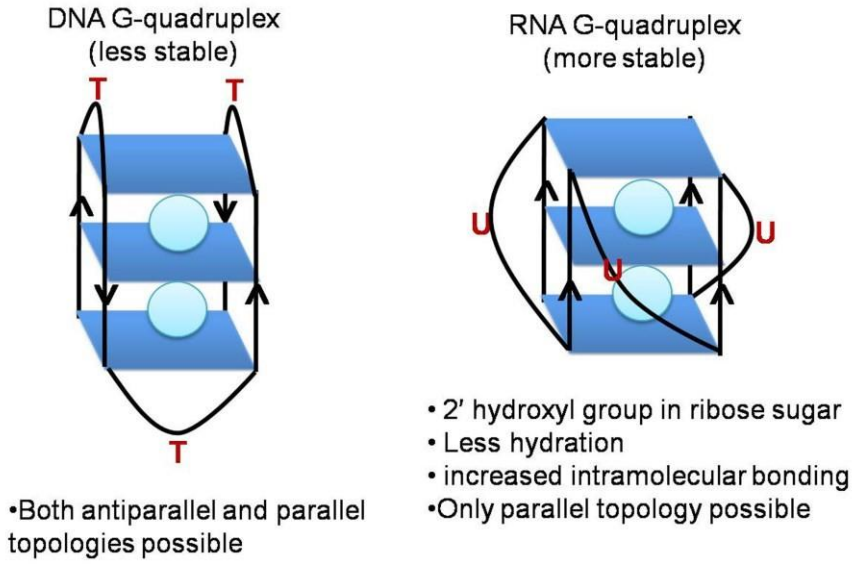

Figure 3 - Differences between DNA G4s and RNA G4s. Adapted from [15]

1.3 - Biological relevance of G4

There are guanine-rich domains in many regions of the human genome. This raises the possibility of these regions containing sequences that could form G4s. Huppart and Balasubramanian investigated how many sequences that could potentially form G4s, dubbed putative G4 sequences (PQS), existed in the human genome. They devised a "folding rule", which could determine whether it could form a G4 structure or not. By using an algorithm they created, the "quadparser", they identified 188.836 G-patterns and 187.610 C-patterns (a G-pattern in a complementary strand), which add up to 376.446 PQS throughout the entire human genome. [14] Remarkably, many of these PQS are not in random regions of the genome, but rather in specific ones, such as telomeric DNA, promotor regions of oncogenes, untranslated regions of RNA and immunoglobulin class switch regions. [15]

This would suggest that G4s are connected to many biological processes, predominantly around DNA replication and transcription, and indeed, many studies have confirmed it. Thus, G4s present themselves as important targets to improve genome stability andcontrol gene expression. [16]

One of the primary ways they are targeted is through the use of G4 ligands. These ligands are small synthesized or natural molecules that have high affinity and preference for the G4 structure over other DNA structures. They can bind to the G4 structure in a multitude of ways, that fit into three types: stacking, groove binding, or mixed binding (a mix between the two previous binding modes). They can affect the biological function of G4s in many ways. [17]

In regards to G4s abilities as a transcription regulator, several studies have shown that G4 motifs in the promotors of a certain gene can be targeted in order to modulate the transcription of the gene. Whether these changes are caused by G4 stabilization and subsequent mediation effects on the gene, or whether G4-ligand binding causes DNA damage is unclear however. [18]

Several studies revealed that the physiological role of G4s in diseased states is altered such as in cancer. For instance, many hallmark genes of cancer have their expression regulated by RNA G4s. As such, it has been suggested that these structures could pose an interesting therapeutic target for diagnosis and therapy of cancer, as well as other diseases, such as immune deficiency and development disorders. [19]

Certain studies suggest that G4s play an important role in maintaining DNA stability, as they can impede the progression of DNA polymerases. Further studies reveal that cancer genomes likely possess more G4 structures than normal ones, increasing their potential value as either cancer biomarkers and/or therapeutic agents. [18]

Indeed, Cogoi et al. suggested that some guanine-rich domains in an human oncogene, KRAS, exist in an equilibrium between a double-stranded and a folded G4 form. The G4 form inhibited the transcription of the gene, while the double-stranded form promoted it. Inhibiting the transcription of an oncogene carries an enormous potential in being able to prevent the progression of cancer. [20] Many important oncogenes, such as c-kit, c-myc and bcl-2 have already been singled out as potential targets for such therapies, with promising early results. [21]–[24]

Another possible way in which G4 can play a part in antitumor therapies is the inhibition of telomerase. This enzyme is overexpressed in cancer cells, with an activity that is 85-90% higher in cancer cells than in normal cells. The principle of this approach is that, by stabilizing the G4 structure that can be formed in many telomeric DNA sequences, telomerase activity can be inhibited, thus halting cancer growth. [25]

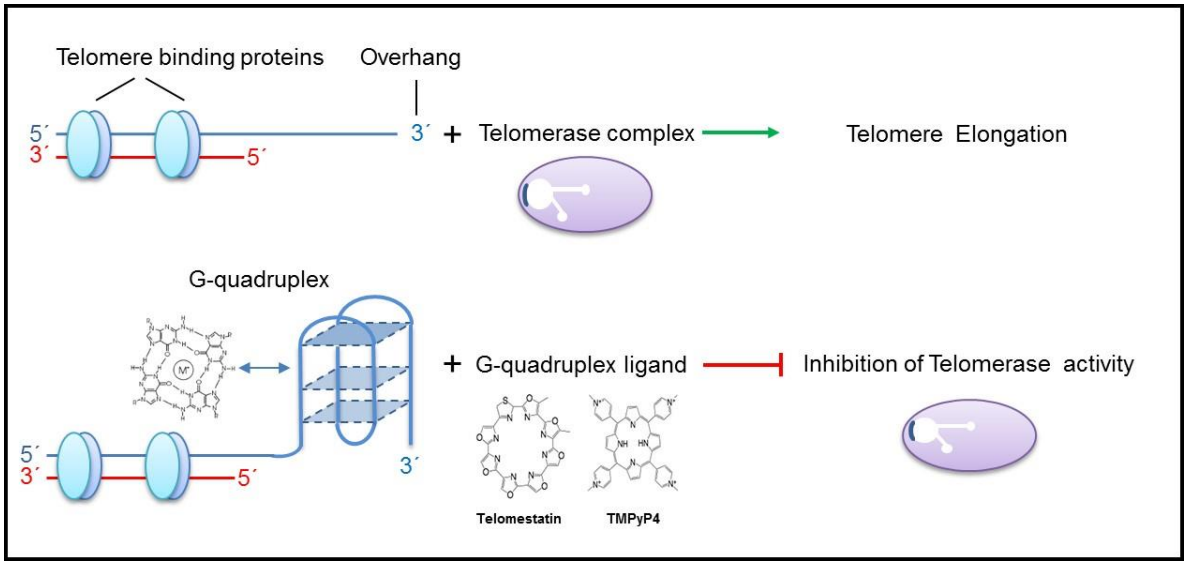

Figure 4 – The formation of a G4 structure by stabilizing it with a ligand inhibits telomerase activity. Adapted from [26].

Another way in which G4s can play a part in anticancer therapies lies in processes involving microRNAs (miRNAs). These miRNAs consist of short RNA sequences that play an important role in regulating gene expression, as well as maintaining the homeostasis of the cell. Their biogenesis involves Dicer-mediated cleavage, and this step is critical for the regulation of miRNA levels at a cellular level, with many disease conditions, such as cancer, being correlated with faulty regulation. [26], [27]

The biogenesis of miRNA involves a series of highly coordinated enzymatic cleavages that convert the primary forms of miRNAs (pri-miRNA) to a premature intermediate form (pre-miRNA), and finally into the final miRNA form. Pre-miRNAs assume a stem-loop/hairpin form, and this form is recognized by Dicer, which then proceeds to cleave it. If the stem-loop structure is somehow modified, as is the case of the G4 forming pre-miRNA sequence, the ensuing Dicer activity is significantly reduced. As such, there is a lower miRNA formation, and their regulatory effects are unremarkable. Thus, these pre-miRNAs are considered very good targets for therapeutic action against diseases caused by the aforementioned deregulations. [26], [27]

Arachchilage et al. tested what were the effects that the G4 formation has on the stem-loop structure of the pre-miRNA. Firstly, they identified 298 pre-miRNAs with a high likelihood of developing a G4 structure. They determined that in the presence of K^+, as opposed to the absence of the same ion, or the presence of Li^+, the G4 is formed, and there is an ensuing increase in the cleavage of the pre-miRNA. This proved that there is an unwinding of the stem loop structure when the G4 is formed. These findings suggest that the formation of G4s could indeed be related to pre-miRNA maturation, and thus could play a significant role in normal

cellular functions, as pre-miRNAs are involved in the silencing of many genes, such as cancer suppressor ones. [26]

1.4 - G4s aptamers

Oligonucleotide aptamers are a class of nucleic acids that bind to various targets. So far, they have shown great potential in being able to diagnose and/or treat several diseases. They are similar to antibodies in this regard, but show numerous advantages, such as their stability at room temperature, and low immunogenicity and toxicity. [28]

Some G4 aptamers have shown great potential as effective therapeutic agents, with promising anticancer and anti-HIV activity. For example, the G4 aptamer AGRO100 (also called AS1411) is a 26-mer oligonucleotide which forms a mixture of G4 structures and has cancer selective anti-proliferative activity, as it binds to nucleolin, a protein highly present in the surface of cancer cells. Numerous research groups have used it as a targeting agent for cancer cells, using it to deliver nanoparticles, oligonucleotides, and small ligands into them. [29]

The aptamers T30695 and 93del can inhibit HIV-1 integrase *in vitro*, being regarded as potential anti-HIV therapeutic agents. Both AGRO100 and T30865 were found to be hemin-binders, increasing their range of biomedical applications. [30] Thrombin-binding aptamers (TBAs), the most widely studied G4 forming aptamer class, can bind to α-thrombin, which can help to prevent thrombosis, and G4 stability and conformation details contribute greatly to the pharmacological properties of these aptamers. [31]

These aptamers' properties can be improved in many ways, such as by changing the backbone, sugars and bases of the aptamers in specific locations in the chain. [32] Overall, the biophysical and biological properties of the aptamers have been determined to stem from the G4s conformation and stability. [33]

1.5 - Targeting the G4 with small ligands

As abovementioned, many possible anticancer, anti-HIV, and other therapeutic strategies rely on the stabilization of G4 structures. Additionally, the stabilization of the aptamer's G4 structure may enhance its biological properties.

Many G4 binding ligands have been identified as having potential antiviral activity, such as: Braco-19 (which has shown anticancer activity *in vivo* via telomerase inhibition, along with antiviral activity against EBV and HIV-1), TMPyP4 (which also shows similar activities, despite being less selective than Braco-19), perylenes and naphthalene diimides (very versatile, suitable for chemical modifications to improve their pharmacological properties), pyridostatin (with excellent specificity to G4s but not duplex-DNA, also shown to have excellent anticancer properties in cell lines, along with antiviral applications against EBV) and bisquinolinium

derivatives (Phen-DC3, the best representative of this class has shown antireplicative activity against EBV and KSHV). [34]

Different ligands bind to G4 structures in different ways, as presented in Figure 5. The most frequently found binding mode is end-stacking, which many ligands adopt, such as porphyrins, telomestatin, cyanine dyes, quinazoline derivatives and others. Another form of binding often reported is groove-binding, with BMVC, distamycin-A, and other ligands, although they are fewer in number than the end-stackers. [35]

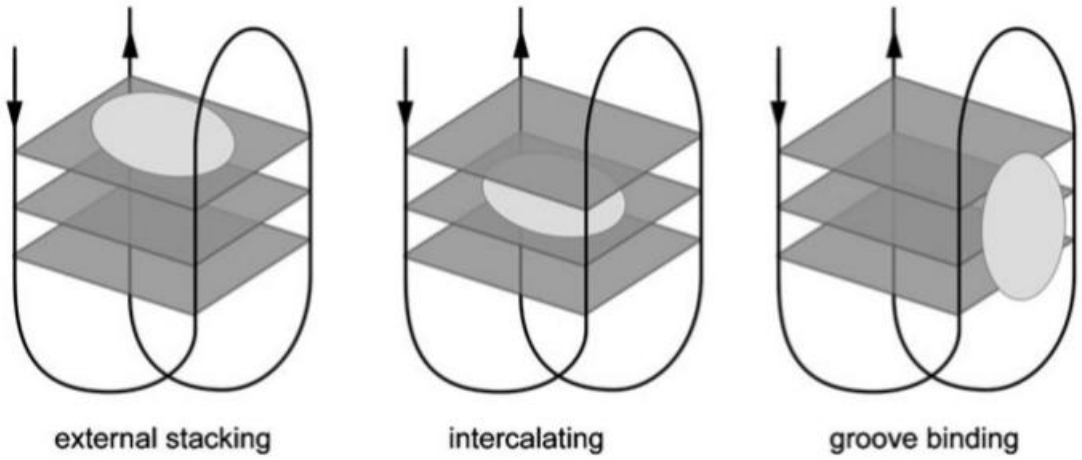

Figure 5 – The different binding modes in a ligand-G4 complex, with (from left to right) external stacking, intercalating and groove binding. [37]

One of the main challenges in targeting G4s with small ligands *in vitro* is that the ligand must bind exclusively to the G4 structure, perhaps only to a certain sequence, while avoiding G4s with different sequences and duplex DNA that may be present in far more abundant quantities. G4s are very geometrically different from B-DNA, and molecules tend to bind with the later *via* intercalation with the base-pairs, groove-binding, and electrostatic interaction with the backbone. [36]

Throughout the past 20 years, different ligands have been studied, with different properties. [37] One of the main parameters used to classify them comes from their elemental components – mostly, they are either organic or metal-based. Within these groups, they are also classified based on other parameters, such as, for organic ligands, ring structure or positive charge introduction into the scaffold, or in the case of metal ligands, the geometry and type of metal present in the ligand. [37]

1.6 - G4 structure and binding characterization

There are many techniques that can be used to study G4s' structure, and its interactions with a ligand, in order to evaluate thermodynamic, conformational, and other aspects of the ensuing G4-ligand complex. [18]

In the context of this dissertation, we focused on biophysical essays, namely circular dichroism and fluorescence spectroscopy. High-resolution magic angle spinning (HRMAS) NMR technique was also used for characterizing the chromatographic support. Finally, affinity chromatography was used to compare the retention times of different G4-forming sequences, and a duplex sequence using the affinity ligand C_8-NH_2. Those techniques will now be approached on a theoretical basis.

1.6.1 - Circular dichroism spectroscopy

Circular dichroism is a commonly used technique for studying many areas of G4s, such as their structure and polymorphisms, cation effect, thermal denaturation, and ligand binding. It is a robust, easy to use, and requires only small amounts of nucleic acid, as low as 25µg. [38],[39]

This technique was developed in the second half of the last century with the goal of assigning an absolute configuration of chiral molecules and has now become a means to study subtle conformation changes and supramolecular interactions. [40] It relies on irradiation of either a chiral or achiral complex in a chiral environment with circularly polarized light. With the passing of the light through the sample, there is a differential absorption of right- and left-handed circularly polarized light, resulting in a CD absorption spectrum or signature. [41]

The G4 structure assumes different topologies and have peaks characteristic to them. For example, G4s with an antiparallel topology show a strong negative peak around 260 nm and a strong positive peak at around 295 nm. G4s with a parallel topology show a weak negative peak around 240 nm and a strong positive peak around 260 nm. Some G4s have mixtures of both topologies, called hybrids, in which case the values just described are in a mixture of the above.

Circular dichroism can also determine the effects of ligand binding on the G4 structure. Depending on the changes of the G4 spectra after ligand binding, data about the stability of the structure can be inferred. If the peaks in the spectra remain unchanged or undergo enhancement, the ligand can be assumed to have stabilized the structure. If the peak positioning changes somehow (peaks partially/completely disappear, new peaks appear), then the ligand either destabilized the structure or changed it into a new topology. [42]

Melting studies are another way of determining the stability of a complex formed by a G4 and a ligand. These consist of monitoring the absorbance of polarized light at the maximum wavelength of the G4 as the temperature is slowly increased, causing the denaturation of the structure, and a subsequent absorbance decrease. A ligand has a greater stabilizing effect on the G4 the more the melting temperature increases. [43]

1.6.2 - Fluorescence spectroscopy

This technique is based on the physical phenomenon of fluorescence. It consists on the emission of a photon upon relaxation from an electronically excited singlet state after the absorption of

a photon by the fluorescent molecule, the fluorophore. This emission process happens much slower than the absorption, meaning that many parameters can alter the fluorescence emission spectrum. Fluorescence, due to its sensitivity to the fluorophore environment, is the most efficient optical spectroscopy to follow ligand binding activity, and shares characteristics with other optical spectroscopies, like being quick, not destructive, not expensive, and requires much lower macromolecule and ligand concentrations compared to other techniques. [44]

Many G4 ligands show intrinsic fluorescence activity and usually, that activity changes upon binding to a G4 structure. As such, titration experiments can allow one to track the ligand's fluorescence activity as a function of G4 concentration. By obtaining the ligand's fluorescence intensity at the maximum of its emission spectrum, and by plotting it against G4 concentration, a binding curve is obtained. Upon fitting that curve to a suitable binding model, the binding stoichiometry and affinity constant can be obtained. [45]

This method has some disadvantages however. For instance, the changes in fluorescence upon G4 binding are not correlated with the ligand binding mode in a simple way. Often, further experiments are necessary in order to better understand the binding mode occurring in the G4-ligand complex. [44]

1.6.3 - High-resolution magic angle spinning (HRMAS) NMR technique

One method for structure elucidation of the support is High-resolution magic angle spinning (HRMAS) Nuclear Magnetic Resonance (NMR) spectroscopy. This technique overcomes weaknesses of Solution-state NMR, such as not working with large proteins, membrane proteins, or protein complexes as these situations would lead to broadening or disappearance of resonances. This technique overcomes this drawback by spinning the sample at a "magic angle", leading to broadening resonances (caused by dipolar coupling and chemical shift anisotropy) disappearing. [46]

There are numerous examples in the literature of this technique being used to determine if molecules are immobilized in a matrix. [47] For this work, this technique was used to evaluate if the ligand C8-NH_2 is immobilized in a previously activated Sepharose medium. [48]

1.6.4 – Chromatography experiments

Affinity chromatography is a method to separate a molecule/group of molecules from complex mixtures, relying upon specific interactions between the target molecules and a chromatography support. The stationary phase is functionalized with specific molecules (the ligand) while the target molecule is part of the complex mixture, in a mobile phase. The target binds to the ligand in a step called the "capture step", while the remaining components of the mixture are washed and eluted. This allows for the recovery of the desired molecule, and thus, this processed was often used either for the purification of samples, to concentrate substances

present at low concentrations, and separating proteins based on biological function. Recently, this process has become a prized method to study biological interactions. [49]

Chang et al., for instance, used affinity chromatography to test the retention times of different G4 sequences with 4 different affinity matrixes, deducing many important characteristics of G4-ligand binding, such as the role of π-π stacking interacting in selective isolation of G4s, and electrostatic interactions between DNA and matrix. [50] Musumeci et al. created a technique for the screening of G-quadruplex ligands that could be reused several times without loss of efficiency and reproducibility, with promising results. [51]

Thus, affinity chromatography is established as a promising method to examine the interactions between the C8-NH_2 support and different G4 sequences versus duplex.

1.6.5 - Docking

Molecular docking has gained much relevancy in the design of new drugs but can also be applied to other purposes. For instance, it can help determine which among a list of ligands would be best suited for a certain macromolecule by comparing binding energies. [52]

Molecular docking is an approach to bioinformatics modelling that simulates the interaction of two or more molecules. It can be done in many ways, but most of them share some elements between them: there is a receptor and a ligand, sampling, and scoring functions. [53]

Overall, the docking process consists of predicting the conformation and orientation of a ligand in a targeted binding site. There are a few parameters involved in this, however, identifying the molecular features responsible for specific biological recognition, and simulating them on a computer is very challenging. Thus, the process is often split into multiple steps. First, docking algorithms pose the ligand in an active site, sampling the degrees of freedom of said ligand to identify the conformation that best fits the receptor structure. The algorithms are assisted by scoring functions, which are design to evaluate the interactions between ligands and targets and predict ensuing biological activity. [54] There are two different approaches by which the ligand/macromolecule binding activity can be explored, which are, the simulation approach and the shape complementarity approach. Afterwards, sampling algorithms and scoring functions will be approached in more depth.

In the simulation approach, seen in Figure 6, the ligand and target are separated by physical distance, and the ligand is then allowed to bind to the target, through moves in its dimensional space (these moves can be internal, through torsional angle rotations, or external, with translations and rotations of the ligand). The binding energy is calculated in the amount of moves the ligand must do to bind to the ligand in a stable conformation. This approach better reflects ligand flexibility but takes more time and

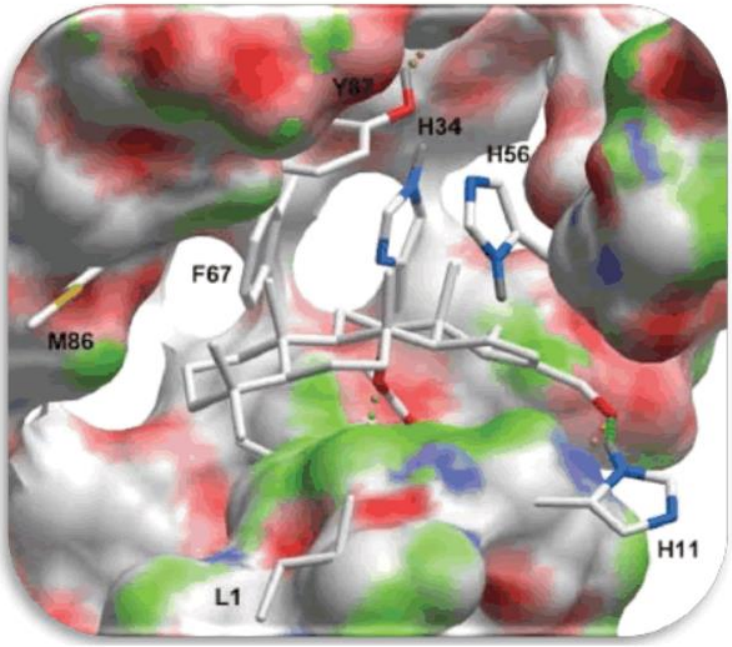

Figure 6 – Two adducts in a simulation approach, with the ligand and target separated by some distance, interacting through H-bonds. [55]

The shape complementarity approach considers the macromolecule and the ligand as a set of features that make them dockable, such as molecular surface descriptors. The macromolecule's accessible surface area and the ligand's matching surface description are used for shape-matching, to find a complementary pose that fits them together. This approach is much faster, simpler and less computationally taxing to use, but it often does not portray any dynamic changes in the macromolecule/ligand conformations, such as their movement, or ligand's flexibility. [55]

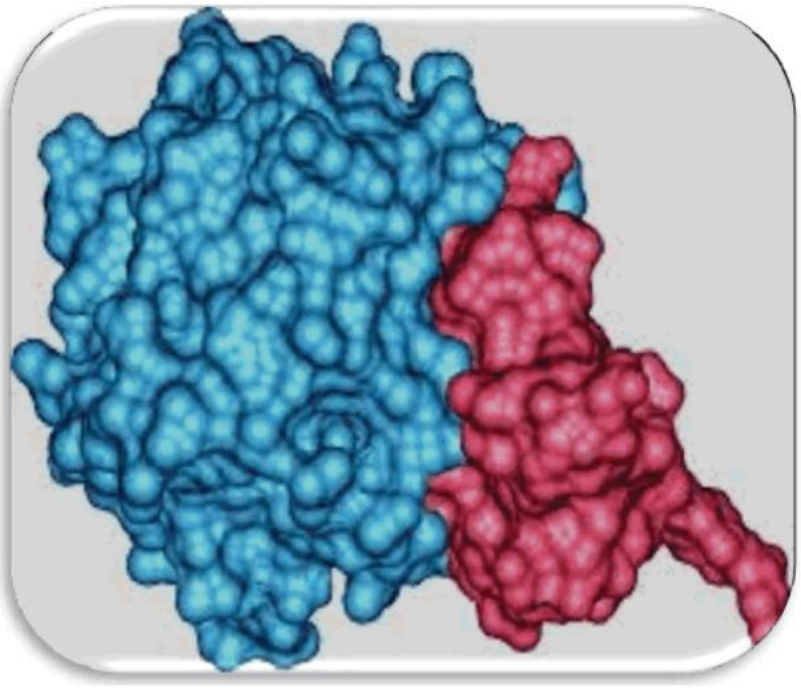

Figure 7 – A shape complementarity approach, with the shape of the ligand and the macromolecule being fitted for interactions according to their geometry. [55]

The amount of possible binding modes between a ligand and a macromolecule, considering the six degrees of translational and rotational freedom, as well as the conformational degrees of freedom of both, is tremendous. As such, it would be too computationally demanding to generate every possible conformation. Sampling algorithm help to narrow down the best ones and can do so in a variety of ways. We will explore Matching Algorithms (MA), Incremental Construction (IC) methods, Multiple Copy Simultaneous Search (MCSS) methods, LUDI, Monte Carlo methods, Genetic Algorithms (GA), and Molecular Dynamics (MD).

MA's are based on molecular shape, matching a ligand into an active site of a macromolecule in terms of shape features, such as distances between the atoms of both molecules, which are considered pharmacophores in this method, and chemical information, such as hydrogen-bond donation and acceptance from both parts. These algorithms are fast, and available in a variety of programs (DOCK, FLOG, LibDock, etc.).

IC methods insert the ligand into an active site in a more fragmented and incremental fashion. The ligand is broken up into several fragments, via breaking its rotatable bonds and inserting those fragments into the docking site first one at a time. The biggest fragment, or the piece with more significant functional role or interaction with the macromolecule is placed first, and the remaining fragments are added incrementally, with different orientations, reflecting the flexibility of the ligand.

MCSS and LUDI also follow a fragment-based approach. MSCC makes many copies (1000 to 5000) of a functional group and places them randomly in the binding site of interest, subjecting them to simultaneous energy minimization and/or quenched molecular dynamics in the forcefield of the macromolecule (none of the copies interact with each other, only the macromolecule). Based on the interaction energies, favorable binding sites and orientations are mapped. The process is repeated with different functional groups, resulting in a different map. Molecules that match the binding site perfectly are thus designed. LUDI focuses instead on hydrogen bonds and hydrophobic contacts formed between the ligand and the macromolecule, and generates interaction sites (positions in space suitable for the formation of hydrogen bonds and filling hydrophobic pockets) by searching databases or using the rules. Fragments of functional groups are then fitted to those sites, and then collected to form a single molecule.

Monte Carlo methods and GA's fit under a broader term of stochastic methods, which search the conformational space by randomly modifying a ligand conformation or a population of ligands. Monte Carlo for instance generates poses for the ligand through bond rotation, and tests the conformation achieved with an energy-based selection criterion. If it passes the test, it is saved and modified further, until a certain pre-determined number of conformations is achieved. GA's, which stem from Darwin's theory of evolution, encode degrees of freedom of the ligand into binary strings called genes, making a "chromosome" which represents the pose of the ligand. "Mutations" (where the genes change randomly) and "crossovers" (in which genes get exchanged between chromosomes) then occur, resulting in new structures, which get assessed by scoring functions. If they survive, they move on to the next generation.

MD is used as a simulation approach, being one of the most powerful of that kind in the field of molecular modeling. It consists of moving every atom of each molecule separately and assessing the resulting structure stability according to scoring functions. It is the method that most reliably represents the flexibility of the ligand and macromolecule in comparison with other methods, but progress in very small steps and are very computationally demanding. [56]

There are many different kinds of scoring functions, with them being divided as force-field based, empirical and knowledge-based. Force-field-based scoring functions assess the binding energy by calculating the sum of non-bonded interactions, such as van der Waals and electrostatics. Empirical scoring functions decompose binding energy into several energy components (hydrogen bonds, ionic interaction, hydrophobic effect and binding entropy). Knowledge-based scoring functions use statistical analysis of already produced crystal structures, based on the principle that the more stable a structure is, the greater the frequency of outcome. [56]

1.7 - Work summary

Based on the aforementioned data, we can divide the following work in three parts: i) performing computational docking and molecular modelling analysis on seven promising ligands identified in the literature, to determine their binding modes and most importantly affinity to a predetermined G4-forming sequence; ii) performing biophysical assays to test the binding affinity of the same G4-forming sequence and ligands; and finally, iii) concluding as to the efficacy of using molecular modelling to draw experimental conclusions as towards the binding affinity of ligands.

By using widely available chemical sketching and rendering software, we can simulate the 3D structures of seven ligands and the pre-miR-149 G4 structure computationally. Through molecular docking and molecular dynamics simulation software, we simulate docking interactions between each ligand and the G4 structure, to determine the most favorable binding site, conformation, and pose the ligand can assume while binding to the G4. Most importantly, the free binding energy, which can be used as a measure of the binding affinity between the ligand and G4 can also be obtained. Ligands would be ranked in accordance with their binding energy values, and thus, their affinity towards the pre-miR-149 G4 structure.

Next, biophysical techniques, such as circular dichroism and fluorescence assays are used to test the ligand-G4 binding affinity in an experimental setting. Affinity chromatography experiments could be used to determine if some of the ligands explored could also be used as ligands towards other G4-forming sequences. At the end, ligands would be ranked according to their binding affinities to the G4 structure, as they were in the docking tests. As these results would be experimental, and thus, far more reliable, comparisons could be drawn between the results obtained via docking and molecular dynamics simulations and those obtained experimentally.

The last part would involve assessing these comparisons, and determining whether molecular modelling techniques could be used in the future as reliable methods to determine ligands' binding affinities towards this or other G4 structures in the future, opening new areas of investigation in this topic.

Chapter II

Objectives

The main objective of this dissertation is summarized as follows: test seven different ligands using both computational and biophysical methods in order to determine which have the best binding affinities to the pre-miR-149 RNA G-quadruplex.

This will be approached in the following steps:

1. Computationally generating the structure of a predetermined G4-forming sequence (pre-miR-149) and seven potent ligands described in literature;
2. Running docking simulations between each ligand and the G4 structure, obtaining the best conformation, binding site and free binding energy for each ligand;
3. Running molecular dynamics simulations with each ligand and the G4 structure, to obtain presumably more realistic binding conformations and energy;
4. Using circular dichroism spectroscopy and melting studies to determine the ΔT_m ,
5. Using fluorescence spectroscopy techniques to determine the binding affinity between each ligand and the G4 structure;
6. Using affinity chromatography technique to determine if the ligand C8-NH_2 retains G4 structures with similar or different conformations;
7. Comparing the binding affinity values obtained from the two different techniques and comparing them to the ones obtained via docking;
8. Establishing which of the seven ligands can be considered to be the best ligand.

Chapter III

Materials and Methods

3.1 - Materials

All oligonucleotides were obtained from Eurofins MWG Operon (Ebersberg, Germany), STAB VIDA Genomins Lab (Lisbon, Portugal) and Eurogen ech (Leven, Belgium). The following oligonucleotide sequence used was a DNA sequence, with the following sequence of nucleotides: GGGAGGGAGGGACGG. Stock solutions of approximately 500 μM were prepared using Milli-Q water and stored at -20°C until used. The synthesis and purification of 10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium iodide (C_8) and 3,6-diamino-10-(8-aminooctyl)acridin-10-ium (C_8-NH_2) was performed as described by Pereira *et al.* [57]. [16]phenN_2 and [32]phen$_2N_4$ were synthesized as described in previous work [58] [59]. Stock solutions of the ligands were prepared as 10 mM in DMSO. The dilutions for fluorescence and affinity chromatography were done using an annealing buffer, consisting of 100 mM KCl, 30 mM phosphate buffer (15 mM potassium hydrogen phosphate + 15 mM dihydrogen phosphate). The dilutions for dichroism were done using an annealing buffer, consisting of 2 mM KCl.

3.2 - Methods

3.2.1 - Synthesis of affinity chromatography supports

The affinity chromatography support synthetized was C_8-NH_2 Sepharose, obtained through the immobilization of C_8-NH_2 ligand to the bifunctional oxiranes.

The following steps to produce both supports are:

i) Sepharose activation

Sepharose quantity between 3.5-5 g was weighted and washed on a glass-filter funnel with 500 mL of milli-Q water. For every gram of Sepharose, 1 mL of diglycidyl ether, and 1 mL of a 0.6M sodium hydroxide solution, containing 2 milligrams of sodium borohydride for every sodium hydroxide millilitre was mixed. This suspension was then kept in continuous rotation overnight (>15h) at 25ºC, and then washed with at least 500 mL of milli-Q water on a glass-filter funnel.

ii) Coupling

The ligand C8-$_{NH2}$ was dissolved in a coupling buffer, consisting of a 2M sodium carbonate solution at pH 10.58. The resulting coupling solution was mixed with the activated Sepharose

medium in a vessel, which was then placed in continuous rotation for 5 days, at 50ºC and 135 rpm.

3.2.2 - Circular dichroism (CD) spectroscopy

CD spectra were obtained in a Jasco J-815 spectrometer (Jaco, Easton, MD, USA), using a Peltier temperature controller (model CDF-426S/15). The induction of G4 formation was assessed by titrating the sample with increasing concentrations of NaCl and KCl, and the quartz rectangular cell used for such purposes had an optical path length of 0.1 cm.

The spectra were acquired at 20ºC using an instrument scanning speed of 10 nm min^{-1} with a response time of 1 s over wavelengths ranging from 220 to 340 nm. The recording bandwidth was 1 nm with a step size of 1 nm. Data was collected in quadruplicate, with the average values of the spectrum presented for each sample, after accounting for the buffer contribution.

CD data was converted into ellipticity, and noise was smoothed by the Jasco J-815 software, including a fast Fourier transform algorithm, allowing the enhancement of the noisiest spectra without distorting the peak shape.

The CD melting experiments were performed at a temperature range of 20-100 ºC, with a heating rate of 2 ºC per min, monitoring the ellipticity at 260/265 nm.

Spectral acquisition was performed in the presence of 16 molar equivalents of a ligand. The resulting data was converted into fraction folded plots (theta) using the following equation:

$$\theta = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}}$$

With CD being the ellipticity of the monitored wavelength at each temperature, while CD^{min} and CD^{max} are the lowest and highest ellipticity, respectively. Data points were fitted to a Boltzmann distribution (OriginPro 2015) and the melting temperatures were determined from the two-state transition model using the first derivative method.

3.2.3 - Fluorescence spectroscopy binding studies

Fluorescence spectra were recorded using a Horiba FloroMax4 fluorometer equipped with a temperature control system. Samples and references were scanned using a quartz cuvette with a path length of 1 cm with a volume of 800 µL. All spectra were scanned with an integration time of 0.5 s, emission and excitation slit width of 10 nm, and a step size of 1 nm.

Of the seven different ligands, only four had fluorescence activity and could undergo this technique: [16]phenN_2, [32]phen$_2N_4$, C_8, C_8-NH_2. The G4 sequence was introduced into a 10mM lithium cocodylate + 2 mM KCl buffer (pH 7.4), then heated to 95ºC for 10 minutes, followed

by a cooling period for 20 minutes in ice, in order to induce the formation of a G4 structure. The interaction of the G4 structure with the ligand was determined by measuring the fluorescence emission values in a titration of the ligand with increasing G4 molar equivalents.

Data was corrected to account for the screening effect (the amount of excitation light absorbed by oligonucleotide at 221 nm).

The ligand [16]phenN_2 was excited at 270 nm, and emission spectra were acquired between 285-600 nm. The ligand [32]phen$_2N_4$ was excited at 272 nm and emission spectra were acquired between 285-600 nm. The ligand C_8 was excited at 498 nm and emission spectra were acquired between 500-700 nm. The ligand C_8-NH_2 was excited at 272 nm and emission spectra were acquired between 530-700 nm.

Fluorescence intensity (F_0/F) was plotted against DNA concentration and fitted to the Stern-Volmer equation to determine binding constants using OriginPro 2017.

Two equations were used depending on the quenching mechanism and resulting curves.

$$\frac{F_0}{F} = 1 + K_{SV}[Q]$$

$$\frac{F_0}{F} = Ae^{K_{SV}[Q]}$$

, where F and F_0 are the fluorescent intensities of the ligand in the presence and absence of the quencher, respectively, A is the amplitude, K_{SV} is the binding constant and Q is the DNA concentration.

3.2.4 – HRMAS NMR spectroscopy

All NMR experiments were performed at room temperature using a Bruker Avance III 400 operating at 400.15 MHz for protons, equipped with a 4-mm triple resonance (HNC) HR MAS probehead. Approximately 10 mg of the support sample was placed in a 4-mm MAS zirconia rotor (50 µL). Samples were spun at the magic angle at a rate of 4.0 kHz, and all spectra were acquired under field-frequency locked conditions using that probe channel with the spectrometer's lock hardware. Spectra were processed using Bruker Topspin 4.1. Unless otherwise stated, all 1H NMR spectra were referenced internally to the residual 1H signal of DMF-d7, which also serves as the swelling agent for the polymer beads (~0.05 mL). Carr-Purcell-Meiboom-Gill (CPMG) sequence with an echo time of 1.5 ms was used to suppress the broad signals of the polymer, experiments were acquired in 256 transients.

3.2.5 – Affinity chromatography

The chromatographic retention times of G4 and duplex sequences were determined through affinity chromatography experiments, performed on a standard 10 mm diameter × 200 mm length column, packed with 2.5 cm (approximately 2 mL) of epoxy-activated Sepharose CL-6B matrix and AO derivative C_8-NH_2 support. The experiments were performed at room temperature in a column that was first equilibrated with in a buffer consisting of 10 mM Tris HCl buffer (pH 8.0) at 1 mL/min. A 200 µL volume (10 µg/mL) of the various oligonucleotides samples was loaded onto the column and it was applied a 10 min linear gradient from 0.075 M to 1.5 M KCl or NaCl. The absorbance of the eluate was continuously monitored at 260 nm.

3.2.6 - Molecular dynamics

3.2.6.1 - Model construction

A model of pre-miR-149 was constructed by modifying an existing human telomeric G4 structure (PDB ID: 4G0F), with the sequence AGGGTTAGGGTTAGGGTTAGGG. Using Swiss PDB Viewer, the starting sequence was edited to assume the pre-miR-149 sequence, without destroying the underlying structure. Thus, the thymine residues in positions 5, 6, 11, 12 and 17 were deleted, along with the adenine residue in position 1. The thymine residue in position 18 was mutated to an adenine, and the adenine residue in position 19 was changed to cytosine. In addition, since the starting sequence represents a DNA G4, whereas the final one is an RNA G4, the resulting pdb files were manually edited, adding an O2' group to every C2' group of every remaining nitrogen base. As many residues were removed, the structure was left unstable, and further molecular dynamics simulations were undertaken to stabilise it.

3.2.6.2- Model stabilization

The following molecular dynamics simulations were performed with Gromacs 2016.3.

After being submitted to the Amber94 forcefield (in order to generate a topology of the structure), the system was placed in a box containing TIP3P water molecules with an 8Å cut-off value. The total charge of the system was then neutralized by adding 13 K^+ ions. A 1000-step energy minimization was performed, with the RNA fixed by 500 kcal mol^{-1} A^{-2}, followed by a 5000-step minimization with no restraints. Equilibration was then performed. Firstly, a 100 ps equilibration was performed, under a modified Berendsen thermostat, in which the system was heated from 0 to 300 K, with 10 kcal mol^{-1} A^{-2} restraints on the solute, as well as position restraints. Another 100 ps of equilibration were done under both the aforementioned thermostat and a Parrinello-Raman barometer, to make sure that the system was stable under certain pressure conditions.

Finally, 20 ns of MD simulations were carried out, with no restraints on the molecule, in order to allow the structure to assume a form more naturally occurring, and the final structure

following that time period was then obtained and used as a starting point for any further experiments.

3.2.6.3 - Molecular docking

Molecular docking simulations were performed with AutoDock 4.0 and Autogrid 4.0, in conjunction with the graphical user interface AutoDockTools (1.5.6). The best conformation was then isolated, and molecular dynamics was performed with Gromacs, followed by cluster analysis, RMSD calculation and other tests.

A total of 7 ligands were tested: L-arginine, C8 and C8-NH2, [16]phenN_2, [32]phen$_2N_4$, Phen-DC3 and pyridostatin.

The docking process performed with AutoDockTools consisted of two phases: setting the parameters for the process (which includes preparing the ligand and macromolecule for the process), followed by performing the procedure, and extracting the results.

3.2.6.3.1 - Parameter preparation

The G4 structure was pre-processed UCSF Chimera 1.11.2 prior to undergoing the docking procedure. Using the software's "Dock Prep" tool, hydrogens were added to the structure in a histidine protonation state, Gasteiger charges were added and certain nomenclature changes were enacted (for compatibility between the two programs). In AutoDockTools, to ensure all proper criteria were met, hydrogens were again added, atoms were assigned the AD4 type, Gasteiger charges were computed and non-polar hydrogens were merged. The structure was then saved as a pdbqt file.

The ligand's molecular structure was created with MarvinSketch, then cleaned in both 2D and 3D in the software. After being examined in Chimera for any structural defects, it was saved as a pdbqt file in AutoDockTools.

The ligand and macromolecule were both rendered in AutoDockTools, and a grid box was created, centered on the G4, being large enough to contain it, but not the ligand. Since the parameters for K^+ ions are unknown by the program, they were added manually, and a grid parameter file (.gpf) was created with all this information.

Then, a docking parameter file (.dpf) was prepared, with the pre-miR-149 G4 structure and ligand defined. A genetic algorithm, more specifically, a Lamarckian GA (4.2) was chosen, with a total of 25 runs being performed. Each run had a high number of evals, corresponding to the "long" software option (25 000 000 evals).

3.2.6.3.2 - Docking performance and results extraction/analysis

AutoGrid4 was then run with the .gpf file given as input, generating a grid log (.glg) file. Afterwards, AutoDock4 was run with the .dpf file given as input, generating a docking log (.dlg) file.

Afterwards, using the "Analyse" tool, the .dlg file was analysed, with all 25 obtained conformations being rendered in increasing order of binding energy (decreasing order of structural stability). The binding energy of each conformation was the main parameter used in order to determine the structural stability of the macromolecule-ligand complex.

The same procedure was repeated for all 7 ligands, and at this stage, the binding energies were analysed in order to determine which was the superior ligand.

3.2.6.3.3 - Molecular dynamics runs

The best conformation of each ligand was then chosen to undergo molecular dynamics runs in Gromacs 2016.3.

Due to limitations inherent in the software, the complex could not be submitted to the forcefield in itself. As such, the complex was split into two separate files, one for the ligand and other for the G4. The G4 was subjected to the Amber94 forcefield, and the ligand was subjected separately to an online available tool, acpype, which applied the AMBER forcefield to the ligand and generated topology, position restraint and molecular structure files. The molecular structure and the topology files of both molecules were merged into two files representing the molecular structure and topology of the complex. A solution environment was simulated by inserting the complex into an octahedron with 5.2 nm sides and filling the octahedron with TIP3P water molecules. 13 K^+ ions were added to the box to neutralize the G4 negative charge. The complex was then minimized until the total energy was below 250 kJ/mol/nm.

The system then underwent thermal stabilization at 300K for 49 ns, with a modified Berendsen thermometer, with the G4/ligand complex and the water + ion molecules being the two different coupling groups considered. The G4 and ligand had position restraints throughout the process. After that, the system underwent pressure stabilization for 1 ns, at 1 bar and 300 K, with a Parrinello-Rahman barometer, G4 and ligand restrained again.

After the equilibration steps, the system was put through 10 ns of unrestrained molecular dynamics (mdrun), and the final positions of the ligand and G4 were obtained and further analysed.

3.2.6.3.4 - Mdrun results analysis

Firstly, the resulting trajectory file from the mdrun was unsuitable for analysis, due to periodic boundary conditions (pbc) causing the molecule to seemingly “jump” out of the box it was placed in. As such, the trajectory had to be converted. Using the “trjconv” command in Gromacs, the trajectory file was changed to display correct pbc with the -pbc option, and the -nojump option was used to center the trajectory on the molecule. Every further analysis was executed on this modified trajectory file.

The types of analysis performed on these files were the following: root mean square deviation (RMSD) analysis, cluster and thermodynamic analysis.

The RMSD analysis was performed with the “rms” module in Gromacs and was used to determine the RMSD throughout the entire mdrun. If a stable RMSD value was maintained throughout, the structure was deemed stable. Otherwise, the run was repeated with more nanoseconds added until a stable structure was achieved.

The cluster analysis was performed with the “cluster” module in Gromacs, and it was used to determine the average structures assumed by the G4/ligand complex throughout the run.

The thermodynamic analysis was performed with an imported Gromacs module called mmpbsa. [60] It helped determine the binding energy of the complex, and by comparing the different binding energy values obtained for each ligand. This could be drawn towards their affinity to the G4 structure, to supplement the results obtained via AutoDockTools.

Chapter IV

Results and Discussion

The results obtained from the experiments described in the previous chapter will now be shown.

Firstly, the docking experiment results will be shown, beginning with the binding energy values obtained via AutoDockTools. Selected ligand-quadruplex docking conformations will be discussed individually in order of increasing binding energy (lower affinity to the quadruplex).

Afterwards, the RMSD graphs of the 10 ns (or more in some cases) molecular dynamics simulations will be analyzed in the same order as the docking experiments, also stating how many clusters were detected for each run, and how many were considered significant. Then, representative cluster structures will be analyzed, comparing them to the conformations obtained by AutoDockTools. Finally, the binding energies calculated through the MM/BBSA method will be displayed in Table 2, and compared to the ones obtained in the docking experiments.

The biophysical techniques' results will then be shown, starting with circular dichroism spectral and melting temperatures data. The melting temperature variations (ΔT_m) will be shown in Table 3, in descending order from the highest value observed, and the spectral data and melting temperature graph will be displayed in the same order for each ligand tested. Then, the fluorescence experiments' results will be shown, with the obtained dissociation constants being displayed in Table 4, in ascending order of value (descending order of binding affinity) and the corresponding obtained graphs will be discussed. A comparison will be drawn between the results obtained through the all biophysical techniques and those obtained by docking and molecular dynamics simulations.

One of the ligands (AO derivative C_8-NH_2) is further characterized with NMR spectroscopy (result shown in Figure 26) and then used as a ligand for an affinity chromatography support, to better assess the interactions of this ligand and other quadruplex-forming sequences, with different quadruplex conformations to the sequence in question, so that conclusions can be drawn about the ligand's ability to bind to different G4 structures.

4.1 - Docking experiments

The pdb file selected for this purpose was 4G0F, a file containing a crystal structure of the complex of a human telomeric repeat G4 and N-methyl mesoporphyrin IX (P6). It is available online, and the G4 sequence has the following nucleotide sequence: AGGGTTAGGGTTAGGGTTAGGG. The sequences were edited to form the target sequence, and

to better simulate an RNA structure (as the structure in question was of a DNA sequence) O2' groups were added to every nucleotide. The structure was then minimized, and the obtained G4 structure was used in all further computational tests.

Docking experiments were performed on seven ligands: pyridostatin, acridine orange (AO) derivatives C_8 and C_8-NH_2, [16]phenN_2, [32]phen$_2N_4$, Phen-DC3 and L-arginine, in order to assess their binding modes with the pre-miR-149 G4 structure.

First, the G4 structure for the pre-miR-149 sequence had to be generated. As there is no protein database (pdb) file available for this structure as of the time of writing, it had no be simulated through different means. As was done in other cases in the literature, a pdb file of another sequence was edited to simulate the desired structure. [61]

The binding energies obtained from AutoDockTools simulations are presented in Table 1. It is worth noting that these simulations were performed in vacuum and are not meant to reflect real Gibbs energy values. In addition, the results obtained for [16]phenN_2 and [32]phen$_2N_4$ were mostly ignored for reasons explained further ahead.

Table 1 – Binding energy, in kJ/mol, of the most favorable ligand-quadruplex complexes obtained in AutoDockTools.

Ligand	Binding energy (kJ/mol)
Pyridostatin	-9.53
Phen-DC3	-8.21
C_8	-5.99
C_8-NH_2	-4.88
L-arginine	-4.59
[32]phen$_2N_4$	-11.07
[16]phenN_2	-7.81

The corresponding docking conformations are depicted in Figure 8.

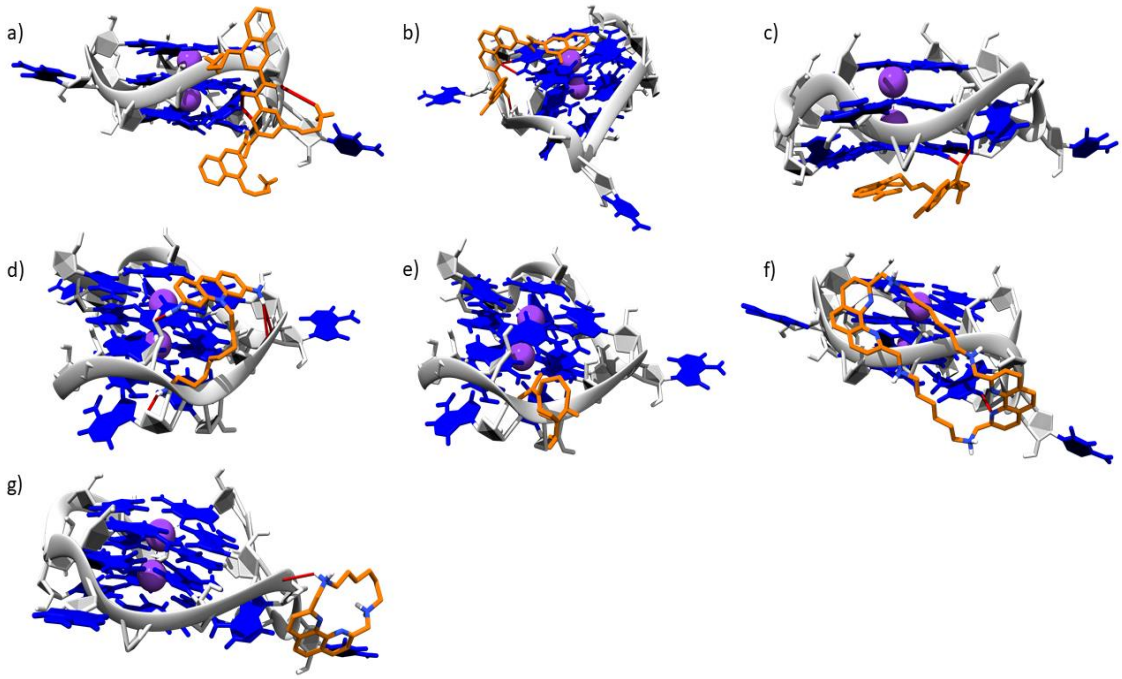

Figure 8 – Best docking conformations obtained between the pre-miR-149 G4 structure and a) Pyridostatin; b) Phen-DC3; c) C8; d) C8-NH_2; e) L-arginine; f) [32]$phen_2N_4$; g) [16]$phenN_2$. Ligand shown in orange for each case, hydrogen bonds shown in red.

4.1.1 – Docking conformation analysis

Of the seven ligands, pyridostatin showed the lowest binding energy, with a value of -9.53 KJ/mol, preliminarily suggesting it to be the ligand that binds more tightly to the G4 structure. The most favorable of all 25 tested binding conformations is shown in Figure 8a and is associated with the G4 through groove binding interactions at the loop formed by the nucleotides G11, A12 and C13. In this conformation, there are hydrogen bonds between the ligand pyridine moiety (a hydrogen group, specifically) and the cytosine C13 O1P group in the backbone, and between both the pyridine O2 and N2 in the aliphatic arm and the adenine A H3 group in the backbone.

Phen-DC3 had the second-best binding energy out of the seven, with a value of -8.21 KJ/mol in its best conformation. It has a mixed mode binding with the G4, exhibiting both stacking and groove binding. One of its quinoline moieties has a stacking association with the top tetrad, whereas the rest of the ligand's structure is in a groove binding association with the guanines G9, G10 and G11 (in the second loop of the quadruplex). There are also two hydrogen bonds between the N atom in the phenanthroline moiety and the H3 and H22 in the guanine G7. Another hydrogen bond appears between the amine group of the naphthalene moiety and the O2' of the guanine G9.

The third best ligand is AO derivative C8, with a binding energy of -5.99 KJ/mol, and has a stack binding association, between the acridine moiety and the bottom tetrad. Two hydrogen bonds

can be spotted between the O1 atom of the C8 aliphatic arm, and both the guanine G9 H22 atom and the adenine A12 H62 atom.

The other AO derivative C_8-NH_2 presented a binding energy of -4.88 KJ/mol. It shows two different binding modes: stacking between the acridine moiety and the top tetrad, and groove binding between the aliphatic arm and adenine A4. Multiple hydrogen bonds are present. There is one hydrogen bond between the NH_2 group of the ligand's aliphatic arm and the O4' atom in A4 of the G4, two bonds connecting the H11 atom of the N3 group of the ligand's acridine moiety connecting to two distinct atoms in the G7 nucleotide in the G4 structure (one connecting to the O2P' group, and another to the O5' group), and a bond connecting the ligand's H9 atom of the N2 group of the acridine moiety to the O2' group of G3.

L-arginine showed a binding energy of -4.59 KJ/mol in its' best conformation. Of the five ligands whose binding energy values are considered valid, it has the highest value. Therefore, it is considered the poorest ligand out of the five. The best conformation is in groove binding to the first loop formed by bases xyz. No hydrogen bonds are observed. Despite repeated attempts, the rendering software would not show any hydrogen bonds, despite the ligand being situated in a place with apparent possibility to form these, namely between the ligand's carboxyl group O atoms and both the guanine G2 H22 atom and the guanine G3 H3 atom. Further results show them in more detail, and these results were considered preliminary for reasons which shall be explained ahead, so the absence of hydrogen bonds can be disregarded.

[32]$phen_2N_4$ and [16]$phenN_2$ both achieved results which were considered to be unreliable. The reason for this is that they both exhibit macrocyclic structures, which is a type of structure very hard to simulate in AutoDockTools. The program cannot compute both bond lengths and torsion angles very accurately – it considered the ligand structures provided to have no torsion angles, which is invalid for macrocyclic structures with considerable rotatable bonds. When the torsion angles were introduced manually, it produced unnatural bond lengths which, after several algorithm runs, made the connections disrupt. Indeed, a review of the literature indicates that this is one of the biggest limitations of AutoDockTools: the ligands are treated as rigid bodies throughout, and not accounting for the flexibility of cyclic or macrocyclic ligands. [62]

Although there are tentative protocols that aim to address and circumvent this issue [62], none of them worked in this case without producing unnatural long bond lengths in some atomic bonds. As such, the results shown clearly suffer from the problem at hand. Both the structures assume the same conformation as they had when introduced into the program, to the point of being untrustworthy, and every conformation other than the one shown displays the ligand in virtually the same position, with practically the same binding energy. Thus, due to the unreliability of these results, the binding energy values are largely to be ignored, and the conclusions regarding binding modes are to be considered as "suggestions" as to how the

structures might bind to the G4 in vacuum conditions. The upcoming molecular dynamics simulations are more reliable.

[32]phen$_2$N$_4$ showed a binding energy of -11.07 KJ/mol in its best conformation. This would make it the best ligand of the seven, were it not for the previously mentioned facts, and the fact that every other conformation has similar values, compounding the unreliable nature of this value. Its main binding mode appears to be groove binding to the loop formed by nucleotides G10, G11, A12 and C13 of the G4 structure. A hydrogen bond is spotted between the right-most N atom in the phenanthroline moiety closest to the third loop, and the G4 A12 N3 group's H3 atom.

[16]phenN$_2$ showed a binding energy of -7.81 KJ/mol in its best conformation. Again, this score is likely overestimated and should not be considered as grounds to consider it superior over ligands with higher energy based on this value alone. Its main binding mode consists of groove binding to the G4 loop formed by nucleotides G11, A12 and C13. A hydrogen bond is spotted between the ligand's HN4 atom on the left-most side of the phenanthroline moiety and the O2P group from the G4 A12.

4.1.2 - RMSD graph analysis

As was specified in the previous chapter, the conformations shown above underwent 10 nanoseconds (or more, depending on whether a stable structure was obtained) of molecular dynamics simulations in order to determine the ligand conformations and positions that would happen in conditions simulating the liquid-stated of the structures, namely in terms of solvent, ionic strength, pressure and temperature.

Throughout these simulations, the ligand conformation and positions changed significantly throughout the runs. In these cases, the root mean square deviation (RMSD), a measure of the average distance of superimposed macromolecules is used. This measure shows the differences that molecules assume compared to their original conformation. [63] RMSD values were calculated for the ligand, the G4 structure, and the complex formed by both, to show how they changed from their original positions and conformations since the start of the simulation.

The RMSD graphs of every ligand will be shown, as well as an ensemble of structures that occurred during the run. Every graph will be discussed in the order in which it appears in Figure 9. For each ligand, the amount of clusters observed will be revealed, as well as how many of those clusters were considered significant (when the combined total of its structures was over 5% of the amount of all structures observed throughout the run).

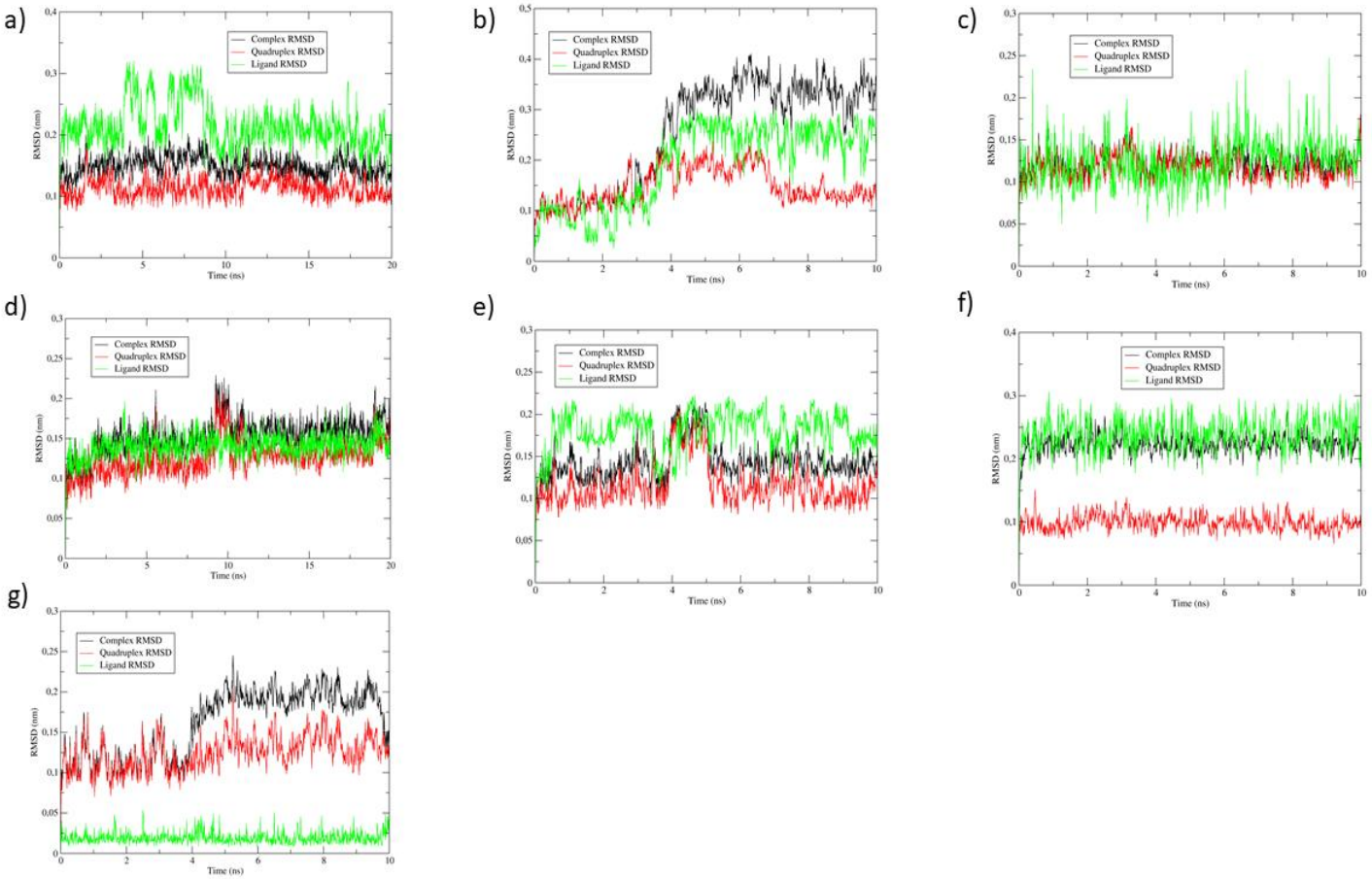

Figure 9 – RMSD graphs of the ligand a) Pyridostatin; b) Phen-DC3; c) C_8; d) C_8-NH_2; e) L-arginine; f) [32]$phen_2N_4$; g) [16]$phenN_2$, quadruplex and complex throughout the simulation. Ligand-quadruplex complex RMSD shown in black, G4 in red, ligand in green.

Pyridostatin was peculiar, as both the quadruplex and the complex's RMSD values changed relatively little throughout the simulation, but the ligand's varied significantly at one point. Around 4 ns, the ligand RMSD values varied from 0.2 to 0.3 nm and fluctuated between those two values until 10 ns, at which point they stayed at the same values until the end of the simulation. Despite this fluctuation, cluster analysis only showed one significant cluster of x structures being formed, with 37 smaller ones being identified, mostly around the time of the aforementioned fluctuations. The fluctuation may have been due to movements of the ligand's two sidechains interacting with the quadruplex, or solvent interactions, more likely the latter, due to the fact that the complex RMSD values do not change significantly along with the ligand's.

Phen-DC3 showed much higher variations in the RMSD values compared to the other ligands. The quadruplex, barring an increase at 3 ns followed by a decrease at 7 ns, does not change much from the baseline in significant ways. The ligand and complex however show a substantial increase at around 4 ns, followed by substantial variations throughout the rest of the run, despite a major conformation is indicated until the end of the simulation. A total of 131 clusters were identified, but of those, only 3 were considered significant. One main cluster is observed

from the beginning of the simulation until the aforementioned rapid increase at 4 ns. Then, the complex assumes another major cluster until about 7 ns, followed by another cluster that lasts until the end of the simulation. These clusters will be analyzed in the following section to determine why the complex's structure varied between them instead of remaining as either of one of them for an extended period of time.

In regards to C_8, all three entities showed RMSD values in the same range, and did not show any drastic variations in relation to the baseline throughout the simulation. Thus, although 28 clusters were identified, only one was considered significant enough. However, all three of the structures RMSD values, but especially the ligand's, show high variance throughout the run. The ligand, in fact, shows RMSD values between 0.05 – 0.25 nm. This may be due to the high degree of rotatability of the ligand aliphatic arm, which provided conformational variability throughout the run.. Additionally, it is very likely that the aliphatic arm underwent many different interactions with the groove it interacted with.

C_8-NH_2 RMSD data was different from C_8, despite the fact that the two differ only in the ending group to the aliphatic arm. The ligand RMSD values remain stable throughout the run of the simulation, but the G4's and the complex's show 3 deviations from an otherwise stable baseline. While the complex's RMSD values are overall higher than the quadruplex's, both exhibit the deviations in the same time stamps. There is an increase around 4 ns, a rapid increase around 9 ns followed by an equally rapid decrease around 10 ns, and a final, more subdued increase around 19 ns. Despite these changes, the cluster analysis revealed that only one major cluster was significant with x structures, with 17 other minor clusters being ignored.

Comparing these results to the ones obtained for the previous ligand, which has the same overall structure, differing only in the functional group at the end of the aliphatic arm, there is a slight difference, but the explanation is likely found in the differing nature of these functional groups. The functional group at the end of the aliphatic arm consists of an aromatic iodobenzene ring, whereas C_8-NH_2 has an NH2 terminal group.

L-arginine shows significant ligand RMSD value variation throughout the simulation, but both the complex and the quadruplex RMSD values follow the same patterns and are reasonably stable for the duration of the run. Both show a sudden spike around 4 ns but then return to the previous observed values at around 5 ns. Only 19 clusters were found, with only 2 of those being significant. In light of that, the ligand's variability can likely be explained due to the structure of the ligand. The lack of any π-π interactions with the structure (due to the ligand having no aromatic ring) makes it so that this ligand is solely a loop/groove binder, leaving it much more vulnerable to solvent interactions and conformation changes than the previously observed ligands. Additionally, the ligand size being relatively small may contribute to the RSMD variation throughout the run as the ligand is virtually capable of fitting multiple binding sites in the G4 structure.

In regards to [32]phen$_2$N$_4$, both the ligand, the G4 structure and the complex formed by both stayed relatively stable throughout the simulation, without undergoing major shifts. It showed 13 clusters, but only one was considered significant enough to be considered. Of note is the fact that the ligand overall showed equal or higher RMSD values than the complex, which is a sign that the initial conformation it was in was not very stable and quickly changed to a more stable form upon the beginning of the simulation.

Finally, in regards to [16]phenN$_2$, the ligand and G4 structures in themselves showed good stability, but the complex's data shows that around 4 ns, the structure underwent a shift. The complex then seemingly adopted this new structure throughout a great part of the simulation, only returning to the previous one around the end. This allows us to establish two main significant clusters for this ligand, out of 23 clusters in total.

4.1.3 - Cluster analysis

Representative structures of the ligand-quadruplex complex are shown in the figures below.

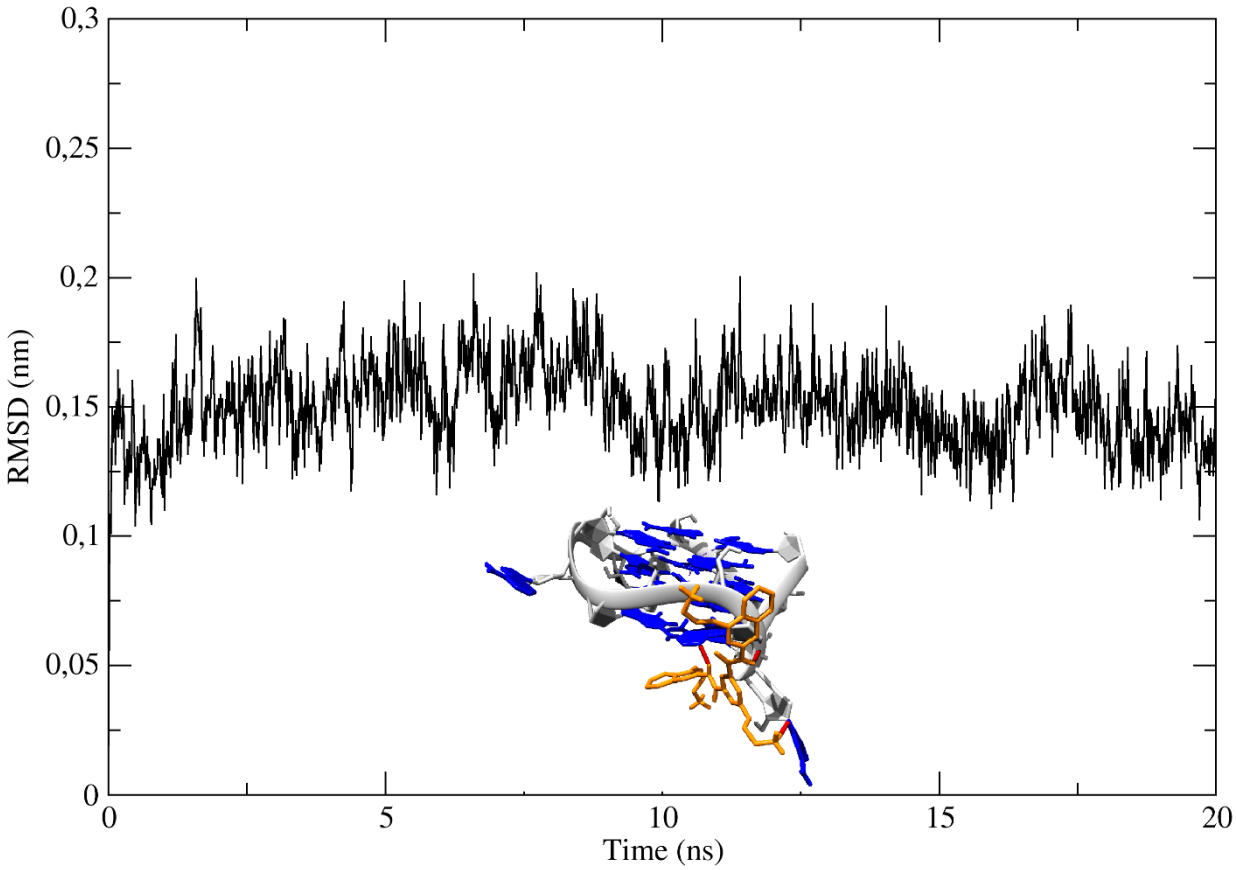

Figure 10 – Representative structure of the main cluster of structures of the ligand pyridostatin MD run.

In regards to pyridostatin, there are many changes between the docking binding pose and the MD representative structure shown in Figure 10. Although the ligand binds to the same region of the quadruplex structure, the biggest change is in the ligand's conformation. In the docking

simulations, most of the aromatic ring moieties were coplanar with each other, whereas in the cluster ones, none of them are.

In the docking conformations, there are two hydrogen bonds between the H3 atom in the quadruplex's A12 N3 group and the ligand (one between the ligand's central aromatic ring's N atom, and another between the oxygen atom to the left of said aromatic ring). Another hydrogen bond is observed between the central aromatic ring's aliphatic arm and the quadruplex's C13 O1P group. In the MD runs, these interactions change. The aforementioned aliphatic arm's HN9 atom is connected to the C13 O2 group, the HN1 atom to the left of the central aromatic ring is connected to the quadruplex's C13 O1P group, and the HN group to the right of the central aromatic ring is connected to the quadruplex's A12 N3 atom.

Overall, compared to the complex obtained via docking simulations, the ligand conformation obtained in the MD runs seems to have wider angles and torsions, favoring a better binding to the quadruplex's C13 nucleotide. In the MD runs, there also appear to be π-π stacking interactions between the ligand's rightmost aromatic ring moieties and the quadruplex's lower tetrad. Thus, this latter conformation is likely more stable, as additional interactions are observed.

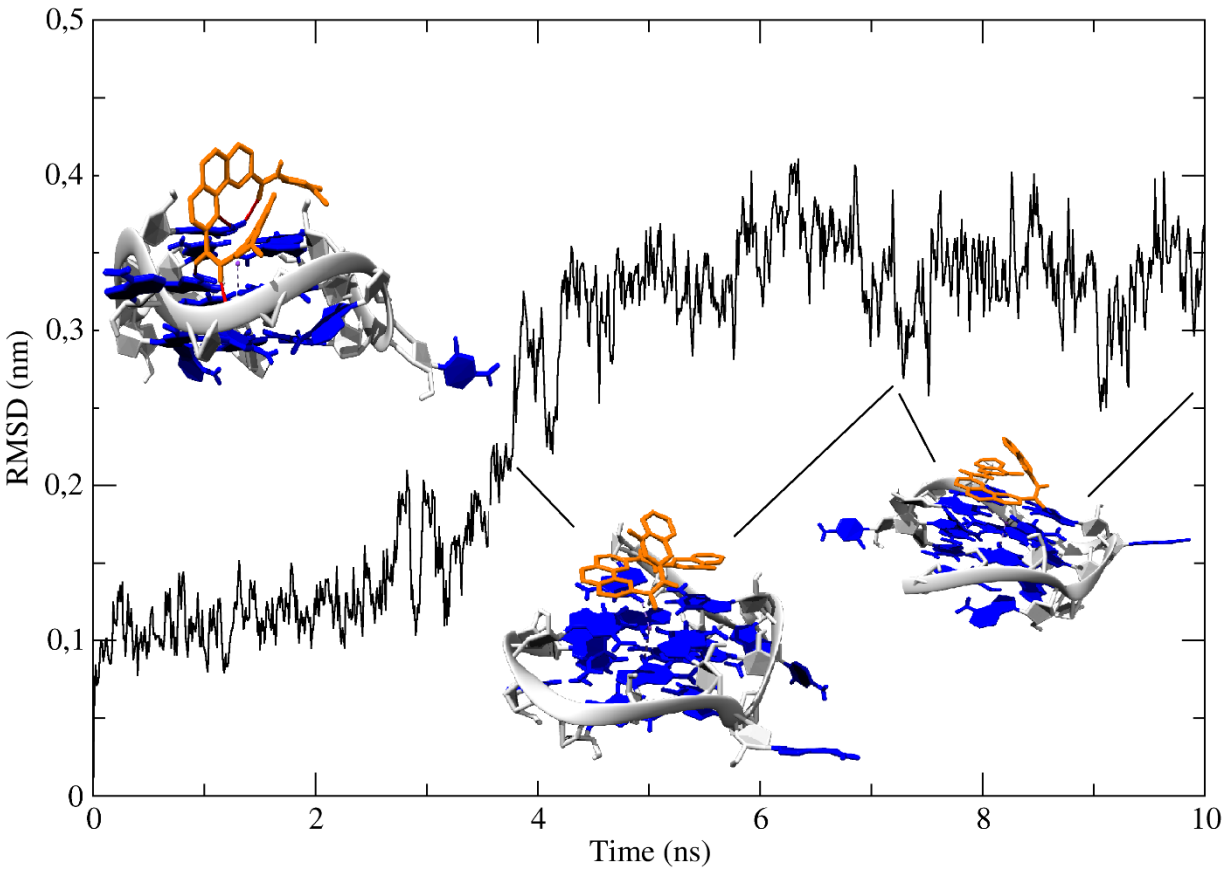

Figure 11 – Representative clusters of the ligand Phen-DC3.

In regards to Phen DC3, the three major clusters it assumes vary between each other greatly in terms of structure and even positioning. The first cluster is similar to the one obtained via docking simulations, which is expected, considering it is the one assumed at the beginning of the simulation. The ligand establishes hydrogen bonds with the H atoms in the quadruplex's G6, G7, and G9 groups, and appears to engage in both groove binding with the second loop of the quadruplex and stacking with the top tetrad.

The second cluster is vastly different. The ligand is now roughly equidistant between the loops of the quadruplex and appears to exclusively stack in the quadruplex's upper tetrad. No hydrogen bonds are observed, meaning π-π stacking and other non-covalent interactions are likely the only interactions it has with the quadruplex. The overall conformation of the ligand seems similar to the first cluster, with one of the side quinoline moieties appearing slanted upwards, facing the other side phenantroline moiety in a roughly 45° angle,

The third cluster is very similar. The only major observable difference is that the ligand's side phenantroline moieties now appear to be closer to the third loop of the quadruplex, which appears to have adjusted its shape as well to better fit the ligand.

Of all simulations, this has been the only one where the ligand significantly changed its location to better interact with the quadruplex. It is likely that this is due to the fact that AutoDockTools accounted for no solvent interactions. The conformation generated in those conditions would indeed allow for far closer interactions between such a large ligand and the quadruplex structure.

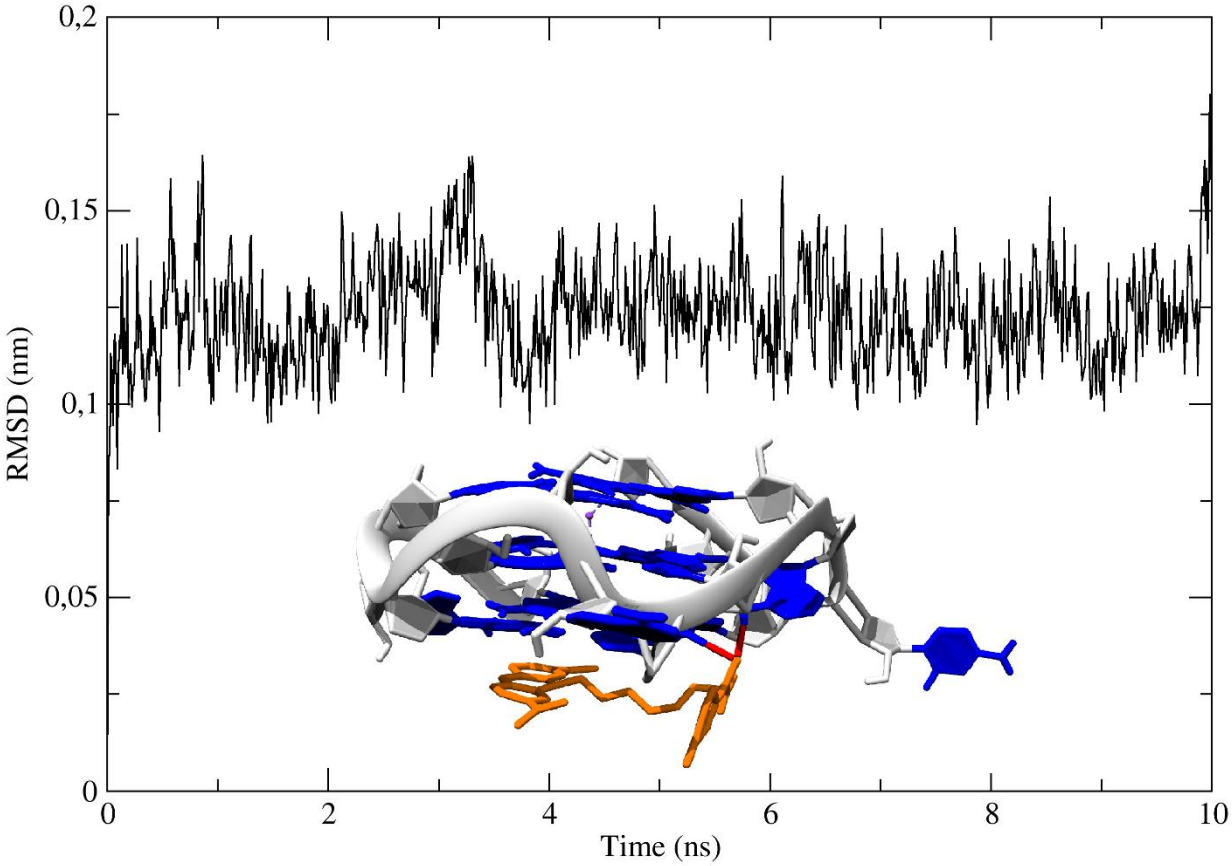

Figure 12 – Representative clusters of the ligand C8.

In regards to C_8, the cluster obtained is nearly identical to the conformation obtained in the docking simulations. The same hydrogen bonds are maintained, and the ligand binds to the quadruplex in the same place. The only change that can be apprehended is that the terminal aromatic ring moiety is nearly perpendicular to the acridine moiety in the cluster, whereas in the docking conformation, it is slanted at about 45 degrees. The variation in RMSD values, as postulated in the previous section, is due to the aliphatic arm slightly changing its position, although none of the clusters that show this are large enough to be considered and discussed in themselves.

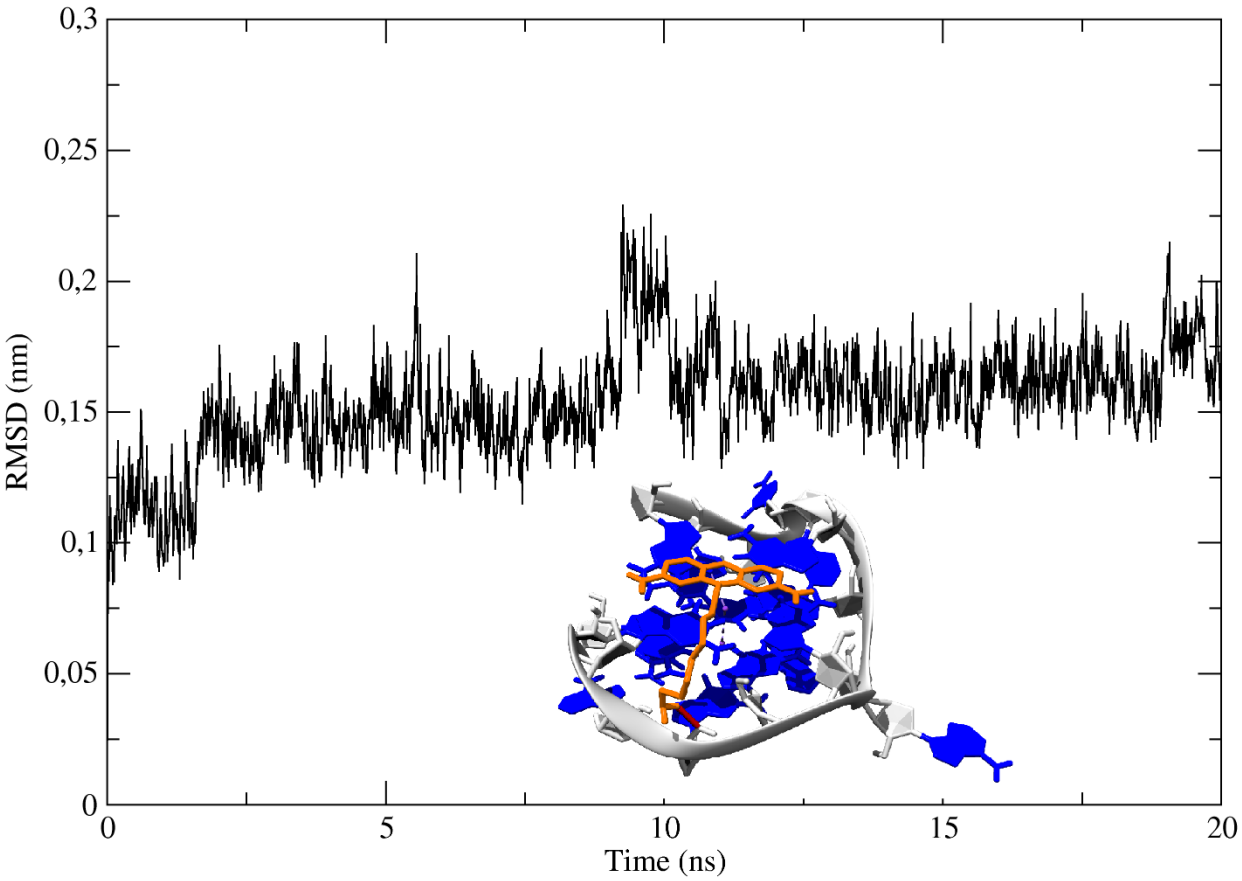

Figure 13 – Representative clusters of the ligand C8-NH2.

C_8-NH_2 has results very consistent with those obtained in the docking simulations. The ligand binds in the same loop, the acridine moiety shows π-π stacking with the top G4 tetrad, and the aliphatic arm shows groove binding with the quadruplex's G5. There are no hydrogen bonds other than one between the terminal NH2 group and the quadruplex G5's O2' atom, despite the fact that the docking simulations showed other ones between the NH groups connected to the acridine moiety and G3 and G7 bases. The cluster contains structures generated throughout the whole simulation.

Closer investigation shows that the hypothesis postulated as to the sudden fluctuations in RMSD values were likely correct, as the short (and mostly insignificant) clusters around those areas show the aliphatic arm tentatively interacting with other parts of the quadruplex backbone. Despite that, those clusters are not significant, as they have a very small number of structures, and thus are not shown.

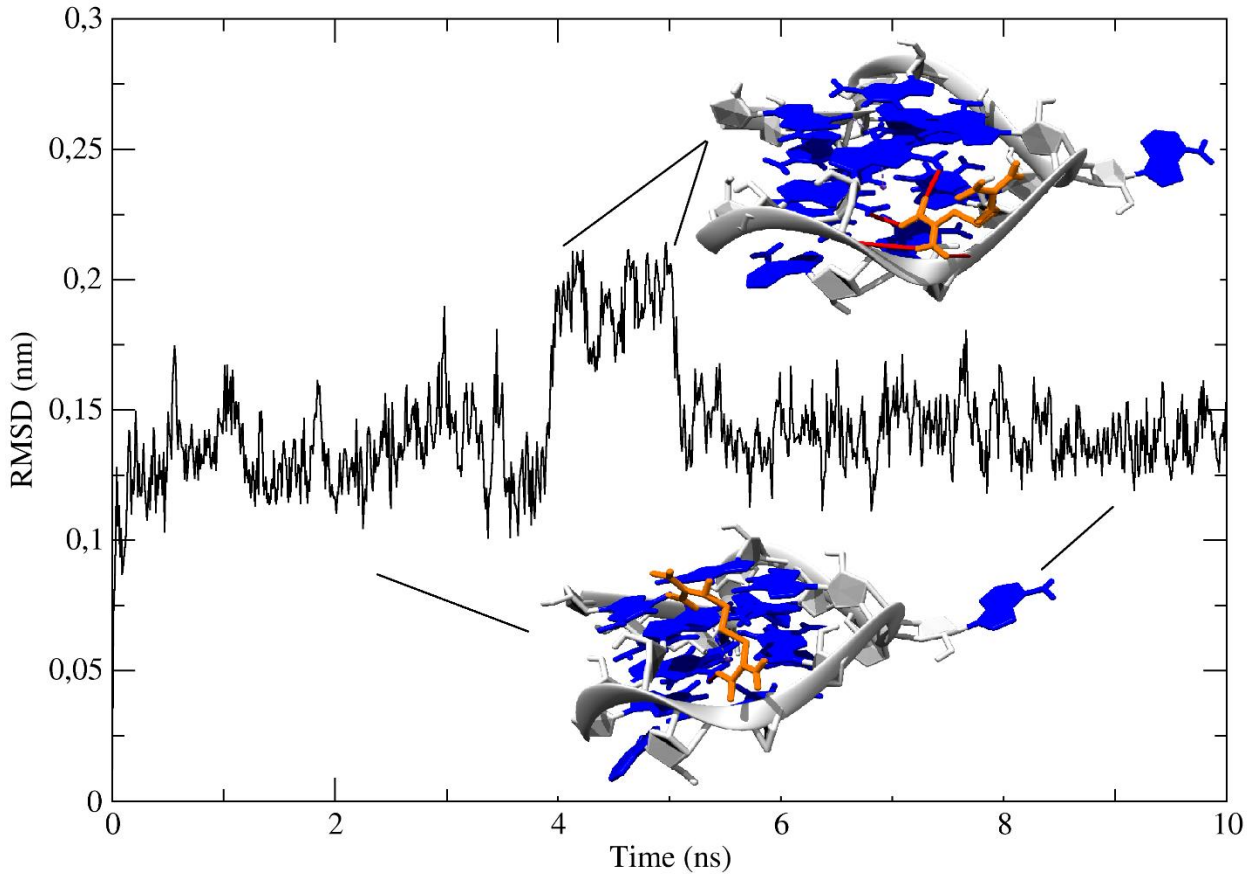

Figure 14 – Representative clusters of the ligand L-arginine.

In regards to the L-arginine, the two clusters obtained are very different from the results obtained via docking. In the docking simulations the ligand is folded upon itself and doesn't appear to establish any hydrogen bonds with the structure.

In the MD runs, it is much more expanded and stretched in a way that allows it to make hydrogen bonds with the structure. The first cluster which is the most prevalent throughout the simulation has the ligand interacting with the quadruplex with a significant part of its surface. A hydrogen bond is spotted between the H atom of its carboxyl group and the H22 atom of the quadruplex's G2 base. In the second cluster, the aliphatic arm of the ligand's radical group seems to point away from the structure, while the carboxyl and amine groups establish 4 hydrogen bonds with the G2 and G3 tetrads hydrogens atoms and with the O1P and O2P of A4 and G6, respectively.

Overall, the structures obtained with the docking simulation seem far different to the ones obtained by cluster analysis after molecular dynamics.

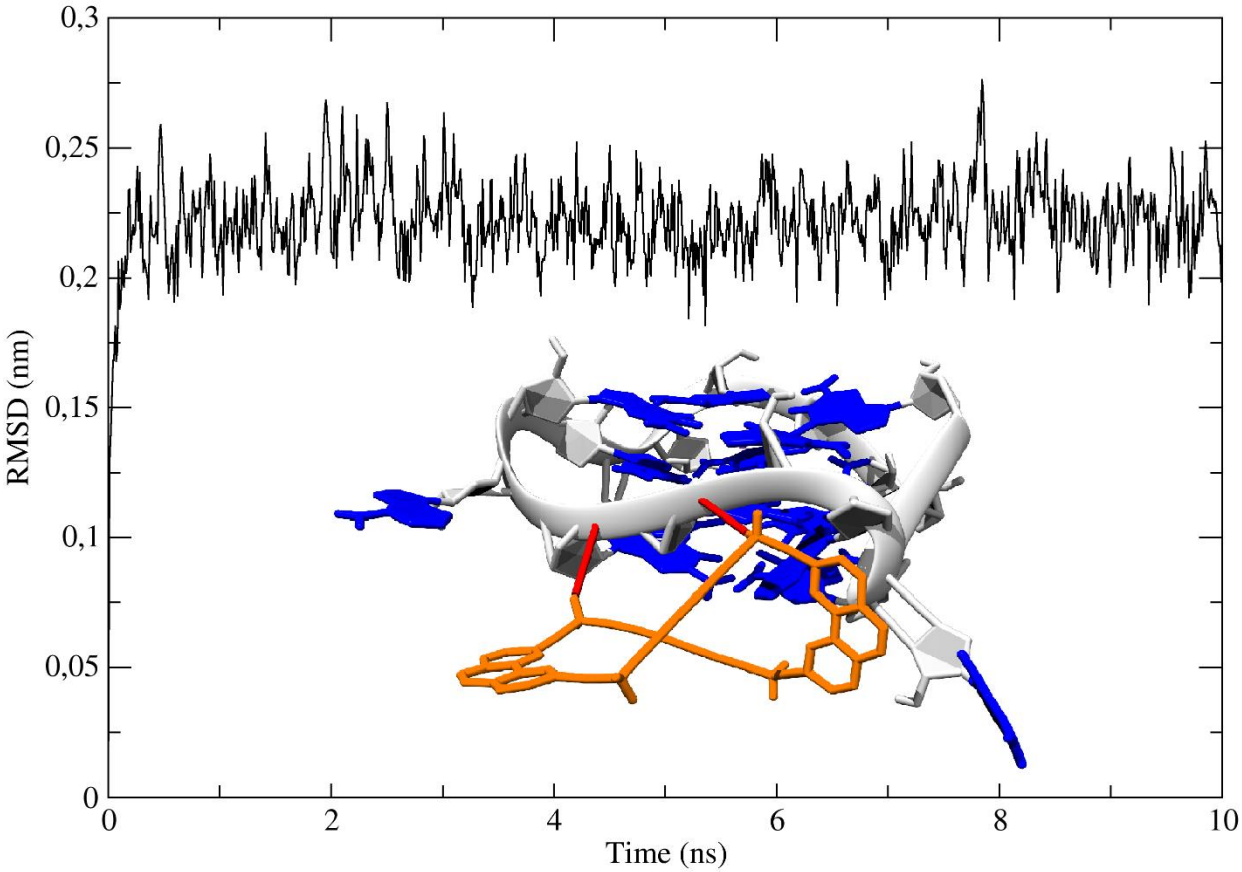

Figure 15 – Representative clusters of the ligand [32]phen$_2$N$_4$.

[32]phen$_2$N$_4$ shows results quite different from the ones obtained via docking simulations. While the placement of the ligand in the complex is in the same loop, and one of the phenantroline moieties is quite close to the quadruplex's C13, the other phenantroline moiety seems to separate itself more from the quadruplex, and the carbon chains in the macrocycle twist far more than in the docking simulations, with the planes of both phenantroline moieties becoming perpendicular to each other.

As the cluster remains stable throughout the rest of the simulation, it is very likely that this was the ideal conformation, and that the aforementioned torsion problems of AutoDockTools were responsible for it not assuming that conformation in the docking simulations. One other problem has surfaced, in the sense that the side arms seem too stretched. This suggests that neither software is fit to render macrocycles with such long side arms.

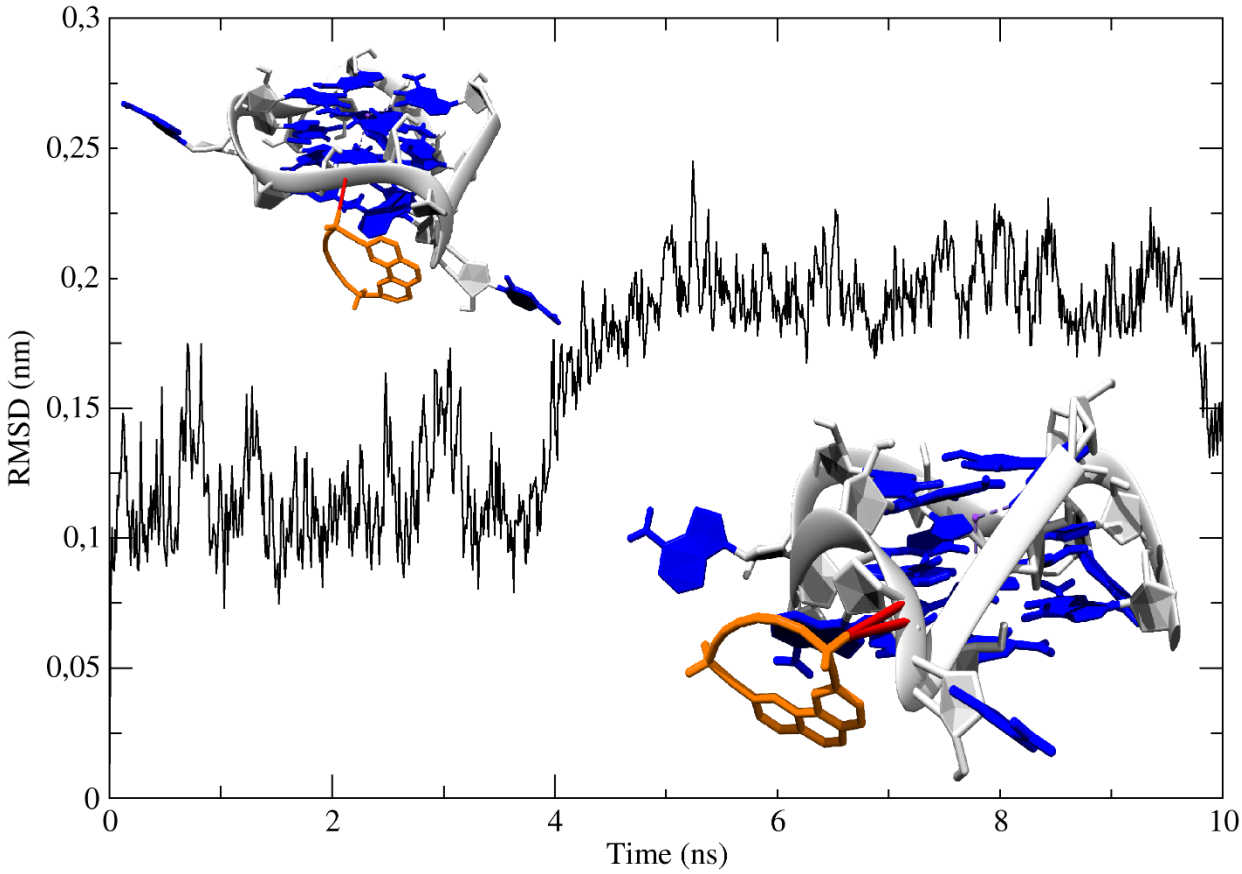

Figure 16 – Representative clusters of the ligand [16]phenN_2.

In the case of [16]phenN_2, in both representative structures of the most significant clusters, the ligand binds to the quadruplex in the same loop as the AutoDockTools simulations predicted, in the third loop of the quadruplex. Despite that, there are some differences to be noted. The structure obtained in first cluster, as Figure 16 shows, is different from the one obtained via docking simulations, as the overall macrocycle structure appear to have rotated 90 degrees counter-clockwise in an axis perpendicular to the quadruplex loop. Hydrogen bonds occur between a hydrogen atom in the leftmost nitrogen group of the phenanthroline moiety and the G4 G13. The next cluster shows a conformation closer to the one obtained via docking, with the macrocycle's phenantroline moiety facing downwards along with the loop's path. Furthermore, the phenantroline moiety appears to have π-π stacking with the sugar base of the quadruplex's A12, and there are hydrogen bonds between the aforementioned hydrogen atoms and the quadruplex's C13.

Overall, this second cluster lasted more time throughout the simulation, and appears to be the the most favorable conformation.

4.1.4 - MM/PBSA analysis

The final conformations of each MD run were then analyzed in terms of binding free energy. The method chosen to perform these calculations was MM/PBSA. However, as Gromacs 2016.3 does not have any built-in tools with which to perform these calculations, the external tool

g_mmpbsa was imported into Gromacs. This tool calculates the average binding energy of the structure in a 3 step process. First, the potential energy in a vacuum is calculated. Then, the polar solvation energy of the complex, followed by the non-polar solvation energy (using a SASA model) is calculated. Then finally, the pbsa calculations are performed to determine the average binding energy of the G4-ligand complex.

The values obtained for every ligand is now shown in Table 2.

Table 2 – Binding energy obtained per ligand, in KJ/mol. Presented in the same order as the previous one.

Ligand	Binding energy (kJ/mol)
Pyridostatin	-1576.230
Phen-DC3	-752.393
C_8	-498.470
C_8-NH_2	-443.278
L-arginine	-629.570
[32]$phen_2N_4$	-1229.646
[16]$phenN_2$	-1217.482

It is worth noting that these values are of a much higher order than the ones previously reported for AutoDockTools. However, the numbers themselves are only used as qualitative data, to establish comparisons between each ligand's binding strength to the G4, and should not be taken as absolute values.

The binding energies obtained through this method correlate well to the ones obtained in AutoDockTools, with some changes being noted.

Firstly, pyridostatin shows the lowest binding energy value in this method, whereas it was the second best behind [32]$phen_2N_4$ in the last one (though the latter was disregarded due to torsion simulation issues inherent to the software). Thus, pyridostatin would be considered the best ligand.

[32]$phen_2N_4$ and [16]$phenN_2$ follow second and third place respectively.

Surprisingly, L-arginine shows a better result in this test than in AutoDockTools. It is very possible that this may also be due to AutoDock's aforementioned problem at calculating torsions and angles correctly. As a groove binder, it's very possible that some unconsidered interactions in the ligand would allow for more stable conformations that the program inherently could not analyze. Indeed, during MD runs, L-arginine seemed to establish additional H-H bonds with the G4 structure thus presenting a relatively lower binding energy.

C_8 and C_8-NH_2 shown the worst binding energies, which would allow them to be considered the worst ligands out of the seven, were this solely the only test considered. Again, the problem with torsion angles may have made it so that the program misjudged some possible conformations that the aliphatic arm would allow these ligands to assume.

Overall, it is worth noting that neither the docking experiments nor the molecular dynamics runs should be considered in isolation. The literature suggest that the molecular dynamics runs should be better considered as a method to discover the best binding poses for the ligands, as it better simulates ligand flexibility and binding conditions, while docking experiments would be better in determining relative binding energy between ligands. [64] In this case, both methods were performed, and while there are slight differences between the results obtained in terms of binding energies, the ranking of the ligands based on their binding strenght remains similar, with molecular docking rating them as [32]$phen_2N_4$ > Pyridostatin > Phen-DC3 > [16]$phenN_2$ > C8 > C_8-NH_2 > L-arginine, and MM/PBSA as Pyridostatin > [32]$phen_2N_4$ > [16]$phenN_2$ > Phen-DC3 > L-arginine > C_8 > C_8-NH_2.

Further tests were necessary to further assess the G4-binding potential of these ligands.

4.2 - Circular dichroism

The interactions between the seven ligands and the pre-miR 149 G4 structure were studied through biophysical studies, namely, circular dichroism and fluorescence spectroscopy. The main objective was to determine whether the ligands induced the formation of the G4, as well as any further conformational changes and overall stabilization. This way, the accuracy of the *in silico* results could be verified.

CD titrations (with ligand titrating the G4 sequence) and thermal denaturing or "melting" studies were performed, with each experiment adding molar equivalents of each ligand to the sequence. An initial screening was performed to evaluate the formation of G4 structure in presence of two monovalent cations: K^+ and Na^+. The coordination of monovalent cations in the G4 is essential for its formation and determines the stability of structure. The pre-miR-149 G4 structure was confirmed to have a parallel conformation in presence of both monovalent cations, with a positive signal around 260 nm, and a negative one around 240 nm.

Firstly, the spectral data and melting temperature graphs of the ions will be examined, to draw general conclusions from the screening. Then, the same data for the ligands will be analyzed, followed by a comparison of all melting temperature variations (ΔT_m) obtained for each ligand, which are shown in Table 3. Afterwards, the spectrum of the titration and melting studies are presented and examined individually.

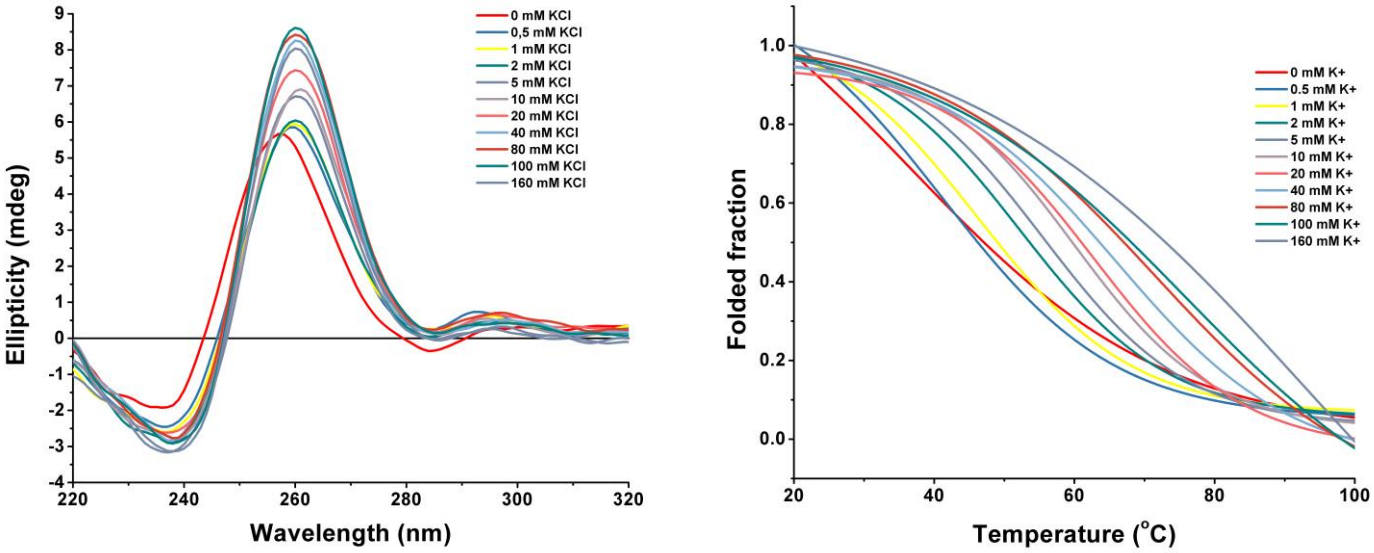

Figure 17 – Spectral data (left) and melting temperature variation spectra (right) for K+.

Regarding K^+, the increasing ellipticity at 260 nm likely indicates that the G4 structure was increasingly stabilizing. [65]

The melting data for K^+ showed an increase in the melting temperature consistent with increasing ion concentrations, with a ΔT_m of 18.89ºC from 0 to 160 mM K^+. the melting curves stayed in the expected shape.

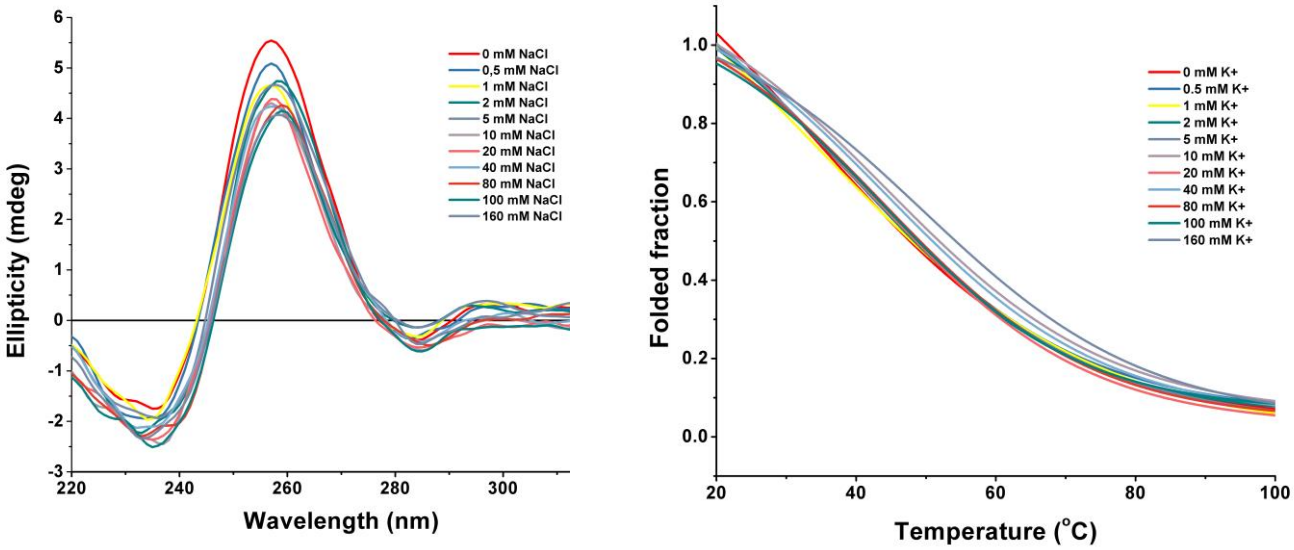

Figure 18 – Spectral data (left) and melting temperature variation spectra (right) for Na+.

Na^+ spectral data shows a slight red shift at 260 nm as molar equivalents of ligand increase. However, the ellipticity values also show a steady decrease.

The melting data for Na^+, shown in Figure 18, displayed the same correlation between the melting temperatures and ion concentrations. Despite that, the resulting ΔT_m shows a value of 5.2ºC in the 0 - 0.8 molar equivalent range, which is far below that of K^+. Like the K^+ results, the curves retained an acceptable shape throughout the experiment. Regardless, it has a much lesser stabilizing effect compared to K^+, which is consistent with what is so far known about ionic stabilization of G4es. As was discussed in Chapter 1, different ions have different stabilizing effects, and in the sequence shown, K^+ is indeed known to have a greater stabilizing effect than Na^+.

As specified in the previous chapter, monitoring these spectra allows one to assess the effect the ligands had on the secondary structure. If the ellipticity of each peak were enhanced or maintained, the ligand can be considered to have a stabilizing effect on the structure. If any of the peaks change drastically, disappear, or if new peaks were to appear altogether, then the ligand would have changed the structure.

Overall, the ligands showed the same results in regards to positive and negative peak wavelengths. The negative and positive signals are all at or around the 240 and 260 nm, respectively. That means that the conformation of the G4 structure wasn't changed in any of the cases, and every ligand possibly had a stabilizing effect on the structure.

Melting studies were performed to better compare the ligands based on the stabilizing effect they had on the G4 structure. These studies consisted of titrating the G4 structure with

increasing amounts of ligand, followed by a linear temperature increase. The folding of the structure was monitored, to determine the melting temperature (T_m), which was defined as the temperature of the midtransition point. The longer it took for the structure to denature, the more the T_m increased, and thus, greater the stabilizing effect. As molar equivalents of the ligand were added, the melting temperature increased.

The melting temperature variation (ΔT_m) was defined as the subtraction of the highest obtained T_m by the starting T_m of the solution containing G4 without ligand. A higher ΔT_m implies a better ability of the ligand to stabilize the structure.

It is worthy of note that, at some points for all ligands, the molar equivalents added stopped having the predicted effect. The melting curves obtained should have a sigmoidal shape to be considered valid, whereas upon reaching some molar equivalent values in the studies, they started having abnormal shapes, such as parabolic shapes or even linear. At that point, any further molar equivalents after that cease to be considered valid.

The ellipticity and thermal spectra obtained for each ligand shall now be shown in Figures 19-22.

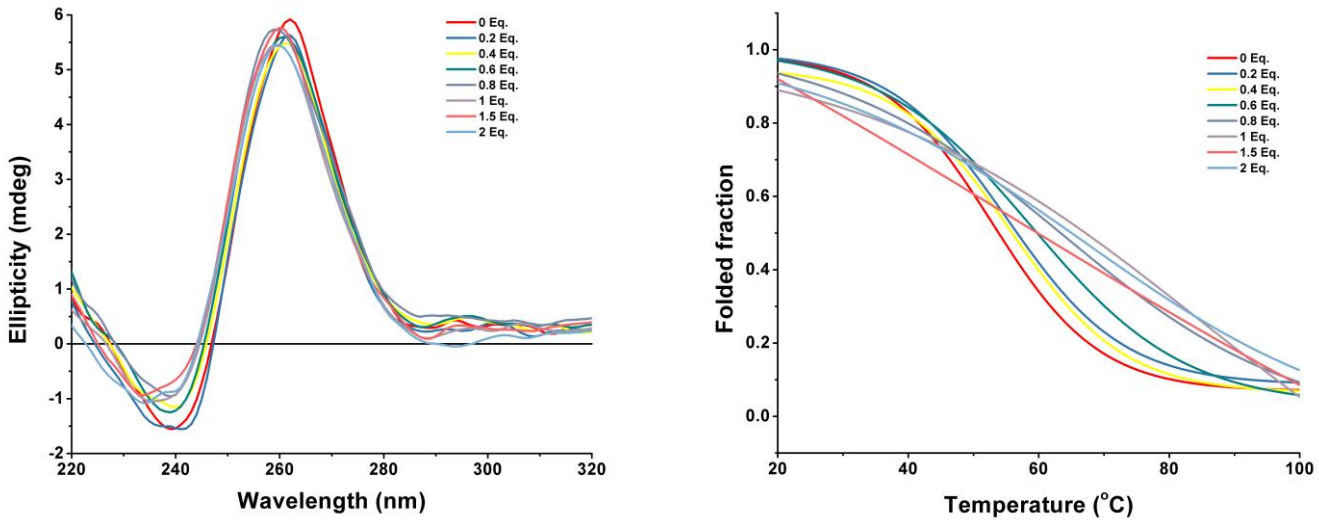

Figure 19 - Spectral data (left) and melting temperature variation spectra (right) for pyridostatin.

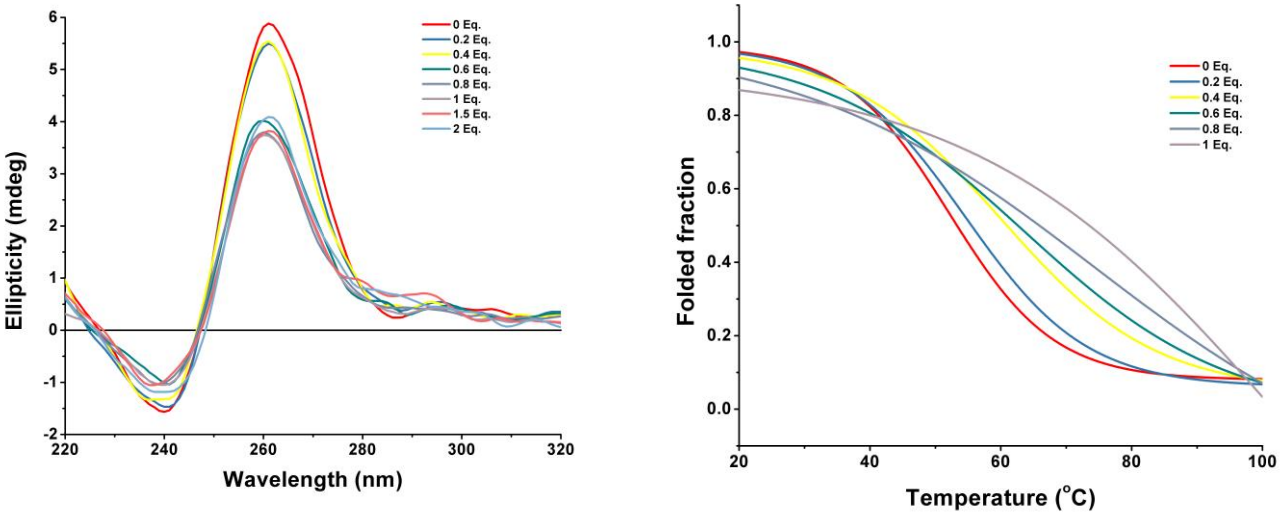

Figure 20 - Spectral data (left) and melting temperature variation spectra (right) for C8.

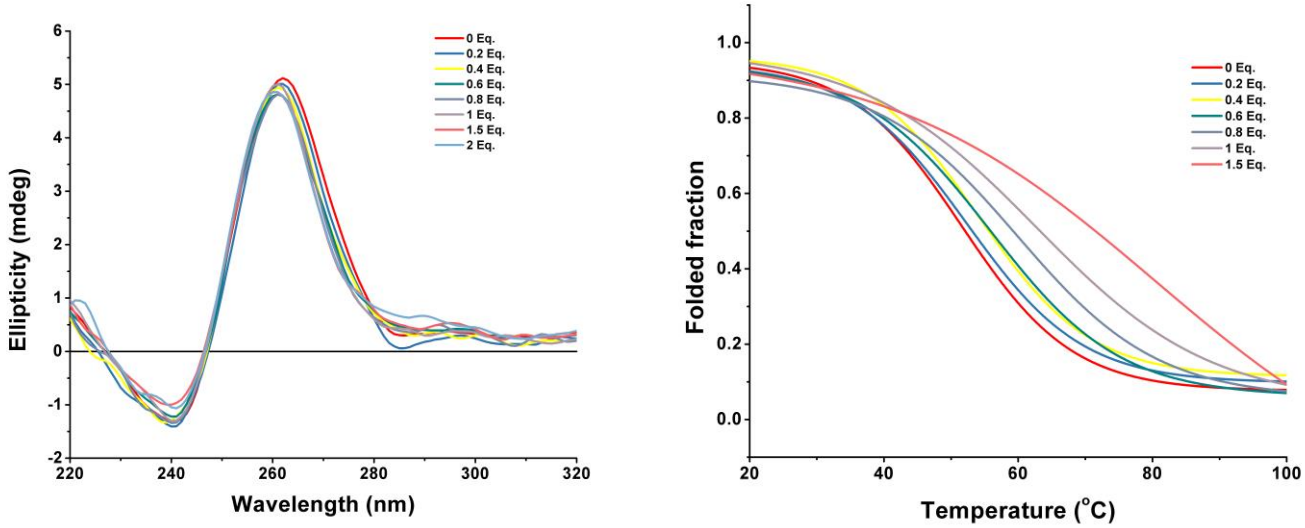

Figure 21 - Spectral data (left) and melting temperature variation spectra (right) for C8-NH2.

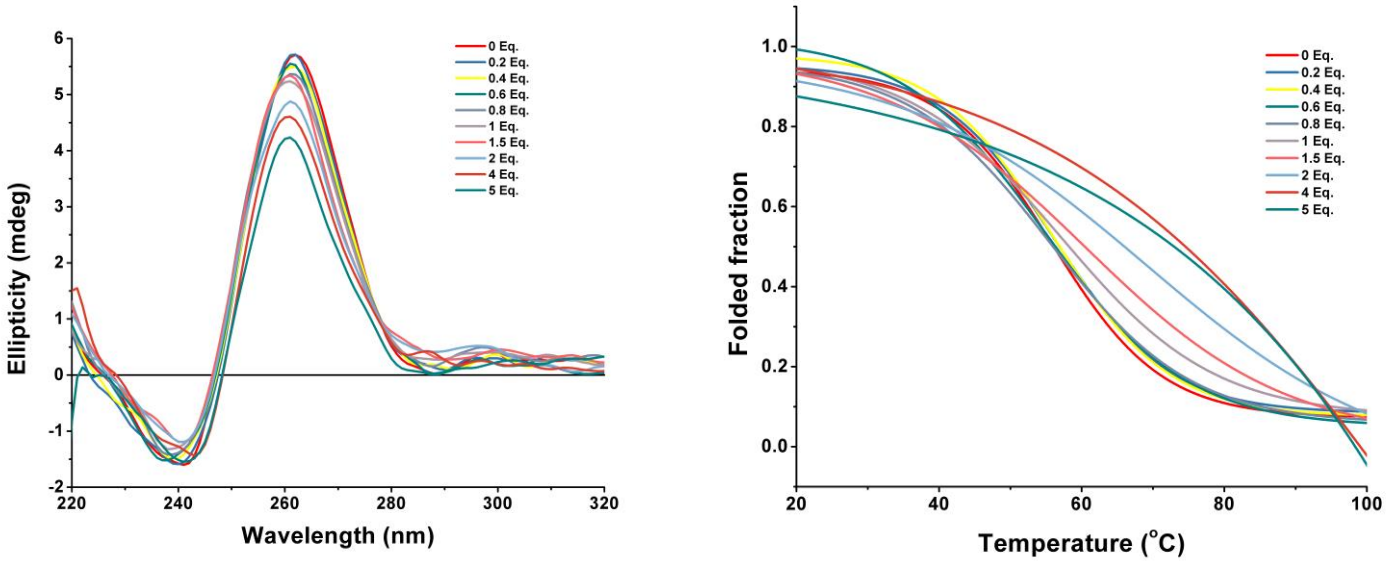

Figure 22 - Spectral data (left) and melting temperature variation spectra (right) for [16]phenN_2.

In the case of C8, as mentioned, the negative and positive signal placement indicates that the quadruplex remained in a parallel conformation. The ellipticity values at 260 nm steadily decreased as more molar equivalents were added, which is very likely due to increasing ligand-quadruplex interactions. [66]

Regarding melting temperatures, C_8 showed the best results, with a ΔT_m value of 22.11ºC. Starting from a baseline of 52.10ºC, at 0 molar equivalents, it increased to 74.20ºC at 0.8 molar equivalents. Although the experiment continues until 1 molar equivalent was reached, it was stopped, due to the parabolic shape that the curve was assuming. Regardless, the ΔT_m value shown means that the ligand has an extremely good stabilizing effect on the quadruplex structure.

In regards to pyridostatin's ellipticity values, increasing concentrations of ligand induced a slight hypsochromic shift. The shift is so small (around 1 nm) to the point of being negligible. Ellipticity values stayed mostly constant at 260 nm, which likely indicates the presence of stack binding.

For pyridostatin, the T_m rises from a base value of 52.83ºC, at 0 molar equivalents of ligand, to 63.69ªC, at 0.8 molar equivalent. That shows a ΔT_m value of 10.95ºC. Again, past 0.8 molar equivalents, the shape of the curve becomes parabolic. Regardless, we can conclude that pyridostatin has a moderate stabilizing effect on the structure, albeit lower than C_8's in the same conditions.

C_8-NH_2 showed a slight bathochromic shift in the spectral data, but again, to such a low extent as to be considered negligible. In regards to ellipticity, C_8-NH_2 remained mostly stable

throughout the run of the experiment. This is expected, as molecular dynamics experiments showed that this ligand's preferred binding mode consists of stack binding, which is consistent with a lack of change in ellipticity. [66]

Regarding the melting studies, the T_m rises from a value of 51.57ºC at 0 molar equivalents to 59.92ºC at 0.8 molar equivalents, with a ΔT_m value of 8.36. These results place C_8-NH_2 in an intermediary position in regards to thermal stabilization compared to other ligands.

[16]phenN_2's spectral data shows a drop in ellipticity at 260 nm, which is likely due to ligand-quadruplex interactions. [66] According to molecular dynamics simulations, this ligand's binding mode is mostly groove binding, which further verifies that hypothesis.

[16]phenN_2 has a starting T_m value of 55.81ºC, at 0 molar equivalents, it achieves a value of 57.13ºC at 1 molar equivalent, with almost negligible improvements until then, and with further molar equivalents' melting curves being unreliable. The resulting ΔT_m has a value of 1.32ºC, and with such a low value, the stabilizing effect it had on the structure can be said to be almost negligible by this parameter.

The following table (Table 3) shows the ΔT_m obtained per ligand.

Table 3 – Melting temperature variation values obtained per ligand

Ligand	**ΔT_m / ºC**
Pyridostatin	10.95
C_8	22.11
C_8-NH_2	8.36
[16]phenN_2	1.32

According to all data collected, the trend in descending G4 ligand affinity appears to be C_8 > Pyridostatin > C_8-NH_2 > [16]phenN_2.

4.3 – Fluorescence essays

The ligands interaction with the pre-miR-149 G4 structure was further tested using fluorescence spectroscopy. This technique, as discussed in Chapter III, consisted of titrating a ligand with the G4 sequence, and monitoring the emission spectra in the ligands excitation wavelength, which varies from ligand to ligand. After proper data fitting, this procedure allows for the determination of the apparent dissociation constants (K_D) for each of the ligands. The lower the dissociation constant obtained, the better the affinity between the ligand and the G4 structure. Many ligands were ineligible for this technique, due to their non-fluorescent nature. The only ligands that could be used for this technique were [16]phenN_2, and AO derivatives C_8 and C_8-NH_2.

C_8-NH_2 showed a K_D of 1.73 × 10^{-6}, meaning it was the ligand with the highest affinity for the G4 structure. The spectrum is shown in Figure 23. From 0 molar equivalents to 1 molar equivalent of G4, it shows fluorescent enhancement, as the fluorescence intensity increases. From 1 molar equivalent to 2 molar equivalents, the binding mode changes, as the fluorescence intensity decreases, showing fluorescent quenching. The turning point for this biphasic binding mode is at a ratio of 1:1 of G4 structure to ligand.

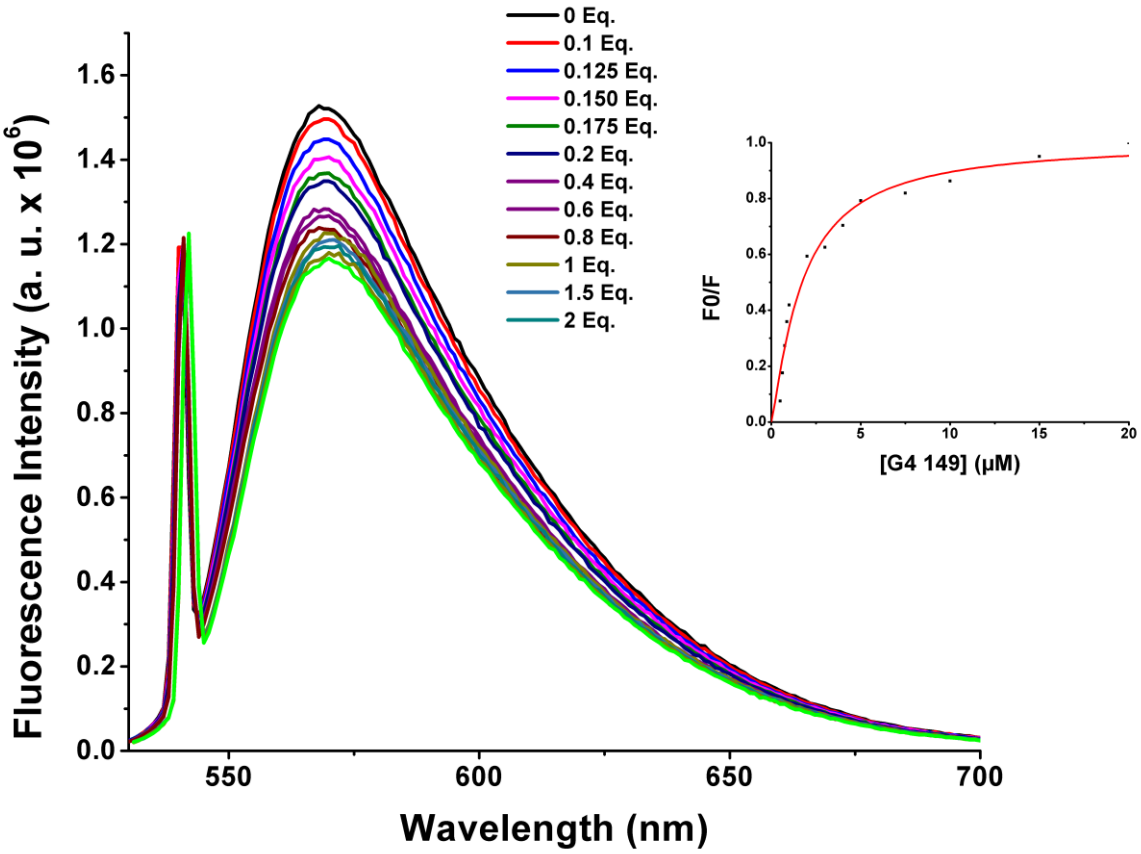

Figure 23 – Fluorescence emission spectra of C8-NH2. Fitting graph is shown at the top right corner.

C_8 showed a K_D of 8.24 × 10^{-7} M, meaning it was the one with the best affinity for the G4 structure. The graph is shown in Figure 24. Upon titration with the G4 solution, an up to sevenfold fluorescence enhancement was observed, after a very slight quenching at the first molar equivalent. This shows that there is a biphasic binding mode between pre-miR-149 and this ligand, with the turning point from fluorescent quenching to fluorescent enhancement being at a ratio of 1:10 of G4 structure to ligand. The binding was cooperative, with a Hill constant of 5.56.

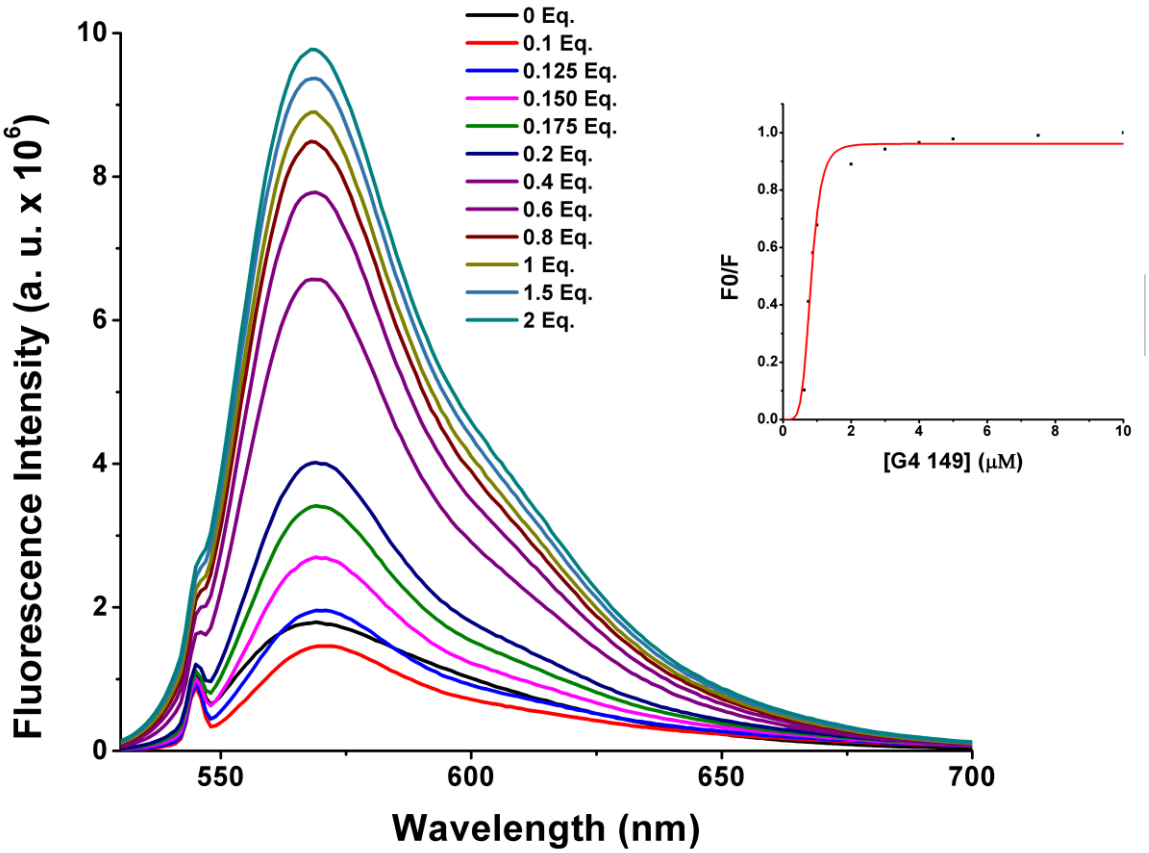

Figure 24 - Fluorescence emission spectra of C8. Fitting graph is shown at the top right corner.

[16]phenN_2 had a K_D value of 2.8×10^{-6}, meaning it had the worst binding affinity of the three ligands. The graph is shown in Figure 25. It exhibited different fluorescent activity compared to the two previous ligands. A biphasic binding mode is observed, with two dramatically different stages. The first stage, from 0 to 0.15 molar equivalents of ligand, showed fluorescent quenching, with a dramatic decrease of fluorescence intensity. From that point onwards, to 3.5 molar equivalents of G4, there was fluorescent enhancement, with a small to moderate increase of fluorescence intensity.

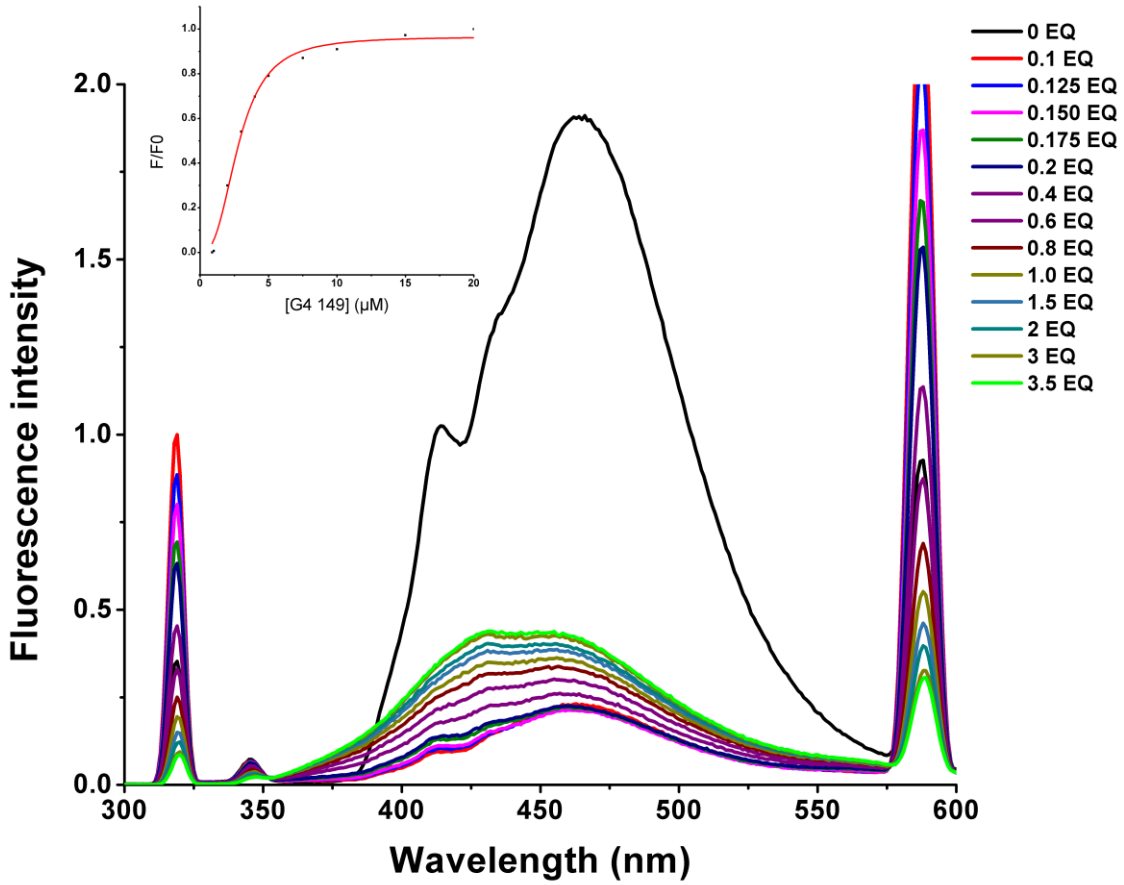

Figure 25 - Fluorescence emission spectra of [16]phenN_2. Fitting graph shown near the top left corner.

Table 4 shows the dissociation constants for these ligands, in increasing order of K_D (decreasing order of affinity).

Table 4 – Dissociation constants (K_D) per ligand.

Ligand	K_D / M^{-1}
C_8	8.24 x 10^{-7}
C_8-NH_2	1.73 x 10^{-6}
[16]phenN_2	2.80 x 10^{-6}

In any case, the ligands can be ordered in terms of binding affinity to the quadruplex structure in this sequence, by these results: C_8 > C_8-NH_2 > [16]phenN_2. AO derivatives again seem to be the most promising ligands of the group, compared to the macrocyclic [16]phenN_2.

Considering all these results, again there is a significant difference between the results obtained by molecular dynamics simulations and those obtained by these techniques. Of these three ligands, MM/PBSA tests in a previous section established that [16]phenN_2 had the lowest, and thus most favorable free binding energy, followed by C_8 and C_8-NH_2. However, in these

fluorescence titration essays, the conclusions were different, with C_8 as the best ligand in terms of binding affinity, followed by C_8-NH_2, with [16]phenN_2 being the worst. This suggests that the molecular dynamics simulations approach has some flaws to it.

4.4 - Incongruencies between computational and biophysical experiments

Indeed, such problems have been encountered before in the literature, regarding the MM/PBSA method. Likely, the problem is endemic to the approach, such as crude approximations, like a lack of conformational entropy and number and free energy of water molecules in the binding site. Although our approach tried to simulate an environment as close to real world conditions as possible, the method's flaws still stand. There are many different variants of the method used in this work, with performances varying significantly based on the system in question. [67] As we applied the same method to seven different ligands with vastly different parameters, it's likely that the method was not suited to some particular cases, such as the AO derivatives, and as such the given values were erroneous.

Other researchers also tried different methods to improve the MM/PBSA method in these regards, such as simulating only a small part of the ligand-macromolecule complex, using minimized structures instead of molecular dynamics snapshots, or but in most of these cases only revealed deeper problems with the method, such as extreme difficulties describing intermolecular interactions. [68] This reveals deeper problems with the method which likely means it is unsuitable for comparing vastly different ligands with each other in terms of free binding energy.

Indeed, even when comparing binding energies between enantiomers, or compounds with identical molecular weights, this method seems to come up short, with the authors of the study concluding that cases where binding energies were correctly predicted were attributable to chance instead of accuracy. [69] Thus, the results obtained by docking in regards to binding energies will be mostly discarded (while the results obtained regarding docking conformations and molecular dynamics should still be considered valid).

Thus, considering only the results obtained from the biophysical techniques in terms of ligand affinity, C_8 and C_8-NH_2 can be considered the ligands with the greatest affinity to the pre-miR-149 G4 structure, and should be considered in any further areas of research on this subject as promising candidates, for instance, for drug-delivery systems and other related topics.

Further testing was done to test whether C_8-NH_2 can be considered a promising ligand for other quadruplex-forming sequences. Results for NMR characterization of the C_8-NH_2 chromatographic support and subsequent affinity chromatography experiments follow.

4.5 - Affinity chromatography

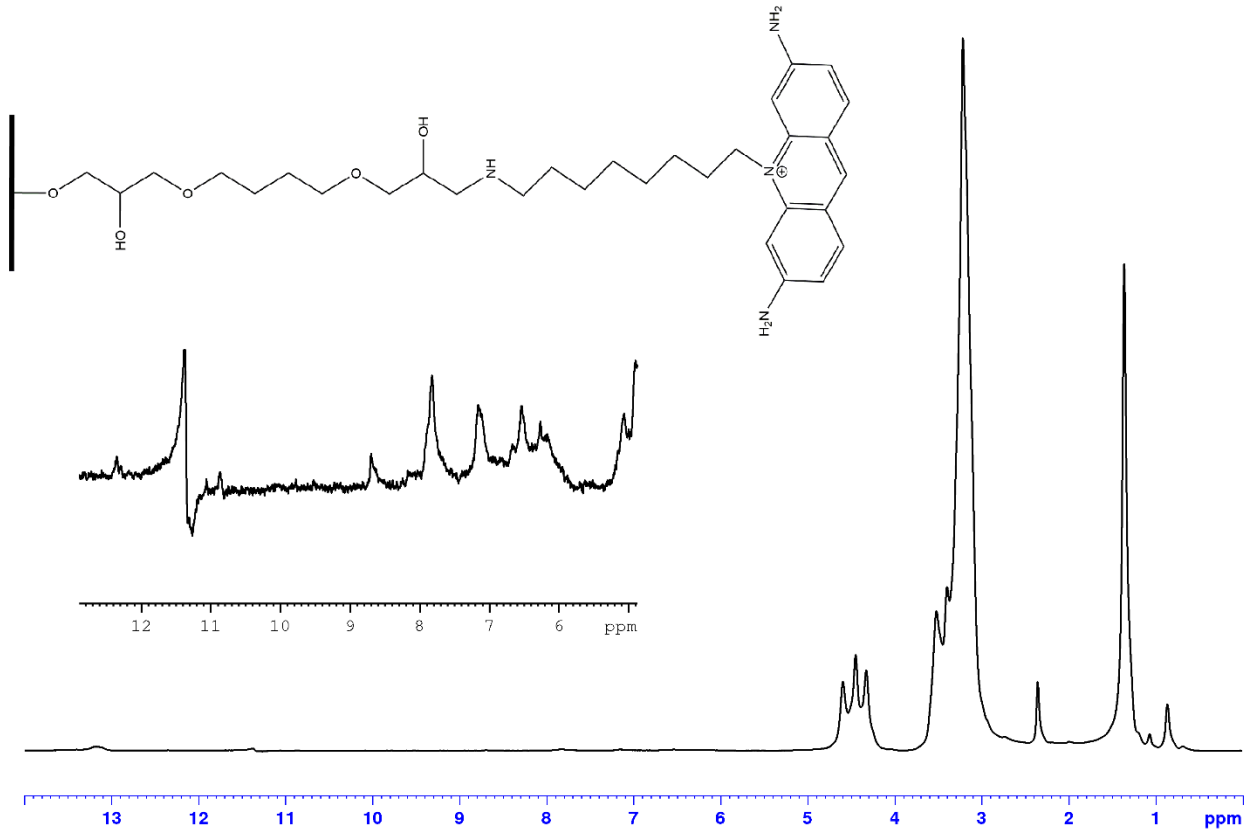

Figure 26 – NMR spectra of the C8-NH2 ligand bound to Sepharose 6B.

The high-resolution NMR spectroscopy combined with magic angle spinning (HR-MAS) reduces the broad Sepharose CL-6B background signals thus resulting in better resolved spectra, possibly contributing to the identification and quantification of compounds linked to the matrix. [70]

As is shown in Figure 26, the main chemical shifts from Sepharose-6B are spotted between 0 and 5 ppm. The resonances from the acridine moiety can be seen from 6 to 9 ppm.

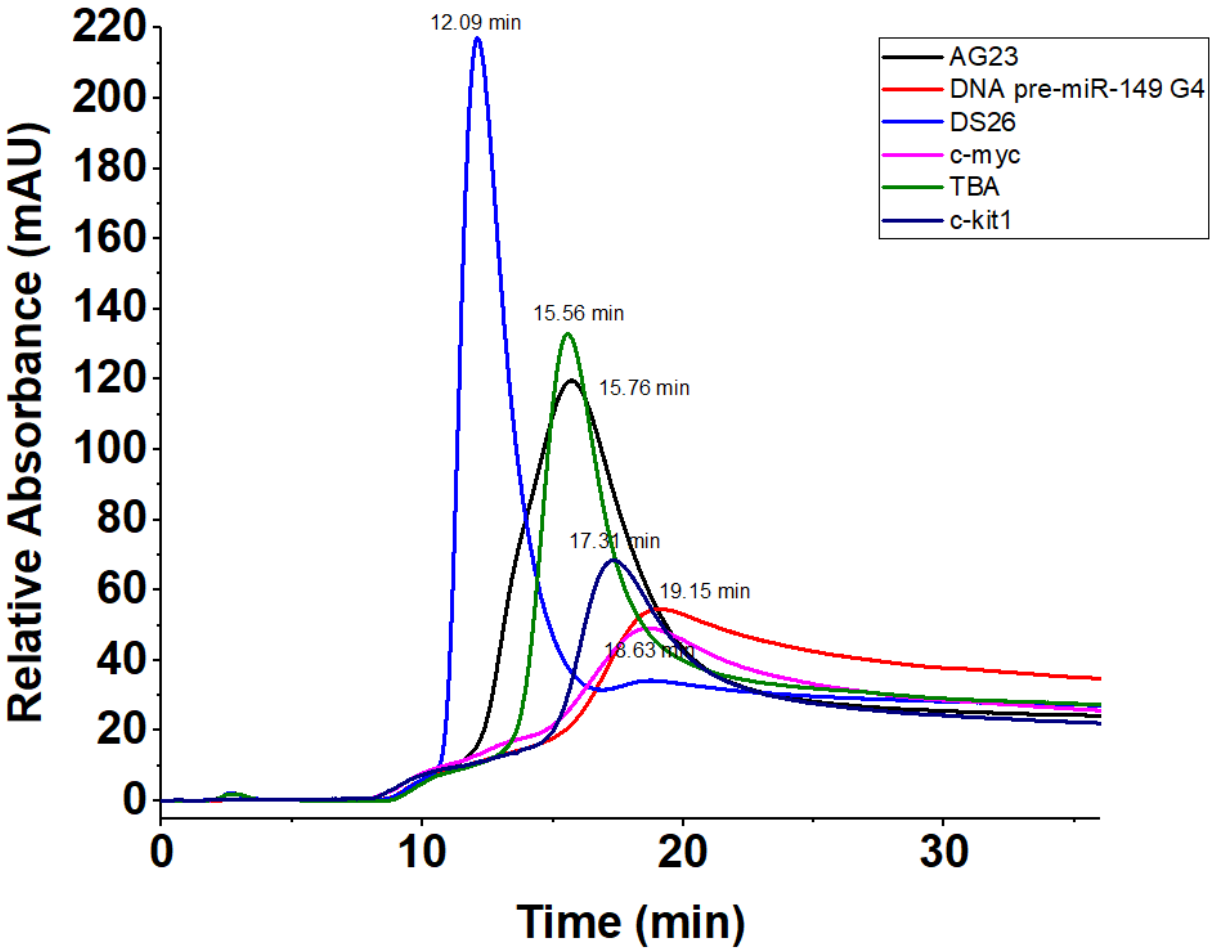

Figure 27 – Chromatographic spectrum for every sequence tested. Every sequence shown in different color, with retention times shown at the top of each peak.

Affinity chromatography was employed in order to evaluate the binding and retention behaviors of different G4s to C8-NH_2 support. Several chromatographic experiments were performed to determine the retention times of different G4-forming oligonucleotides (AG23, c-MYC, TBA, c-KIT1 and pre-miR-149 G4) and a duplex oligonucleotide (DS26).

The oligonucleotides were annealed in 30 mM K-phosphate buffer, pH 7.1 and 100 mM KCl at 95ºC for 10 min, followed by ice cooling until room temperature. Thereafter, oligonucleotides were injected onto the column and a linear gradient from 10 mM Tris-HCl pH8 to 1 M NaCl and 100 mM L-arginine was applied to elute the sequences. They eluted in the following order of increasing retention time: DS26 (12.09 min), TBA (15.56 min), AG23 (15.76 min), c-kit (17.31 min), c-myc (18.63 min) and pre-miR-149 (19.15 min). Thus, pre-miR-149 showed the greatest retention time of all 6 oligonucleotides. The c-myc and c-kit retention times showed differences of, respectively, 0.52 min and 1.38 min towards the ligand's retention time. The remaining 3 sequences have much more sizable differences. This confirms that the ligand C_8-NH_2 has good affinity with this particular pre-miR-149 sequence, followed by c-myc and c-kit, with DS26, TBA and AG23 having less affinity.

Furthermore, of the six sequences tested, AG23 and TBA have an anti-parallel conformation upon forming a G4 structure [71], and while c-MYC, c-KIT1 and pre-miR-149 have a parallel conformation [72] [73]. The G4 oligonucleotides that adopt anti-parallel conformation showed lesser retention times, meaning their affinity to the C8-NH2 support is lower than the others with parallel conformations. This suggests that the ligand binds better to G4s with parallel conformations. As the docking analysis showed, the ligand has π-π interactions with the top tetrad and the aliphatic arm interacts with the grooves. In an anti-parallel conformation, the grooves offer much less contact points and a greater possibility for steric clashes, meaning the parallel conformation is more favorable for this particular ligand. [74]

Chapter V

Conclusions

The interest in G4s by the scientific community is on the rise. They open many new therapeutic approaches to cancer, both as a way of stopping cancer spread in the body, and as a potential target for drug delivery systems. In both of these approaches, the necessity of finding suitable ligands is paramount in advancing this area of research.

In this work, we endeavored to test seven prominent ligands found in the literature, in the hopes of establishing qualitative comparisons between their affinity to the chosen G4 structure. Whereas in previous work, these comparisons were made mostly through biophysical techniques, in this one, molecular modelling techniques were added, based on the simulation of the structures of both ligand and G4 and their ensuing interactions, to hopefully discern more in depth mechanisms of the ligand-G4 interactions, such as the ligand binding site, the binding conformation, etc. Tentative binding energy values were also obtained in each case, to better assist the ranking of the ligands based on their affinity to the structure. Molecular dynamics simulations then simulated the behavior of such complexes in a system with explicit solvent, to determine their stability, and cluster analysis was used, to determine the most often occurring conformations throughout the simulation. Thermodynamic studies were then conducted to determine the binding energies for each ligand throughout the run, supplementing the previously obtained results.

Traditional biophysical techniques used in these studies were then used to experimentally verify the docking results. Circular dichroism spectroscopy and melting studies were performed to determine the stabilizing effect the ligands had on the structure, and fluorescence essays were performed to determine binding affinities between each ligand and the G4 structure. Affinity chromatography experiments were also held to compare the effects of one of the most promising ligands to G4 sequences with different topologies and a duplex sequence.

The seven ligands chosen were: AO derivatives C_8 and C_8-NH_2, [16]phenN_2, [32]phen$_2N_4$, Phen-DC3, L-arginine and pyridostatin. The molecular modelling tests and biophysical techniques had vastly different results in regards to determining the ligand with the best binding affinity.

The docking studies established the following trend, in order of descending ligand affinity to the G4 structure: Pyridostatin > [32]phen$_2N_4$ > [16]phenN_2 > Phen-DC3 > L-arginine > C_8 > C_8-NH_2. The biophysical techniques, despite not being performed on all ligands, established a different trend: C_8 > Pyridostatin (only tested on circular dichroism spectroscopy and melting studies) > C_8-NH_2 > [16]phenN_2.

Such a difference between results obtained is evidenced in other cases in the literature, who point out many inherent flaws of the computational simulations, such as not correctly accounting for the binding energy of water molecules on the binding site and lack of conformational entropy. Thus, it seems that, at least for the currently available software, the binding energy results are unreliable and should be discarded in the face of contrary experimental data. Ligand conformations obtained through docking also seem mostly reliable, except in the case of macrocyclic ligands, where the program notably falls short in correctly predicting the flexibility of the structure, as reported in the literature. Molecular dynamics simulations should in any case be performed upon those obtained conformations, to test if the ligand maintains them throughout the run, as that attests to their reliability.

Affinity chromatography shows that in the case of C_8-NH_2, one of the ligands proven to have a greater affinity to the G-quadruplex structure, it binds better to G4s with a parallel conformation, in contrast to antiparallel conformations, let alone duplex structures. Its selectivity to the sequence in question is also superior to every other sequence tested.

Pyridostatin also showed promising results in circular dichroism studies, though due to its lack of fluorescent activity, it could not be used in fluorescence essays.

[16]PhenN4, while still showing high affinity to the G4 structure, had less affinity than any other ligand tested with biophysical techniques. [32]$phen_2N_4$, while showing tentatively good results in the docking studies, could not be tested with biophysical techniques due to a lack of resources.

However, if one ligand were to be considered the best out of all seven of the ones tested, it would be C_8. It excelled in both circular dichroism and fluorescence titration essays, and as C_8-NH_2 is a chemical precursor to it, it is very likely that it would show similar or superior results to the ones obtained in affinity chromatography. In the data obtained by computational methods, it is suggested that the interaction between the ligand and the complex occurs through a π-π stacking interaction between the acridine moiety and lower G4 tetrad, with the aliphatic arm having hydrogen bond interactions with the guanine G9 and the adenine A12, which also form part of the lower tetrad.

Thus, out of all 7 ligands, pyridostatin, C_8 and C_8-NH_2 are the ones who show most affinity to the pre-miR-149 G4 structure, and should be considered first and foremost candidates in any further research in this area, as well as tentatively being considered potent G4 ligands in general.

Chapter VI

Future perspectives

The current work could be improved upon in many ways. Every phase of this work will be examined in light of the changes that could be made to improve it.

Firstly, as was highlighted during the work, the computational approaches are lacking in some regards. The docking conformation approach was limited by the software used, as it failed to correctly represent the flexibility of macrocycles in the docking phase, and the binding energy values it calculated proved entirely wrong in regards to the AO derivative ligands. In regards to the molecular dynamics simulations, the binding energy values calculated at the end also misrepresented the AO derivatives, and were overall proven unreliable.

In regards to the software itself, the only thing that can be done is using the most updated software and accompanying the current literature, to be aware of the pitfalls of this method, and not to rely upon the binding energy values calculated. There are also many workarounds to some of the problems that were not used throughout the course of this work that could perhaps improve the obtained results had they been implemented. The software used for the docking simulations, AutoDockTools, was also somewhat outdated at the time of writing, with many current docking studies focusing on AutoDock Vina, which shows some signs of performing better in some regards.

Lastly, another approach for the molecular dynamics studies would be to start the simulations with the ligand and G4 structure placed randomly in a box and allowing them to bind together naturally, instead of starting from a predetermined conformation, as was done in this work. This was done due to technical difficulties, and to save computational power, but may have influenced the results in unknown ways.

In regards to the biophysical techniques, the main pitfall of this work was that not all ligands were tested. This was due to lack of supplies and material constraints, and in the process much data that may have impacted the work may have been lost. The main conclusions of the work would probably remain the same, as the remaining ligands were unlikely to be as potent as pyridostatin and the AO derivatives, but empirical data would have confirmed this thoroughly.

Regardless, the current work has identified three promising ligands that may be used to stabilize this particular G4 sequence, and more G4 structures in general, particularly those with a parallel topology. This could lead to many new medical advancements in the future, such as

anticancer therapies revolving around drug delivery systems, and stopping cancer proliferation through telomerase stabilization.

Chapter VII

References

[1] R. Dahm, "Friedrich Miescher and the discovery of DNA," *Dev. Biol.*, vol. 278, no. 2, pp. 274–288, 2005.

[2] L. A. Pray, "Discovery of DNA Structure and Function: Watson and Crick," *Nat. Educ.*, vol. 1, no. 1, p. 6, 2008.

[3] M. Cobb, "Oswald Avery, DNA, and the transformation of biology," *Curr. Biol.*, vol. 24, no. 2, pp. R55–R60, 2014.

[4] R. Rudner, J. D. Karkas, and E. Chargaff, "Separation of microbial deoxyribonucleic acids into complementary strands.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 63, no. 1, pp. 152–9, 1969.

[5] M. Cobb, "1953: When genes became 'information,'" *Cell*, vol. 153, no. 3, pp. 503–506, 2013.

[6] F. H. C. Crick and J. D. Watson, "The Complementary Structure of Deoxyribonucleic Acid," *Proc. R. Soc. Math. Phys. Eng. Sci.*, vol. 223, no. 1152, pp. 80–96, 1954.

[7] A. Rich, "DNA comes in many forms," *Gene*, vol. 135, no. 1–2, pp. 99–109, 1993.

[8] T. M. Bryan and P. Baumann, "G-quadruplexes: From guanine gels to chemotherapeutics," *Mol. Biotechnol.*, vol. 49, no. 2, pp. 198–208, 2011.

[9] E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, and E. H. Blackburn, "Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs," *Cell*, vol. 51, no. 6, pp. 899–908, 1987.

[10] J. R. Williamson, M. K. Raghuraman, and T. R. Cech, "Monovalent cation-induced structure of telomeric DNA: The G-quartet model," *Cell*, vol. 59, no. 5, pp. 871–880, 1989.

[11] M. M. Fay, S. M. Lyons, and P. Ivanov, "RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms," *J. Mol. Biol.*, vol. 429, no. 14, pp. 2127–2147, 2017.

[12] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, and S. Neidle, "Quadruplex DNA: Sequence, topology and structure," *Nucleic Acids Res.*, vol. 34, no. 19, pp. 5402–5415, 2006.

[13] A. Tua, "Human telomeric G-quadruplex : structures of DNA and RNA sequences," vol. 277, pp. 1107–1117, 2010.

[14] J. L. Huppert, "Prevalence of quadruplexes in the human genome," *Nucleic Acids Res.*, vol. 33, no. 9, pp. 2908–2916, May 2005.

[15] J. L. Huppert, "Structure, location and interactions of G-quadruplexes: Structure, location and interactions of G-quadruplexes," *FEBS J.*, vol. 277, no. 17, pp. 3452–3458, Sep. 2010.

[16] C. K. Kwok and C. J. Merrick, "G-Quadruplexes: Prediction, Characterization, and Biological Application," *Trends Biotechnol.*, vol. xx, pp. 1–17, 2017.

[17] N. S. Ilyinsky, A. M. Varizhuk, A. D. Beniaminov, M. A. Puzanov, A. K. Shchyolkina, and D. N. Kaluzhny, "G-quadruplex ligands: Mechanisms of anticancer action and target binding," *Mol. Biol.*, vol. 48, no. 6, pp. 778–794, Nov. 2014.

[18] D. Rhodes and H. J. Lipps, "G-quadruplexes and their regulatory roles in biology," *Nucleic Acids Res.*, vol. 43, no. 18, pp. 8627–8637, Oct. 2015.

[19] Y. Wu and R. M. Brosh, "G-quadruplex nucleic acids and human disease," *FEBS J.*, vol. 277, no. 17, pp. 3470–3488, Sep. 2010.

[20] S. Cogoi and L. E. Xodo, "G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription," *Nucleic Acids Res.*, vol. 34, no. 9, pp. 2536–2549, Jan. 2006.

[21] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, and L. H. Hurley, "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.*, vol. 99, no. 18, pp. 11593–11598, Sep. 2002.

[22] S. Rankin *et al.*, "Putative DNA Quadruplex Formation within the Human c-kit Oncogene," *J. Am. Chem. Soc.*, vol. 127, no. 30, pp. 10584–10589, Aug. 2005.

[23] J. Dai *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.*, vol. 128, no. 4, pp. 1096–1098, Feb. 2006.

[24] S. Balasubramanian, L. H. Hurley, and S. Neidle, "Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?," *Nat. Rev. Drug Discov.*, vol. 10, no. 4, pp. 261–275, Apr. 2011.

[25] R.-J. Man, L.-W. Chen, and H.-L. Zhu, "Telomerase inhibitors: a patent review (2010–2015)," *Expert Opin. Ther. Pat.*, vol. 26, no. 6, pp. 679–688, Jun. 2016.

[26] G. Mirihana Arachchilage, A. C. Dassanayake, and S. Basu, "A Potassium Ion-Dependent RNA Structural Switch Regulates Human Pre-miRNA 92b Maturation," *Chem. Biol.*, vol. 22, no. 2, pp. 262–272, Feb. 2015.

[27] S. Pandey, P. Agarwala, G. G. Jayaraj, R. Gargallo, and S. Maiti, "The RNA Stem-Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell," *Biochemistry (Mosc.)*, vol. 54, no. 48, pp. 7067–7078, 2015.

[28] Q. Pan, F. Luo, M. Liu, and X.-L. Zhang, "Oligonucleotide aptamers: promising and powerful diagnostic and therapeutic tools for infectious diseases," *J. Infect.*, May 2018.

[29] P. J. Bates, E. M. Reyes-Reyes, M. T. Malik, E. M. Murphy, M. G. O'Toole, and J. O. Trent, "G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms," *Biochim. Biophys. Acta BBA - Gen. Subj.*, vol. 1861, no. 5, pp. 1414–1428, May 2017.

[30] T. Li, L. Shi, E. Wang, and S. Dong, "Multifunctional G-Quadruplex Aptamers and Their Application to Protein Detection," *Chem. - Eur. J.*, vol. 15, no. 4, pp. 1036–1042, Jan. 2009.

[31] R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe, and J. Feigon, "Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution.," *Proc. Natl. Acad. Sci.*, vol. 90, no. 8, pp. 3745–3749, Apr. 1993.

[32] W. O. Tucker, K. T. Shum, and J. A. Tanner, "G-quadruplex DNA Aptamers and their Ligands: Structure, Function and Application," *Curr. Pharm. Des.*, vol. 18, no. 14, pp. 2014–2026, Mar. 2012.

[33] B. Pagano, L. Martino, A. Randazzo, and C. Giancola, "Stability and Binding Properties of a Modified Thrombin Binding Aptamer," *Biophys. J.*, vol. 94, no. 2, pp. 562–569, Jan. 2008.

[34] E. Ruggiero and S. N. Richter, "G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy," *Nucleic Acids Res.*, vol. 46, no. 7, pp. 3270–3283, Apr. 2018.

[35] W. Gai *et al.*, "A dual-site simultaneous binding mode in the interaction between parallel-stranded G-quadruplex [d(TGGGGT)] 4 and cyanine dye 2,2′-diethyl-9-methyl-selenacarbocyanine bromide," *Nucleic Acids Res.*, vol. 41, no. 4, pp. 2709–2722, Feb. 2013.

[36] J. Cuesta, M. Read, and S. Neidle, "The Design of G-quadruplex Ligands as Telomerase Inhibitors," *Mini-Rev. Med. Chem.*, vol. 3, no. 1, pp. 11–21, Feb. 2003.

[37] D. Monchaud and M.-P. Teulade-Fichou, "A hitchhiker's guide to G-quadruplex ligands," *Org Biomol Chem*, vol. 6, no. 4, pp. 627–636, 2008.

[38] J. Carvalho, J. A. Queiroz, and C. Cruz, "Circular Dichroism of G-Quadruplex: a Laboratory Experiment for the Study of Topology and Ligand Binding," *J. Chem. Educ.*, vol. 94, no. 10, pp. 1547–1551, Oct. 2017.

[39] J. Kypr, I. Kejnovska, D. Renciuk, and M. Vorlickova, "Circular dichroism and conformational polymorphism of DNA," *Nucleic Acids Res.*, vol. 37, no. 6, pp. 1713–1725, Jan. 2009.

[40] A. Randazzo, G. P. Spada, and M. W. da Silva, "Circular Dichroism of Quadruplex Structures," in *Quadruplex Nucleic Acids*, vol. 330, J. B. Chaires and D. Graves, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 67–86.

[41] P. J. Stephens, "Circular Dichroism: Principles and Applications. Second Edition Edited by Nina Berova (Columbia University), Koji Nakanishi (Columbia University), and Robert W. Woody (Colorado State University). Wiley-VCH: New York. 2000. xxii + 878 pp. $195.00. ISBN: 0-471-33003-5.," *J. Am. Chem. Soc.*, vol. 124, no. 27, pp. 8182–8182, Jul. 2002.

[42] S. Paramasivan, I. Rujan, and P. H. Bolton, "Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding," *Methods*, vol. 43, no. 4, pp. 324–331, Dec. 2007.

[43] P. A. Rachwal and K. R. Fox, "Quadruplex melting," *Methods*, vol. 43, no. 4, pp. 291–301, Dec. 2007.

[44] B. Pagano *et al.*, "State-of-the-Art Methodologies for the Discovery and Characterization of DNA G-Quadruplex Binders," *Curr. Pharm. Des.*, vol. 18, no. 14, pp. 1880–1899, Mar. 2012.

[45] J. Jaumot and R. Gargallo, "Experimental Methods for Studying the Interactions between G-Quadruplex Structures and Ligands," *Curr. Pharm. Des.*, vol. 18, no. 14, pp. 1900–1916, Mar. 2012.

[46] C. Li *et al.*, "Magnetic Resonance Spectroscopy as a Tool for Assessing Macromolecular Structure and Function in Living Cells," *Annu. Rev. Anal. Chem.*, vol. 10, no. 1, pp. 157–182, Jun. 2017.

[47] J. Klein, R. Meinecke, M. Mayer, and B. Meyer, "Detecting Binding Affinity to Immobilized Receptor Proteins in Compound Libraries by HR-MAS STD NMR," *J. Am. Chem. Soc.*, vol. 121, no. 22, pp. 5336–5337, Jun. 1999.

[48] J. Ferreira *et al.*, "Naphthalene amine support for G-quadruplex isolation," *The Analyst*, vol. 142, no. 16, pp. 2982–2994, 2017.

[49] M. Urh, D. Simpson, and K. Zhao, "Chapter 26 Affinity Chromatography," in *Methods in Enzymology*, vol. 463, Elsevier, 2009, pp. 417–438.

[50] T. Chang, X. Liu, X. Cheng, C. Qi, H. Mei, and D. Shangguan, "Selective isolation of G-quadruplexes by affinity chromatography," *J. Chromatogr. A*, vol. 1246, pp. 62–68, Jul. 2012.

[51] D. Musumeci *et al.*, "G-Quadruplex on Oligo Affinity Support (G4-OAS): An Easy Affinity Chromatography-Based Assay for the Screening of G-Quadruplex Ligands," *Anal. Chem.*, vol. 86, no. 9, pp. 4126–4130, May 2014.

[52] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, "Principles of docking: An overview of search algorithms and a guide to scoring functions," *Proteins Struct. Funct. Genet.*, vol. 47, no. 4, pp. 409–443, Jun. 2002.

[53] A. M. Dar and S. Mir, "Molecular Docking: Approaches, Types, Applications and Basic Challenges," *J. Anal. Bioanal. Tech.*, vol. 08, no. 02, 2017.

[54] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nat. Rev. Drug Discov.*, vol. 3, no. 11, pp. 935–949, Nov. 2004.

[55] B. K. Shoichet, I. D. Kuntz, and D. L. Bodian, "Molecular docking using shape descriptors," *J. Comput. Chem.*, vol. 13, no. 3, pp. 380–397, Apr. 1992.

[56] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery," *Curr. Comput. Aided-Drug Des.*, vol. 7, no. 2, pp. 146–157, Jun. 2011.

[57] E. Pereira *et al.*, "Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017.

[58] C. Cruz, E. Cairrao, S. Silvestre, L. Breitenfeld, P. Almeida, and J. A. Queiroz, "Targeting of Mitochondria-Endoplasmic Reticulum by Fluorescent Macrocyclic Compounds," *PLoS ONE*, vol. 6, no. 11, p. e27078, Nov. 2011.

[59] C. Cruz, V. Calisto, R. Delgado, and V. Félix, "Design of Protonated Polyazamacrocycles Based on Phenanthroline Motifs for Selective Uptake of Aromatic Carboxylate Anions and Herbicides," *Chem. - Eur. J.*, vol. 15, no. 13, pp. 3277–3289, Mar. 2009.

[60] R. Kumari, R. Kumar, Open Source Drug Discovery Consortium, and A. Lynn, "g_mmpbsa —A GROMACS Tool for High-Throughput MM-PBSA Calculations," *J. Chem. Inf. Model.*, vol. 54, no. 7, pp. 1951–1962, Jul. 2014.

[61] W. Tan, L. Zhang, J. Zhou, H. Chen, and G. Yuan, "Formation of a G-quadruplex structure from human mature miR-5196-5p," *Spectrosc. Lett.*, vol. 50, no. 9, pp. 489–493, Oct. 2017.

[62] S. Forli and M. Botta, "Lennard-Jones Potential and Dummy Atom Settings to Overcome the AUTODOCK Limitation in Treating Flexible Ring Systems," *J. Chem. Inf. Model.*, vol. 47, no. 4, pp. 1481–1492, Jul. 2007.

[63] I. Kufareva and R. Abagyan, "Methods of Protein Structure Comparison," in *Homology Modeling*, vol. 857, A. J. W. Orry and R. Abagyan, Eds. Totowa, NJ: Humana Press, 2011, pp. 231–257.

[64] A. Khandelwal, V. Lukacova, D. Comez, D. M. Kroll, S. Raha, and S. Balaz, "A Combination of Docking, QM/MM Methods, and MD Simulation for Binding Affinity Estimation of Metalloprotein Ligands," *J. Med. Chem.*, vol. 48, no. 17, pp. 5437–5447, Aug. 2005.

[65] A. J. Stevens and M. A. Kennedy, "Structural Analysis of G-Quadruplex Formation at the Human MEST Promoter," *PLOS ONE*, vol. 12, no. 1, p. e0169433, Jan. 2017.

[66] J. Carvalho *et al.*, "Phenanthroline polyazamacrocycles as G-quadruplex DNA binders," *Org. Biomol. Chem.*, vol. 16, no. 15, pp. 2776–2786, 2018.

[67] S. Genheden and U. Ryde, "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities," *Expert Opin. Drug Discov.*, vol. 10, no. 5, pp. 449–461, May 2015.

[68] A. Weis, K. Katebzadeh, P. Söderhjelm, I. Nilsson, and U. Ryde, "Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field," *J. Med. Chem.*, vol. 49, no. 22, pp. 6596–6606, Nov. 2006.

[69] D. Ramírez and J. Caballero, "Is It Reliable to Use Common Molecular Docking Methods for Comparing the Binding Affinities of Enantiomer Pairs for Their Protein Target?," *Int. J. Mol. Sci.*, vol. 17, no. 4, p. 525, Apr. 2016.

[70] R. Warrass, J.-M. Wieruszeski, C. Boutillon, and G. Lippens, "High-Resolution Magic Angle Spinning NMR Study of Resin-Bound Polyalanine Peptides," *J. Am. Chem. Soc.*, vol. 122, no. 8, pp. 1789–1795, Mar. 2000.

[71] X. Zhao, B. Liu, J. Yan, Y. Yuan, L. An, and Y. Guan, "Structure variations of TBA G-quadruplex induced by 2′-O-methyl nucleotide in K+ and Ca2+ environments," *Acta Biochim. Biophys. Sin.*, vol. 46, no. 10, pp. 837–850, Oct. 2014.

[72] K. V. Diveshkumar, S. Sakrikar, F. Rosu, S. Harikrishna, V. Gabelica, and P. I. Pradeepkumar, "Specific Stabilization of c-MYC and c-KIT G-Quadruplex DNA Structures by Indolylmethyleneindanone Scaffolds," *Biochemistry (Mosc.)*, vol. 55, no. 25, pp. 3571–3585, Jun. 2016.

[73] C. K. Kwok, A. B. Sahakyan, and S. Balasubramanian, “Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA,” *Angew. Chem. Int. Ed.*, vol. 55, no. 31, pp. 8958–8961, Jul. 2016.

[74] P. K. Mandal, B. Baptiste, B. Langlois d’Estaintot, B. Kauffmann, and I. Huc, “Multivalent Interactions between an Aromatic Helical Foldamer and a DNA G-Quadruplex in the Solid State,” *ChemBioChem*, vol. 17, no. 20, pp. 1911–1914, Oct. 2016.