UNIVERSIDADE
BEIRA INTERIOR

Recognition of G-quadruplexes in microRNA precursors by nucleolin and its implications in cancer

Tiago André Afonso dos Santos

Tese para obtenção do Grau de Doutor em
Biomedicina
(3º ciclo de estudos)

Orientador: Prof.ª Doutora Carla Patrícia Alves Freire Madeira Cruz
Co-orientador: Prof. Doutor Eurico José da Silva Cabrita

Júri:
Prof. Doutor Ilídio Joaquim Sobreira Correia
Profª. Doutora Maria João Romão
Profª. Doutora Cândida Ascenção Teixeira Tomaz
Profª. Doutora Fani Pereira de Sousa
Profª. Doutora Maria Eduarda Romãozinho de Almeida Esteves Mendes
Profª. Doutora Alexandra da Silva Paulo
Profª. Doutora Carla Patrícia Alves Freire Madeira da Cruz
Dr. Aldino José Martins Viegas

outubro de 2022

Declaração de Integridade

Eu, Tiago André Afonso dos Santos, que abaixo assino, estudante com o número de inscrição D2258 do 3º ciclo em Biomedicina da Faculdade de Ciências da Saúde, declaro ter desenvolvido o presente trabalho e elaborado o presente texto em total consonância com o **Código de Integridades da Universidade da Beira Interior**.

Mais concretamente afirmo não ter incorrido em qualquer das variedades de Fraude Académica, e que aqui declaro conhecer, que em particular atendi à exigida referenciação de frases, extratos, imagens e outras formas de trabalho intelectual, e assumindo assim na íntegra as responsabilidades da autoria.

Universidade da Beira Interior, Covilhã 31/10/2022

Assinado por: **Tiago André Afonso dos Santos**
Num. de Identificação: 13786877
Data: 2022.10.31 06:46:28+00'00'

Acknowledgments

Esta Tese de Doutoramento é o culminar de um ciclo, não só de estudos, mas também pessoal, que envolveu muito trabalho, superação, resiliência, sacrifício e perseverança. Ao longo deste percurso enfrentei diversos obstáculos que contribuíram inequivocamente para o meu crescimento académico, profissional e pessoal. Esta conquista só é possível com o contributo de várias pessoas, às quais não posso deixar de expressar o meu sentido agradecimento.

Em primeiro lugar expresso o meu agradecimento à minha orientadora, a Professora Carla Cruz. Com plena consciência que foi um grande desafio, com diversos percalços e dificuldades, próprios das grandes conquistas. O seu acompanhamento no desenvolvimento desta Tese foi crucial, e por isso gostaria de lhe agradecer pela confiança, disponibilidade, compreensão e paciência. O que me transmitiu durante este percurso foi, não só a nível académico e profissional mas, também a nível pessoal, o meu profundo obrigado.

Ao Professor Eurico Cabrita, pelo seu apoio e preocupação, particularmente por partilhar comigo o seu espírito crítico e conhecimento científico. Obrigado por me estimular a encontrar soluções, mesmo quando elas pareciam escassear. O seu contributo e ajuda foram valiosos para que esta Tese fosse desenvolvida com sucesso, a si o meu sincero muito obrigado.

Ao Professor Gilmar Salgado, por me ter permitido alargar os meus horizontes científicos num país com uma cultura diferente. Durante a minha estadia no seu laboratório fez-me sentir parte integrante da sua equipa. A sua confiança e amizade foram fundamentais. Obrigado pela sua preocupação constante, disponibilidade em ajudar e por me estimular sempre a melhorar.

Gostaria também de agradecer à Universidade da Beira Interior, Universidade Nova de Lisboa e Universidade de Bordéus, em particular ao Centro de Investigação em Ciências da Saúde, Unidade de Ciências Biomoleculares Aplicadas – UCIBIO - FCT/UNL e Institut Européen de Chimie et Biologie por me facultarem as instalações e equipamentos necessários ao desenvolvimento da Tese.

A todas as pessoas com as quais contactei durante este percurso (técnicos, colegas, professores e investigadores). Aos colegas de laboratório aos quais agradeço a amizade,

apoio, paciência e colaboração em todos os momentos de trabalho. Não será justo especificar, a todos vocês o meu agradecimento. Aos colegas do programa doutoral PTNMR que me acompanharam inicialmente neste percurso, a vossa companhia e companheirismo foram também importantes. A todos vocês o meu agradecimento.

Aos meus amigos mais chegados, nem sempre estive tão presente como gostaria, no entanto, sei que todos vocês compreendem que o trabalho e a responsabilidade assim o exigem, concluo sabendo que também para vocês esta Tese é um motivo de orgulho.

À minha namorada e companheira, Sofia, pelo seu apoio incondicional. Obrigado pela tua dedicação, compreensão, cumplicidade e amor. Obrigado por estares a meu lado em todos os momentos, e por me incentivares incessantemente a seguir os meus objetivos. Esta conquista também é tua, sei que foste das pessoas que sofreu mais com os desaires mas, também das que vibrou efusivamente com as conquistas.

Um agradecimento muito especial e profundo à minha família pela força, amor e carinho. Aos meus pais por serem a minha muralha, os valores e princípios que me transmitiram foram fundamentais para atingir este objetivo. Ao meu irmão por ser inexcedível no seu apoio, por me fazer sempre acreditar e por partilhar comigo todos os momentos. À minha avó por ser um exemplo de força e lucidez. Aos meus tios e primos por estarem sempre presentes.

Por último gostaria de agradecer à FCT pelo financiamento (PD/BD/142851/2018).

Preface

Não tenhamos pressa, mas não percamos tempo.
José Saramago

Nucleolin is a multifunctional protein highly conserved in eukaryotes and involved in a series of important biological processes, including transcription, translation, synthesis and assembly of ribosomes, and microRNA (miRNA) biogenesis. Through its diverse biological functions, nucleolin is increasingly implicated in pathological processes, with more substantial relevance for its role in cancer, where it is overexpressed on the surface of cells. Furthermore, nucleolin has been shown to exhibit a remarkable binding preference towards G-quadruplex structures (G4s), a secondary architecture of nucleic acids. Upon binding to G4s, the structure and function of nucleolin can be influenced, and vice versa. As a result of the biological significance of the nucleolin/G4 partnership, its modulation for therapeutic and/or diagnostic purposes has been successfully achieved. Overall, the work presented in this thesis contributes to advance our knowledge of nucleolin/G4 partnership in cancer diagnostics and therapeutics, providing new tools and frameworks to deal with future research.

This thesis describes the work developed under the supervision of Dr. Carla Cruz and the co-supervision of Prof. Eurico Cabrita. Furthermore, national, and international collaborations, including the Health Sciences Research Centre (CICS-UBI, Portugal), Applied Molecular Biosciences Unit (UCIBIO-FCT-UNL, Portugal), European Institute of Chemistry and Biology (IECB, France) and Institute of Structural Biology (IBS, Grenoble), who have contributed significantly for understanding the structural and molecular features of nucleolin/G4 interactions.

In this thesis, we exploited the potential of miRNA precursors (pre-miRNAs) as G4-forming sequences, especially pre-miRNA 149, and their recognition by small compounds (G4 ligands) and nucleolin. Besides pre-miRNA 149, we also studied two other cancer-relevant pre-miRNAs that comprise G4 sequences (pre-miRNA 92b and let 7e), with distinct structural features.

The thesis is subdivided into three main sections: the aims and introduction section (**Chapters 1-4**), the experimental research section (**Chapters 5-10**), and the conclusions and future perspectives section (**Chapter 11**).

The first chapter focuses on the aims and outline of this thesis.

Chapter two presents a literature review of the current ligands used to stabilize G4s in non-coding RNAs. Some of those ligands are reported to affect the binding of nucleolin towards G4 structures, like structures found in pre-miRNAs let 7e, 92b, and 149.

Chapter three goes deeper into the principles and applications of biophysical methods that are currently available to explore G4/ligand interactions. This review discusses those molecular interactions in terms of their chemical nature.

Chapter four provides the state-of-the-art about the structure and function of nucleolin and its binding to G4 structures. In addition, current and future approaches for modulating nucleolin/G4 interaction are also discussed.

Chapter five unveils the formation of a G4 by a sequence in pre-miRNA let-7e and its targeting with ligands and nucleolin. We also discuss how this strategy could open new opportunities to control the miRNA biogenesis of let 7e miRNA.

Chapter six discloses the molecular interaction of the G4 sequence identified in pre-miRNA-92b with acridine orange derivatives. This chapter also shows the ability to use the G4 as a recognition agent for nucleolin in complex samples.

Chapter seven presents the binding modes of the predicted G4 in pre-miRNA 149 with several well-known ligands through molecular docking and molecular modelling approaches.

Chapter eight evaluates the binding and stabilization of the G4 of pre-miRNA 149 by acridine derivative ligands, C_8 and C_8-NH_2, and displays the potential of the complex G4/C_8 to target nucleolin on the surface of prostate cancer cells.

Chapter nine describes the ability of the full-length pre-miRNA 149 sequence to adopt a G4 structure, bind to G4 ligands, and capture nucleolin on a microfluidic device.

Chapter ten unveils the contacts at the atomic level between the G4 of pre-miRNA 149 and the acridine orange derivative, C_8. In addition, the amino acid residues of nucleolin involved in the interaction with the G4 and G4/C_8 complex are shown and discussed.

Lastly, **chapter eleven** presents the concluding remarks of this thesis as well as their impact and future perspectives in the context of nucleolin/G4-related pathologies, namely cancer, and their treatment and diagnosis.

Resumo

O cancro encontra-se entre as patologias mais prevalentes, sendo responsável por elevadas taxas de mortalidade, que tendem a aumentar de ano para ano. De modo a restituir o bem-estar dos cidadãos que padecem desta patologia, têm sido efetuados inúmeros esforços no desenvolvimento de novas abordagens de diagnóstico e terapêutica. No entanto, de modo a assegurar a eficiência e eficácia destas novas abordagens é necessário obter maior conhecimento sobre os mecanismos e interações moleculares que ocorrem durante o desenvolvimento de cancro.

Nas últimas décadas foram identificados e estudados diversos mecanismos moleculares relacionados com o desenvolvimento de diversos cancros. O seu estudo permitiu identificar novas estruturas de ácidos nucleicos, as quais podem adotar conformações secundárias que têm implicações no cancro. Entre estas estruturas encontram-se os G-quadruplex (G4), estruturas tridimensionais, que se formam em regiões do genoma ou transcriptoma ricas em guaninas. Os DNA G4 localizam-se em regiões fundamentais do genoma, as quais incluem as extremidades dos telómeros e os promotores oncogénicos, e receberam desde há duas décadas considerável atenção por parte da comunidade científica. Por outro lado, os RNA G4 que se encontram em regiões não codificantes, apenas recentemente captaram esse interesse, devido à sua crescente importância na regulação de diversos processos biológicos. Neste sentido, os intermediários da biogénese dos microRNAs (miRNAs) com capacidade para adotar G4 têm sido estudados nos últimos anos, com especial destaque para os precursores de microRNAs (pre-miRNAs).

O desenvolvimento e/ou estudo de pequenas moléculas (ligandos de G4, peso molecular < 500 Da) com capacidade para se ligar e estabilizar/desestabilizar as estruturas G4 pode, inequivocamente, contribuir para o controlo da biogénese dos miRNAs. Os ligandos de G4 têm demonstrado elevado potencial para ser aplicados ao diagnóstico ou à terapia de doenças neurodegenerativas, cancro e infeções virais. Os derivados de laranja de acridina têm sido amplamente testados em ligar e estabilizar diferentes DNA e RNA G4s. Além das pequenas moléculas que se ligam às estruturas G4, existem outras abordagens promissoras para controlar a biogénese dos miRNAs. O facto de ser um processo, essencialmente, mediado por proteínas revela o seu potencial para ser regulado por esta via. A abordagem mais utilizada visa influenciar a atividade da Dicer, uma

proteína com atividade enzimática que cliva o pre-miRNA em miRNA. No entanto, existem outras abordagens, igualmente relevantes, que podem ser aplicadas.

A nucleolina é uma proteína com elevada expressão em células cancerígenas. A proteína está envolvida em diversos processos celulares entre os quais se destacam a tumorigénese, angiogénese e vias de sinalização extracelular. A proteína localiza-se principalmente no nucléolo, no entanto pode também localizar-se no nucleoplasma, citoplasma e superfície celular. Em células saudáveis, a expressão da nucleolina ocorre, maioritariamente, na sua forma nuclear, enquanto em células cancerígenas ocorre de forma exacerbada na superfície da célula. Recentemente, a localização da nucleolina foi associada à biogénese dos miRNAs, através da sua interação com o complexo microprocessador. No entanto, uma vez que é uma proteína com elevada afinidade para estruturas G4 paralelas, e que em células cancerígenas efetua o transporte de diversas moléculas entre o citoplasma e o núcleo, a sua ação nesta fase da biogénese é uma hipótese plausível.

Assim, esta Tese visa numa primeira fase abordar a interação das sequências G4 presentes nos pre-miRNA let 7e, 92b e 149, com ligandos e a nucleolina. A desregulação nos níveis de expressão destes pre-miRNAs e miRNAs tem sido associada a diversos tipos de cancro. Além disso, as implicações biológicas e potenciais aplicações do reconhecimento destas sequências pelos ligandos e a nucleolina foram abordadas.

A formação de estruturas G4 encontradas nas sequências dos pre-miRNAs let 7e, 92b e 149 foi avaliada em diferentes condições experimentais (concentração, força iónica e temperatura). O efeito estabilizador/desestabilizador dos ligandos foi estudado e difere consoante a estrutura do G4 e as condições experimentais. O ligando C_8 apesar de ter um efeito modesto na estabilização do G4 do pre-miRNA let 7e, demonstrou elevada capacidade para estabilizar os G4s dos pre-miRNAs 92b e 149.

Uma vez estabelecido o potencial dos ligandos G4 para estabilizar os G4s presentes nos miRNAs, procedeu-se à avaliação da interação G4/nucleolina na presença e ausência de ligandos. No caso particular da estrutura G4 do pre-miRNA let 7e, a formação do complexo ternário G4/ligando/nucleolina foi observada, exceto na presença dos ligandos PhenDC3 e TMPyP4, os quais parecem desestabilizar a formação de G4. Deste modo, estes dois ligandos apresentam potencial para controlar a biogénese do miRNA let 7e, aumentando os seus níveis de expressão. Na presença do ligando C_8, o reconhecimento da nucleolina e consequente formação do complexo ternário não foram afetados. Nos estudos realizados com a estrutura G4 do pre-miRNA 92b, os derivados de laranja de

acridina demonstraram um elevado potencial na estabilização da estrutura, nomeadamente o C_8. Além disso, o ligando não afeta significativamente o reconhecimento da nucleolina pelo G4. Desta forma, a sequência G4 complexada com o ligando C_8 foi testada num dispositivo microfluídico e detetou a nucleolina em amostras de plasma de doentes com cancro da próstata.

Apesar da sua enorme relevância em diversos processos e mecanismos associados ao cancro, a sequência G4 do pre-miRNA 149 tem sido menos estudada que as anteriormente descritas. Além disso, sobrepõe-se quase na totalidade com a sequência do miRNA 149-3p, o que apresenta uma oportunidade para atuar em diferentes etapas da biogénese do miRNA 149. Por estas razões, a sequência G4 do pre-miRNA 149 foi estudada em termos da sua interação com ligandos, nomeadamente os derivados de laranja de acridina e a nucleolina. Além disso, devido ao seu elevado potencial, a estrutura G4 foi testada como potencial estratégia de reconhecimento e deteção da nucleolina na superfície de células cancerígenas. A formação da estrutura G4 foi confirmada através de diversos métodos biofísicos, e revelou uma estrutura G4 de elevada complexidade estrutural. Posteriormente, o modo de ligação e interação dos ligandos com a estrutura G4 foram analisados e os resultados demonstraram um elevado potencial do ligando C_8 para estabilizar a estrutura G4. A sequência completa do pre-miRNA 149 permitiu detetar a nucleolina através de um sistema microfluídico fabricado pelo INESC-MN. Por fim, a interação do G4 e do complexo G4/C_8 com a nucleolina foram também investigados e revelaram o local de ligação na estrutura 3D dos domínios 1 e 2 da nucleolina.

No geral, estes resultados revelaram o potencial de aplicação biológica das sequências G4 nos pre-miRNA let 7e, 92b e 149. A regulação estrutural das G4 presentes nos pre-miRNAs pode possibilitar o desenvolvimento de aplicações no diagnóstico ou terapia do cancro. No futuro, é expectável o surgimento de novos estudos estruturais da interação pre-miRNA G4/nucleolina e de desenvolvimento de novos ligandos com efeito inibidor ou facilitador da interação.

Palavras-chave

Cancro;G-quadruplex;nucleolina;ligandos de G-quadruplex;precursores de microRNAs

Resumo alargado

Atualmente, a saúde é um dos mais poderosos fatores socioeconómicos, sendo que o seu objetivo global é proporcionar bem-estar aos cidadãos. Um dos maiores desafios reside no diagnóstico e tratamento de patologias, como é o caso do cancro. O cancro encontra-se entre as patologias mais prevalentes, sendo responsável por elevadas taxas de mortalidade, que tendem a aumentar de ano para ano. De modo a restituir o bem-estar dos cidadãos que padecem desta patologia, têm sido efetuados inúmeros esforços no desenvolvimento de novas abordagens de diagnóstico e terapêuticas. No entanto, de modo a assegurar a eficácia destas novas abordagens é necessário obter maior conhecimento sobre os mecanismos e interações moleculares que ocorrem durante o desenvolvimento de cancro.

Nas últimas décadas foram identificados e estudados diversos mecanismos moleculares relacionados com o desenvolvimento de diversos cancros. O seu estudo permitiu identificar novas estruturas de ácidos nucleicos, as quais podem adotar conformações secundárias. Entre estas estruturas encontram-se os G-quadruplex (G4), estruturas tridimensionais, que se formam em regiões do genoma ou transcriptoma ricas em guaninas. Quando quatro guaninas estabelecem ligações de hidrogénio nos pares Hoogsteen entre si originam uma estrutura planar designada por tétrada. Sempre que duas ou mais tétradas se empilham por interações π-π, originam uma estrutura G4. Além disso, os G4s são estabilizados por catiões como o potássio (K^+) e o sódio (Na^+). Os DNA G4 localizam-se em regiões fundamentais do genoma, as quais incluem as extremidades dos telómeros e os promotores oncogénicos, receberam desde há duas décadas considerável atenção por parte da comunidade científica. Por outro lado, os RNA G4 que se encontram em regiões não codificantes, apenas recentemente captaram esse interesse, devido à sua crescente importância na regulação de diversos processos biológicos. Neste sentido, os intermediários da biogénese dos microRNAs (miRNAs) com capacidade para adotar G4 têm sido estudados nos últimos anos, com especial destaque para os precursores de microRNAs (pre-miRNAs).

Os miRNAs são pequenas moléculas de RNA (18-22 nucleótidos) não codificantes envolvidas no controlo da expressão génica a um nível pós-transcricional. A sua biogénese consiste numa série de reações enzimáticas e intermediários, entre os quais se incluem os pre-miRNAs. Em alguns pre-miRNAs é possível encontrar sequências ricas em guaninas com capacidade para formar G4s, como é o caso dos pre-miRNA let 7e, 92b

e 149. Além desta particularidade, estes pre-miRNAs partilham o facto de serem anormalmente expressos em diversos tipos de cancro.

A família dos miRNA let 7 é reconhecida por controlar diversos mecanismos moleculares importantes ao normal funcionamento das células. A baixa expressão, entre os quais se inclui o miRNA let 7e, está associada à proliferação e invasão das células cancerígenas. A expressão desregulada do miRNA 92b tem sido extensivamente estudada no cancro do pulmão de não pequenas células, e está relacionada com diversos mecanismos oncogénicos e com a resistência à terapia. O miRNA 149 está envolvido em diversos aspetos da oncogénese, entre os quais se incluem a progressão do tumor e a sua resistência à terapia, no entanto tem sido menos estudado que os miRNAs anteriormente mencionados. Além disso, as isoformas do miRNA 149 conferem-lhe um impacto duplo, podendo atuar como supressor tumoral ou oncogene, dependendo de fatores como, por exemplo, o tipo ou estadio do cancro. Possui também a particularidade e vantagem de a sequência G4 presente no pre-miRNA 149 se sobrepor, quase na totalidade, com a sequência do miRNA 149-3p, o que permite desenvolver potenciais estratégias terapêuticas que atuam em diferentes etapas da biogénese. Assim, a caraterização estrutural, molecular e celular dos G4 presentes nestas regiões é de fundamental importância para desenvolver novas estratégias de diagnóstico e terapia do cancro.

O desenvolvimento e/ou estudo de pequenas moléculas (ligandos de G4, peso molecular < 500 Da) com capacidade para se ligar e estabilizar/desestabilizar as estruturas G4 pode, inequivocamente, contribuir para o controlo da biogénese dos miRNAs. Os ligandos de G4 têm demonstrado elevado potencial, tanto no diagnóstico como na terapia de doenças neurodegenerativas, cancro e infeções virais. Os derivados de laranja de acridina detêm um potencial ímpar para ser aplicados ao diagnóstico e terapia, e por essa razão têm sido amplamente testados em ligar e estabilizar diferentes DNA e RNA G4s. Além das pequenas moléculas que se ligam às estruturas G4, existem outras abordagens promissoras para controlar a biogénese dos miRNAs. O facto de ser um processo, essencialmente, mediado por proteínas revela o seu potencial para ser regulado por esta via. A abordagem mais utilizada visa influenciar a atividade da Dicer, uma proteína com atividade enzimática que cliva o pre-miRNA em miRNA. A utilização de ligandos direcionados para os pre-miRNAs G4 pode impedir ou favorecer a atividade da Dicer através da modulação da estrutura secundária dos pre-miRNAs, controlando desta forma os níveis de miRNAs. No entanto, existem outras abordagens, igualmente relevantes, que podem ser aplicadas, nomeadamente ao nível da conversão de precursores primários de miRNAs (pri-miRNAs) em pre-miRNAs pelo complexo

microprocessador formado pela Drosha (endonuclease da família da RNase III) e pela DGCR8 (do inglês, *DiGeorge syndrome critical region gene 8*).

A nucleolina é uma proteína multifuncional com elevada expressão em células cancerígenas. A proteína está envolvida em diversos processos celulares entre os quais se destacam a tumorigénese, angiogénese e vias de sinalização extracelular. A proteína localiza-se principalmente no nucléolo, no entanto pode também localizar-se no nucleoplasma, citoplasma e superfície celular. No nucléolo, a proteína está envolvida no metabolismo do DNA e RNA; no citoplasma participa no transporte de moléculas de e para o núcleo, controlando deste modo a regulação pós-transcricional; e na superfície da célula funciona como recetor de diversos ligandos extracelulares. Em células saudáveis, a expressão da nucleolina ocorre, maioritariamente, na sua forma nuclear, enquanto em células cancerígenas ocorre de forma exacerbada na superfície da célula. Recentemente, a localização da nucleolina foi associada à biogénese dos miRNAs, através da sua interação com o complexo microprocessador. No entanto, uma vez que é uma proteína com elevada afinidade para estruturas G4 paralelas, e que em células cancerígenas efetua o transporte de diversas moléculas entre o citoplasma e o núcleo, a sua ação nesta fase da biogénese é uma hipótese plausível. Assim, uma estratégia que englobe a utilização de ligandos de G4 e a nucleolina poderá ser vantajosa para o desenvolvimento de novas estratégias de diagnóstico e terapia do cancro. Deste modo, esta Tese visa abordar a interação das sequências G4 presentes nos pre-miRNA let 7e, 92b e 149, com ligandos e a nucleolina.

Numa primeira fase, foi realizada revisão da literatura que culminou na introdução desta Tese. No capítulo 2, identificaram-se os ligandos de G4 mencionados na literatura que interagem com RNA não codificantes. De seguida, no capítulo 3 foram abordadas as técnicas biofísicas utilizadas para caraterizar estruturalmente a interação G4/ligandos, bem como as suas vantagens e desvantagens. De notar que a maioria das técnicas descritas neste capítulo podem ser aplicadas também à interação G4/proteína. Por fim, no capítulo 4 foram escrutinadas a estrutura e funções da nucleolina, bem como os efeitos biológicos da sua interação preferencial com estruturas G4.

Tendo-se identificado as sequências G4 dos pre-miRNAs let 7e, 92b e 149 como relevantes na biologia do cancro, iniciou-se a sua caraterização, modulação e aplicação. Numa primeira abordagem foi estudada a sequência G4 localizada no pre-miRNA let-7e (5'-GGGCUGAGGUAGGAGG-3'), a qual possui capacidade para formar um G4 composto por duas tétradas. Após abordagem inicial para avaliar a formação de G4, oito ligandos de G4 foram testados quanto à sua capacidade para se ligar e estabilizar/destabilizar a

estrutura do G4. Os ligandos PDS e 360A demonstraram efeitos promissores na estabilização da estrutura. Adicionalmente, a interação do G4 com a nucleolina foi também avaliada na presença e ausência dos ligandos. A formação do complexo ternário G4/ligando/nucleolina foi observada, exceto na presença dos ligandos derivados de fenantrolina e porfirina, PhenDC3 e TMPyP4, respetivamente. No geral, este estudo demonstrou o potencial dos ligandos para regular a formação de G4 no pre-miRNA let-7e e a sua interação com a nucleolina. Além disso, estas evidências revelam novas oportunidade de regulação da biogénese dos miRNAs.

De seguida procedeu-se à caraterização e avaliação da sequência G4 localizada no pre-miRNA 92b (5'-GGGCGGGCGGGAGGG-3'). Esta estrutura apresenta três tétradas e *loops* com apenas um nucleótido, similar à sequência localizada no pre-miRNA 149 (5'-GGGAGGGAGGGACGGG-3'). Tendencialmente os G4s constituídos por *loops* com menor número de nucleótidos apresentam maior estabilidade térmica e resistência às nucleases, uma vez que são estruturalmente mais compactos. Este G4 foi utilizado para reconhecimento da nucleolina em amostras complexas, pelo que a sua estabilidade e resistência a nucleases foi crucial para o sucesso da aplicação. De notar que o ligando de laranja de acridina, C_8, demonstrou potencial para estabilizar as estruturas dos G4 e aumentar a resistência às nucleases, sem afetar o reconhecimento da nucleolina. A sequência G4 foi posteriormente imobilizada num sistema microfluídico com o objetivo de detetar a nucleolina em amostras de plasma de doentes com cancro da próstata.

Os resultados encorajadores com as sequências G4 dos pre-miRNA let 7e e 92b estimularam-nos a avançar para a caraterização da sequência G4 do pre-miRNA 149. Numa primeira fase foi efetuado um estudo *in silico*, no qual se construiu um modelo 3D da estrutura G4 do pre-miRNA 149. O modo de ligação e interação de seis ligandos diferentes com a estrutura G4 foram analisados, através de simulações dinâmicas moleculares. Além disso, foram também calculados os parâmetros termodinâmicos que regem a interação de cada complexo G4/ligando. Esta estratégia permitiu concluir que os ligandos com grupos funcionais aromáticos e cadeias laterais carregadas positivamente podem interagir com a estrutura G4 através de um modo misto, através de interações por empilhamento π-π, electroestáticas e pontes de hidrogénio. Este estudo *in silico* revelou que o ligando C_8 interage com o G4, maioritariamente, através de interações por empilhamento π-π.

Uma vez previstos os principais locais e modos de ligação, avaliaram-se os efeitos biológicos da formação do complexo G4 do pre-miRNA 149/C_8/nucleolina. De modo notável, o ligando melhorou a estabilidade térmica e de resistência às nucleases do G4.

Além disso, o ligando detém uma elevada afinidade pelo G4, promovendo o reconhecimento da nucleolina. O G4 foi também utilizado numa estratégia simples de conjugação não covalente com o ligando C_8 para a sua entrega direcionada à nucleolina expressa na superfície de células cancerígenas da próstata. Após ter sido demonstrado o potencial do complexo G4/C_8, foi investigada a formação da estrutura G4 na sequência completa do pre-miRNA 149, e avaliada a sua capacidade de deteção da nucleolina num sistema microfluídico. Os resultados demonstraram que em condições semelhantes ao ambiente fisiológico, a sequência completa do pre-miRNA 149 reconhece a nucleolina.

Por fim, partindo da premissa de que estrutura e função biológica se encontram intrinsecamente conectadas, procedeu-se à caraterização estrutural da interação da estrutura G4 do pre-miRNA 149 com o ligando C_8 e com a nucleolina, através da espectroscopia de RMN. Este estudo permitiu observar os contactos estabelecidos entre a estrutura G4 do pre-miRNA 149, o ligando C_8, e a nucleolina.

A otimização de diversas técnicas biofísicas complementares durante esta tese de doutoramento permitiu a caraterização detalhada das sequências G4 localizadas nos pre-miRNAs let 7e, 92b e 149. No entanto, devido às caraterísticas peculiares do pre-miRNA 149 e dos miRNA 149-5p e miRNA 149-3p, e ao facto estarem menos estudados, o foco deste trabalho incidiu maioritariamente na sua avaliação e possíveis aplicações no diagnóstico e terapia do cancro. A caraterização estrutural dos G4 também possibilitou a utilização de novas abordagens para controlar a biogénese dos miRNAs através da utilização de ligandos. Além disso, os resultados indicam elevada afinidade da nucleolina pelas estruturas G4 localizadas nos pre-miRNAs. Ainda assim, é necessário continuar a investigar o envolvimento da nucleolina na biogénese dos miRNAs. Outro fator que poderá revelar-se fundamental, num futuro próximo, é a determinação estrutural da nucleolina com todos os seus domínios. A previsão da estrutura 3D foi recentemente calculada pelo AlphaFold, um algoritmo inovador de inteligência artificial que possibilita a previsão de estruturas de proteínas complexas. Este progresso, pode desencadear novos estudos estruturais (Cristalografia de raios-X, Ressonância Magnética Nuclear e Crio-Microscopia Eletrónica) dos complexos pre-miRNA G4/nucleolina e à emergência de novos ligandos que podem ter um efeito inibidor ou favorecer a interação.

Palavras-chave

Cancro;G-quadruplex;nucleolina;ligandos de G-quadruplex;precursores de microRNAs

Abstract

Cancer is among the most frequent pathologies, and it is responsible for high mortality rates, which tend to increase year by year. In order to restore the well-being of citizens, numerous efforts have been performed to develop new diagnostic and therapeutic approaches. However, ensuring the effectiveness of these new approaches is a challenging task. Indeed, it is necessary to study the mechanisms and molecular interactions during cancer development.

Several molecular mechanisms related to cancer development have been studied in the past decades. Their study encouraged the identification of new nucleic acid structures, which can adopt alternative secondary conformations. Among these structures are G-quadruplexes (G4), three-dimensional structures formed in the genome or transcriptome of regions rich in guanines. G4 DNAs are located in key regions of the genome, including the ends of telomeres and oncogenic promoters, and have received considerable attention from the scientific community in the past two decades. On the other hand, the G4 RNAs found in non-coding regions have recently received this interest due to their importance in controlling multiple biological processes. In this way, the intermediates of microRNA biogenesis (miRNAs) that can adopt a G4 structure have been more intensely studied, focusing on microRNA precursors (pre-miRNAs).

The development and/or study of small molecules (G4 ligands, molecular weight < 500 Da) with the ability to bind and stabilize/destabilize G4 structures could strongly contribute to the modulation of miRNA biogenesis. G4 ligands have shown high potential to be applied to the diagnosis and therapy of several pathologies such as neurodegenerative diseases, cancer, and viral infections. Acridine orange derivatives have been demonstrated to bind and stabilize different DNA and RNA G4 structures. Besides small molecules that bind to G4 structures, there are other approaches to control the biogenesis of miRNAs, which are also attractive. Indeed, the process is essentially modulated by proteins, which reveals its potential to be modulated by interfering with those molecules. In this way, the most widely used approach aims to influence Dicer activity, a protein with enzymatic activity that cleaves pre-miRNAs into miRNAs. However, other equally relevant approaches could be explored.

Nucleolin is a protein highly expressed in cancer cells. The protein is involved in several cellular processes, including tumorigenesis, angiogenesis, and extracellular signaling pathways. Nucleolin is mainly located in the nucleolus but can also be found in the

nucleoplasm, cytoplasm, and cell surface. Recently, it has been described that nucleolin is involved in miRNA biogenesis through its interaction with the microprocessor complex. However, since has been described as a protein with a high affinity for parallel G4 structures, and in cancer cells performs the transport of several molecules between the cytoplasm and the nucleus, its action at this stage of biogenesis is a robust hypothesis.

In this way, this thesis aims to deal with the interaction of the G4s adopted by pre-miRNA let 7e, 92b, and 149, with ligands and nucleolin. Indeed, the expression levels of these pre-miRNAs and miRNAs have been found dysregulated in several types of cancer. In addition, the potential biological applications of the recognition of these sequences by ligands and nucleolin were addressed.

The formation of G4s in the sequences of the pre-miRNA let 7e, 92b, and 149 were evaluated under different experimental conditions (concentration, ionic strength, and temperature). In addition, the stabilizing/destabilizing effect of the ligands on the G4s was evaluated and they are dependent on the experimental conditions. For example, the ligand $C_{8,}$ depicts a modest impact in stabilizing the G4 of pre-miRNA let 7e, but stabilized the G4s of pre-miRNAs 92b and 149.

Once we established the potential of G4 ligands to stabilize the G4s present in miRNAs, we evaluated the G4/nucleolin molecular interactions in the presence and absence of ligands. In the case of the G4 in pre-miRNA let 7e, the formation of the ternary complex G4/ligand/nucleolin was observed, except in the presence of the ligands PhenDC3 and TMPyP4, which seem to destabilize the G4 structure. Thus, these two ligands have the potential to control the miRNA biogenesis by increasing miRNA let 7e expression levels. The formation of the ternary complex was not affected by the presence of C_8. In the studies performed with the G4 structure of pre-miRNA 92b, the acridine derivatives showed high potential to stabilize the structure, namely C_8. Also, this ligand does not significantly affect the recognition of nucleolin by G4. Therefore, the G4/C_8 complex was tested in a microfluidic device to sense nucleolin in plasma samples from prostate cancer patients.

Finally, even known for its enormous relevance in several cancer-associated processes and mechanisms, the G4 sequence of pre-miRNA 149 has been less studied than those previously described. Furthermore, the G4 sequence overlaps almost entirely with the miRNA 149-3p, which presents an opportunity to act at different stages of miRNA 149 biogenesis. Considering this evidence, we aim to study in-depth the G4 sequence of pre-miRNA 149 in terms of its interaction with ligands and nucleolin. Furthermore, due to

its high potential, the G4 structure was tested as a potential strategy to recognize and detect nucleolin on the surface of cancer cells. First, the formation of the G4 structure was confirmed by several biophysical techniques and revealed a G4 structure of high structural complexity. Then, the binding mode and interaction of different G4 ligands with the G4 structure were analyzed, and the results showed a high potential of the ligand C_8 to stabilize the G4 structure. We also analyzed the complete sequence of pre-miRNA 149 and demonstrated its ability to detect nucleolin using a microfluidic system manufactured by INESC-MN. Finally, the interaction of G4 and the G4/C_8 complex with nucleolin was also investigated and revealed a binding pocket in the 3D structure of nucleolin domains 1 and 2.

Overall, these results revealed the biological potential of the G4 sequences in pre-miRNA let 7e, 92b, and 149. The structural regulation of G4 sequences present in pre-miRNAs can enable the development of applications for cancer diagnosis and therapy. Therefore, we anticipated that structural studies of the G4/nucleolin interaction would be available in the future, leading to the emergence of new ligands with inhibitory or enhancing effects on the interaction.

Keywords

Cancer;G-quadruplex;nucleolin;G-quadruplex ligands;microRNA precursors

Table of contents

Section I 1

Chapter 1 3
Aims and outline of the thesis 5

Chapter 2 9
Chapter overview 9
Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs 11
Abstract 11
Keywords 11
Introduction 11
rG4s in Telomere Long ncRNAs 15
rG4s in pri-miRNAs 17
rG4s in pre-miRNAs 18
rG4s in miRNAs 22
Conclusions 26
Funding 27
Acknowledgments 27
References 28

Chapter 3 33
Chapter overview 33
G-quadruplexes and their ligands: Biophysical methods to unravel G-quadruplex/ligand interactions 35
Abstract 35
Keywords 35
Introduction 35
Overview of G4-Interacting Ligands 40
DNA G4-Interacting Ligands 41
RNA G4-Interacting Ligands 45
Methods to Characterize G4/Ligand Interactions **49**
Structure-Based Methods to Investigate G4/Ligand Interactions 50
Circular Dichroism (CD) 50
Nuclear Magnetic Resonance (NMR) 53
X-ray Crystallography 57

Affinity- and Apparent Affinity-Based Methods to Investigate G4/Ligand Interactions 60
Surface Plasmon Resonance (SPR) 60
Isothermal Titration Calorimetry (ITC) 62
Mass Spectrometry (MS) 63
High-Throughput Methods to Investigate G4/Ligand Interactions 66
FRET-Melting 66
G4-FID Screening 69
Affinity Chromatography Screening 71
Microarrays-Based Screening 73
Conclusions 74
Funding 77
References 78

Chapter 4 93
Chapter overview 93
Nucleolin: A binding partner of G-quadruplex structures 95
Abstract 95
Keywords 95
Nucleolin: relationship between structure and functional role 95
Nucleolin/G-quadruplex partnership 96
Targeting of nucleolin/G-quadruplex partnership 98
Concluding remarks 100
Acknowledgments 101
References 101

Section II 103

Chapter 5 105
Chapter overview 105
Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin 107
Abstract 107
Keywords 107
Introduction 108
Materials and Methods 109
Oligonucleotides and ligands 109
Cloning, cell-free and purification of NCL RBD1,2 110

UV Absorption spectroscopy 111
Circular dichroism spectroscopy 111
NMR spectroscopy 112
Fluorescence Resonance Energy Transfer (FRET) Melting 112
Surface Plasmon Resonance (SPR) Biosensor 113
Non-denaturing polyacrylamide gel electrophoresis 113
Confocal microscopy 114
Results 114
Putative rG4 sequence in let-7e pre-miRNA adopt a G4 structure in presence of K^+ 114
Stabilization of rG4-let-7e by G4 ligands 116
Binding affinity of G4 ligands towards rG4-let-7e 118
Molecularity and structural nuances of rG4-let-7e in the presence of G4 ligands 119
Binding of NCL RBD1,2 to rG4-let-7e structure 120
Retention of rG4 structure in presence of ligands and ligands/NCL RBD1,2 122
Targeting of cell surface NCL by rG4-let-7e and their complexes with ligands 123
Discussion 124
Conclusions 127
Acknowledgments 127
References 128
Supplementary Information 133

Chapter 6 143
Chapter overview 143
Recognition of nucleolin through interaction with RNA G-quadruplex 145
Abstract 145
Keywords 145
Introduction 146
Materials and Methods 147
Oligonucleotides and ligands 147
Circular dichroism 148
NMR spectroscopy 149
Non-denaturing polyacrylamide gel electrophoresis 149
Fluorescence Resonance Energy Transfer (FRET)-melting 149
pre-miR-92b G4 model preparation 150
Molecular dynamics 150

Fluorescence titrations 151
Nuclease stability assays 152
Fabrication of microfluidic device 152
Microfluidic structure beads packing 153
Microfluidic experiments and image acquisition 153
Results 154
rG4 formation and stabilization in the presence of K^+ ions 154
Acridine orange derivatives stabilize pre-miR-92b rG4 155
Structural insights into pre-miR-92b rG4 structure by NMR spectroscopy studies 156
Molecularity of pre-miR-92b rG4 and pre-miR-92b rG4/ligands complex by non-denaturing PAGE 158
Structural prediction by molecular docking and MD simulations 159
pre-miR-92b rG4/NCL RBD1,2 complex evaluation by FRET-melting assay 161
Determination of binding affinity by fluorescence titrations 162
Stability studies 163
NCL detection by microfluidic assays 165
Discussion 166
Acknowledgments 169
References 170

Chapter 7 173
Chapter overview 173
Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics 175
Abstract 175
Keywords 175
Introduction 175
Methods 177
pre-miR-149 G4 model preparation 177
Molecular docking 178
Molecular dynamics 179
Thermodynamics calculations 179
Results and discussion 180
pre-miRNA 149 rG4 model structural nuances 180
Molecular docking of the set of chosen ligands 182
Molecular Dynamics simulations 184

Solvent accessible surface area determination 188
Free energy of binding calculations (MM/PBSA) 189
Conclusions 191
Acknowledgments and Funding 192
References 193
Supplementary Information 197

Chapter 8 205
Chapter overview 205
RNA G-quadruplex as supramolecular carrier for cancer-selective delivery 207
Abstract 207
Keywords 207
Introduction 208
Materials and methods 209
Oligonucleotides and ligands 209
Thermal difference spectroscopy (TDS) experiments.....210
Circular dichroism spectroscopy.....210
Fluorescence spectroscopy titrations 211
Viability/Proliferation assays..... 211
Nuclease stability assay.....212
Confocal microscopy studies 212
Results and discussion 213
Binding and stabilization of the rG4s.....213
Fluorescence binding studies with rG4 sequences 214
Fluorescence binding studies with NCL.....215
Cytotoxic profile and antiproliferative effect of the complexes 216
Stability of rG4-C_8 complex.....217
In vitro cellular uptake and intracellular distribution of the rG4-C_8 complex...218
Conclusions 219
Acknowledgments 220
References 220
Supplementary Information 223

Chapter 9.....229
Chapter overview 229
Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin 231

Abstract 231
Keywords 232
Introduction 232
Materials and methods 233
Bioinformatic analysis 233
Oligonucleotides and ligands 234
Cloning, cell-free expression, and purification 234
NMR spectroscopy 235
Circular dichroism 235
Molecular docking and molecular dynamics simulations 235
Non-denaturing polyacrylamide gel electrophoresis 236
Surface Plasmon Resonance (SPR) 237
Fluorescein isothiocyanate (FITC) labelling 238
Confocal microscopy 238
Microfluidic experiments 239
Results 239
Bioinformatic analysis of pre-miRNA-149 G4 239
pre-miRNA-149 G4 formation 240
Evaluation of K^+ dependence in the formation of pre-miRNA-149 G4 241
Effect of G4 ligands on pre-miRNA-149 stability 242
In silico analysis of the interaction between NCL RBD1,2 and G4 motif found in pre-miRNA-149 243
Evaluation of the formation of the pre-miRNA-149/ligand, pre-miRNA-149/NCL and pre-miRNA-149/ligand/NCL molecular complexes 244
Binding affinity of pre-miRNA-149 and pre-miRNA-149/ligand complexes towards NCL RBD1,2 246
Co-localization of NCL with pre-miRNA-149 in cancer cells 246
Capture of NCL RBD1,2 by a pre-miRNA-149-based microfluidic platform 247
Discussion 248
Conclusions 252
Acknowledgments 253
References 253
Supplementary Information 257

Chapter 10 261
Chapter overview 261

Insights into the interaction of an RNA G-quadruplex with a ligand and nucleolin: A structural perspective.....263
Abstract 263
Keywords 263
Introduction 263
Materials and methods 264
Oligonucleotides and ligands 264
Protein production and purification 264
Non-denaturing polyacrylamide gel electrophoresis 265
Isothermal Titration Calorimetry (ITC) 265
NMR experiments 266
Results and discussion 267
Acknowledgments 276
References 276
Supplementary Information 279

Section III 283

Chapter 11.....285
General Conclusions and Future Perspectives.....287
General Conclusions 287
Future Perspectives.....291

List of Figures

Figure 2.1. Illustration of G-tetrad formation through Hoogsteen bonds and G-tetrads stacking 12

Figure 2.2. (A) Schematic representation of different G4 topologies (parallel, hybrid, and anti-parallel). **(B)** Conformations of G4s according to strands polarity when polarity differences are associated with an angle between the G-bases and the pentose, namely to the 2' hydroxyl group. 13

Figure 2.3. Illustration of G4RP method protocol for isolation of G4 targets from human cell extracts and the influence of the ligands in gene expression. 14

Figure 2.4. Chemical structures of the G4 ligands presented in this review. Small molecules are divided according to their chemical families. Structures were designed with ChemDraw 20.0 ®—PerkinElmer. 16

Figure 2.5. Schematic representation of TERRA rG4 interaction with TRF2 GAR domain required for telomere stability and integrity. Treatment of human melanoma cells with N-methyl mesoporphyrin IX (NMM) leads to disruption of TERRA and the induction of γH2AX associated telomeric DNA damage. 17

Figure 2.6. Representation of the influence of G-rich sequence in normal **(A)** and reduced **(B)** levels of miR-23b, miR 27b, and miR-24-1. 18

Figure 2.7. Representation of the ionic profiling to induce G4 formation by unwinding the stem–loop of pre-miR-92b 19

Figure 2.8. Representation of SHALiPE method 20

Figure 2.9. Biogenesis of miRNAs. Comparison of natural biogenesis versus interference of the rG4s in the biogenesis and function of miRNAs. 23

Figure 2.10. Representation of G4 formation in miR-3620-5p and the inhibition effect of sanguinarine on the base pair formation of miR-3620-5p with its target sequence 24

Figure 3.1. (**A**) Guanine-rich sequence with the potential to form a three-tetrad G4. (**B**) Chemical structure of G-tetrad formed by the Hoogsteen hydrogen-bonded guanines and central cation (colored in gray) coordinated to oxygen atoms. Schematic representation of common unimolecular G4s based on the strand direction: (**C**) parallel, (**D**) antiparallel, and (**E**) hybrid. Representative PDB structures of (**F**) parallel (PDB ID: 2M4P), (**G**) antiparallel (PDB ID: 1I34), and (**H**) hybrid (2JPZ) G4 structures. The different loops (propeller, diagonal and lateral) and a bulge were also shown. 36

Figure 3.2. List of some examples of G4-interacting ligands mentioned in this review showing the common name of the ligand, chemical structure, and family of the compound (chemical backbone). 39

Figure 3.3. Examples of selective DNA G4 ligands. (**A**) Chemical structures and family of the ligands TOxaPy and NDI 3. (**B**) Chemical structure of IZNP-1 and proposed binding model of the intercalation of IZNP-1 into two telomeric G4 units. 43

Figure 3.4. (**A**) Chemical structure of the acridine derivative C_8. (**B**–**E**) Computational models of complex C_8/*KRAS*22-RT generated by molecular docking after energy minimization. The figures (**B**–**E**) depicted conformers obtained with different binding free energies, −9.32, −7.24, −9.17, and −8.40 kcal/mol, respectively. G4 structure is depicted as a light orange surface with the 5′ tetrad highlighted in darker orange. C_8 is represented in blue..... 44

Figure 3.5. Fluorescent G4 probes. (**A**) Common name and chemical structures of IZFL-2, NBTE, and IMT. (**B**) Chemical structure of SiR-PyPDS and single-molecule fluorescence imaging of G4s in living cells using the fluorescent probe SiR-PyPDS 45

Figure 3.6. RNA G4-interacting ligands mentioned in this review showing the common name of the ligand and chemical structure. 47

Figure 3.7. Computational models showing the lowest energy binding poses of the complexes between the prediction model of RNA G4 formed into pre-miRNA-149 and the ligands (**A**) C_8, (**B**) C_8-NH_2, (**C**) (16)phenN_2, (**D**) (32)phen$_2N_4$, (**E**) PhenDC3 and (**F**) PDS. (**G**) Alternative binding pose of the complex formed by pre-miRNA 149 and PDS after 100 ns of MD simulations. G4 structure is depicted in blue, while ligands are highlighted in green and red 47

Figure 3.8. (**A**) Chemical structure of QUMA-1; confocal microscopy images of fixed HeLa cells (**B**) stained with QUMA-1; (**C**) stained with QUMA-1 and treated with RNase

A; (**D**) stained with QUMA-1 and treated with DNase I; (**E**) stained with QUMA-1 and treated with CX-5461 and (**F**) stained with QUMA-1 and treated with α-Amanitin. (**G**) Fluorescence intensity per cell. 48

Figure 3.9. Schematic representation of GTFH probe. (**A**) GTFH probe design principle. (**B**) Chemical structure of ISCH-oa1. (**C**) GTFH probe preparation..... 48

Figure 3.10. Chemical structures of PyroTASQ and NaphtoTASQ. 49

Figure 3.11. (**A**) CD spectra of G4 structures: c-MYC (10 μM) in 100 mM KCl (parallel); 22AG (10 μM) in 100 mM NaCl (antiparallel) and 22AG (10 μM) in 100 mM KCl (hybrid). The final volume of each sample in the 1 cm path-length quartz cell was 300 μL. The buffer used for the experiments was 10 mM phosphate buffer pH 7.1. (**B**) CD melting curves for c-MYC G4 (10 μM) in 100 mM LiCl in the absence and presence of 2 molar equivalents of TMPyP4. The ellipticity was monitored at a wavelength of 263 nm. The buffer used for the experiments was 10 mM phosphate buffer pH 7.1..... 52

Figure 3.12. (**A**) Chemical structure of 2,4,6-triarylpyridine (TAP). (**B**) Imino and aromatic proton regions titration spectra of 500 μM 22AG with different concentrations of TAP in 200 μL of sodium phosphate buffer containing 20 mM Na_2HPO_4/NaH_2PO_4, 70 mM NaCl at pH 6.9. (**C**) Expanded H8/H6-H1′ 300 ms 2D–NOESY spectrum of the 2.5:1 TAP:22AG complex. (**D**) Expanded H1-H8/H2 region from a 300 ms 2D–NOESY spectrum of the TAP:22AG complex at 2.5:1 stoichiometry.....55

Figure 3.13. High-resolution structures of Myc2345_T23 in complex with PEQ, Quindoline-I, DC-34, and BMVC. (**A**) Common names and chemical structures of the ligands. Binding of each ligand to the (**B**) 5′-G-tetrad and (**C**) 3′-G-tetrad of Myc2345_T23..... 56

Figure 3.14. ^{19}F NMR spectra of ^{19}F-labeled 0.5 mM RNA (^{19}F-5′-UAGGGUUAGGGU-3′) in 150 μL of 90% H_2O/10% D_2O, 10 mM Tris-HCl buffer (pH 7.0), 50 mM KCl. titrated with PDS 57

Figure 3.15. Crystal structure of the T7-NMM complex. (**A**) Representative crystal morphologies: half-hexagonal and hexagonal. (**B**) Schematic representation of the folding topology with the numbering scheme for T1. (**C**) Electron density maps of the T7-NMM dimer. (**D**) Crystal structure of the T7-NMM complex. The T7-NMM sample was prepared by annealing T7 with 1 eq. of NMM at 0.65 mM in 10 mM lithium cacodylate pH 7.2 and 20 mM KCl. Drops were set by the TTP Labtech Mosquito Crystal liquid

handler equipped with a humidity chamber at 0.1 μL DNA sample and 0.1 μL of the crystallization condition. Small hexagonal crystals grew within three weeks to 80 μm in the largest dimension from condition C5 of the Natrix screen (Hampton Research): 4.0 M LiCl, 0.01 M MgCl2, and 0.05 M HEPES sodium pH 7.0. Crystals were cryoprotected in the base condition supplemented with 15% ethylene glycol before being flash-frozen in liquid nitrogen..... 59

Figure 3.16. Schematic representation of (**A**) immobilization of biotin-labeled RNA G4 on CM5 chip, (**B**) biomolecular recognition in the flow cell with changes in the angle of incidence light, and (**C**) SPR sensorgrams..... 61

Figure 3.17. (**A**) Sensorgrams with increasing concentrations of the diamidine derivative DB1246. (**B**) Binding curve showing a two-site binding model with K_{D1} = 3 ± 0.8 nM and K_{D2} = 80 ± 12 nM. For the immobilization procedure, ~100 μL of 25 nM of TERRA G4 were prepared in HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 05% *v*/*v* polysorbate 20). The compound DB-1246 (300 μL) was injected into the system from low (10 nM) to high concentration (250 nM)..... 62

Figure 3.18. Representative ITC titration profiles for the titration experiments between hTel22 and (**A**) Cu(II)-TMPyP4, (**B**) Ni(II)-TMPyP4, (**C**) Co(III)-TMPyP4, and (**D**) Zn(II)-TMPyP4 in 150 mM K^+ BPES buffer. ITC experiments were performed of either 28 (10 μL) or 56 (5 μL) injections of a dilute ligand solution into a dilute hTel22 DNA solution..... 63

Figure 3.19. (**A**) Chemical equilibria for ligand binding to a G4 forming sequence. Mass spectra at different temperatures of 10 μM of human telomeric DNA sequence (**B**,**C**: 24TTG; **D**,**E**: 22GT) in the presence of 1 or 3 equivalents of (**B**) Phen-DC3, (**C**) TMPyP4, (**D**) Cu-ttpy, and (**E**) Pt-ttpy. The buffer used was 100 mM TMAA and 1 mM KCl. The volume of oligonucleotide solution was approximately 10 μL..... 65

Figure 3.20. Schematic general representation of FRET-melting experiments..... 67

Figure 3.21. (**A**) Chemical structures of (16)phenN_4 and (32)phen$_2N_4$ and FRET-melting stabilization by (16)phenN_4 and (32)phen$_2N_4$ (1 μM) of the labeled F21T sequence (0.2 μM) in buffer 10 mM KCl + 90 mM LiCl and in the presence of increasing concentrations of competitor ds26 (0, 15 and 50 equivalents). (**B**) Chemical structures of indolo(3,2-c)quinoline derivatives 3d and 3e, and FRET-melting competition assay with 3e (1 μM) in complex with *KRAS*21R and F21T G4s (0.2 μM), with increasing concentrations of non-labeled 26ds, HT21 or *KRAS*21R (0.4 to 25 μM) competitor, in

cacodylate buffer, pH 7.4 supplemented with K^+. (**C**) Chemical structure of N-methylquinolinium derivative 2a and FRET-melting competition assay results for compounds 2a–d at 0.5 μM (2a) and 5 μM (2b–d) complexed with *KRAS* G4 (0.2 μM), challenged with increasing concentrations of non-labeled 26ds DNA (dsDNA) (0.4 μM to 25 μM) competitor. Experiments were performed in cacodylate buffer, pH 7.4, supplemented with 60 mM K^+..... 68

Figure 3.22. Schematic general representation of G4-FID experiments. Chemical structures of TO, TO-PRO-3, Hoechst 33258, and PhenDV. 70

Figure 3.23. (**A**) Representative G4-FID curves obtained with MMQ1 (gray), MMQ3 (brown), Pt-ttpy (red), 360A (blue) and Phen-DC3 (green) with 22AG.K (squares) and ds26 (circles) using TO as fluorescent probe. (**B**) Representative G4-FID curves of TMPyP4 with 22AG.Na (blue), 22AG.K (red), c-kit1 (brown), c-kit2 (gray), and ds26 (green) with TO-PRO-3 as fluorescent probe. For each well of the microplate were added: (i) K^+- or Na^+-buffer solution; (ii) oligonucleotides (5 μM) and fluorescent probe (TO or TO-PRO-3; 10 μM for G4-DNA or 15 μM for ds26) and (iii) an extemporaneously prepared 5 μM ligand solution in K^+- or Na^+-buffer. The final volume in each well was 200 μL..... 70

Figure 3.24. (**A**) Scheme of affinity chromatography strategy to purify G4 sequences based on a ligand affinity chromatography support. (**B**) Scheme of affinity chromatography strategy to purify G4 ligands based on a G4 affinity chromatography support.....72

Figure 3.25. Amount of the released ligands expressed as a percentage of the quantity loaded (**A**) on OAS support and (**B**) on tel26-OAS support, as a function of the washing solution volume (100 mM KCl/15% DMSO) 73

Figure 4.1. (A) Schematic illustration of nucleolin primary structure and sites of post-translational modifications. **(B)** Predicted 3D structure of nucleolin via the AlphaFold algorithm. α-Helixes are depicted in orange, β-sheets are highlighted in green, and random coils are shown in gray. **(C)** G-quadruplex (G4) structures are among the main nucleolin binding partners that assist it in several cellular functions. The activity of nucleolin can be regulated upon G4 recognition and vice versa. Recent advances propose important roles for the nucleolin/G4 partnership in neurodegenerative disease, cancer, and viral infection. In cancer, the binding of nucleolin to G4s can both trigger or repress transcription, translation, and/or miRNA biogenesis. Cell-surface nucleolin has a key

role in cancer because it can be the target of extracellular ligands. In neurodegeneration, G4s in the *C9orf72* gene bind to nucleolin, which triggers nucleocytoplasmic trafficking, altered liquid-liquid phase separation, and aberrant stress-granule formation. In the case of viral infections, the nucleolin/G4 partnership plays a crucial role in virus attachment, episomal latency, viral replication, and nucleolin sequestration. Abbreviation: RBD, RNA-binding domain 97

Figure 5.1. Chemical structure, common name, and family of the ligands used in this work (chemical backbone). 110

Figure 5.2. (A) IDS steady state of rG4-let-7e in the presence of increasing amounts of KCl. **(B)** TDS spectrum in the presence of 20 mM phosphate buffer, pH 7.1 supplemented with 100 mM KCl **(C)** CD spectra of the putative rG4, found in let-7e pre-miRNA, at 10 μM in 20 mM phosphate buffer, pH 7.1 in the absence and presence of 100 mM KCl. **(D)** CD-melting curves of the putative rG4-let-7e at 10 μM in 20 mM phosphate buffer, pH 7.1 in the absence and presence of 100 mM KCl. 115

Figure 5.3. ^{1}H NMR spectra of the imino and aromatic proton region of the putative rG4-let-7e (100 μM) in Phosphate buffer containing 20 mM K_2HPO_4/KH_2PO_4, and different concentrations of KCl in the range 0-100 mM. All the ^{1}H NMR spectra were acquired at 27 °C. 116

Figure 5.4. T_m radar plots of rG4-let-7e in the presence of different molar equivalents of G4 ligands, obtained by **(A)** FRET-melting and **(B)** CD-melting experiments. 117

Figure 5.5. ^{1}H NMR spectra of the imino and aromatic proton region of the rG4-let-7e (100 μM) upon titration with **(A)** 360A, **(B)** PDS, and **(C)** TMPyP4. Spectra were acquired in phosphate buffer containing 20 mM K_2HPO_4/KH_2PO_4 at 27 °C. 120

Figure 5.6. Molecular interaction of rG4-let-7e with NCL RBD1,2. **(A)** CD spectra of rG4-let-7e in the absence and presence of increasing amounts of NCL RBD1,2 (0-10 μM). **(B)** CD-melting curves of rG4-let-7e in the absence and presence of NCL RBD1,2. **(C)** Equilibrium binding curve of rG4-let-7e upon addition of increasing concentrations of NCL RBD1,2. 121

Figure 5.7. (A) Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) in presence of different concentrations of NCL RBD1,2 (Lanes: M – Marker (90, 60, 30, 21 and 15 nt.); 1 – NCL RBD1,2 (2 μM) ; 2 – rG4-let-7e without NCL RBD1,2; 3 – 0.1 μM; 4 - 0.250 μM; 5 – 0.5 μM; 6 – 1 μM; 7 - 2 μM; 8 – 4 μM: 9 – 8 μM; 10 – 16 μM and 11 – 32 μM). **(B)**

Relative proportions of NCL RBD1,2-bound rG4-let-7e to unbound rG4-let-7e were quantified and plotted. **(C)** Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) / ligand (2 μM) / NCL RBD1,2 (2 μM). (Lanes: Lanes: M – Marker (90, 60, 30, 21 and 15 nt.); 1 – rG4-let-7e in 20 mM Phosphate buffer supplemented with 100 mM KCl; 2 – rG4-let-7e (2 μM) / NCL RBD1,2 (2 μM); 3 – 10 – ligands (2 μM) (3 - 360A; 4 – BioTASQ; 5 – BRACO-19; 6 – C_8 and 7 – c-PDS; 8 – PDS; 9 - PhenDC3; 10 - TMPyP4)) 122

Figure 5.8. (A) Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) without or with ligand (2 μM) in presence of RNase H (0.3 U/μL). (Lanes: M – Marker (60, 30, 21 and 15 nt.); 1 – rG4-let-7e in water; 2 – rG4-let-7e in 20 mM Phosphate buffer supplemented with 100 mM KCl; 3 – 10 – ligands (2 μM) (3 - 360A; 4 – BioTASQ; 5 – BRACO-19; 6 – C_8 and 7 – c-PDS; 8 – PDS; 9 - PhenDC3; 10 - TMPyP4)). **(B)** Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) / ligand (2 μM) / NCL RBD1,2 (2 μM) in presence of RNase H (0.3 U/μL). (Lanes: Lanes: M – Marker (60, 30, 21 and 15 nt.); 1 – rG4-let-7e in water; 2 – rG4-let-7e in 20 mM Phosphate buffer supplemented with 100 mM KCl; 3 – rG4-let-7e (2 μM) / NCL RBD1,2 (2 μM); 4 – 10 – ligands (2 μM) (4 - 360A; 5 – BioTASQ; 6 – BRACO-19; 7 – C_8 and 8 – c-PDS; 9 – PDS; 10 - PhenDC3; 11 - TMPyP4)). 123

Figure 5.9. Confocal microscopy of A549 cells incubated with **(A)** rG4-let-7e and **(B)** rG4-let-7e/PDS complex. For each panel, images showed the cells stained with Hoechst 33342® nuclear probe (2 μM, blue); rG4-let-7e Cy3.5 (1 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:500). Arrows showed co-localization of rG4-let-7e or rG4-let-7e/PDS complex and NCL.....124

Figure 6.1. (A) Chemical structure of the acridine orange derivative ligands used in this study. **(B)** pre-miR-92b rG4 sequence (5'-GGGCGGGCGGGAGGG-3') (mutated residues are highlighted in green). **(C)** Microchannel design of the two-height microfluidic structure.....154

Figure 6.2. (A) CD titration spectra of pre-miR-92b rG4 at 10 μM with increasing concentrations of K^+. **(B)** CD melting curves at 10 μM in the absence and presence of increasing concentrations of K^+. Data points were recorded at 261 nm. 155

Figure 6.3. (A) CD titration spectra of pre-miR-92b rG4 in the absence and presence of ligands. **(B)** CD melting curves of pre-miR-92b rG4 in the absence and presence of 2 M equivalents of each of the ligands. Data points were recorded at 261 nm. 156

Figure 6.4. 1D ^{1}H spectra showed the imino region of pre-miR-92b rG4 (0.1 mM) upon KCl titration of **(A)** 100 mM, **(B)** 50 mM, **(C)** 10 mM, **(D)** 5 mM, **(E)** 1 mM and **(F)** 0.1 mM KCl. The spectra were recorded in 90% H_2O, 10% D_2O at 27.7 °C. 157

Figure 6.5. 1D ^{1}H spectra showed the imino region of pre-miR-92b rG4 (0.1 mM) upon **(A)** 0.1 mM; **(B)** 1 mM; **(C)** 5 mM; **(D)** 10 mM; **(E)** 50 mM and **(F)** 100 mM KCl titrations. The spectra were recorded in 90% H_2O, 10% D_2O at different temperatures, 17 °C (blue), 27.7 °C (red), 40 °C (green), 50 °C (purple) and 60 °C (grey). 158

Figure 6.6. (A) pre-miR-92b rG4 detected on non-denaturing polyacrylamide gel electrophoresis. The samples were loaded on a non-denaturing 15% acrylamide gel supplemented with 10 mM KCl. Migration was performed at 20 °C. Migration markers were single-stranded (n = 9, 15, 21, 30, 60, and 90). The diffuse intense band below corresponds to the dyes bromophenol blue and xylene cyanol used to follow the migration. After electrophoresis, the gels were stained and visualized by GE Typhoon Trio Imager Scanner. **(B)** Percentage of pre-miR-92b rG4/NCL RBD1,2 complexes formed in different concentrations of protein (50, 100, 250, and 500 nM). **(C)** Percentage of pre-miR-92b rG4/ligand/NCL RBD1,2 complexes formed in the presence of NCL RBD1,2 (500 nM) and each ligand (30 μM). Three independent experiments were conducted. Data are shown as mean with standard error of the mean. A one-way analysis of variance (ANOVA) was performed to evaluate the significance of the differences between the complex formation values of the different ligands and between the protein concentrations (the asterisks indicate $p < 0.05$)..... 159

Figure 6.7. (A) RMSD plot for the pre-miR-92b rG4 (green); NCL RBD1,2 (black) and pre-miR-92b rG4/NCL RBD1,2 complex (red) structure during 100 ns MD simulation. Final snapshots of pre-miR-92b rG4 and pre-miR-92b rG4/NCL RBD1,2 complex after 100 ns MD production. **(B)** Side view showing a parallel-stranded rG4 **(C)** Top view showing a parallel-stranded G4 with propeller-type loops (depicted in gray). **(D)** pre-miR-92b rG4/NCL RBD1,2 complex. Propeller-type loops are depicted in gray. K^+ cations are depicted as purple spheres. NCL RBD1,2 is represented in green and contact residues of rG4 and NCL RBD1,2 are highlighted in orange. H-bonds are shown in red. Ligplot + image showing site-specific hydrogen bonding and hydrophobic interactions of NCL RBD1,2 residues and **(E)** pre-miR-92b rG4 C4 residue and **(F)** pre-miR-92b rG4 A12 and G13 residues. NCL RBD1,2 residues are represented in orange. pre-miR-rG4 residues with hydrogen bonding interactions are depicted in blue. pre-miR-92b rG4 residues without hydrogen bonding with NCL RBD1,2 are highlighted in green. H-bonds are shown in green dashed lines. 160

Figure 6.8. Plots of the fraction of bound NCL, fitted to the saturation binding equation. **(A)** pre-miR-92b rG4-Cy3.5 and **(B)** pre-miR-92b rG4-Cy3.5/C_8 complex at 1 μM with increasing concentrations of NCL RBD1,2 in 10 mM lithium cacodylate buffer pH 7.2, supplemented with 0.1 mM KCl.....163

Figure 6.9. Stability assays of **(A)** control RNA sequence and **(B)** pre-miR-92b rG4 structure in 10 mM lithium cacodylate buffer pH 7.2, supplemented with 0.1 mM KCl at 1, 3, 12, 24, and 48 h. The experiments were performed at 37 °C in absence and presence of 10% FBS. The degradation pattern was monitored by 20% non-denaturing polyacrylamide gel supplemented with 10 mM KCl. Migration markers were single-stranded oligonucleotides (n = 9, 15, 21, 30, 60, and 90). Note that the diffuse bottom bands of the migration marker correspond to bromophenol blue and xylene cyanol used to follow the migration. Normalized band intensities of **(C)** control RNA sequence and **(D)** pre-miR-92b rG4 structure in the presence of C_8, NCL RBD1,2, and both were analyzed by densitometry. Three independent experiments were conducted. Data are shown as mean with standard error of the mean. A one-way analysis of variance was performed to evaluate the significance of the differences between different groups (the asterisks indicate $p < 0.05$). 164

Figure 6.10. Microfluidic assay studies for NCL RBD1,2 measurement with pre-miR-92b rG4 labeled with Cy3.5 and complexed with C_8. **(A)** Relative fluorescence intensity of the pre-miR-92b rG4-Cy3.5/C_8/NCL RBD1,2 complex and control proteins PSA and HSA. **(B)** Fluorescence images acquired with an exposure time of 2 s in an Axio Imager Z2 microscope and processed in Zeiss Zen Software (SP2, 2010). 166

Figure 7.1. Structure of the G4 ligands used in this study. 177

Figure 7.2. Predicted structural model of pre-miRNA 149 rG4 after 20 ns MD production. Side view (left) and top view (right) show a parallel stranded rG4 with propeller-type loops (depicted in gray). K^+ cations are depicted as purple spheres. ... 181

Figure 7.3. RMSD plot of the 20 ns simulation of pre-miRNA 149 G4 model. 181

Figure 7.4. RMSF plot of each nucleotide in the pre-miRNA 149 G4 during the full 20 ns MD simulation. 182

Figure 7.5. Selected binding poses in terms of the lowest binding free energy of the complex formed by pre-miRNA 149 rG4 and **(A)** C_8, **(B)** C_8-NH_2, **(C)** [16]phenN_2, **(D)** [32]phen$_2N_4$, **(E)** PhenDC_3 and **(F)** PDS. G4 structure is depicted as a blue surface with

K^+ cations depicted as purple spheres, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. 183

Figure 7.6. RMSD plot of the 20 ns simulation of pre-miRNA 149 - C_8 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. 184

Figure 7.7. RMSD plot of the 20 ns simulation of pre-miRNA 149 - C_8-NH_2 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. 185

Figure 7.8. RMSD plot of the 20 ns simulation of pre-miRNA 149 - [16]phenN_2 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. Inset, corner: zoomed view of the H-bonds with the rG4 phosphate backbone..... 185

Figure 7.9. RMSD plot of the 20 ns simulation of pre-miRNA 149 - [32]phen$_2$$N_4$ complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. Inset, corner: zoomed view of the H-bonds with the rG4 phosphate backbone..... 186

Figure 7.10. RMSD plot of the 20 ns simulation of pre-miRNA 149 – PhenDC_3 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. 187

Figure 7.11. RMSD plot of the 20 ns simulation of pre-miRNA 149 - PDS complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. Inset, corner: zoomed view of the H-bonds with the rG4 structure. 188

Figure 7.12. Normalized frequency data plots of SASA of the ligand-free pre-miRNA 149 rG4 and rG4-ligand complexes..... 189

Figure 8.1. CD titration spectra of short loop (112) rG4 at 10 μM with increasing concentrations of **(A)** C_8 and **(B)** C_8-NH_2. Spectra acquisitions were performed at 20 °C in 10 mM lithium cacodylate and 0.1 mM KCl; **(C)** CD melting curves at 10 μM in the absence and presence of 2 M equivalents of ligands in 10 mM lithium cacodylate and 0.1 mM KCl, respectively. Data points were recorded at 263 nm. Temperature corresponds to the temperature set by the Peltier system..... 214

Figure 8.2. Fluorescence emission spectra of **(A)** C_8 and **(B)** C_8-NH_2 at 5 μM with different concentrations of the short loop (112) rG4 in 10 mM lithium cacodylate buffer and 0.1 mM KCl. Insets: fraction of ligand bound plots fitted to the saturation binding equation. 215

Figure 8.3. Plots of the fraction of bound NCL to **(A)** short loop rG4-Cy3.5 and **(B)** short loop rG4-Cy3.5-C_8 complex at 1 μM in 10 mM lithium cacodylate buffer and 0.1 mM KCl. 216

Figure 8.4. Relative cell viability of **(A)** PC-3 and **(B)** PNT1A cell lines incubated for 7 days with short loop (112) rG4, and rG4-C_8 complex at C_8 and rG4 concentrations of 1 μM and 15 μM, respectively. Untreated cells were used as negative controls for cytotoxicity. Mean percentage values relative to untreated cells and standard error of the mean in 3 independent experiments are shown. The bars represent the mean, and the lines represent the SEM associated. $^{****}P < 0.0001$ (one-way ANOVA). 217

Figure 8.5. Confocal laser scanning microscopy images showing the intracellular uptake of short loop rG4-C_8 complex from 0 to 60 min. For each panel, images showed the cells with nuclear staining by Hoechst 33342® (1 μM, blue); rG4 Cy3.5-C_8 complex (1:1 ratio, yellow); C_8 (1 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:1000). **(A)** Representative z-stacks data showing PC-3 cell line and **(B)** Representative z-stacks data showing PNT1A cell line. 219

Figure 9.1. (A) Chemical structures of C_8, C_8-NH_2, and PhenDC3. **(B)** G4 scores of pre-miRNA-149 predicted by using QGRS-Mapper, G4Hunter, cGcC, and G4NN. **(C)** Predicted secondary structure of pre-miRNA-149 showing miRNA-149-5p, miRNA-149-3p, and the rG4 region. **(D)** Pairwise alignment and conservation of pre-miRNA-149 sequences found in several species 240

Figure 9.2. (A) 1D ^1H spectra of pre-miRNA-149 in the presence of K^+ showing the spectrum region of the Watson-Crick and G4 structures. **(B)** CD spectra of pre-miRNA-149 before and after annealing and in the absence and presence of K^+. Inset: Zoom of the maximum positive band. 241

Figure 9.3. (A) CD spectra of pre-miRNA-149 in 20 mM lithium cacodylate buffer in the absence and presence of increasing concentrations of KCl. **(B)** CD melting curves at

10 μM in the absence and presence of increasing concentrations of KCl. Data points were recorded at 262 nm. 242

Figure 9.4. (A) CD spectra of pre-miRNA-149 in 20 mM lithium cacodylate and 1 mM KCl in the absence and presence of increasing concentrations of **(A)** C_8, **(B)** C_8-NH_2, and **(C)** PhenDC3. CD melting normalized curves at 10 μM in the absence and presence of increasing concentrations of **(D)** C_8 and (E) C_8-NH_2. CD melting curve without normalization at 10 μM in the absence and presence of increasing concentrations of (F) PhenDC3. Data points were recorded at 262 nm..... 243

Figure 9.5. *In silico* analysis of G4/NCL RBD1,2 complex. **(A)** Final snapshots of side and top view of the G4/NCL RBD1,2 complex after 100 ns MD production. G4 structure is depicted in blue, with the backbone in light gray. K^+ cations are depicted as purple spheres. NCL RBD1,2 is represented in green and contact residues of the protein are highlighted in orange. H-bonds are shown in cyan, while labels are shown in red. **(B)** RMSD plot of the G4/protein (green), protein (black), and G4 (red) for complex G4/NCL RBD1,2. **(C)** Ligplot + image showing site-specific hydrogen bonding and hydrophobic interactions of NCL RBD1,2 residues and rG4 residues A4 and G5. Hydrogen bonds are depicted in dashed green lines and hydrophobic interactions in red. 244

Figure. 9.6. PAGE under native conditions of pre-miRNA-149 in **(A)** absence (Lane M - migration marker; lane 1 – pre-miRNA-149 in the presence of 20 mM K_2HPO_4/KH_2PO_4; lane 2 – pre-miRNA-149 in the presence of 20 mM K_2HPO_4/KH_2PO_4 supplemented with 100 mM KCl; lanes 3 to 5 – pre-miRNA-149/ligand complexes (with 1 molar eq. of C_8, C_8-NH_2 , and PhenDC3, respectively) and in the **(B)** presence of NCL RBD1,2 (1 μM) (Lane M - migration marker; lane 1 – NCL RBD1,2 (1 μM); lane 2 – pre-miRNA-149 in the presence of 20 mM K_2HPO_4/KH_2PO_4 supplemented with 100 mM KCl; lanes 3 to 5 – pre-miRNA-149/ligand complexes (with 1 molar eq. of C_8, C_8-NH_2 , and PhenDC3, respectively). The gel was supplemented with 10 mM KCl. Migration markers were oligothymidylate single-stranded DNA (n = 30, 60, and 90). 245

Figure 9.7. Confocal microscopy images showing co-localization of NCL and pre-miRNA-149, labeled with FITC. The images showed the cells with nuclear staining by Hoechst 33342® (1 μM, blue); pre-miRNA-149 labeled with FITC (10 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:1000). 247

Figure 9.8. Fluorescence intensity of NCL RBD1,2 labeled with FITC and captured by pre-miRNA-149 or by pre-miRNA-149/PhenDC3 complex immobilized in streptavidin beads. Different concentrations of protein were flowed through the microfluidic channel (0.25, 0.5, 1.25, 2.5, and 5 μM)..... 248

Figure 10.1. a) Schematic illustration showing the rG4 sequence in pre-miRNA 149. **b)** Electrophoretic profile of rG4 structure. M – Marker; 1 – rG4 in the presence of H_2O; 2 - rG4 in the presence of 100 μM KCl and 3 – rG4 in the presence of 100 mM KCl **c)** Effect of KCl salt on pre-miR-149 as monitored by 1D ^{1}H NMR spectroscopy. Imino protons of rG4 solution (100 μM) with increasing concentrations of KCl (as indicated on the right of the spectrum). The spectra were recorded in H_2O Milli-Q, 10 μM TSP-d_4, and 10 % D_2O at 293.15 K. **d)** ^{1}H NMR spectra titration of the rG4 structure with C_8. Imino region of ^{1}H spectra of rG4 solution (100 μM) with increasing concentrations of C_8 (rG4/C_8 ratio and intensity scale are indicated on the right of the spectrum). Imino protons of the rG4 are indicated with asterisks (*), and those in the complex are indicated with #. The spectra were recorded in water containing 100 μM KCl, 10 μM TSP-*d4*, 10 % D_2O at 293.2 K. The annealing procedure was performed for each sample before measurement..... 268

Figure 10.2. ^{1}H NMR spectra titration of C_8 with rG4. **a)** Chemical structure of C_8. **b)** Aromatic region of ^{1}H spectra of C_8 solution (400 μM) with increasing concentrations of rG4 (rG4/ligand ratio). **c)** ^{1}H NMR spectra overlay of rG4 without and with 0.01 molar eq. of C_8. The spectra were recorded in water containing 100 μM KCl, 10 μM TSP-*d4*, 10 % D_2O and 4% DMSO-d_6 at 293.15 K. The annealing procedure was performed for each sample before measurement..... 269

Figure 10.3. STD-NMR of C_8 in the presence of the rG4. **a)** Top - reference spectrum (off-resonance) with resonance assignments; **b)** Bottom – STD NMR spectrum of 400 μM C_8 with 4 μM rG4 (irradiation at 12 ppm) and relative STD NMR intensities in percentage. **c)** Epitope mapping was obtained for the C_8 ligand in the presence of the rG4 271

Figure 10.4. Affinity and thermodynamic parameters were measured by isothermal titration calorimetry (ITC) using the rG4 sequence (5'-GGGAGGGAGGGACGGG-3') as a ligand of NCL RBD1,2. The ΔH value was -18.25 kcal/mol, whereas $T\Delta S$ value was -0.75 kcal/mol..... 272

Figure 10.5. NMR titrations of NCL RBD1,2 with unlabeled rG4 and rG4/C_8 complex. **a)** Overlay of three ^{1}H-^{15}N HSQC spectra of ^{15}N ^{13}C NCL RBD1,2 with increasing concentrations of unlabeled rG4 and rG4/C_8 complex. **b)** Plot of ratio of peak intensities of complex (I) vs free (I_o) for the NCL RBD1,2/rG4 and NCL RBD1,2/rG4/C_8 complex at 1/0.25 molar ratio shown with the average peak intensity ratio (dashed gray line) and one standard deviation from the average (1σ, dashed orange line)**. c)** Most affected residues in the presence of 0.25 molar eq. rG4 and rG4/C_8 (colored red) were mapped onto the solution NMR structure of NCL RBD1,2 (colored gray). Data were acquired with 200 μM ^{15}N, ^{13}C-labelled NCL in buffer 50 mM potassium phosphate pH 6.2, 0.1 % NaN_3, 20 μM TSP-d4, 10 % D_2O with 400 mM KCl. 274

List of Tables

Table 2.1. List of G4 sequences in ncRNA nucleic acids discussed in this review 25

Table 3.1. Advantages and limitations of structural-based methods for investigating G4/ligand interactions..... 59

Table 3.2. Advantages and limitations of affinity and apparent affinity-based methods for investigating G4/ligand interactions. 66

Table 3.3. Advantages and limitations of high-throughput methods for investigating G4/ligand interactions..... 74

Table 4.1. Ligands and aptamers that interfere with nucleolin/G4 binding..... 99

Table 5.1. K_D constant values of rG4-let-7e in the presence of G4 ligands measured by SPR biosensor 118

Table 6.1. Thermal stabilization of pre-miR-92b rG4 92b with increasing concentration of KCl measured by CD melting experiments..... 155

Table 6.2. Pre-miR-92b rG4 binder-induced thermal stabilization (ΔT_m) measured by FRET melting experiments..... 162

Table 7.1. Oligonucleotide sequences of the human telomeric G4 and pre-miR-149 rG4. 178

Table 7.2. Binding-free energy components of rG4-ligand complexes calculated from the MD simulations. The binding energies were calculated using MM/PBSA while entropy contribution was determined by normal mode analysis in GROMACS. All values reported in kcal mol^{-1} with standard deviations (SD) from three different calculations. 191

Table 9.1. K_D (M) constant values of pre-miRNA-149 measured by SPR biosensor and obtained by the Langmuir isotherm 246

List of Abbreviations

^{13}C	Carbon
^{15}N	Nitrogem
^{19}F	Fluorine
^{1}H	Hydrogen
ADP	Adenosine diphosphate
AKT1	V-akt murine thymoma viral oncogene homolog 1
ALIS	Affinity Selection-Mass Spectrometry
ALT	Alternative Lengthening of Telomeres
ANOVA	One-way analysis of variance
ATP	Adenosine triphosphate
BCL2	B-cell lymphoma 2
BHQ1	Black Hole Quencher 1
BP	Bandpass
BRCA 1	Breast cancer 1
BRCA 2	Breast cancer 2
BSA	Bovine serum albumin
C9orf72	Chromosome 9 open reading frame 72
CD	Circular Dichroism
ChIP–seq	chromatin immunoprecipitation sequencing
CK2	Casein kinase
CLSM	Confocal Laser Scanning Microscopy
CM	Carboxymethyl
CM5	carboxymethyl dextran-coated sensor chip
CO_2	Carbon dioxide
COSY	Correlated spectroscopy
c-PDS / cPDS	Carboxy pyridostatin
CPG	Controlled pore glass
Cy3.5	Cyanine3.5
Da	Dalton
DEPC	Diethyl pyrocarbonate
DFQ-COSY	Double quantum filter- Correlated spectroscopy
DGCR8	DiGeorge syndrome critical region 8
DHX36	DEAH-Box Helicase 36

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-stranded DNA breakage
dsDNA	double-stranded DNA
EBNA1	Epstein-Barr nuclear antigen 1
EBV	Epstein–Barr virus
EDP	Electron photodetachment dissociation
EDTA	Ethylenediamine tetraacetic acid
EpCAM	Epithelial cell adhesion molecule
ESI	Electrospray ionization
FAM	6-carboxyfluorescein
FBS	Fetal bovine serum
FITC	Fluorescein isothiocyanate
FOXM1	Forkhead Box M1
FRET	Förster resonance energy transfer
FRET-MC	FRET-melting competition
G4	G-quadruplex
G4-FID	G4-fluorescent intercalator displacement assay
G4NN	G4 Neural Network
G4RP	G4-RNA-specific precipitation
G4RP-seq	G4-RNA-specific precipitation with sequencing
G4-seq	G-quadruplex sequencing
GAR	Glycine–arginine-rich domain
GTFH	G4-triggered fluorogenic hybridization
HCA	Hierarchical cluster analysis
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HeLa	Human cervical adenocarcinoma cells
HIV	Human Immunodeficiency Virus
HMQC	Heteronuclear multiple quantum coherence
HnRNP	Heterogeneous nuclear ribonucleoprotein
HPV	Human Papilloma Virus
HSA	Human serum albumin
HSP90	Heat shock protein 90
HSQC	Heteronuclear Single Quantum Coherence
hTERC	Human telomerase RNA component
ICD	Induced Circular Dichroism

ID	Identification
IDS	Isothermal difference spectra
ITC	Isothermal titration calorimetry
K^+	Potassium
K_2HPO_4	di-Potassium hydrogen phosphate
KCl	Potassium chloride
K_D	Dissociation constant
KH_2PO_4	Potassium dihydrogen phosphate
Kpnb1	Karyopherin Subunit Beta 1
KRAS	Kirsten rat sarcoma viral oncogene homolog
KSHV	Kaposi's sarcoma-associated herpesvirus
let 7	Lethal 7
Li^+	Lithium
LiCaco	Lithium cacodylate
LiCl	Lithium chloride
lincRNAs	Long intergenic non-coding RNAs
LNA	Locked nucleic acid
lncRNA	Long non-coding RNA
LOGSY	Ligand observed via gradient spectroscopy
LSM	Laser Scanning microscopy
MD	Molecular dynamics
Mg^{2+}	Magnesium
$MgCl_2$	Sodium chloride
miRNA /miR	microRNA
MM	Molecular mechanics
mRNA	Messenger RNA
MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide salt
Na^+	Sodium
Na_2HPO_4	di-Sodium hydrogen phosphate
NaCl	Sodium chloride
NaH_2PO_4	Sodium dihydrogen phosphate
NAT	Natural antisense transcripts
NCL	nucleolin
ncRNA	Non-coding RNA

NGS	Next-generation sequencing
NH_4^+	Ammonium
NMM	N-methyl mesoporphyrin IX
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NRAS	Neuroblastoma RAS Viral Oncogene Homolog
NSCLC	Non-small cell lung cancer
nt.	nucleotide
OAS	Oligo-Affinity Support
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PBSA	Poisson Boltzmann surface area
PC3 / PC-3	Human prostatic adenocarcinoma cell line
PCA	Principal component analysis
PCa	Prostate cancer
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PDMS	Polydimethylsiloxane
PDS	Pyridostatin
PEG	Polyethylene glycol
PGMEA	Propylene glycol monomethyl ether acetate
PME	Particle Mesh Ewald
PNT1A	Human prostatic cell line
pre-miRNA / pre-miR	microRNA precursor
pre-rRNA	RNA ribosomal precursor
pri-miRNA / pri-miR	Primary miRNA precursor
PSA	Prostate-specific antigen
PTEN	Phosphatase and tensin homolog
QGRS	Quadruplex forming G-rich sequences
qPCR	Quantitative polymerase chain reaction
qRRM	quasi-RNA recognition motifs
RBD	RNA-binding domain
rDNA	Ribosomal DNA
R-FID	Rapid fluorescent intercalator displacement assay
rG4	RNA G-quadruplex
RGG	arginine-glycine-glycine domain

RISC	RNA-induced silencing complex
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RNA	Ribonucleic acid
rNTP	Ribonucleoside tri-phosphate
RPMI	Roswell Park Memorial Institute
RT-PCR	Real-time reverse-transcription polymerase chain reaction
RU	Resonance units
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SASA	Solvent-accessible surface area
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEM	Standard error of the mean
SHALiPE	Selective 2′-hydroxyl acylation lithium ion based primer extension
SMMs	Small-molecule microarrays
SORL1	Sortilin Related Receptor 1
SOX2	SRY-Box Transcription Factor 2
SPR	Surface Plasmon Resonance
STD	Saturation transfer difference
TAMRA	6-carboxy tetramethylrhodamine
TAP	2,4,6-triarylpyridine
TBA	Thrombin aptamer
TDS	Thermal difference spectra
TERRA	telomeric repeat-containing RNA
T_m	Melting temperature
TMAA	Trimethylammonium acetate
TO	Thiazole orange
TOCSY	Total correlation spectroscopy
TRAP	Telomerase Repeated Amplification Protocol
TRF2	Telomeric repeat binding factor 2
TSP	Trimethylsilylpropanoic acid
UBE2N	Ubiquitin Conjugating Enzyme E2 N
UTR	Untranslated regions
UV	Ultraviolet
VEGF	Vascular endothelial growth factor

VMD	Visual molecular dynamics
γH2AX	Phosphorylated histone H2AX
ΔG	Gibbs Free Energy
ΔH	Entalphy

List of Scientific Publications

Publications included in this thesis:

Figueiredo, J.[‡]; **Santos, T.**[‡]; Miranda, A.[‡]; Alexandre, D.; Teixeira, B.; Simões, P.; Lopes-Nunes, J. and Cruz, C. (2021) Ligands as stabilizers of G-quadruplexes in non-coding RNAs. *Molecules* 26, 6164. DOI: 10.3390/molecules26206164 – IF: 4.412 – **Chapter 2**

Santos, T.; Salgado, G.F.; Cabrita, E.J. and Cruz, C. (2021) G-quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions. *Pharmaceuticals* 14, 769. DOI: 10.3390/ph14080769 – IF: 5.863 – **Chapter 3**

Santos, T.; Salgado, G.F., Cabrita, E.J. and Cruz. C. (2022) Nucleolin: a binding partner of G-quadruplex structures. *Trends Cell Biol.* DOI: 10.1016/j.tcb.2022.03.003 – IF: 20.808 – **Chapter 4**

Santos, T.; Miranda, A.; Imbert, L.; Monchaud, D.; Salgado G. F.; Cabrita E.J. and Cruz, C. (2022) Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin. *J. Pharm. Biomed. Anal.* 215, 114757. DOI: 10.1016/j.jpba.2022.114757 – IF: 3.935– **Chapter 5**

Santos, T.; Miranda, A.; Campello, M.P.C.; Paulo, A.; Salgado, G.; Cabrita, E.J. and Cruz, C. (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* 189, 114208. DOI: 10.1016/j.bcp.2020.114208 – IF: 5.858 – **Chapter 6**

Carvalho, J.; **Santos, T.**; Carrilho, R.; Sousa, F.; Salgado, G.F.; Queiroz, J.A. and Cruz, C. (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286. DOI: 10.1080/07391102.2019.1632743 – IF: 3.392 – **Chapter 7**

Santos, T.; Pereira, P.; Campello, M.P.C.; Paulo, A.; Queiroz, J.A.; Cabrita, E. and Cruz, C. (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479. DOI: 10.1016/j.ejpb.2019.07.017 – IF: 5.571 – **Chapter 8**

Santos, T.; Miranda, A.; Imbert, L.; Jardim, A.; Caneira, C.R.F.; Chu, V.; Conde, J.P.; Campello, M.P.C.; Paulo, A.; Salgado, G.; Cabrita, E.J. and Cruz. C. (2022) Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin. *Eur. J. Pharm. Sci.* 169, 106093. DOI: 10.1016/J.EJPS.2021.106093 – IF: 4.384 – **Chapter 9**

[‡] Equal contribution

Publications not included in this thesis:

Santos, T.‡; Lopes-Nunes, J.‡; Alexandre, D.; Miranda, A.; Figueiredo, J.; Silva, M. S.; Mergny, J-L. and Cruz, C. (2022) Stabilization of a DNA aptamer by ligand binding. *Biochimie* 200, 8–18 – IF: 4.079

Ramos, C. I. V.; Almodôvar, V. A. S.; Candeias, N.; **Santos, T.**; Cruz, C.; Neves, M. G. P. M. S. and Tomé, A. C. (2022) Diketopyrrolo[3,4-c]pyrrole derivative as a promising ligand for the stabilization of G-quadruplex DNA structures. *Bioorg. Chem.* 122, 105703 – IF: 5.275

Carvalho, J.; Lopes-nunes, J.; Figueiredo, J.; **Santos, T.**; Miranda, A.; Riscado, M.; Sousa, F.; Duarte, A.P.; Socorro, S.; Tomaz, C.T.; Felgueiras, M.; Teixeira, R.; Faria, C. and Cruz, C. (2021) Molecular beacon assay development for severe acute respiratory syndrome coronavirus 2 detection. *Sensors* 21, 7015. DOI:10.3390/s21217015 – IF: 3.576

Miranda, A.; **Santos, T.**; Largy, E.; Cruz, C. (2021) Locking up the as1411 aptamer with a flanking duplex: Towards an improved nucleolin-targeting. *Pharmaceuticals* 14, 1–20. DOI:10.3390/ph14020121 – IF: 5.863

Miranda, A.‡; **Santos, T.**‡; Carvalho, J.‡; Alexandre, D.; Jardim, A.; Caneira, C.R.F.; Vaz, V.; Pereira, B.; Godinho, R.; Brito, D.; Chu, V.; Conde, J.P. and Cruz, C. (2021) Aptamer-based approaches to detect nucleolin in prostate cancer. *Talanta* 226, 122037. DOI: 10.1016/j.talanta.2020.122037 – IF: 6.057

Cruz, C.; Pedro, A.Q.; Carvalho, J.; **Santos, T.**; Talhada, D.; Paiva, A.; Queiroz, J.A.; Andrade, M.; Pinto, M.; Montenegro, L.; Delgado, L. and Pereira, P. (2021) Nucleolin as a potential biomarker for canine malignant neoplasia. *Res. Vet. Sci.* 135, 297–303. DOI: 10.1016/j.rvsc.2020.10.004 – IF: 2.534

Santos, T.; Pereira, P.; Queiroz, J.A.; Cruz, C.; Sousa, F. (2019) Plasmid production and purification: An integrated experiment-based biochemistry and biotechnology laboratory course. *Biochem. Mol. Biol. Educ.* 47, 638–643 DOI: 10.1002/bmb.21290 – IF: 1.160

Santos, T.; Brito, A.; Boto, R.; Sousa, P.; Almeida, P.; Cruz, C. and Tomaz, C. (2018) Influenza DNA vaccine purification using pHEMA cryogel support. *Sep. Purif. Technol.* 206, 192–198. DOI: 10.1016/j.seppur.2018.06.002 – IF: 7.312

‡ Equal contribution

Book chapters not included in this thesis:

Cruz, C.; Miranda, A.; **Santos, T.** Aptamer-based microfluidics for circulating tumor cells, Elsevier, Proofs

SECTION I

Chapter 1

Aims and outline of the thesis

Aims and outline of the thesis

While DNA G4s have attracted considerable interest over the last two decades, G4s in non-coding RNAs have recently emerged as significant agents in biology. The formation of G4s has been described in the regulation of RNA metabolism, including the miRNA pathway, where such structures play essential roles in every step of the miRNA biogenesis and function. Indeed, miRNA biogenesis can be regulated by influencing the equilibrium G4-stem-loop of pre-miRNAs. Developing new ligands that can specifically and selectively target those G4s and modulate their structure and function has recently attracted the attention of researchers. However, the structure and function of G4s can also be regulated by endogenous proteins and vice versa. For example, nucleolin is an essential RNA-binding protein with a high affinity and preference for DNA and RNA G4 architectures. The influence of nucleolin on microRNA biogenesis has also been studied, and since some pre-miRNAs have the ability to adopt a G4 structure, it is expected that the protein binds tightly to these structures. Moreover, upon binding to G4s, the structure and functions of nucleolin can be changed, and vice versa. As a result of the biological significance of the nucleolin/G4 partnership, its modulation for diagnostic and/or therapeutic purposes has tremendous potential.

In this context, one of the main goals of this doctoral thesis is the study of G4 sequences in pre-miRNAs, namely pre-miRNAs let 7e, 92b, and 149. The resulting miRNAs have been found dysregulated in several types of cancer. Our focus is on the G4 found in pre-miRNA 149, since it produces miRNA 149-5p and miRNA 149-3p, which act as antitumor or oncogenic miRNAs by targeting several oncogenic genes, including AKT1, SOX2, and FOXM1. Furthermore, the G4 sequence partially overlaps with miRNA 149-3p, and the strategy could be employed at different stages of miRNA biogenesis. The control of biogenesis could be performed by using ligands that influence the equilibrium between G4-stem loop structures. In some cases, the stabilization of the G4 structure could be beneficial, while in others, the opposite approach of destabilization can be useful in the biological context. Thus, we intend to describe ligands' binding modes to G4 structures in the aforementioned pre-miRNAs. Furthermore, since nucleolin is involved in miRNA biogenesis, this thesis aims to characterize the molecular interactions between nucleolin and G4 sequences found in pre-miRNAs, namely pre-miRNA 149. Indeed, one of the prerequisites to unveil the molecular determinants of protein-RNA complexes is to characterize the binding modes and biological interactions. Finally, this work aims to

provide new insights into nucleolin/G4-related pathologies, namely cancer, and applications in diagnosis and treatment.

Therefore, the specific aims of this doctoral thesis are as follows:

1. To study the formation of G4 structures by the putative G4-forming sequences found in pre-miRNAs let 7e, 92b, and 149.
2. To evaluate the binding modes and molecular interactions of the ligands targeting pre-miRNAs.
3. To characterize the partnership nucleolin/G4 and the use of ligands to interfere with the binding and characterize this at structural and cellular levels.
4. To explore potential diagnostic and therapeutic approaches based on the recognition of G4s in pre-miRNAs by nucleolin.

In line with these main global goals, this thesis was developed by following an integrated strategy that involved different studies with clearly defined goals.

In **Chapter 2,** we present the state-of-the-art of the G4 structures found in non-coding RNAs. The presence and biological influence of G4 structures in pri-miRNAs, pre-miRNAs, and miRNAs are discussed in detail. G4s in those regions play important roles in every step of miRNA biogenesis and function. Also, the modulation of G4s in non-coding RNAs by several classes of ligands is discussed. Overall, this chapter aims to provide a complete framework of the different ligands used.

The focus of **chapter 3** is the elucidation of the different biophysical methods used to characterize G4/ligand interactions. The strengths and weaknesses of each technique are pinpointed, and several examples of applications are given. Furthermore, the binding mode and nature of the interactions between G4-forming sequences and their ligand are discussed. Overall, this chapter provides an overview of the most commonly used approaches to characterize G4/ligands. All of these experimental methods have the possibility to be transposed to protein/ligand characterization.

Chapter 4 gives a comprehensive overview of the nucleolin/G4 partnership and its involvement in different pathologies. The structure of nucleolin and its relationship with the function and localization are scrutinized. The options currently used to modulate nucleolin/G4 interaction are also introduced. Therefore, the most promising approaches include the use of ligands to interfere with the partnership. Finally, we discuss the most relevant achievements in the last few years and the future trends in the field.

Chapter 5 explores the targeting of the G4 sequence (5'-GGGCUGAGGUAGGAGG-3') found in pre-miRNA let 7e by eight well-known G4 ligands and nucleolin. This study introduces an approach to target the G4 with two tetrads found in pre-miRNA let-7e and opens up a new framework to control the miRNA biogenesis of miRNA let-7e.

C**hapter 6** provides the biophysical characterization of the G4 sequence (5'-GGGCGGGCGGGAGGG-3') found in pre-miRNA 92b and their interaction with acridine orange derivatives. This study also found the G4 of pre-miRNA 92b as a specific and selective recognition agent of nucleolin in complex samples.

Chapter 7 describes molecular docking and molecular dynamics to fully characterize the predicted binding modes and types of molecular interaction that drive the binding of six well-known G4 binders/ligands towards the G4 sequence (5'-GGGAGGGAGGGACGGG-3') found in pre-miRNA 149.

Chapter 8 focuses on the use of the G4 structure in pre-miRNA 149 as a selective supramolecular carrier of C_8, which shows anti-tumour activity in prostate cancer cells.

Chapter 9 describes the formation of the G4 structure in the full-length pre-miRNA 149 sequence and the possibility of using it as a capture agent for nucleolin in a microfluidic device.

Chapter 10 presents the structural determinants of the interaction between the G4 structure in pre-miRNA 149 with the ligand, C_8 and/or nucleolin.

Chapter 11 summarizes the main findings and achievements of this thesis and introduces novel perspectives on the use of G4s as potential diagnostic and therapeutic agents for nucleolin/G4-related pathologies, namely, cancer.

Chapter 2

Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs

This chapter was published in:
*Figueiredo, J.[‡]; **Santos, T.**[‡]; Miranda, A.[‡]; Alexandre, D.; Teixeira, B.; Simões, P.; Lopes-Nunes, J. and Cruz, C. (2021) Ligands as stabilizers of G-quadruplexes in non-coding RNAs. Molecules 26, 6164. DOI: 10.3390/molecules26206164 – IF: 4.412*

Chapter overview

This chapter presents the state-of-the-art of the different G4 ligands used to bind and stabilize the G-quadruplexes located in the different classes of non-coding RNAs. With the recent developments that unveil the function of G4s in those regions, the research in this field promptly evolves, providing increasing evidence of the biological activity of old and new G4 ligands. Therefore, this review aims to provide an updated overview of the current G4 ligands used to bind and stabilize non-coding G4 RNAs and their biological implications. Furthermore, we also discussed the polymorphic nature of RNA molecules, and how this feature could be used to modulate cellular functions. Indeed, it is known that the same RNA sequence can have the ability to adopt multiple conformations under different circumstances. The control of each RNA conformation can open a new opportunity to regulate important cellular processes such as gene expression, miRNA biogenesis, and miRNA-mRNA interactions.

Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs

Abstract

The non-coding RNAs (ncRNA) are RNA transcripts with different sizes, structures, and biological functions that do not encode functional proteins. RNA G4s (rG4s) have been found in small and long ncRNAs. The existence of an equilibrium between rG4 and stem–loop structures in ncRNAs and its effect on biological processes remains unexplored. For example, deviation from the stem–loop leads to deregulated mature miRNA levels, demonstrating that miRNA biogenesis can be modulated by ions or small molecules. In light of this, we report several examples of rG4s in certain types of ncRNAs, and the implications of G4 stabilization using small molecules, also known as G4 ligands, in the regulation of gene expression, miRNA biogenesis, and miRNA–mRNA interactions. Until now, different G4 ligand scaffolds have been synthesized for these targets. The regulatory role of the above-mentioned rG4s in ncRNAs can be used as novel therapeutic approaches for adjusting miRNA levels.

Keywords

Non-coding RNAs; miRNA biogenesis; RNA G-quadruplex; ligands

Introduction

The ncRNAs are RNAs that are not translated into proteins and are divided into two main categories based on their size: small ncRNAs with size < 200 nt. (e.g., miRNA, piRNA, and sRNAs) and long ncRNAs with size ≥ 200 nt. (e.g., lincRNA and NAT) [1]. ncRNAs can fold into complex structures and interact with proteins, DNA, and other RNAs, modulating the activity of DNA targets and multiprotein complexes [2]. Thus, ncRNAs can be regulators of different diseases (cancer, cardiovascular and neurological disorders) and open new avenues to novel therapeutic approaches [3,4]. More recently, RNA G4s (rG4) were found in both types of ncRNAs, and their dysregulation has been proposed to impact disease [**5**]. rG4s are thermodynamically stable secondary structures in which the guanines, which are linked via Hoogsteen hydrogen bonds, engage the N1, N7, O6, and N2 atoms from each base **(Figure 2.1)** [6,7].

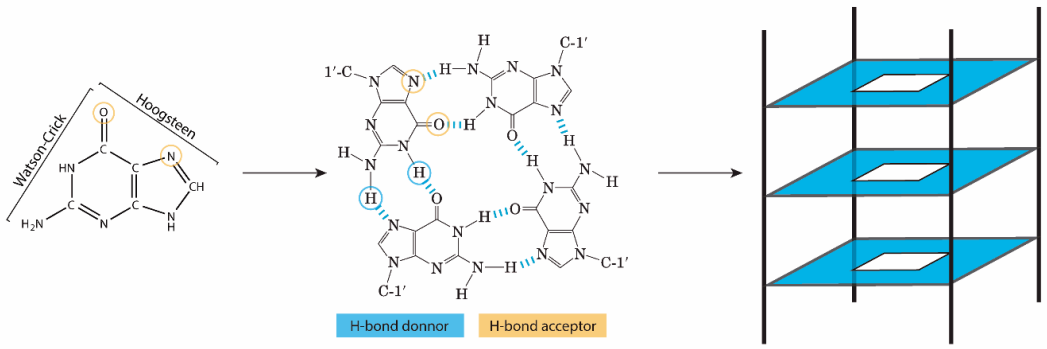

Figure 2.1. Illustration of G-tetrad formation through Hoogsteen bonds and G-tetrads stacking.

The stability of the negatively charged core of the G-quartet composed of O_6 atoms, the coordination between the G-quartets, and the base—stacking interactions are governed by intercalated monovalent cations [7,8,9,10].

In case of cation absence, sequence folding towards G4 conformation is electronically unachievable, because of the negative regions of the G-quartets [10,11]. Moreover, G4's folding and stability are directly correlated with the cation species, inherent properties, and the working concentration [12]. K^+ and Na^+ are the most common cations used for G4 enhancement, due to their biological role and amount in cells [13]. K^+ has a large ionic radius, which makes it able to settle in between multiple G-quartet layers; comparatively, Na^+ is a smaller cation capable of linking up within the core of individual G-quartets [11,14,15]. Further cations, such as Li^+ and Mg^{2+}, can behave as destabilizing or neutral ions in regard to G4 formation [11].

Parallel G4 topology prevails in rG4s, wherein all strands participating in the G4 network follow the same direction **(Figure 2.2A)**. Upon half the strands presenting a reverse orientation, G4 is defined as antiparallel. Once only one strand is in a contrary direction, the structure is classified as a hybrid G4 **(Figure 2.2A)** [16,17]. The 2' hydroxyl group in rG4s generates extra interactions compared to DNA G4s, leading to a more stable, compact, and less hydrated profile in rG4s [18] **(Figure 2.2B)**.

In addition, the folding of rG4 can also be controlled by small-molecule ligands [11,14,15,19,20], suggesting that rG4 can be druggable, being used *in vivo* to modulate ncRNA biological functions. Usually, the ligands' structure is planar chromophores, which bind rG4 via π–π stacking to a terminal G-quartet, and is composed of one or more flexible substituents with a cationic charge that binds G4 grooves and loops, that is: (i) fused aromatic polycyclic systems (e.g., berberine, quarfloxin, PhenDC3, RHPS4, and BRACO-19), (ii) macrocycles (e.g., telomestatin, pyridostatin and carboxypyridostatin), and (iii) non-fused aromatic systems with flexible structural motifs.

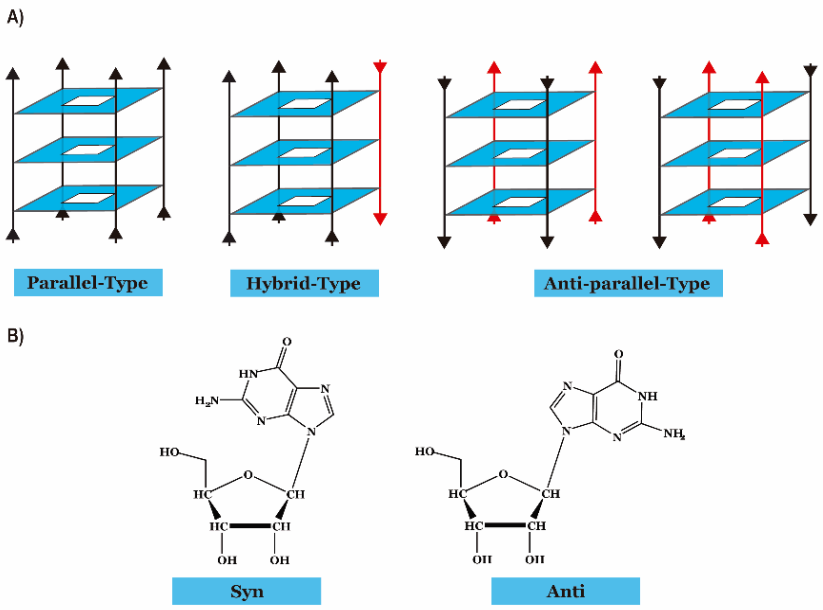

Figure 2.2. (A) Schematic representation of different G4 topologies (parallel, hybrid, and anti-parallel). **(B)** Conformations of G4s according to strands polarity when polarity differences are associated with an angle between the G-bases and the pentose, namely to the 2' hydroxyl group.

Biophysical studies highlighted several parameters influencing rG4 conformations, including the number of G-quartet stacks, the length/sequence composition of the loops, the occurrence of bulges, the availability/nature of the central ion, the sequence in flanking regions, and the ligands interaction/stabilization of the rG4s [12]. Moreover, computational algorithms and next-generation sequencing (NGS), as well as the use of fluorescent light-up probes, have highlighted the location and biological functions of rG4s in untranslated regions (UTR) of mRNA sequences [21], and more recently in ncRNAs [22]. This was key to driving subsequent functional analyses that revealed the biological relevance of rG4s in the post-transcriptional control of gene expression impacting cellular processes. Although those computational predictions have been helpful in proposing the importance of G4s in the transcriptome, caution should be taken, since this could lead to false positives or negatives [23]. Therefore, the use of *in vitro* and *in vivo* approaches has been growing in the last few years to probe and validate the existence of those structures.

Balasubramanian and co-workers developed an approach called rG4-seq, which is a profiling method that couples rG4-mediated reverse transcriptase stalling with NGS [24]. Firstly, the technique was employed *in vitro* in RNA extracted from HeLa cells, and later *in vivo* in bacteria and plants. Later, Monchaud and co-workers developed an approach called G4RP-seq, which combines rG4-specific precipitation (G4RP) with sequencing, to identify rG4s in human cells [25] **(Figure 2.3)**. The presence of rG4s *in vivo* was also proved by using the G4-specific antibody BG4 and some fluorescent light-

up probes, which allowed for the tracking of the folding/unfolding of G4s in living cells [26].

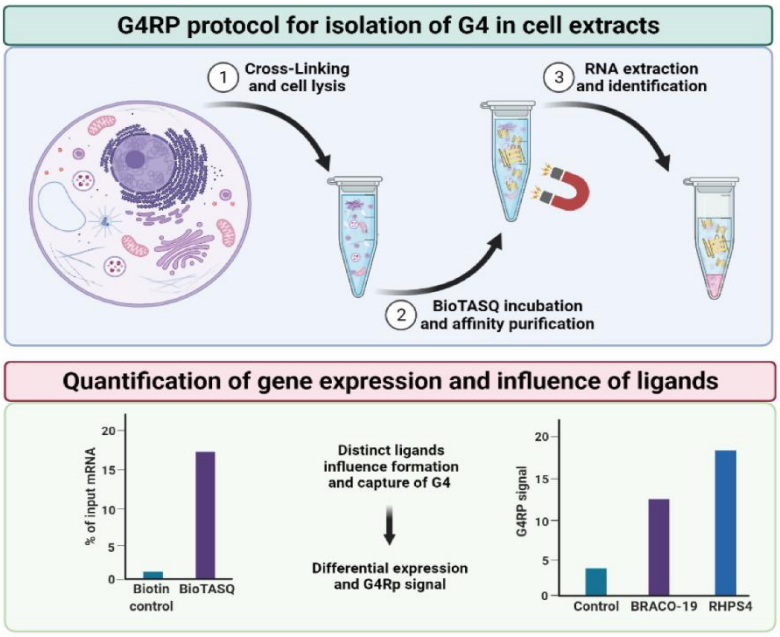

Figure 2.3. Illustration of G4RP method protocol for isolation of G4 targets from human cell extracts and the influence of the ligands in gene expression. Adapted from ref. [25]. Copyright (2018), with permission from Springer Nature.

In those studies, it was observed that rG4s are significantly present in mRNAs (5′- and 3′-UTRs, and introns of pre-mRNAs), as well as in many classes of long and short ncRNAs, which gained significant attention in the last few years due to its ability to control several crucial biological processes, such as transcription and gene expression. An interesting review of Richter and co-workers has recently described the biological relevance and therapeutic potential of G4 structures in the human noncoding transcriptome [5]. An increasing number of reports on the presence of rG4s in ncRNAs, such as microRNAs (miRNAs) and long ncRNAs, suggest that rG4 mediated the regulation of gene expression and the control of miRNA biogenesis, opening a wide range of possibilities for new therapeutic approaches [5].

Herein we report recent findings suggesting that rG4s exist in the folded conformation of ncRNAs in living cells by using small molecules, termed rG4 small molecules ligands, which are able to modulate rG4 conformation **(Table 2.1)**. Furthermore, we will explore its therapeutic potential to interfere with disease-associated ncRNAs.

In the following sections, we detail the presence of rG4s in some types of ncRNAs and their implications in the biological context.

rG4s in Telomere Long ncRNAs

rG4s have been reported in telomere-associated long ncRNAs [27,28]. Telomeres are essential to protect chromosome ends and impose a finite lifespan on cells and tissues. Dysfunctional telomeres cause severe chromosome instability, thereby unleashing cascades of cellular reactions that are common hallmarks of cancer and premature aging [29]. Indeed, upon extensive shortening, critically short telomeres accumulate in cells and emanate an irreversible DNA damage signal, causing permanent growth arrest and, eventually, cell death. To gain unlimited replicative potential, 85–90% of human cancer cells reactivate the reverse transcriptase telomerase, which utilizes an associated RNA moiety to produce telomeric DNA [30,31]. The remaining 10–15% of human cancers elongate telomeres through homology-directed repair pathways collectively known as Alternative Lengthening of Telomeres (ALT) [32,33]. ALT can thus be considered a specialized DNA repair mechanism, assuring cell immortality. Several studies have been conducted to understand how telomeres execute their protective functions and how telomeric dysfunctions are mechanistically linked to pathological conditions. Indeed, the fact that hTERC forms rG4s structures at the 5′-terminal has several implications in telomerase activity and the function of cancer cells [28,32].

Lacroix and co-workers reported that G4 formation interferes with P1 helix formation in hTERC, a critical structural element of telomerase activity. On the other hand, DHX36, an RNA helicase member of the DEXH box family, can bind and unwind hTERC rG4s in the presence of ATP, which enables the formation of the P1 helix structure necessary for telomerase function [27]. HnRNP F/H possesses three quasi-RNA recognition motifs (qRRM) [34], which preferentially bind to poly(G)-rich sequences RNA and were reported to bind G4 sequences in the 5′ tract of hTERC to regulate telomerase activity and telomere length. However, whether other RNA binding proteins exist that can associate with the 5′-end region of hTERC to regulate telomerase function is still elusive [34].

Another telomere-associated lncRNA is the telomeric repeat-containing RNA (TERRA), which was also reported to form rG4s and be involved in the regulation of telomerase activity and DNA telomere length [35]. TERRA RNA is localized to chromosome ends in the nucleus, suggesting a link between TERRA rG4s and telomere function [36]. Several biological functions have been associated with TERRA tandem repeats, namely, modulation of heterochromatin regulation, telomerase inhibition, telomere length regulation, and telomere protection. For instance, TERRA rG4s can interact with telomeric DNA G4s to form intermolecular hybrid G4 structures and suppress

telomerase activity [35]. A study of Balasubramanian and co-workers showed that TERRA recruits TRF2, an important protein that regulates the association of TERRA and telomeric DNA [37]. Recently, a study developed by Lieberman and co-workers reported that the G4 structure of TERRA is a recognition element for the TRF2 GAR domain [37,38]. They reported that the interaction between TERRA and TRF2 GAR domain is of utmost importance to maintain telomere stability and regulation. In their study, they used the well-known G4 ligand N-methyl mesoporphyrin IX (NMM: represented in **Figure 2.4**) to disrupt the TERRA-TRF2 GAR complex [39]. Only NMM showed preferential binding to TERRA. The results showed the loss of TERRA, and the induction of γH2AX-associated telomeric DNA damage associated with decreased telomere length, and increased telomere aberrations **(Figure 2.5)** [39].

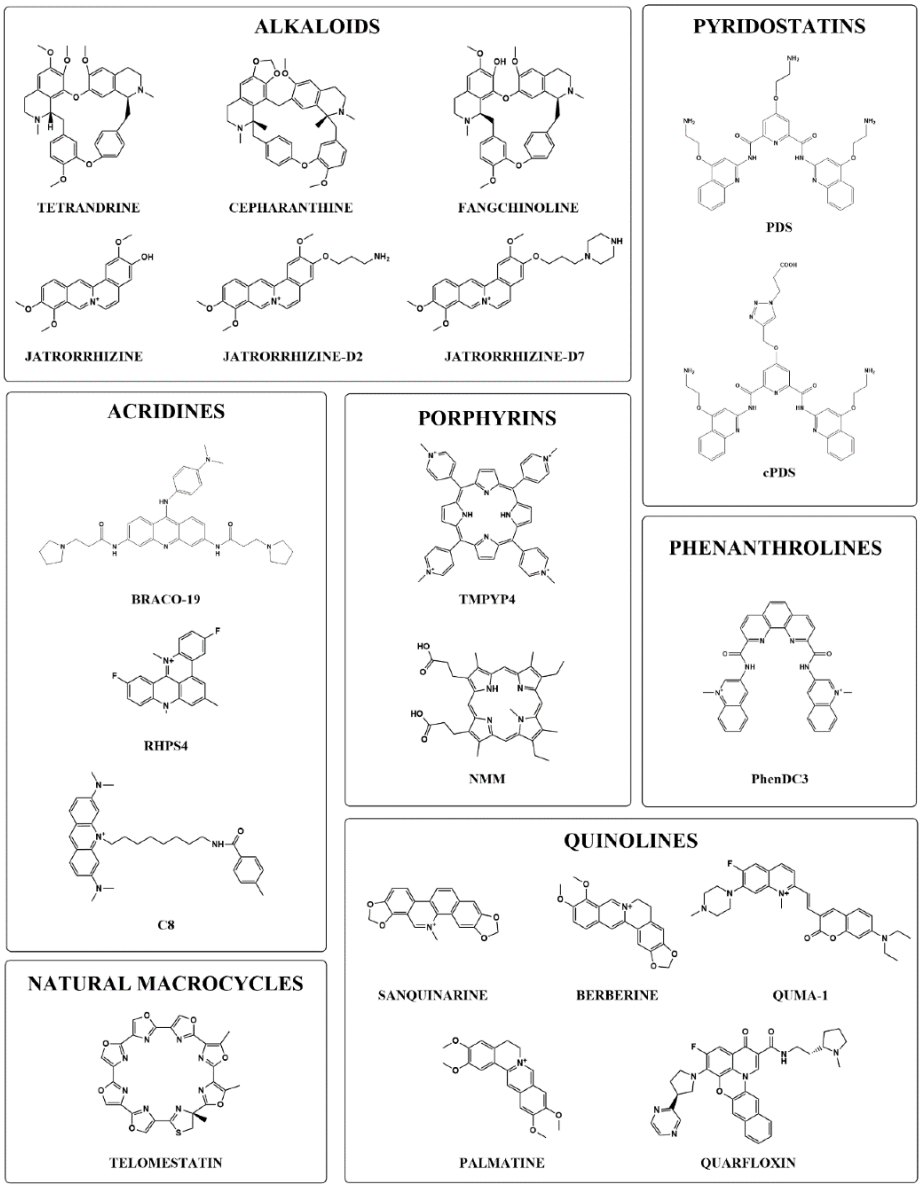

Figure 2.4. Chemical structures of the G4 ligands presented in this review. Small molecules are divided according to their chemical families. Structures were designed with ChemDraw 20.0 ®—PerkinElmer.

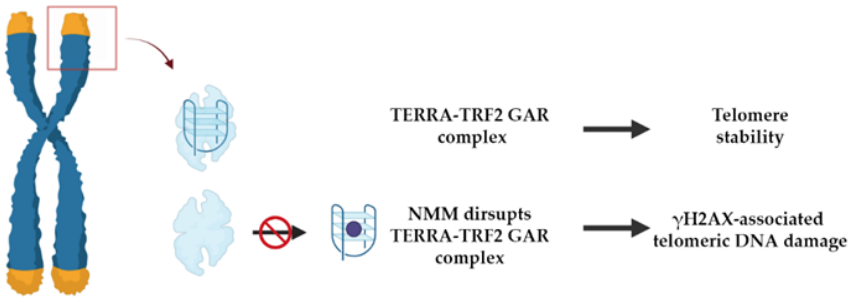

Figure 2.5. Schematic representation of TERRA rG4 interaction with TRF2 GAR domain that is required for telomere stability and integrity. Treatment of human melanoma cells with N-methyl mesoporphyrin IX (NMM) leads to disruption of TERRA and the induction of γH2AX associated telomeric DNA damage.

rG4s in pri-miRNAs

miRNAs are produced by a highly coordinated series of enzymatic cleavages from pri-miRNA to its mature form miRNA via its premature form (pre-miRNA) [40]. There are a few reports which have described G4s in pri-miRNAs compared to those found in pre-miRNAs, which have paved the way to improving the knowledge on pri-miRNA processing and the outcome in the miRNA biogenesis pathway [5].

Perreault et al. identified G4 located near the Drosha cleavage site in three distinct pri-miRNAs (pri-mir200c, pri-mir451a, and pri-mir497) by Reverse Transcription Stalling (RTS) [41]. The G4 folding in pri-miRNAs was unstable, being detected by *in vitro* methods only in the presence of ligands cPDS and PhenDC3 **(Figure 2.4)**. Nevertheless, the authors showed that mutations disrupting G4 folding had an impact on the processing of the pri-miRNAs, leading to low levels of miRNAs. These were explained by the dynamic shifts between G4 and hairpin folding, which influences the pri-miRNA processing [41]. Considering the tumor suppressor nature of the mentioned pri-miRNAs, this approach could be further explored to develop novel anticancer therapies. Moreover, Perreault et al. showed that binding small antisense oligonucleotides to the pri-miRNA can modulate mature miRNA levels [41]. Considering G4's impact on pri-miRNA, a processing analysis of potassium ion dependence could be the best starting point when identifying and validating G4 formation in pri-miRNAs.

A recent report by Ming Xu showed that G4 formation *in vitro* and cells prevented Drosha-DGCR8 binding and processing of the pri-miR, suppressing the biogenesis of the three miRs (miR-23b/27b/24-1) which **(Figure 2.6A)** are related to cardiac disease [42]. The formation of rG4 could negatively regulate the production of the miRNAs encoded by the cluster and, in the presence of tetrandrine, a significant reduction of all three miRNAs occurs. Disruption of this intragenic G4 in these pri-miRs increased the

production of all three miRs. Interestingly, tetrandrine also affects the expression levels of pre-miRNAs instead of pri-miRNAs, suggesting that its mechanism of action in miRNA biogenesis takes place during the generation of pre-miRNA from pri-miRNA [42] **(Figure 2.6B)**. Conversely, the binding of ligand tetrandrine to G4 stabilized the structure and suppressed miRs production in human and rodent cardiomyocytes **(Figure 2.5 and Figure 2.6)** [42].

However, to date, these ligands did not recognize just one G4; and for these reasons, finding molecules specific for G4 in the pri-miRNA structure may prove more difficult, but the field is underexplored, and thus applications such as microarrays or high-throughput screening methods towards different rG4 structures may provide new insights for G4 in pri-miRNAs.

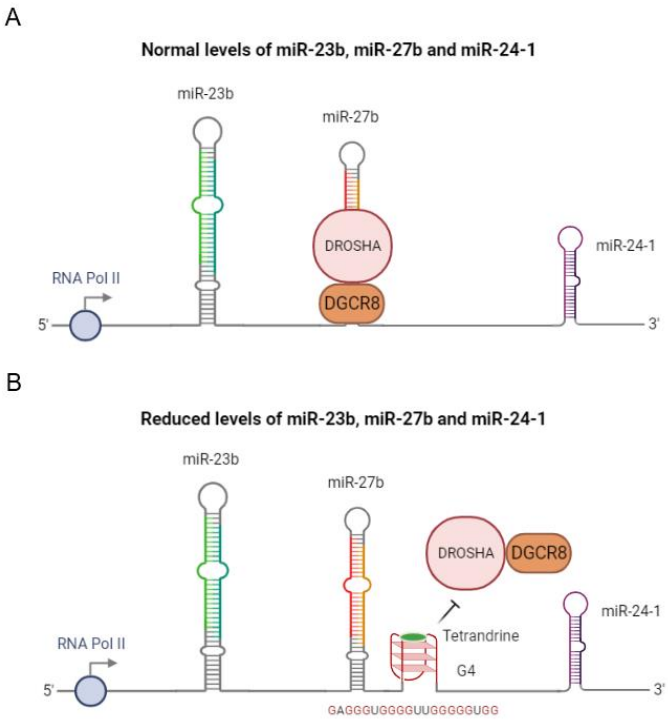

Figure 2.6. Representation of the influence of G-rich sequence in normal **(A)** and reduced **(B)** levels of miR-23b, miR 27b, and miR-24-1. Adapted from ref. [42]. Copyright (2021), with permission from Oxford University Press.

rG4s in pre-miRNAs

In the last few years, the presence of G4 structures in pre-miRNAs has been studied more, comparing their presence to those found in pri-miRNAs. In 2015, bioinformatics predictive tools, using a QGRS mapper [43], suggested the presence of at least one rG4 motif in 16% of human pre-miRNA stem–loop regions [44]. Recently, Perreault and co-

workers, using a new approach, reported that only 2% of pre-miRNA contain a putative rG4 motif [22]. This analysis confirms a competition between G4, and the secondary structures recognized by Dicer identifying that rG4 overlapping the Dicer cleavage site in 9% of the miRNAs. In this sense, several studies have reported the presence of the equilibrium between G4 and stem–loop in pre-miRNA and its effect in the biological process [22].

Arachchilage et al. reported the presence of a very stable G4 structure in the human pre-miR-92b [44]. The human miR-92b is a clinically important miRNA and is significantly upregulated in several human cancers [45,46,47,48]. The G4 structure of pre-miR-92b is very stable and has six G stretches, each containing three G4. Using several biochemical and biophysics approaches, the G4 motif in pre-miR-92b was proven in cytoplasmatic K^+ concentrations **(Figure 2.7)**. The presence of the rG4 structure exists in equilibrium with the canonical stem–loop structure, and this equilibrium regulates the maturation of pre-miR-92b [44]. In this sense, a rational design of locked nucleic acid (LNA) able to bind specifically the rG4 motif conformation of pre-miR-92b was carried out [49]. The LNA treatment demonstrated the importance of increasing PTEN levels in NSCLC through suppression by miR-92b maturation [49].

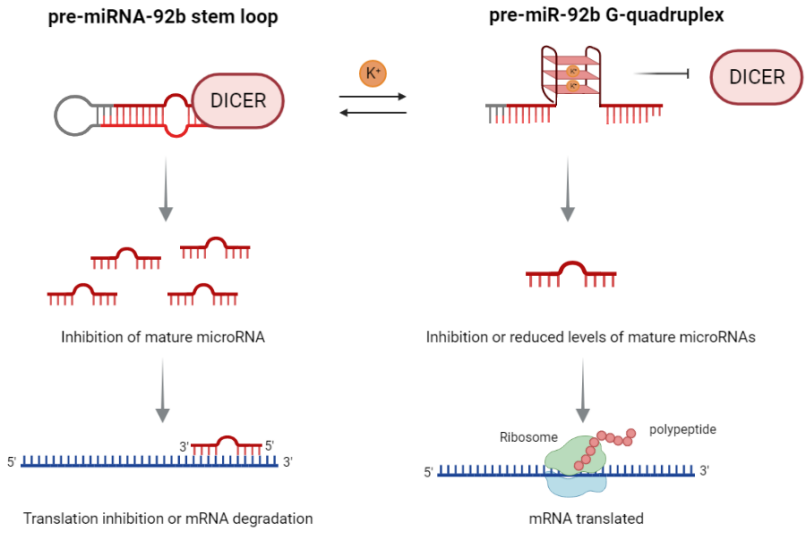

Figure 2.7. Representation of the ionic profiling to induce G4 formation by unwinding the stem–loop of pre-miR-92b Adapted from [44] copyright (2015), with permission from Elsevier.

Cruz and co-workers used a labeled pre-miR-92b sequence as a molecular recognition probe of a microfluidic platform for the recognition of the nucleolin (NCL), an important biomarker, in biological samples. The G4 motif present in pre-miR-92b was used for the molecular detection of NCL. The additional stability of the rG4 structure was provided

by ligand C_8 **(Figure 2.4)** (Please see Chapter 6) [50]. C_8 is an acridine orange derivative [51] that has been reported as a promising G4 binder [52,53,54,55].

Additionally, Pandey et al. used the pre-miR-let-7e to study the equilibrium between G4 structure and stem–loop in RNA in a more biologically relevant environment [56]. Pre-miR-let-7e contains a G2 with a total loop length of seven nucleotides, and the Dicer cleavage assay suggests a role for the G4 structure at physiologically relevant concentrations of Mg^{2+} and K^+, leading to the reduction of miR-let-7e levels. To study the importance of the potassium-dependent G4 structure, cells were treated with TMPyP4, leading to an increase in miRNA levels. It is probable that TMPyP4 promotes the disruption of G4 and shifts the equilibrium between G4 and the stem–loop [56].

Balasubramanian and co-workers used a new approach to map the G4 formation in RNAs. This new method, called SHALiPE, couples selective 2′-hydroxyl acylation with a lithium ion-based primer extension, and recognizes the structural characteristics for rG4 mapping. They used SHALiPE to prove the formation of an rG4 structure in pre-miR-149. Additional biophysics studies support the SHALiPE and have proven that rG4 has a parallel topology with high thermal stabilization under physiological K^+ conditions **(Figure 2.8)**. The authors still used the G4 ligand and pyridostatin (PDS) **(Figure 2.4)**, and the results provided strong evidence that pre-miR-149 can form G4 structure in the presence of K^+ and PDS. The presence and stabilization of the G4 structure inhibits the Dicer processing activity *in vitro* and decreases the miR-149 production [57].

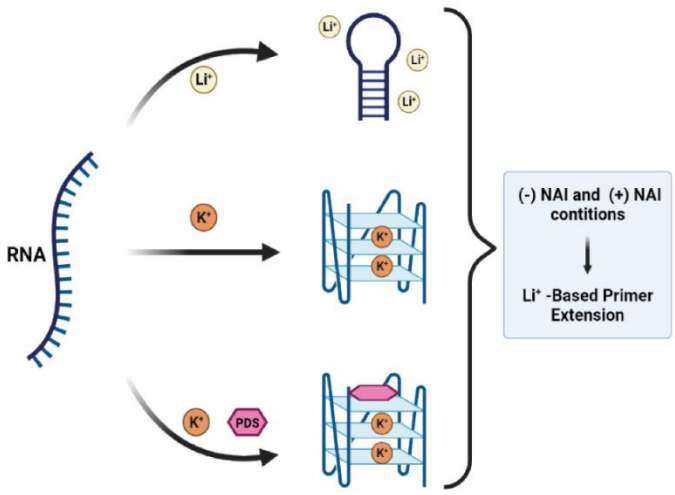

Figure 2.8. Representation of SHALiPE method [57]. Copyright (2016), with permission from Wiley.

Additionally, Cruz and co-workers studied the thermal stabilization of the pre-miR-149 G4 in the presence of two acridine orange derivatives, C_8 and C_8-NH_2 **(Figure 2.4)** [58].

The results showed that the G4 structure can be stabilized by the C_8 derivative with more effectiveness than C_8-NH_2 (Please see Chapter 8, page 215, figure 8.1). These results were anticipated by molecular dynamics in an in silico study (Please see chapter 7) [59]. Ghosh et al. used the porphyrin TMPyP4 **(Figure 2.4)** to destabilize the G4 structure formed from the pre-miR-149. Biophysics studies, such as CD, UV-melting, ITC, and NMR showed that TMPyP4 binds strongly to G4 and unfolds it, and upon addition of TMPyP4, the dynamic change occurs to the hairpin structure, enhancing the transcript levels of mature miR-149 in cells [60].

Mihailescu and co-workers reported the presence of a parallel intramolecular G4 structure with six G-tracts in pre-miR-1229. They used several biophysical techniques to show an equilibrium between the G4 structure and the canonical hairpin. G4 structure mature miR-1229 has been shown to directly control the expression of SORL1, one protein responsible for the processing and trafficking of β-amyloid proteins [61]. Whilst the G4 structure within pre-miR-1229 has not been stabilized by any ligand, it could potentially become a therapeutic target in Alzheimer's disease, through ligands that are able to stabilize the G4 and consequently reduce the levels of mature miR-1229. Additionally, the G4 forming motif could be an important therapeutic target in other pathologies such as colorectal [62] and breast cancer [63].

Similarly, miR-26a has been described in several physiological processes, is emerging as a therapeutic target for human diseases, and generally functions as a tumor suppressor [64,65]. Fu and co-workers identify a guanine-rich sequence in pre-miR-26a-1 that can fold into the metastable G4 structure containing 2-quartet *in vitro* and *in vivo* [66]. This sequence could regulate miR-26a function and expression levels. Using one helicase (DHX36), they are able to bind and unwind the G4, promoting miR-26a maturation. Moreover, the treatment of cancer cell lines with PDS and carboxy pyridostatin (cPDS) **(Figure 2.4)** significantly decreases miR-26a levels, suggesting that G4 structure inhibits endogenous miR-26a [66].

Recently, Cruz and co-workers reported the formation of G4 structure in the human pre-miR-150 [67]. Circular dichroism studies indicated the formation of a parallel G4 in the presence of K^+ and PhenDC3 **(Figure 2.4)**. Biophysics studies have shown that the pre-miR150 G4-forming sequence recognizes and interacts with the NCL protein. This sequence still shows co-localization with NCL in a lung cancer cell line and in peripheral blood mononuclear cells [67]. Considering that human miR-150 plays an important role in several cancer developments [68], the G4-forming region found in precursor pre-miR-150 can be an attractive therapeutic target.

rG4s in miRNAs

rG4s play roles in every step of the miRNA biogenesis and function **(Figure 2.9)** [5]. Pri-mRNA rG4s regulate Drosha-mediated processing, pre-miRNAs rG4s inhibit DICER-mediated maturation, and miRNA rG4s abolish its loading onto RISC **(Figure 2.9)** [69,70,71]. Understanding the mechanisms that control miRNA biogenesis may allow for the development of tools to modulate the expression of specific miRNAs, which is crucial for developing novel therapies for human disorders derived from the aberrant expression of miRNAs.

Several studies have demonstrated the *in vitro* formation of rG4s in miRNAs and the competition with the RNA stem–loop conformation, which have crucial roles in post-transcriptional regulation [27,70,71]. Indeed, these studies revealed the important role that rG4 structures have in impairing miRNA–mRNA interaction [72]. However, the current structural information and cellular roles of the rG4s in miRNAs remain elusive [73]. The latest release of miRBase (version 22) deposited 38,589 hairpin precursors and 48,860 mature microRNAs from 271 organisms [74]. So far, by bioinformatics analysis, 166 rG4-positive human miRNAs sequences (6%) using two different G4 prediction tools (G4-detection algorithm G4NN and the Quadparser-based $(G_{2+}N_{1-12})_3G_2$) were identified [75]. In the last few years, the miRNA rG4 formation, and targeting by different G4 ligands, revealed a new possible mechanism for regulating microRNA functions with therapeutic potential.

The first research relating the rG4 in miRNA and ligands was published in 2015 by Ming Xu and collaborators [76]. This study evaluated the formation of G4 in the genomic cluster near three distinct microRNAs, namely, miR-23b, miR-27b, and miR-24 (cluster miR-23b-27b-24-1). The miR clusters can display tumor-suppressive or oncogenic roles depending on their context, as well as the individual members of the cluster that are able to adopt suppressor or oncogenic promoter behaviors according to the cancer type [77]. The authors found three potential G4 sequences capable of folding in antiparallel, and one in parallel conformations and tested six small molecules (tetrandrine, fangchinoline, palmatine, jatrorrhizine, berberine and cepharanthine; **Figure 2.4**) [76]. Surprisingly, some of the tested ligands (tetrandrine, fangchinoline, and cepharanthine) evidenced selective binding affinity to parallel rG4 instead of antiparallel conformations. This effect can be explained due to the unusual chemical structure (non-planar molecules) of these natural alkaloids, which allows for binding via locating on the groove formed by the loops [78].

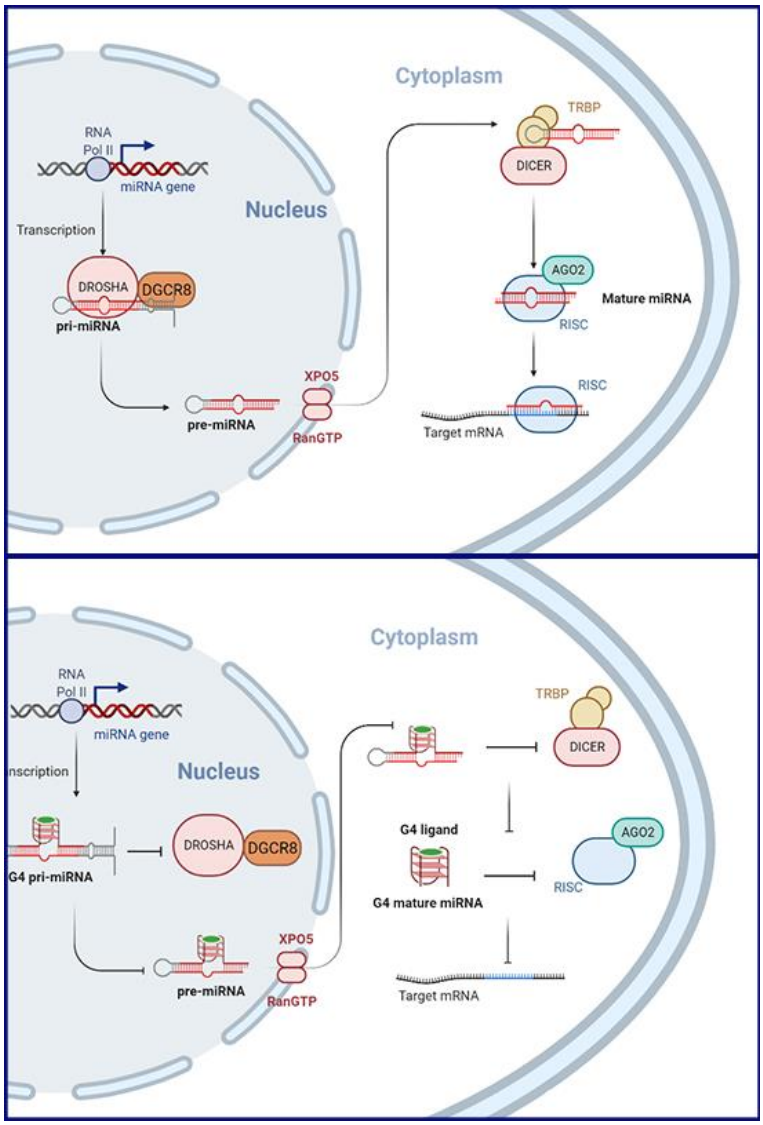

Figure 2.9. Biogenesis of miRNAs. Comparison of natural biogenesis versus interference of the rG4s in the biogenesis and function of miRNAs.

Yuan et al. published a second study reporting the formation and folding of a stable parallel rG4 in miR-3620-5p in physiological conditions by using a set of biophysical techniques such as ESI-MS, CD, NMR, and SPR [73] **(Figure 2.10)**. This miR is reported to be associated with different kinds of cancers, such as gastric cancer and metastasis [79]. Additionally, the same authors concluded that sanguinarine, a natural alkaloid used in traditional Chinese medicine, binds and stabilizes the rG4 in miR-3620-5p [73]. Furthermore, the study revealed that this interaction among rG4 and sanguinarine can block the base pairing between miR-3620-5p and its target sequence, proposing a mechanism to regulate miRNA function, affecting the miRNA–mRNA interactions through the miRNA rG4 formation and targeting by ligands [73].

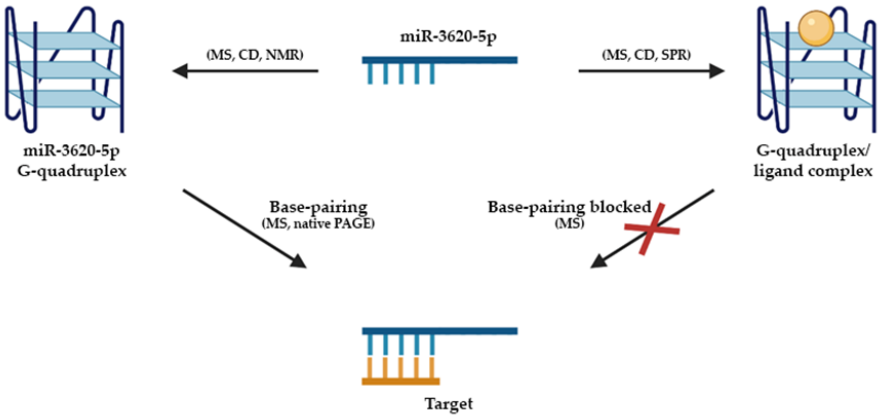

Figure 2.10. Representation of G4 formation in miR-3620-5p and the inhibition effect of sanguinarine on the base pair formation of miR-3620-5p with its target sequence. Adapted from ref. [73]. Copyright (2016), with permission from Elsevier.

Later, the same research group also proved the formation of the rG4 miRNA-1587 [80,81]. This miRNA is responsible for cell proliferation, migration, and tumorigenicity in glioma and breast cancer [82,83]. It is noteworthy that miR-1587 formed a dimeric G4 through the 3′-to-3′ stacking of two monomeric G4 subunits with one NH_4^+ sandwiched between the interfaces [80]. The dimeric rG4 was also adopted under molecular crowding conditions or in the presence of a jatrorrhizine ligand [80]. Another study conducted by Yuan and collaborators, which focused on miR-1587, showed that the formation of a secondary structure affects the ability of miR-1587 to bind the target mRNA sequence and, consequently, reduces its expression levels [81]. This conclusion shows the potential for therapeutic applications. Moreover, the pseudopalmatine ligand prevents interaction with the target gene via G4 stabilization and has the same effect on the suppression and modulation of gene expression. Contrarily, TMPyP4 revealed the disruption of rG4, promoting the binding of miR15-87 to its target, allowing for mRNA processing [81]. In addition, through bioinformatic analysis, the authors identified other miRNAs able to form rG4, namely, miR-197-5p, miR-765, miR-3620-5p, and miR-5196-5p [81].

Using the G4NN algorithm, specially developed to detect RNA G4s, Kwok and co-workers identified up to 478 putative RNA G4 human miRNAs [76]. However, when they increased the stringency by using Quadparser, only 166 human miRNAs were positive for the presence of a putative rG4 structure. From those 166 human miRNAs, they selected four miRNAs (miR-149, miR-197, miR-432, and miR-765), taking into account the biological relevance, conservation, as well as confidence of G4 formation based on G4NN parameters. It is noteworthy that these miRNAs have been linked to several types of cancer. For instance, dysregulation of miRNA-149 has been implicated in cancer

migration and invasion [84]. Therefore, their study provides a strong basis for the involvement of RNA G4s in post-transcriptional regulation and paves the way for future investigations [75]. Moreover, they showed that the NMM ligand stabilizes *in vitro* rG4 of these four miRNAs in a specific manner, modulating gene expression [75].

Zhou et al. identified and characterized rG4 in miR-92a by ESI-MS and CD [85], which is associated with retinoblastoma, cervical and non-small cell lung cancers [86,87,88]. The study evidenced the formation of a parallel rG4 in distinct ionic conditions and adopted a dimeric structure in the presence of NH_4^+. Moreover, the palmatine ligand has a high binding affinity and stabilization effect (increasing its thermal stability by 10 °C) on the dimeric rG4 in miR-92a [85].

Recently, Vivekanandan et al. performed an analysis of G4 upstream of herpesvirus miRNAs, namely in Kaposi's sarcoma-associated Herpesvirus miR-K12–1-9,11 cluster and Human Cytomegalovirus miR-US33 [89]. They also characterized a parallel rG4 in these miRNAs and showed that TMPyP4 and PDS interfered with rG4 stabilization modulating the promoter activities of KSHV miR-K12 and HCMV miR-US33.

Other examples of miRNA rG4s have also been reported [75,69]. A list with the G4 sequences in ncRNA nucleic acids discussed in this review, as well as their ligands, biological target, and effect of the G4 formation is highlighted in **Table 2.1**. Overall, these results proved the influence of rG4s in miRNA biogenesis and consequently in several important biological processes such as regulation of gene expression and transcription.

Table 2.1. List of G4 sequences in ncRNA nucleic acids discussed in this review.

		rG4 forming sequence (5' – 3')	**Ligands**	**Biological Target**	**Effect**	**Refs**
Gene Cluster	mir-23b	GAGGGTGGGGGGTTG GGGGTGG	Tetrandrine	Drosha-DGCR8	Decreased of pri-miR levels	[45]
	mir-27b					
	mir-24-1					
	miR-K12–1-9	GGGTGGGAGGAAGGA TGTGGGGGTGGG	$TmPyP_4$, PDS	promoter activity	Transcriptional regulation of virus encode miRNA.	[92]
Pri-miRNA	Pri-mir200c	GGUGGGCGGGCUGG GCGGGGG	cPDS, PhenDC3	Drosha-DGCR8	Reverse Transcription Stalling	[44]
	Pri-mir451a	GGGCACUUGGGAAUG GCAAGG				
	Pri-mir497	GGAGGGGGUGGG				

Table 2.1. List of G4 sequences in ncRNA nucleic acids discussed in this review. (Continued)

		rG4 forming sequence (5' – 3')	Ligands	Biological Target	Effect	Refs
Pre-miRNA	pre-miR-92b	GGGCGGGCGGGAGGG	LNA, C_8	Dicer	Decrease mature levels of miRNA	[52, 53]
	pre-miR-let-7e		$TmPyP_4$		Increase mature levels of miRNA	[59]
	pre-miR-149	GGGAGGGAGGGACGGGGG	PDS, C_8		Decrease mature levels of miRNA	[60, 61]
	pre-miR-1229	GGGUAGGGUUUGGGGGAG AGCGUGGGCUGGGGUUCAGGGACA		Dicer		[64]
	pre-miR-26a-1	GGAUAGGCUGUGCAGGUCCCAAUGGG	PDS, cPDS		Decrease mature levels of miRNA	[69]
	pre-miR-150	GGCCUGGGGGACAGGGACCUGGG	PhenDC3			[70]
miRNA	miR-3620-5p	GUGGGCUGGGCUGGGCUGGGCC	Sanguinarine	Blocked the base pairing with the target sequence	Depression of miRNAs	[76]
	miRNA-1587	UUGGGCUGGGCUGGGUUGGG	Jatrorrhizine derivatives, pseudopalmatine			[83, 84]
	miRNA-149	AGGGAGGGACGGGGGCUGUGC	NMM			[78]
	miRNA-197	CGGGUAGAGAGGGCAGUGGGAGG				
	miRNA-432	UCUUGGAGUAGGUCAUUGGGUGG				
	miRNA-765	UGGAGGAGAAGGAAGGUGAUG				
	miR-92a	TAATGGGGTGGGGGCTGGGAA	Palmatine			[88]

Conclusions

The bioinformatics tools showed that most of the human transcriptome is composed of ncRNAs with the widespread occurrence of potential G4-forming sequences. The representative roles of ncRNAs rG4 on a myriad of biological processes in cells needs to continue to be verified to illustrate their utility in association with diseases. The potential for exploiting rG4 small molecules to control miRNA-regulated target gene expression

based on high-order rG4 structures must be exploited *in vitro* and *in vivo* to modulate ncRNA biological functions.

Overall, the developments described in this review showed that the research on ncRNA G4s needs to be protracted, and the transient nature of the rG4 structures at the cellular level, which are quite unstable, can be surpassed by the presence of small molecules. Indeed, multiple overlapping structures can exist within the same RNA molecule, and these structures may undergo equilibrium shifts as per the cellular requirement. In this context, the use of small molecules that selectively recognize G4s over other structures may prove effective. However, their regulatory role over ncRNA targets and the effect on resultant protein products remain to be elucidated. However, the research on rG4s is expanding, and new rG4 small molecules with new functions at the cellular level will be available, allowing the development of novel therapeutic approaches in pathological rG4 ncRNAs-linked contexts.

Funding

This work was supported by the project UIDP/00709/2020, PESSOA program ref. 5079 and project “Projeto de Investigação Exploratória” ref. IF/00959/2015 entitled “NCL targeting by G-quadruplex aptamers for cervical cancer therapy” financed by Fundo Social Europeu e Programa Operacional Potencial Humano. Thanks are due to FCT/MCT for the financial support to CICS-UBI UIDB/00709/2020 research unit and the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016), through national funds and, where applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and PIDDAC.

Acknowledgments

Joana Figueiredo acknowledges a doctoral fellowship grant from the Foundation for Science and Technology (FCT) ref. SFRH/BD/145106/2019. Daniela Alexandre and André Miranda acknowledge the FCT fellowship ref. 2021.07695.BD and ref. PINFRA/22161/2016-B4, respectively. Tiago Santos acknowledges FCT for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013). Jéssica Lopes-Nunes acknowledges a doctoral fellowship grant from FCT ref. 2020.05329.BD. C. Cruz acknowledges the grant from FCT ref. UIDP/00709/2020.

References

1 Baptista, B. *et al.* (2021) Non-coding RNAs: Emerging from the discovery to therapeutic applications. *Biochem. Pharmacol.* 189, 114469

2 Zampetaki, A. *et al.* (2018) Long Non-coding RNA Structure and Function: Is There a Link? *Front. Physiol.* 0, 1201

3 Anastasiadou, E. *et al.* (2018) Non-coding RNA networks in cancer. *Nat. Rev. Cancer* 18, 5–18

4 Peng, Y. *et al.* (2019) Cancer and non-coding RNAs. In *Nutritional Epigenomics* pp. 119–132, Elsevier

5 Tassinari, M. *et al.* (2021) Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* 49, 3617–3633

6 Belmont, P. *et al.* (2001) Nucleic acid conformation diversity: from structure to function and regulation. *Chem. Soc. Rev.* 30, 70–81

7 Spiegel, J. *et al.* (2020) The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* 2, 123–136

8 Kaushik, M. *et al.* (2016) A bouquet of DNA structures: Emerging diversity. *Biochem. Biophys. Reports* 5, 388–395

9 Sharawy, M. and Consta, S. (2018) Effect of the chemical environment of the DNA guanine quadruplex on the free energy of binding of Na and K ions. *J. Chem. Phys.* 149, 225102

10 Davis, J.T. (2004) G-Quartets 40 Years Later: From 5′-GMP to Molecular Biology and Supramolecular Chemistry. *Angew. Chemie - Int. Ed.* 43, 668–698

11 Bhattacharyya, D. *et al.* (2016) Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* 4, 1–14

12 Carvalho, J. *et al.* (2020) G-quadruplex, Friend or Foe: The Role of the G-quartet in Anticancer Strategies. *Trends Mol. Med.* 26, 848–861

13 Neidle, S. and Balasubramanian, S. (2006) *Quadruplex Nucleic Acids*, (1st edn) Cambridge University Press.

14 Huppert, J.L. (2008) Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* 37, 1375

15 Mergny, J.L. and Sen, D. (2019) DNA quadruple helices in nanotechnology. *Chem. Rev.* 119, 6290–6325

16 Popenda, M. *et al.* (2020) Topology-based classification of tetrads and quadruplex structures. *Bioinformatics* 36, 1129–1134

17 Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* 43, 8627–8637

18 Prachi Agarwala *et al.* (2015) The tale of RNA G-quadruplex. *Org. Biomol. Chem.* 13, 5570–5585

19 Fujii, T. *et al.* (2017) Effects of metal ions and cosolutes on G-quadruplex topology. *J. Inorg. Biochem.* 166, 190–198

20 Sanchez-martin, V. *et al.* (2021) Quadruplex Ligands in Cancer Therapy. *Cancers (Basel)*. 13,

21 Huppert, J.L. *et al.* (2008) G-quadruplexes: The beginning and end of UTRs. *Nucleic Acids Res*. 36, 6260–6268

22 Vannutelli, A. *et al.* (2020) Where are G-quadruplexes located in the human transcriptome? *NAR Genomics Bioinforma*. 2, lqaa035

23 Puig Lombardi, E. and Londoño-Vallejo, A. (2020) A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res*. 48, 1–15

24 Kwok, C.K. *et al.* (2016) RG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* 13, 841–844

25 Yang, S.Y. *et al.* (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun*. 9, 4730

26 G, B. *et al.* (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem*. 6, 75–80

27 Lyu, K. *et al.* (2021) RNA G-quadruplexes (rG4s): Genomics and biological functions. *Nucleic Acids Res*. 49, 5426–5450

28 Xu, Y. *et al.* (2008) G-Quadruplex Formation by Human Telomeric Repeats-Containing RNA in Na + Solution. *J. Am. Chem. Soc*. 130, 11179–11184

29 O'Sullivan, R.J. and Karlseder, J. (2010) Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* 11, 171–181

30 Deng, Y. and Chang, S. (2007) Role of telomeres and telomerase in genomic instability, senescence and cancer. *Lab. Investig*. 87, 1071–1076

31 Ram, R. *et al.* (2005) The importance of the telomere and telomerase system in hematological malignancies. *Leuk. Lymphoma* 46, 1121–1135

32 Xu, Y. *et al.* (2010) Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci.* 107, 14579–14584

33 Cong, Y.-S. *et al.* (2002) Human Telomerase and Its Regulation. *Microbiol. Mol. Biol. Rev*. 66, 407–425

34 Xu, C. *et al.* (2020) HnRNP F/H associate with hTERC and telomerase holoenzyme to modulate telomerase function and promote cell proliferation. *Cell Death Differ*. 27, 1998–2013

35 Martadinata, H. and Phan, A.T. (2013) Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K + Solution. *Biochemistry* 52, 2176–2183

36 Cusanelli, E. and Chartrand, P. (2015) Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front. Genet*. 6, 143

37 Biffi, G. *et al.* (2012) An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2. *J. Am. Chem. Soc.* 134, 11974–11976

38 Deng, Z. *et al.* (2009) TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol. Cell* 35, 403–413

39 Mei, Y. *et al.* (2021) TERRA G-quadruplex RNA interaction with TRF2 GAR domain is

required for telomere integrity. *Sci. Rep*. 11, 3509

40 Ha, M. and Kim, V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524

41 Rouleau, S.G. *et al.* (2018) G-Quadruplexes influence pri-microRNA processing. *RNA Biol.* 15, 198–206

42 Zhu, M. *et al.* (2021) Novel roles of an intragenic G-quadruplex in controlling microRNA expression and cardiac function. *Nucleic Acids Res.* 49, 2522–2536

43 Kikin, O. *et al.* (2006) QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34, W676–W682

44 Mirihana Arachchilage, G. *et al.* (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* 22, 262–272

45 Li, Y. *et al.* (2013) MiR-92b regulates the cell growth, cisplatin chemosensitivity of A549 non small cell lung cancer cell line and target PTEN. *Biochem. Biophys. Res. Commun.* 440, 604–610

46 Guo, J.H. *et al.* (2020) MicroRNA-92b acts as an oncogene by targeting PTEN/AKT in NSCLC. *Cell Biochem. Funct.* 38, 1100–1110

47 Song, H. *et al.* (2016) miR-92b regulates glioma cells proliferation, migration, invasion, and apoptosis via PTEN/Akt signaling pathway. *J. Physiol. Biochem.* 72, 201–211

48 Ni, Q. *et al.* (2020) miR-92b promotes gastric cancer growth by activating the DAB2IP-mediated PI3K/AKT signalling pathway. *Cell Prolif.* 53, e12630

49 Mirihana Arachchilage, G. *et al.* (2018) Targeting of G-Quadruplex Harboring Pre-miRNA 92b by LNA Rescues PTEN Expression in NSCL Cancer Cells. *ACS Chem. Biol.* 13, 909–914

50 Santos, T. *et al.* (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* 189, 114208

51 Pereira, E. *et al.* (2017) Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA. *Sci. Rep*. 7, 42544

52 Carvalho, J. *et al.* (2018) Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex. *Biochimie* 144, 144–152

53 Josué Carvalho *et al.* (2019) Aptamer-guided acridine derivatives for cervical cancer. *Org. Biomol. Chem.* 17, 2992–3002

54 Figueiredo, J. *et al.* (2019) AS1411 derivatives as carriers of G-quadruplex ligands for cervical cancer cells. *Int. J. Pharm.* 568, 118511

55 Carvalho, J. *et al.* (2019) Aptamer-based Targeted Delivery of a G-quadruplex Ligand in Cervical Cancer Cells. *Sci. Rep.* 9, 7945

56 Pandey, S. *et al.* (2015) The RNA Stem–Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. *Biochemistry* 54, 7067–7078

57 Kwok, C.K. *et al.* (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. *Angew. Chemie - Int. Ed.* 55, 8958–8961

58 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective

delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

59 Carvalho, J. *et al.* (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286

60 Ghosh, A. *et al.* (2019) Restoration of miRNA-149 Expression by TmPyP4 Induced Unfolding of Quadruplex within Its Precursor. *Biochemistry* 58, 514–525

61 Imperatore, J.A. *et al.* (2020) Characterization of a G-Quadruplex Structure in Pre-miRNA-1229 and in Its Alzheimer's Disease-Associated Variant rs2291418: Implications for miRNA-1229 Maturation. *Int. J. Mol. Sci. 2020, Vol. 21, Page 767* 21, 767

62 Hu, H.Y. *et al.* (2019) Exosomal miR-1229 derived from colorectal cancer cells promotes angiogenesis by targeting HIPK2. *Int. J. Biol. Macromol.* 132, 470–477

63 Tan, Z. *et al.* (2016) MicroRNA-1229 overexpression promotes cell proliferation and tumorigenicity and activates Wnt/β-catenin signaling in breast cancer. *Oncotarget* 7, 24076–24087

64 Fu, X. *et al.* (2015) MicroRNA-26a regulates insulin sensitivity and metabolism of glucose and lipids. *J. Clin. Invest.* 125, 2497–2509

65 Fu, X. *et al.* (2013) MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation. *Proc. Natl. Acad. Sci.* 110, 17892–17897

66 Liu, G. *et al.* (2020) RNA G-quadruplex regulates microRNA-26a biogenesis and function. *J. Hepatol.* 73, 371–382

67 Figueiredo, J. *et al.* (2021) Targeting nucleolin by RNA G-quadruplex-forming motif. *Biochem. Pharmacol.* 189, 114418

68 Wang, F. *et al.* (2015) Role of microRNA-150 in solid tumors (review). *Oncol. Lett.* 10, 11–16

69 Michlewski, G. and Cáceres, J.F. (2019) Post-transcriptional control of miRNA biogenesis. *RNA* 25, 1–16

70 Slezak-Prochazka, I. *et al.* (2010) MicroRNAs, macrocontrol: Regulation of miRNA processing. *RNA* 16, 1087–1095

71 Shukla, G.C. *et al.* (2011) MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. *Mol. Cell. Pharmacol.* 3, 83

72 Rouleau, S. *et al.* (2017) 3′ UTR G-quadruplexes regulate miRNA binding. *RNA* 23, 1172–1179

73 Tan, W. *et al.* (2016) Probing the G-quadruplex from hsa-miR-3620-5p and inhibition of its interaction with the target sequence. *Talanta* 154, 560–566

74 Kozomara, A. and Griffiths-Jones, S. (2014) MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73

75 Chan, K.L. *et al.* (2018) Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem. Commun.* 54, 10878–10881

76 Qi, Y. *et al.* (2015) The genomic sequences near the mir-23b-27b-24-1 cluster form G-quadruplexes and are selectively bound by the natural alkaloid tetrandrine. *Rapid Commun. Mass Spectrom.* 29, 1611–1616

77 Stark, V.A. *et al.* (2021) The Role of miRNAs, miRNA Clusters, and isomiRs in

Development of Cancer Stem Cell Populations in Colorectal Cancer. *Int. J. Mol. Sci.* 22, 1424

78 Cui, X. *et al.* (2012) Spectroscopic probing of recognition of the G-quadruplex in c-kit promoter by small-molecule natural products. *Int. J. Biol. Macromol.* 50, 996–1001

79 Kim, S. *et al.* (2021) MicroRNA signatures associated with lymph node metastasis in intramucosal gastric cancer. *Mod. Pathol.* 34, 672–683

80 Tan, W. *et al.* (2018) Hsa-miR-1587 G-quadruplex formation and dimerization induced by NH4+, molecular crowding environment and jatrorrhizine derivatives. *Talanta* 179, 337–343

81 Li, F. *et al.* (2019) Up- and downregulation of mature miR-1587 function by modulating its G-quadruplex structure and using small molecules. *Int. J. Biol. Macromol.* 121, 127–134

82 Figueroa, J. *et al.* (2017) Exosomes from Glioma-Associated Mesenchymal Stem Cells Increase the Tumorigenicity of Glioma Stem-like Cells via Transfer of miR-1587. *Cancer Res.* 77, 5808–5819

83 Tu, D. *et al.* (2021) M2 macrophages contribute to cell proliferation and migration of breast cancer. *Cell Biol. Int.* 45, 831–838

84 Zhi, Y. *et al.* (2018) Regulation and functions of MicroRNA-149 in human cancers. *Cell Prolif.* 51, e12465

85 Xi, M. *et al.* (2020) Exploration of the formation and structure characteristics of a miR-92a promoter G-quadruplex by ESI-MS and CD. *Talanta* 211, 120708

86 H, Y. *et al.* (2021) Expression of miR-92a, miR-224 and miR-25 in non-small cell lung cancer and their correlation with clinical characteristics. *Am. J. Transl. Res.* 13, 5561–5567

87 Wang, Y. *et al.* (2021) miR-92a promotes cervical cancer cell proliferation, invasion, and migration by directly targeting PIK3R1. *J. Clin. Lab. Anal.* 35, e23893

88 Chen, S. *et al.* (2021) Retinoblastoma cell-derived exosomes promote angiogenesis of human vesicle endothelial cells through microRNA-92a-3p. *Cell Death Dis.* 12, 695

89 Kumar, S. *et al.* (2020) Analysis of G-quadruplexes upstream of herpesvirus miRNAs: evidence of G-quadruplex mediated regulation of KSHV miR-K12–1-9,11 cluster and HCMV miR-US33. *BMC Mol. Cell Biol.* 21, 67

Chapter 3

G-quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions

This chapter was published in:
***Santos, T.**; Salgado, G.F.; Cabrita, E.J., and Cruz, C. (2021) G-quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions. Pharmaceuticals 14, 769. DOI: 10.3390/ph14080769 – IF: 5.863*

Chapter overview

In this chapter, we deeply explored the molecular interactions between G4s and ligands, as well as the most common biophysical methods used to characterize G4/ligand interactions. The techniques described here are classified regarding the nature of the information provided, and their advantages and drawbacks are extensively discussed. This important review concludes that a complete framework of the interaction between G4s and ligands can only be achieved by combining the data obtained from different experimental methods. A "partnership" between different methods is always beneficial for fully characterizing G4/ligand interactions.

G-quadruplexes and their ligands: biophysical methods to unravel G-quadruplex/ligand interactions

Abstract

Progress in the design of G-quadruplex (G4) binding ligands relies on the availability of approaches that assess the binding mode and nature of the interactions between G4 forming sequences and their putative ligands. The experimental approaches used to characterize G4/ligand interactions can be categorized into structure-based methods (circular dichroism (CD), nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography), affinity and apparent affinity-based methods (surface plasmon resonance (SPR), isothermal titration calorimetry (ITC) and mass spectrometry (MS)), and high-throughput methods (fluorescence resonance energy transfer (FRET)-melting, G4-fluorescent intercalator displacement assay (G4-FID), affinity chromatography and microarrays. Each method has unique advantages and drawbacks, which makes it essential to select the ideal strategies for the biological question being addressed. The structural- and affinity and apparent affinity-based methods are in several cases complex and/or time-consuming and can be combined with fast and cheap high-throughput approaches to improve the design and development of new potential G4 ligands. In recent years, the joint use of these techniques permitted the discovery of a huge number of G4 ligands investigated for diagnostic and therapeutic purposes. Overall, this review article highlights in detail the most commonly used approaches to characterize the G4/ligand interactions, as well as the applications and types of information that can be obtained from the use of each technique.

Keywords

G-quadruplex, ligands, molecular interactions, biophysical methods

Introduction

The human genome and transcriptome contain several guanine-rich sequences, which have stimulated considerable interest from researchers since the first reports of their being folded into non-classical structural motifs known as G-quadruplexes (G4s) [1,2,3]

(Figure 3.1A). These structures are characterized by the presence of two or more stacks of four guanines organized in a coplanar manner [4]. Each set of four guanines forms a building block, usually called a G-tetrad, that are stabilized by Hoogsteen hydrogen base-pairing in physiological conditions, π–π interactions, as well as in the presence of positively charged monovalent cations (usually K^+ and Na^+) **(Figure 3.1B)** [5]. G4s are highly polymorphic and can adopt a wide variety of structures based on strand molecularity and strand direction, as well as length and loop composition [6]. According to molecularity, the structures may be distinguished as intramolecular or intermolecular [6]. Considering the direction of the strands, G4 structures may be classified as parallel, antiparallel, and hybrid **(Figure 3.1C–H)**. The loops are generally divided into three main groups: propeller, lateral, and diagonal [6]. Recently, some structural studies demonstrated the formation of G4 structures with longer loop lengths and bulges, opening the framework for the development of novel diagnostic and therapeutic approaches based on those features [7,8].

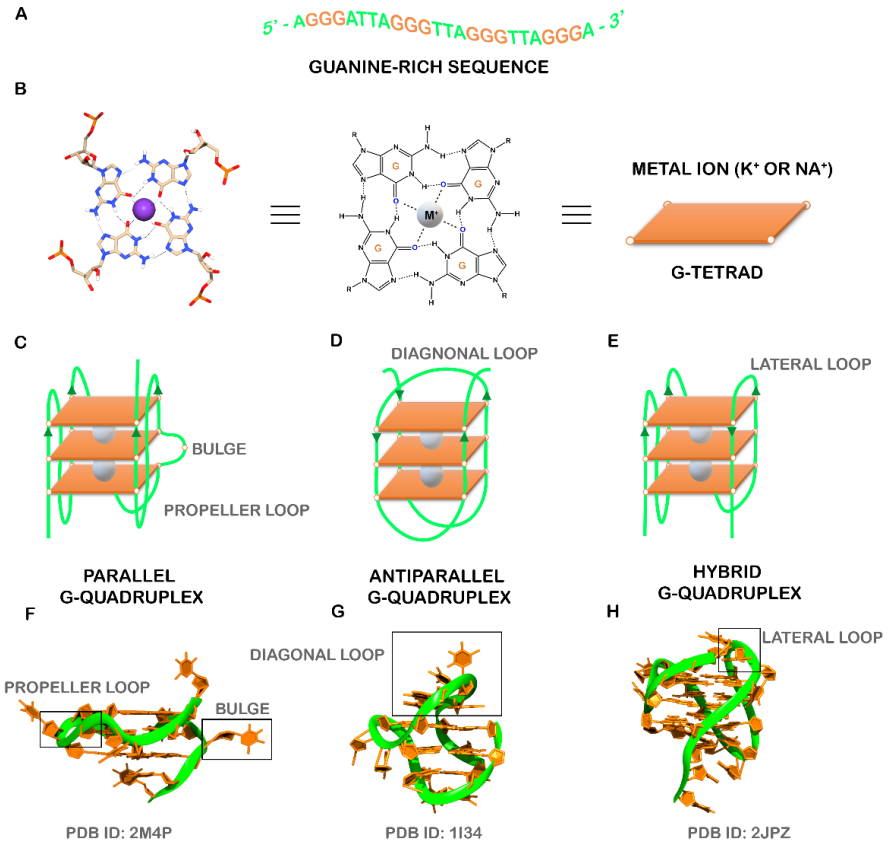

Figure 3.1. (**A**) Guanine-rich sequence with potential to form a three-tetrad G4. (**B**) Chemical structure of G-tetrad formed by the Hoogsteen hydrogen-bonded guanines and central cation (colored in gray) coordinated to oxygen atoms. Schematic representation of common unimolecular G4s based on the strand direction: (**C**) parallel, (**D**) antiparallel, and (**E**) hybrid. Representative PDB structures of (**F**) parallel (PDB ID: 2M4P), (**G**) antiparallel (PDB ID: 1I34), and (**H**) hybrid (2JPZ) G4 structures. The different loops (propeller, diagonal and lateral) and a bulge were also shown.

Computational algorithms were developed to predict the location of specific G4 sequence motifs in the human genome [9,10]. Such predictors consisted of the general motif $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$ and identified over 370,000 sequences with the potential to fold into G4 structures [11]. However, the early algorithms were not accurate and lacked the flexibility to accommodate divergences from the canonical pattern. In order to overcome these disadvantages, novel approaches were developed to compute the G4 propensity score by quantifying G-richness (reflecting the fraction of guanines in the sequence) and G-skewness (reflecting G/C asymmetry between the complementary nucleic acid strands) of a given sequence [12,13], or by summing the binding affinities of smaller regions within the G4 and penalizing with the destabilizing effect of loops [14]. Recently, new machine learning approaches were employed to map active G4s based on sequence features and trained using newly available genome wide mapping of G4s *in vitro* and *in vivo* [15,16].

In recent years, the development of high-throughput sequencing methods, such as G4-seq, has enabled the identification of over 716,000 DNA guanine-rich sequences across the human genome with the ability to fold into G4 structures in the presence of the well-known G4 ligand, pyridostatin (PDS) **(Figure 3.2)** [17]. PDS has an important role in next-generation sequencing (NGS) since it stabilizes G4s and induces polymerase stalling. Those DNA guanine-rich sequences are non-randomly distributed and are mainly located in clusters of immunoglobulin switch regions [18], telomeres [19], and promoter regions of oncogenes [20]. Several reports have described the formation of G4 structures within endogenous chromatin and their ability to recruit transcription factors to promote active transcription [21,22,23,24,25,26,27,28]. The location of those G4 structures was revealed using an antibody-based G4 chromatin immunoprecipitation sequencing (G4 ChIP–seq) approach [21] and suggests that they play a crucial role in critical cellular processes such as DNA replication [29,30], DNA damage repair [26], transcription [22,23], translation [31] and epigenetic modifications [32]. By using G4 ChIP–seq, Hänsel-Hertsch et al. showed a reduction in the number of detected DNA G4s (10,000) in the genome [21]. These results are not surprising since transient G4 structures strongly depend on chromatin relaxation and cell status [21]. Recently, an improved version of the G4-seq method was developed and made available the G4 map of 12 different species [33].

RNA guanine-rich sequences came into the research trends in the last few years due to their intrinsic features and strengths. RNA G4s are more compact, less hydrated, and more thermodynamically stable than their DNA counterparts [34]. Furthermore, the presence of the 2′-OH group in the ribose ring favors the parallel topology, making them

more attractive as target molecules [34]. To date, using computational approaches, more than 1.1 million guanine-rich sequences h with the ability to fold into RNA G4 have been identified [35]. RNA G4s were shown to exist in human cells by using the specific G4 antibody BG4 [36] and, in the same way as DNA G4s, those sequences are non-randomly distributed in the transcriptome [37]. Those sequences are mainly located in both 5′ and 3′UTR, as well as at the splicing junction of mRNA and noncoding RNAs, being of utmost importance in regulatory post-transcriptional mechanisms [37]. In the last few years, several reports have highlighted the importance of G4s in the transcriptome by employing G4 sequencing high-throughput approaches [38,39,40,41]. rG4-seq was initially applied to map G4s in RNA extracted from HeLa cells [38] and later to plants [40] and bacteria [41]. G4RP-seq was also used to characterize the G4 transcriptomic landscape *in vivo* [39]. Yang et al. developed a biotinylated template-assembled synthetic G-quartet (TASQ) derivative (BioTASQ v.1) **(Figure 3.2)** and captured G4 RNAs from breast cancer cells in log-phase growth, followed by target identification by sequencing [39]. The effect of BRACO-19 and RHPS4 **(Figure 3.2)** treatment was also evaluated [39]. They found that those ligands can change the G4 transcriptome in a more remarkable way in long non-coding RNAs [39]. More recently, the same research group developed a new BioTASQ prototype called BioTASQ v.2 **(Figure 3.2)** and performed an in-depth study of both ligands [42]. Those studies are of utmost importance and revealed the strong relevance that G4 ligands could have in cell biology.

Therefore, the location of G4s at both DNA and RNA levels suggests an active role in the development of diseases such as cancer and neurological disorders [43]. Several evidence suggests that G4s play an important role in promoting genomic instability by triggering DNA damage [44,45,46]. The G4 ligand PDS induces DNA damage as shown by the formation of γH2AX foci, a marker of double-stranded DNA breakage (DSB) [47]. Furthermore, ChIP-seq has shown that PDS accumulates at genes containing clusters of G4 structures and that accumulation is transcription-dependent [47]. Recently, De Magis et al. showed that the G4 ligands PDS, BRACO-19, and bis-guanylhydrazone derivative of diimidazo(1,2-a:1,2-c)pyrimidine 1 (FG) **(Figure 3.2)** induced the formation of R-loops, another noncanonical secondary of a DNA:RNA hybrid compatible with the formation of a G4, and promote DNA damage as a consequence of that formation [44]. They also found that the mechanism of genome instability and cell killing by G4 ligands was particularly efficient in *BRCA2*-depleted cancer cells [44]. This study could open up new investigation possibilities and lead to the development of new anticancer approaches.

Although the G4s present in eukaryotic species have been extensively studied, their presence in bacteria and viruses has only attracted attention in the last few years [48,49,50,51,52]. In bacteria, G4s are found in regulatory regions that play important functions in replication, radioresistance, antigenic variation, and latency [51]. G4s in viruses have important regulatory roles in key viral steps [53]. Recent studies have demonstrated the formation and function of G4s in pathogens responsible for serious diseases. Among them are *Mycobacterium tuberculosis* [54], *Pseudomonas aeruginosa* [41], Human Papilloma Virus (HPV) [55], Human Immunodeficiency Virus (HIV) [53], and SARS-CoV-2 [56].

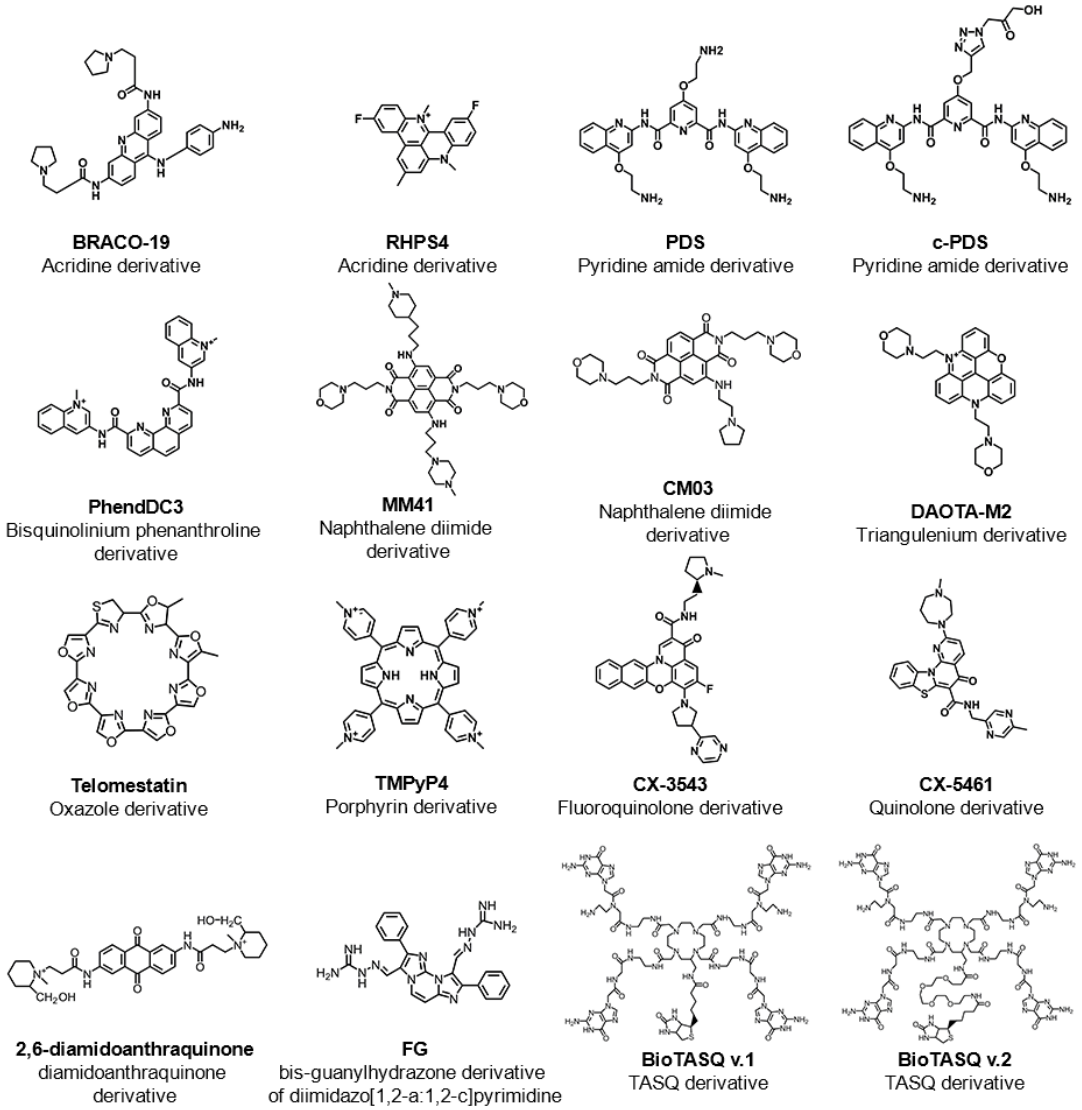

Figure 3.2. List of some examples of G4-interacting ligands mentioned in this review showing the common name of the ligand, chemical structure and family of the compound (chemical backbone).

Therefore, the recognition of the biological significance of G4s has promoted the research and development of ligands that interact with G4s and regulate their structure and function. The most well-known G4 ligands were initially developed to target DNA G4s, but many of them have also been employed to target of RNA G4s [57]. Despite some significant progress in the field, the main challenge remains on the trade-off between affinity and selectivity, which could be achieved with the full characterization of G4/ligand interactions. Since the discovery of the first G4 ligands (disubstituted amidoanthraquinones) **(Figure 3.2)** [58], methods such as circular dichroism, surface plasmon resonance, isothermal titration calorimetry, mass spectrometry, nuclear magnetic resonance, and X-ray crystallography have been used to characterize the molecular interactions of the G4/ligand pair. However, despite the utility of those methods, they are in general, time-consuming and/or costly for the first screening of G4/ligand interactions. Following the general tendency, high-throughput approaches such as FRET-melting, G4-Fluorescence intercalator displacement (G4-FID), affinity chromatography, and microarrays have emerged as rapid and efficient methods to detect the binding and interaction of ligands with their G4 targets.

Overall, this review describes the most well-known G4 ligands and highlights the importance of the most recently developed experimental methods for characterizing G4/ligand complex interactions.

Overview of G4-Interacting Ligands

The increasing evidence that DNA and RNA G4s can regulate a variety of physiological functions in cancer and neurological disorders has encouraged the design and development of new G4-interacting ligands that may act as therapeutic agents [59]. However, besides acting as potential therapeutic agents, ligands can be utilized as molecular agents in biosensing and bioimaging for diagnostic purposes [60].

To date, a significant number of ligands targeting G4s have been investigated, and most of them have been deposited in the G4 Ligands Database 2.1 (http://www.g4ldb.com/ (accessed on 3 August 2021)) [61]. Some G4 ligands are highlighted in **Figure 3.2**. Most of these well-known G4 ligands share common structural features such as an aromatic core, which permits π–π stacking interactions with planar G-tetrads, and one or more positive moieties that may interact with DNA or RNA backbone phosphate groups in grooves and loops [62]. The dissociation constant (K_D) that measures the binding affinity is a key parameter of affinity in G4/ligand interaction and, for most of the G4/ligand complexes, is lower than 10^{-6} M [62].

However, in addition to affinity and stabilization, the G4 ligands must display selectivity between G4 and duplexes' topologies. It is worth noting that some of the most well-known G4 ligands, such as BRACO-19 and TMPyP4 **(Figure 3.2)**, display low selectivity to G4 structures over duplex DNA [63,64]. It was recently proposed that structure-activity relationship studies could significantly improve the physicochemical properties of ligands and enable the optimum trade-off between affinity and selectivity [65]. A recent review by Asamitsu et al. describes in great detail the ligand design and development to acquire specificity and selectivity without compromising affinity [66]. However, there is a significant lack of knowledge of the binding of ligands to G4s when they are embedded with duplex DNA or mRNA, and in chromatin contexts [67].

The early evidence that G4s can be found at the end of telomeres and on oncogene promoter regions led to much attention in the G4 field focusing on developing DNA G4 ligands [66] against cancer. Indeed, because of the low pharmacological properties, which include poor selectivity, no G4 ligand has advanced beyond Phase II trials. Until now, the only ligand that reached this phase was the fluoroquinolone derivative, quarfloxin (CX-3543), which binds to G4s located in ribosomal DNA (rDNA) and disrupts nucleolin interaction [68]. CX-5461 is another G4 ligand that is currently at advanced phase I clinical trials for patients with *BRCA1/2* deficient tumors [69]. Recently, Masud et al. demonstrated that inhibition of the critical member of the DNA damage response, *UBE2N*, acted synergistically with CX-5461 increasing cell toxicity [70]. Further, this compound has shown the potential to suppress pulmonary arterial hypertension and associated vascular remodeling and pulmonary inflammation by inhibiting the RNA polymerase I [71]. For those reasons, the unique chemical qualities of RNA G4s, together with their location in key regions of the human transcriptome, have spurred the design and development of specific and selective RNA G4 ligands. Most of those ligands have been transposed from DNA G4s, and others have been optimized with functional groups to obtain better affinity and selectivity. In this section, we describe some of the compounds that are lastly reported as DNA or RNA G4 ligands, as well as their binding mode and interactions.

DNA G4-Interacting Ligands

As previously mentioned, since guanine-rich sequences of telomeric ends and oncogene promoter regions have been considered attractive targets for cancer therapeutics, a significant number of ligands targeting DNA G4s have been reported [20,72].

Telomeres are nucleoprotein structures that are crucial in genome stability and cell growth due to protecting chromosome ends. In humans, telomeres consist of tandem repeats of the highly conserved repetitive sequence d(TTAGGG), ending in a shorter G-rich single-stranded 3′-overhang. Those tandem repeats can form G4 structures that could impact telomerase activity. BRACO-19 is a 3,6,9-trisubstituted acridine derivative and one of the first G4 ligands developed with specific telomerase inhibitory activity by targeting G4s in telomeres [73]. Despite promising *in vitro* and *in vivo* anticancer activity, its potential use in clinical settings was hindered by its low selectivity toward DNA G4s over duplex [73]. Recently, the binding modes of BRACO-19 toward different telomeric DNA G4 topologies were tested by molecular dynamics simulations with an explicit solvent and revealed an end-stacking mode for the parallel G4s, bottom stacking mode for the antiparallel G4s, and top stacking mode for hybrid G4s [74]. The lack of preferential binding selectivity of BRACO-19 toward G4s over duplex could be explained by the similar binding affinity of groove binding mode for both G4 and duplex structures [74]. In order to tackle the issue of selectivity without compromising binding affinity, one approach that has been used is the use of loops and grooves with the ability to establish different interactions in distinct environments. The pyridine-oxazole derivative TOxaPy **(Figure 3.3A)** shows preferential binding to telomeric G4s with antiparallel topology over telomeric G4s with parallel topology [75]. The results also confirmed the much stronger binding affinity of the ligand in Na^+ than in K^+. The results in Na^+ showed a stoichiometry of 1:1 and $K_D = 2\times10^{-7}\,M^{-1}$. Further evaluation of the interactions by molecular docking suggests a specific groove bind. The naphthalene diimide derivative NDI 3 **(Figure 3.3A)** was developed as a ligand with specificity for *c-KIT2* G4, and the preference for this interaction possibly relies on the specific contact with the loops or grooves, which confer binding preferences [76]. Another approach to gain selectivity is the use of G4 ligands that permit favorable discrimination of dimeric G4 forms from monomeric ones. A new triaryl-substituted imidazole derivative called IZNP-1 **(Figure 3.3B)** was reported and through several biophysical and *in vitro* techniques it was proven that the ligand stabilizes telomeric G4s specifically [77]. In addition, the binding mode was unveiled and revealed intercalation of the ligand into the pocket between the two G4 units **(Figure 3.3B)** [77].

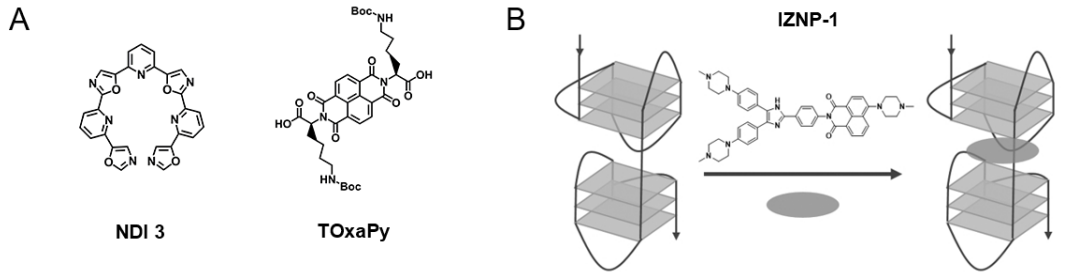

Figure 3.3. Examples of selective DNA G4 ligands. (**A**) Chemical structures and family of the ligands TOxaPy and NDI 3. (**B**) Chemical structure of IZNP-1 and proposed binding model of the intercalation of IZNP-1 into two telomeric G4 units. Adapted from ref. [77].

Previous studies support the evidence that G4s located in promoter regions of *hTERT*, *c-MYC*, *BCL2*, *KRAS*, *HSP90* , and *VEGF*, can serve as potential targets in cancer therapeutics [78]. Therefore, most of the G4 ligands used to hinder telomerase activity were also investigated for their ability to block the transcriptional activity of polymerase in promoter regions of oncogenes [78]. Besides those ligands, the naphthalene diimide derivatives, MM41 and CM03 **(Figure 3.2)**, are binders and stabilizers of G4s in both telomeric ends and oncogene promoter regions (*HSP90, BCL2, and KRAS*), and exert a noteworthy antiproliferative effect in pancreatic cancer cells [79,80]. Furthermore, the computational and experimental approaches revealed that pyrrolidino side chains interact with G4s mainly through groove binding using electrostatic interactions.

The acridine orange derivative C_8 **(Figure 3.4)** was recently investigated for its ability to bind and strongly stabilize G4 structures in the *KRAS* oncogene promoter [81]. The NMR and docking studies demonstrated that the ligand interacts with 22-mer *KRAS* 22RT G4 via π–π stacking and groove/loop interactions [81]. In addition, the acridine core permits end-stacking interactions with external G-tetrads, while alkylamide side chains establish contact points with grooves/loops. Interestingly, the positively charged nitrogen group may occupy the cavity, usually occupied by one of the three central ions through electrostatic interactions. The computational studies using NOE distance restraints revealed details about the ligand structural features in the interaction with *KRAS*-22RT G4 **(Figure 3.4)**.

Nevertheless, the binding modes of other well-known G4 ligands such as telomestatin [82,83], RHPS4 [84,85], and naphthalene diimide derivatives [86,87] **(Figure 3.2)** are well-characterized by several computational and experimental approaches.

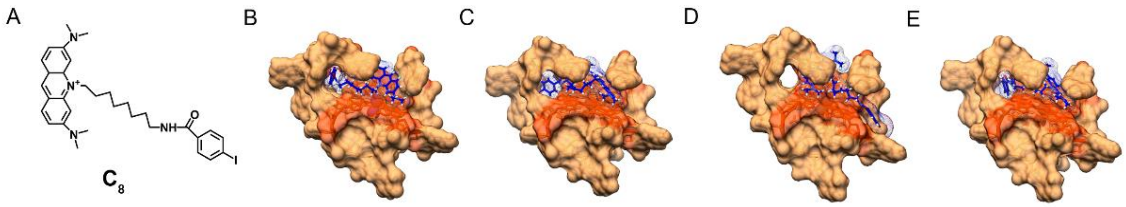

Figure 3.4. (**A**) Chemical structure of the acridine derivative C8. (**B**–**E**) Computational models of complex C8/*KRAS*22-RT generated by molecular docking after energy minimization. The figures (**B**–**E**) depicted conformers obtained with different binding free energies, −9.32, −7.24, −9.17 and −8.40 kcal/mol, respectively. G4 structure is depicted as a light orange surface with the 5′ tetrad highlighted in darker orange. C8 is represented in blue. Adapted from ref. [81].

Meanwhile, at the same time that G4 ligands have been used for therapeutic purposes, a range of G4 optical probes, suitable for diagnostic purposes, have been developed to recognize G4s selectively [60]. Hu et al. developed a specific fluorescent G4 probe that distinguishes *c-MYC* G4 from other G4 structures [88]. In addition, the authors described the binding model of IZFL-2 **(Figure 3.5A)**, which is mainly based on π–π stacking and additional hydrogen bonding interactions [88]. Further, several reviews report G4-specific molecules endowed with fluorescence properties. These molecules represent a variety of structural scaffolds, a mechanism of G4-recognition and fluorescence signal transduction. The G4 selectivity and *in vivo* imaging potential of these molecules place them uniquely as G4-theranostic agents in the predominantly cancer therapeutic context of G4-selective ligands. Moreover, several probes were described to study the function and mechanism of G4 formation in mammalian cells [60,89,90,91,92]. Shivalingam et al. also reported a new G4 optical probe, DAOTA-M2 **(Figure 3.2)**, which localizes and interacts with G4s [93]. Later, the binding mechanism of the ligand toward *c-MYC* G4 was thoroughly investigated [94]. Their study reveals that the DAOTA-M2 polyaromatic core establishes stacking interactions with external G-tetrads of the *c-MYC* G4. In fact, a wide range of DNA G4 optical probes (e.g., NBTE [95], IMT [96] and SiR-PyPDS [97]) **(Figure 3.5)** are increasingly being developed and characterized in recent years. The tripodal cationic fluorescent probe NBTE detected a percentage of G4 DNA in live cells and found 4-fold G4 DNA in cancer cells than in normal cells, suggesting the potential applications of this probe in cancer cell detection [95]. The benzothiazole derivative IMT is a fluorescent probe, reported by Zhang et al., that tracked DNA G4 changes at different points in cell progression and aphidicolin and hydroxyurea treatment, revealing more information on the roles of DNA G4s in biological systems [96].

Another example of a G4-specific fluorescent probe is SiR-PyPDS, which enables single-molecule and real-time detection of individual G4 structures in living cells without

perturbing G4 formation and dynamics **(Figure 3.5B)** [97]. The authors also demonstrated that G4 formation in live cells is cell-cycle-dependent and disrupted by chemical inhibition of transcription and replication [97].

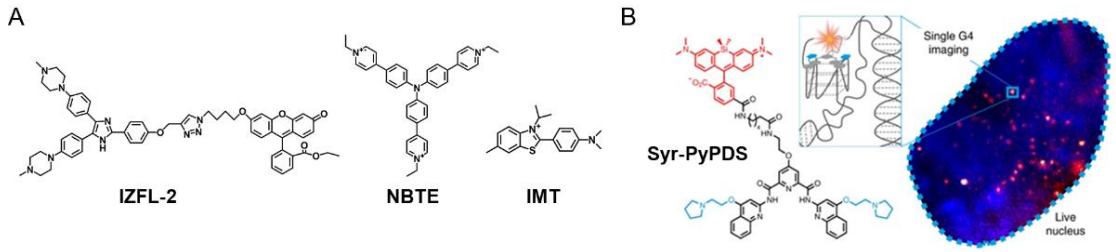

Figure 3.5. Fluorescent G4 probes. (**A**) Common name and chemical structures of IZFL-2, NBTE and IMT. (**B**) Chemical structure of SiR-PyPDS and single-molecule fluorescence imaging of G4s in living cells using the fluorescent probe SiR-PyPDS. Adapted with permission from ref. [97] Copyright 2020 Springer Nature.

RNA G4-Interacting Ligands

Despite a growing number of RNA G4s being identified and characterized, reports on their interaction with G4 ligands are less common. Nevertheless, other RNA secondary structures have been extensively studied for therapeutic and diagnostic purposes [37]. Since RNA G4s are non-randomly distributed in transcriptome and present in several important genes and regulatory regions, designing G4 ligands that stabilize or destabilize such structures is seen as an attractive therapeutic and diagnostic strategy for various diseases such as cancer and neurological disorders [37]. The destabilization of G4 structures by G4 ligands is less commonly described; however, in recent years some studies have been reported for RNA G4s. For instance, some studies involving TMPyP4 have reported that the ligand destabilizes RNA G4s [98,99,100,101,102]. Despite this intriguing observation by different research groups, the mechanism of the unfolding of RNA G4s by TMPyP4 is still unknown.

Until recently, most RNA G4 structures were thought to conform in relatively simple RNA G4 structures, in which guanines stacking within the G4 would also be contiguous in sequence (e.g., four successive guanine trinucleotide tracts separated by loop nucleotides) [34]. Recently, crystallographic and NMR structure determinations of some *in vitro* selected RNA aptamers have revealed RNA G4 structures of unprecedented complexity [103]. Indeed, Sc1 and spiegelmer aptamers have nucleotides in syn conformation, locally inverted strand polarity, and nucleotide quartets that are not all-G. Common to these new structures, the sequences folding into G4s do not conform to the requirement that guanine stacks arise from consecutive nucleotides [103]. On the other hand, the G4 ligands design and development that provides recognition and

selectivity exclusively for RNA has been laborious. In fact, most of the G4 ligands that have been originally designed and developed for DNA G4s have been later investigated to target RNA G4s. An interesting review by Tao et al. describes in detail the recent advances in the development of ligands toward RNA G4 structures [104].

Several pieces of evidence have pointed out that, similar to the DNA G4s located in gene promoters, mRNA 5′UTR RNA G4s could be responsible for the co-regulation of the expression levels of mRNAs with different functions [105]. Currently, multiple efforts are being driven toward the design and development of ligands that would target G4s in specific mRNA transcripts [106]. Halder et al. reported the interaction of bisquinolinium ligands (PhenDC3, PhenDC6 and 360A) **(Figure 3.6)** with RNA G4s present in the 5′UTR *TRF2* mRNA [107]. The driving mechanism for the binding of bisquinolium ligands toward G4s was proposed to be π–π stacking with external G-tetrads [107]. Miglietta et al. identified anthrafurandione derivatives **(Figure 3.6)** as potential ligands of 5′UTR *KRAS* G4 [108]. The binding mechanism probably occurs through the tetracyclic cores of ligands that promote π–π stacking interactions with G4 tetrads, whereas the cationic side chains bind to grooves and loops via electrostatic interactions [108].

Emerging evidence indicates that telomeric repeat-containing RNAs (TERRA) actively participate in the mechanisms regulating telomere maintenance and chromosome end protection. In light of this evidence, several ligands have been developed to target TERRA G4s. The interaction of the polyaromatic molecule, RGB-1 **(Figure 3.6)**, with TERRA and *NRAS* mRNA G4s has been described [109]. The authors proposed that RGB-1 could selectively recognize RNA G4s due to hydrogen bonding acceptors that interact with 2′-OH group of the RNA G4. However, a more deeply structural analysis of the complex RGB-1/TERRA RNA G4 is necessary to understand the binding mechanisms that drive the strong binding affinity and selectivity for RNA over DNA G4s. Carboxy-PDS (c-PDS) **(Figure 3.2)** is recognized by strongly exhibit high specificity for RNA [110]. In a study by Rocca et al., they found that c-PDS established π–π stacking interactions with external TERRA G4 tetrads, and several hydrogen bonds with guanine residues [111]. Furthermore, this ligand showed a noteworthy stabilizing effect on TERRA RNA G4 (ΔT_m = 20.7 °C), and the stabilization was not affected by the addition of up to 100 equivalents of a DNA G4 competitor [36]. c-PDS has been successfully applied for the selective stabilization of endogenous RNA G4s in cells [36].

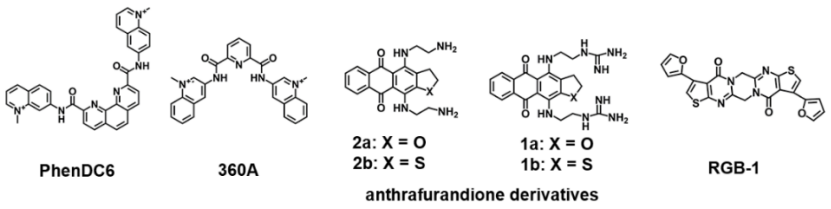

Figure 3.6. RNA G4-interacting ligands mentioned in this review showing the common name of the ligand and chemical structure.

Meanwhile, various G4 ligands were computationally and experimentally investigated for their ability to interact with pre-miRNA G4 structures, which have been previously described as key regulatory agents of miRNA biogenesis. Kwok et al. developed a distinguished experimental approach, which they designated SHALiPE, to detect and structurally map RNA G4s at single-nucleotide resolution. The approach is derived from SHAPE method and consists of the coupling of selective 2′-hydroxyl acylation with a lithium ion-based primer extension [112]. They found that pre-miRNA-149 was able to adopt a G4 form in the presence of PDS, which regulates the miRNA-149 biogenesis. Ghosh et al. showed that in the presence of TMPyP4, the pre-miRNA-149 G4 structure was disrupted, leading to a restoration of miRNA-149 biogenesis [102]. The RNA G4 region of pre-miRNA-149 and pre-miRNA-92b were recently characterized, and their interaction with acridine orange derivatives was evaluated (Please see Chapters 6 and 8) [113,114]. Further, six G4 ligands, which have the potential to target the RNA G4 in pre-miR-149 were investigated by means of molecular docking and molecular modelling **(Figure 3.7)** [115]. The results suggest that large aromatic surfaces with smaller positively charged side chains capable of performing mixed-mode interactions are preferred on the targeting of RNA G4 formed in pre-miR-149 (Please see Chapter 7) [115].

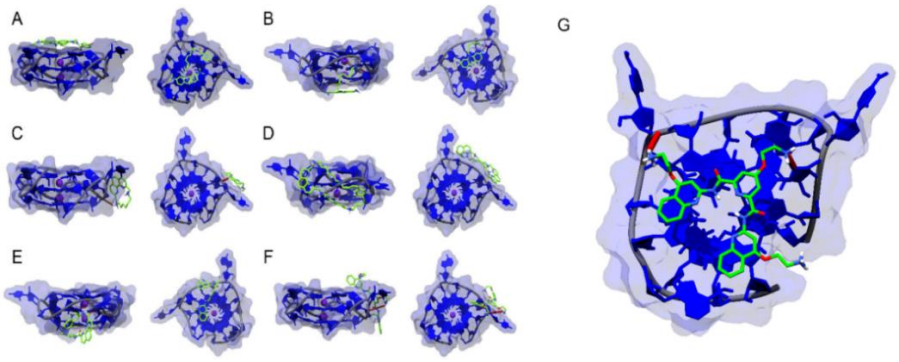

Figure 3.7. Computational models showing the lowest energy binding poses of the complexes between the prediction model of RNA G4 formed into pre-miRNA-149 and the ligands (**A**) C_8, (**B**) C_8-NH_2, (**C**) (16)$phenN_2$, (**D**) (32)$phen_2N_4$, (**E**) PhenDC3 and (**F**) PDS. (**G**) Alternative binding pose of the complex formed by pre-miRNA 149 and PDS after 100 ns of MD simulations. G4 structure is depicted in blue, while ligands are highlighted in green and red. Reprinted with permission from ref. [115] Copyright 2020 Taylor & Francis Group.

As in DNA G4s, a significant number of optical probes targeting RNA G4s have been synthesized in recent years [104]. Chen et al. developed QUMA-1 **(Figure 3.8A)**, a highly selective fluorescence light-on probe, for tracking and monitoring the folding and unfolding of RNA G4s in live cells **(Figure 3.8)** [116]. The interaction between QUMA-1 and RNA G4s was demonstrated to be caused by a rotational restriction that reorganizes the ligand conformation [116].

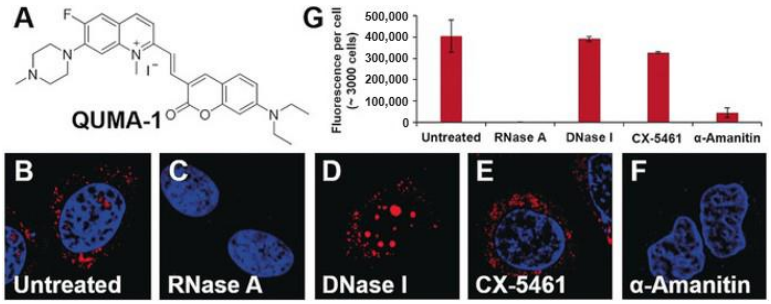

Figure 3.8. (**A**) Chemical structure of QUMA-1; confocal microscopy images of fixed HeLa cells (**B**) stained with QUMA-1; (**C**) stained with QUMA-1 and treated with RNase A; (**D**) stained with QUMA-1 and treated with DNase I; (**E**) stained with QUMA-1 and treated with CX-5461 and (**F**) stained with QUMA-1 and treated with α-Amanitin. (**G**) Fluorescence intensity per cell. Copyright (2018) Wiley. Adapted with permission from [116] Copyright 2018 John Wiley and Sons, Inc.

Remarkably, the same research group employed more sophisticated engineered approaches to detect G4s in *NRAS* mRNA, with the development of the molecular probe ISCH-0a1 **(Figure 3.9)** [117]. They combined the G4 light-on ligand ISCH-1 with an oligonucleotide complementary to the adjacent sequence of the *NRAS* G4 sequence to form an ISCH-*nras*1 ligand that can selectively bind a specific region in *NRAS* mRNA (G4-triggered fluorogenic hybridization (GTFH) probe) **(Figure 3.9)** [117]. This approach of sequence-specific recognition of *NRAS* mRNA adjacent to an intended G4 provides a gain of specificity in the transcriptomic context.

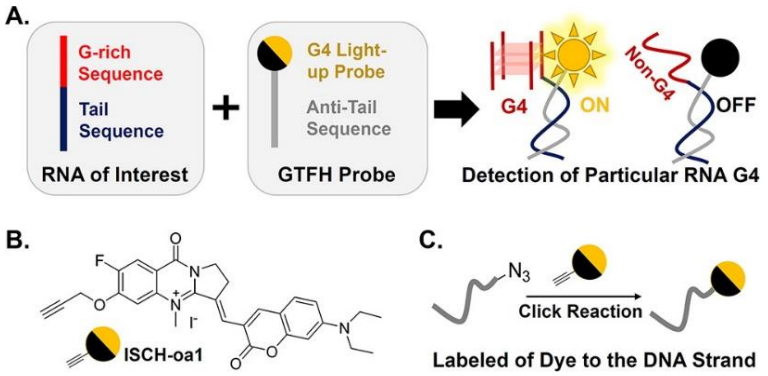

Figure 3.9. Schematic representation of GTFH probe. (**A**) GTFH probe design principle. (**B**) Chemical structure of ISCH-0a1. (**C**) GTFH probe preparation. Reprinted with permission from ref. [117]. Copyright 2016 American Chemical Society.

Laguerre et al. also described a new multifunctional compound called PyroTASQ **(Figure 3.10)**, which belongs to both ligands and fluorescent probes [118]. The ligand was designed using the template-assembled synthetic G-quartets method. This multifunctional compound binds to both DNA and RNA G4s in a mechanism that promotes the reorganization of the ligand in a structure that leads to the release of fluorescence by the pyrene moiety [118]. However, *in vivo* application of PyroTASQ is limited due to its aggregation in cells. Later, in order to overcome this drawback, the same research group substituted the pyrene moiety with naphthalene [119]. The new compound was named NaphtoTASQ **(Figure 3.10)** and it was used to visualize DNA and RNA G4s in living cells [120].

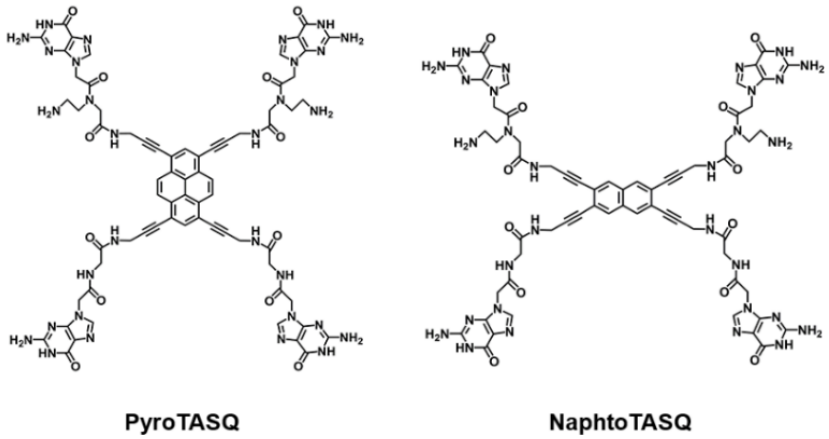

Figure 3.10. Chemical structures of PyroTASQ and NaphtoTASQ.

Methods to Characterize G4/Ligand Interactions

In the past few years, the progress in the design and development of numerous methods has helped researchers to assess the binding modes and interactions between G4s and their ligands. These approaches could be different in their nature and more straightforward or complex, but all of them give important information about binding interactions between G4s and their ligands. Many of the methods earlier used to characterize dsDNA interactions have been employed and modified to highlight the contact points between G4s and their binders [121,122]. Biophysical techniques are usually the primary choice in the study of physical interactions and can be divided into three main categories: structure-based methods; affinity and apparent affinity-based methods; and high-throughput methods. In this review, we highlight in detail structure-based methods (circular dichroism (CD), nuclear magnetic resonance (NMR), and X-ray crystallography); affinity and apparent affinity-based methods (surface plasmon resonance (SPR), isothermal titration calorimetry (ITC) and mass spectrometry (MS));

and high-throughput methods (fluorescence resonance energy transfer (FRET)-melting, fluorescence intercalator displacement (G4-FID), affinity chromatography and microarrays). Each method has advantages and drawbacks associated with its features and usually, several methods are necessary to achieve complementary information about G4/ligand interaction [121,122]. This review section discusses biophysical techniques and how they are important to characterize *in vitro* G4/ligand interactions without cellular relationship.

Structure-Based Methods to Investigate G4/Ligand Interactions

Circular Dichroism (CD)

CD is a standard biophysical method for evaluating the secondary structure, folding, and binding properties of chiral molecules. The method is based on the difference in absorbance by a substance of right- and left-handed circularly polarized light. The intrinsic features of CD make it a powerful tool for the study of the secondary structures and conformations adopted by nucleic acids [123]. The CD spectral signature of nucleic acids arises from (i) asymmetric backbone sugars; (ii) helicity of the secondary nucleic acid structures and (iii) long-range tertiary ordering of nucleic acids in some specific solution conditions [123]. Those important features make CD a method of choice for the study of G4 structures. CD spectroscopy is commonly used to characterize G4 topology, cation effect, G4/ligand interactions and ligand-induced thermal stabilization [124].

By using qualitative rules-of-thumb, CD can be easily applied to distinguish different G4 topologies [124,125]. Therefore, the CD spectral features of a given G4 structure are indicative of its topology: parallel G4s have a positive band ≈ 260 nm and a negative band ≈ 240 nm; antiparallel G4s are characteristic by a positive band at ≈ 295 nm and a negative one at ≈ 260 nm and hybrid G4s depict two positive bands (≈ 295 nm and ≈ 260 nm), and a negative one ≈ 245 nm **(Figure 3.11A)** [124,125].

Furthermore, the technique has emerged as a strong tool to characterize the interaction of G4s with ligands, being a low-resolution complement to high-resolution methods. NMR spectroscopy and X-ray crystallography are used for high-resolution structural analysis of G4/ligand interactions but are costly, time-consuming, and require a large amount of the sample [124,125,126,127]. CD provides a less detailed analysis of the structure, but requires less of the sample. Further, measurements are always performed in solution and are not limited by the molecular weight or size of a molecule [126,127]. Furthermore, the technique provides a fast method to screen G4/ligand interactions,

particularly when the timescale of a particular interaction or orientation is short [127]. These advantages significantly contributed to increase the number of studies that employed the technique to analyze the effect of ligands on G4s.

Some specific G4 ligands showed the ability to induce conformational switches in G4s [6,128]. This feature is of utmost importance since it open-up an important framework to control G4 structure and function [128]. In many cases, the new G4 topology is primarily assigned by CD. One of the first examples of a ligand that induced a switch in G4 topology was the triamino-anthracene derivative, which was first designed to interact with G-tetrads by stacking and with the central ion channel with the amine moieties present in the side-chain [129]. Rodriguez et al. monitored the interaction of the ligand with telo24 G4 via circular dichroism spectrum and found that in a sodium buffer, the ligand changes the G4 topology from antiparallel to parallel in about two hours [129]. Interestingly, a porphyrazine ligand can reverse the equilibrium by favoring the antiparallel topology [129]. However, despite some reported examples of ligands that induced a switch in G4 topology centered mainly on the CD data [130,131,132], care must be taken before attributing a new folding topology based on CD results alone. Recent reports have employed CD as a complementary method to evaluate switches in G4 topology [133,134]. Marchand et al. used NMR spectroscopy and mass spectrometry to prove that 360A, PhenDC3, and PDS switch the G4 folding topology of telo22, telo23, telo24 and telo26 [134]. CD spectroscopy was further employed and revealed an antiparallel G4 structure upon ligand binding [134].

In general, G4 ligands are non-chiral and therefore have no CD signal by themselves in solution [135]. However, upon interaction with G4s, those ligands can eventually acquire an induced CD (ICD) signal through the coupling of electric transition moments of the ligand and G4 bases. The observation of an ICD signal is indicative of G4/ligand interactions, but can also perturb the apparent signature of a given G4 topology [135]. As previously mentioned, in those particular cases care must be taken to avoid mistakes in the interpretation of CD results. Nevertheless, the ICD signal could be used to obtain insights into the mode of interaction and geometry of the binding [135].

Ligands that bind G4s by end-stacking or intercalation display only very weak or even no ICD signals [136]. A new class of 9-O-substituted berberine derivatives has been synthesized and their interactions with several G4 structures are characterized by CD titrations [137]. Moreover, in the same study, they also investigated the orientation of the ligand with G4s by analyzing the ICD signals [137]. The same research group reported the binding modes of a selective fluorescent probe of DNA G4s [138].

In the case of groove binders, larger positive ICD signals are generally observed [139]. Zuffo et al. reported that core-extended naphthalene diimides display intense ICD positive signals in the presence of antiparallel and hybrid G4s [139]. In their study, the ligands do not display a detectable ICD signal in the presence of parallel G4s, suggesting a different binding mode that was hypothesized to be end-stacking [139].

CD is also used to determine the binding stoichiometry of G4/ligand complexes and the folding of the G4 structures induced by the ligands. Recently, Głuszyńska et al. employed CD titrations to probe binding stoichiometries and the effect of a fluorescent carbazole derivative in the folding of G4 structures [140]. However, in these particular cases, CD just provides qualitative information about G4/ligand interactions, being frequently a complement for other methods.

Additionally, almost all reported G4 ligands have been characterized by means of CD-melting experiments. The measurement of the melting temperature of the G4 structure by CD is usually performed at the maximum ellipticity wavelength (260 nm or 295 nm), which decreases upon denaturation **(Figure 3.11B)** [124]. Thus, depending on the effect of the ligand, the melting temperature of G4s could decrease or increase. In recent decades, the considerable advantages of using CD-melting to study G4/ligand interactions have made it a method widely used by almost all researchers in the field. Indeed, the method is usually one of the primary choices to evaluate the potential of new ligands to stabilize G4 structures.

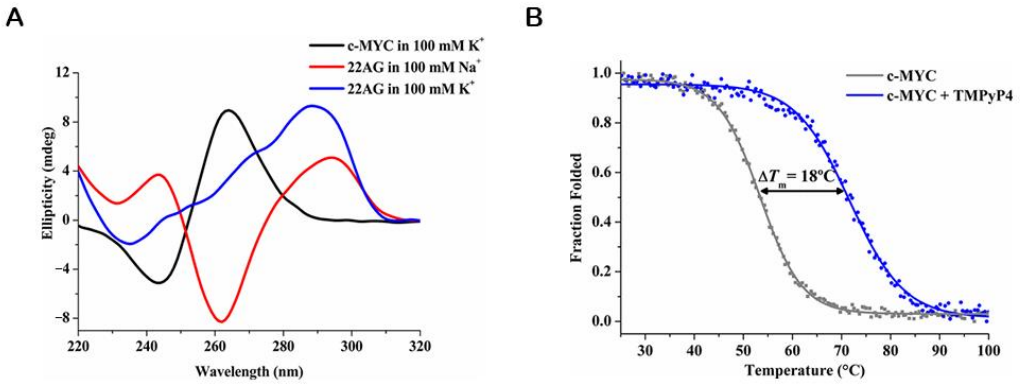

Figure 3.11. (**A**) CD spectra of G4 structures: c-MYC (10 μM) in 100 mM KCl (parallel); 22AG (10 μM) in 100 mM NaCl (antiparallel) and 22AG (10 μM) in 100 mM KCl (hybrid). The final volume of each sample in the 1 cm path-length quartz cell was 300 μL. The buffer used for the experiments was 10 mM phosphate buffer pH 7.1. (**B**) CD melting curves for c-MYC G4 (10 μM) in 100 mM LiCl in the absence and presence of 2 molar equivalents of TMPyP4. The ellipticity was monitored at a wavelength of 263 nm. The buffer used for the experiments was 10 mM phosphate buffer pH 7.1. Reprinted with permission from ref. [124] Copyright 2017 American Chemical Society.

Nuclear Magnetic Resonance (NMR)

NMR has proved to be fundamental to study the structural, kinetics, and dynamics of G4/ligand complexes. The technique provides information about G4/ligand binding under *in vitro* conditions. In some specific cases, the conditions were set to mimic cellular media with the addition of molecular crowding agents such as PEG and Ficoll [141], and more recently closer to the conditions observed in the living cells using cytoplasmatic extractions or directly inside living cells such as oocytes [142,143,144]. Using 2D $^{1}H–^{15}N$ SOFAST-HMQC type experiments to follow changes in chemical shift and intensities, Salgado et al. demonstrated the assembly of the tetramolecular G4 model $d(TG_4T)_4$, inside living *Xenopus laevis* oocytes [143]. They also showed, for the first time, that is possible to study the interaction of ligands with G4s inside living cells [143]. This study opened a new framework to study G4/ligand interactions under *in vivo* conditions with atomic resolution. Recently, Krafcikova et al. evaluated DNA/ligand interactions inside the nuclei of living human cells [142]. They tested two model DNA fragments and four ligands and the results revealed a strong influence of environmental conditions on the stability of DNA/ligand complexes, especially for one ligand that presented off-target interactions with genomic DNA and cellular metabolites [142]. Their study proved the suitability of the technique to validate ligands, after an *in vitro* assessment of promising ligands.

The choice of NMR methods to study G4/ligand complexes depends strongly not only on the exchange regime of the individual molecules and the complex, but as well on the ligand properties such as those related with hydrophobicity and immiscibility in water-based solutions. In most cases, the polyaromatic properties of G4 targeting ligands require that stocks are prepared in DMSO and other solvents best adapted for hydrophobic compounds. For the vast majority of reported cases, the NMR experiments probe the ligand interaction based on the chemical shift, peak height, and linewidth properties of the observed signals. Those experiments are easy to implement and rapid to acquire and analyze. More elaborate experiments based on polarization transfer between spins systems, such as the nuclear Overhauser effect (NOE) [145], Transferred-NOESY [146], Saturation Transfer Difference (STD) [147], and Water-LOGSY [148] methods require more time and elaborated controls. In addition, the abovementioned methods are usually complemented with indirect NMR studies such as H_2O/D_2O exchange-rate measurement or paramagnetic spin labels attached in different parts of the G4. As a first approach, most studies take advantage of the G4 imino (Hoogsteen base pairing) region near 10–12.5 ppm for initial sampling of ligands. Nevertheless, precautions should be taken in the case of loose G4-ends and loops that do not present

imino signals. The advantage is enormous because the aromatic regions (~7–8.5 ppm) are often crowded by many more nuclei from the G4 and the ligand itself, and a direct 1D-visualization and analysis it is not possible. Nevertheless, the imino regions of G4 guanines provide a direct and clear monitoring system for the formation of a G4 structure and its ligand binding interactions.

A study by Yang's research group showed, by using 2D NMR experiments (DQF-COSY, NOESY and TOCSY), the interaction modes and dynamic binding of BMVC to *c-MYC* G4 [149]. The authors revealed that upon the addition of lower equivalents of BMVC (0–0.7 equivalents), a new set of imino protons appeared, suggesting the formation of a BMVC/*c-MYC* G4 complex [149]. At 1:1 ratio, the imino region showed only one set of imino protons, suggesting a 1:1 complex formation. After the addition of higher equivalents (>1 equivalent) of BMVC, a second set of peaks appeared, which represents the formation of a second complex [149].

However, for high-resolution structural analysis of the complex G4/ligand, 2D NMR and molecular modeling were combined to unambiguously obtain information to construct the 3D model [150]. In addition, the intermolecular NOE interactions can give substantial evidence of the position of the ligand in the 3D structure of the G4 [151]. There are some chemical moieties that are preferable for analyzing the complex G4/ligand, such as methyl groups once they present strong and individual NOE interactions [152].

Kerkour et al. demonstrated the binding interaction between the G4 ligand, 2,4,6-triarylpyridine (TAP) **(Figure 3.12A)**, and the 22AG G4 structure **(Figure 3.12B–D)** [150]. The ligand presented some signal dispersion indicative of loose binding in multiple conformations. The aromatic protons were not well resolved and the unambiguous restraints used in structural studies originated from the aliphatic protons **(Figure 3.12C)**. Their NMR data were supplemented with restrained-docking studies to determine the 3D model of the complex TAP-22AG G4. The results obtained with NMR spectroscopy are consistent with those obtained with molecular docking experiments and revealed two binding sites in the 22AG structure, with the most favored site being the lateral loop formed by T17, T18 and A19. In addition, the results depicted a low predisposition of TAP to bind through π–π stacking, preferring the grooves and loops interface.

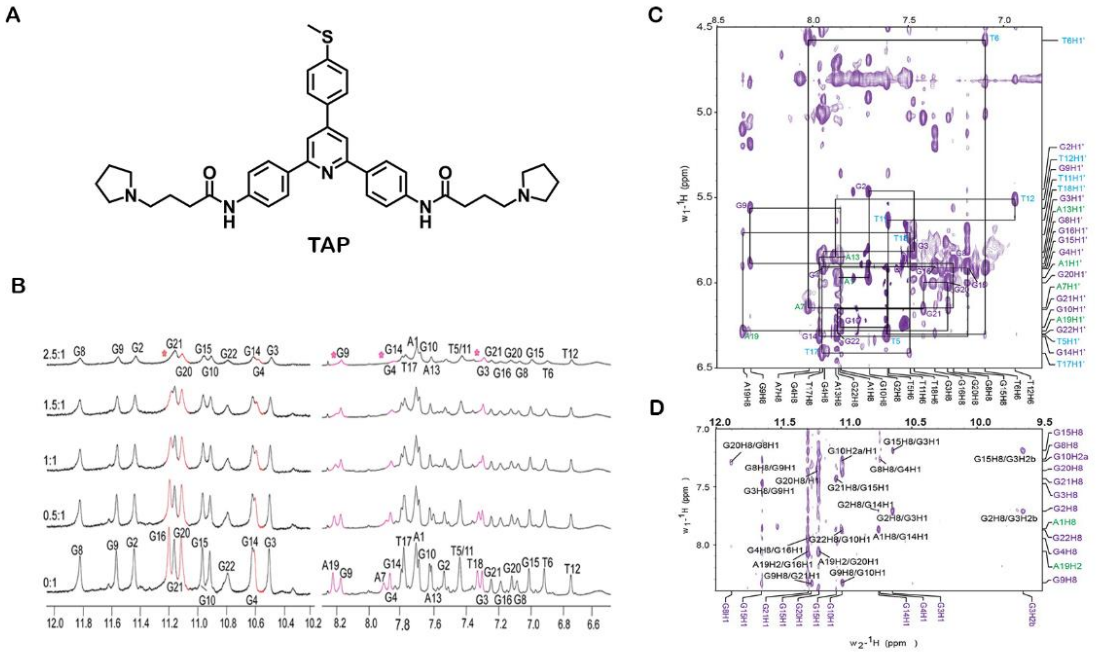

Figure 3.12. (**A**) Chemical structure of 2,4,6-triarylpyridine (TAP). (**B**) Imino and aromatic proton regions titration spectra of 500 μM 22AG with different concentrations of TAP in 200 μL of sodium phosphate buffer containing 20 mM Na_2HPO_4/NaH_2PO_4, 70 mM NaCl at pH 6.9. (**C**) Expanded H8/H6-H1′ 300 ms 2D–NOESY spectrum of the 2.5:1 TAP:22AG complex. (**D**) Expanded H1-H8/H2 region from a 300 ms 2D–NOESY spectrum of the TAP:22AG complex at 2.5:1 stoichiometry. Reprinted with permission from ref. [150] Copyright 2017 Elsevier.

Recently, Yang's research group reported, by using NOESY, HSQC, and DQF-COSY experiments, the NMR solution structures of *c-MYC* complexed with PEQ **(Figure 3.13)**, a specific *c-MYC* G4 binder with druglike properties [153]. Chemical shift perturbation was used to elucidate macromolecular interactions in the complex G4/PEQ [153]. Furthermore, they have performed a comparison of four available complex structures involving the *Myc*2345_T23 structure and ligands PEQ, BMVC, Quindoline-I (Qi,) and DC-34 **(Figure 3.13)** [153]. The *Myc*2345_T23 sequence bears a G23-to-T mutation in the 3′-flanking at position 23, in addition to the G14-to-T mutation in the second propeller loop [153]. This structure can be used as a model of parallel G4 structures in promoters and RNA G4s. The same approaches were also implemented to evaluate the complex formation of several G4 ligands and various G4 structures [81,154,155,156].

In the last few years, in order to obtain valuable structural information about G4/ligand complexes, Fluorine-19 (^{19}F) NMR has become a very useful tool [157]. The advantages of ^{19}F NMR spectroscopy include simplicity and sensitivity as well as its ability to study large complexes that cannot easily be probed by conventional NMR experiments [157].

Moreover, it has been applied to study G4 structures, due to the high sensitivity of the ^{19}F chemical shift to the environment [158,159].

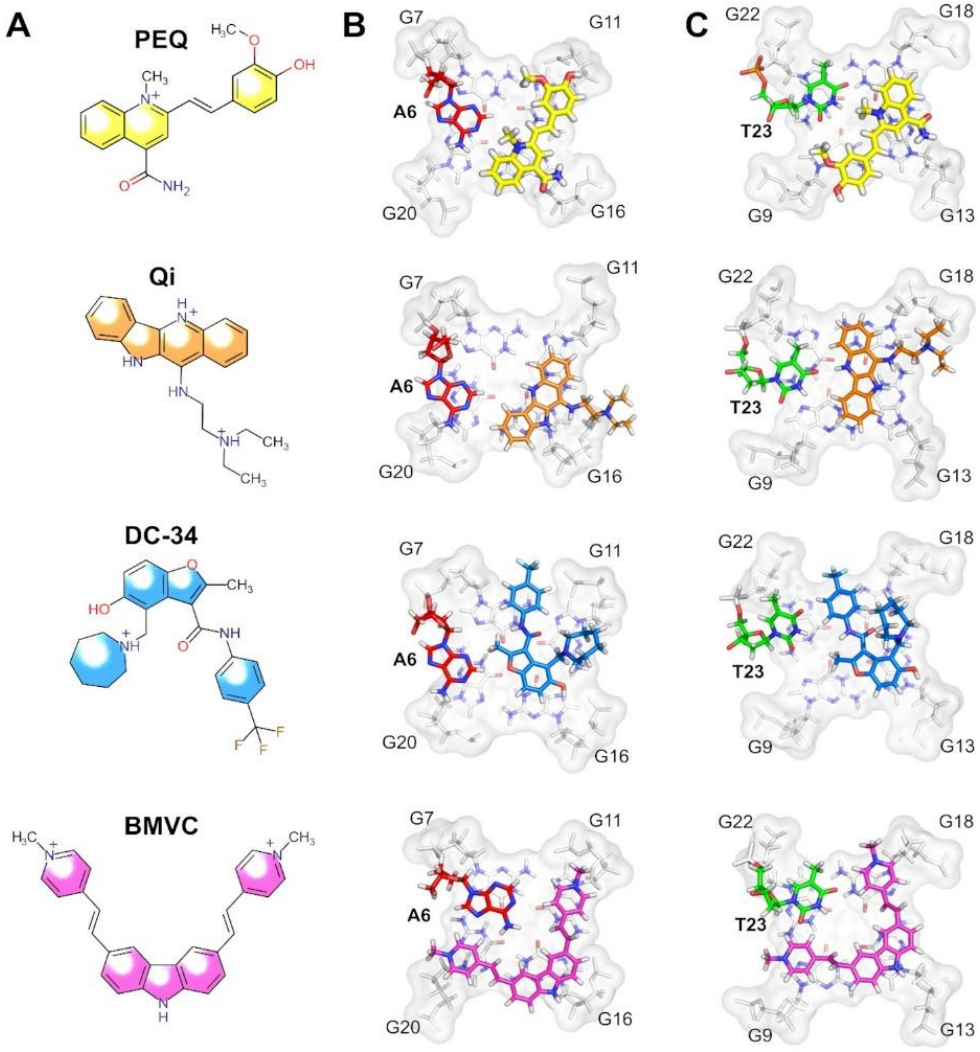

Figure 3.13. High-resolution structures of Myc2345_T23 in complex with PEQ, Quindoline-I, DC-34 and BMVC. (**A**) Common names and chemical structures of the ligands. Binding of each ligand to the (**B**) 5′-G-tetrad and (**C**) 3′-G-tetrad of Myc2345_T23. Reprinted from ref. [153].

Bao et al. demonstrated that ^{19}F could be a powerful technique to analyze the interaction of RNA G4s with ligands [159]. They analyzed the complex structure between the fluorinated telomeric RNA G4 (^{19}F-5′-UAGGGUUAGGGU-3′) and the well-known G4 ligand PDS **(Figure 3.14)**. ^{19}F NMR enables researchers to distinguish the free and complexed RNA G4, because it is highly dependent on the structural environment of the ^{19}F label [159]. Their results revealed that PDS interacts with dimeric G4s but also with high-order G4s with two subunits stacked [159]. Furthermore, ^{19}F NMR is an attractive tool to determine K_D and T_m of the complex RNA G4 and PDS [159]. Binding stoichiometry of PDS to dimeric G4 was 2:1 in a ratio of 6:1 of the ligand to RNA, and

the K_Ds were 0.63 mM and 1.24 mM for the first and second binding events, respectively [159]. The ΔT_m values of dimeric and high-order G4s were >20 °C and 10.7 °C, respectively [159]. The same research group also demonstrated the utility of ^{19}F NMR for the observation and quantitative thermodynamic characterization of telomeric RNA and DNA G4s within in living cells [160,161,162]. Overall, ^{19}F NMR offers a suitable and sensible approach to obtain structural insights of DNA and RNA G4s and their complexes with ligands.

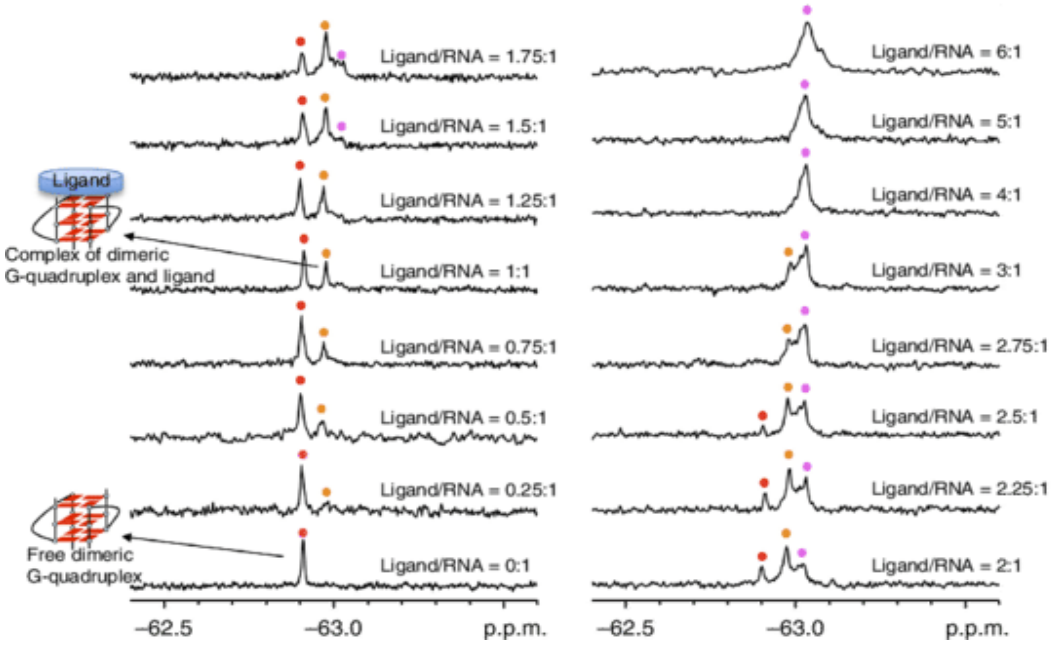

Figure 3.14. ^{19}F NMR spectra of ^{19}F-labeled 0.5 mM RNA (^{19}F-5′-UAGGGUUAGGGU-3′) in 150 µL of 90% H_2O/10% D_2O, 10 mM Tris-HCl buffer (pH 7.0), 50 mM KCl. titrated with PDS. Copyright (2017) Wiley. Reprinted with permission from ref. [159] Copyright 2017 John Wiley and Sons, Inc.

X-ray Crystallography

The application of X-ray crystallography toward a structural understanding of G4/ligand complexes enables a detailed structural and spatial characterization of the intermolecular interactions at the atomic level. Furthermore, the considerable advances in X-ray crystallography, such as the miniaturization and automation of crystallization trials, have converted the technique into a method that can be used for screening purposes. One of the advantages of X-ray diffraction over the other techniques is that ligand binding sites can be precisely determined through electron density maps [163]. Furthermore, high-quality X-ray data (i.e., better than 2.5 Å) make the following steps more reliable and allow the visualization and quantification of G4/ligand hydrogen bonding and electrostatic interactions [163]. Moreover, the role of water molecules in the G4/ligand interaction can be determined. However, some important limitations of the

technique should be highlighted, such as, (i) the need of a high amount of sample; (ii) the use of cryogenic temperature that can damage artifacts. (iii) the sample must be crystallizable; and iv) it cannot be performed in solution, which limits its application in biological samples. Nevertheless, the technique could be used to obtain detailed information on G4/ligand interactions and has been used in the last year by several researchers in the field.

The earliest reported crystal structures of G4/ligand complexes were the tetramolecular G4/daunomycin complex [164] and the telomeric G4 from the Oxytricha nova/acridine derivative complex [165]. Since then, several G4/ligand complexes have been characterized using X-ray crystallography [166,167,168,169].

Recently, Bazzicalupi et al. unveiled the crystal structure of human telomeric DNA G4 complexed with berberine [169]. The results showed that berberine binds toward the G4 telomeric structure in a 2:1 molar ratio through π-stacking interactions. The authors stated that strong interaction can be attributable to the coplanarity of the G5/G11 and G17/G23 pairs and to the modified conformation of the 3′-end tetrad.

Lin et al. investigated the interaction of the G4 forming sequence, 5′-GGGTTGGGTTGGGTTGGG-3′ (T1) and several other sequences based on T1, with the G4 ligand, N-methyl mesoporphyrin IX (NMM) by biophysical and X-ray crystallographic studies [166]. They solved the crystal structure of T1- and 5′-TGGGTTGGGTTGGGTTGGGT-3′ (T7)-NMM complexes. Both structures produced large hexagonal crystals and the T1-NMM crystal structure was determined with a resolution of 2.39 Å, while the T7-NMM crystal structure was solved with a resolution of 2.34 Å **(Figure 3.15)**. The results showed that both complexes form a 5′-5′ dimer of parallel G4s, which interact with NMM in 3′ G-tetrad through stacking interactions. However, taking into account that one of the main limitations of X-ray crystallography is its non-dynamic nature, they investigated the correlation of the results obtained by X-ray crystallography with techniques performed in solution, namely, CD and PAGE. The T1-NMM complex is concentration-dependent and exists as a mixture of monomer and dimer, whereas the T7-NMM complex only exists as a monomer at low and high concentrations. These data suggest that the T7-NMM complex determined by crystallographic studies could be an artifact of the crystal packing.

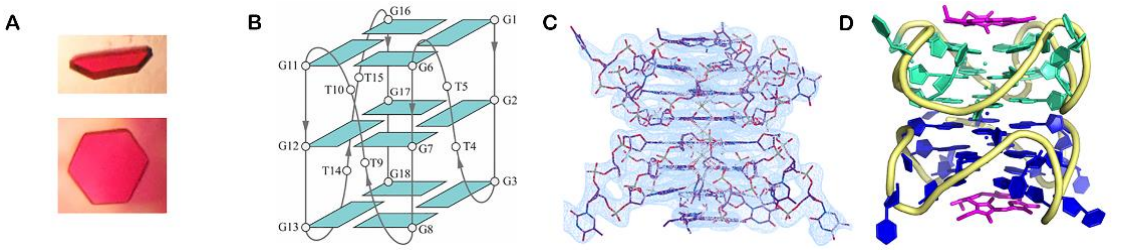

Figure 3.15. Crystal structure of the T7-NMM complex. (**A**) Representative crystal morphologies: half-hexagonal and hexagonal. (**B**) Schematic representation of the folding topology with the numbering scheme for T1. (**C**) Electron density maps of the T7-NMM dimer. (**D**) Crystal structure of the T7-NMM complex. The T7-NMM sample was prepared by annealing T7 with 1 eq. of NMM at 0.65 mM in 10 mM lithium cacodylate pH 7.2 and 20 mM KCl. Drops were set by the TTP Labtech Mosquito Crystal liquid handler equipped with a humidity chamber at 0.1 µL DNA sample and 0.1 µL of the crystallization condition. Small hexagonal crystals grew within three weeks to 80 µm in the largest dimension from condition C5 of the Natrix screen (Hampton Research): 4.0 M LiCl, 0.01 M MgCl2, and 0.05 M HEPES sodium pH 7.0. Crystals were cryoprotected in the base condition supplemented with 15% ethylene glycol before being flash frozen in liquid nitrogen. Adapted from ref. [166].

The most commonly used techniques to structurally characterize, in detail, the G4/ligands interactions are above described and can provide crucial information about structure-function relationships. The major advantages and limitations of each structure-based technique are summarized in **Table 3.1**.

Table 3.1. Advantages and limitations of structural-based methods for investigating G4/ligand interactions.

Method	**Advantages**	**Limitations**
CD	Simplicity	Low-resolution
	Small amount of sample	
	No need of sample labelling	
	Not limited by the molecular weight or size of the molecules	Most of the ligands are non-optically active
	Can easily provide melting temperature curves and global folding changes	
	Most suitable method for finding the polarity of chains	
NMR	Provides atomic-resolution characterization of a G4/ligand complex	High amount of sample
	Detailed pairs of atoms contacts between ligand and receptor	Time-consuming
	Three-dimensional structures in their natural state can be measured in solution	Limited by size or atomic weight
		Need of isotopic labelling

Table 3.1. Advantages and limitations of structural-based methods for investigating G4/ligand interactions. (Continued)

X-ray	Relatively cheap and simple	High amount of sample
	Provides atomic-resolution characterization of a G4/ligand complex	Cryogenic temperature can induce altered contacts
	Provides void electronic areas in the receptor that can be used to improve ligands design	The sample must be crystallizable
	Not limited by size or atomic weight	Only provides static three-dimensional analysis

Affinity- and Apparent Affinity-Based Methods to Investigate G4/Ligand Interactions

Surface Plasmon Resonance (SPR)

SPR is a label-free powerful method that combines microfluidics and optics to study the interactions of G4s with ligands **(Figure 3.16)**. The method provides a real-time equilibrium and kinetic characterization for G4 interactions with small amounts of materials and no external probe. The G4 sequence is immobilized on the surface and the ligand is flowed through a microfluidic system. The interaction of the ligand with the G4 is measured by following the changes in the surface refractive index, which modify the angle of the incidence of light. The surface refractive index changes are converted to resonance units (RU) and the results are presented as sensorgrams or binding curves. The advantages of using SPR are as follows: (i) acquisition of data in real-time; (ii) time efficiency; (iii) no need for sample labeling and (iv) high sensitivity, which permits the use of a small amount of the sample [170]. However, some disadvantages have been pointed out to the technique: (i) it requires sophisticated instrumentation and it is costly; (ii) the high dependence on the experimental conditions; (iii) the requirement of maintaining the G4 structure intact after immobilization; and (iv) great care has to be taken in the analysis of the collected data. Indeed, significant efforts have been employed to overcome these issues [170]. The most reliable and suitable immobilization strategy to evaluate the G4/ligand interactions is to have the biotin covalently attached to the 5′-end of the G4 sequence. This strategy ensures the proper assembly of the G4 structure after immobilization and prevents any perturbation that might happen by steric hindrance. However, care must be taken since the addition of biotin covalently attached to the 5′-end may impact the G4 conformation, as modifications at the 5′ or 3′ very often do. The use of other complementary experimental techniques must be considered to rule out this hypothesis. Using this immobilization strategy, Perenon et al. used SPR experiments to investigate the interaction between N-methyl mesoporphyrin IX (NMM)

and different G4 topologies. They determined the dissociation constant of parallel G4s and showed that it was at least ten times lower than for other topologies [171]. Recently, the binding affinities of the well-known G4 ligands, PhenDC3, 360A, and PDS, with the G4 aptamer AS1411 and its derivative AS1411-N6, which have six nucleotides to the 5′-end that are complementary to other six nucleotides at the 3′-end, were assessed by means of SPR [172]. The same strategy of immobilization was used with the oligonucleotides labeled at 5′-end with biotin. The results of SPR measurements are in line with the results obtained by melting experiments [172].

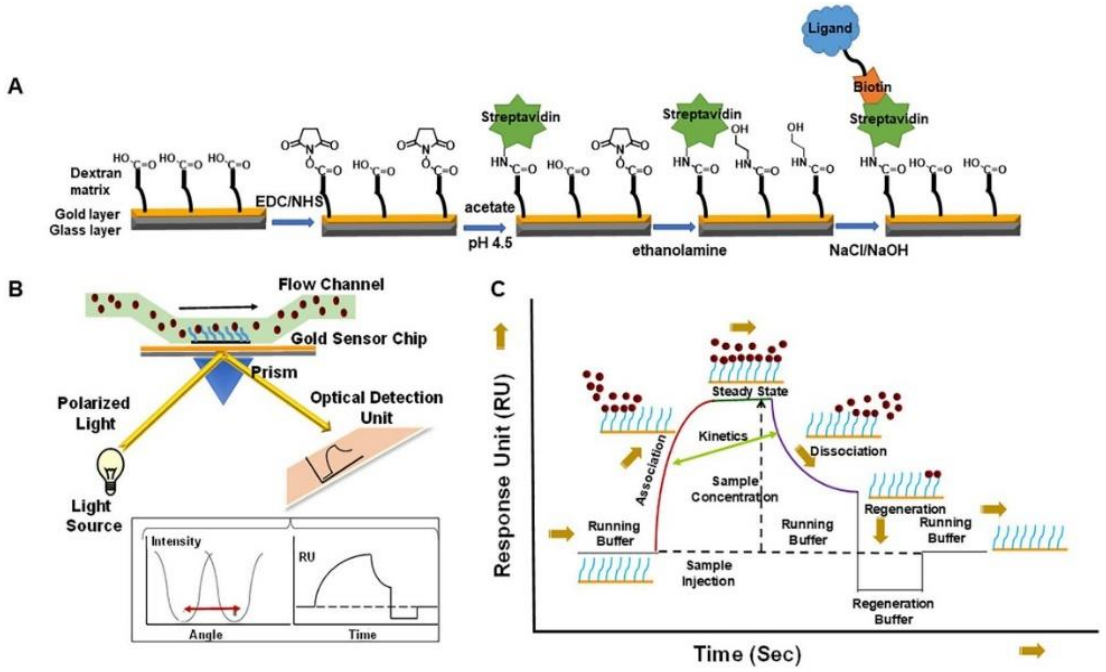

Figure 3.16. Schematic representation of (**A**) immobilization of biotin labelled RNA G4 on CM5 chip, (**B**) biomolecular recognition in the flow cell with changes in the angle of incidence light and (**C**) SPR sensorgrams. Reprinted with permission from ref. [173] Copyright 2019 Elsevier.

Similar to other techniques, the first SPR studies were carried out to evaluate DNA G4/ligand interactions. However, considering the growing interest in RNA G4s, the SPR studies to assess the interaction of these molecules with ligands have increased [173]. The interaction of diamidine derivative ligands with TERRA was recently reported as an example of using the SPR experiments to assess G4/ligand interactions [173]. The results showed a two-site binding model of diamidine derivative, DB1246, toward TERRA G4 **(Figure 3.17)**. The binding affinity was in the nanomolar range with K_{D1} = 3 ± 0.8 nM and K_{D2}= 80 ± 12 nM [173]. Tan et al. also reported the use of SPR experiments to determine the binding affinity of miR-3620-5p G4 toward the ligand sanguinarine [174]. They demonstrated the existence of two binding sites for the binding of sanguinarine to miR-3620-5p G4 structure, and revealed that dissociation constants for those binding

sites are 0.12 μM and 4.83 μM. In general, SPR provides valuable kinetic and affinity information about G4/ligand complexes.

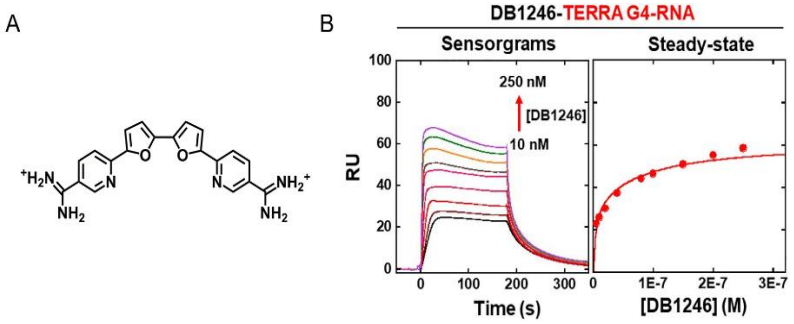

Figure 3.17. (**A**) Sensorgrams with increasing concentrations of the diamidine derivative DB1246. (**B**) Binding curve showing a two-site binding model with K_{D1} = 3 ± 0.8 nM and K_{D2} = 80 ± 12 nM. For the immobilization procedure, ~100 μL of 25 nM of TERRA G4 were prepared in HBS-EP buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 05% *v*/*v* polysorbate 20). The compound DB-1246 (300 μL) was injected into the system from low (10 nM) to high concentration (250 nM). Reprinted with permission from ref. [173] Copyright 2019 Elsevier.

Isothermal Titration Calorimetry (ITC)

ITC is an experimental method that is especially useful to characterize the thermodynamics of molecular interactions. The technique measures the heat generated or consumed upon the interaction of G4s with ligands [175]. While the three-dimensional structure of a G4/ligand complex may reveal specific interactions, ITC provides insights into the molecular forces that drive the association of G4 with ligands [175,176]. Furthermore, it is a robust technique that measures the binding enthalpy (ΔH^o), the free energy of Gibbs (ΔG^o), and entropy (ΔS^o) in a single experiment, and it is a high-accuracy method for measuring binding affinities and stoichiometry [177].

Recently, Funke et al. investigated via ITC the thermodynamic parameters of the binding of indoloquinoline derivatives to G4s located at the *c-MYC* promoter region [178,179,180]. The binding of a ligand to a G4 structure in multiple sites is frequently reported; however, in their recent study, they reported a less common phenomenon that involves the observation of a 2:1 G4/ligand association [180]. Those findings could be explored to control the formation of G4 aggregates.

The isotherms profile of the interaction of TMPyP4 with telomeric G4s was characterized by two well-differentiated binding events [63,181]. Dupont et al. studied the binding of TMPyP4 and its metallated complexes (Co(III), Ni(II), Cu(II), and Zn(II)) toward the 22-mer oligonucleotide of the telomeric G4 [182]. Different metal center coordination

geometries showed distinct thermodynamic signatures **(Figure 3.18)** [182]. Their results also showed that TMPyP4, Ni(II)-TMPyP4, and Cu(II)-TMPyP4 bind tightly to G4 structures through an end-stacking/intercalation mode, whereas Co(III)-TMPyP4 and Zn(II)-TMPyP4 bind poorly to the G4 structure in a mechanism consistent with end-stacking.

A comparative thermodynamic study carried out by Bončina et al. showed a different thermodynamic profile for the binding of ligands with low G4 selectivity over duplexes and ligands highly selective for G4 structures [183,184]. The determination of thermodynamic parameters for RNA G4/ligand interactions is less common and there are only a few reports in the literature. One of those reports characterized the binding of TMPyP4 to DNA and RNA repeats of C9orf72 [185]. For DNA repeats, the ITC experiments revealed that TMPyP4 binds through end-stacking and intercalation, whereas for RNA repeats it binds through two different modes, consistent with groove binding and intercalation.

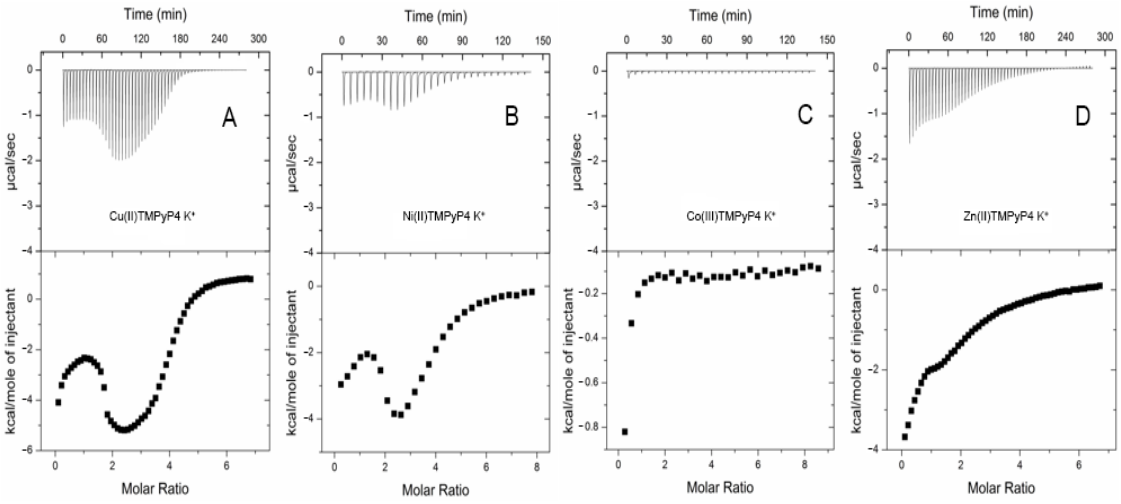

Figure 3.18. Representative ITC titration profiles for the titration experiments between hTel22 and (**A**) Cu(II)-TMPyP4, (**B**) Ni(II)-TMPyP4, (**C**) Co(III)-TMPyP4, and (**D**) Zn(II)-TMPyP4 in 150 mM K^+ BPES buffer. ITC experiments were performed of either 28 (10 µL) or 56 (5 µL) injections of a dilute ligand solution into a dilute hTel22 DNA solution. Adapted with permission from ref. [182] Copyright 2016 Elsevier.

Mass Spectrometry (MS)

MS is an analytical method that is used to measure the mass-to-charge ratio of ions. The technique is highly sensitive, accurate, and versatile in its application and can be used to specifically identify multistranded nucleic acid structures, such as G4s [186,187]. In particular, electrospray ionization mass spectrometry (ESI-MS) has been widely applied to probe G4/ligand interactions [134,186,187]. The main advantage of ESI-MS relies on

the preservation of non-covalent interactions in the gas phase. From a single spectrum and with low amounts of sample, it is possible to determine detailed structural information (e.g., formation, binding affinity, and stoichiometry) of a G4/ligand complex [187]. Monitoring cation binding by ESI-MS provides useful information on ligand binding modes. The number of bound K^+ in the free and ligand-bound forms of G4s is indicative of the number of G-quartets formed for each ligand binding stoichiometry [188]. Furthermore, the changes in the number of K^+ upon ligand binding hint at structural conformational changes [188]. Lecours et al. investigated the interaction of PhenDC3, 360A, and PDS with several G4 structures through this approach [188].

Marchand et al. also investigated the interaction of $(Cu(ttpy)(NO_3)_2)$ with 21 different G4 sequences by ESI-MS [189]. They obtained evidence for the selective and cooperative 2:1 binding of the complex to telomeric G4s. Their results also showed a preference for antiparallel structures with diagonal loops and/or wide-medium–narrow-medium groove-width order. The cooperative binding comes with a conversion of the G4 structure to an antiparallel conformation with three G-quartets.

Ceschi et al. employed ESI-MS to study the binding of perylene derivatives, PIPER and K20, to *KIT2* G4 [190]. Their results revealed a binding stoichiometry of 1:1 on both ligands. They also demonstrated the incapacity of the ligand to bind the unfolded oligonucleotide [190]. Moreover, the preferential binding of K20 to a two-quartet topology was demonstrated, since the most represented complex was the one having a single selectively coordinated K^+ ion, while only a small amount of the complex having two K^+ was detected [190]. On the other hand, PIPER showed a reduced preference for a two-quartet topology since the complexes with one and two K^+ ions were equally represented [190].

Moreover, MS-melting experiments are significantly helpful in explaining inconsistent results between isothermal and melting experiments. The research group of Gabelica used a temperature-controlled nanoelectrospray source to determine the contribution of the thermodynamic parameters to the formation of each stoichiometry G4/ligand [191]. They tested different G4 oligonucleotides in the presence of various G4 ligands such as, PhenDC3, TMPyP4, Cu-ttpy and Pt-ttpy **(Figure 3.19)**. They reported that ligand stoichiometries differ from ligand to ligand [191]. The ability to distinguish between unfolded, partially folded (two-quartet) and fully folded (three-quartet) ligands is helpful since some ligands have the ability to induce significant changes in topology toward partially folded structures [191].

Recently, the same research group demonstrated the applicability of a top–down mass spectrometry sequencing approach, called electron photodetachment dissociation (EDP), in the study of G4/ligand interactions [192]. They found that the G4 ligands PhenDC3 and 360A can replace a K^+ ion and bind close to the central loop of telomeric G4 sequences with the ability to form three-tetrad G4s [192].

Scalabrini et al. developed an approach that increases the G4 sensitivity without affecting physiological folding [193]. They optimized the solvent conditions and reported that isopropanol in a triethylamine/hexafluoroisopropanol mixture significantly increases the sensitivity of the technique [193]. The increased sensitivity enables the use of low concentrations of G4s and their ligands. This study could pave the way for future developments of the technique that permit the high-throughput screening of new G4 ligands [193]. Overall, MS is a powerful quantitative technique for characterizing the G4/ligand complexes structurally.

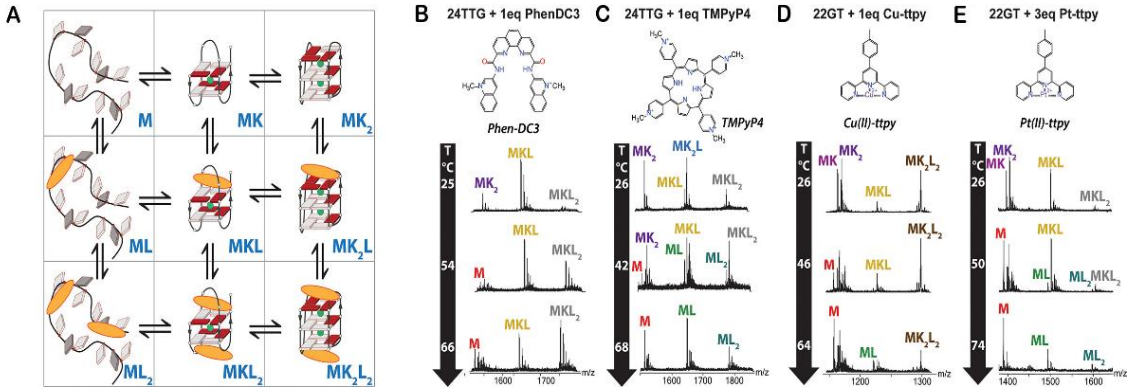

Figure 3.19. (**A**) Chemical equilibria for ligand binding to a G4 forming sequence. Mass spectra at different temperatures of 10 μM of human telomeric DNA sequence (**B**,**C**: 24TTG; **D**,**E**: 22GT) in the presence of 1 or 3 equivalents of (**B**) Phen-DC3, (**C**) TMPyP4, (**D**) Cu-ttpy, and (**E**) Pt-ttpy. The buffer used was 100 mM TMAA and 1 mM KCl. The volume of oligonucleotide solution was approximately 10 μL. Reprinted with permission from ref. [191] Copyright 2017 American Chemical Society.

Taking into account that one of the key future challenges of the G4 field still remains in the design and development of ligands that possess selectivity and affinity for DNA or RNA G4s, the techniques here described offer important comparative information on those parameters. The advantages and disadvantages of each affinity and apparent affinity-based methods were summarized in **Table 3.2**.

Table 3.2. Advantages and limitations of affinity and apparent affinity-based methods for investigating G4/ligand interactions.

Method	Advantages	Limitations
SPR	Acquisition of data in real-time	Requires sophisticated instrumentation and it is costly
	Time efficiency	High dependence of experimental conditions
	High sensitivity	SPR often requires labeling with biotin
	Provides valuable kinetic and affinity information (association (K_a or K_{on}), dissociation (K_d or K_{off}) and equilibrium (K_D) constants)	Requirement of maintaining the G4 structure intact after immobilization
ITC	Provides insights of molecular forces that drive the interaction	High amount of sample
	Provides kinetic and thermodynamically parameters	Ligands should be soluble in water
MS	Provides information on formation, stoichiometry, and binding affinity of G4/ligand complex	Limitations regarding the media

High-Throughput Methods to Investigate G4/Ligand Interactions

FRET-Melting

The Förster resonance energy transfer (FRET) spectroscopy is one of the most used high-throughput methods to study G4/ligand interactions. The technique is highly suitable to evaluate the changes in the G4 folding and unfolding processes **(Figure 3.20)**. In FRET experiments, G4 oligonucleotides are labeled at the 5′- and 3′-end with a donor and acceptor fluorophore, respectively [194]. Briefly, the excited state energy is transferred from the donor to the acceptor via dipole-dipole interactions. Changes in the distance between the two labeled sites result in a measurable energy transfer. The most common FRET pairs include 6-carboxyfluorescein (FAM) as a donor and 6-carboxy tetramethylrhodamine (TAMRA) or Black Hole Quencher 1 (BHQ1) as acceptors. However, care must be taken because modifications at 5′ and/or 3′ may alter G4 folding process. Besides that, the method only measures ligand-induced G4 stabilization and does not provide information about other types of interactions.

Despite this, the method has significant advantages and has been widely used to investigate the binding of several classes of G4 ligands, such as phenanthrolines [195], acridines [114], indoloquinolines [196,197] and pyridine-2,6-dicarboxamide derivatives [198] to various G4 structures.

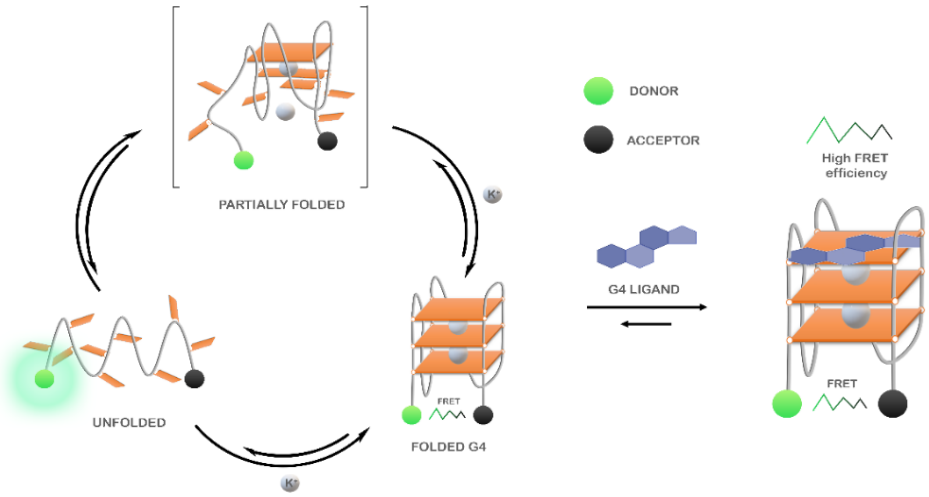

Figure 3.20. Schematic general representation of FRET-melting experiments.

The ability of four phenanthroline polyazamacrocycles to bind and stabilize G4 structures was assessed by FRET-melting [195]. In this study, the ligand (32)phen$_2$N$_4$ showed a higher thermal stabilization effect toward c-*MYC* and 22AG G4 sequences at 10 molar equivalents (17.2 °C and 20.3 °C, respectively). The ligands' selectivity was also evaluated toward other G4s located in oncogene promotors such as *KRAS*-21R, *VEGF* and *c-KIT87*. The results revealed a preference of the ligands toward *KRAS*-21R over *VEGF* and *c-KIT87*. Furthermore, the four ligands demonstrated selectivity and specificity toward G4s over duplex DNA. Indeed, in the presence of 50 molar equivalents of a competitor double-stranded sequence, the stabilizer effect of the ligands (16)phenN$_4$ and (32)phen$_2$N$_4$ in 22AG does not change significantly **(Figure 3.21A)**.

The stabilization and selectivity of acridine derivatives toward a G4 sequence found in pre-miRNA-92b were recently assessed using FRET-melting [114]. The acridine derivative C$_8$, at one molar equivalent, showed the ability to stabilize the structure in 6.3 °C. However, in the presence of 500 nM of RNA-binding domains 1 and 2 of nucleolin, the G4 sequence was stabilized in 22 °C. The remarkable synergistic effect can be attributed to an additional stabilization of the protein by C$_8$.

Lavrado et al. also reported the synthesis of novel indolo(3,2-b)quinolines with one, two, or three basic side chains and evaluated its ability to bind and stabilize G4 sequences by FRET-melting [196]. For all the ligands, the thermal stabilization is concentration-dependent. Later, they reported the synthesis of indolo(3,2-c)quinolines and evaluated their interaction toward *KRAS*-21R sequence and a 26-mer hairpin double-stranded DNA sequence (dsDNA) [197]. FRET-melting results revealed that compounds 3d and 3e have promising stabilizing effects in *KRAS*-21R (19.9 °C and 22.0

°C, respectively) sequence but not in the dsDNA (5.6 °C and 6.5 °C, respectively). They also reported the selectivity of the compound 3e in the presence of increasing concentrations of different non-fluorescent competitors (26mer ds-DNA (ds26) or the human telomeric sequence (HT21)). The results revealed a remarkable selectivity of the compound 3e towards *KRAS*-21R **(Figure 3.21B)**.

In a recent study, Cadoni et al. reported the synthesis of Pyridine-2,6-dicarboxamide derivatives and evaluated their ability to stabilize G4 structures using FRET-melting [198]. Their results revealed that all the N-methylquinolinium derivatives displayed a preference for G4 sequences when compared to a duplex DNA sequence (T-loop). Moreover, the results showed a noteworthy ability of compound 2a to stabilize G4 structures even at lower concentrations of the ligand. In fact, at 2.5 molar equivalents of 2a, the sequence of the *KRAS* G4 was stabilized in 32.8 ± 0.2 °C. They also evaluated the selectivity of the methylated compounds in the presence of double-stranded DNA. The data showed a neglectable effect of the competitor in the G4/ligand complex, which proved the high selectivity of the ligands toward G4 structures **(Figure 3.21C)**.

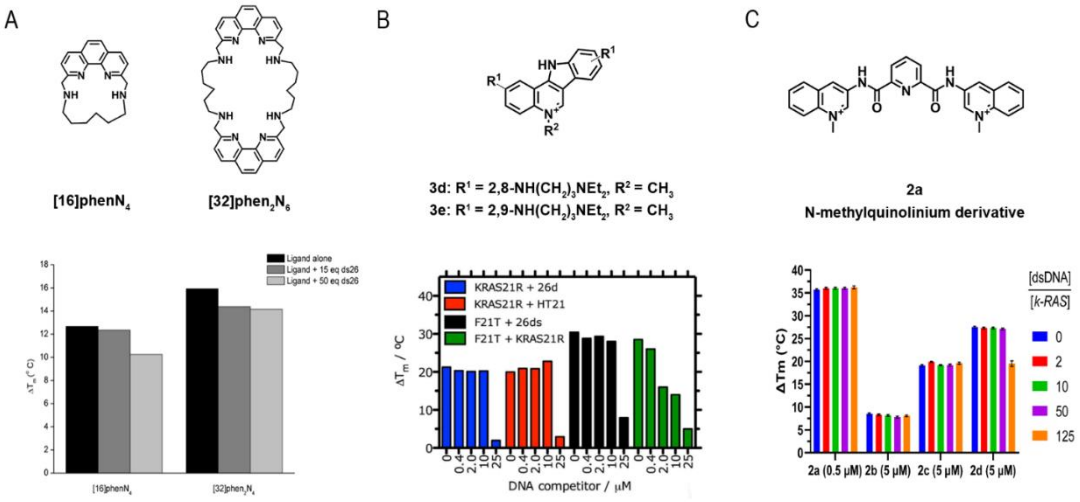

Figure 3.21. (**A**) Chemical structures of (16)phenN_4 and (32)phen$_2N_4$ and FRET-melting stabilization by (16)phenN_4 and (32)phen$_2N_4$ (1 μM) of the labelled F21T sequence (0.2 μM) in buffer 10 mM KCl + 90 mM LiCl and in the presence of increasing concentrations of competitor ds26 (0, 15 and 50 equivalents). Reproduced with permission from [195]. Copyright 2018 Royal Society of Chemistry. (**B**) Chemical structures of indolo(3,2-c)quinoline derivatives 3d and 3e, and FRET-melting competition assay with 3e (1 μM) in complex with *KRAS*21R and F21T G4s (0.2 μM), with increasing concentrations of non-labeled 26ds, HT21 or *KRAS*21R (0.4 to 25 μM) competitor, in cacodylate buffer, pH 7.4 supplemented with K^+. Reprinted from ref. [197], (**C**) Chemical structure of N-methylquinolinium derivative 2a and FRET-melting competition assay results for compounds 2a–d at 0.5 μM (2a) and 5 μM (2b–d) complexed with *KRAS* G4 (0.2 μM), challenged with increasing concentrations of non-labeled 26ds DNA (dsDNA) (0.4 μM to 25 μM) competitor. Experiments were performed in cacodylate buffer, pH 7.4, supplemented with 60 mM K^+. Reprinted from ref. [198].

In the last few years, some modifications of the method have been carried out to improve the data output and surpass the early limitations of the technique.

A modified transition-FRET method was employed by Noureini et al. to estimate the selectivity of isoquinoline alkaloids from *Chelidomium majus* toward a human telomeric G4 [199]. The novel FRET method consists of two transition steps: the first transition step is the melting temperature of the double helix structure, whereas the second transition state belongs to G4 denaturation. Therefore, in the presence of selective G4 ligands, an increase in the melting temperature of the second step is observed, as well as a decrease in the fluorescence intensity of the first step.

Rakers et al. developed a novel microfluidic platform to synthesize novel G4 binders and evaluate their affinity and using a real-time online FRET assay [200]. Furthermore, their strategy enabled the high-throughput online analysis of several G4 binders. Thus, this study could trigger novel strategies that combine the synthesis and the assessment of G4 binders in a unique device.

An important feature of FRET-melting is its remarkable competitive nature [201], when compared with the other techniques mentioned in this review. Indeed, most of the techniques used to investigate G4/ligand interactions are mainly comparative, which makes it a strong competitive tool to analyze selectivity and specificity.

In a very recent report, Luo et al. also explored the competitive nature of FRET-melting and studied whether the interaction between a fluorescent G4-forming oligonucleotide and PhenDC3 is modified by a competitor sequence added in excess [202]. Sixty-five sequences with a known structure were tested to validate this FRET-melting competition (FRET-MC) assay. In this case, if the unlabeled competitor has a high affinity for PhenDC3, it will sequester a significant fraction of the compound, which will be no longer available for fluorescent G4-forming oligonucleotide stabilization, leading to a decrease in T_m. On the other hand, if the competitor sequence is unable to sequester the ligand, T_m remains high.

G4-FID Screening

The fluorescence intercalator displacement (FID) assay is based on the competitive displacement of thiazole orange (TO) fluorescent probe from various G4 and duplex by increasing amounts of the ligand subjected to evaluation.

Monchaud et al. have optimized the original FID assay to study G4/ligand complexes [203,204,205]. The technique does not require modified oligonucleotides and it is based on the loss of fluorescence of the bound intercalator, thiazole orange (TO), upon displacement induced by a G4 ligand **(Figure 3.22)**.

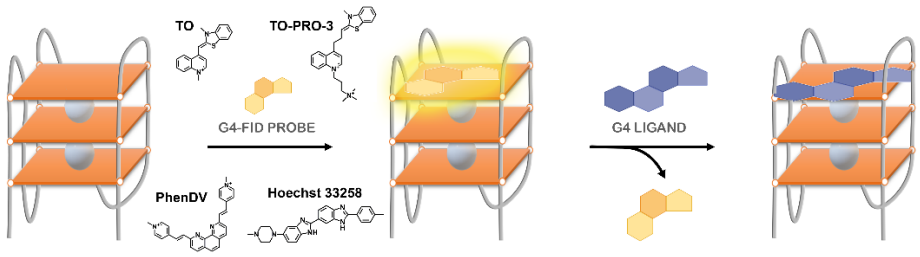

Figure 3.22. Schematic general representation of G4-FID experiments. Chemical structures of TO, TO-PRO-3, Hoechst 33258 and PhenDV.

However, the traditional method had some limitations that were correlated with its applicability. To surpass those drawbacks, Largy et al. developed the high-throughput version of the G4-FID assay by analyzing an extensive set of G4 ligands in a 96-well plate **(Figure 3.23)** [206]. Further, the spectral range of the assay was expanded by using two other fluorescent probes, Hoechst 33258 and TO-PRO-3, a minor groove binder that emits blue fluorescence and an intercalator that emits red fluorescence, respectively [206]. These fluorescent light-up probes emit low fluorescence when free in a solution, but present a 1000-fold increase when complexed with DNA. The binding affinity of several ligands toward a significant number of DNA/RNA sequences was assessed by G4-FID assay [207]. However, considering the modest selectivity of the fluorescent probes for DNA G4s over duplex DNA and their ambiguous binding mode, alternative dyes have been proposed. PhenDV-based is a bisvinylpyridinium dye that would bind to DNA G4s in a similar mode to that of PhenDC3 [208]. The readout was carried out by fluorescence enhancement and not by quenching [208].

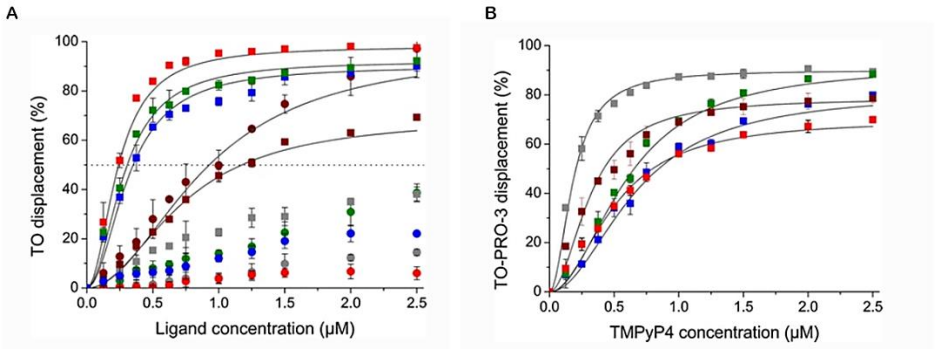

Figure 3.23. (**A**) Representative G4-FID curves obtained with MMQ1 (gray), MMQ3 (brown), Pt-ttpy (red), 360A (blue) and Phen-DC3 (green) with 22AG.K (squares) and ds26 (circles) using TO as fluorescent probe. (**B**) Representative G4-FID curves of TMPyP4 with 22AG.Na (blue), 22AG.K (red), c-kit1 (brown), c-kit2 (gray), and ds26 (green) with TO-PRO-3 as fluorescent probe. For each well of the microplate were added: (i) K^+- or Na^+-buffer solution; (ii) oligonucleotides (5 μM) and fluorescent probe (TO or TO-PRO-3; 10 μM for G4-DNA or 15 μM for ds26) and (iii) an extemporaneously prepared 5 μM ligand solution in K^+- or Na^+-buffer. The final volume in each well was 200 μL. Reprinted with permission from ref. [206] Copyright 2011 Springer Nature.

The same strategy of fluorescence enhancement was employed by using the dimeric cyanine small molecule, B6,5, which can be used as a probe in fluorescence displacement assays to screen G4 specific ligands for DNA and RNA G4s [209].
The methods used for processing a huge quantity of data are often ineffective for easily visualizing the binding selectivity of G4 ligands. Villar-Guerra et al. described a rapid FID (R-FID) assay which, when combined with a novel application of principal component analysis (PCA) and hierarchical cluster analysis (HCA), represents a powerful tool for the analysis of large FID data sets [210].
Most of the efforts to improve the G4-FID experiments have been focused on developing novel highly fluorescent probes that could be more selective and specific than the existing alternatives.

Affinity Chromatography Screening

Affinity chromatography is a highly effective technique that enables the selective isolation of target molecules [211]. The method relies on selective and reversible interactions that undergo with the binding agent and the target. Thus, the release of the target occurs under specific conditions, where it does not have an opportunity to rebound, and according to its binding affinity toward the recognition agent. Nowadays, the technique has been used in a wide variety of applications across different fields, which include the isolation of G4s from complex mixtures (plasma, serum) [212] **(Figure 3.24A)** and the identification of G4 ligands [213] **(Figure 3.24B)**. The G4 ligand, N-methyl mesoporphyrin IX (NMM) was coupled to a Sepharose matrix and demonstrated the ability to selectively bind DNA (albeit with distinct binding affinities for different topologies) [212]. This work demonstrated the feasibility and applicability of the method but is somewhat limited by the incomplete recovery of the bound DNA, the use of a ligand with a lack of selectivity for G4 structures, and the employment of batch affinity adsorption rather than column affinity chromatography. Later, the perylene derivative ligand (*N*,*N*′-Bis-(2-(amino)ethyl)-3,4,9,10-perylenetetracarboxylic acid diimide, Pery01) was coupled to a matrix of polyglycidylmethacrylate (PGMA) to isolate DNA G4s by column affinity chromatography [214]. The support displays a preference for parallel G4s and demonstrated capability to isolate parallel G4s from plasma. Ferreira et al., by using the same column affinity chromatography strategy, functionalized a matrix of Sepharose CL-6B with a new naphthalene derivative (L1) and the ability of the support to selectively bind to different G4 sequences was evaluated [213]. The results showed a remarkable binding affinity for parallel G4s over non-parallel G4s as well as other conformations.

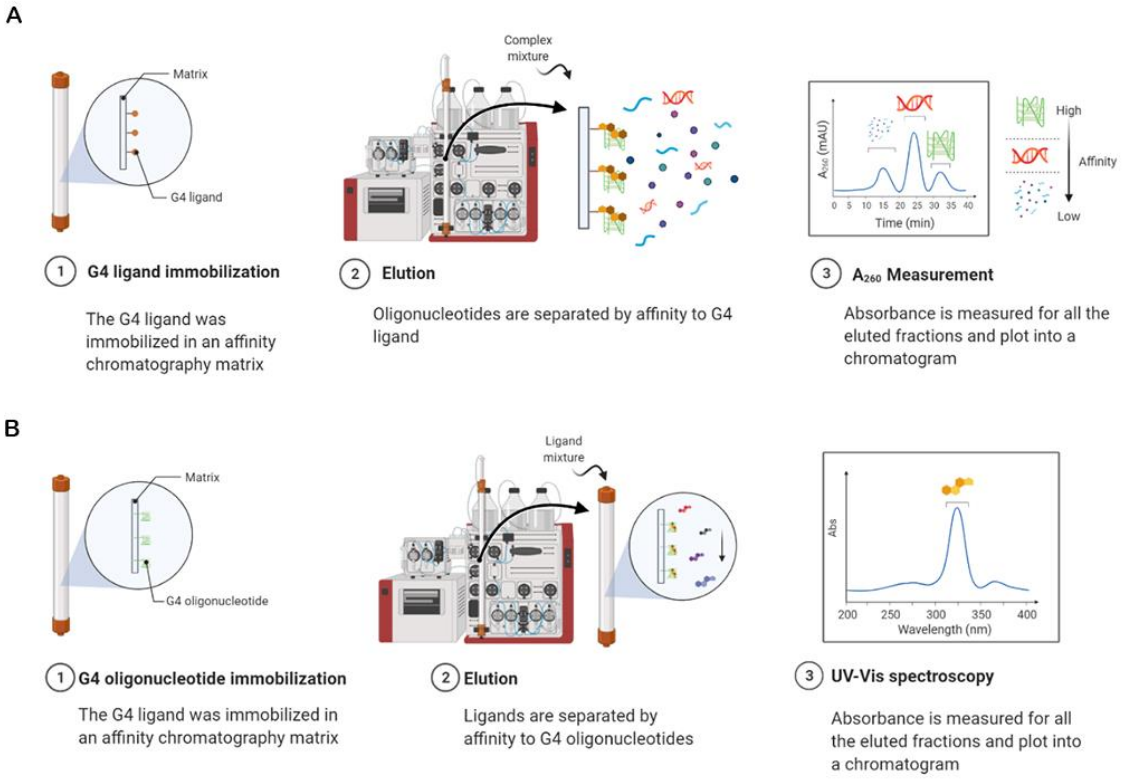

Figure 3.24. (**A**) Scheme of affinity chromatography strategy to purify G4 sequences based on a ligand affinity chromatography support. (**B**) Scheme of affinity chromatography strategy to purify G4 ligands based on a G4 affinity chromatography support.

Musumeci et al. developed an identical approach, named G4 on Oligo Affinity Support (G4-OAS) assay [215]. However, instead of flowing throughout the column a complex mixture containing G4-forming sequences, the method consists of flowing solutions of probable G4 binders through a polystyrene resin functionalized with a G4-forming sequence. The compounds having a high affinity for the G4 sequence are retained in the support, whereas those with no or low affinity are eluted first and quantified by UV measurements **(Figure 3.25)**. The method is simple and rapid, but there are some important limitations to point out: (i) the unspecific binding of the ligand to the resin and (ii) the absence of structural information about G4 conformation. The high-throughput nature of the method was proved later by using the virtual screening and the G4-OAS assay in tandem [216]. Recently, to overcome those initial limitations, the authors developed a novel functionalization method by employing an approach of solid-phase synthesis of the oligonucleotides on Controlled Pore Glass (CPG) support [217]. The strategy was further employed to analyze a set of naphthalene diimide derivatives by using the sequence d[AGGG(TTAGGG)7] (tel46), which folds into two consecutive G4s [218]. The G4-CPG screening identified the ligand NDI-5 as the most notable ligand of the tel46 sequence.

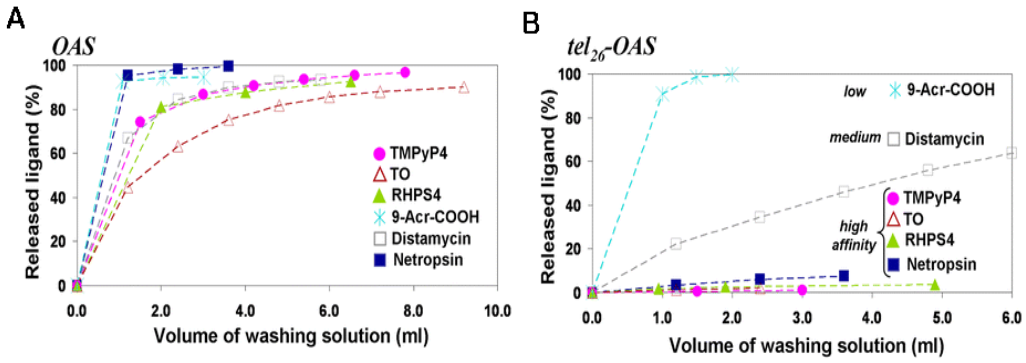

Figure 3.25. Amount of the released ligands, expressed as percentage of the quantity loaded (**A**) on OAS support and (**B**) on tel26-OAS support, as a function of the washing solution volume (100 mM KCl/15% DMSO). Reprinted from ref. [215].

Microarrays-Based Screening

Microarrays are a fast, simple, and high-throughput method that could be employed to screen ligand interactions with DNA and RNA sequences. The well-known G4 ligand PDS has recently been fluorescently labeled and evaluated toward a hundred thousand G4 sequences, which covalently attached to the surface of glass slides [219]. The binding selectivity of the G4 ligand BMVC to several G4 structures was also assessed by a custom G4 microarrays platform and revealed a preference for parallel G4s [220].

The reverse strategy was employed in small-molecule microarrays (SMMs). The technology relies on the immobilization of the ligands by covalent interaction followed by incubation with the fluorescently labeled oligonucleotide. Although the method has been broadly used to identify protein ligands, it has only recently been transposed to analyze G4/ligand interactions. Felsenstein et al. employed SMMs for the first time in the G4 field and identified a benzofuran derivative as a specific *c-MYC* G4 binder among 20,000 compounds [221]. This initial study has led to the optimization of the ligand in a more potent analog (DC-34), which showed interesting *in vivo* potency and pharmacological properties in cancers expressing *c-MYC* [222].

High throughput screening is an essential strategy of early-stage drug discovery, and methods and technologies have seen many improvements and innovations over recent years. In many cases, the early high-throughput methods have been optimized and recently, some improvements have been reported in the characterization of G4/ligand interactions. The advantages and disadvantages of the high-throughput methods described in this review are summarized in **Table 3.3**.

Table 3.3. Advantages and limitations of high-throughput methods for investigating G4/ligand interactions.

Method	Advantages	Limitations
FRET-melting	Simplicity	Fluorescently labelled oligonucleotides It only measures ligand-induced G4 stabilization, while other types of interactions are not detected
	Small amount of sample	
	Real time monitoring	
G4-FID	Simplicity	Ambiguous binding mode of used probes
	Small amount of sample	Compatibility of the oligonucleotides with the fluorescent probe
Affinity Chromatography	Selective and reversible interactions that undergo with the ligand and G4	Unspecific binding of the ligand to the resin
	Column chromatography allow real-time monitoring	The absence of structural information about G4 conformation
Microarrays	Small amount of sample	Expensive Specialized equipment Fluorescently labeled molecules
	Massive parallel screening	

In addition to the techniques here approached and used for investigating G4/ligand interactions, other robust and efficient biophysical, biochemical and molecular biology techniques are available to describe G4/ligand interactions, such as surface-enhanced Raman spectroscopy [223], single-molecule fluorescence imaging [97,224], equilibrium dialysis [225], gel electrophoresis [114,226], qPCR-stop assay [227], Taq polymerase stop assay [228] and TRAP assay [229]. Furthermore, other high-throughput methods are emerging, such as pull-down assays [230] and Affinity Selection-Mass Spectrometry (ALIS) [231]. These are not described in detail in this review but, depending on the applicability, could be employed to evaluate the molecular interactions between G4s and ligands.

Conclusions

G4s are recognized as important structural motifs in a wide range of biological processes. Over the past two/three decades, G4s have been intensively studied in the presence of a diverse family of ligands. As a result of these studies, important atomic-detailed information concerning the principal forces governing G4/ligand interactions was revealed. In addition, the results also point out that fundamental chemical aspects remain to be fine-tuned concerning the design and development of ligands that possess selectivity and affinity for DNA or RNA G4s.

From a structural point of view, CD, NMR and X-ray crystallography are techniques with many strengths, with individual limitations rather well complemented among them.

The simple and fast nature of CD makes it one of the primary choices to obtain low-resolution information about structural details. CD is commonly used at an initial stage of the G4/ligand interaction characterization. Users can easily obtain information about the G4 topology and often preliminary information on the binding modes of the ligands. Furthermore, almost all the studied G4 ligands were investigated in terms of their ability to stabilize G4s by CD-melting experiments. Despite the valuable information provided by CD, the full structural characterization of a given G4/ligand complex can only be achieved by using high-resolution structural methods such as NMR or X-ray Crystallography.

For instance, NMR is a solution-based technique that can analyze the dynamics of the complexes on a broad range of time scale (ps-s), whereas X-ray crystallography provides insight into the most thermodynamically stable form of a complex. Under some circumstances using both may be the best option, as some ligands tend to induce aggregates that precipitate in the NMR tube over time and resolution is lost. For some polymorphic oligonucleotides the other way around is preferable, and for those cases where both techniques can be used, they represent the strongest solution to investigate G4/ligand complexes at atomic detail. However, when compared to CD, both NMR and X-ray crystallography have the disadvantage that they often require milligram amounts of material. In the last decade, the lower concentration limit in NMR has been significantly reduced due to the combination of increased magnetic field strength and other sensitivity enhancement technologies, so we expect that the determining of the high-resolution structure of G4s and complexes will continue to be a very active field of research.

Joint analysis using high-resolution NMR and X-ray studies have provided invaluable structural insights into the interaction and binding modes of G4/ligand complexes in the biological context. These structural studies have shown the atomic details of the interaction between G4s and their ligands, allowing researchers to analyze structure–function relationships to rationally design and develop new specific and selective G4 ligands to treat a variety of diseases. However, these structural techniques are not sufficient to characterize completely the driving forces of a given interaction. Indeed, one key challenge still remaining to be addressed is the design and development of ligands that possess selectivity and affinity for DNA or RNA G4s. Therefore, other affinity- and

apparent affinity-based methods such as SPR, ITC and MS have been intensively used in the G4 field.

SPR was shown to be a preferable choice in comparison to other *in vitro* techniques such as ELISA, which does not allow an absolute measurement of kinetic constants, is less sensitive and more expensive in terms of required samples and time. However, the SPR technology requires that one of the analyzed molecules is immobilized on a functionalized metal surface and this often leads to the heterogeneous orientation of the ligand molecule and the alteration of its activity, and a lot of parameters must be carefully controlled, increasing the difficulty of the analysis of the collected data.

Until now, ITC is the only available experimental method to directly measure binding enthalpies without any chemical modification or immobilization of the interacting components. Besides that, ITC also provides a complete characterization of the thermodynamic parameters of G4/ligand interactions.

MS is a powerful technique that provides a direct readout of the stoichiometries of G4/ligands complexes versus duplex formed, including minor species, and it is a very attractive characteristic of this technique, as well as which it does not require modified oligonucleotides and allows using a broad variety of G4s versus duplex. However, electrospray is incompatible with the presence of alkaline salts such as NaCl or KCl and must be replaced by ammonium acetate, entailing changes in the G4 structures and consequently in the binding of ligands. ITC and MS, as well as SPR, require specific know-how and relatively expensive equipment.

The newly emerged high-throughput methods combined with other biophysical approaches created a new framework, which allows the fast and cheap development and characterization of G4/ligand complexes.

For instance, FRET-melting, which is carried out in presence of a competitive duplex sequence, has been adapted for an overwhelming number of G4 ligands. This method is rapid and convenient but requires modified oligonucleotides, and possible interferences between ligands and the fluorescent labels represent the two main limitations of FRET-melting. A suitable alternative/complementary method of FRET-melting is G4-FID, which allows the screening of an important number of unmodified G4 structures with wide-spread equipment i.e., a qPCR devices, fluorimeters or microplate readers.

Affinity chromatography emerged as a simple and fast method to analyze the interaction of ligands with G4s. The method relies on simple and efficient assays by immobilizing

G4s on a solid support and is particularly useful to select G4 ligands with specificity and selectivity for a given G4 structure. Further, the method could be used to separate those G4 ligands from a complex mixture. The reverse strategy, in which the ligand is immobilized, could also be employed to separate G4s from complex biological samples.

Microarrays have only recently been used to characterize G4/ligand interactions and few reports are as yet available in the literature. The method is fast and straightforward and can be used to analyze a huge quantity of G4 structures, ligands, and conditions. However, its use as a routine technique to analyze G4/ligand interactions could be hindered due to its high cost and need for specialized equipment.

In general, the modification of low-throughput methods to become high-throughput could significantly impact the field. Such progress will naturally occur and lead to the characterization of a great number of G4/ligand complexes.

One of the main challenges of the field is that, at the moment, despite the significant advances in recent years, none of the techniques provides a full characterization of G4/ligand complexes. The complete framework is only achieved with the use of several techniques, which can sometimes be difficult to assemble in a single research institution.

Overall, this review intends to improve the knowledge and understanding of the currently developed methods used to assess G4/ligand interactions, and at the same time provide a broad framework of the chemical features of the G4 ligands available for diagnostic and therapeutic purposes.

Funding

Tiago Santos acknowledges Fundação para a Ciência e Tecnologia (FCT) for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013). This work was supported by PESSOA program ref. 5079 and project “Projeto de Investigação Exploratória” ref. IF/00959/2015 entitled “NCL targeting by G-quadruplex aptamers for cervical cancer therapy” financed by Fundo Social Europeu e Programa Operacional Potencial Humano. Thanks are due to FCT/MCT for the financial support of the CICS-UBI UIDB/00709/2020 research unit and to the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016), through national funds and, where applicable, supported by European Investment Funds FEDER through COMPETE 2020, POCI, PORL and PIDDAC.

References

1 Spiegel, J. *et al.* (2020) The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* 2, 123–136

2 Gellert, M. *et al.* (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U. S. A.* 48, 2013–2018

3 Arnott, S. *et al.* (1974) Structures for polyinosinic acid and polyguanylic acid. *Biochem. J.* 141, 537–543

4 Burge, S. *et al.* (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* 34, 5402–5415

5 Largy, E. *et al.* (2016) Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In *Metal Ions in Life Sciences* 16pp. 203–258

6 Ma, Y. *et al.* (2020) Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem. Biophys. Res. Commun.* 531, 3–17

7 Ngoc Nguyen, T.Q. *et al.* (2020) Duplex formation in a G-quadruplex bulge. *Nucleic Acids Res.* 48, 10567–10575

8 Meier, M. *et al.* (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res.* 46, 5319–5331

9 Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* 33, 2908–2916

10 Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 35, 406–413

11 Maizels, N. and Gray, L.T. (2013) The G4 Genome. *PLoS Genet.* 9, e1003468

12 Bedrat, A. *et al.* (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746–1759

13 Brázda, V. *et al.* (2019) G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495

14 Hon, J. *et al.* (2017) pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics* 33, 3373–3379

15 Klimentova, E. *et al.* (2020) PENGUINN: Precise Exploration of Nuclear G-Quadruplexes Using Interpretable Neural Networks. *Front. Genet.* 11, 1287

16 Garant, J.M. *et al.* (2017) Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* 33, 3532–3537

17 Chambers, V.S. *et al.* (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33, 877–881

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

19 Bryan, T.M. (2020) G-quadruplexes at telomeres: Friend or foe? *Molecules* 25, 3686

20 Balasubramanian, S. *et al.* (2011) Targeting G-quadruplexes in gene promoters: A novel anticancer strategy? *Nat. Rev. Drug Discov.* 10, 261–275

21 Hänsel-Hertsch, R. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* 48, 1267–1272

22 Spiegel, J. *et al.* (2021) G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* 22, 117

23 Lago, S. *et al.* (2021) Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* 12, 3885

24 Jara-Espejo, M. and Peres Line, S.R. (2020) DNA G-quadruplex stability, position and chromatin accessibility are associated with CpG island methylation. *FEBS J.* 287, 483–495

25 Shen, J. *et al.* (2021) Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol.* 22, 143

26 Komůrková, D. *et al.* (2021) G-quadruplex structures colocalize with transcription factories and nuclear speckles surrounded by acetylated and dimethylated histones H3. *Int. J. Mol. Sci.* 22, 1–18

27 Hoffmann, R.F. *et al.* (2016) Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res.* 44, 152–163

28 Varshney, D. *et al.* (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 21, 459–474

29 Lee, W.T.C. *et al.* (2021) Single-molecule imaging reveals replication fork coupled formation of G-quadruplex structures hinders local replication stress signaling. *Nat. Commun.* 12, 2525

30 Tran, P.L.T. *et al.* (2021) Folding and persistence times of intramolecular G-quadruplexes transiently embedded in a DNA duplex. *Nucleic Acids Res.* 49, 5189–5201

31 Murat, P. *et al.* (2018) RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. *Genome Biol.* 19, 229

32 Dutta, A. *et al.* (2021) Promoter G-quadruplex favours epigenetic reprogramming-induced atypical expression of ZEB1 in cancer cells. *Biochim. Biophys. Acta - Gen. Subj.* 1865, 129899

33 Marsico, G. *et al.* (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* 47, 3862–3874

34 Zaccaria, F. and Fonseca Guerra, C. (2018) RNA versus DNA G-Quadruplex: The Origin of Increased Stability. *Chem. - A Eur. J.* 24, 16315–16322

35 Vannutelli, A. *et al.* (2020) Where are G-quadruplexes located in the human transcriptome? *NAR Genomics Bioinforma.* 2, lqaa035

36 Biffi, G. *et al.* (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* 6, 75–80

37 Tassinari, M. *et al.* (2021) Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* 49, 3617–3633

38 Kwok, C.K. *et al.* (2016) RG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* 13, 841–844

39 Yang, S.Y. *et al.* (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* 9, 4730

40 Yang, X. *et al.* (2020) RNA G-quadruplex structures exist and function *in vivo* in plants. *Genome Biol.* 21, 226

41 Shao, X. *et al.* (2020) RNA G-quadruplex structures mediate gene regulation in bacteria. *MBio* 11, e02926-19

42 Renard, I. *et al.* (2019) Small-molecule affinity capture of DNA/RNA quadruplexes and their identification *in vitro* and *in vivo* through the G4RP protocol. *Nucleic Acids Res.* 47, 502–510

43 Maizels, N. (2015) G4-associated human diseases. *EMBO Rep.* 16, 910–922

44 De Magis, A. *et al.* (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 116, 816–825

45 Zell, J. *et al.* (2021) DNA folds threaten genetic stability and can be leveraged for chemotherapy. *RSC Chem. Biol.* 2, 47–76

46 Maffia, A. *et al.* (2020) From R-loops to G-quadruplexes: Emerging new threats for the replication fork. *Int. J. Mol. Sci.* 21, 1506

47 Rodriguez, R. *et al.* (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* 8, 301–310

48 Yadav, P. *et al.* (2021) G-quadruplex structures in bacteria: Biological relevance and potential as an antimicrobial target. *J. Bacteriol.* 203, e0057720

49 Ruggiero, E. and Richter, S.N. (2020) Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. In *Annual Reports in Medicinal Chemistry* 54pp. 101–131

50 Abiri, A. *et al.* (2021) Unlocking G-quadruplexes as antiviral targets. *Pharmacol. Rev.* 73, 897–923

51 Saranathan, N. and Vivekanandan, P. (2019) G-Quadruplexes: More Than Just a Kink in Microbial Genomes. *Trends Microbiol.* 27, 148–163

52 Seifert, H.S. (2018) Above and beyond Watson and Crick: Guanine Quadruplex Structures and Microbes. *Annu. Rev. Microbiol.* 72, 49–69

53 Metifiot, M. *et al.* (2014) G-quadruplexes in viruses: Function and potential therapeutic applications. *Nucleic Acids Res.* 42, 12352–12366

54 Perrone, R. *et al.* (2017) Mapping and characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions. *Sci. Rep.* 7, 5743

55 Tlučková, K. *et al.* (2013) Human papillomavirus G-quadruplexes. *Biochemistry* 52, 7207–7216

56 Zhao, C. *et al.* (2021) Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19? *Angew. Chemie - Int. Ed.* 60, 432–438

57 Asamitsu, S. *et al.* (2019) Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules* 24, 429

58 Sun, D. *et al.* (1997) Inhibition of human telomerase by a G-Quadruplex-Interactive compound. *J. Med. Chem.* 40, 2113–2116

59 Duarte, A.R. *et al.* (2018) Design of Modular G-quadruplex Ligands. *ChemMedChem* 13, 869–893

60 Umar, M.I. *et al.* (2019) G-quadruplex-based fluorescent turn-on ligands and aptamers: From development to applications. *Molecules* 24, 2416

61 Li, Q. *et al.* (2013) G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* 41, D1115–D1123

62 Sun, Z.Y. *et al.* (2019) Developing novel G-quadruplex ligands: From interaction with nucleic acids to interfering with nucleic acid–protein interaction. *Molecules* 24, 396

63 Martino, L. *et al.* (2009) Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *J. Phys. Chem. B* 113, 14779–14786

64 Read, M. *et al.* (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4844–4849

65 Zuffo, M. *et al.* (2018) More is not always better: Finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res.* 46, e115

66 Asamitsu, S. *et al.* (2019) Ligand Design to Acquire Specificity to Intended G-Quadruplex Structures. *Chem. - A Eur. J.* 25, 417–430

67 Neidle, S. (2020) Challenges in developing small-molecule quadruplex therapeutics. In *Annual Reports in Medicinal Chemistry* 54pp. 517–546

68 Drygin, D. *et al.* (2009) Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis. *Cancer Res.* 69, 7653–7661

69 Xu, H. *et al.* (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* 8, 14432

70 Masud, T. *et al.* (2021) Ubiquitin-mediated DNA damage response is synthetic lethal with G-quadruplex stabilizer CX-5461. *Sci. Rep.* 11, 9812

71 Xu, X. *et al.* (2021) Therapeutic efficacy of the novel selective RNA polymerase I inhibitor CX-5461 on pulmonary arterial hypertension and associated vascular remodelling. *Br. J. Pharmacol.* 178, 1605–1619

72 Neidle, S. (2010) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* 277, 1118–1125

73 Burger, A.M. *et al.* (2005) The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* 65, 1489–1496

74 Machireddy, B. *et al.* (2019) Binding of BRACO19 to a telomeric G-quadruplex DNA probed by all-atom molecular dynamics simulations with explicit solvent. *Molecules* 24, 1010

75 Hamon, F. *et al.* (2011) An acyclic oligoheteroaryle that discriminates strongly between diverse G-Quadruplex topologies. *Angew. Chemie - Int. Ed.* 50, 8745–8749

76 Răsădean, D.M. *et al.* (2017) Amino-Acid-Derived Naphthalenediimides as Versatile G-Quadruplex Binders. *Chem. - A Eur. J.* 23, 8491–8499

77 Hu, M.H. *et al.* (2017) Specific targeting of telomeric multimeric G-quadruplexes by a new triaryl-substituted imidazole. *Nucleic Acids Res.* 45, 1606–1618

78 Rigo, R. *et al.* (2017) G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1399–1413

79 Micco, M. *et al.* (2013) Structure-based design and evaluation of naphthalene diimide G-quadruplex ligands as telomere targeting agents in pancreatic cancer cells. *J. Med. Chem.* 56, 2959–2974

80 Marchetti, C. *et al.* (2018) Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. *J. Med. Chem.* 61, 2500–2517

81 Carvalho, J. *et al.* (2018) Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex. *Biochimie* 144, 144–152

82 Shin-ya, K. *et al.* (2001) Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. *J. Am. Chem. Soc.* 123, 1262–1263

83 Sullivan, H.J. *et al.* (2018) Binding of Telomestatin, TMPyP4, BSU6037, and BRACO19 to a Telomeric G-Quadruplex-Duplex Hybrid Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *ACS Omega* 3, 14788–14806

84 Gavathiotis, E. *et al.* (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat. *J. Mol. Biol.* 334, 25–36

85 Mulholland, K. *et al.* (2017) Binding modes and pathway of RHPS4 to human telomeric G-quadruplex and duplex DNA probed by all-atom molecular dynamics simulations with explicit solvent. *Phys. Chem. Chem. Phys.* 19, 18685–18694

86 Muoio, D. *et al.* (2018) Naphthalene diimide-derivatives G-quadruplex ligands induce cell proliferation inhibition, mild telomeric dysfunction and cell cycle perturbation in U251MG glioma cells. *FEBS J.* 285, 3769–3785

87 Marchetti, C. *et al.* (2015) Macrocyclic naphthalene diimides as G-quadruplex binders. *Bioorganic Med. Chem.* 23, 3819–3830

88 Hu, M.H. *et al.* (2019) Development of a Smart Fluorescent Sensor That Specifically Recognizes the c-MYC G-Quadruplex. *Anal. Chem.* 91, 2480–2487

89 Vummidi, B.R. *et al.* (2013) Fluorescent Probes for G-Quadruplex Structures. *ChemBioChem* 14, 540–558

90 Largy, E. *et al.* (2012) Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes. In *Quadruplex Nucleic Acids* pp. 111–177

91 Kwok, C.K. and Merrick, C.J. (2017) G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.* 35, 997–1013

92 Chilka, P. *et al.* (2019) Small molecule fluorescent probes for G-quadruplex visualization as potential cancer theranostic agents. *Molecules* 24, 752

93 Shivalingam, A. *et al.* (2015) The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* 6, 8178

94 Kotar, A. *et al.* (2016) NMR Structure of a Triangulenium-Based Long-Lived Fluorescence Probe Bound to a G-Quadruplex. *Angew. Chemie - Int. Ed.* 55, 12508–12511

95 Liu, L.Y. *et al.* (2020) Quantitative Detection of G-Quadruplex DNA in Live Cells Based on Photon Counts and Complex Structure Discrimination. *Angew. Chemie - Int. Ed.* 59, 9719–9726

96 Zhang, S. *et al.* (2018) Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule fluorescent probe. *Nucleic Acids Res.* 46, 7522–7532

97 Di Antonio, M. *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* 12, 832–837

98 Ofer, N. *et al.* (2009) The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res.* 37, 2712–2722

99 Morris, M.J. *et al.* (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.* 40, 4137–4145

100 Zamiri, B. *et al.* (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNAbinding proteins. *J. Biol. Chem.* 289, 4653–4659

101 Huang, H. *et al.* (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.* 31, 2296–2309

102 Ghosh, A. *et al.* (2019) Restoration of miRNA-149 Expression by TmPyP4 Induced Unfolding of Quadruplex within Its Precursor. *Biochemistry* 58, 514–525

103 Banco, M.T. and Ferré-D'Amaré, A.R. (2021) The emerging structural complexity of G-quadruplex RNAs. *Rna* 27, 390–402

104 Tao, Y. *et al.* (2021) Recent advances in the development of small molecules targeting RNA G-quadruplexes for drug discovery. *Bioorg. Chem.* 110, 104804

105 Song, J. *et al.* (2016) RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* 4, e1244031

106 Fay, M.M. *et al.* (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* 429, 2127–2147

107 Halder, K. *et al.* (2011) Efficient Suppression of Gene Expression by Targeting 5'-UTR-Based RNA Quadruplexes with Bisquinolinium Compounds. *ChemBioChem* 12, 1663–1668

108 Miglietta, G. *et al.* (2017) RNA G-Quadruplexes in Kirsten Ras (KRAS) Oncogene as Targets for Small Molecules Inhibiting Translation. *J. Med. Chem.* 60, 9448–9461

109 Katsuda, Y. *et al.* (2016) A Small Molecule That Represses Translation of G-Quadruplex-Containing mRNA. *J. Am. Chem. Soc.* 138, 9037–9040

110 Di Antonio, M. *et al.* (2012) Selective RNA versus DNA G-quadruplex targeting by situ click chemistry. *Angew. Chemie - Int. Ed.* 51, 11073–11078

111 Rocca, R. *et al.* (2017) Molecular recognition of a carboxy pyridostatin toward G-quadruplex structures: Why does it prefer RNA? *Chem. Biol. Drug Des.* 90, 919–925

112 Kwok, C.K. *et al.* (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. *Angew. Chemie - Int. Ed.* 55, 8958–8961

113 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

114 Santos, T. *et al.* (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* 189, 114208

115 Carvalho, J. *et al.* (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286

116 Chen, X.C. *et al.* (2018) Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells. *Angew. Chemie - Int. Ed.* 57, 4702–4706

117 Chen, S. Bin *et al.* (2016) Visualization of NRAS RNA G-Quadruplex Structures in Cells with an Engineered Fluorogenic Hybridization Probe. *J. Am. Chem. Soc.* 138, 10382–10385

118 Laguerre, A. *et al.* (2014) A Twice-As-smart synthetic G-quartet: PyroTASQ is both a smart quadruplex ligand and a smart fluorescent probe. *J. Am. Chem. Soc.* 136, 12406–12414

119 Laguerre, A. *et al.* (2015) Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.* 137, 8521–8525

120 Laguerre, A. *et al.* (2016) Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci. Rep.* 6, 32141

121 Murat, P. *et al.* (2011) Methods for investigating G-quadruplex DNA/ligand interactions. *Chem. Soc. Rev.* 40, 5293–5307

122 Jaumot, J. and Gargallo, R. (2012) Experimental Methods for Studying the Interactions between G-Quadruplex Structures and Ligands. *Curr. Pharm. Des.* 18, 1900–1916

123 Vorlíčková, M. *et al.* (2012) Circular dichroism spectroscopy of DNA: From duplexes to quadruplexes. *Chirality* 24, 691–698

124 Carvalho, J. *et al.* (2017) Circular dichroism of G-Quadruplex: A laboratory experiment for the study of topology and ligand binding. *J. Chem. Educ.* 94, 1547–1551

125 del Villar-Guerra, R. *et al.* (2018) G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chemie - Int. Ed.* 57, 7171–7175

126 Eriksson, M. and Nordén, B. (2001) Linear and circular dichroism of drug-nucleic acid complexes. *Methods Enzymol.* 340, 68–98

127 Garbett, N.C. *et al.* (2007) Circular dichroism to determine binding mode and affinity of ligand-dna interactions. *Nat. Protoc.* 2, 3166–3172

128 O'Hagan, M.P. *et al.* (2019) Binding and Beyond: What Else Can G-Quadruplex Ligands Do? *European J. Org. Chem.* 2019, 4995–5017

129 Rodriguez, R. *et al.* (2007) Ligand-driven G-quadruplex conformational switching by using an unusual mode of interaction. *Angew. Chemie - Int. Ed.* 46, 5405–5407

130 He, J.H. *et al.* (2013) New quinazoline derivatives for telomeric G-quadruplex DNA: Effects of an added phenyl group on quadruplex binding ability. *Eur. J. Med. Chem.* 63, 1–13

131 Cousins, A.R.O. *et al.* (2014) Ligand selectivity in stabilising tandem parallel folded G-quadruplex motifs in human telomeric DNA sequences. *Chem. Commun.* 50, 15202–15205

132 Xing, X. *et al.* (2011) Light-driven conformational regulation of human telomeric G-quadruplex DNA in physiological conditions. *Org. Biomol. Chem.* 9, 6639–6645

133 Wang, Z.F. *et al.* (2016) A novel transition pathway of ligand-induced topological conversion from hybrid forms to parallel forms of human telomeric G-quadruplexes. *Nucleic Acids Res.* 44, 3958–3968

134 Marchand, A. *et al.* (2015) Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* 137, 750–756

135 Šmidlehner, T. *et al.* (2017) Polarization spectroscopy methods in the determination of interactions of small molecules with nucleic acids - Tutorial. *Beilstein J. Org. Chem.* 14, 84–105

136 Nanjunda, R. *et al.* (2012) Heterocyclic Dications as a New Class of Telomeric G-Quadruplex Targeting Agents. *Curr. Pharm. Des.* 18, 1934–1947

137 Becher, J. *et al.* (2020) Synthesis and investigation of quadruplex-DNA-binding, 9-O-substituted berberine derivatives. *Beilstein J. Org. Chem.* 16, 2795–2806

138 Wickhorst, P.J. and Ihmels, H. (2021) Berberrubine phosphate: A selective fluorescent probe for quadruplex dna. *Molecules* 26, 2566

139 Zuffo, M. *et al.* (2017) G-quadruplex fluorescence sensing by core-extended naphthalene diimides. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1303–1311

140 Głuszyńska, A. *et al.* (2018) Binding study of the fluorescent carbazole derivative with human telomeric G-quadruplexes. *Molecules* 23, 3154

141 Yaku, H. *et al.* (2013) Study on effects of molecular crowding on G-quadruplex-ligand binding and ligand-mediated telomerase inhibition. *Methods* 64, 19–27

142 Krafcikova, M. *et al.* (2019) Monitoring DNA-Ligand Interactions in Living Human Cells Using NMR Spectroscopy. *J. Am. Chem. Soc.* 141, 13281–13285

143 Salgado, G.F. *et al.* (2015) G-quadruplex DNA and ligand interaction in living cells using NMR spectroscopy. *Chem. Sci.* 6, 3314–3320

144 Krafcikova, M. *et al.* (2019) In Cell NMR Spectroscopy: Investigation of G-Quadruplex Structures Inside Living Xenopus laevis Oocytes. In *Methods in Molecular Biology* 2035pp. 397–405

145 Carver, T.R. and Slichter, C.P. (1953) Polarization of nuclear spins in metals. *Phys. Rev.* 92, 212–213

146 Ni, F. (1994) Recent developments in transferred NOE methods. *Prog. Nucl. Magn. Reson. Spectrosc.* 26, 517–606

147 Mayer, M. and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* 38, 1784–1788

148 Dalvit, C. *et al.* (2001) WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomol. NMR* 21, 349–359

149 Liu, W. *et al.* (2019) Structures of 1:1 and 2:1 complexes of BMVC and MYC promoter G-quadruplex reveal a mechanism of ligand conformation adjustment for G4-recognition. *Nucleic Acids Res.* 47, 11931–11942

150 Kerkour, A. *et al.* (2017) NMR based model of human telomeric repeat G-quadruplex in complex with 2,4,6-triarylpyridine family ligand. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1293–1302

151 Kerkour, A. *et al.* (2017) High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *J. Biol. Chem.* 292, 8082–8091

152 Ferreira, R. *et al.* (2013) Structure and Stability of Human Telomeric G-Quadruplex with Preclinical 9-Amino Acridines. *PLoS One* 8, e57701

153 Dickerhoff, J. *et al.* (2021) Structural recognition of the MYC promoter G-quadruplex by a quinoline derivative: Insights into molecular targeting of parallel G-quadruplexes. *Nucleic Acids Res.* 49, 5905–5915

154 Tawani, A. *et al.* (2017) Structural insight for the recognition of G-quadruplex structure at human c-myc promoter sequence by flavonoid Quercetin. *Sci. Rep.* 7, 3600

155 Chung, W.J. *et al.* (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew. Chemie - Int. Ed.* 53, 999–1002

156 Liu, W. *et al.* (2018) Solution structures of multiple G-quadruplex complexes induced by a platinum(II)-based tripod reveal dynamic binding. *Nat. Commun.* 9, 3496

157 Gimenez, D. *et al.* (2021) 19F NMR as a tool in chemical biology. *Beilstein J. Org. Chem.* 17, 293–318

158 Ishizuka, T. *et al.* (2019) 19F NMR Spectroscopy for the Analysis of DNA G-Quadruplex Structures Using 19F-Labeled Nucleobase. In *Methods in Molecular Biology* 2035pp. 407–433

159 Bao, H.L. *et al.* (2017) A Simple and Sensitive 19F NMR Approach for Studying the Interaction of RNA G-Quadruplex with Ligand Molecule and Protein. *ChemistrySelect* 2, 4170–4175

160 Bao, H.L. *et al.* (2017) Characterization of human telomere RNA G-quadruplex structures *in vitro* and in living cells using 19F NMR spectroscopy. *Nucleic Acids Res.* 45, 5501–5511

161 Bao, H.L. *et al.* (2019) Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res.* 47, 4940–4947

162 Bao, H.L. and Xu, Y. (2020) Telomeric DNA-RNA-hybrid G-quadruplex exists in environmental conditions of HeLa cells. *Chem. Commun.* 56, 6547–6550

163 Parkinson, G.N. and Collie, G.W. (2019) X-Ray Crystallographic Studies of G-Quadruplex Structures. In *Methods in Molecular Biology* 2035pp. 131–155

164 Clark, G.R. *et al.* (2003) Structure of the first parallel DNA quadruplex-drug complex. *J. Am. Chem. Soc.* 125, 4066–4067

165 Haider, S.M. *et al.* (2003) Structure of a G-quadruplex-ligand complex. *J. Mol. Biol.* 326, 117–125

166 Lin, L.Y. *et al.* (2020) Biophysical and X-ray structural studies of the (GGGTT)3GGG G-quadruplex in complex with N-methyl mesoporphyrin IX. *PLoS One* 15, e0241513

167 McQuaid, K. *et al.* (2019) Structural Studies Reveal Enantiospecific Recognition of a DNA G-Quadruplex by a Ruthenium Polypyridyl Complex. *Angew. Chemie - Int. Ed.* 58, 9881–9885

168 Guarra, F. *et al.* (2018) Interaction of a gold(i) dicarbene anticancer drug with human telomeric DNA G-quadruplex: solution and computationally aided X-ray diffraction analysis. *Dalt. Trans.* 47, 16132–16138

169 Bazzicalupi, C. *et al.* (2013) The crystal structure of human telomeric DNA complexed with berberine: An interesting case of stacked ligand to G-tetrad ratio higher than 1:1. *Nucleic Acids Res.* 41, 632–638

170 Prado, E. *et al.* (2018) Influence of the SPR Experimental Conditions on the G-Quadruplex DNA Recognition by Porphyrin Derivatives. *Langmuir* 34, 13057–13064

171 Perenon, M. *et al.* (2020) Surface plasmon resonance study of the interaction of: N - methyl mesoporphyrin IX with G-quadruplex DNA. *Phys. Chem. Chem. Phys.* 22, 4158–4164

172 Miranda, A. *et al.* (2021) Locking up the as1411 aptamer with a flanking duplex: Towards an improved nucleolin-targeting. *Pharmaceuticals* 14, 1–20

173 Vo, T. *et al.* (2019) Biosensor-surface plasmon resonance: A strategy to help establish a new generation RNA-specific small molecules. *Methods* 167, 15–27

174 Tan, W. *et al.* (2016) Probing the G-quadruplex from hsa-miR-3620-5p and inhibition of its interaction with the target sequence. *Talanta* 154, 560–566

175 Pagano, B. *et al.* (2009) Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes. *Int. J. Mol. Sci.* 10, 2935–2957

176 Giancola, C. and Pagano, B. (2013) Energetics of ligand binding to G-quadruplexes. *Top. Curr. Chem.* 330, 211–242

177 Funke, A. and Weisz, K. (2019) Revealing the Energetics of Ligand-Quadruplex Interactions Using Isothermal Titration Calorimetry. In *Methods in Molecular Biology* 2035pp. 45–61

178 Funke, A. *et al.* (2016) Towards the Development of Structure-Selective G-Quadruplex-Binding Indolo[3,2-b]quinolines. *Chem. - A Eur. J.* 22, 3170–3181

179 Funke, A. and Weisz, K. (2017) Comprehensive Thermodynamic Profiling for the Binding of a G-Quadruplex Selective Indoloquinoline. *J. Phys. Chem. B* 121, 5735–5743

180 Funke, A. *et al.* (2018) Ligand-Induced Dimerization of a Truncated Parallel MYC G-Quadruplex. *ChemBioChem* 19, 505–512

181 Pérez-Arnaiz, C. *et al.* (2018) Kinetic evidence for interaction of TMPyP4 with two different G-quadruplex conformations of human telomeric DNA. *Biochim. Biophys. Acta - Gen. Subj.* 1862, 522–531

182 Dupont, J.I. *et al.* (2016) Calorimetric and spectroscopic investigations of the binding of metallated porphyrins to G-quadruplex DNA. *Biochim. Biophys. Acta - Gen. Subj.* 1860, 902–909

183 Bončina, M. *et al.* (2015) Thermodynamic fingerprints of ligand binding to human telomeric G-quadruplexes. *Nucleic Acids Res.* 43, 10376–10386

184 Bončina, M. *et al.* (2015) Dominant Driving Forces in Human Telomere Quadruplex Binding-Induced Structural Alterations. *Biophys. J.* 108, 2903–2911

185 Alniss, H. *et al.* (2018) Thermodynamic and spectroscopic investigations of TMPyP4 association with guanine- and cytosine-rich DNA and RNA repeats of C9orf72. *Biochem. Biophys. Res. Commun.* 495, 2410–2417

186 Yuan, G. *et al.* (2011) Mass spectrometry of G-quadruplex DNA: Formation, recognition, property, conversion, and conformation. *Mass Spectrom. Rev.* 30, 1121–1142

187 Li, H. (2019) Mass Spectroscopic Study of G-Quadruplex. In *Methods in Molecular Biology* 2035pp. 105–116

188 Lecours, M.J. *et al.* (2017) What stoichiometries determined by mass spectrometry reveal about the ligand binding mode to G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1353–1361

189 Marchand, A. *et al.* (2016) Selective and Cooperative Ligand Binding to Antiparallel Human Telomeric DNA G-Quadruplexes. *Chem. - A Eur. J.* 22, 9551–9555

190 Ceschi, S. *et al.* (2020) A two-quartet G-quadruplex topology of human KIT2 is conformationally selected by a perylene derivative. *Biochimie* 179, 77–84

191 Marchand, A. *et al.* (2018) Thermal Denaturation of DNA G-Quadruplexes and Their Complexes with Ligands: Thermodynamic Analysis of the Multiple States Revealed by Mass Spectrometry. *J. Am. Chem. Soc.* 140, 12553–12565

192 Paul, D. *et al.* (2019) Probing ligand and cation binding sites in G-quadruplex nucleic acids by mass spectrometry and electron photodetachment dissociation sequencing. *Analyst* 144, 3518–3524

193 Scalabrin, M. *et al.* (2017) Highly Improved Electrospray Ionization-Mass Spectrometry Detection of G-Quadruplex-Folded Oligonucleotides and Their Complexes with Small Molecules. *Anal. Chem.* 89, 8632–8637

194 Carvalho, J. and Cruz, C. (2020) Forster resonance energy transfer for studying nucleic acids denaturation: A chemical and biological sciences laboratory experiment. *Biochem. Mol. Biol. Educ.* 48, 329–336

195 Carvalho, J. *et al.* (2018) Phenanthroline polyazamacrocycles as G-quadruplex DNA binders. *Org. Biomol. Chem.* 16, 2776–2786

196 Lavrado, J. *et al.* (2013) Synthesis, G-quadruplex stabilisation, docking studies, and effect on cancer cells of indolo[3,2-b]quinolines with one, two, or three basic side chains. *ChemMedChem* 8, 1648–1661

197 Lavrado, J. *et al.* (2015) KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines. *Sci. Rep.* 5, 9696

198 Cadoni, E. *et al.* New (Iso) quinolinyl-pyridine-2 , 6-dicarboxamide G-Quadruplex Stabilizers . A Structure-Activity Relationship Study. , *Pharmaceuticals* , 14. (2021) , 1–22

199 Noureini, S.K. *et al.* (2017) Selectivity of major isoquinoline alkaloids from Chelidonium majus towards telomeric G-quadruplex: A study using a transition-FRET (t-FRET) assay. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 2020–2030

200 Rakers, V. *et al.* (2018) Development of microfluidic platforms for the synthesis of metal complexes and evaluation of their DNA affinity using online FRET melting assays. *Chem. Sci.* 9, 3459–3469

201 De Cian, A. *et al.* (2005) Affinity and selectivity of G4 ligands measured by FRET. *Nucleic Acids Symp. Ser. (Oxf).* 49, 235–236

202 Luo, Y. *et al.* (2021) FRET-MC: A fluorescence melting competition assay for studying G4 structures *in vitro*. *Biopolymers* 112, e23415

203 Monchaud, D. *et al.* (2006) Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorganic Med. Chem. Lett.* 16, 4842–4845

204 Monchaud, D. and Teulade-Fichou, M.P. (2010) G4-FID: a fluorescent DNA probe displacement assay for rapid evaluation of quadruplex ligands. *Methods Mol. Biol.* 608, 257–271

205 Monchaud, D. *et al.* (2008) Ligands playing musical chairs with G-quadruplex DNA: A rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* 90, 1207–1223

206 Largy, E. *et al.* (2011) Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal. Bioanal. Chem.* 400, 3419–3427

207 Tran, P.L.T. *et al.* (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie* 93, 1288–1296

208 Beauvineau, C. *et al.* (2017) PhenDV, a turn-off fluorescent quadruplex DNA probe for improving the sensitivity of drug screening assays. *Org. Biomol. Chem.* 15, 7117–7121

209 Desai, N. *et al.* (2021) Assessing G4-binding ligands *in vitro* and in cellulo using dimeric carbocyanine dye displacement assay. *Molecules* 26, 1400

210 del Villar-Guerra, R. *et al.* (2018) A rapid fluorescent indicator displacement assay and principal component/cluster data analysis for determination of ligand–nucleic acid structural selectivity. *Nucleic Acids Res.* 46, e41

211 Santos, T. *et al.* (2016) Purification of supercoiled G-quadruplex pDNA for *in vitro* transcription. *Sep. Purif. Technol.* 163, 59–71

212 Smith, J.S. and Johnson, F.B. (2010) Isolation of G-quadruplex DNA using NMM-sepharose affinity chromatography. *Methods Mol. Biol.* 608, 207–221

213 Ferreira, J. *et al.* (2017) Naphthalene amine support for G-quadruplex isolation. *Analyst* 142, 2982–2994

214 Chang, T. *et al.* (2012) Selective isolation of G-quadruplexes by affinity chromatography. *J. Chromatogr. A* 1246, 62–68

215 Musumeci, D. *et al.* (2014) G-quadruplex on oligo affinity support (G4-OAS): An easy affinity chromatography-based assay for the screening of G-quadruplex ligands. *Anal. Chem.* 86, 4126–4130

216 Musumeci, D. *et al.* (2017) Tandem application of ligand-based virtual screening and G4-OAS assay to identify novel G-quadruplex-targeting chemotypes. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1341–1352

217 Platella, C. *et al.* (2018) Controlled Pore Glass-based oligonucleotide affinity support: towards High Throughput Screening methods for the identification of conformation-selective G-quadruplex ligands. *Anal. Chim. Acta* 1030, 133–141

218 Pirota, V. *et al.* (2021) On the binding of naphthalene diimides to a human telomeric G-quadruplex multimer model. *Int. J. Biol. Macromol.* 166, 1320–1334

219 Ray, S. *et al.* (2020) Custom DNA Microarrays Reveal Diverse Binding Preferences of Proteins and Small Molecules to Thousands of G-Quadruplexes. *ACS Chem. Biol.* 15, 925–935

220 Wu, G. *et al.* (2020) Custom G4 microarrays reveal selective G-quadruplex recognition of small molecule BMVC: A large-scale assessment of ligand binding selectivity. *Molecules* 25, 3465

221 Felsenstein, K.M. *et al.* (2016) Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression. *ACS Chem. Biol.* 11, 138–148

222 Calabrese, D.R. *et al.* (2018) Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex. *Nat. Commun.* 9, 4229

223 Gracie, K. *et al.* (2014) Qualitative SERS analysis of G-quadruplex DNAs using selective stabilising ligands. *Analyst* 139, 4458–4465

224 Aznauryan, M. *et al.* (2021) Ligand Binding to Dynamically Populated G-Quadruplex DNA. *ChemBioChem* 22, 1811–1817

225 Rosu, F. *et al.* (2003) Selective interaction of ethidium derivatives with quadruplexes: An equilibrium dialysis and electrospray ionization mass spectrometry analysis. *Biochemistry* 42, 10361–10371

226 Saad, M. *et al.* (2019) Mapping and characterization of G-quadruplexes in the genome of the social amoeba Dictyostelium discoideum. *Nucleic Acids Res.* 47, 4363–4374

227 Jamroskovic, J. *et al.* (2019) Identification of putative G-quadruplex DNA structures in S. pombe genome by quantitative PCR stop assay. *DNA Repair (Amst).* 82, 102678

228 Wu, G. and Han, H. (2019) A DNA Polymerase Stop Assay for Characterization of G-Quadruplex Formation and Identification of G-Quadruplex-Interactive Compounds. In *Methods in Molecular Biology* 2035pp. 223–231

229 Gomez, D. *et al.* (2002) Detection of telomerase inhibitors based on G-quadruplex ligands by a modified telomeric repeat amplification protocol assay. *Cancer Res.* 62, 3365–3368

230 Panda, D. *et al.* (2019) A Competitive Pull-Down Assay Using G-quadruplex DNA Linked Magnetic Nanoparticles to Determine Specificity of G-quadruplex Ligands. *Anal. Chem.* 91, 7705–7711

231 Flusberg, D.A. *et al.* (2019) Identification of G-Quadruplex-Binding Inhibitors of Myc Expression through Affinity Selection–Mass Spectrometry. *SLAS Discov.* 24, 142–157

Chapter 4

Nucleolin: A binding partner of G-quadruplex structures

This chapter was published in:
***Santos, T.**; Salgado, G.F., Cabrita, E.J. and Cruz. C. (2022) Nucleolin: a binding partner of G-quadruplex structures. Trends Cell Biol. DOI: 10.1016/j.tcb.2022.03.003 – IF: 20.808*

Chapter overview

This chapter gathers the recent developments and trends of the involvement of nucleolin/G-quadruplex partnership in different pathologies such as cancer, viral infections, and neurodegenerative diseases. This literature overview was developed because even knowing the importance of the nucleolin/G4 partnership, the structural and molecular interaction studies involving this partnership have been substantially stalled due to the puzzling structure of the full-length nucleolin. Moreover, some attempts have been made to discover and characterize small molecules that could impact the partnership nucleolin/G4. The biological effects of some of those molecules are mentioned in this chapter. Overall, we sought to fill the gap in the available literature and encourage reflections and new insights into the field.

Nucleolin: A binding partner of G-quadruplex structures

Abstract

Nucleolin protein is involved in a plethora of cellular pathways across the nucleolus, nucleus, and cytoplasm. The association of its RNA-binding domain (RBD) and its RGG (arginine–glycine–glycine-rich) domain allows it to interact with G-quadruplex structures in nucleic acids. We highlight evidence that the nucleolin/G-quadruplex partnership is of extensive relevance to neurodegenerative diseases, cancer, and viral infections.

Keywords

Nucleolin; G-quadruplex; Ligands; Interaction

Nucleolin: relationship between structure and functional role

Nucleolin is a multifunctional protein whose cellular function is dependent on its intricate structure [1]. It contains an intrinsically disordered N-terminal region, followed by four RBDs and an RGG domain at the C terminus (**Figure 4.1**) [1]. The N-terminal region is a long (300 amino acids) domain which, in addition to its reported interactions with other proteins, could aid in the binding specificity of RBDs and RGG domains to nucleic acids [2]. Specifically, the association of the N-terminal domain with the RBDs and RGG domains of the protein can modulate its binding affinity and specificity for nucleic acids.

The RGG domain of nucleolin drives also its subcellular localization via liquid-liquid phase separation, protein-protein interactions, and plasma-membrane interactions [3]. Liquid-liquid phase separation is an intracellular mechanism that generates condensed liquid-like protein droplets and allows the assembly of membraneless compartments such as nucleoli. Recent findings demonstrate that the nucleolin RGG domain is crucial for axonal trafficking of *Kpnb1* (Importin β1) and *mTOR* mRNAs as well as for oncogene and viral promoter activity regulation [3].

The complex structure of nucleolin comprises multiple sites that can undergo post-translational modifications including phosphorylation, methylation, glycosylation, and ADP-ribosylation [4]. These influence nucleolin protein structure and affect its cellular localization and biological activity [4]. For instance, phosphorylation of the N-terminal domain of nucleolin by casein kinase (CK2) is involved in the regulation of rDNA transcription, whereas methylation of the arginine residues in the RGG domain impact on its interaction with nucleic acids [4]. In turn, *N*-glycosylation is essential for the expression of cell-surface nucleolin [4]. Although predominantly found in the dense fibrillar and granular regions of the nucleolus, following stress stimuli, nucleolin is known to move to other regions, including the cytoplasm and the cell surface [5]. In both tumor growth and viral infections, nucleolin shuttles from the nucleus to the cell surface and acts as the target for extracellular ligands. In neurodegenerative diseases, altered liquid-liquid phase separation and the formation of aberrant stress granules lead to the expulsion of nucleolin from the nucleoli and shuttling into the nucleoplasm and cytoplasm. In the nucleolus, nucleolin is involved in diverse aspects of ribosome biogenesis, including transcription by RNA polymerase I, processing of pre-rRNA, and ribosome assembly [5]. Nucleolin in nucleoplasm maintains genome stability and controls oncogene expression and has also been shown to be essential for DNA break repair and binds to oncogene promoters [5]. The functions of nucleolin in the cytoplasm have been mainly linked to protein expression regulation through mRNA and pre-miRNA interactions [5]. Cell-surface nucleolin plays a key role in disease since it acts as an essential receptor for numerous ligands and mediates their biological effects [5]. Among several well-known nucleolin partners, guanine-rich sequences have shown remarkable binding affinity and specificity [2].

Nucleolin/G-quadruplex partnership

Human cells contain many guanine-rich sequences that play an important regulatory role in diverse cellular processes such as splicing, polyadenylation, transcription termination, translation, and miRNA biogenesis [2,6]. These sequences have the potential to fold into G-quadruplexes (G4s) – four-stranded structures formed by self-recognition of guanines to generate stacked G-quartets [2,6]. G4s have been found in regions of biological relevance such as telomeres, oncogene promoters, 5′-untranslated regions, and miRNAs [6]. Given their biological significance, G4 structures have potential clinical relevance as both therapeutics and targets.

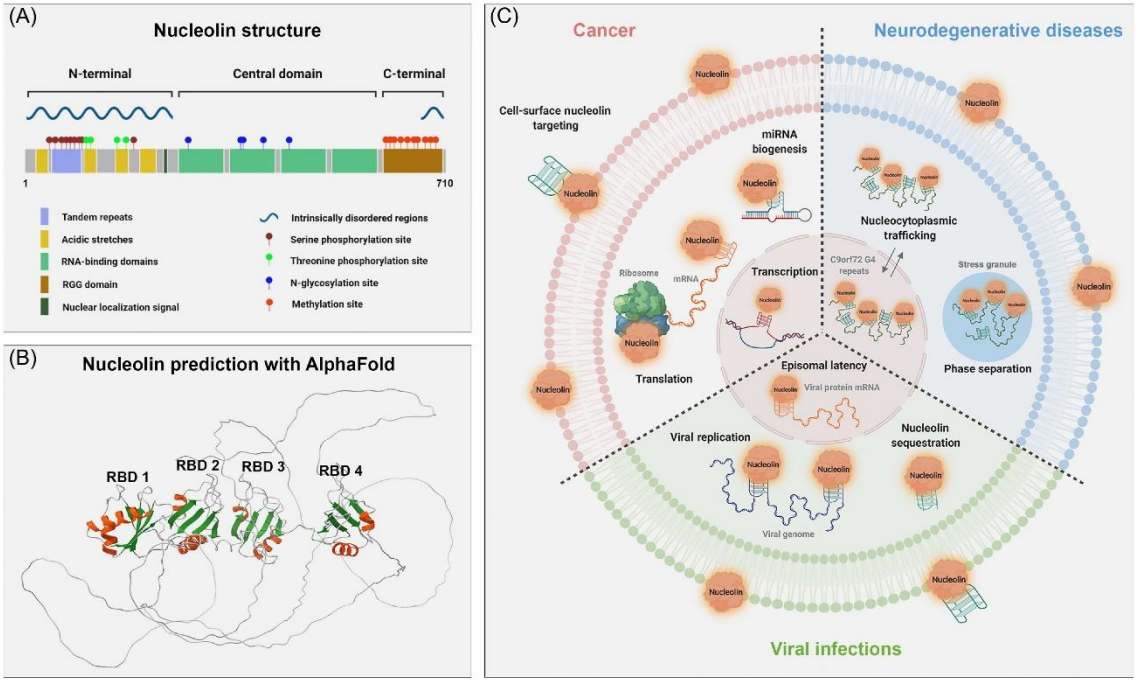

Figure 4.1. (A) Schematic illustration of nucleolin primary structure and sites of post-translational modifications. **(B)** Predicted 3D structure of nucleolin via the AlphaFold algorithm. α-Helixes are depicted in orange, β-sheets are highlighted in green, and random coils are shown in gray. **(C)** G-quadruplex (G4) structures are among the main nucleolin binding partners that assist it in several cellular functions. The activity of nucleolin can be regulated upon G4 recognition and vice versa. Recent advances propose important roles for the nucleolin/G4 partnership in neurodegenerative disease, cancer, and viral infection. In cancer, the binding of nucleolin to G4s can both trigger or repress transcription, translation, and/or miRNA biogenesis [15]. Cell-surface nucleolin has a key role in cancer because it can be the target of extracellular ligands. In neurodegeneration, G4s in the *C9orf72* gene bind to nucleolin, which triggers nucleocytoplasmic trafficking, altered liquid-liquid phase separation, and aberrant stress-granule formation [12]. In the case of viral infections, the nucleolin/G4 partnership plays a crucial role in virus attachment, episomal latency, viral replication, and nucleolin sequestration [10]. Abbreviation: RBD, RNA-binding domain.

G4s are transient and switch dynamically over time according to the local biological conditions, indicating that endogenous proteins can regulate G4 regions. Although several proteins have been described as G4 partners, nucleolin is one of the most relevant due to its prevalent role in G4 interactions [2]. In addition to the well-known binding of nucleolin to non-G4 sequences, nucleolin has been shown to exhibit a remarkable binding preference for G4 structures [2]. Recent evidence suggests that the RBDs of nucleolin are responsible for the interaction with G-rich sequences, whereas the RGG domain contributes to G4 folding [7]. This protein also displays an interesting trend of binding specifically and tightly to DNA G4s with a long central loop [8]. This general rule also seems to apply to RNA G4s with a long central loop (Please see Chapter 8) [9], but more evidence is needed. In addition, nucleolin can modulate the function and stability of endogenous and exogenous G4 structures, although the molecular structural

determinants of the nucleolin/G4 interaction have not been established. There are so far no high-resolution structural data for full-length nucleolin, although solution structures of the RBDs have already been solved by nuclear magnetic resonance (NMR) spectroscopy (PDB ID: RBD1-1FJ7, RBD2-1FJC, RBD3-2FC9, RBD4-2FC8).

Bearing in mind the relevance of this molecular interaction in a biological context, the nucleolin/G4 partnership has been increasingly implicated in G4-related pathological processes, notably in neurodegeneration, oncogenesis, and viral infection (**Figure 4.1C**).

Targeting of nucleolin/G-quadruplex partnership

The nucleolin/G4 partnership is a potential target for the diagnosis and/or therapy of several diseases. One strategy to modulate nucleolin/G4 regulation is to interfere with the binding interaction [10]. Some ligands may induce conformational changes in the G4 target, thus preventing or disrupting the interaction with nucleolin, whereas others do not modify the G4 structure. For instance, upon incubation of the c-*MYC* G4 (Pu27), a well-known binding partner of nucleolin, with TMPyP4 (a G4 ligand), the G4 structure switches from a parallel to a hybrid conformation [11]. Alternatively, the ligands may bind to an allosteric site of the G4 target that does not interfere with nucleolin recognition. There is also the particular case of the pyridostatin (PDS) derivative, carboxy-PDS, that binds to G4s without affecting their recognition by antibodies and proteins. Finally, the ligands may compete with nucleolin for the same G4 binding site, thus preventing the G4/nucleolin interaction (**Table 4.1**).

Nucleolin binds to *C9orf72* G_4C_2 repeats in a G4-dependent manner [12]. TMPyP4 destabilizes G4s in the *C9orf72* G_4C_2 repeat, mitigates neurotoxicity by disrupting the G4 structure, altering liquid-liquid phase separation, and promoting aberrant stress-granule formation [11]. The ligand CX-3543 was the first to reach Phase II clinical trials for the treatment of several tumor types. CX-3543 promotes the dissociation of nucleolin/rDNA G4 complexes, inhibits polymerase I transcription, and causes cancer cell apoptosis [1]. Moreover, the redistribution of non-bound nucleolin into the nucleoplasm facilitates nucleolin binding to and stabilization of *MYC* G4, thus repressing gene transcription [1]. The notion that ligands may also modulate nucleolin binding in infections with viruses such as Epstein–Barr virus (EBV) and hepatitis C virus (HCV) has emerged in the past few years. In the case of EBV, PhenDC3 disrupts the nucleolin/G4 EBNA1 mRNA interaction, enhances EBNA1 expression, and overcomes immune evasion by EBV, whereas PDS has the opposite effect [13]. PhenDC3 has a much higher affinity than PDS

for this specific G4 structure, and this could potentially explain their distinct behaviors. An alternative is that PDS binding may not interfere with nucleolin binding, whereas PhenDC3 might bind to the same site that is targeted by nucleolin. Finally, PhenDC3 could induce a strong conformational change in the G4 structure, whereas PDS does not induce structural changes. In addition, nucleolin expression is increased following HCV infection, and it binds to viral core G4s, resulting in suppressed HCV replication. PDS and its analog PDP compete for the same nucleolin binding sites and activate HCV replication, whereas TMPyP4 does not produce the same effect [14]. These results led to the conclusion that G4 stabilization and the ability to interfere with nucleolin/G4 partnership are two distinct molecular events.

The AS1411 aptamer was discovered by chance, and it binds to nucleolin with high affinity and specificity [1]. The aptamer is a G4-forming oligonucleotide that is internalized into cells and hijacks nucleolin away from other processes. Furthermore, the antitumoral and antiviral activities of AS1411 were ascribed to nucleolin-dependent biological processes [1]. These lines of evidence open up a framework for developing novel diagnostics and/or therapeutics based on AS1411 nanoplatforms.

Table 4.1. Ligands and aptamers that interfere with nucleolin/G4 binding.

Ligand	**Target G4s**	**Type of disease**	**Biological effects**	**Molecular mode of action**	**Refs**
PhenDC3, PhenDH2 and PyDH2	EBNA1[a] mRNA	EBV	Disruption of nucleolin/G4 EBNA1 mRNA interaction Induction of EBNA1 expression Induction of immune evasion of EBV	Competition or conformational change	[11,14]
PDS	EBNA1 mRNA and HCV RNA genome	EBV and HCV	Suppression of EBNA1 mRNA translation Downregulation of EBNA1 expression Induction of HCV replication	EBV: allosteric HCV: competition	[14,15]
PDP	HCV RNA genome	HCV	Induction of HCV replication	Competition	[15]
CX-3543	Ribosomal DNA	Cancer	Disruption of nucleolin/rDNA G4 interaction Inhibition of Polymerase I transcription Induction of Apoptosis and DNA damage Downregulation of *c-MYC* expression	Conformational change (disrupt G4 structures)	[4]

Table 4.1. Ligands and aptamers that interfere with nucleolin/G4 binding. (Continued)

TMPyP4 and GSA0932	Androgen receptor promotor	Prostate cancer	Downregulation of androgen receptor mRNA and protein expression	Allosteric	[1]
SYUIQ-05	Androgen receptor promotor	Prostate cancer	Induction of androgen receptor mRNA and protein expression	Conformational change	[1]
C_8	pre-miRNA 149 G4	Prostate cancer	Induction of cytotoxicity for cancer cells Remarkable cellular uptake	Allosteric	[12]
TMPyP4	*C9orf72* G_4C_2 repeat RNA	ALS/FTD	Disruption of G4s from *C9orf72* G_4C_2 repeat Mitigation of neurotoxicity Inhibition of aberrant stress granule formation	Conformational change (disrupt G4 structures)	[13]
AS1411	-	Neurodegenerative diseases, cancer, and virus	Inhibition of pro-survival NF-κB signalling pathway Induction of cell cycle arrest and apoptosis Inhibition of binding of nucleolin to *Bcl-2* mRNA Induction of nucleolin sequestration Perturbation of association of nucleolin with the Kif5a motor complex Downregulation of axonal *Inpp5f* mRNA	Competition (hijacks nucleolin away from other cellular processes)	[3,4,6]

Abbreviations: ALS, amyotrophic lateral sclerosis; EBNA1, EBV nuclear antigen 1; EBV, Epstein–Barr virus; FTD, frontotemporal dementia; HCV, hepatitis C virus; Inpp5f, inositol polyphosphate 5-phosphatase (also known as Sac2); Kif5a, molecular motor kinesin 1.

Concluding remarks

Accurate molecular models of the interaction between nucleolin and G4 structures could provide valuable information for the development of novel diagnostic and therapeutic strategies in neurodegenerative disease, cancer, and viral infection. To date, technical limitations have hindered the resolution of the 3D structure of full-length nucleolin and its complexes with G4s. One plausible way to overcome these limitations is the 'divide and conquer' strategy in which single domains of the protein are studied separately. A predicted 3D model of the protein via the AlphaFold algorithm was recently generated

and represents a significant breakthrough. Efforts have also been made to develop powerful bioinformatic tools to predict nucleic acid-protein interactions. However, much remains to be learned about nucleolin interactions, including how nucleolin forms complexes with other proteins and how it interacts with G4s. These recent findings represent a new paradigm for developing novel diagnostic and therapeutic approaches targeting the nucleolin/G4 partnership.

Acknowledgments

This work was supported by the Pessoa program (ref. 5079) and project Projeto de Investigação Exploratória (ref. IF/00959/2015, NCL Targeting by G-Quadruplex Aptamers for Cervical Cancer Therapy) financed by the Fundo Social Europeu (FSE) and the Programa Operacional Potencial Humano (POPH). Thanks are due to the Fundação para a Ciência e a Tecnologia (FCT)/Ministério para a Ciência e Tecnologia (MCT) for financial support to the CICS-UBI UIDB/00709/2020 research unit, POCI-01-0145-FEDER-022122 research unit Portuguese Platform of BioImaging (PPBI), and to the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016) through national funds, where applicable cofinanced by the Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE 2020, POCI, PORL, and PIDDAC. T.S. thanks the Fundação para a Ciência e Tecnologia (FCT) for a doctoral fellowship (PD/BD/142851/2018 integrated in the PhD Programme in NMR Applied to Chemistry, Materials, and Biosciences; PD/00065/2013) cofinanced by the Fundo Social Europeu and the Programa Operacional Regional Centro . C.C. acknowledges FCT grant UIDP/00709/2020.

References

1 Ugrinova, I. *et al.* (2018) Multifaceted nucleolin protein and its molecular partners in oncogenesis. In *Advances in Protein Chemistry and Structural Biology* (Donev, R., ed), pp. 133–164, Elsevier

2 Lago, S. *et al.* (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1371–1381

3 Doron-Mandel, E. *et al.* (2021) The glycine arginine-rich domain of the RNA-binding protein nucleolin regulates its subcellular localization. *EMBO J.* 40, e107158

4 Abdelmohsen, K. and Gorospe, M. (2012) RNA-binding protein nucleolin in disease. *RNA Biol.* 9, 799–808

5 Jia, W. *et al.* (2017) New perspectives of physiological and pathological functions of nucleolin (NCL). *Life Sci.* 186, 1–10

6 Varshney, D. *et al.* (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 21, 459–474

7 Masuzawa, T. and Oyoshi, T. (2020) Roles of the RGG domain and RNA recognition motif of nucleolin in G-quadruplex stabilization. *ACS Appl. Mater. Interfaces* 5, 5202–5208

8 Saha, A. *et al.* (2020) Nucleolin discriminates drastically between long-loop and short-loop quadruplexes. *Biochemistry* 59, 1261–1272

9 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

10 Ruggiero, E. and Richter, S.N. (2020) Viral G-quadruplexes: New frontiers in virus pathogenesis and antiviral therapy. In *Annual Reports in Medicinal Chemistry* 54pp. 101–131

11 Mori, K. *et al.* (2021) The porphyrin TMPyP4 inhibits elongation during the noncanonical translation of the FTLD/ALS-associated GGGGCC repeat in the C9orf72 gene. *J. Biol. Chem.* 297, 101120

12 Haeusler, A.R. *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195–200

13 Lista, M.J. *et al.* (2017) Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat. Commun.* 8, 16043

14 Bian, W.X. *et al.* (2019) Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res.* 47, 56–68

15 Miranti, C.K. *et al.* (2020) Nucleolin represses transcription of the androgen receptor gene through a G-quadruplex. *Oncotarget* 11, 1758–1776

SECTION II

Chapter 5

Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin

This chapter was published in:
***Santos, T.**; Miranda, A.; Imbert, L.; Monchaud, D.; Salgado G. F.; Cabrita E.J. and Cruz, C. (2022) Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin. J. Pharm. Biomed. Anal. 215, 114757. DOI: 10.1016/j.jpba.2022.114757 – IF: 3.935*

Chapter overview

Understanding the molecular mechanisms that control miRNA biogenesis may allow the development of tools to modulate the expression of miRNAs. Many miRNAs are crucial for the development of several diseases, and consequently for discovering novel diagnostic and therapeutic strategies for human disorders. Recently, it has been shown that precursor miRNAs that could adopt a G4 structure may control miRNA biogenesis by dynamically switching between the stem-loop and G4 structures. Therefore, molecules that interfere with this switching mechanism, like G4 ligands and proteins, may be used as molecular agents to regulate this pathway. Pre-miRNA let 7e has the potential to adopt a G4 structure and recently, its critical roles in oncology have been the focus of some important studies, as it is now known that it is frequently dysregulated in several cancer types. Herein, we have studied the stabilization/destabilization of the G4 structure of miRNA let 7e by eight different ligands and nucleolin. The ligands demonstrated distinct structural effects on the G4 structure. For instance, PDS stabilizes the G4 structure, while TMPyP4 disrupts/aggregates the G4 structure. Overall, the results presented in this chapter allowed the understanding of the complex biology of miRNA and could pave the way toward the control of miRNA biogenesis.

Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin

Abstract

Let-7e precursor microRNA has the potential to adopt a G-quadruplex (rG4) structure and recently, its roles in oncology have been the focus of much attention, as it is now known that let-7e pre-miRNA is frequently dysregulated in cancers. Therefore, it is crucial to unveil and fully characterize its ability to adopt a rG4 structure, which could be stabilized or destabilized by small molecules and proteins such as nucleolin, a protein that is deeply associated with miRNA biogenesis. Herein, by combining a set of different methods such as circular dichroism (CD), nuclear magnetic resonance (NMR), UV spectroscopy (thermal difference spectra (TDS) and isothermal difference spectra (IDS)) and polyacrylamide gel electrophoresis (PAGE), we demonstrate the formation of the rG4 structure found in let-7e pre-miRNA sequence in the presence of K^+ (5'-GGGCUGAGGUAGGAGG-3'). The ability of eight small molecules (or ligands) to bind to and stabilize this rG4 structure was also fully assessed. The dissociation constants for each RNA G-quadruplex/ligand complex, determined by surface plasmon resonance (SPR), ranged in the 10^{-6} to 10^{-9} M range. Lastly, the binding of the rG4 structure to nucleolin in the presence and absence of ligands was evaluated *via* CD, SPR, PAGE, and confocal microscopy. The small molecules 360 A and PDS demonstrated attractive properties to target the rG4 structure of let-7e pre-miRNA and control its biology. Our findings also highlighted that the interaction of TMPyP4 with the G-quadruplex of let-7e precursor miRNA could block the formation of the complex between the rG4 and nucleolin. Overall, this study introduces an approach to target the rG4 found in let-7e pre-miRNA which opens up a new opportunity to control the microRNA biogenesis.

Keywords

let-7e pre-miRNA; G-quadruplex; small molecules; nucleolin; interaction.

Introduction

RNA G-quadruplexes (rG4s) have been studied in the last few years due to intrinsic features that confer them, under certain circumstances, some advantages over their DNA counterparts (Please see Chapter 3) [1], [2]. Among these advantages, single-strand RNA lacks a complementary strand, which facilitates the intramolecular folding into secondary structures (including rG4s) known to dictate their cellular functions [3]. Recently the structural complexity of rG4s has been discussed [4] but the vast majority of rG4s adopt a parallel topology (in which all RNA strands are oriented in a parallel manner), which simplifies the development of small molecules to target these structures (or ligands) [5].

To date, much attention has been paid to the study of rG4s in precursor microRNAs (pre-miRNAs), suggesting a possible role of rG4s in the regulation of miRNA biogenesis [6]. Previous works have described the formation of G4s in pre-miRNA-149 [7], -92b [8], -26a [9], and -1229 [10], including our recent studies on rG4-forming sequences of pre-miRNA-149 (5'-GGGAGGGAGGGACGGG-3') (Please see Chapter 8) [11] and -92b (5'-GGGCGGGCGGGAGGG-3') (Please see Chapter 6) [12]. Furthermore, some studies explored and demonstrated the formation of rG4 structures in miRNA-149 [13], -197 [13], -432 [13], -765 [13], -1587 [14] and -3620 [15]. The formation of the rG4 in let-7e pre-miRNA (rG4-let-7e) has also been reported [16] by Pandey *et al.*, who showed that rG4 rG4-let-7e could significantly influence miRNA biogenesis [16] notably *via* a Dicer cleavage assay. In fact, a recent study demonstrated that both DNA and RNA G4s could inhibit the cleavage of pre-miRNAs by Dicer [17]. However, a deeper biophysical characterization of the pre-miRNA G4s is of utmost importance to pave the way for developing new therapeutic approaches.

Let-7 family is deregulated in various cancers [18] but its exact function is not yet fully understood. Several lines of evidence have shown the key roles of let-7 in oncology, as it is now known that let-7e miRNA is frequently down-regulated in cancers. In addition, let-7e is associated with shorter overall survival of cancer patients [19], [20], is implicated in the modulation of drug sensitivity in certain cancers [21], and is one of the miRNAs that influenced γH2AX *foci* formation, a marker of DNA double-strand breaks [22]. With this in mind, an appealing therapeutic strategy would be to control miRNA biogenesis *via* the use of ligands that could modulate the equilibrium between rG4 and stem–loop structures in pre-miRNAs [23], [24], [25]. However, this approach is, to date, underexplored since few rG4 structures have been characterized in terms of biological functions and structures.

Another possible approach is to consider the protein partners of miRNA, chief among them is nucleolin (NCL), known to be deeply involved in miRNA biogenesis [26]. NCL is predominantly located in the nucleolus but, in a cancerous context, can be found in the cytoplasm and cell surface [27], [28]. As it is established that NCL binds to and promotes the folding of G4 structures [29], its cytoplasmic location in cancer cells opens the possibility that NCL modulates the formation of G4 structures in pre-miRNAs.

Here, we sought out to investigate the rG4 biology and partners of a rG4-forming sequence found in let-7e pre-miRNA (5'-GGGCUGAGGUAGGAGG-3'). To this end, we first characterized the rG4 structure by circular dichroism (CD), UV absorption and nuclear magnetic resonance (NMR). Next, we studied the interacting properties of 8 well-known G4 ligands by CD-melting, Fluorescence Resonance Energy Transfer (FRET)-melting and Surface Plasmon Resonance (SPR) measurements. Finally, we investigated the binding of NCL in the presence and absence of ligand by CD, SPR, polyacrylamide gel electrophoresis (PAGE) and confocal microscopy.

Materials and Methods

Oligonucleotides and ligands

Oligonucleotides were obtained from Eurogentec (Belgium) with HPLC-grade purification. rG4-let-7e oligonucleotides used in this work are 5'-GGGCUGAGGUAGGAGG-3', 5' biotin-GGGCUGAGGUAGGAGG-3', and 5′-FAM GGGCUGAGGUAGGAGG TAMRA-3′. The duplex sequence d(TATAGCTAT-hexaethyleneglycol-TATAGCTATA) labeled with FAM and TAMRA was also used. Stock solutions of approximately 1 mM were prepared using DEPC water (Sigma-Aldrich, USA) and stored at − 80 °C until used. The concentration of oligonucleotide samples was determined from the absorbance at 260 nm by using the molar extinction coefficient. Annealing of oligonucleotide sequences was carried out by heating the samples for 10 min at 95 °C and cooling them on ice for 20 min before the experiments. Synthesis and purification of the ligand C_8 (10-(8-(4-iodobenzamide)octyl))− 3,6-bis(dimethylamine) acridinium iodide was performed as previously described [30]. The detailed synthesis and characterization of BioTASQ were performed as previously depicted [31]. The ligands PhenDC3 (3,3′-[1,10-phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate [32]; CAS: (929895−45−4), PDS (3-{1-[3-(dimethylamino)propyl]− 2-methyl-1 H-indol-3yl)− 1 H-pyrrole-2,5-dione [33]; CAS: (1085412−37−8), carboxyPDS 4-[[[2,6-Bis[[[4-(2-aminoethoxy)− 2-quinolinyl]amino]carbonyl]− 4-pyridinyl]oxy]methyl]− 1 H-1,2,3-

triazole-1-propanoic [34]; CAS: (1417638–60–8), BRACO-19 (N,N′-(9-(4-(dimethylamino)phenylamino)acridine-3,6-diyl)bis(3-(pyrrolidin-1-yl-propanamide) [35]; CAS: (1177798–88–7), TMPyP4 (tetra-(N-methyl-4-pyridyl)porphyrin [36]; CAS: 36951–72–1) and 360 A (2,6-N,N′-methyl-quinolinio-3-yl-pyridine dicarboxamide) [37]; CAS: (794458–56–3) were obtained from Sigma-Aldrich (USA). The chemical structures of each ligand are depicted in **Figure 5.1**. Stock solutions of the compounds were prepared as 10 mM solutions in dimethyl sulfoxide (DMSO, Thermo Fisher Scientific, USA) and their subsequent dilution was done using nuclease-free water.

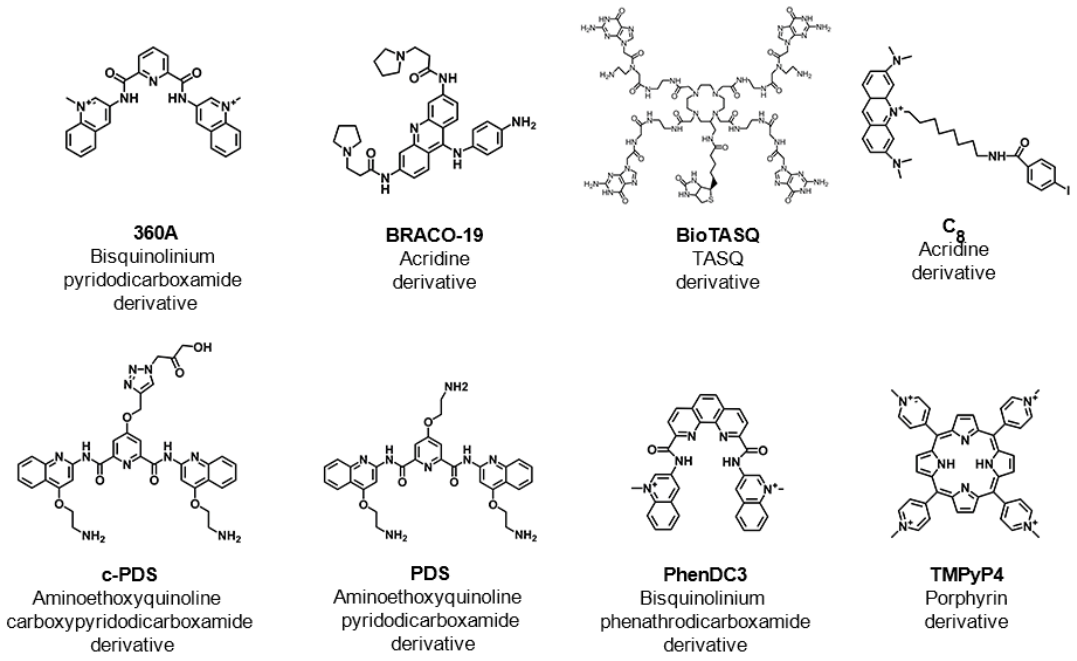

Figure 5.1. Chemical structure, common name and family of the ligands used in this work (chemical backbone).

Cloning, cell-free and purification of NCL RBD1,2

The sequence corresponding to NCL RNA binding domain (RBD) 1 and 2 was cloned into a pIVEX 2.4D vector. NCL RBD1,2 was synthesized *in vitro* using a cell-free expression system (ISBG, Grenoble). Briefly, NCL RBD1,2 was expressed under RNAse-free conditions in dialysis mode for 16 h at 23 °C under gentle agitation. The cell-free mixture contained 16 µg/mL of pIVEX 2.4D plasmid encoding the NCL RBD1,2 sequence, 1 mM of each essential amino acid, 0.8 mM of each rNTPs (guanosine-, uracil-, and cytidine-5'-triphosphate, 55 mM HEPES (pH 7.5)), 68 µM folinic acid, 0.64 mM cyclic adenosine monophosphate, 3.4 mM dithiothreitol, 27.5 mM ammonium acetate, 2 mM spermidine, 80 mM creatine phosphate, 208 mM potassium glutamate,

16 mM magnesium acetate, 250 μg/mL creatine kinase, 27 μg/mL T7 RNA polymerase, 0.175 μg/mL tRNA, and 400 μL/mL S30 E. coli bacterial extract. After incubation, the reaction mixture was diluted in binding buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, and 10 mM imidazole) to a final volume of 45 mL and centrifuged for 45 min at 36,000 g at 4 °C. Thereafter, the supernatant was applied onto a 5 mL Ni-NTA column that had been previously equilibrated in binding buffer (50 mM Hepes (pH 7.5), 300 mM NaCl and 10 mM Imidazole) at 4 °C. The column was washed with 5% of elution buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, and 500 mM imidazole) to eliminate residual contaminants and the protein was eluted with 50% of elution buffer. The fraction containing the NCL RBD1,2 was pooled and concentrated on a 10-kDa cut-off membrane. The purity of each fraction was analyzed by SDS-PAGE and the protein was identified through western blot analysis by using the primary anti-NCL antibody (Thermo Fisher, ref. PA3–16875).

UV Absorption spectroscopy

UV experiments were performed using a Thermo Scientific TM Evolution TM201 UV–Vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and recorded between 220 and 340 nm. Spectra acquisition was conducted with a 600 nm/min scanning rate, 1 nm data intervals, and 0.05 s of integration time. The rG4-let-7e sequence was diluted to 5 μM in lithium cacodylate buffer (10 mM, pH = 7.2). Thermal difference spectra (TDS) were carried out at 95 °C and 25 °C, corresponding to the unfolded and folded states, respectively. The TDS spectrum was calculated by subtracting the 25 °C spectra from the obtained at 95 °C. The difference spectrum was normalized relative to the maximum absorbance. The experiment was performed in 20 mM phosphate buffer pH 7.1 supplemented with 100 mM KCl. Isothermal difference spectra (IDS) were acquired at 25 °C and calculated by subtraction of the UV spectra of oligonucleotides, acquired in the absence or presence of increasing amounts of KCl.

Circular dichroism spectroscopy

CD spectra were acquired in a Jasco J-815 spectrometer (Jasco, USA), using a Peltier temperature controller (model CDF-426S/15). rG4-let-7e sequence was annealed as previously described. A 1 mm path-length quartz cuvette (Hellma, Germany) was used with rG4-let-7e at 10 μM in 10 mM lithium cacodylate buffer (Sigma-Aldrich, USA) at pH 7.2, supplemented with KCl (Thermo Fisher Scientific, USA). The required volume for the titrations was added directly to the quartz cell. The CD melting experiments were performed in the temperature range of 20–100 °C, with a heating rate of 2 °C/min by

monitoring the ellipticity at 262 nm. Spectra acquisition was performed in the absence and presence of increasing concentrations of KCl or ligands. Data was converted into fraction folded (f) plots.

$$f = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \quad \text{(Equation 5.1)}$$

Where CD is the ellipticity of the monitored wavelength at each temperature and CD^{min} and CD^{max} are the lowest and highest ellipticity values, respectively. Data points were then fitted to a Boltzmann distribution equation (OriginPro 2016) and the melting temperatures were determined.

NMR spectroscopy

Standard zgesgp pulse sequence was used to acquire individual ^{1}H NMR spectra with water suppression using excitation sculpting, on a 600 MHz Bruker Avance III spectrometer equiped with a QCI cryoprobe. rG4-let-7e sequence was used at a concentration of 100 μM with a total volume of 200 μL in a 3 mm NMR tube, annealed as described above and supplemented with 10% D_2O (Eurisotop, France). The required volume for titrations was added directly to NMR tubes. The spectra of rG4-let-7e sequence in 20 mM phosphate buffer pH 7.1 supplemented with 100 mM KCl were acquired at different temperatures (17, 27, 37 and 47 °C). All the spectra were acquired and processed with the software Topspin 3.1. Figures were prepared using TopSpin 4.0.6. Chemical shifts (δ) are reported in ppm.

Fluorescence Resonance Energy Transfer (FRET) Melting

FRET melting experiments were performed using a CFX Connect™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA), equipped with a FAM filter (λ_{ex} = 492 nm; λ_{em} = 516 nm). Oligonucleotides at 0.2 μM were annealed in lithium cacodylate (10 mM, pH 7.2) supplemented with 100 mM KCl before the experiment as described in the above sections. Each experimental condition was tested in duplicate in three separate plates. For each condition, 20 μL of oligonucleotides was aliquoted into each strip, followed by 5 μL of ligands solutions, at five different final concentrations (1, 2 and 5 eq.). Then, this was followed by an incubation time of 30 min at room temperature. The thermocycler was parametrized to measure and acquire the FAM emission after each step with a stepwise increase of 1 °C every 1 min, from 25 °C to 95 C. Through the fluorescence normalized curves, specifically to values when normalized emission is 0.5, the T_m values were ascertained for each type and concentration of ligand.

Surface Plasmon Resonance (SPR) Biosensor

SPR analysis was conducted on a Biacore T200 (Biacore, GE Healthcare, Uppsala, Sweden) with a SA sensor chip (streptavidin-coated sensor chip) (GE Healthcare, Sweden). The biotin-labeled rG4-let-7e (25 nM dissolved in 20 mM phosphate buffer supplemented with 100 mM KCl) was annealed as previously stated. The sensor chip was equilibrated with running buffer (20 mM phosphate buffer supplemented with 100 mM KCl) at 25 µL/min for 1 h. The flow cell was activated by injection of 1 M NaCl, 50 mM NaOH for 3 min. The injection was repeated seven times to remove unbound streptavidin from the sensor chip. Finally, to ensure surface stability, two primes with running buffer were performed and the buffer was flowed for 10 min at 1 µL/min to obtain a stable baseline. The biotin-labeled rG4-let-7e (25 nM) was immobilized at 1 µL/min until it reached approximately 250 RU.

For kinetic/affinity analysis, each ligand was serially diluted in running buffer, ranging the concentration from 1 nM to 1 µM. All experiments were performed in triplicate at 25 °C. Each ligand was injected from low to high concentrations during 75 s with a flow rate of 50 µL/min, followed by dissociation of 600 s. Surface regeneration was achieved by injecting two pulses of 30 s of 10 mM glycine/HCl pH 1.5, and the next three 60 s injections of running buffer to remove any trace of regeneration solution.

BiaEvaluation Software was used for data analysis and the likelihood of fittings was assessed through the statistical parameters of Chi2 and U-value. All sensorgrams were double corrected for non-specific binding and refractive index changes (bulk effect) by subtracting the signals of an equivalent injection across the reference flow cell 1. Dissociation constants were obtained from the 1:1 affinity model of sensorgrams.

Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel (15%) electrophoresis was used to visualize the oligonucleotides. rG4-let-7e samples were prepared at a concentration of approximately 2 µM. Sucrose (Sigma-Aldrich, USA) was added to the samples at a final concentration of 23%. The oligonucleotide marker was loaded in parallel on the gel. The rG4-let-7e sequence was injected with and without KCl. A molar ratio of 1:1 of rG4-let-7e/ligand, rG4-let-7e/NCL RBD1,2 and rG4-let-7e/TBA was prepared, and the mixture was incubated for 30 min. The supramolecular complexes of rG4-let-7e/ligand/NCL RBD1,2 were prepared at a molar ratio of 1:1:1. The samples digested with RNase H (0.3-U/µL) (NZYtech, Portugal) were incubated for 3 h at 37 °C. Electrophoresis was performed at

120 V with a temperature close to 20 °C. After electrophoresis, the gel was stained by SYBR Gold (Invitrogen, USA) for 10 min under gentle agitation and visualized using ChemiDoc™ MP Imaging System (Bio-Rad, USA).

Confocal microscopy

A549 cell line was grown in Ham's F12 medium supplemented with 10% (v/v) FBS and 1% (v/v) penicillin-streptomycin. Cultures were maintained in a humidified chamber at 37 °C and 5% CO_2. The cells were subsequently seeded in µ-Slide 8-well flat bottom imaging plates (Ibidi GmbH, Germany) at a plating density of 5×10^4 cells/well and incubated for cell adhesion in a humidified atmosphere at 37 °C and 5% CO_2. After 24 h, cells were incubated for 2 h at 37 °C with primary anti-NCL antibody (PA3–16875, Invitrogen, USA; dilution of 1: 100). Following primary anti-NCL antibody incubation, cells were washed 3 × with fresh serum-free medium and incubated with secondary antibody anti-rabbit IgG conjugated with Alexa Fluor® 647 (Thermo Scientific, USA; dilution of 1:500) for 1 h at 37 °C. Thereafter, cells were washed 3 × with fresh serum-free medium and incubated with rG4-let-7e labelled at the 5' terminus with Cyanine-3.5 (Cy3.5) (1 µM) for 1 h. Then, cells were washed 3 × with fresh serum-free medium and stained with Hoechst 33342® nuclear probe (Thermo Scientific, USA; 2 µM) for 15 min. Then, cells were imaged using a Zeiss LSM 710 confocal laser scanning microscope (Carl Zeiss, Germany) and processed using the blue edition of ZEN 2012 software (Carl Zeiss, Germany).

Results

Putative rG4 sequence in let-7e pre-miRNA adopt a G4 structure in presence of K^+

To support the formation of the rG4-let-7e, we carried out standard spectroscopic assays such as UV isothermal difference spectra (IDS), UV thermal difference spectra (TDS), CD and NMR of the short 16-nt-long G-rich sequence (5'-GGGCUGAGGUAGGAGG-3').

Using IDS, we demonstrated the cation-dependent nature of the rG4-let-7e. The IDS spectra showed negative peaks around 295 nm and positive peaks around 275 nm in the presence of KCl concentration above 10 mM, which indicated the formation of a rG4 structure (**Figure 5.2A**). Next, TDS signatures were recorded on rG4-let-7e in the presence of 100 mM KCl, by subtracting the absorbance spectrum at 25 °C (rG4 is fully folded) from the absorbance spectrum at 95 °C (rG4 is fully unfolded). In line with IDS,

a prominent negative peak at 295 nm and a positive peak at 275 nm were observed (**Figure 5.2B**), indicative of the rG4 formation [38]. The intense negative peak suggests that the rG4-let-7e sequence had a high propensity to form a rG4 structure.

CD measurements were also performed to confirm the rG4 signature and evaluate the effect of KCl. Results seen in **Figure 5.2C** showed an increase in the ellipticity upon addition of 100 mM KCl, with a CD signature typical of a parallel G4 (positive band ~ 260 nm and a negative band ~ 240 nm). In order to evaluate the influence of KCl in the thermal stabilization of the rG4 structure, CD-melting experiments were performed in the absence or presence of 100 mM KCl: without KCl, the mid-transition temperature was 39.9 ± 0.4 °C, while the presence of 100 mM KCl increased it to 51.1 ± 0.2 °C (**Figure 5.2D**). Altogether, these biophysical results confirmed that rG4-let-7e sequence folds into a stable rG4 structure in presence of 100 mM KCl.

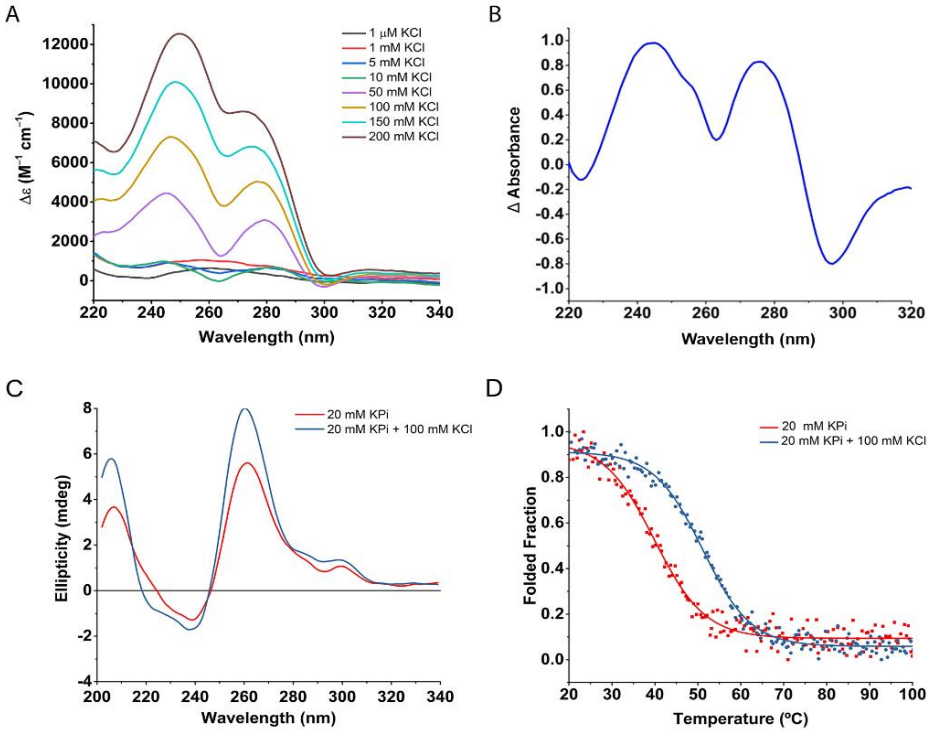

Figure 5.2. (A) IDS steady state of rG4-let-7e in the presence of increasing amounts of KCl. **(B)** TDS spectrum in the presence of 20 mM phosphate buffer, pH 7.1 supplemented with 100 mM KCl **(C)** CD spectra of the putative rG4, found in let-7e pre-miRNA, at 10 µM in 20 mM phosphate buffer, pH 7.1 in the absence and presence of 100 mM KCl. **(D)** CD-melting curves of the putative rG4-let-7e at 10 µM in 20 mM phosphate buffer, pH 7.1 in the absence and presence of 100 mM KCl.

To better characterize the structure of rG4-let-7e, ^{1}H NMR spectroscopy was employed. rG4 structures present a set of imino protons in the 10–12 ppm range, characteristic of Hoogsteen base pairing [39]. The experiments were performed in the absence and

presence of increasing concentrations of K^+ and the results seen in **Figure 5.3** revealed that at low K^+ concentrations (< 5 mM KCl), the sequence does not fold into a rG4 structure. However, the imino protons are easily observable at 10 mM of KCl but are not well-resolved, while a distinct set of 8 imino proton signals can be observed in concentrations of KCl > 50 mM, indicating a single G4 structure with two G-tetrads. The imino proton pattern does not change significantly between 50 and 100 mM KCl, further suggesting that the rG4 structure is fully folded in the presence of 50 mM KCl. Variable temperature ^{1}H NMR experiments (**Figure S5.1**) confirmed the stability of rG4-let-7e since the structure is maintained up to 47 °C These results are thus fully in line with those obtained by IDS and TDS experiments.

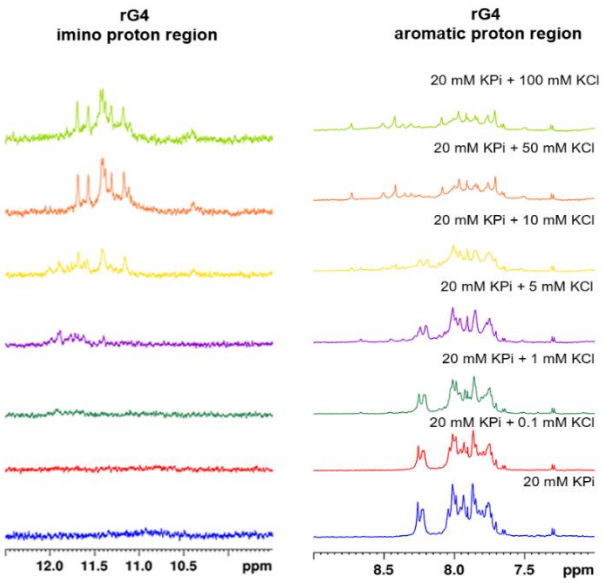

Figure 5.3. ^{1}H NMR spectra of the imino and aromatic proton region of the putative rG4-let-7e (100 μM) in Phosphate buffer containing 20 mM K_2HPO_4/KH_2PO_4, and different concentrations of KCl in the range 0-100 mM. All the ^{1}H NMR spectra were acquired at 27 °C.

Stabilization of rG4-let-7e by G4 ligands

Next, we investigated the interaction of well-known G4 ligands with the rG4-let-7e, *via* Fluorescence Resonance Energy Transfer (FRET)-melting experiments performed with TAMRA/FAM-labelled rG4-let-7e and a selection of ligands very heterogeneous in nature (**Figure 5.1**). The experiments were carried out at different ligand concentrations. The results seen in **Figure 5.4A** and **Table S5.1** revealed that melting temperature of the doubly labelled rG4 in the presence of 100 mM KCl was 57.3 °C. The highest stabilizations were obtained with 360 A, PDS and TMPyP4, with ΔT_m = 24, 15.8 and 22.7 °C at 5 mol. equiv., respectively. As further discussed, hereafter,

TMPyP4 has an ambivalent impact on G4 stabilization, depending on whether it is used at 1, 2 or 5 mol. equiv.; this is also the case for C_8, which displays a slight destabilizing effect at 2 mol. equiv. but a small stabilizing effect at 5 mol. eq. Both established rG4 ligands, BioTASQ and carboxyPDS (c-PDS) failed to stabilize rG4-let-7e in our conditions, which is not surprising for the former as the biotin appendage is known to 'poison' one of the G arms when free (that is, when not embedded in a biotin/streptavidin complex) [40].

CD spectroscopy was then implemented with each rG4/ligand pair (seen in **Figure S5.2**) and showed a distinct ellipticity behavior: at 2 mol. equiv. concentration, 360A, BioTASQ, C_8 and PhenDC3 did not affect the CD signature, while a higher concentration led to a slight increase in ellipticity; conversely, 2 mol. equiv. of BRACO-19, c-PDS, PDS and TMPyP4 triggereda decrease in ellipticity, while keeping the overall rG4 topology. These observations are in line with previous results notably for BRACO-19, whose quite surprising G4-destabilizing properties (at low concentration) were confirmed through a series of *in vitro* techniques, and for TMPyP4, which triggers both partial rG4 unfolding (at low concentration), stabilization and then, rG4 aggregation at higher concentrations [41], [42], [43].

CD-melting was also performed and collected results (**Figure 5.4B, Figure S5.3** and **Table S5.3**) confirmed a strong stabilizing effect for 360 A, PDS, PhenDC3 and TMPyP4, with ΔT_m > 20 °C. The other ligands (BRACO-19, BioTASQ, C_8 and c-PDS) showed weak stabilizing effects only. Overall, these results are in line with those obtained by FRET-melting experiments, with the notable exception of PhenDC3, found to be a modest rG4 binder by FRET-melting assay and an excellent one in CD-melting experiments.

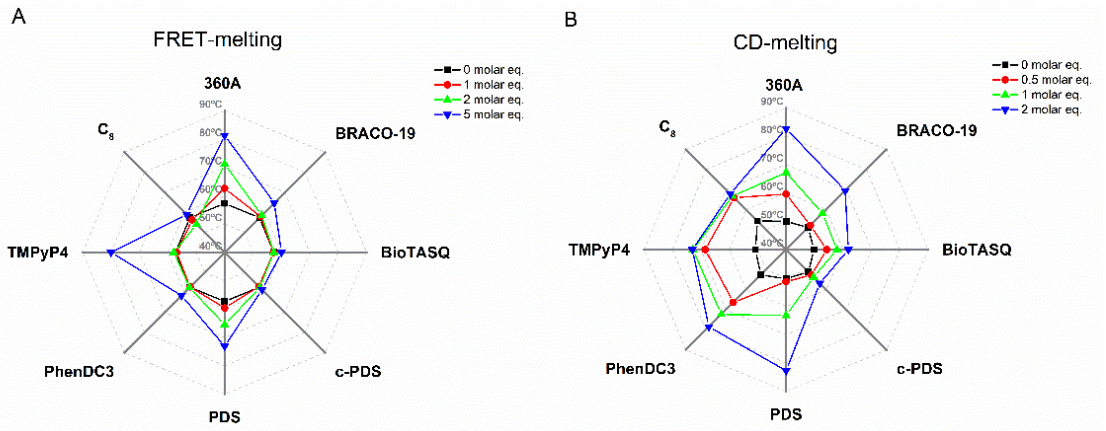

Figure 5.4. T_m radar plots of rG4-let-7e in the presence of different molar equivalents of G4 ligands, obtained by **(A)** FRET-melting and **(B)** CD-melting experiments.

Binding affinity of G4 ligands towards rG4-let-7e

We further quantified the rG4 binding affinities of these ligands by SPR. The dissociation constant (K_D) was determined using a biotin-labeled rG4-let-7e sequence immobilized on the surface of streptavidin sensor chips. The SPR signal responses related to the specific interaction with the rG4-let-7e were obtained after subtraction of the signals recorded on the reference flow-cell and the running buffer injection by applying a double referencing procedure. The K_D was obtained by fitting the steady-state response *vs*. the ligand concentration by Langmuir isotherm according to a 1:1 binding stoichiometry (**Figure S5.4**). The obtained K_D values, seen in **Table 5.1**, revealed that all ligands bind to rG4-let-7e with high affinity. These results do not fully agree with melting-based results, notably for C_8 and c-PDS, found to be modest thermal stabilizers but to display high binding affinity by SPR (K_D = 3.09 × 10^{-9} and 6.43 × 10^{-8} for C_8 and c-PDS, respectively). On the opposite, 360 A and PDS, which were found good thermal stabilizers provide rather low K_D values (6.56 × 10^{-6} and 3.81 × 10^{-6} M, respectively). These results highlight that a great caution must be exercised when dealing with *in vitro* investigations to determine the binding behavior of ligands as results can be found to be strongly dependent on the technique implemented, thereby emphasizing the need to use and compare results collected with assays relying on different biophysical features.

Table 5.1. K_D constant values of rG4-let-7e in presence of G4 ligands measured by SPR biosensor.

Ligand	**K_D ± SE (M)**
360A	6.56 × 10^{-6} ± 1.1 × 10^{-6}
BioTASQ	1.17 × 10^{-6} ± 2.6 × 10^{-7}
BRACO-19	2.68 × 10^{-7} ± 7.3 × 10^{-8}
C_8	3.09 × 10^{-9} ± 1.1 × 10^{-9}
c-PDS	6.43 × 10^{-8} ± 1.6 × 10^{-8}
PDS	3.81 × 10^{-6} ± 1.3 × 10^{-6}
PhenDC3	5.37 × 10^{-8} ± 1.8 × 10^{-8}
TMPyP4	2.49 × 10^{-7} ± 2.9 × 10^{-8}

Molecularity and structural nuances of rG4-let-7e in the presence of G4 ligands

The molecularity of the rG4-let-7e and rG4-let-7e/NCL RBD 1,2 complexes was next evaluated by polyacrylamide gel electrophoresis (PAGE). The oligonucleotides were diluted to 2 μM and complexes with ligands and/or NCL RBD 1,2 were prepared at a 1:1 molar ratio. The electrophoretic profiles seen in **Figure. S5.5** indicated that rG4-let-7e displays a major conformation in both K^+-free and K^+-rich conditions but in water the electrophoretic profile showed a smear along the run and two bands with less intensity at molecular weight around 60 nt. In the presence of BioTASQ, BRACO-19, C_8, c-PDS, PDS and PhenDC3, the electrophoretic bands retained the same intensity than that observed in K^+-rich conditions and, while TMPyP4 triggered smeared bands, in line with its aggregation properties (as previously described [42]). c-PDS led to the same major conformation with similar intensity, but a small smear was observed at high molecular weight (above 90 nt.), suggesting the formation of a molecular specie not structurally defined enough to yield to an isolated, well-defined band. Of note, 360 A association resulted in a band with less intensity and low molecular weight, indicating a strong association.

These results prompted us to further investigate the interactions between rG4-let-7e and 360A, PDS and TMPyP4 by ^{1}H NMR. Titrations performed with 360A (**Figure 5.5A**) resulted in a broadening of the imino protons, suggesting a good but not well-defined binding of the ligand to the G4 structure [44]. PDS elicited the same behavior (**Figure 5.5B**), while the imino proton signals of the rG4 totally disappeared in the presence of 0.5 molar eq. of TMPyP4 (**Figure 5.5C**), which again advocated for a ligand-induced aggregation.

In silico studies (molecular docking and molecular dynamics (MD) simulations) can provide in some cases interesting insights into the rG4 binding mode of ligands at the molecular level. Since the solution structure of rG4-let-7e is not available, we first used the 3D-NuS algorithm for generating the rG4-let-7e 3D structure [45]. The model was re-optimized by running fully solvated MD simulation during 20 ns. The representative model of the MD simulation is shown in **Figure S5.6**: in this model, the rG4 is made of two stacked G-tetrads (G1-G8-G12-G15 and G2-G9-G13-G16) interconnected by one long, 5-nt loop (GCUGA) and two short loops (UA and A). The residues belonging to the long loop appear to point inward, while that of the short loops outward. The long loop can facilitate the binding of the ligands (as recently demonstrated for porphyrinic ligands) [46], while end-stacking interactions can be observed since no end-capping residues are present. The binding mode of each ligand to the 3D structure of rG4-let-7e

was also investigated by molecular docking and molecular dynamics (**Figures S5.7-S5.8**); however, the results were poorly conclusive only.

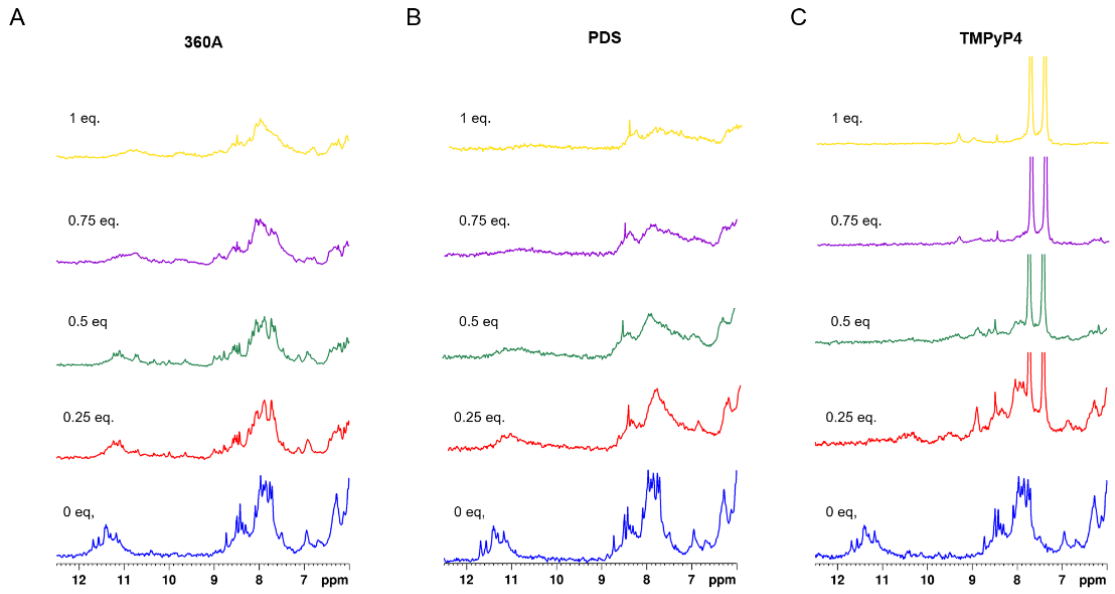

Figure 5.5. ^{1}H NMR spectra of the imino and aromatic proton region of the rG4-let-7e (100 µM) upon titration with **(A)** 360A, **(B)** PDS and **(C)** TMPyP4. Spectra were acquired in phosphate buffer containing 20 mM K_2HPO_4/KH_2PO_4 at 27 °C.

Binding of NCL RBD1,2 to rG4-let-7e structure

We thoroughly described (above) the interactions that take place between rG4-let-7e and a selected panel of ligands; it was thus of interest to investigate the interactions of both rG4 and rG4/ligand complexes with the NCL RBD 1,2 *via* CD, SPR and PAGE experiments.

CD measurements were carried out to assess the influence of the protein on the overall rG4-let-7e topology. Spectra seen in **Figure 5.6A** revealed a slight decrease in the ellipticity of the maximum positive band of the rG4 (264 nm), suggesting an interaction of the protein with the rG4 structure, and a strong increase of the negative bands, as a result of the contribution of the secondary structure of the NCL (the CD signature of NCL RBD1,2 displays a double negative band at around 210 and 220 nm) [47]. The thermal stability of the rG4-let-7e was evaluated by CD-melting experiments and showed a slight increase of the rG4 melting temperature in the presence of 10 µM NCL RBD1,2 (ΔT_m = 2.9 °C, **Figure 5.6B**), indicating a weak but detectable interaction between rG4-let-7e and NCL RBD1,2. This interaction was further characterized by SPR: the equilibrium binding affinity curve (**Figure 5.6C**) revealed that the K_D of the association

of NCL RBD1,2 with rG4-let-7e was in the micromolar range ($1.91 \times 10^{-6} \pm 3.6 \times 10^{-7}$ M), again indicating a weak but reliably detectable interaction.

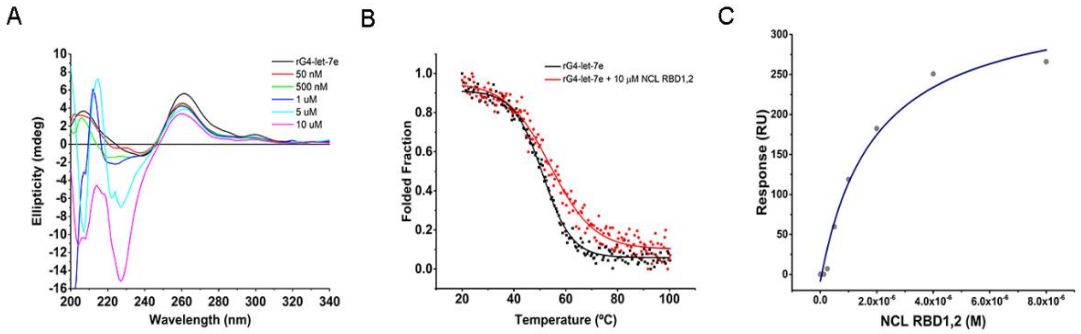

Figure 5.6. Molecular interaction of rG4-let-7e with NCL RBD1,2. **(A)** CD spectra of rG4-let-7e in the absence and presence of increasing amounts of NCL RBD1,2 (0-10 µM). **(B)** CD-melting curves of rG4-let-7e in the absence and presence of NCL RBD1,2. **(C)** Equilibrium binding curve of rG4-let-7e upon addition of increasing concentrations of NCL RBD1,2.

PAGE experiments were performed to evaluate the formation of the rG4-let-7/NCL RBD1,2 complex and the possible ternary rG4-let-7e/ligand/NCL RBD1,2 complexes. Results seen in **Figures 5.7A–B** showed a clear, concentration-dependent formation of the rG4/protein complex, illustrated by the increase of band intensity above 90 nt. Moreover, in the presence of 32 µM NCL RBD1,2, the band corresponding to rG4-let-7e almost disappeared, indicating a complete association of the rG4 structure with NCL RBD1,2. The presence of all ligands does not modify this electrophoretic profile, suggesting the possible formation of the ternary complex, with the notable exception of TMPyP4, which triggers the complete disappearance of the electrophoretic bands corresponding to the complex with NCL RBD1,2, which again advocated for a possible aggregation (**Figure 5.7C**). Since NCL was described to preferentially bind to parallel G4s, we also tested its ability to bind to a two G-tetrad antiparallel G4, namely that adopted by thrombin-binding aptamer (TBA). The results seen in **Figure S5.9** revealed that in the presence of 32 µM NCL RBD1,2, the band corresponding to free TBA remains visible proving the less efficient binding of NCL RBD1,2 to antiparallel G4s.

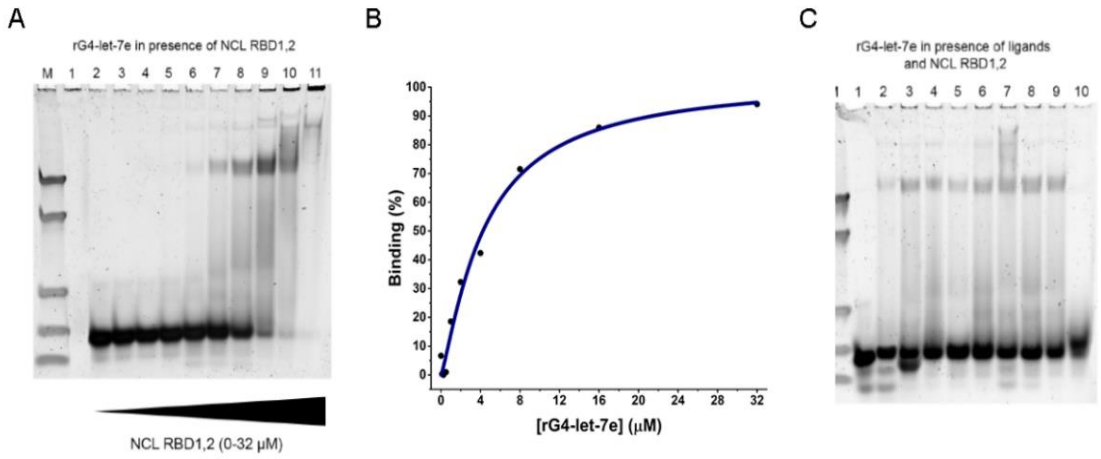

Figure 5.7. (A) Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) in presence of different concentrations of NCL RBD1,2 (Lanes: M – Marker (90, 60, 30, 21 and 15 nt.); 1 – NCL RBD1,2 (2 μM) ; 2 – rG4-let-7e without NCL RBD1,2; 3 – 0.1 μM; 4 - 0.250 μM; 5 – 0.5 μM; 6 – 1 μM; 7 - 2 μM; 8 – 4 μM: 9 – 8 μM; 10 – 16 μM and 11 – 32 μM). **(B)** Relative proportions of NCL RBD1,2-bound rG4-let-7e to unbound rG4-let-7e were quantified and plotted. **(C)** Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) / ligand (2 μM) / NCL RBD1,2 (2 μM). (Lanes: Lanes: M – Marker (90, 60, 30, 21 and 15 nt.); 1 – rG4-let-7e in 20 mM Phosphate buffer supplemented with 100 mM KCl; 2 – rG4-let-7e (2 μM) / NCL RBD1,2 (2 μM); 3 – 10 – ligands (2 μM) (3 - 360A; 4 – BioTASQ; 5 – BRACO-19; 6 – C_8 and 7 – c-PDS; 8 – PDS; 9 - PhenDC3; 10 - TMPyP4)).

Retention of rG4 structure in presence of ligands and ligands/NCL RBD1,2

To go a step further, we decided to digest rG4 samples with RNAse H, in absence and presence of K^+, ligands and/or both ligands and NCL RBD1,2. The PAGE experiments seen in **Figure 5.7** and **Figure S5.5** were used as controls for these experiments: as seen in **Figure 5.8A**, the RNase H digestion was partial only in our conditions, since it resulted in a decrease in the intensity of the bands when compared with the control. The presence of K^+ favoured a fully folded and stable rG4, while that of ligands resulted in different situations: results obtained with BioTASQ, BRACO-19, C_8, c-PDS and PDS are comparable to that obtained without RNase H incubation, suggesting that these ligands do not affect the rG4 structure; PhenDC3 and TMPyP4 displayed less intense bands, suggesting an aggregation or an increased sensitivity to RNase H, which could be rationalized for TMPyP4 (owing to its possible G4 destabilizing properties) but not for PhenDC3, thus supporting an aggregation. Only 360A provided a distinct signature, highlighting again its unique G4-interacting properties. We repeated these experiments in the presence of NCL RBD1,2 (**Figure 5.8B**): the rG4/NCL RBD1,2 was found to withstand digestion and, as above, the presence of BioTASQ, BRACO-19, C_8, c-PDS and PDS does not modify the electrophoretic profiles. However, noticeable differences were obtained with 360A, PhenDC3 and TMPyP4: the intensity of the corresponding bands

strongly decreased and only the band corresponding to low molecular weight complexes were visible with 360A and TMPyP4, suggesting that these three ligands, 360 A, PhenDC3 and TMPyP4, avoid the formation–or favour the disassembly–of the complex they formed with rG4 and NCL RBD1,2, and thus help RNase H processivity (to be compared with **Figure 5.7C**). These results thus offer a new perspective on the possible use of G4-targeting ligands to modulate miRNA biology.

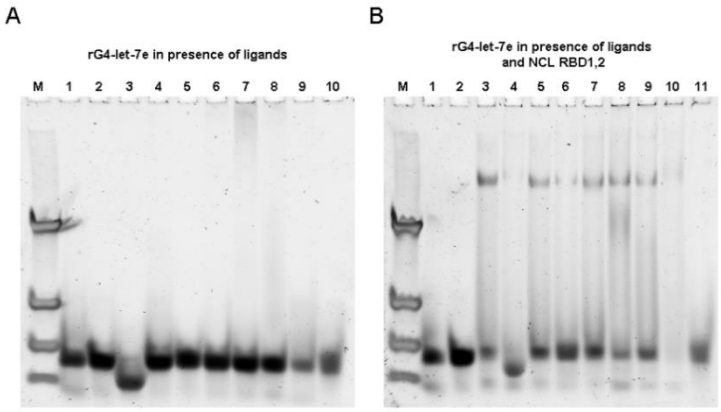

Figure 5.8. (A) Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) without or with ligand (2 μM) in presence of RNase H (0.3 U/μL). (Lanes: M – Marker (60, 30, 21 and 15 nt.); 1 – rG4-let-7e in water; 2 – rG4-let-7e in 20 mM Phosphate buffer supplemented with 100 mM KCl; 3 – 10 – ligands (2 μM) (3 - 360A; 4 – BioTASQ; 5 – BRACO-19; 6 – C_8 and 7 – c-PDS; 8 – PDS; 9 - PhenDC3; 10 - TMPyP4)). **(B)** Non-denaturing gel electrophoresis of rG4-let-7e (2 μM) / ligand (2 μM) / NCL RBD1,2 (2 μM) in presence of RNase H (0.3 U/μL). (Lanes: Lanes: M – Marker (60, 30, 21 and 15 nt.); 1 – rG4-let-7e in water; 2 – rG4-let-7e in 20 mM Phosphate buffer supplemented with 100 mM KCl; 3 – rG4-let-7e (2 μM) / NCL RBD1,2 (2 μM); 4 – 10 – ligands (2 μM) (4 - 360A; 5 – BioTASQ; 6 – BRACO-19; 7 – C_8 and 8 – c-PDS; 9 – PDS; 10 - PhenDC3; 11 - TMPyP4)).

Targeting of cell surface NCL by rG4-let-7e and their complexes with ligands

Finally, we investigated the binding of rG4-let-7e to NCL in a cellular context *via* confocal laser scanning microscopy (CLSM), using the non-small cell lung cancer (NSCLC) cell line A549, known to overexpress NCL at the cell surface. Cells were incubated with rG4-let-7e labelled with Cy3.5 to monitor its cellular distribution. CLSM images seen in **Figure 5.9** confirmed the presence of NCL at the cell surface of NSCLC cells (arrows); upon incubation with rG4-let-7e/PDS (**Figure 5.9B**) and rG4-let-7e/TMPyP4 complexes (**Fig. S5.10**), this colocalization seems to be redistributed by the ligands.

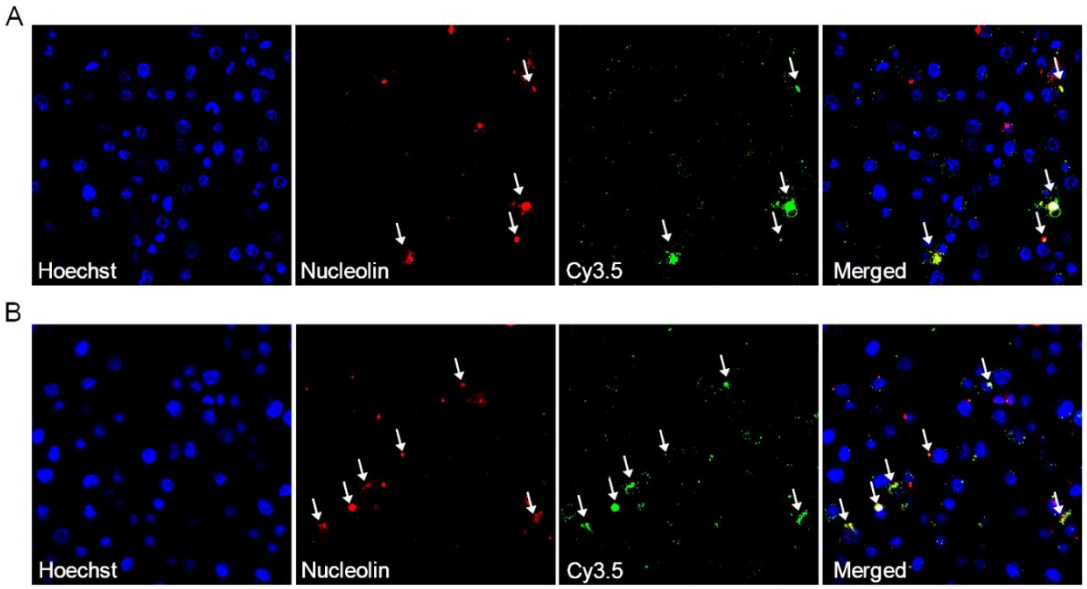

Figure 5.9. Confocal microscopy of A549 cells incubated with **(A)** rG4-let-7e and **(B)** rG4-let-7e/PDS complex. For each panel, images showed the cells stained with Hoechst 33342® nuclear probe (2 µM, blue); rG4-let-7e Cy3.5 (1 µM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:500). Arrows showed co-localization of rG4-let-7e or rG4-let-7e/PDS complex and NCL.

Discussion

Among all tumor suppressor microRNAs, depletion of miR-let-7e expression frequently occurs in cancers and is strongly correlated with poor overall survival rates in cancer patients [48]. The short 16-nt-long G-rich sequence (5'-GGGCUGAGGUAGGAGG-3') found in let-7e pri- and pre-miRNA is a possible rG4-forming sequence that partially overlaps with let-7e-5p sequence [16], [49]. This was demonstrated by Pandey *et al.* [16] who first showed the formation of a rG4 structure in let-7e pre-miRNA in physiological conditions [16]. They notably transfected MCF-7 cells using both the rG4-forming wild-type let-7e pre-miRNA sequence and the mutant pre-let7e, for which the rG4 formation is precluded: their RT-qPCR analyses revealed a remarkable difference in mature miRNA levels, with a 5-fold decrease in mature let-7e in cells transfected with wild-type let-7e pre-miRNA. These results suggested an inhibitory effect of the rG4 structure in dicer activity. They also evaluated the impact of the ligand TMPyP4 on rG4 and indicated that the ligand might disrupt the rG4 structure. This effort was recently continued by Pandolfini *et al.* who explored the connection between the rG4 formation in let-7e pre-miRNA and G methylation (7-methylguanosine, m7G) [49]. They demonstrated that the methylation of G11 negatively impacted rG4 formation, likely affecting the structural equilibrium towards the stem-loop conformation. These results

prompted us to further investigate the folding and stability of rG4-let-7e structure and its interaction with well-known G4 ligands.

To this end, we applied a set of biophysical methods routinely used to characterize rG4 folding. Collectively, the results of CD, IDS and TDS experiments confirmed the formation of a parallel rG4 structure, found to be strongly dependent on K^+ concentration. Furthermore, in line with recent evidence on the unexpected structural complexity of rG4 structures, [4] we found an intriguing CD signature for rG4-let-7e, with a slight positive band around 295 nm: on the basis of what was described for the G4 structure found in *c-KIT* gene [50], notably its ability to form a parallel G4 containing an external loop, [50], [51] we can postulate that rG4-let-7e might also have such a loop [52]. The structure of this rG4 was further characterized by ^{1}H NMR spectroscopy, which evidenced the formation of a two-quartet G4 core with 8 characteristic signals of Hoogsteen hydrogen bonds in the 10–12 ppm region [39] for KCl concentration higher than 5 mM.

Next, the ability of eight well-known G4-ligands to form a rG4-let-7e/ligand complex was investigated by CD- and FRET-melting experiments. Both pyridodicarboxamide (PDC)-based ligands, 360A and PDS, strongly stabilized rG4-let-7e, in line with results described by Kwok *et al.* who already described the stabilization of the rG4 found in pre-miRNA-149 by PDS [7]. In sharp contrast, the behavior of TMPyP4 was found to be dual, thermally stabilizing rG4-let-7e (by CD- and FRET-melting) while triggering disruption and/or aggregation of rG4-let-7e in isothermal experiments (by CD and PAGE). Again, these results fully complied with investigations recently performed by some of us, which described the dual behavior of TMPyP4 as a function of its concentration (disrupting G4 at low concentrations while triggering aggregation at high concentrations) [42]. The properties of PhenDC3 were more puzzling, as it poorly stabilized rG4 (ΔT_m = 4.2 °C at 5 molar equiv.) in FRET-melting experiments and strongly stabilized it (ΔT_m = 26 °C) in CD-melting experiments. This considerable difference is likely due to the use of fluorophores in FRET-melting experiments, which could lead to unspecific interactions with the aromatic moities of the ligand [1], [53]. The other ligands (BioTASQ, BRACO-19, C_8 and c-PDS) were found to be modest stabilizers in comparison. Again, the case of C_8 was unexpected since it was described as a strong stabilizer of the rG4s found in pre-miRNA-149 [11] and pre-miRNA-92b [12]. This difference can be ascribed to the structural features of these rG4s, pre-miRNA-149 and pre-miRNA-92b forming 3-tetrad rG4s with short loops (1–2 nt) while rG4-let-7e is a 2-tetrad rG4 with a long external loop (5 nt).

One of the conspicuous aspects of the present study is the variations of results between different *in vitro* techniques. Two illustrative examples are C_8 and c-PDS, with moderate stabilizing properties but excellent SPR results (with K_D values in the nanomolar range, 3.09×10^{-9} and 6.43×10^{-8}, respectively). On the other hand, 360A and PDS, which display high stabilizing properties, elicited K_D values in the micromolar range only (6.56×10^{-6} and 3.81×10^{-6} M, respectively). This brightly illustrates the need to implement different and complementary biophysical techniques when searching for new ligands, to discard false positives and negatives in the most reliable fashion possible. Along this line, PAGE is an interesting assay as it uniquely allows for a direct analysis of the molecularity of nucleic acid structures [54]. The results obtained with rG4-let-7e in the presence of ligands revealed a strong association of 360A and confirmed the ability of TMPyP4 to aggregate rG4 (these results were corroborated by ^{1}H NMR titrations), while the results collected with the other ligands were less conclusive.

In cells, the biological activity of nucleic acids is mediated by their protein partners. We thus focused on NCL, which is directly involved in several steps of miRNA biogenesis and found to be involved in their aberrant processing linked to many cancers [55]. It is known that NCL can play a chaperone role in assembling G4 structures and binds more tightly to parallel topology [29], [56]. We verified this by investigating the binding of NCL to the parallel rG4-let-7e and the antiparallel G4 TBA *via* a series of PAGE experiments, which confirmed the tighter binding of NCL to the former, with an affinity constant in the micromolar range ($1.91 \times 10^{-6} \pm 3.6 \times 10^{-7}$ M). These results were in line with CD, IDS and TDS results, which indicated a parallel G4 topology of rG4-let-7e in the presence of K^+. PAGE experiments also revealed the formation of the complex rG4-let-7e/NCL RBD1,2 and rG4-let-7e/ligand/NCL RBD1,2 for most, if not all ligands, with the notable exception of TMPyP4, which aggregates the rG4/NCL RBD1,2 complex.

The formation and stability of the rG4-let-7e structure in the cellular context are fundamental aspects of its functions in miRNA biogenesis. To assess this, we conducted an RNase H assay with the rG4 structure in the presence of K^+, ligands and NCL. This assay confirmed the proper folding, and thus, the stability of the rG4-let-7e structure in a biologically relevant context (where K^+ is present in a concentration around 140–150 mM) (Please see Chapter 9) [57], [58]; next, we showed that PhenDC3 and TMPyP4 made rG4 more sensitive to enzymatic digestion, on the basis of either G4 disruption or a possible aggregation/precipitation in the condition of the assay [42]. With the notable exception of 360A, the other ligands poorly affect RNase H activity, even in the presence of NCL RBD1,2.

Finally, we demonstrated by CLSM the binding of rG4-let-7e to NCL present at the surface of A549 cells using a fluorescently labelled rG4-let-7e. These images revealed co-localization spots of rG4-let-7e and NCL, which can be modulated by ligands, with PDS that seems to favor this association and TMPyP4 that seems to preclude it, on the basis of our previous investigations in which cell surface NCL was targeted by the rG4 found in pre-miRNA-149 in the presence of the ligand C_8 [11].

Altogether, these results demonstrated the binding of the rG4 structure to ligands, NCL and ligand/NCL complex both *in vitro* and in cells, which supports the hypothesis according to which the biogenesis of let-7e miRNA could be modulated by targeting its rG4 structure with small molecules.

Conclusions

In summary, we used here a series of biophysical methods to unravel the structure of rG4-let-7e. The biological relevance of let-7e is widely recognized and linked to several hallmarks of cancer. We demonstrated here that the 16-nt sequence found in let-7e pre-miRNA (5'-GGGCUGAGGUAGGAGG-3') does fold into a stable rG4 structure with two G-tetrads. The ability of 8 well-known ligands to interact with rG4-let-7e was assessed by CD- and FRET-melting experiments and highlighted the enticing interacting properties of 360A and PDS, while PhenDC3 and TMPyP4 provided more intricate results. We also investigated here the effect of NCL RBD1,2 in absence and presence of ligands and showed the possible existence of a ternary G4/ligand/protein complex. Altogether these results contribute to decipher the complex biology of miRNA and could pave the way toward the control of miRNA biogenesis by targeting rG4 structures in pre-miRNAs by ligands and protein surrogates.

Acknowledgments

This work was supported by PESSOA program ref. 5079 and project "Projeto de Investigação Exploratória" ref. IF/00959/2015 entitled "NCL targeting by G-quadruplex aptamers for cervical cancer therapy" financed by Fundo Social Europeu e Programa Operacional Potencial Humano. This work benefited from access to the Cell-Free platform of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL), an Instruct-ERIC centre, within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). Financial support was provided by Instruct-ERIC (PID:

10168 “Production of the full-length nucleolin for structural studies”). Thanks are due to FCT/MCT for the financial support to CICS-UBI UIDB/00709/2020 research unit, POCI-01-0145-FEDER-022122 research unit PPBI-Portuguese Platform of BioImaging, and to the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016), through national funds and, where applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and PIDDAC. Tiago Santos acknowledges Fundação para a Ciência e Tecnologia (FCT) for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013) co-financed by Fundo Social Europeu. André Miranda acknowledges the research fellowship “Rede Nacional de Ressonância Magnética Nuclear” ref. PINFRA/22161/2016-B4 and the doctoral fellowship grant from FCT – Foundation for Science and Technology (ref. 2021.04785.BD). C. Cruz acknowledges the grant from FCT ref. UIDP/00709/2020. The authors thank A. Paulo and M. P. Campello for the synthesis and supply of compound C_8.

References

1 Santos, T. *et al.* (2021) G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* 14, 769

2 Lyu, K. *et al.* (2021) RNA G-quadruplexes (rG4s): Genomics and biological functions. *Nucleic Acids Res.* 49, 5426–5450

3 Ganser, L.R. *et al.* (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* 20, 474–489

4 Banco, M.T. and Ferré-D’Amaré, A.R. (2021) The emerging structural complexity of G-quadruplex RNAs. *Rna* 27, 390–402

5 Fay, M.M. *et al.* (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* 429, 2127–2147

6 Tassinari, M. *et al.* (2021) Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* 49, 3617–3633

7 Kwok, C.K. *et al.* (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. *Angew. Chemie - Int. Ed.* 55, 8958–8961

8 Mirihana Arachchilage, G. *et al.* (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* 22, 262–272

9 Liu, G. *et al.* (2020) RNA G-quadruplex regulates microRNA-26a biogenesis and function. *J. Hepatol.* 73, 371–382

10 Imperatore, J.A. *et al.* (2020) Characterization of a G-Quadruplex Structure in Pre-miRNA-1229 and in Its Alzheimer’s Disease-Associated Variant rs2291418: Implications for miRNA-1229 Maturation. *Int. J. Mol. Sci. 2020, Vol. 21, Page 767* 21, 767

11 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

12 Santos, T. *et al.* (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* DOI: 10.1016/j.bcp.2020.114208

13 Chan, K.L. *et al.* (2018) Structural analysis reveals the formation and role of RNA G-quadruplex structures in human mature microRNAs. *Chem. Commun.* 54, 10878–10881

14 Tan, W. *et al.* (2018) Hsa-miR-1587 G-quadruplex formation and dimerization induced by NH4+, molecular crowding environment and jatrorrhizine derivatives. *Talanta* 179, 337–343

15 Tan, W. *et al.* (2016) Probing the G-quadruplex from hsa-miR-3620-5p and inhibition of its interaction with the target sequence. *Talanta* 154, 560–566

16 Pandey, S. *et al.* (2015) The RNA Stem-Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. *Biochemistry* DOI: 10.1021/acs.biochem.5b00574

17 Koralewska, N. *et al.* (2021) RNA and DNA G-quadruplexes bind to human dicer and inhibit its activity. *Cell. Mol. Life Sci.* 78, 3709–3724

18 Barh, D. *et al.* MicroRNA let-7: An emerging next-generation cancer therapeutic. , *Current Oncology*. (2010)

19 Zhu, W.Y. *et al.* (2014) Differential expression of miR-125a-5p and let-7e predicts the progression and prognosis of non-small cell lung cancer. *Cancer Invest.* 32, 394–401

20 Chirshev, E. *et al.* (2019) Let - 7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. *Clin. Transl. Med.* 8, 24

21 Boyerinas, B. *et al.* (2010) The role of let-7 in cell differentiation and cancer. *Endocr. Relat. Cancer* 17, 19–36

22 Xiao, M. *et al.* (2017) Let-7e sensitizes epithelial ovarian cancer to cisplatin through repressing DNA double strand break repair. *J. Ovarian Res.* 10, 24

23 Figueiredo, J. *et al.* (2021) Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs. *Molecules* 26, 6164

24 Henn, A. *et al.* (2008) Inhibition of dicing of guanosine-rich shRNAs by quadruplex-binding compounds. *Chembiochem* 9, 2722–2729

25 Bugaut, A. *et al.* (2012) An RNA hairpin to g-quadruplex conformational transition. *J. Am. Chem. Soc.* DOI: 10.1021/ja308665g

26 Pickering, B.F. *et al.* (2011) Nucleolin protein interacts with microprocessor complex to affect biogenesis of microRNAs 15a and 16. *J. Biol. Chem.* 286, 44095–44103

27 Jia, W. *et al.* (2017) New perspectives of physiological and pathological functions of nucleolin (NCL). *Life Sci.* 186, 1–10

28 Ugrinova, I. *et al.* (2018) Multifaceted nucleolin protein and its molecular partners in oncogenesis. In *Advances in Protein Chemistry and Structural Biology* (Donev, R., ed), pp. 133–164, Elsevier

29 Lago, S. *et al.* (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1371–1381

30 Pereira, E. *et al.* (2017) Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA. *Sci. Rep.* 7, 42544

31 Yang, S.Y. *et al.* (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* 9, 4730

32 De Cian, A. *et al.* (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.* 129, 1856–1857

33 Rodriguez, R. *et al.* (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* 130, 15758–15759

34 Di Antonio, M. *et al.* (2012) Selective RNA versus DNA G-quadruplex targeting by situ click chemistry. *Angew. Chemie - Int. Ed.* 51, 11073–11078

35 Read, M. *et al.* (2001) Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4844–4849

36 Han, F.X. *et al.* (1999) Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *J. Am. Chem. Soc.* 121, 3561–3570

37 Pennarun, G. *et al.* (2005) Apoptosis related to telomere instability and cell cycle alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene* 24, 2917–2928

38 Mergny, J.-L. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* 33, e138–e138

39 Webba da Silva, M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods* 43, 264–277

40 Renard, I. *et al.* (2019) Small-molecule affinity capture of DNA/RNA quadruplexes and their identification *in vitro* and *in vivo* through the G4RP protocol. *Nucleic Acids Res.* 47, 502–510

41 Haldar, S. *et al.* (2021) Ligand-induced unfolding mechanism of an RNA G-quadruplex. *bioRxiv* DOI: 10.1101/2021.10.26.465985

42 Mitteaux, J. *et al.* (2021) Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J. Am. Chem. Soc.* 143, 12567–12577

43 Lejault, P. *et al.* (2021) How to untie G-quadruplex knots and why? *Cell Chem. Biol.* 28, 436–455

44 Kerkour, A. *et al.* (2017) NMR based model of human telomeric repeat G-quadruplex in complex with 2,4,6-triarylpyridine family ligand. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1293–1302

45 Patro, L.P.P. *et al.* (2017) 3D-NuS: A Web Server for Automated Modeling and Visualization of Non-Canonical 3-Dimensional Nucleic Acid Structures. *J. Mol. Biol.* 429, 2438–2448

46 Stadlbauer, P. *et al.* (2021) Insights into G-Quadruplex-Hemin Dynamics Using Atomistic Simulations: Implications for Reactivity and Folding. *J. Chem. Theory Comput.* 17, 1883–1899

47 Miranda, A. *et al.* (2021) Aptamer-based approaches to detect nucleolin in prostate cancer. *Talanta* 226, 122037

48 Balzeau, J. *et al.* (2017) The LIN28/let-7 pathway in cancer. *Front. Genet.* 8, 31

49 Pandolfini, L. *et al.* (2019) METTL1 Promotes let-7 MicroRNA Processing via m7G Methylation. *Mol. Cell* 74, 1278-1290.e9

50 Manaye, S. *et al.* (2012) Porphyrin binding mechanism is altered by protonation at the loops in G-quadruplex DNA formed near the transcriptional activation site of the human c-kit gene. *Biochim. Biophys. Acta - Gen. Subj.* 1820, 1987–1996

51 Hsu, S.T.D. *et al.* (2009) A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *J. Am. Chem. Soc.* 131, 13399–13409

52 Małgowska, M. *et al.* (2013) How to study G-quadruplex structures. *Biotechnologia* 93, 381–390

53 Wang, K. *et al.* (2019) High-Throughput Screening of G-Quadruplex Ligands by FRET Assay. In *Methods in Molecular Biology* pp. 323–331

54 Saad, M. *et al.* (2019) Mapping and characterization of G-quadruplexes in the genome of the social amoeba Dictyostelium discoideum. *Nucleic Acids Res.* 47, 4363–4374

55 San, A. *et al.* (2021) In silico study predicts a key role of RNA-binding domains 3 and 4 in nucleolin-miRNA interactions. *bioRxiv* DOI: 10.1101/2021.06.09.447752

56 González, V. *et al.* (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J. Biol. Chem.* 284, 23622–23635

57 Zacchia, M. *et al.* (2016) Potassium: From Physiology to Clinical Implications. *Kidney Dis.* 2, 72–79

58 Santos, T. *et al.* (2022) Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin. *Eur. J. Pharm. Sci.* 169, 106093

Supplementary Information

Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin

***Santos, T.**; Miranda, A.; Imbert, L.; Monchaud, D.; Salgado G. F.; Cabrita E.J. and Cruz, C. (2022) Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin. J. Pharm. Biomed. Anal. 215, 114757. DOI: 10.1016/j.jpba.2022.114757 – IF: 3.935*

Materials and methods

rG4-let-7e 3D structural modeling

rG4-let-7e 3D structure was modeled using the 3D-NuS web server (https://iith.ac.in/3dnus/ accessed on 12 July 2021) [1]. The energy-minimized 3D model in PDB format was visualized using the UCSF Chimera 1.11.2 software. The built structure was optimized for further experiments by running fully solvated MD. The rG4 model was initially centered in a cubic periodic box, with a solute-box distance of 1 nm and filled with TIP3P water molecules and K^+ atoms. 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. A 100 ps equilibration was then 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 was done under the thermostat mentioned above and a Parrinello–Raman barometer. Finally, 10 ns of MD simulations were carried out and the most representative model of rG4-let-7e was used for further experiments. All molecular images were rendered using the UCSF ChimeraX 1.2.5 software.

Molecular docking studies

The minimized model of rG4-let-7e was used as a target for molecular docking studies. Both the rG4-let-7e structure and the ligand were optimized for docking using the Dock Prep tool of Chimera 1.11.2. After assigning polar hydrogens and Gasteiger charges, docking simulations were carried with AutoDock 4.2.6 software using Lamarckian genetic algorithm. The size of the box was constrained to 100 × 100 × 100 Å along the x, y, and z axes, respectively, with a grid spacing of 0.375 Å. For each ligand, 10 runs were performed with an initial population of 150 random individuals, a maximum number of evaluations set to 2.5×10^7, the rate of mutation and crossover set to 0.02 and 0.8, respectively, and an elitism value of 1. The ligands were allowed total flexibility, while the rG4-let-7e structure was kept rigid. The best conformers were selected based on the binding free energy and further processed with MD simulations.

Molecular dynamics simulations

The most energetically favorable binding poses were energy minimized using GROMACS 2020.6 package. Ligand topology was generated using ACPYPE server (https://www.bio2byte.be/acpype/) [2]. Energy minimization was carried out using

AMBER DNA OL15 force field [3] and TIP3P water model. RNA-ligand complexes were solvated in a 10 Å dodecahedron box. K^+ counterions were added to neutralize the system's total charge. Energy minimization (steepest descent algorithm) was run for a maximum of 50000 steps until the maximum force was < 10.0 kJ/mol. Each replicate was then simulated for 10 ns, using the particle mesh Ewald (PME) method to treat long-range electrostatic interactions with a Fourier grid spacing of 0.16 nm and a cutoff of 1.2 nm for direct contributions. Lennard–Jones interactions were calculated using a neighbor pair list with a cutoff of 1.2 nm and using a Verlet scheme. The temperature was maintained at 300 K using a modified Berendsen thermostat and the pressure was kept constant by Parrinello-Rahman barostat. Coordinates were collected in trajectory files every 10 ps. MD trajectories were analyzed with VMD 1.9.3 and representative images were rendered using UCSF ChimeraX 1.2.5 software.

Results

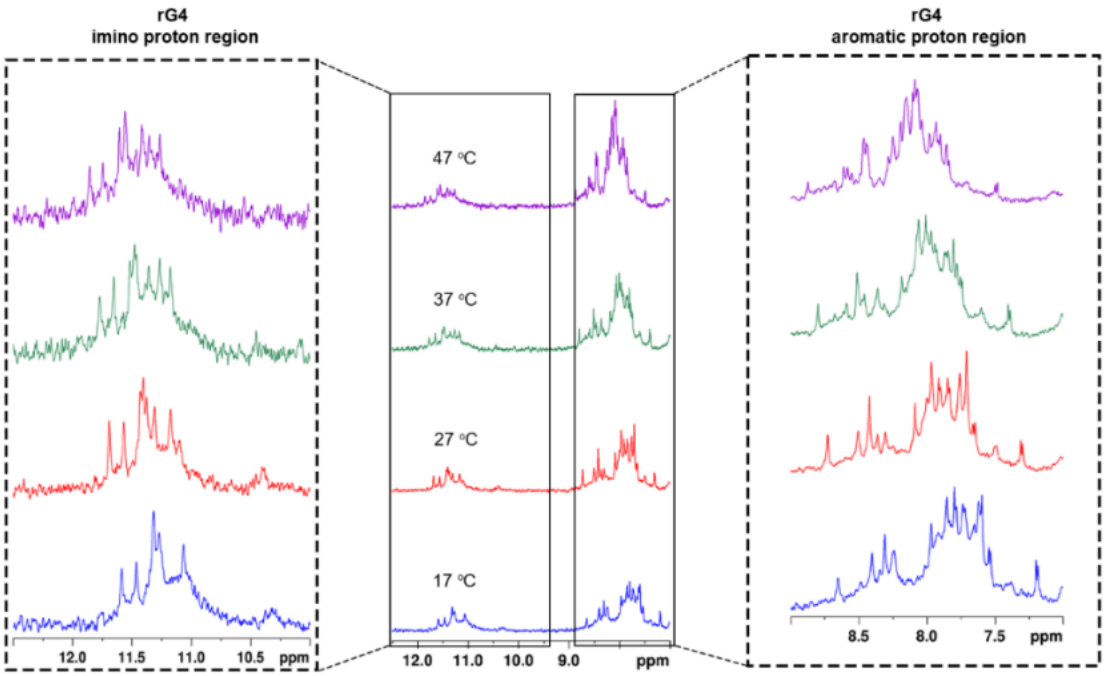

Figure S5.1. ^{1}H NMR spectra of the imino and aromatic proton region of rG4-let-7e (100 µM) in phosphate buffer containing 20 mM phosphate buffer pH 7.1, supplemented with 100 mM KCl. The spectra were acquired at different temperatures (17, 27, 37 and 47 °C).

Table S5.1. Melting temperature variations (ΔT_m) values for F-rG4-let-7e-T in the presence of several G4 ligands, obtained by FRET-melting assay.

	ΔT_m (°C)		
Ligand	**1 molar eq.**	**2 molar eq.**	**5 molar eq.**
360A	5.3 ± 1.0	13.9 ± 2.2	24 ± 0.4
BioTASQ	>0.5	>0.5	2.7 ± 0.3
BRACO-19	0.6 ± 0.8	1.2 ± 0.2	7.4 ± 1.2
C_8	>0.5	>0.5	1.6 ± 0.1
c-PDS	>0.5	>0.5	1.3 ± 0.7
PDS	2.3 ± 0.3	8.2 ± 0.7	15.8 ± 0.4
PhenDC3	>0.5	>0.5	4.2 ± 0.3
$TMPyP_4$	>0.5	0.6 ± 0.6	22.7 ± 1.3

Table S5.2. Melting temperature variations (ΔT_m) values for FdxT in the presence of several G4 ligands, obtained by FRET-melting assay.

	ΔT_m (°C)		
Ligand	**1 molar eq.**	**2 molar eq.**	**5 molar eq.**
360A	-	-	1.7 ± 0.7
BioTASQ	-	-	-
BRACO-19	-	-	-
C_8	-	-	-
c-PDS	-	-	-
PDS	-	-	-
PhenDC3	-	-	-
$TMPyP_4$	-	2.8 ± 0.6	12.1 ± 1.1

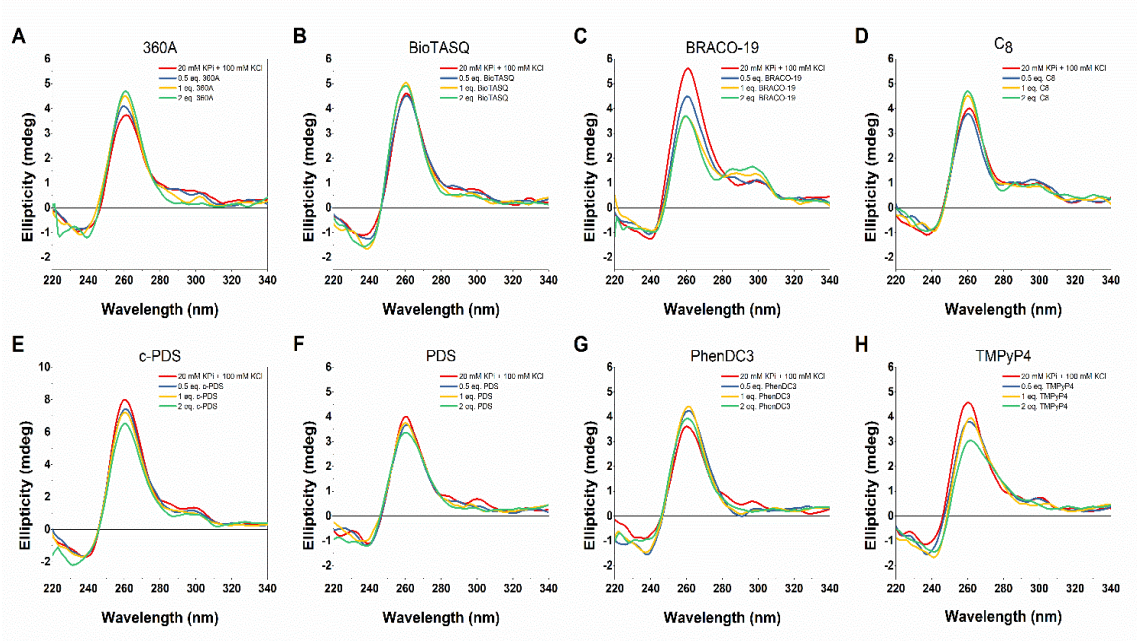

Figure S5.2. CD titration spectra of rG4-let-7e at 10 µM with increasing concentrations of ligands. Spectra acquisitions were performed at 20 °C in 20 mM phosphate buffer pH 7.1, supplemented with 100 mM KCl.

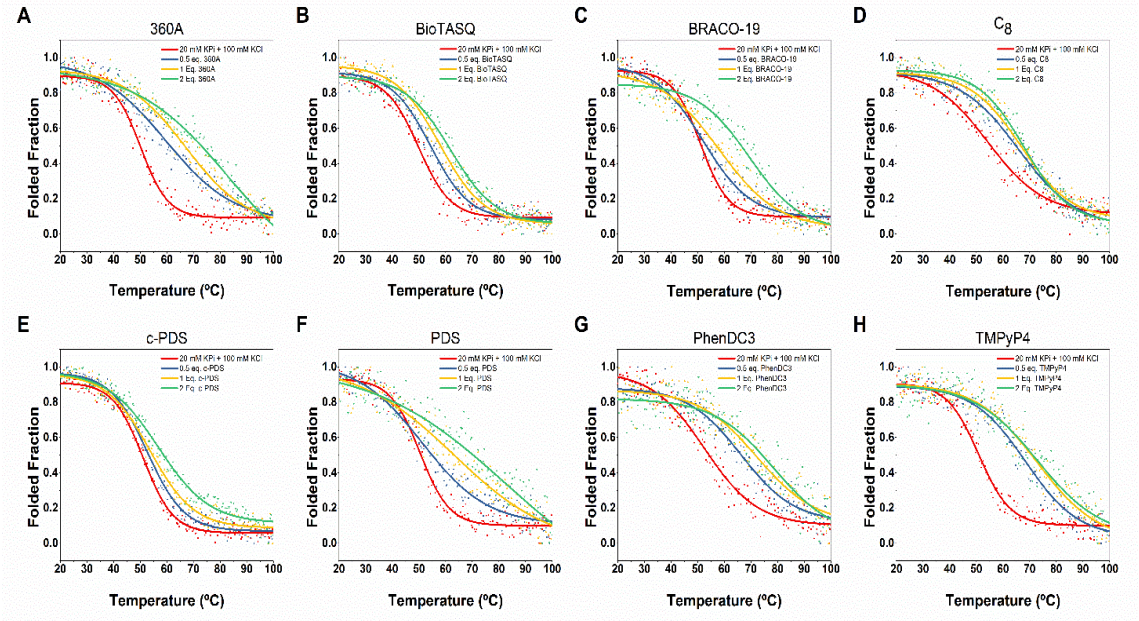

Figure S5.3. CD-melting curves of rG4-let-7e at 10 µM with increasing concentrations of ligands. Spectra acquisitions were performed in 20 mM phosphate buffer pH 7.1, supplemented with 100 mM KCl.

Table S5.3. Melting temperature variations (ΔT_m) values for the tested sequences in the presence of several G4 ligands, obtained by CD-melting experiments.

	ΔT_m (°C)		
Ligand	**0.5 molar eq.**	**1 molar eq.**	**2 molar eq.**
360A	9.6	17.09	>30
BioTASQ	4.6	8.1	12.1
BRACO-19	1.0	7.09	18.4
C_8	11.6	12.4	13.5
c-PDS	1.5	2.6	5.7
PDS	1.0	12.9	>30
PhenDC3	13.8	19.8	26.0
$TMPyP_4$	17.6	22.0	22.1

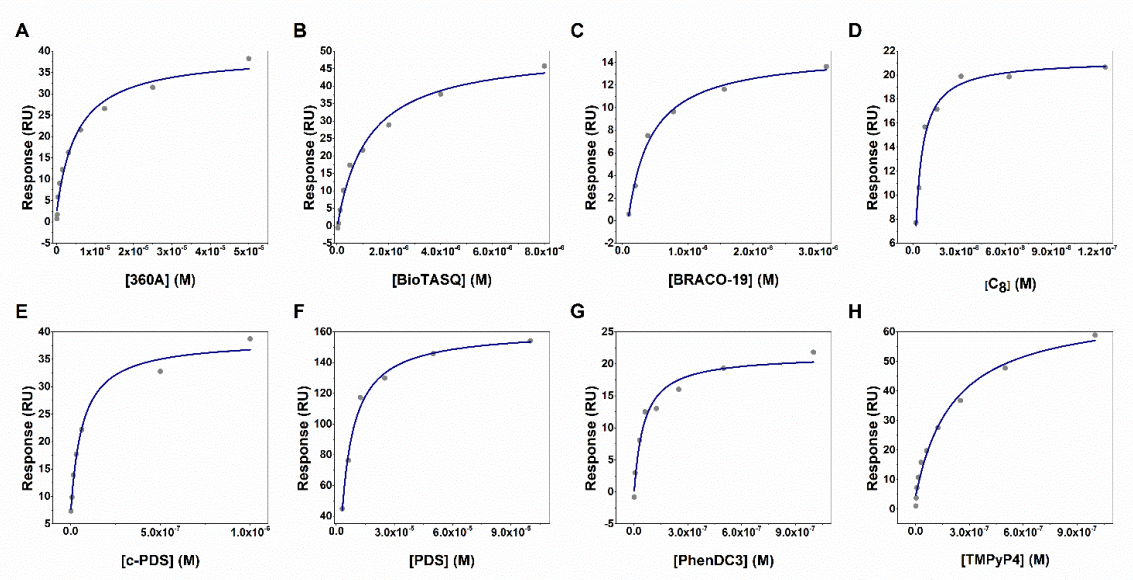

Figure S5.4. Equilibrium binding curves of rG4-let-7e upon addition of increasing concentrations of each ligand. K_D values were determined by fitting equilibrium binding data using a one-site saturation binding model.

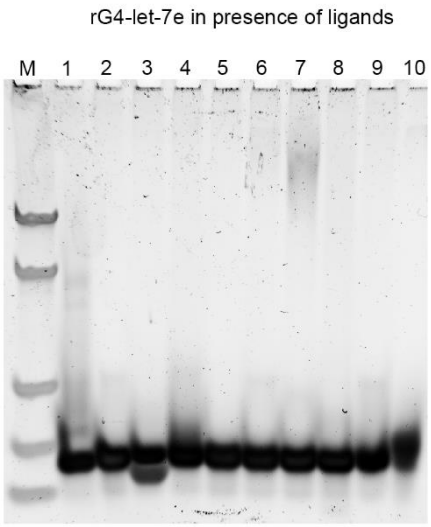

Figure S5.5. Polyacrylamide gel electrophoresis under native conditions of rG4-let-7e. The gel was supplemented with 10 mM KCl. Migration markers were oligothymidylate single-stranded DNA (n = 15, 21, 30, 60 and 90). Lane M - migration marker; lane 1 - rG4-let-7e in the presence of ultrapure water; lane 2 - rG4-let-7e in the presence of 20 mM K_2HPO_4/KH_2PO_4 and 100 mM KCl; lanes 3 to 10 - rG4/ligand complexes (with 1 molar eq. of BioTASQ, BRACO-19, C_8, c-PDS, PDS, PhenDC3, TMPyP4 and 360A, respectively).

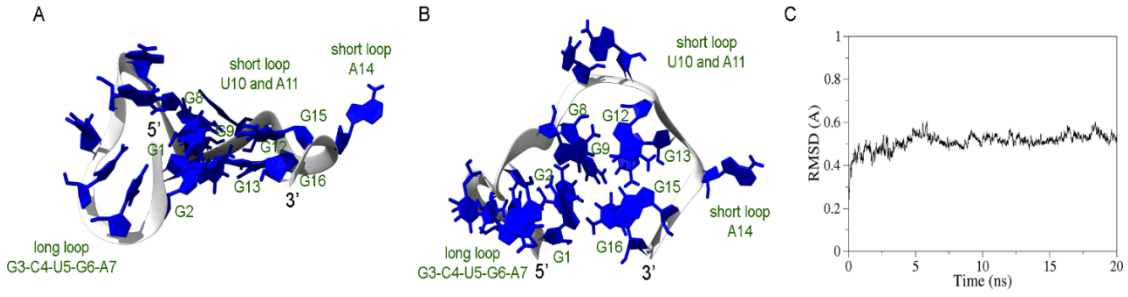

Figure S5.6. Predicted structural model of rG4-let-7e after 20 ns MD simulation. (A) Side view and (B) top view show a parallel stranded rG4. (C) RMSD plot of the 20 ns simulation rG4-let-7e model.

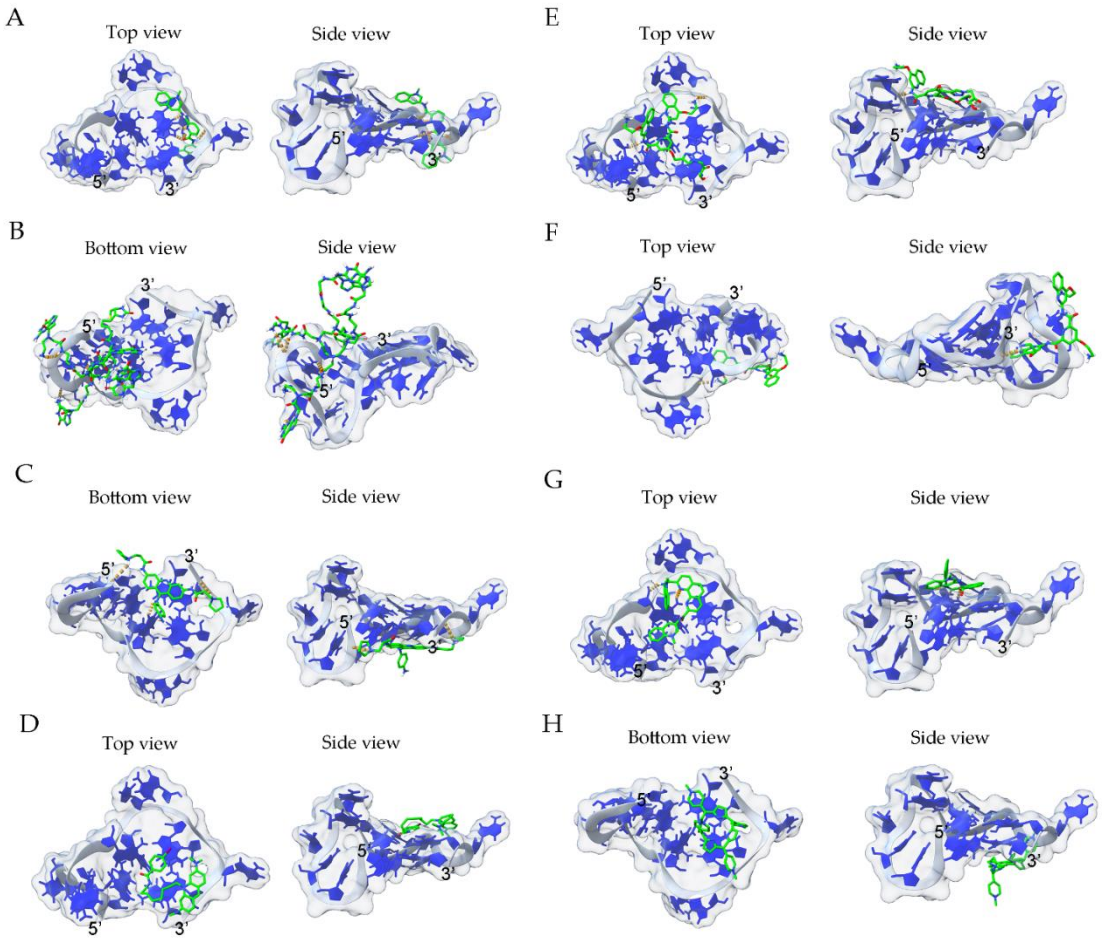

Figure S5.7. Selected molecular conformers in terms of lowest binding free energy of the complex formed by rG4-let-7e and ligands (A) 360A, (B) BioTASQ, (C) BRACO-19, (D) C_8, (E) c-PDS and (F) PDS, (G) PhenDC3 and (H) TMPyP4. G4 structure is shown in blue, while the ligand is depicted in green. Hydrogen bonds are depicted as orange dashed lines.

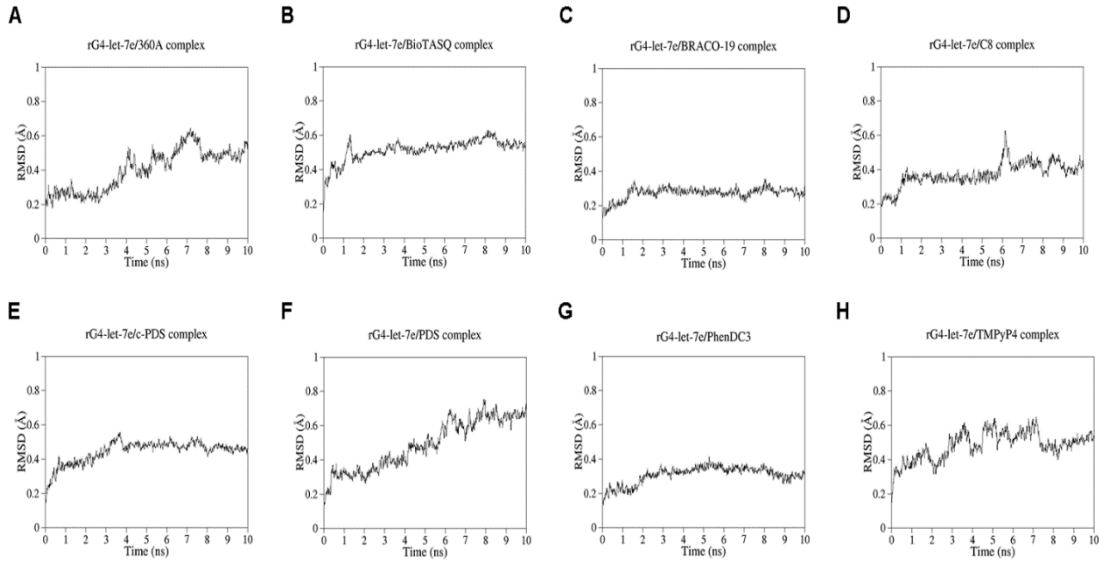

Figure S5.8. RMSD plots of the 10 ns MD simulation of the complex formed by rG4-let-7e and ligands (A) 360A, (B) BioTASQ (C) BRACO-19, (D) C_8, (E) c-PDS, (F) PDS, (G) PhenDC3 and (H) TMPyP4.

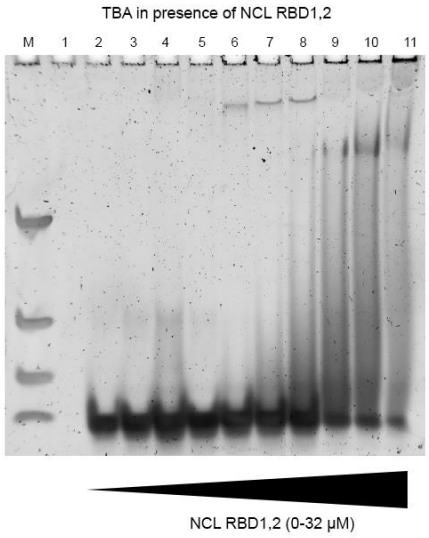

Figure S5.9. Non-denaturing gel electrophoresis of TBA (2 μM) in presence of different concentrations of NCL RBD1,2 (Lanes: M – Marker; 1 – NCL RBD1,2 (2 μM) ; 2 – rG4-let-7e without NCL RBD1,2; 3 – 0.1 μM; 4 - 0.250 μM; 5 – 0.5 μM; 6 – 1 μM; 7 - 2 μM; 8 – 4 μM: 9 – 8 μM; 10 – 16 μM and 11 – 32 μM).

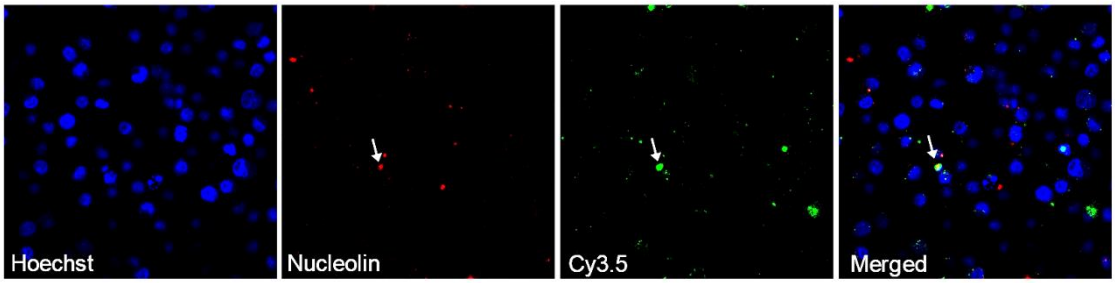

Figure S5.10. Confocal microscopy of A549 cells incubated with rG4-let-7e/TMPyP4 complex. For each panel, images showed the cells stained with Hoechst 33342® nuclear probe (2 μM, blue); rG4-let-7e Cy3.5/TMPyP4 complex (1 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:500). Arrows showed co-localization of rG4-let-7e Cy3.5/TMPyP4 and NCL.

References

1 Patro, L.P.P. *et al.* (2017) 3D-NuS: A Web Server for Automated Modeling and Visualization of Non-Canonical 3-Dimensional Nucleic Acid Structures. *J. Mol. Biol.* 429, 2438–2448

2 Sousa Da Silva, A.W. and Vranken, W.F. (2012) ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* 5, 367

3 Zgarbová, M. *et al.* (2015) Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* 11, 5723–5736

Chapter 6

Recognition of nucleolin through interaction with RNA G-quadruplex

This chapter was published in:
***Santos, T.**; Miranda, A.; Campello, M.P.C.; Paulo, A.; Salgado, G.; Cabrita, E.J. and Cruz, C. (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. Biochem. Pharmacol. 189, 114208. DOI: 10.1016/j.bcp.2020.114208 – IF: 5.858*

Chapter overview

Besides being used as regulatory elements to control miRNA biogenesis, G4s in pre-miRNAs can also be used for diagnostic purposes, namely as molecular recognition agents to detect their targets. Nucleolin was recently described as a promising biomarker for the early-stage diagnosis of prostate cancer. Moreover, the protein was found to bind with high-affinity G4 structures. Considering that few studies have been conducted to characterize G4s in pre-miRNAs and their binding to ligands and/or nucleolin, we investigated the formation of these complexes. Herein, we studied the G4 formation in pre-miRNA 92b and unveiled its stabilization by acridine orange derivatives. Using a set of computational and biophysical approaches, we disclosed that the acridine orange derivative C_8 bound G4 parallel motif with high affinity and stabilized the structure. This proof-of-concept study demonstrated that the G4 in pre-miRNA 92b is resistant to nucleases and can capture and detect nucleolin with high specificity from complex samples.

Recognition of nucleolin through interaction with RNA G-quadruplex

Abstract

The development of novel biomarkers for early-stage diagnosis of prostate cancer (PCa) has attracted the attention of researchers in the last decade. Nucleolin (NCL) has emerged as a possible biomarker of PCa due to its high expression levels in the surface of PCa cells and affinity towards parallel G4s since it contains four RNA-binding domains (RBDs). Herein, we developed a novel strategy based on a microfluidic platform for the detection of NCL in biological samples, such as human plasma. The RNA G4 (rG4) sequence found in human precursor microRNA 92b (pre-miR-92b) was used as a molecular recognition probe since it forms a single dominant parallel rG4 conformation in the presence of 0.1 mM K^+ as confirmed by NMR spectroscopy. The additional stability of the rG4 structure was provided by the acridine orange derivative ligand C_8, which stabilizes the pre-miR-92b rG4 structure, as denoted by an increase in more than 30 °C of its melting temperature. FRET-melting assay revealed a remarkable synergistic effect of NCL RBD1,2 and C_8 on the stabilization of the pre-miR-92b rG4. The binding of pre-miR-92b to NCL RBD1,2 was determined by *in silico* studies, which revealed a binding pocket formed by a 12-residue linker between RBD1 and RBD2. Both, pre-miR-92b rG4 and pre-miR-92b rG4/C_8 complex demonstrated high affinity towards NCL RBD1,2, as proved by fluorimetric titrations (K_D range between 10^{-12} and 10^{-9} M). The stability and nuclease resistance of pre-miR-92b rG4 and pre-miR-92b rG4/C_8 complex were evaluated as molecular recognition probes to capture and detect NCL. Finally, the microfluidic platform detects NCL in complex biological samples, such as human plasma. Overall, this work demonstrates the usefulness of the microfluidic platform based on the pre-miR-92b to detect NCL and the possibility to be used as a valuable biomedical tool in PCa diagnosis.

Keywords

Nucleolin; RNA G-quadruplex; Acridine ligands; Microfluidics and Prostate cancer

Introduction

The recent development of valid biomarkers for prostate cancer (PCa) diagnosis, prognosis, and therapeutic monitoring have a tremendous impact in the clinical management of the disease. By 2030 new PCa cases diagnosed worldwide are expected to increase to 1,700,000, together with a predicted increase of approximately 500,000 deaths [1]. Early-stage cancer detection is the key strategy to reduce PCa death rates significantly in the long term. Prostate-Specific Antigen (PSA) testing became a widespread test for early-stage detection worldwide. However, there are inherent limitations of PSA testing, as the lack of specificity and the high percentage of false-positive or false-negative results [2]. Besides PCa, benign prostatic hyperplasia and prostatitis can also increase PSA levels [3]. Promisingly, a new generation of PCa biomarkers is emerging, consisting of serum, plasma, and tissue-based experiments that may supplement or complement PSA testing [4].

Nucleolin (NCL) is a multifunctional RNA-binding phosphoprotein that has been implicated in many mechanisms of cell biology, being mainly localized at the cell nucleolus and nucleoplasm [5]. However, it is also found outside the nucleolus, in the cytoplasm, as well as on the cell membrane [5]. The overexpression of NCL and increased localization at the cell membrane were reported in PCa cell lines [6]. NCL contains four RNA-binding domains (RBDs), which mediate the interaction with nucleic acids [7]. It is known that NCL plays a direct role in the biogenesis of specific oncogenic and suppressor miRNAs [8]; however, there is a significant lack of knowledge of the interaction pre-miRNA/NCL.

Cell surface NCL staining pattern and expression level were described as suitable markers for the sensitive and specific detection of circulating PCa cells [9]. In order to develop a novel point of care (PoC) platform, a microfluidic device functionalized with LNA aptamers targeting epithelial cell adhesion molecule (EpCAM) and NCL was developed for the capture of circulating tumour cells [10]. Because of their small dimensions, microfluidic systems have become promising PoC devices for early-stage detection of prostate cancer [11], [12]. They present several advantages over conventional devices, such as reduced sample and reagent consumption and simple process and portability [13]. Therefore, it is not surprising that nucleic acid oligonucleotides have emerged as novel recognition molecular probes in microfluidic platforms [14].

RNA sequences with the potential to form G4s are attractive molecular recognition probes, and cell surface NCL has been reported as a target for the RNA G4 (rG4) found in human precursor microRNA 149 (pre-miR-149) (Please see Chapter 8) [15]. A similar rG4 sequence (5′ GGGCGGGCGGGAGGG 3′) was found in the human precursor microRNA 92b (pre-miR-92b) [16]. As previously described, the formation of the rG4 parallel structure at the 5′-end disrupts the stem-loop and leaves a single-stranded sequence in the 3′-end [16], [17]. Its formation open-up the understanding of the miR-92b biogenesis. miR-92b is upregulated in several cancers and acts as an oncomiR in certain malignancies, including glioma [18], non-small cell lung cancer [19], oral squamous cell carcinoma [20], and prostate cancer [21]. These properties make NCL a promising biomarker for using as a target of the rG4 founded in pre-miR-92b.

Overall, our study provides a new strategy based on a pre-miR-92b as a molecular recognition probe of a microfluidic platform for the detection of NCL. The use of a fluorescent labelled probe allows the prompt recognition of NCL. The pre-miR-92b rG4 structure and its interaction with NCL were characterized by biophysical methods like circular dichroism (CD), fluorescence resonance energy transfer (FRET) melting assay, fluorescence titrations, and Nuclear Magnetic Resonance (NMR) spectroscopy. Additionally, the predicted complex between pre-miR-92b rG4 and NCL RBD1,2 was also determined by *in silico* studies. The detection of NCL RBD1,2 was achieved in 10% human plasma samples spiked with NCL RBD1,2 through a microfluidic system, by taking advantage of the intrinsic fluorescence of the labelled pre-miR-92b rG4. The results showed the suitability of the developed microfluidic platform for the detection of the model protein NCL RBD1,2 in 10% human plasma samples.

Materials and Methods

Oligonucleotides and ligands

All oligonucleotides were obtained from Eurogentec (Belgium) or IDT technologies (Belgium) with HPLC-grade purification. The pre-miR-92b rG4 sequences used in this work are 5′-GGGCGGGCGGGAGGG – 3′, 5′- FAM GGGCGGGCGGGAGGG TAMRA – 3′ and 5′- Cy3.5 GGGCGGGCGGGAGGG – 3′. The duplex sequence d(TATAGCTAT-hexaethyleneglycol-TATAGCTATA) labelled with FAM and TAMRA was used. The RNA sequence 5′- CCUGGUUUGACAGG – 3′ was used as control for the stability assays. Stock solutions of approximately 1 mM were prepared using DEPC water (Sigma-Aldrich, USA) and stored at −80 °C until used. The concentration of oligonucleotide samples was determined from the absorbance at 260 nm by using the molar extinction coefficient.

Annealing of oligonucleotide sequences was performed by heating the samples for 10 min at 95 °C and slowly cooling on ice for 30 min before the experiments. Synthesis and purification of the ligands 10-(5-(4-iodobenzamide)pentyl))-3,6-bis(dimethylamino) acridinium iodide (C_5), 10-(5-(4-iodobenzamide)pentyl))-3,6-bis(dimethylamino) acridinium (C_5-NH_2), 10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium iodide (C_8) and 10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium (C_8-NH_2) were performed as previously described [22]. 3,3′-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate (PhenDC3) was obtained from Sigma-Aldrich (USA). The chemical structures of each ligand are depicted in **Figure 6.1A**. Stock solutions of the compounds were prepared as 10 mM solutions in DMSO (Thermo Fisher Scientific, USA) and their subsequent dilution was done using nuclease-free water. Recombinant NCL protein (RBD1,2) was purchased from NZYTech, Lda (Portugal), and Bovine Serum Albumin (BSA) and Human Serum Albumin (HSA) were obtained from Sigma-Aldrich (USA).

Circular dichroism

CD spectra were acquired in a Jasco J-815 spectrometer (Jasco, USA), using a Peltier temperature controller (model CDF-426S/15). pre-miR-92b rG4 sequence was annealed as previously described. Unless otherwise stated, a 1 mm path-length quartz cuvette (Hellma, Germany) was used with pre-miR-92b rG4 at 10 μM in 10 mM lithium cacodylate (Sigma-Aldrich, USA) at pH 7.2, supplemented with 0.1 mM KCl (Thermo Fisher Scientific, USA). The required volume for the titrations was added directly in the quartz cell. The CD melting experiments were performed in the temperature range 20–100 °C, with a heating rate of 2 °C/min by monitoring the ellipticity at 261 nm. Spectra acquisition was performed in the absence and presence of increasing concentrations of KCl or 2 M equivalents of ligands. Data was converted into fraction folded (f) plots.

$$f = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \qquad \text{(Equation 6.1)}$$

Data points were then fitted to a Boltzmann distribution or a Biphasic dose–response equation (OriginPro 2016) and the melting temperatures were determined from the two-state transition model using the first derivative method.

NMR spectroscopy

Standard zgesgp ^{1}H NMR spectra were acquired on a 600 MHz Bruker Avance III spectrometer with a QCI cryoprobe operating at a proton frequency of 600 MHz at different temperatures. Pre-miR-92b rG4 was used at a concentration of 100 μM with a total volume of 190 μL in a 3 mm NMR tube, annealed as described above and supplemented with 10% D_2O (Eurisotop, France). NMR titrations were carried out by adding increasing amounts of KCl directly to the 3 mm NMR tube. All spectra were acquired and processed with the software Topspin 3.1. Images were prepared using TopSpin 4.0.6. Chemical shifts (δ) are reported in ppm.

Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel (15%) electrophoresis was used to visualize the oligonucleotides. pre-miR-92b rG4 samples were prepared at a concentration of 30 μM. Sucrose (Sigma-Aldrich, USA) was added to the samples at a final concentration of 23%. Oligonucleotide marker was purchased from NZYTech, Lda (9, 15, 21, 30, 60 and 90 bp) and was loaded in parallel on the gel. Different concentrations of NCL RBD1,2 (50, 100, 250 and 500 nM) were used to evaluate the formation of supramolecular complexes with pre-miR-92b rG4. For the formation of supramolecular complexes with ligands, a molar ratio of 1:1 of pre-miR-92b rG4/ligand was prepared and incubated for 30 min before addition of NCL RBD1,2 (500 nM). Electrophoresis was performed at 4 W per gel to reach a temperature close to 20 °C. After electrophoresis, the gels were stained by Stains-All solution (Sigma-Aldrich, USA) for 10 min under gentle agitation. Thereafter, they were incubated overnight in water for discolouration under the light before visualization on GE Typhoon Trio Imager Scanner (GE healthcare, USA).

Fluorescence Resonance Energy Transfer (FRET)-melting

FRET-melting experiments were performed on a CFX Connect™ Real-Time PCR Detection System (Bio-Rad, USA), equipped with a FAM filter (λ_{ex} = 492 nm; λ_{em} = 516 nm). The fluorescence measurements were performed in 10 mM lithium cacodylate pH 7.2, supplemented with 0.1 mM KCl. The experiments were carried out with the pre-miR-92b rG4 concentration at 0.2 μM. pre-miR-92b rG4 solution was annealed before the experiment, as described previously. In order to assess the selectivity of each ligand to rG4 sequences, the duplex sequence d(TATAGCTAT-hexaethyleneglycol-TATAGCTATA) labelled with FAM and TAMRA was used. Unless otherwise stated, the samples were prepared by aliquoting 15 μL of oligonucleotide

solution into each strip, 5 µL of ligands solution at 0.2 µM and 5 µL NCL RBD1,2 at four different concentrations (50, 100, 250 and 500 nM). BSA was used as negative control at the same four concentrations (50, 100, 250 and 500 nM). Subsequently, samples were incubated 30 min at room temperature. After an initial incubation at 25 °C for 5 min, the thermocycler was set to perform a stepwise increase of 1 °C every 1 min, from 25 °C to 95 °C, and measurements of FAM emission were acquired after each step. The melting temperatures were determined from the normalized curves as the temperature for which the normalized emission was 0.5.

pre-miR-92b G4 model preparation

The predicted structure of pre-miR-92b rG4 was built based on the model of the parallel-stranded telomeric G4 AGGG(TTAGGG)3 (PDB: 4G0F), as previously reported (Please see Chapter 7) [23]. First, the telomeric DNA sequence was modified in order to correspond to the nucleotide sequence of pre-miR-92b rG4 (5′- GGGCGGGCGGGAGGG – 3′) **(Figure 6.1B)**. Adenine residue in position 1 and thymine residues in positions 5, 6, 11, 12, 17 and 18 were deleted. Then, adenine residues in positions 7 and 13 were mutated to cytosines. The changes were performed in the loop residues to maintain the integrity and stability of the three G-tetrads. Finally, the PDB structure file was manually edited at the 2′ position of each sugar base to convert deoxyribose into ribose by modifying 2′-H to 2′–OH. All modifications were performed with UCSF Chimera 1.11.2. The built structure of pre-miR-92b was optimized for further experiments by running fully solvated molecular dynamics (MD) simulations using GROMACS 2016.3 with the following parameters. The pre-miR-92b rG4 structure was initially centred in an octahedral solute box and filled with TIP3P water molecules and K^+ atoms. The structure was then submitted to 5000-step energy minimization with no restraints. A 100 ps equilibration was then performed under a modified Berendsen thermostat, with a temperature gradient from 0 to 300 K, followed by another 100 ps of equilibration done under aforementioned thermostat and a Parrinello-Raman barometer. Finally, 100 ns of MD were carried out. The final snapshot of the 100 ns simulation was used as an optimized model for further experiments. 2D diagrams of protein/ligand complex were determined by using LigPlot + v.2.2. All molecular images were rendered using UCSF Chimera 1.11.2.

Molecular dynamics

The minimized model of pre-miR-92b rG4 was used as a target for molecular docking studies. Both the pre-miR-92b rG4 structure and the NCL RBD1,2 (PDB: 2KRR) were

optimized for docking using Dock Prep tool of Chimera 1.11.2. After assigning polar hydrogens and Gasteiger charges, docking simulations were carried with AutoDock 4.0 program using Lamarckian genetic algorithm. The size of the box was constrained to 100 × 100 × 100 Å along the ×, y, and z axes, respectively, with a grid spacing of 0.375 Å. A total of 25 runs were performed with an initial population of 150 random individuals, a maximum number of evaluations set to 2.5 × 107, the rate of mutation and crossover set to 0.02 and 0.8, respectively, and elitism value of 1. The most representative structures were selected based on the binding free energy and further processed with MD simulations the all-atom force field AMBER99SB of GROMACS 2016.3. For MD simulations, the pre-miR-92b rG4/NCL RBD1,2 complex was centred in an octahedral box and solvated with TIP3P water molecules and K atoms to neutralize the system. After an energy minimization of 1000 steps using the steepest descent algorithm, the system was gradually heated to 300 K in 100 ps under the control of Berendsen thermostat followed by 100 ps isobaric simulation under the control of the Parrinello-Rahman barostat. Finally, explicit solvent MD runs of 100 ns were performed for pre-miR/92b rG4-NCL RBD1,2 complex at constant temperature (300 K) and pressure (1 atm). The Particle mesh Ewald (PME) method was used for calculating long range electrostatic interactions. A 1 nm cut-off was applied to short range Lennard-Jones interactions. Coordinates were collected in trajectory files every 10 ps. All molecular images were rendered using UCSF Chimera 1.11.2.

Fluorescence titrations

All steady-state fluorescence measurements were collected at 25 °C on a FluoroMax4 (Horiba, Japan). Samples and references were scanned using 1 cm path length quartz microvolume cuvettes (Hellma, Germany) with an optimal volume of 700 µL. All spectra were scanned with an integration time of 0.5 s, an emission and excitation slit width of 1.7 nm and step size of 1 nm. The association between pre-miR-92b rG4 or pre-miR-92b rG4/C_8 complex with NCL RBD1,2 was followed by titrating the protein and measuring the change in fluorescence. The titration was performed by adding NCL RBD1,2 from stock solutions, followed by 5 min for equilibration. The obtained data were converted into a fraction of bound ligand (α) plots using the following equation:

$$\alpha = \frac{I - I_\lambda^{free}}{I_\lambda^{bound} - I_\lambda^{free}} \qquad \text{(Equation 6.2)}$$

where *I* is the fluorescence intensity of pre-miR-92b rG4-Cy3.5/NCL RBD1,2 ratio and, I_λ is the fluorescence intensity of the free and fully bound pre-miR-92b rG4. Data points

were then fitted to a two-site binding function (OriginPro 2016) and two K_D values were determined.

Nuclease stability assays

Nuclease stability assays of pre-miR-92b rG4, pre-miR-92b rG4/C_8, pre-miR-92b rG4/NCL RBD1,2 and pre-miR-92b rG4/C_8/NCL RBD1,2 complexes were conducted in the absence and presence of cell culture medium RPMI (Merck, Germany) 1640 with phenol red supplemented with 10% (m/v) FBS (Thermo Fisher Scientific, USA) and 1% (m/v) Penicillin-Streptomycin (Thermo Fisher Scientific, USA) for 1, 3, 12, 24 and 48 h at 37 °C. Stock solution of the RNA hairpin (5′ CCUGGUUUGACAGG 3′), which was used as a control, and pre-miR-92b rG4 was prepared in 10 mM lithium cacodylate buffer pH 7.2, supplemented with 0.1 mM KCl. The complexes of control RNA and pre-miR-92b rG4 with C_8 were prepared at molar ratio of 1:1. Thereafter, RNA sequences were annealed in the absence and presence of C_8 by heating to 95 °C for 10 min, following by a slowly cooling in ice for 30 min. After that, NCL RBD1,2 at 500 nM was added to the complex's solution. The degradation pattern were observed by 20% non-denaturing PAGE supplemented with 10 mM KCl [24]. Electrophoresis was performed at 4 W per gel to reach a temperature close to 20 °C. After that, the gels were stained by stains-all solution (Sigma-Aldrich, USA) for 10 min under gentle agitation. Thereafter, they were incubated overnight in water for discolouration under the light before visualization. The densitometry analysis of band intensities was performed with ImageJ (National Institute of Health, USA). A one-way analysis of variance (ANOVA) was performed with Prism (GraphPad Software, USA) to evaluate the significance of the differences between the groups.

Fabrication of microfluidic device

Polydimethylsiloxane (PDMS) microfluidic devices were fabricated using standard mold replication techniques, as previously described [25]. The device was designed to have two-channel heights to allow the holding of the beads inside the channel. Briefly, to obtain the master mold two aluminium masks were fabricated by direct-write lithography and wet etching to define each of the 20 and 100 μm tall structures. The 20 μm tall structure was fabricated using SU-8 2015 (Microchem Corp., USA), spin-coated on top of a silicon substrate. After development with propylene glycol monomethyl ether acetate (PGMEA) (Sigma-Aldrich, USA), the second layer of SU-8 50 for the 100 μm structure layer was spin-coated on top of the first layer. The second hard mask was manually aligned with the 20 μm features prior to exposure to UV

light. Finally, exposed SU-8 was developed in PGMEA and hard-baked at 150 °C for 15 min. The PDMS structures were fabricated by mixing PDMS pre-polymer with curing agent in a 10:1 wt ratio (Sylgard 184 silicone elastomer kit, Dow Corning, USA). The PDMS was poured on top of the master mold and cured at 70 °C for 90 min. After peeling PDMS from the master mold, access holes were punched using a 14 ga and 20 ga blunt needles for the inlets and outlets, respectively. The device was sealed against a 500 μm PDMS slab after treating both surfaces with an oxygen plasma (Harrick Plasma, USA). The scheme of the microfluidic device channels is shown in **Figure 6.1C**.

Microfluidic structure beads packing

The description of the microfluidic structure CM Sephadex C-25 beads (Sigma-Aldrich, USA) packing were carried out as previously detailed [25]. Briefly, the liquids were driven inside the microchannel using a syringe pump exerting a negative pressure at the outlet (NE-1002X, New Era Pump System Inc., USA). The beads were first suspended in a solution of 20% (w/w) polyethylene glycol 8000 (PEG) (Sigma-Aldrich, USA) and packed inside the channel at a flow rate of 7 μL/min, followed by a washing step with PBS 1 × at a flow rate of 7 μL/min.

Microfluidic experiments and image acquisition

Firstly, the pre-miR-92b rG4 (5 μM) was incubated with C_8 at 0.1 M equivalents during 10 min. After that, different concentrations of NCL RBD1,2 were added and incubated for 30 min. PSA and human serum albumin (HSA) were used as control proteins. After that, 0.1 M acetate buffer, pH 4.5 was added to change the surface charge of the pre-miR-92b rG4/C_8/NCL RBD1,2 complex. The experiment was carried out at a flow rate of 1 μL/min for 30 min. Finally, the channels were washed with PBS 1 × at a flow rate of 5 μL/min for 2 min to remove non-specifically bound molecules. All fluorescence images were acquired with an exposure time of 2 s in an Axio Imager Z2 microscope using 579 nm as the excitation wavelength (BP 565/30) and collecting the emission at 591 nm (BP 620/60) and processed in Zeiss Zen Software (SP2, 2010) and ImageJ (National Institute of Health, USA). Fluorescence emission values were obtained by averaging the entire end-section of the micro-columns. For every individual experiment, new structures with fresh functionalization steps were performed. ImageJ was used to analyze the fluorescence images.

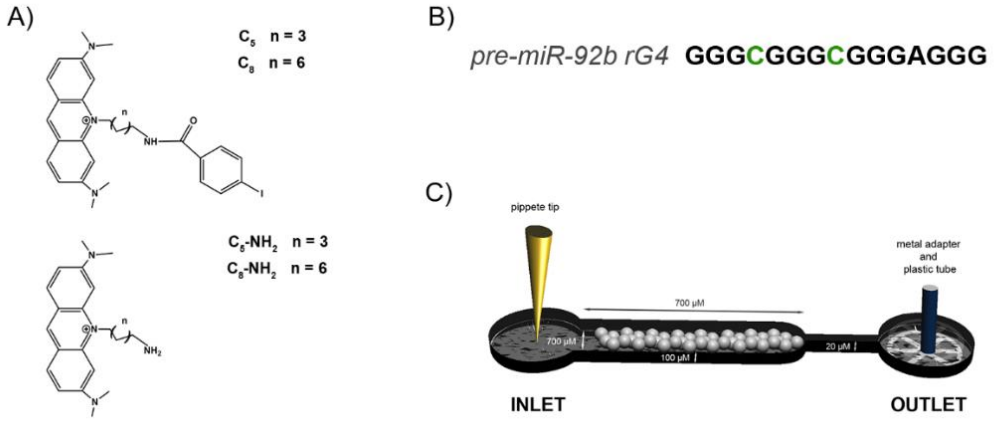

Figure 6.1. (A) Chemical structure of the acridine orange derivative ligands used in this study. **(B)** pre-miR-92b rG4 sequence (5'-GGGCGGGCGGGAGGG-3') (mutated residues are highlighted in green). **(C)** Microchannel design of the two-height microfluidic structure.

Results

Herein, we used the rG4 sequence (5′-GGGCGGGCGGGAGGG-3′) previously identified in pre-miR-92b, as a molecular recognition probe. Hence, we employed CD and NMR spectroscopies and non-denaturing PAGE to investigate the formation and topology of the rG4 sequence found in pre-miR-92b. The techniques are remarkably sensitive to changes in the conformational modifications of the G4 structures and provide useful structural information [26], [27]. The interaction of pre-miR-92b rG4 with NCL RBD1,2 was also studied by fluorescence titrations, FRET melting, molecular docking and molecular dynamics studies. Since the pre-miR-92b rG4 sequence can form a stable parallel G4, its applicability to recognize NCL in microfluidic devices is of utmost importance concerning the advantages of those systems. Overall, we intend to detect NCL in biological samples for the diagnosis of PCa.

rG4 formation and stabilization in the presence of K^+ ions

CD measurements of the pre-miR-92b rG4 in the absence and presence of increasing amounts of KCl were performed to assess the rG4 topology. As seen in **Figure 6.2A**, upon titrating KCl into the pre-miR-92b rG4 solution, a parallel-stranded rG4 topology was formed, showing a positive band around 260 nm and a negative band around 240 nm. During titration, the ellipticity increases, while the typical CD signature for parallel-stranded rG4s was conserved. The influence of K^+ on stabilization was further investigated by thermal denaturation using CD melting experiments. The pre-miR-92b rG4 melting temperatures (T_m) were determined from the maxima of the first derivatives

plots of absorbance at 261 nm versus temperature. **Figure 6.2B** represents the CD melting curves with increasing KCl concentration. The addition of KCl stabilized the pre-miR-92b rG4 structure by increasing the T_m values **(Table 6.1)**. Interestingly, one melting point was observed in the presence of 0.1 mM KCl, while two distinct melting points were observed for KCl concentration in a range between 1 and 100 mM KCl.

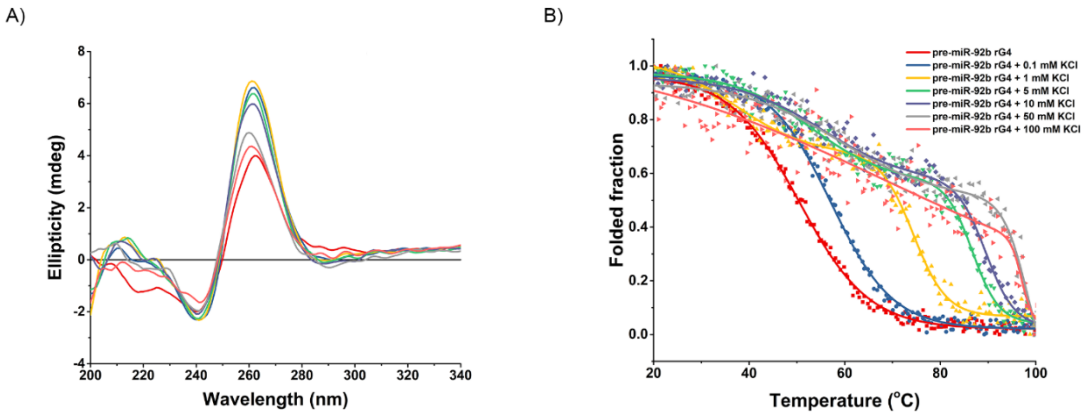

Figure 6.2. (A) CD titration spectra of pre-miR-92b rG4 at 10 μM with increasing concentrations of K^+. **(B)** CD melting curves at 10 μM in the absence and presence of increasing concentrations of K^+. Data points were recorded at 261 nm.

Table 6.1. Thermal stabilization of pre-miR-92b rG4 92b with increasing concentration of KCl measured by CD melting experiments.

	rG4 92b	
[K^+] (mM)	**T_m^1 (°C)**	**T_m^2 (°C)**
0	50.5 ± 0.1	-
0.1	57.1 ± 0.1	-
1	37.2 ± 1.3	74.4 ± 0.3
5	53.0 ± 0.9	86.7 ± 0.3
10	55.3 ± 1.2	89.6 ± 0.3
50	59.1 ± 1.3	>90
100	68.4 ± 16.4	>90

Acridine orange derivatives stabilize pre-miR-92b rG4

CD experiments were used to evaluate the ability of acridine ligands, C_5, C_5-NH_2, C_8, C_8-NH_2 to bind and stabilize the pre-miR-92b rG4 structure. The CD spectra were acquired

in the presence and absence of 2 M equivalents of ligands. As shown in **Figure 6.3A**, upon addition of ligands, no significant variations on the pre-miR-92b rG4 ellipticity were observed, suggesting that ligands preserved the rG4 parallel-stranded topology. Then, as demonstrated by **Figure 6.3B**, the influence of ligands on the thermal stability of the pre-miR-92b rG4 was determined via CD melting experiments by measuring the change in the melting temperature. Initially, the saline solution of the pre-miR-92b rG4 was adjusted for melting temperature, which was 57.1 °C in the presence of 10 mM LiCaco pH 7.2 supplemented with 0.1 mM KCl. Upon addition of 2 M equivalents of ligands, C_8, C_5 and C_5-NH_2 demonstrated a higher stabilization capability, with ΔT_m values higher than + 30 °C. On the other hand, C_8-NH_2 only showed a modest stabilizing effect with a ΔT_m value of + 6.8 °C.

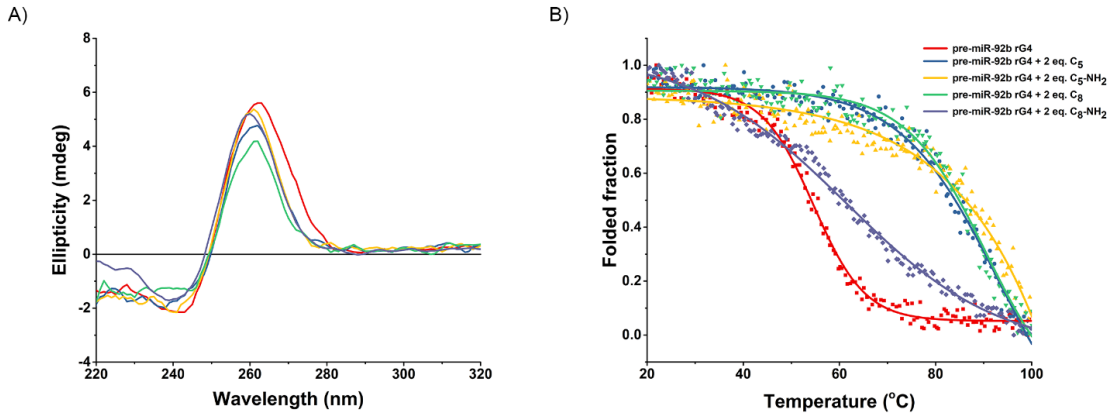

Figure 6.3. (A) CD titration spectra of pre-miR-92b rG4 in the absence and presence of ligands. **(B)** CD melting curves of pre-miR-92b rG4 in the absence and presence of 2 M equivalents of each of the ligands. Data points were recorded at 261 nm.

Structural insights into pre-miR-92b rG4 structure by NMR spectroscopy studies

To obtain more structural insights into the pre-miR-92b rG4 structure, we employed ^{1}H NMR spectroscopy. The ^{1}H NMR spectra of the pre-miR-92b rG4 upon KCl titration are shown in **Figure 6.4**. A distinct set of nine signals assignable to the guanine imino protons of the pre-miR-92b rG4 structure is visible in the imino proton region (10–12 ppm) in the presence of 0.1 mM KCl. At this ionic strength, pre-miR-92b rG4 seems to fold into a single dominant G4 conformation. However, when the ionic strength was increased in a range between 1 and 100 mM KCl, other conformations seems to be formed, and two sets of signals became visible in the imino proton region, one around 11.5 ppm and another around 10.5 ppm. Additionally, the second set of signals became more prominent as ionic strength increases, which suggests an increase in the stability of the rG4 conformations.

The effect of temperature on rG4 structure was further investigated by increasing the temperature and recording ^{1}H NMR spectra. The results are shown in **Figure 6.5**. Firstly, the pre-miR-92b rG4 solution was heated from 17 °C to 27 °C in the presence of 0.1 mM KCl, but a single dominant rG4 conformation remains visible, as indicated by the unique set of imino signals. On the other hand, as seen in **Figure 6.5**, in the presence of 1 mM KCl two sets of imino protons are visible at 27.7 °C. Similarly, in the presence of 100 mM KCl, the pre-miR-92b rG4 spectra also indicated two sets of imino signals even upon heating and vanished at 60 °C. The conformational preference strongly depends on the amounts of the K^{+} ion present in solution and temperature.

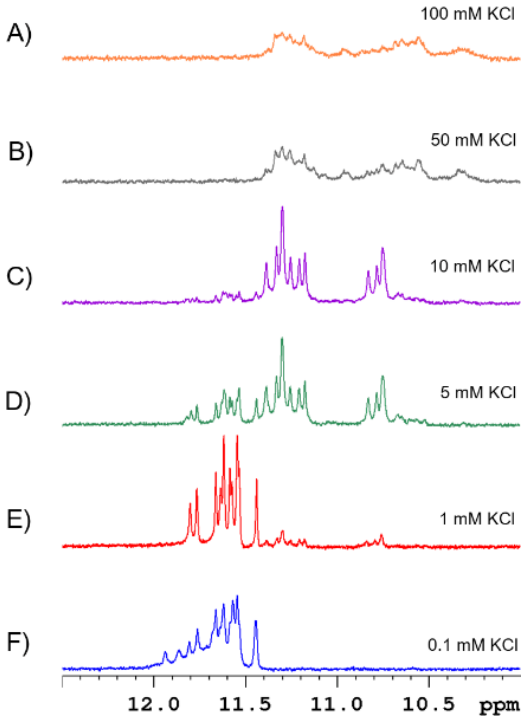

Figure 6.4. 1D ^{1}H spectra showed the imino region of pre-miR-92b rG4 (0.1 mM) upon KCl titration of **(A)** 100 mM, **(B)** 50 mM, **(C)** 10 mM, **(D)** 5 mM, **(E)** 1 mM and **(F)** 0.1 mM KCl. The spectra were recorded in 90% H_2O, 10% D_2O at 27.7 °C.

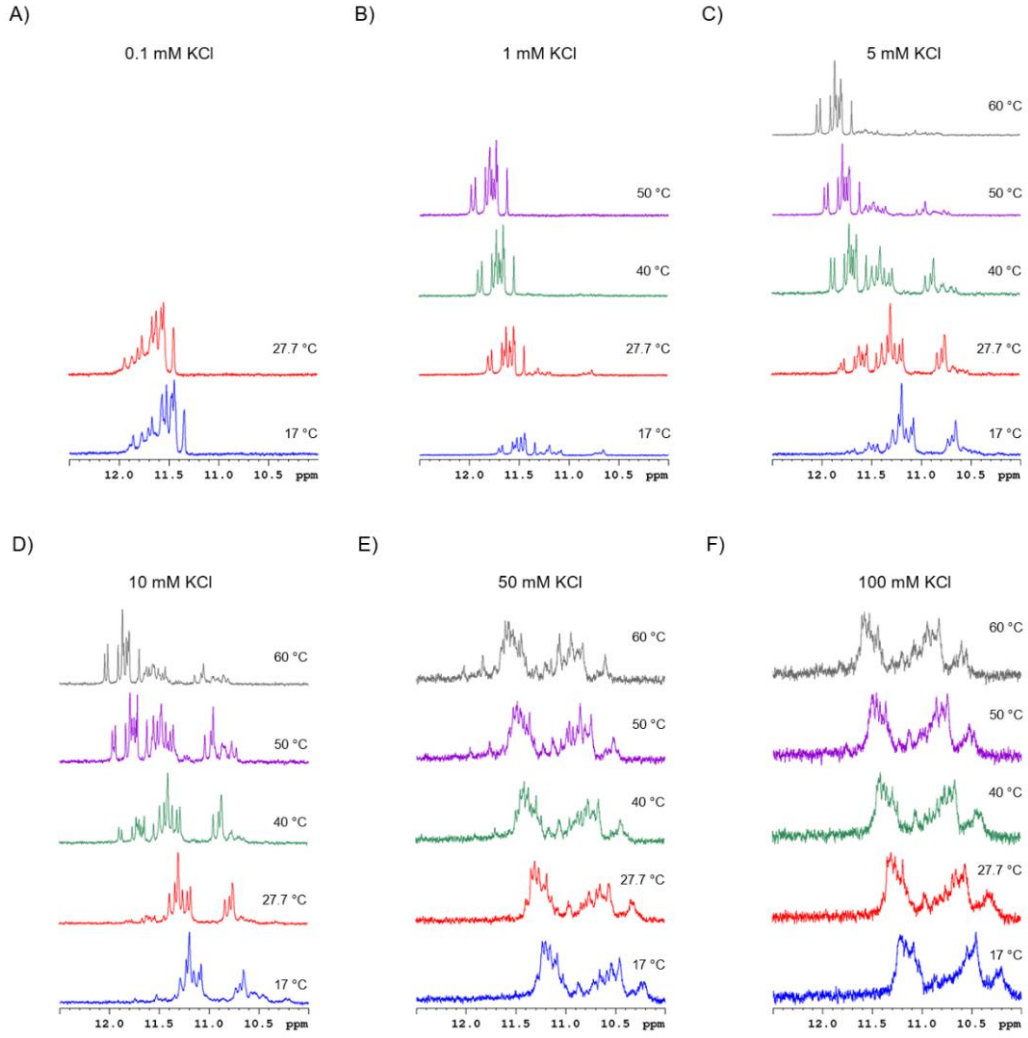

Figure 6.5. 1D ^{1}H spectra showed the imino region of pre-miR-92b rG4 (0.1 mM) upon **(A)** 0.1 mM; **(B)** 1 mM; **(C)** 5 mM; **(D)** 10 mM; **(E)** 50 mM and **(F)** 100 mM KCl titrations. The spectra were recorded in 90% H_2O, 10% D_2O at different temperatures, 17 °C (blue), 27.7 °C (red), 40 °C (green), 50 °C (purple) and 60 °C (grey).

Molecularity of pre-miR-92b rG4 and pre-miR-92b rG4/ligands complex by non-denaturing PAGE

Following the structural studies, we intend to determine the molecularity of the pre-miR-92b rG4 or pre-miR-92b rG4/ligand complex in the absence and presence of NCL RBD1,2 by polyacrylamide gel electrophoresis. The pre-miR-92b rG4 sequences were diluted to 30 µM and complexes with ligands were prepared at a 1:1 M ratio. After that, NCL RBD1,2 was added to each pre-miR-92b rG4/ligand sample at a concentration of 0.5 µM. The complexes pre-miR-92b rG4/NCL RBD1,2 were prepared by adding different concentrations (0.05, 0.1, 0.25 and 0.5 µM) of the protein. Single-stranded oligonucleotides were used as a ladder for comparison. The results of gel

electrophoresis are highlighted in **Figure 6.6**. The results revealed the formation of the rG4 structure even in presence of 10 mM LiCaco pH 7.2 (without KCl). The stained gel also showed an intense band in the presence of 0.1 mM KCl, suggesting a major dominant unimolecular folding form in solution. On the other hand, the multiple intense bands shown in presence of 100 mM KCl suggest the folding into different intermolecular rG4s. The formation of a major dominant unimolecular form was also observed in the presence of ligands and NCL RBD1,2.

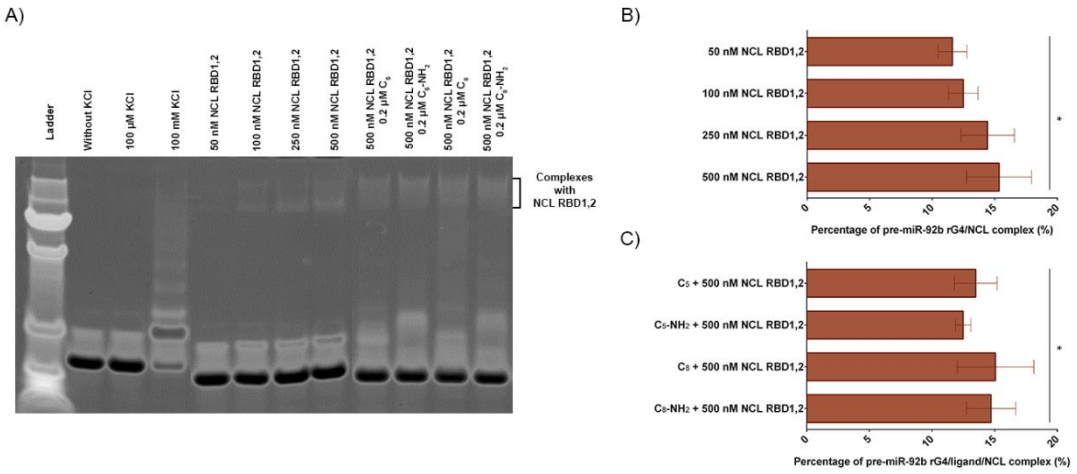

Figure 6.6. (A) pre-miR-92b rG4 detected on non-denaturing polyacrylamide gel electrophoresis. The samples were loaded on a non-denaturing 15% acrylamide gel supplemented with 10 mM KCl. Migration was performed at 20 °C. Migration markers were single-stranded (n = 9, 15, 21, 30, 60 and 90). The diffuse intense band below correspond to the dyes bromophenol blue and xylene cyanol used to follow the migration. After electrophoresis, the gels were stained and visualized by GE Typhoon Trio Imager Scanner. **(B)** Percentage of pre-miR-92b rG4/NCL RBD1,2 complexes, formed in the presence of different concentrations of protein (50, 100, 250 and 500 nM). **(C)** Percentage of pre-miR-92b rG4/ligand/NCL RBD1,2 complexes, formed in the presence of NCL RBD1,2 (500 nM) and each ligand (30 µM). Three independent experiments were conducted. Data are shown as mean with standard error of mean. A one-way analysis of variance (ANOVA) was performed to evaluate the significance of the differences between the complex formation values of the different ligands and between the protein concentrations (the asterisks indicate $p < 0.05$).

Structural prediction by molecular docking and MD simulations

In order to investigate the *in silico* interaction of pre-miR-92b rG4 with NCL RBD1,2, we edited the previous model of the parallel-stranded telomeric G4 AGGG(TTAGGG)3 (PDB: 4G0F). Then, the model of the pre-miR-92b rG4 was optimized for MD studies to remove unfavourable van der Waals interactions and incorrect atom names and connectivities. As seen by RMSD graph in **Figure 6.7A** a stable rG4 structure was

obtained during 100 ns. The final pre-miR-92b rG4 snapshots of the 100 ns MD run are shown in **Figure 6.7** (side view – **Figure 6.7B** and top view – **Figure 6.7C**).

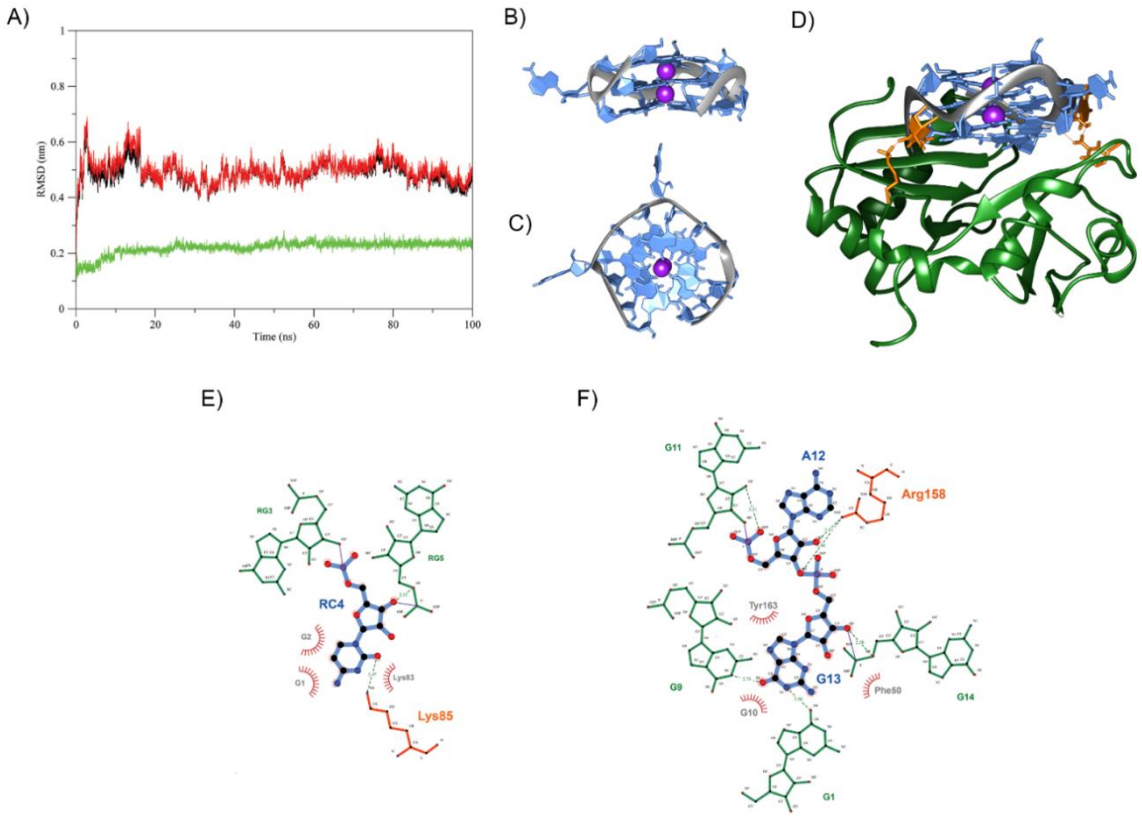

Figure 6.7. (A) RMSD plot for the pre-miR-92b rG4 (green); NCL RBD1,2 (black) and pre-miR-92b rG4/NCL RBD1,2 complex (red) structure during 100 ns MD simulation. Final snapshots of pre-miR-92b rG4 and pre-miR-92b rG4/NCL RBD1,2 complex after 100 ns MD production. **(B)** Side view showing a parallel-stranded rG4 **(C)** Top view showing a parallel-stranded G4 with propeller-type loops (depicted in gray). **(D)** pre-miR-92b rG4/NCL RBD1,2 complex. Propeller-type loops are depicted in gray. K^+ cations are depicted as purple spheres. NCL RBD1,2 is represented in green and contact residues of rG4 and NCL RBD1,2 are highlighted in orange. H-bonds are shown in red. Ligplot + image showing site-specific hydrogen bonding and hydrophobic interactions of NCL RBD1,2 residues and **(E)** pre-miR-92b rG4 C4 residue and **(F)** pre-miR-92b rG4 A12 and G13 residues. NCL RBD1,2 residues are represented in orange. pre-miR-rG4 residues with hydrogen bonding interactions are depicted in blue. pre-miR-92b rG4 residues without hydrogen bonding with NCL RBD1,2 are highlighted in green. H-bonds are shown in green dashed line.

Since no solution structure of full-length NCL is available, we have used the NCL RBD1,2 (PDB: 2KRR) for the pre-miR-92b rG4/NCL RBD1,2 complex simulations. The complex structure after MD simulations was depicted in **Figure 6.7D**. A stable complex was obtained during the 100 ns simulations, as seen by the RMSD graph shown in **Figure 6.7A**. The 12-residue linker between RBD1 and RBD2 seems to create a binding pocket to the rG4 structure, and the main interactions occur in this region. The results revealed the binding of pre-miR-92b rG4 to NCL RBD1,2 via C4, A12 and G13, while the amino

acid residues involved in the interaction are Lys85 and Arg185. The 2D diagram plot depicted in **Figure 6.7E and 6.7F** revealed the site-specific hydrogen bonding and hydrophobic interactions between those residues.

pre-miR-92b rG4/NCL RBD1,2 complex evaluation by FRET-melting assay

Once the ability of pre-miR-92b rG4 structure to bind NCL RBD1,2 was predicted at the atomic level by docking and MD studies, the ability of the protein to stabilize the rG4 structure was assessed by FRET-melting assay. The experiments were conducted in the absence and presence of NCL RBD1,2, and according to the experimental procedures previously developed [28], [29]. The experiments were performed with pre-miR-92b rG4 labelled with FAM and TAMRA fluorophores placed at their 5′- and 3′-ends, respectively. The results are depicted in **Table 6.2** that shows the T_m values of rG4 in the absence and presence of different concentrations of NCL RBD1,2 (50, 100, 250 and 500 nM). The data show the stabilization of the pre-miR-92b rG4 structure in the presence of NCL RBD1,2, as depicted by the increase of the T_m value from 62.6 without protein to 71.0 °C with protein at 0.5 μM. On the other hand, the negative control BSA did not show any stabilization effect in the pre-miR-92b rG4 structure.

Experiments were also carried out in order to study how the ligands together with NCL RBD1,2 affect the stability of the pre-miR-92b rG4. The ligands selected were C_8, C_5, C_8-NH2, C_5-NH_2 and PhenDC3 at 1 M equivalent and in presence and absence of 500 nM NCL RBD1,2 **(Table 6.2)**. The stabilization effect of PhenDC3 (ΔT_m = +1.9 °C) on the pre-miR-92b rG4 is lower than that observed with C_8 (ΔT_m = +6.3 °C). In order to assess the selectivity of acridine orange ligands to pre-miR-92b rG4, we carried out experiments in the presence of the duplex sequence d(TATAGCTAT-hexaethyleneglycol-TATAGCTATA) modified in 5′ and 3′ with FAM and TAMRA (FdxT), respectively. The results are depicted in **Table 6.2**.

The results showed selectivity of all the ligands towards pre-miR-92b rG4, since no stabilization effect was observed for duplex sequence. Moreover, the results highlighted in **Table 6.2** show a remarkable synergistic effect of C_8 together with NCL RBD1,2. The T_m of the pre-miR-92b rG4 sequence increased from 62.6 to 84.6 °C in the presence of C_8 and NCL RBD1,2 at 500 nM. However, the T_m provided by ligands C_5, C_5-NH_2, C_8-NH_2 and PhenDC3 together with NCL RBD1,2 showed a modest effect on the stability or destabilization. The T_m values of C_5, C_5-NH_2, C_8-NH_2 and PhenDC3 are 72.5 °C, 77.3 °C, 70.1 °C and 70.4 °C, respectively. Overall, C_8 together with NCL, showed a remarkable result in stabilizing pre-miR-92b rG4, and was chosen for the following experiments.

Table 6.2. Pre-miR-92b rG4 binder-induced thermal stabilization (ΔT_m) measured by FRET melting experiments.

	T_m (°C) in presence of NCL RBD1,2					
Concentration (nM)	0	50	100	250	500	-
pre-miR-92b rG4	62.6 ± 0.2	63.0 ± 0.2	64.0 ± 0.2	64.3 ± 0.1	71.0 ± 0.3	-
FdxT	54.4 ± 0.2	56.5 ± 0.3	57.2 ± 0.3	59.9 ± 0.2	61.5 ± 0.3	-
	T_m (°C) in presence of ligands (1 molar equivalent)					
	Without ligand	C_8	C_8-NH_2	C_5	C_5-NH_2	PhenDC3
pre-miR-92b rG4	62.6 ± 0.2	68.9 ± 0.2	67.2 ± 0.4	64.4 ± 0.2	67.0 ± 0.6	64.5 ± 0.8
FdxT	54.1 ± 0.2	54.1 ± 0.1	54.2 ± 0.2	53.9 ± 0.5	55.6 ± 0.4	54.5 ± 0.2
	T_m (°C) rG4/ligand/NCL RBD1,2 (500 nM) complex					
	Without ligand	C_8	C_8-NH_2	C_5	C_5-NH_2	PhenDC3
pre-miR-92b rG4	71.0 ± 0.3	84.6 ± 0.5	70.1 ± 1.6	72.5 ± 1.8	77.3 ± 1.7	70.4 ± 0.4
FdxT	60.5 ± 0.3	60.1 ± 0.5	60.4 ± 0.7	58.7 ± 2.1	61.7 ± 1.1	60.2 ± 0.5

Determination of binding affinity by fluorescence titrations

Next, the interaction of NCL RBD1,2 with pre-miR-92b rG4 or pre-miR-92b rG4/C_8 complex was studied by means of fluorimetric titrations and was carried out with pre-miR-92b rG4 labelled with Cy3.5. Fluorescence emission spectra of pre-miR-92b rG4-Cy3.5 and pre-miR-92b rG4-Cy3.5/C_8 complex were recorded at 577 nm, in the absence and presence of increasing amounts of NCL RBD1,2. The binding plots are depicted in **Figure. 6.8**. Both titrations revealed a significant change in the fluorescence intensity at low protein concentrations. In order to determine the dissociation constants (K_D) from each binding curve, a two-site binding model was employed. In a two-site binding model, two K_D values can be determined due to the fact that the protein exhibit two binding sites for the same ligand. The K_D values of the pre-miR-92b rG4 for protein were K_D^1 = 0.014 ± 0.001 nM and K_D^2 = 387.63 ± 51.36 nM, while the obtained affinity values of the pre-miR-92b rG4/C_8 complex for protein were K_D^1 = 0.22 ± 0.07 nM and K_D^2 = 843.60 ± 155.00 nM. pre-miR-92b without C_8 had shown the highest K_D values towards protein.

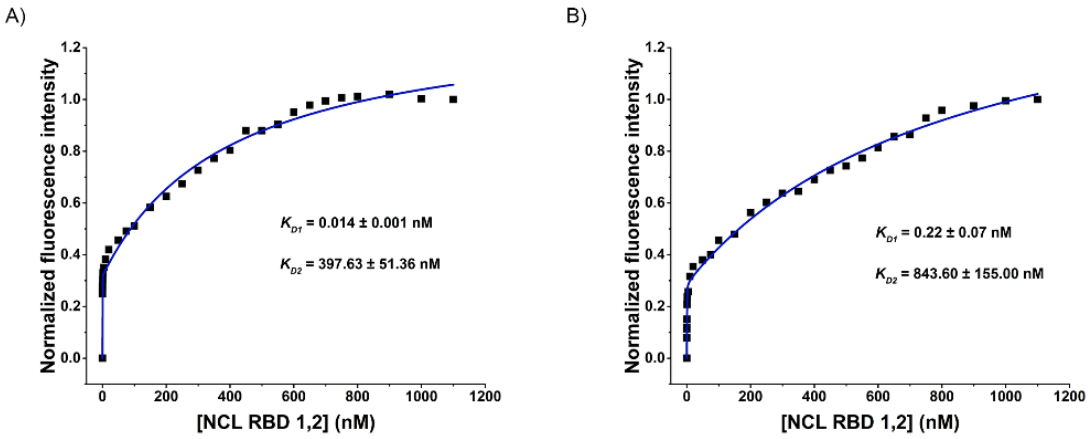

Figure 6.8. Plots of the fraction of bound NCL, fitted to the saturation binding equation. **(A)** pre-miR-92b rG4-Cy3.5 and **(B)** pre-miR-92b rG4-Cy3.5/C_8 complex at 1 μM with increasing concentrations of NCL RBD1,2 in 10 mM lithium cacodylate buffer pH 7.2, supplemented with 0.1 mM KCl.

Stability studies

The degradation of RNA oligonucleotides by nucleases in complex biological samples is a drawback associated with the biological environments. Therefore, the stability of pre-miR-92b rG4 and complexes pre-miR-92b rG4/C_8, pre-miR-92b rG4/NCL RBD1,2 and pre-miR-92b rG4/C_8/NCL RBD1,2 were assessed via 20% non-denaturing PAGE after 1, 3, 12, 24 and 48 h of incubation. For comparison purposes, the degradation pattern of the control RNA, which can form a hairpin structure, was evaluated in the presence of C_8, NCL RBD1,2, and C_8/NCL RBD1,2 **(Figure 6.9A)**. The electrophoretic profile showed significant degradation of the control after 3 h in the presence of 10% FBS. The degradation pattern was maintained even in the presence of C_8 and NCL RBD1,2. On the other hand, the electrophoretic profile of pre-miR-92b rG4 and pre-miR-92b rG4/NCL RBD1,2 in the absence and presence of C_8 showed that structures were resistant after 48 h, even in the presence of 10% FBS **(Figure 6.9B)**. The intense bands in polyacrylamide gels are ascribed to pre-miR-92b rG4, while the top band in the presence of 10% FBS without NCL RBD1,2 was attributed to RPMI 1640 medium, which indicates the possibility of circulating nucleic acids or molecular aggregates presented in FBS solution. Normalized band intensities were quantified by densitometry and the results are highlighted in **Figures 6.9C and 6.9D**. The results revealed that band intensities of pre-miR-92b rG4 were maintained after incubation at 37 °C, but presented a small decrease in the presence of 10% FBS, suggesting a small degradation after 48 h. As expected, band intensities of pre-miR-92b rG4 in the presence of C_8 hypothesize a protector effect of the ligand, since a small difference is observed at 0 and 48 h of incubation with 10% FBS. All experiments were conducted at 37 °C, revealing that pre-

miR-92b rG4 is stable at physiological temperature. The stability of pre-miR-92b rG4 to nucleases is important for microfluidic application using biological samples.

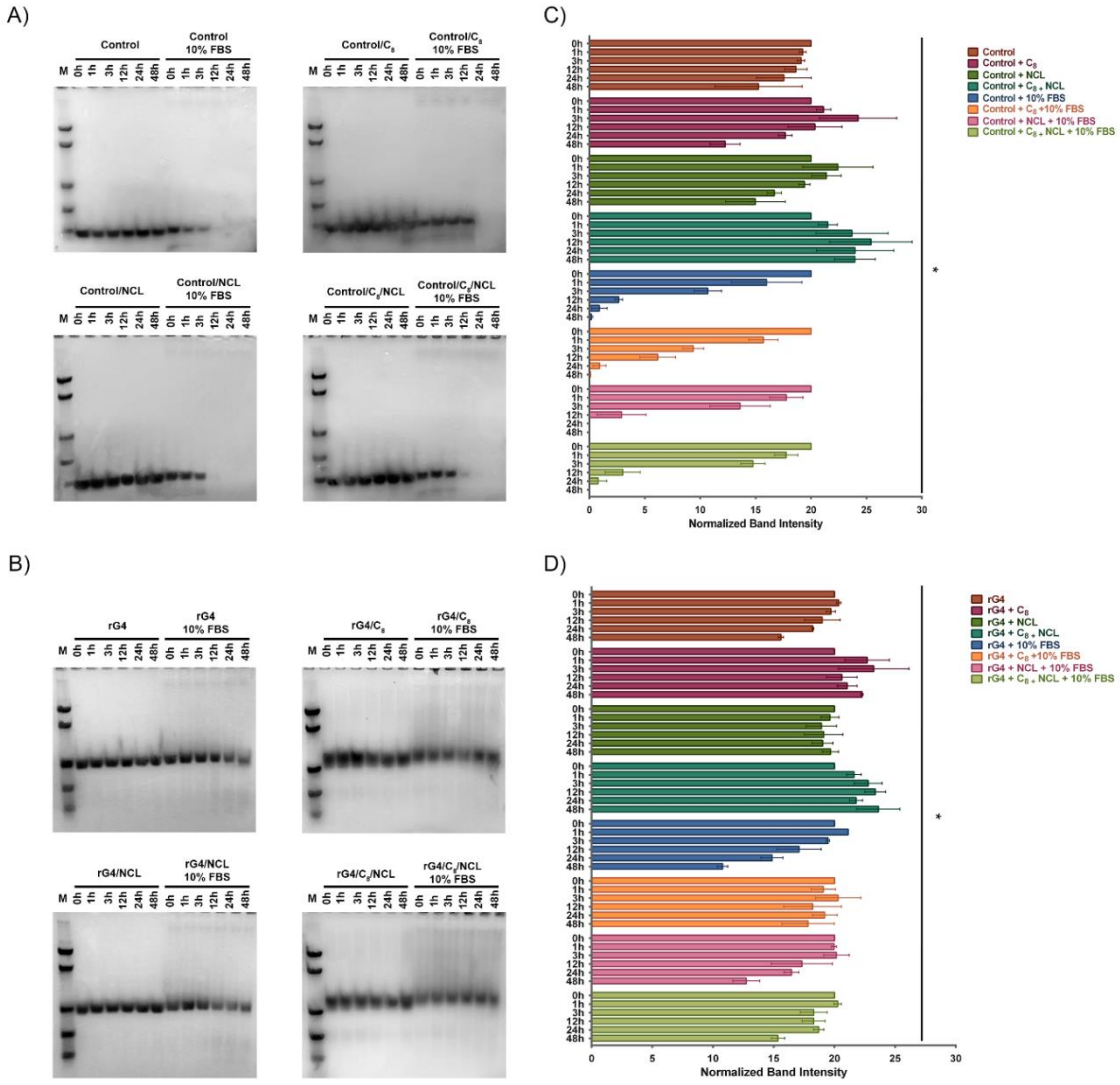

Figure 6.9. Stability assays of **(A)** control RNA sequence and **(B)** pre-miR-92b rG4 structure in 10 mM lithium cacodylate buffer pH 7.2, supplemented with 0.1 mM KCl at 1, 3, 12, 24 and 48 h. The experiments were performed at 37 °C in absence and presence of 10% FBS. The degradation pattern was monitored by 20% non-denaturing polyacrylamide gel supplemented with 10 mM KCl. Migration markers were single-stranded oligonucleotides (n = 9, 15, 21, 30, 60 and 90). Note that the bottom diffuse bands of migration marker correspond to bromophenol blue and xylene cyanol used to follow the migration. Normalized band intensities of **(C)** control RNA sequence and **(D)** pre-miR-92b rG4 structure in presence of C_8, NCL RBD1,2 and both were analysed by densitometry. Three independent experiments were conducted. Data are shown as mean with standard error of mean. A one-way analysis of variance was performed to evaluate the significance of the differences between different groups (the asterisks indicate $p < 0.05$).

NCL detection by microfluidic assays

After demonstrating the stability, binding affinity, and nuclease resistance of pre-miR-92b rG4 complexes, we developed a microfluidic platform based on pre-miR-92b rG4 (acting as a probe) to detect NCL in PCa human plasma samples. The microfluidic assays were carried out with pre-miR-92b rG4 at 5 μM labelled with Cy3.5 and stabilized with 0.1 M equivalents of C_8. Briefly, CM Sephadex C-25 beads were packed in the microfluidic channel and the pre-miR-92b rG4-Cy3.5/C_8 complex was incubated with NCL RBD1,2 in a tube for 30 min. After that, the mixture was flowed through the channel and fluorescence intensity of Cy3.5 fluorophore recorded in an Axio Imager Z2 microscope using 579 nm as the excitation wavelength (BP 565/30) and collecting the emission at 591 nm (BP 620/60). The retention of NCL RBD1,2 onto the beads seems to be due to electrostatic interactions between pre-miR-92b rG4-Cy3.5/C_8 complex and CM Sephadex C-25 beads. The detection of NCL RBD1,2 from a purified sample was first tested in buffer with a concentration range between 0 and 2.5 M equivalents. A significant fluorescence signal was observed for concentrations of protein higher than 0.25 M equivalents, as seen in **Figure 6.10**. However, after addition of 0.5 M equivalents of NCL RBD1,2, the relative fluorescence intensity signal seems to reach the maximum intensity being constant at 0.5, 1 and 2.5 M equivalents. Probably, the saturated behaviour of fluorescence signal is due to CM Sephadex C-25 beads binding capacity, which varies according to a specific protein. Since selectivity is an essential parameter for consistent microfluidic devices, PSA and HSA were used to assess if the molecular recognition probe was selective in the presence of other proteins. HSA was used as control since is the most abundant protein in human plasma and its low levels are directly linked to clinical characteristics and prognostic features associated with PCa [30]. The microfluidic assay was performed using 1 M equivalent of the control proteins, and the resulting fluorescence signal for both proteins was compared with that obtained for blank, as shown in **Figure 6.10**. The fluorescence signal is significantly higher towards NCL RBD1,2 comparing to PSA, HSA and blank, in which the signal was negligible.

The next experiments were focused on evaluating the ability of the microfluidic platform to detect NCL RBD1,2 in PCa human plasma. Briefly, the plasma was diluted 10 times and non-spiked and spiked NCL RBD1,2 were flowed with pre-miR-92b rG4-Cy3.5/C_8 in the microfluidic platform. The results showed a significant difference in the relative fluorescence intensities between the plasma samples with and without NCL RBD1,2, being negligible for plasma without NCL RBD1,2. However, the relative fluorescence intensity of the plasma sample spiked with NCL RBD1,2 is lower than the observed for

NCL RBD1,2 in buffer. This decrease is possibly due to human plasma composition that can promote non-specific interactions with the microfluidic surface.

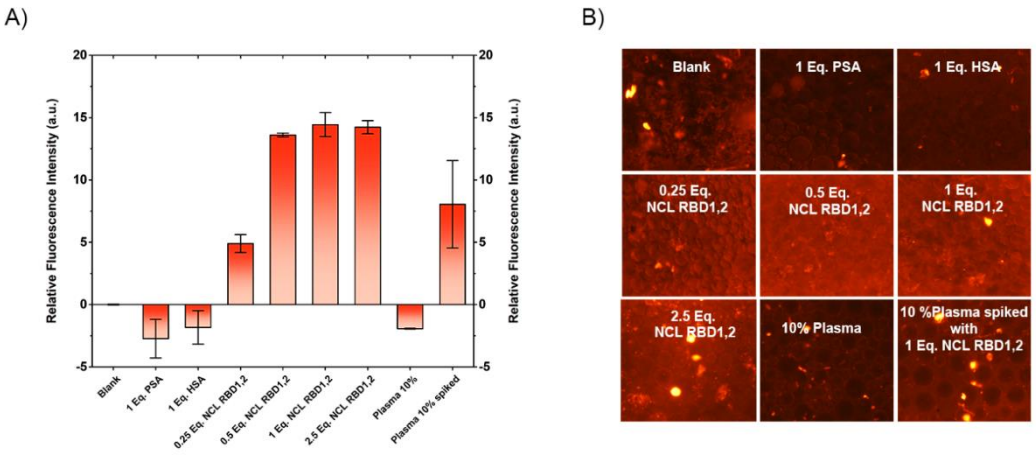

Figure 6.10. Microfluidic assay studies for NCL RBD1,2 measurement with pre-miR-92b rG4 labelled with Cy3.5 and complexed with C_8. **(A)** Relative fluorescence intensity of the pre-miR-92b rG4-Cy3.5/C_8/NCL RBD1,2 complex and control proteins PSA and HSA. **(B)** Fluorescence images acquired with an exposure time of 2 s in an Axio Imager Z2 microscope and processed in Zeiss Zen Software (SP2, 2010).

Discussion

The human precursor pre-miR-92b is 96 nucleotides long and can adopt a stem-loop or a rG4 structure [16]. In our recent work, we showed that rG4 sequence found in pre-miR-149 act as a supramolecular carrier of anticancer drugs for prostate cancer cells (Please see Chapter 8) [15]. Sequences with the ability to form stable parallel rG4s are attractive NCL binders [15]. Therefore, in addition to being used as drug delivery carriers via binding to cell surface NCL, rG4s could be also used as molecular recognition probes to detect NCL [31], [32].

In the present study, we demonstrated the formation of a single dominant pre-miR-92b parallel rG4 conformation in the presence of 0.1 mM K^+, which is an essential feature for NCL complexation, since NCL binds and recognizes with high affinity parallel G4s [33]. Indeed, among other advantages over DNA G4 congeners, the rG4s allows additional hydrogen bonds and preferentially adopt a parallel conformation due to the presence of the 2′-OH group of the ribose sugar [34]. The results of CD melting experiments suggest multiple rG4 conformations or aggregation in a range between 1 and 100 mM KCl. Similar results were demonstrated by Arachchilage et al. for pre-miRNA-92b full sequence, in which the stem-loop coexists with the G4 structure and two melting points were observed in the presence of 5 mM KCl [16]. In their study, the first

melting point was attributed to the rG4, while the second melting point was ascribed to the stem-loop structure [16].

The stabilization through ligand binding is required to maintain the folded state of the pre-miR-92b rG4 structure. Until now, several small molecules able to interact with and stabilize G4 structures have been identified [35], [36]. Acridine orange derivatives C_8, C_5 and C_8-NH_2 were previously described as G4 binders and stabilizers [15], [37]. The CD melting experiments showed the ability of C_5, C_5-NH_2 and C_8 to bind and stabilize the pre-miR-92b rG4 in more than 30 °C. In the presence of the ligands, the T_m increases, but surprisingly the CD melting curves only depicted one melting point, suggesting a lock in a single dominant conformation of pre-miR-92b rG4.

In light of these observations, we have employed ^{1}H NMR spectroscopy to investigate the formation of rG4 conformations. G4 structures exhibit characteristic resonances for the imino protons, whose presence may be used to confirm the existence of the rG4 conformation in solution [38], [39]. The ^{1}H NMR spectra corroborated the previous observations in CD experiments and non-denaturing PAGE, revealing more than one conformation in solution, especially in the presence of KCl higher than 1 mM. The presence of multiple rG4 conformations formed by a single putative rG4 sequence can be confirmed by the guanine imino proton peaks [27]. The effect of temperature in the equilibrium towards a specific rG4 conformation showed that the rG4 formed around 11.5 ppm is more stable than the conformation formed around 10.5 ppm. Overall, we observed a strong dependence of pre-miR-92b rG4 structure and conformation on K^+ concentration and temperature [27], [40], [41].

The predicted structure of the complex formed between rG4 and NCL RBD1,2 represents a paradigm in the understanding of how rG4s binds to NCL at atomic level. MD simulations showed a binding pocket between two N-terminal RBDs 1 and 2 of NCL, where pre-miR-92b rG4 binds to positively charged amino acid residues (Lys 85 and Arg 158) by hydrogen bonding. Generally, positively charged amino acids, lysine and arginine, mediate most of the protein-nucleic acid contacts, while the hydrophobic residues mediate the remaining interactions [42]. This binding pocket was also previously described as the preferential binding site of RNA on the NMR molecular structure of the two N-terminal RBDs 1 and 2 of NCL with the pre-ribosomal RNA target [43].

Using FRET-melting, we showed that NCL RBD1,2 increases the stability of the pre-miR-92b rG4 by 8.4 °C and the C_8 ligand by 6.3 °C. On the other hand, the ligands C_5, C_5-

NH_2 and C_8-NH_2 showed a modest effect on stability. In light of these results, we hypothesize that the increase in the length of the side chain in C_8 allows interactions between the ligand and the grooves/loops and/or phosphate backbone of the pre-miR-92b rG4 structure by H-bonding through the amide groups [44]. Furthermore, when used together, we observed a synergistic effect of C_8 and NCL RBD1,2 of 22 °C on the thermal stability of the pre-miR-92b rG4. In fact, acridine derivatives could interact with proteins directly or as a constituent of the nucleic acid-acridine complex [45]. Taking this into account, we suggest that this noteworthy synergistic effect could be attributed to an additional stabilization of the protein by C_8.

The compound PhenDC3 was previously described as a G4-specific ligand with end stacking binding mode and with *in vivo* biological effects [46]. In this work, the stabilization and selectivity of acridine orange ligands and PhenDC3 were assessed by FRET-melting towards pre-miR-92b rG4 and duplex DNA. C_8 showed the highest thermal stabilization towards pre-miR-92b rG4, while PhenDC3 was a weaker stabilizer. Moreover, PhenDC3 disrupts and prevents the binding of both DNA and RNA G4s to NCL [47], [48].

Altogether, these evidences suggest that C_8 can be used to stabilize the G4 in pre-miR-92b and do not affect the binding of pre-miR-92b rG4 to NCL RBD1,2. Furthermore, PAGE results suggest the formation of the complex rG4/C_8/NCL RBD1,2.

The assessment of binding characteristics of G4s towards their targets is of utmost importance in the selection of suitable molecular recognition probes. Herein, two K_D values were determined for pre-miR-92b and pre-miR-92b/C_8 complex interaction with NCL RBD1,2, which is consistent with a two-site binding model. As previously described for other G4 structures targeting NCL, the data had shown K_D^1 values in the low picomolar range, suggesting an overestimation of the affinity values, probably due to the use of labelled oligonucleotides [37]. On the other hand, K_D^2 values are in the nanomolar range, in agreement with those previously reported for rG4 sequences (Please see Chapter 8, page 217, figure 8.3) [15].

The present study shows that pre-miR-92b rG4 demonstrates high stability and nuclease resistance, which increases when pre-miR-92b rG4 is complexed with C_8. In fact, rG4s display higher stability and resistance to nucleases when compared with their DNA G4 congeners [49]. Furthermore, it has been described that G4 ligands increase the stability and resistance of DNA and RNA G4s in harsh environmental conditions [50].

Since the pre-miR-92b rG4 sequence can form a stable and resistant parallel G4, its applicability to recognize NCL in microfluidic devices is of utmost importance concerning the advantages of those systems [13], [51]. In this study, we have successfully detected NCL with selectivity in purified and complex biological samples, such as PCa human plasma, using pre-miR-92b rG4 stabilized with C_8 as a G4 ligand. Other approaches also reported fluorescence-based methods able to detect specific proteins. Hamaguchi and collaborators described molecular beacon aptamers, which underwent a transition from a stem loop to a compact G4 in the presence of a target protein [52]. These protein-induced changes in aptamer conformation resulted in a change of the fluorescence signal change [52]. Heyduk and collaborators also reported fluorescence based methods able to detect specific proteins in purified and complex biological samples [53], [54].

In conclusion, the results obtained with pre-miR-92b rG4, as a molecular recognition probe, highlight the opportunity to use rG4s to detect NCL RBD1,2 in purified and plasma samples. Our study showed the potential of using the microfluidic assay to detect NCL RBD1,2; however future improvements are necessary to use full NCL.

Acknowledgments

This work was supported by MIT Portugal FCT project BIODEVICE ref. MIT-EXPL/BIO/0008/2017, PESSOA program ref. 5079 and project “Projeto de Investigação Exploratória” ref. IF/00959/2015 entitled “NCL targeting by G-quadruplex aptamers for cervical cancer therapy” financed by Fundo Social Europeu e Programa Operacional Potencial Humano. Thanks are due to FCT/MCT for the financial support to CICS-UBI UIDB/00709/2020 research unit and to the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016), through national founds and, where applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and PIDDAC. Tiago Santos acknowledges Fundação para a Ciência e Tecnologia (FCT) for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013). A. Miranda acknowledges the fellowship grant from PTNMR project, ref. PINFRA/22161/2016-B4. The authors acknowledge University Hospital Center Cova da Beira (study number 93/2018) and IPO Coimbra (study number 31/PI/2019) for providing human blood samples according with the protocols approved by the Ethicals Committee. The authors acknowledge V. Chu and J. P. Conde for the material provided to perform the microfluidic assays.

References

1 Sharma, S. *et al.* (2017) Prostate cancer diagnostics: Clinical challenges and the ongoing need for disruptive and effective diagnostic tools. *Biotechnol. Adv.* 35, 135–149

2 Roobol, M.J. and Carlsson, S. V. (2013) Risk stratification in prostate cancer screening. *Nat. Rev. Urol.* 10, 38–48

3 Saini, S. (2016) PSA and beyond: alternative prostate cancer biomarkers. *Cell. Oncol.* 39, 97–106

4 Wadhwa, K. *et al.* (2015) The post PSA era: new developments in biomarkers, imaging and biopsy techniques in prostate cancer detection. *Trends Urol. Men's Heal.* 6, 25–29

5 Abdelmohsen, K. and Gorospe, M. (2012) RNA-binding protein nucleolin in disease. *RNA Biol.* 9, 799–808

6 Tate, A. *et al.* (2006) Met-independent Hepatocyte Growth Factor-mediated regulation of cell adhesion in human prostate cancer cells. *BMC Cancer* 6, 197

7 Mongelard, F. and Bouvet, P. (2007) Nucleolin: a multiFACeTed protein. *Trends Cell Biol.* 17, 80–86

8 Pichiorri, F. *et al.* (2013) *In vivo* NCL targeting affects breast cancer aggressiveness through miRNA regulation. *J. Exp. Med.* 210, 951–968

9 Chalfin, H.J. *et al.* (2017) Nucleolin Staining May Aid in the Identification of Circulating Prostate Cancer Cells. *Clin. Genitourin. Cancer* 15, e477–e481

10 Maremanda, N.G. *et al.* (2015) Quick chip assay using locked nucleic acid modified epithelial cell adhesion molecule and nucleolin aptamers for the capture of circulating tumor cells. *Biomicrofluidics* 9, 054110

11 Jolly, P. *et al.* (2016) DNA aptamer-based sandwich microfluidic assays for dual quantification and multi-glycan profiling of cancer biomarkers. *Biosens. Bioelectron.* 79, 313–319

12 Madaboosi, N. *et al.* (2015) A microfluidic immunoassay platform for the detection of free prostate specific antigen: A systematic and quantitative approach. *Analyst* 140, 4423–4433

13 Xu, Y. *et al.* (2010) Review: Aptamers in microfluidic chips. *Anal. Chim. Acta* 683, 12–20

14 Khan, N.I. and Song, E. (2020) Lab-on-a-chip systems for aptamer-based biosensing. *Micromachines* 11, 1–30

15 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

16 Mirihana Arachchilage, G. *et al.* (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* 22, 262–272

17 Mirihana Arachchilage, G. *et al.* (2018) Targeting of G-Quadruplex Harboring Pre-miRNA 92b by LNA Rescues PTEN Expression in NSCL Cancer Cells. *ACS Chem. Biol.* 13, 909–914

18 Wang, K. *et al.* (2013) MiR-92b controls glioma proliferation and invasion through regulating Wnt/beta-catenin signaling via Nemo-like kinase. *Neuro. Oncol.* 15, 578–588

19 Lei, L. *et al.* (2014) Inhibition of miR-92b suppresses nonsmall cell lung cancer cells growth and motility by targeting RECK. *Mol. Cell. Biochem.* 387, 171–176

20 Liu, Z. *et al.* (2015) MicroRNA-92b promotes tumor growth and activation of NF-κB signaling via regulation of NLK in oral squamous cell carcinoma. *Oncol. Rep.* 34, 2961–2968

21 Valentino, A. *et al.* (2017) Exosomal microRNAs in liquid biopsies: future biomarkers for prostate cancer. *Clin. Transl. Oncol.* 19, 651–657

22 Pereira, E. *et al.* (2017) Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA. *Sci. Rep.* 7, 42544

23 Carvalho, J. *et al.* (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286

24 Lyons, S.M. *et al.* (2017) Identification of functional tetramolecular RNA G-quadruplexes derived from transfer RNAs. *Nat. Commun.* 8, 1127

25 Pinto, I.F. *et al.* (2017) The application of microbeads to microfluidic systems for enhanced detection and purification of biomolecules. *Methods* 116, 112–124

26 Vorlíčková, M. *et al.* (2012) Circular dichroism and guanine quadruplexes. *Methods* 57, 64–75

27 Adrian, M. *et al.* (2012) NMR spectroscopy of G-quadruplexes. *Methods* 57, 11–24

28 Tosoni, E. *et al.* (2015) Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* 43, 8884–8897

29 González, V. and Hurley, L.H. (2010) The C-terminus of nucleolin promotes the formation of the c-MYC G-quadruplex and inhibits c-MYC promoter activity. *Biochemistry* 49, 9706–9714

30 Spiess, P.E. *et al.* (2007) Treatment outcomes of small cell carcinoma of the prostate. *Cancer* 110, 1729–1737

31 Li, H. *et al.* (2016) Aptamer-based microcantilever biosensor for ultrasensitive detection of tumor marker nucleolin. *Talanta* 146, 727–731

32 Gao, W. *et al.* (2020) Electrochemiluminescence Biosensor for Nucleolin Imaging in a Single Tumor Cell Combined with Synergetic Therapy of Tumor. *ACS Sensors* 5, 1216–1222

33 Lago, S. *et al.* (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1371–1381

34 Zaccaria, F. and Fonseca Guerra, C. (2018) RNA versus DNA G-Quadruplex: The Origin of Increased Stability. *Chem. - A Eur. J.* 24, 16315–16322

35 Folini, M. *et al.* (2010) Remarkable interference with telomeric function by a G-quadruplex selective bisantrene regioisomer. *Biochem. Pharmacol.* 79, 1781–1790

36 Gunaratnam, M. *et al.* (2009) G-quadruplex compounds and cis-platin act synergistically to inhibit cancer cell growth *in vitro* and *in vivo*. *Biochem. Pharmacol.* 78, 115–122

37 Carvalho, J. *et al.* (2019) Aptamer-based Targeted Delivery of a G-quadruplex Ligand in Cervical Cancer Cells. *Sci. Rep.* 9, 7945

38 Bugaut, A. *et al.* (2012) An RNA hairpin to g-quadruplex conformational transition. *J. Am. Chem. Soc.* 134, 19953–19956

39 Ferreira, J. *et al.* (2017) Naphthalene amine support for G-quadruplex isolation. *Analyst* 142, 2982–2994

40 Lim, K.W. *et al.* (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J. Am. Chem. Soc.* 132, 12331–12342

41 Phan, A.T. and Patel, D.J. (2003) Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *J. Am. Chem. Soc.* 125, 15021–15027

42 Hoffman, M.M. *et al.* (2004) AANT: The Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res.* 32, D174–D181

43 Johansson, C. *et al.* (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *J. Mol. Biol.* 337, 799–816

44 Carvalho, J. *et al.* (2018) Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex. *Biochimie* 144, 144–152

45 Ježek, J. *et al.* (2017) Interactions with proteins. In *Progress in Drug Research* 72pp. 73–97, Springer International Publishing

46 Chung, W.J. *et al.* (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew. Chemie - Int. Ed.* 53, 999–1002

47 Lista, M.J. *et al.* (2017) Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat. Commun.* 8, 16043

48 Saha, A. *et al.* (2020) Nucleolin discriminates drastically between long-loop and short-loop quadruplexes. *Biochemistry* 59, 1261–1272

49 Song, J. *et al.* (2016) RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* 4, e1244031

50 Gao, S. *et al.* (2016) Post-SELEX optimization of aptamers. *Anal. Bioanal. Chem.* 408, 4567–4573

51 Fraser, L.A. *et al.* (2019) Microfluidic Technology for Nucleic Acid Aptamer Evolution and Application. *Adv. Biosyst.* 3, e1900012

52 Hamaguchi, N. *et al.* (2001) Aptamer beacons for the direct detection of proteins. *Anal. Biochem.* 294, 126–131

53 Ranallo, S. *et al.* (2015) A Modular, DNA-Based Beacon for Single-Step Fluorescence Detection of Antibodies and Other Proteins. *Angew. Chemie - Int. Ed.* 54, 13214–13218

54 Heyduk, E. and Heyduk, T. (2005) Nucleic acid-based fluorescence sensors for detecting proteins. *Anal. Chem.* 77, 1147–1156

Chapter 7

Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics

This chapter was published in:
Carvalho, J.; **Santos, T.**; Carrilho, R.; Sousa, F.; Salgado, G.F.; Queiroz, J.A. and Cruz, C. (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286. DOI: 10.1080/07391102.2019.1632743 – IF: 3.392

Chapter overview

After the successful characterization of the G4s in pre-miRNAs let 7e and 92b, we move forward to unveil the G4 structure adopted by pre-miRNA 149 and its interaction with ligands. In this chapter, using docking and molecular dynamics simulations, we studied the interaction modes of six well-known ligands with the predicted G4 structure adopted by pre-miRNA 149. MD simulations showed that acridine orange derivatives and PhenDC3 interact with the G4 structure via stacking interactions with the top and bottom G-quartets. Furthermore, these ligands showed an entropic contribution upon ligand binding, which may be due to the displacement of water molecules from the rG4 surface. Overall, this first approach is important for rationalizing the following experimental methods employed in the further chapters.

Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics

Abstract

Using a molecular dynamics approach, the study of the interaction between six different known ligands and a predicted pre-miRNA 149 RNA G-quadruplex (rG4) structure is reported. The stabilization of rG4 structures formed within the pre-miRNA stem-loop regions using small ligands is an attractive anticancer strategy. Particularly, miRNA-149 is upregulated in a variety of cancers such as prostate cancer and is therefore a potential target for drug development. The results show that ligands C_8 and $PhenDC_3$ interact with the rG4 structure via stacking interactions with the end G-quartets. Ligands [16]$phenN_2$, [32]$phen_2N_4$ and pyridostatin on the other hand bind the loops/groove interface of the rG4 being H-bonding and electrostatic interactions the driving force of the interaction. The C_8 precursor, C_8-NH_2, emphasizes the structural nuances of the rG4 short loops as the lack of a large terminal aromatic moiety produced a mixed stacking-groove binding mode. Overall, this study may help the design of specific ligands for pre-miRNA rG4 towards anticancer therapeutics development.

Keywords

miRNA 149; RNA G-quadruplex; Molecular Dynamics; ligand screening; thermodynamics

Introduction

MicroRNAs (miRNAs) consist of short noncoding RNA sequences that play an important role in regulating gene expression at post-transcriptional level, as well as maintaining the homeostasis of the cell. These act by binding to the 3′-untranslated region (3′-UTR) of specific mRNAs, resulting in translational inhibition and/or mRNA degradation [1]. 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 contain a stem-loop

region which is recognized by Dicer, a cytoplasmatic enzyme that cleaves the RNA sequence [2].

The deregulation of miRNAs is linked with several pathologies and abnormalities such as cancer [3]. Due to their important roles in tumorigenesis and tumor progression, premature and mature miRNAs are attractive targets for anticancer drug development. Among these, miRNA 149 was found to be significantly dysregulated in various cancer types, such as lung, gastric, breast and prostate cancers [4]. Particularly in prostate cancer, it has been shown that the inhibition of the upregulated miRNA 149 leads to a reduction in proliferation and increase in apoptosis of cancer cells from a prostate cancer cell line [5].

Recently, Balasubramanian group revealed that the human precursor of miRNA 149 (pre-miRNA 149) is able to form a G-quadruplex (G4) structure [6]. G4 structures are characterized by the planar square arrangement of guanine residues (G-quartets) connected by Hoogsteen hydrogen bonds [7–14]. Stacking of at least two G-quartets forms the G4 structure, linked by loops of different lengths and stabilized through π-π interactions between the aromatic systems of G-quartets and a central monovalent cation (such as K^+ or Na^+). RNA G4s (rG4s) were already shown to form under physiological conditions inside the cells [15] and are comparatively more stable than their DNA counterparts as they adopt almost exclusively a parallel conformation due to the 2'-hydroxyl group [16]. The formation of these non-canonical structures at RNA level is involved in the regulation of several key cellular processes such as telomerase function, pre-mRNA processing, mRNA turnover and targeting, and translation [16]. Particularly, in pre-miRNAs the formation of a G4 motif in the stem-loop region interferes with Dicer activity decreasing mature miRNA production [2,6].

On this basis, targeting pre-miRNA 149 G4 structure with ligands that stabilize the rG4 leading to the inhibition of Dicer processing activity may represent a promising anticancer strategy, as reported for other pre-miRNAs [2,17]. Using an *in silico* approach we studied the interaction between a set of known G4 ligands and the predicted model of pre-miRNA 149 rG4 structure. The ligands used in this study were phenanthrolines $PhenDC_3$ [18], [16]$phenN_2$ and [32]$phen_2N_4$ [19], acridines C_8 and C_8-NH_2 [20], pyridine pyridostatin (PDS) **(Figure 7.1)** [21].

Computer simulations can provide valuable insights on ligand-G4 complexes, and has been successfully employed in the study of several DNA and RNA G4 structures [22–26]. As no NMR or X-ray derived model of pre-miRNA 149 rG4 structure is available, we built

the rG4 model using a parallel-stranded telomere G4 as template. Then a docking-based virtual screening approach was employed to study rG4-ligand interaction. The more promising conformers were further calculated by Molecular Dynamics (MD) simulations and the binding free energy were estimated using both semi-empirical and empirical approaches. This work may aid the identification of potential pre-miRNA 149 maturation inhibitors, by understanding the interactions between ligands and the rG4 structure, which could also be useful in identifying promising scaffolds for the design of novel rG4 ligands.

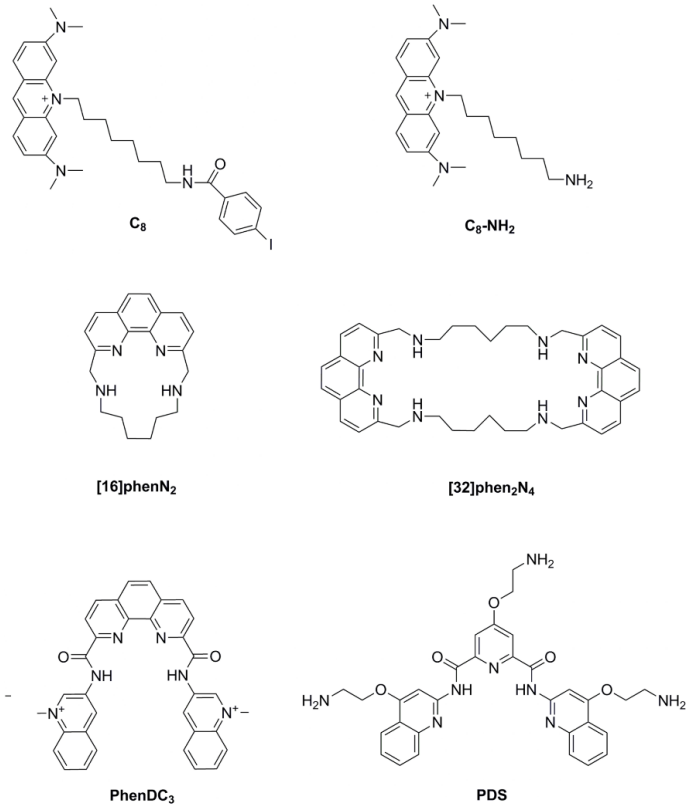

Figure 7.1. Structure of the G4 ligands used in this study.

Methods

pre-miR-149 G4 model preparation

A predicted model of pre-miR-149 G4 was built based on the parallel-stranded telomeric G4 AGGG(TTAGGG)$_3$ (PDB: 4G0F). The DNA structure 4G0F was chosen due to its sequence homology with pre-miR-149 G4 and the fact that it presents a parallel G4 conformation as expected for pre-miR-149 G4 [6]. First, as shown in **Table 7.1**, the 4G0F telomeric DNA sequence was modified so that it would correspond to the nucleotide

sequence of pre-miR-149 $(GGGA)_3CGGG$, using Swiss-PDB Viewer mutation tool. Thymine residues in positions 5, 6, 11, 12 and 17, and the adenine residue in position 1 were deleted (changes highlighted in red, **Table 7.1**). The thymine residue in position 18 was mutated to adenine, while the adenine residue in position 19 was changed to cytosine (changes highlighted in blue, **Table 7.1**). The changes were performed in the loop residues, in order to maintain the G-tetrads integrity. Finally, the PDB structure was manually edited at the 2' position of each sugar base to convert deoxyribose into ribose by modifying 2'-H to 2'-OH, thus turning the sequence a RNA oligonucleotide. The modifications were performed using Swiss-PDB Viewer.

Table 7.1. Oligonucleotide sequences of the human telomeric G4 and pre-miR-149 rG4.

Name	**Sequence (5'-3')** [a]
Tel22	A GGG TT A GGG TT A GGG TT A GGG
pre-miR-149	GGG A GGG A GGG A C GGG
[a] Base deletion is highlighted in red, base mutation is highlighted in blue.	

As the 4G0F was heavily modified to create the pre-miR-149 G4 model, the built structure was optimized for the further experiments by running fully solvated molecular dynamics. To remove unfavorable van der Waals interactions and probe for incorrect atom names and connectivity, the model was re-optimized using GROMACS 2016.3 with the following parameters [27]. The rG4 model was initially centered in a cubic periodic box, with a solute-box distance of 1 nm and filled with TIP3P water molecules and K^+ atoms. 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. A 100 ps equilibration was then 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. Finally, 20 ns of MD simulations were carried out. The final snapshot of the 20 ns simulation was used as pre-miR-149 G4 optimized model for further experiments. All molecular images were rendered using UCSF Chimera 1.11.2 [28].

Molecular docking

The minimized model of pre-miR-149 G4 was used as target for molecular docking studies. Both the rG4 structure and the ligand were optimized for docking using Dock

Prep tool of Chimera 1.11.2. After assigning polar hydrogens and Gasteiger charges, docking simulations were carried with AutoDock 4.0 program using Lamarckian genetic algorithm [29]. The size of the box was constrained to 100 × 100 × 100 Å along the x, y, and z axes, respectively, with a grid spacing of 0.375 Å. For each ligand, 25 runs were performed with an initial population of 150 random individuals, a maximum number of evaluations set to 2.5×10^7, the rate of mutation and crossover set to 0.02 and 0.8, respectively, and elitism value of 1. The ligands were allowed full flexibility, while the rG4 structure was kept rigid. The best conformers were selected based on the binding free energy and further processed with MD simulations.

Molecular dynamics

The best conformers from molecular docking studies underwent molecular dynamics simulations using GROMACS 2016.3. The acpype program [30] was used to generate the ligand's topology and parameters files in GROMACS-compatible format. The complexes were centered in an octahedral box and solvated with TIP3P water molecules and 13 K^+ atoms to neutralize the system. Explicit solvent simulations were performed at 300 K under the control of Berendsen thermostat for 49 ns followed by 1 ns isotropic constant-pressure simulation under the control of the Parrinello-Rahman barostat. The equilibration procedure was performed with bond-length and angle constraints on the complex. The Particle mesh Ewald (PME) method was used for calculating long range electrostatic interactions. A 1 nm cut-off was applied to short range Lennard-Jones interactions. Coordinates were collected in trajectory files every 10 ps. An energy minimization of 1000 steps using the steepest descent algorithm was followed by 20 ns and 100 ns of unrestrained MD simulations. MD representative structures were obtained by cluster analysis performed using the gmx cluster tool implemented in GROMACS. All molecular images were rendered using UCSF Chimera 1.11.2.

Thermodynamics calculations

The binding free energy of each complex was calculated by empirical MM/PBSA method [31]. MM/PBSA method on its hand computes the relative free energies of binding from a trajectory, to provide an average of energies. The change of free energy calculated with the python module *MmPbSaStat* [32] is described through the following equations:

$$\Delta G_{bind} = \Delta G_{total} - T\Delta S \qquad \text{(Equation 7.1)}$$

being:

$$\Delta G_{total} = G_{complex} - G_{G4} - G_{ligand} \quad \text{(Equation 7.2)}$$

$$= \Delta E_{MM} + \Delta G_{solv} \quad \text{(Equation 7.3)}$$

where:

$$\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{elec} \quad \text{(Equation 7.4)}$$

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{SASA} \quad \text{(Equation 7.5)}$$

where $G_{complex}$, G_{rec}, and G_{ligand} are the complex, G4, and ligand free energies, respectively. The polar contribution to solvation was calculated by solving the Poisson–Boltzmann (PB) equation with a grid size of 0.5 Å. The solute dielectric constant was set to 2, and 80 for surrounding water. The nonpolar solvation contribution was calculated under SASA model using the solvent accessible surface area with solvent-probe radius set to 1.4 Å. ΔS was estimated using GROMACS *gmx anaeig* tool, T was 300 K. The free energy of the RNA G4–ligand binding was calculated based on the full 20 ns MD simulations from at least three different experiments. The SASA calculations for rG4 over the full trajectory of the MD simulations were carried out using GROMACS *gmx sasa* tool. H-bonds were computed with *gmx hbond* tool and H-bond occupancy is defined as the fraction of conformations in which the given atom participates in a hydrogen bond.

Results and Discussion

Elucidating the structural basis for the recognition process between a ligand and its molecular target is of great importance in the rational drug design and effective rG4 ligand optimization. In this work we selected a set of 6 G4 ligands and probed their interaction with the predicted rG4 model of pre-miRNA 149. The study workflow was as follows: i) rG4 model construction; ii) molecular docking simulations of all ligands; iii) best conformers in terms of binding energy and visual inspection are further calculated with MD; iv) thermodynamics calculations of the MD trajectories.

pre-miRNA 149 rG4 model structural nuances

Since the solution structure of pre-miRNA 149 rG4 is not available, we built a predicted model based on the parallel-stranded telomeric G4 AGGG(TTAGGG)$_3$ (PDB: 4G0F). The last snapshot of the 20 ns MD simulation is shown on **Figure 7.2**. The rG4 model is

composed of three G-tetrads (G1-G5-G9-G14; G2-G6-G10-G15; G3-G7-G11-G16) interconnected by two single-nucleotide propeller-type loops (A4 and A7) and an AC loop, similarly to propeller type (also known as double chain reversal).

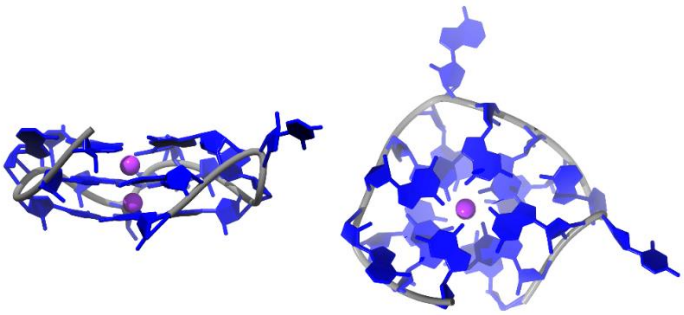

Figure 7.2. Predicted structural model of pre-miRNA 149 rG4 after 20 ns MD production. Side view (left) and top view (right) show a parallel stranded rG4 with propeller-type loops (depicted in gray). K^+ cations are depicted as purple spheres.

This structure is in agreement with the results from SHALiPE analysis reported by Balasubramanian group [6]. A8 and C13 appear pointed outward, while A4 and A12 form hydrogen bonds with the edges of the top tetrad, further stabilizing the rG4 structure as seen for other G4s [33]. The lack of any end-capping residues might favor stacking interactions with the top and bottom tetrads. The result showed that the rG4 structure was rather stable over 20 ns MD simulation as seen by the RMSD graph **(Figure 7.3)**, with two K^+ atoms stabilizing the G-quartets.

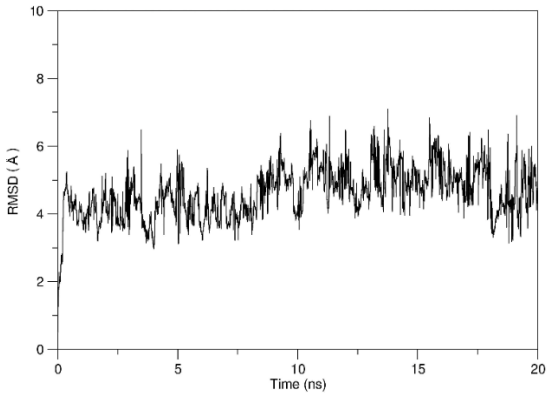

Figure 7.3. RMSD plot of the 20 ns simulation of pre-miRNA 149 G4 model.

The stability of the rG4 structure, particularly the G-tetrads, was further demonstrated by the analysis of the RMSF graphs **(Figures 7.4 and S7.1)**. As suggested by the per-nucleotide RMSF analysis **(Figure 7.4)** the guanines constituting the G-tetrads showed minor fluctuation when compared to the more flexible residues of the loops (adenines and cytosines). **Figure S7.1** shows the RMSF values for all atoms with a similar profile with stable guanine quartets. Additionally, the variation of the radius of gyration (R_g)

was also analyzed as a function of simulation time **(Figure S7.2)**. The average R_g of 11 Å (1.10 nm) is similar to that of other G4 structures reported in the literature [34] and the obtained value supports the compactness of the parallel G4 structure with short loops.

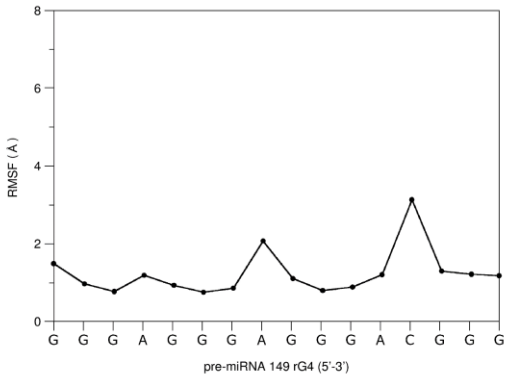

Figure 7.4. RMSF plot of each nucleotide in the pre-miRNA 149 G4 during the full 20 ns MD simulation.

Molecular docking of the set of chosen ligands

Following the model preparation, molecular docking simulations were undertaken using AutoDockTools software. The best binding structures based on the lowest energy of rG4-ligand complex are shown in **Figure 7.5** (the top 5 structures in terms of binding energy are depicted in SI, **Figures S7.3-S7.8**). The acridine orange derivative C_8 was docked on the top tetrad through stacking interactions between the acridine moiety and the guanine bases **(Figure 7.5A)**. The protonated nitrogen atom of the acridine moiety was placed near the central carbonyl groups of the G-tetrads which have negative electrostatic potential. The estimated free binding energy (ΔG) was -5.99 kcal mol^{-1}. Ligand C_8-NH_2 which is a precursor of C_8, lacking the iodobenzene group, presented a different binding mode **(Figure 7.5B)**. Similarly, to C_8, the acridine moiety stacks in one of the outer tetrads (bottom), but the alkylamide arm seems to interact with the A8 loop/groove interface. The estimated ΔG value was -4.88 kcal mol^{-1}. This different behavior when compared to C_8 may be due to the higher accessibility of NH_2 terminal moiety when compared to iodobenzene. Indeed, the pre-miRNA 149 G4 short loops may not create binding pockets able to accommodate the iodide atom. The difference in the ΔG values may be explained by the higher hydrophobicity of C_8 and the possibility of additional π-π interactions by the iodobenzene moiety. Several H-bonds were detected between the terminal NH_2 groups of the acridine moiety and the side arm. While the acridine orange derivatives both showed a preferential end-stacking mode of interaction, the phenanthroline derivatives [16]phenN_2 and [32]phen$_2N_4$ were docked onto the

loop/groove interface **(Figure 7.5C-D)**. The single-phenanthroline ligand [16]phenN_2 was docked in the two-nucleotide loop formed by A12 and C13 **(Figure 7.5C)**. On its hand, the di-phenanthroline ligand [32]phen$_2N_4$ was docked with one of the phenanthroline moieties in the groove formed by the AC loop and the second moiety in the groove formed by A8 loop **(Figure 7.5D)**. The protonable amine groups of the side arms seem to interact with the phosphate backbone of the rG4 structure, namely G10 and G11, probably by electrostatic interactions. Despite both ligands having aromatic moieties capable of stacking onto the large planar tetrads of the rG4 structure, this binding mode was not observed. This may be explained by the difficulty of AutoDock to simulate the ring flexibility of macrocyclic structures as previously reported [35]. Nonetheless, the estimated ΔG values for [16]phenN_2 and [32]phen$_2N_4$ were -7.81 and -11.07 kcal mol^{-1}, respectively. Regarding the bisquinolinium compound PhenDC$_3$ **(Figure 7.5E)**, the ligand was docked in the bottom G-tetrad of the rG4 structure, with one of the quinolinium moieties stacking with the guanine bases and the other interacting with the loop. Oddly, the phenanthroline moiety was not found to stack with the tetrad. This may be due to the full flexibility introduced into the bisquinolinium arms. The estimated ΔG value was -8.21 kcal mol^{-1}. Finally, PDS ligand was docked in the AC loop performing groove binding interactions primarily **(Figure 7.5F)**. The lack of a large central planar moiety and the existence of several protonable amine groups seem to favor such type of interactions with the loop/groove interface. Of all the ligands tested, PDS has the highest flexibility degree, which makes it capable of adapting to the dynamics of the rG4 structure. The estimated ΔG value was -9.53 kcal mol^{-1}.

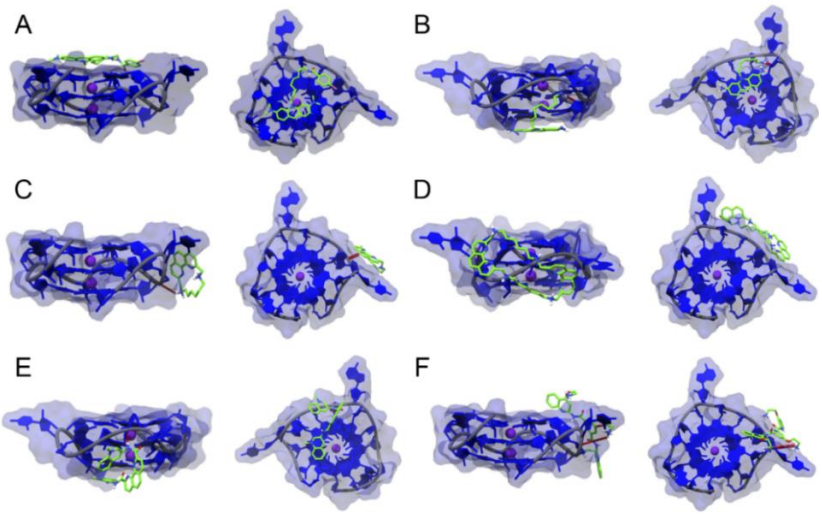

Figure 7.5. Selected binding poses in terms of lowest binding free energy of the complex formed by pre-miRNA 149 rG4 and **(A)** C_8, **(B)** C_8-NH_2, **(C)** [16]phenN_2, **(D)** [32]phen$_2N_4$, **(E)** PhenDC$_3$ and **(F)** PDS. G4 structure is depicted as a blue surface with K^+ cations depicted as purple spheres, while the ligand is represented in green. Hydrogen bonds are depicted as red lines.

Molecular Dynamics simulations

The lowest binding energy conformers abovementioned were further subjected to MD simulations. The conformational stability of the rG4-ligand complexes was assessed by measuring the RMSD during the simulations. The RMSD graphs of the complex, ligands and rG4 structure are shown in **Figure S7.9**. The analysis of the graphs indicates that all atoms were within 1 Å during the course of the MD simulations, and not more than 4 Å relatively to the initial conformation. Furthermore, extending the MD simulations until 100 ns did not render any significant alteration in the stability of the complexes **(Figure S7.8)**. C_8 complex remained stable through 10 ns of MD simulation being the main average structure similar to the initial conformer obtained from docking experiments **(Figures 7.6, S7.9A)**. π-π interactions with the top tetrad seem to be the principal binding mode, with two additional H-bonds formed between the ligand oxygen atom and G9-H22 and A12-H62. The alkylamide side arm relative position is similar to previously observed binding modes of C_8 and KRAS-22RT G4 [20]. As observed in **Figure S7.9A** the ligand RMSD fluctuates more comparatively to the rG4. The analysis of the MD trajectory revealed to be due to the torsion freedom of the alkylamide arm.

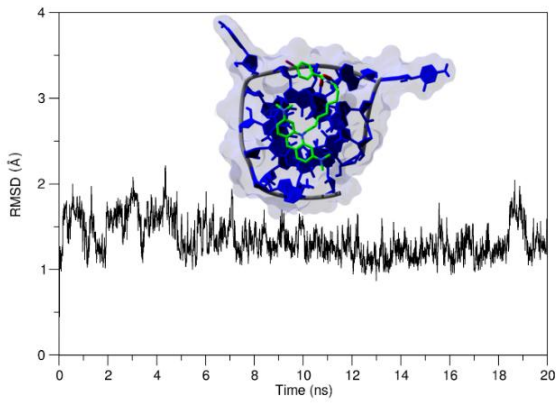

Figure 7.6. RMSD plot of the 20 ns simulation of pre-miRNA 149 - C_8 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines.

C_8-NH_2 on its hand, presented a similar binding mode to that of the docking conformer, with mixed π-π and groove binding interactions, however the alkylamide arm was placed in the A4 loop/groove interface **(Figure 7.7)**. The complex remained stable during 20 ns MD run, with a slight fluctuation of the rG4 RMSD around 10 ns **(Figure S7.9B)** which was found to be due to relaxation on the G-tetrads conformation and the loops bases outwards rotation. Two H-bonds were detected involving acridine orange NH_2 and G3-O2', and the alkylamine NH_2 and G5-O2'. The H-bond occupancy was 55.2 and 77.2% respectively, which indicates moderate bonding. Indeed, H-bond analysis shows that the

terminal NH_2 is also involved in strong H-bonding with A4-O2' (91.7% occupancy), which was not detected in the representative structure obtained from cluster analysis but highlights the flexibility of the alkylamide arm.

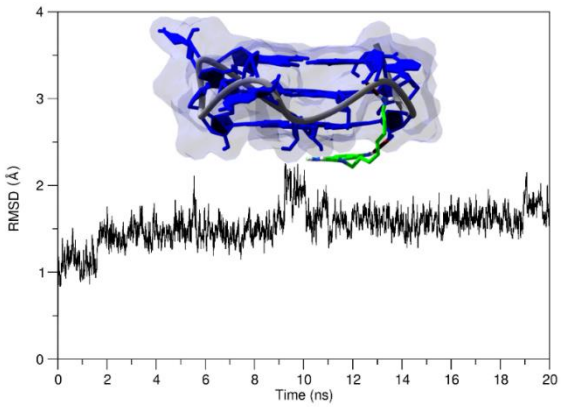

Figure 7.7. RMSD plot of the 20 ns simulation of pre-miRNA 149 - C_8-NH_2 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines.

Phenanthroline derivatives [16]phenN_2 and [32]phen$_2N_4$ showed also to be stably bound to the rG4 structure as seen by the RMSD graphs **(Figures 7.7, 7.8, S7.9C and S7.9D)**. [16]phenN_2 showed a preferential groove binding mode of interaction, establishing two H-bonds between the diamine arm's NH_2 and A12-O1P and C13-O1P phosphate atoms **(Figure 7.8)**. High occupancy percentages of 95.6 and 79.3% indicate that the H-bonds are strong. Similarly, [32]phen$_2N_4$ also interacts with the AC loop with two H-bonds between the side chain NH_2 and G10 and G11 O1P atoms, with 97.2 and 98.8% occupancies **(Figure 7.9)**. Different from the docking conformer, the second phenanthroline moiety was pushed outwards relatively to the rG4 structure. This may be due to the imperfect treatment of macrocycle conformation dynamics inherent to the methods used, which is still a major hurdle in the field [36].

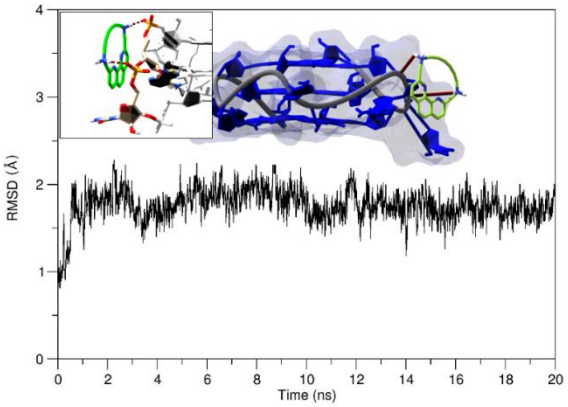

Figure 7.8. RMSD plot of the 20 ns simulation of pre-miRNA 149 - [16]phenN_2 complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. Inset, corner: zoomed view of the H-bonds with the rG4 phosphate backbone.

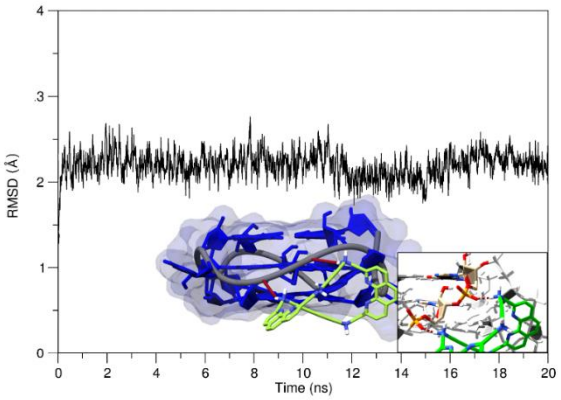

Figure 7.9. RMSD plot of the 20 ns simulation of pre-miRNA 149 - [32]phen$_2$N$_4$ complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. Inset, corner: zoomed view of the H-bonds with the rG4 phosphate backbone.

Regarding PhenDC$_3$ a major structural transition was observed relatively to the starting conformer until a stable complex was attained **(Figure S7.9E)**. Three major clusters were obtained from cluster analysis. Until around 3 ns, the ligand conformation is similar to that of the initial docking conformation, with one of the quinolinium units stacking on the lower tetrad with additional interactions with the loop/groove interface formed by A8 residue **(Figure 7.10)**. Four H-bonds were observed between the side chain O atoms and G6-H22 and G7-H21 (0.3 and 0.8% occupancy, respectively), the phenanthroline NH and G7-H22 (1.6% occupancy), and between the side chain NH and G9-O2' (32.1% occupancy). The residual occupancies confirm the instability and transiency of the complex at the beginning of the MD simulation. Between 3-8 ns, the ligand underwent a conformation transition into a fully stacking binding pose that maintained throughout the remaining MD simulation. During this period, the phenanthroline moiety changed position between residues on the lower tetrad. The final cluster representative structure shows the phenanthroline moiety and one of the quinolinium moieties interacting with the bottom tetrad through π-π stacking, being this binding mode in agreement with what has been reported for PhenDC$_3$ [37]. No H-bonds were observed, meaning π-π stacking is likely the major determinant of the interaction.

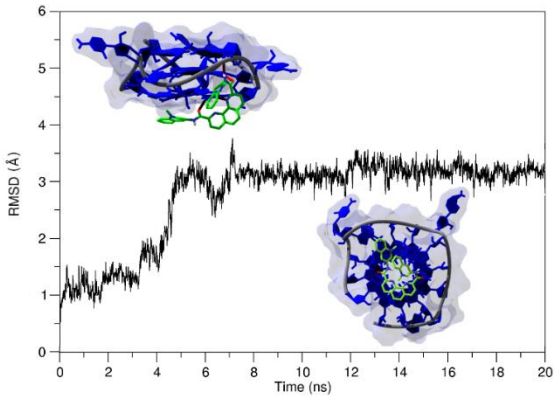

Figure 7.10. RMSD plot of the 20 ns simulation of pre-miRNA 149 – $PhenDC_3$ complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines.

Finally, PDS was stably bound to the AC loop as the initial conformer **(Figures 7.11, S7.9F)**. The ligand interacts with A12 through H-bonding between A12-N3 and PDS NH group (36.5% occupancy), and with C13 with H-bonds observed between C13-O1P and -O2 and ligand's NH and NH_2 groups, respectively (96 and 23.2% occupancy, respectively). Between 4-9 ns, a significant fluctuation on the RMSD of the ligand was observed **(Figure S7.9F)**. This was due to the rotation of the quinoline arms, oscillating between a coplanar state and a perpendicular state of the ligand ring systems. This was followed by the formation of different H-bonds with A12-O1P (44.1%), G11-O1P (14.9%) and A12-N3 (36.5%). The average structure of the major cluster found shows one of the quinoline rings stacking onto A12 while the pyridine ring interacts mostly with C13. Opposingly to the docking conformer, the fact that the quinoline ring systems are coplanar relatively to each other and perpendicular to the pyridine ring favors additional interactions and better fitting in the rG4 structure. In the case of PDS, a significantly different conformation was observed in the course of 100 ns MD simulations. As observed in **Figure S7.11**, the ligand is found stacking in the lower tetrad with additional H-bonds with the loop residues, namely between G9-O2P and G14-O4' and ligand's terminal NH_2 groups. Despite solution structures are still missing for PDS, due to the ligand design, both binding modes are feasible as it contains a potentially planar electron-rich aromatic surface and the ability to participate in H-bonding. Moreover, its rotatable bonds provide large flexibility, which makes PDS capable of adapting to the dynamism of the rG4 structure [38].

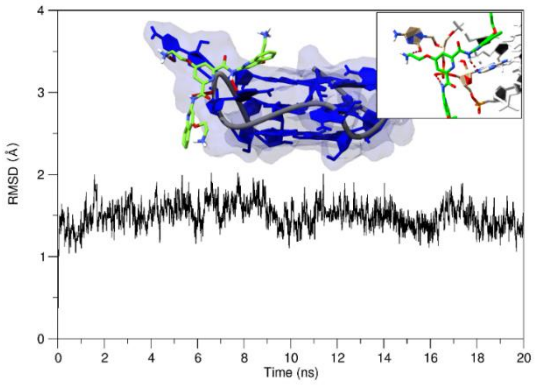

Figure 7.11. RMSD plot of the 20 ns simulation of pre-miRNA 149 - PDS complex. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines. Inset, corner: zoomed view of the H-bonds with the rG4 structure.

As observed for the RMSD plots of [32]phen$_2$N$_4$, PhenDC3 and PDS **(Figures S7.9D, S7.9E and S7.9F)**, the complex RMSD variation was higher than that of the rG4 structure. The higher RMSD values of the complex are mostly due to ligand atoms fluctuation at the beginning of MD simulations until the complex converges to a stable structure, as observed by the similar profile of both the complex and ligand RMSDs (black and green lines). The lack of significant RMSD variation towards the end of simulation after the initial fluctuation (all atoms were within 1 Å) suggests that indeed stable complexes were attained. This was further demonstrated by free energy of binding calculations (see below).

It should be emphasized that RMSD is a measure of similarity [39]. Therefore, the RMSD variation itself only suggests that these stable complexes attained after convergence are somehow different from the initial conformations, which can be perceived by comparing the initial docking conformers with the representative structures taken from MD simulations, particularly for PhenDC3 and PDS **(Figures 7.10 and 7.11)**.

Solvent accessible surface area determination

Following RMSD analysis, the available solvent accessible surface area (SASA) of the free rG4 and ligand-bound rG4 was determined **(Figure 7.12)**. The SASA value of a molecule gives a measure of its contact area with the solvent. Binding of ligands to the external G-quartets (or loops) of the rG4 structure reduces the total SASA of the complex. It has been shown that dehydration appears to stabilize G4 structures, opposingly to duplex DNA which is destabilized by dehydration [34,40]. Therefore, the thermodynamic stability increases if a given ligand minimizes the SASA value. Even though the quantitative relation between available surface area and free energy of binding is elusive,

the SASA can be used to compare different ligands in terms of their ability to minimize the SASA value, thus improving the rG4 structure stability [41]. The normalized frequency of SASA during the course of the simulations is shown in **Figure 7.12**. SASA value for free rG4 was 3589 Å^2 which is within the range 3500–4000 Å^2 observed for folded G4s.[34] Ligand C_8-NH_2 promoted a ΔSASA of -244 Å^2, demonstrating the lower SASA value among all the tested compounds. This may be due to the mixed binding mode demonstrated, with π-π stacking and groove binding interactions, thus minimizing the accessible surface of the rG4. C_8 on its hand, presented a lower ΔSASA of -143 Å^2 which may be due to π-π stacking alone being the preferred mode of binding. Surprisingly, [16]phenN_2 which was found to bind exclusively to the AC loop, promoted a ΔSASA of -191 Å^2. This may be a consequence of the loop residues reorientation upon ligand binding, minimizing the available accessible surface area [42]. [32]phen$_2$N$_4$ sits on the other end in terms of ΔSASA values with 119 Å^2, leading to an increase in the available surface area, possibly destabilizing the rG4 structure. PDS also increased the available surface area by 32 Å^2. As both ligands interact with the same loop, the ΔSASA observed may be due to the reorientation of A12 and C13 residues. PhenDC_3 promoted a ΔSASA of -103 Å^2 which is in agreement with the stacking of the ligand onto the bottom tetrad minimizing the accessible surface.

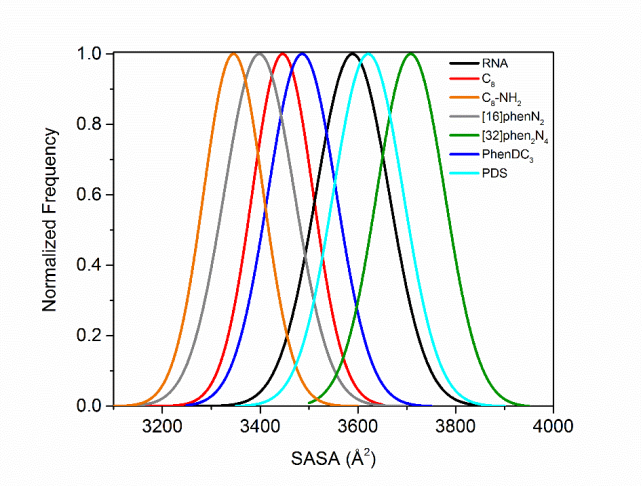

Figure 7.12. Normalized frequency data plots of SASA of the ligand-free pre-miRNA 149 rG4 and rG4-ligand complexes.

Free energy of binding calculations (MM/PBSA)

The thermodynamic parameters driving the interaction between the ligands and pre-miRNA G4 structure were determined using MM/PBSA calculations. All the ligands bind pre-miRNA 149 with overall favorable free energies (ΔG_{bind}) between -266.0 and - 422.3 kcal mol^{-1} (**Table 7.2**). End-stacking ligands C_8 and PhenDC_3 presented ΔG_{bind} values of

-335 and 401.5 kcal mol^{-1}, respectively. Both complexes have a large entropic contribution driving their formation, which can be attributed to the similar end-stacking binding mode and can be ascribed to the dehydration of the G-tetrads upon ligand binding. This is in agreement with the ΔSASA values obtained for both ligands. The binding mode of PhenDC3 for pre-miRNA 149 G4 is similar to the complex with c-MYC G4 determined in solution by NMR [37]. In this study, PhenDC3 binds very tightly DNA and RNA G4s structures (the K_D is in the submicromolar to nanomolar range) [37]. The obtained ΔG_{bind} values suggesting high affinity of PhenDC3 to pre-miRNA 149 G4. C_8 binding mode is also in agreement with previous NMR experiments performed in our group with two different G4 structures [20,43]. Groove binders [16]phenN_2 and PDS (ΔG_{bind} of -304.9 and -422.3, respectively) have smaller $T\Delta S$ values and higher electrostatic energies (ΔE_{elec}) as expected from their binding modes which compensate for the polar solvation term ΔG_{PB}. [16]phenN_2 was previously shown to bind other G4 structures with strong affinity with binding constants in the 10^6 M range, namely telomeric G4s [20]. Despite the MD simulations suggesting that the ligand binds exclusively to the AC loop, two high occupancy H-bonds were detected which may be responsible by the moderate binding strength suggested by the ΔG_{bind} value. Regarding PDS, which along with PhenDC3 is considered one of the most potent G4 ligand discovered to date, the MM/PBSA results suggest it to bind tightly to the rG4 structure. Previous solution experiments already shown PDS to strongly bind and stabilize the pre-miRNA 149 rG4 structure, being able to inhibit Dicer processing activity [6]. The C_8 precursor, C_8-NH_2 despite being able to bind the structure through both π-π stacking and groove binding, presented the lower ΔG_{bind} value (-266.0 kcal mol^{-1}). The entropic contribution is lower than that of C_8, which was not expected due to the broad surface covered requiring the displacement of water molecules from both the lower G-tetrad and the A4 loop/groove interface. Indeed, the ligand presented the higher ΔSASA values. These differences may arise from the lack of the terminal iodobenzene group that may participate in the water exclusion from the structure. [32]phen$_2$$N_4$ on its hand, despite showing a preferential groove binding mode, presented a higher entropic contribution when compared to [16]phenN_2 and PDS, and low van der Waals (ΔE_{vdW}) contribution. This may be due to the existence of two phenanthroline aromatic moieties. [32]phen$_2$$N_4$ bound the structure with a ΔG_{bind} of -349.5 of kcal mol^{-1}, having great affinity towards the rG4 structure, higher than the acridine derivatives. This may be overestimated, given the abovementioned issues with macrocyclic structures representation in GROMACS.

Altogether, the electrostatic (ΔE_{elec}), van der Waals (ΔE_{vdW}), and nonpolar solvation (ΔG_{SASA}) energies contribute favorably to the formation of all the complexes. These

results emphasize the importance of aromatic and positively charged in the interaction with the rG4 structure. It is worth mentioning that the available force fields do not handle well π-π interactions and hydrophobic exclusion [44], despite the major determinant of ligand-G4 interaction being these interactions. Therefore, experimentally derived binding energies may differ from the MM/PBSA calculations.

Table 7.2. Binding-free energy components of rG4-ligand complexes calculated from the MD simulations. The binding energies were calculated using MM/PBSA while entropy contribution was determined by normal mode analysis in GROMACS. All values reported in kcal mol^{-1} with standard deviations (SD) from three different calculations.

	pre-miRNA 149 G4					
	C_8	**C_8-NH_2**	**[16]phen N_2**	**[32]phen$_2$ N_4**	**PhenDC$_3$**	**PDS**
ΔE_{elec}	-188.7 ± 4.7	-206.3 ± 5.8	-390.3 ± 4.8	-361.0 ± 39.9	-298.7 ± 11.9	-507.9 ± 60.3
ΔE_{vdW}	-53.5 ± 3.9	-33.2 ± 1.2	-23.9 ± 1.1	-13.6 ± 1.7	-24.7 ± 1.4	-42.9 ± 1.6
ΔE_{MM} (ΔE_{elec} + ΔE_{vdW})	-242.1 ± 8.1	-239.5 ± 7.1	-414.2 ± 3.7	-374.6 ± 51.5	-323.4 ± 12.9	-550.8 ± 76.0
ΔG_{PB}	68.8 ± 2.5	68.6 ± 2.3	130.1 ± 7.6	82.2 ± 5.7	71.1 ± 2.5	140.1 ± 2.3
ΔG_{SASA}	-3.8 ± 0.4	-2.6 ± 0.84	-2.0 ± 0.1	-1.5 ± 0.4	-1.7 ± 0.6	-3.4 ± 0.8
ΔG_{solv} (ΔG_{PB} + ΔE_{SASA})	65.0 ± 2.8	66.0 ± 2.3	128.2 ± 7.7	80.7 ± 5.1	69.4 ± 2.4	136.7 ± 2.3
ΔG_{total} (ΔE_{MM} + ΔG_{solv})	-177.1 ± 10.8	-173.5 ± 4.8	-286.2 ± 4.0	-293.9 ± 67.03	-253.9 ± 10.4	-414.1 ± 53.1
$T\Delta S$	158.0 ± 11.2	92.5 ± 16.9	18.8 ± 7.1	55.6 ± 2.8	147.6 ± 7.8	8.3 ± 2.9
ΔG_{bind} ($\Delta G_{total} - T\Delta S$)	-335.0 ± 22.1	-266.0 ± 10.9	-304.9 ± 13.0	-349.5 ± 66.8	-401.5 ± 12.7	-422.3 ± 59.4

Conclusions

Targeting rG4 structures is emerging as an attractive approach for the development of anticancer therapeutic compounds. Notably, pre-miRNA 149 which is upregulated in prostate cancer is a suitable target for a ligand-based anticancer strategy. The homology modelling of pre-miRNA 149 rG4 structure indicates a propeller-type parallel stranded topology with short loops and three G-tetrads. The MD simulations showed that ligands C_8 and PhenDC$_3$ interact with the rG4 structure via stacking interactions with the top and bottom G-quartets, respectively. Both ligands showed a great entropic contribution upon ligand binding which may be due to the displacement of water molecules from the

rG4 surface, which is corroborated by ΔSASA. The precursor C_8-NH_2, despite having a lower free energy of binding, revealed the importance of the short loops of the rG4 structure for ligand design. The large aromatic terminal moiety of C_8 was pushed outwards as no binding pockets are produced by the rG4 loops, while the terminal amine of C_8-NH_2 was able to interact with the groove. Ligands [16]phenN_2, [32]phen$_2N_4$ and PDS on the other hand bind the loops/groove interface. Several H-bonds with the loop resides and electrostatic interactions with the phosphate backbone were detected. PDS due to its flexibility and protonable groups was able to adapt to the rG4 dynamics and presented a high free energy of binding. Our study provides useful hints for drug design as the structural nuances herein reported may be further explored for specific ligand development. Large aromatic surfaces such as acridine or phenanthroline with smaller positively charged side chains capable of performing mixed-mode interactions are preferred. In the future these results should be correlated with biophysical studies as to ascertain the validity of the *in silico* data herein reported.

Nonetheless, computational methods and molecular modeling have been increasingly used to predict G4–ligand interactions and the rationalization of biophysical data, which established a strong foundation for the use of these methods with meaningful outcome [22].

Acknowledgments and Funding

J. Carvalho acknowledges a doctoral fellowship grant from the FCT – Foundation for Science and Technology ref. SFRH/BD/122953/2016. T. Santos also acknowledges FCT for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013). This work was supported by project “Ações Integradas Luso-Francesas” ref. TC-15/17, Project “Programa Pessoa” ref. 5079/2019, FCT project ref. IF/00959/2015 financed by Fundo Social Europeu and Programa Operacional Potencial Humano, FCT project ref. UID/Multi/04349/2013, project Fundação Luso-Americana (FLAD) Healthcare 2020 ref. 45/2018, MIT Portugal FCT project BIODEVICE ref. MIT-EXPL/BIO/0008/2017 and UTAustin Portugal Program Exploratory project DREAM ref. UTAP-EXPL/NTec/0015/2017. This work was also supported by POCI – COMPETE 2020 – Operational Programme Competitiveness and Internationalisation in Axis I – Strengthening research, technological development and innovation (project POCI-01-0145-FEDER-007491). The authors acknowledge Jean-Louis Mergny for the valuable discussion and suggestions.

References

1 He, Y. et al. (2018) miR-149 in human cancer: A systemic review. J. Cancer 9, 375–388

2 Pandey, S. et al. (2015) The RNA Stem-Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. Biochemistry 54, 7067–7078

3 Peng, Y. and Croce, C.M. (2016) The role of MicroRNAs in human cancer. Signal Transduct. Target. Ther. 1, 15004

4 Bischoff, A. et al. (2014) MiR149 functions as a tumor suppressor by controlling breast epithelial cell migration and invasion. Cancer Res. 74, 5256–5265

5 Chen, Y. et al. (2016) Downregulated expression of miRNA-149 promotes apoptosis in side population cells sorted from the TSU prostate cancer cell line. Oncol. Rep. 36, 2587–2600

6 Kwok, C.K. et al. (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. Angew. Chemie - Int. Ed. 55, 8958–8961

7 Mitrasinovic, P.M. (2018) Structural insights into the binding of small ligand molecules to a G-quadruplex DNA located in the HIV-1 promoter. J. Biomol. Struct. Dyn. 36, 2292–2302

8 Park, J.H. et al. (2018) Enantioselective light switch effect of Δ- and Λ-[Ru(phenanthroline) 2 dipyrido[3,2-a:2′, 3′-c]phenazine] 2+ bound to G-quadruplex DNA. J. Biomol. Struct. Dyn. 36, 1948–1957

9 Poshteh Shirani, M. et al. (2018) Folate receptor-targeted multimodal fluorescence mesosilica nanoparticles for imaging, delivery palladium complex and *in vitro* G-quadruplex DNA interaction. J. Biomol. Struct. Dyn. 36, 4156–4169

10 Saha, U. et al. (2019) Targeting human telomeric DNA quadruplex with novel berberrubine derivatives: insights from spectroscopic and docking studies. J. Biomol. Struct. Dyn. 37, 1375–1389

11 Singh, A. and Kukreti, S. (2018) A triple stranded G-quadruplex formation in the promoter region of human myosin β (Myh7) gene. J. Biomol. Struct. Dyn. 36, 2773–2786

12 Suganthi, S. et al. (2019) Supramolecular complex binding to G-quadruplex DNA: Berberine encapsulated by a planar side arm–tethered β-cyclodextrin. J. Biomol. Struct. Dyn. 37, 3305–3313

13 Vinnarasi, S. et al. (2020) Structural insights into the anti-cancer activity of quercetin on G-tetrad, mixed G-tetrad, and G-quadruplex DNA using quantum chemical and molecular dynamics simulations. J. Biomol. Struct. Dyn. 38, 317–339

14 Viryasova, G.M. et al. (2019) G-quadruplex-forming oligodeoxyribonucleotides activate leukotriene synthesis in human neutrophils. J. Biomol. Struct. Dyn. 37, 3649–3659

15 Biffi, G. et al. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. Nat. Chem. 6, 75–80

16 Song, J. et al. (2016) RNA G-quadruplexes and their potential regulatory roles in translation. Translation 4, e1244031

17 Mirihana Arachchilage, G. et al. (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. Chem. Biol. 22, 262–272

18 De Cian, A. et al. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. J. Am. Chem. Soc. 129, 1856–1857

19 Carvalho, J. et al. (2018) Phenanthroline polyazamacrocycles as G-quadruplex DNA binders. Org. Biomol. Chem. 16, 2776–2786

20 Carvalho, J. et al. (2018) Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex. Biochimie 144, 144–152

21 Rodriguez, R. et al. (2008) A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. J. Am. Chem. Soc. 130, 15758–15759

22 Haider, S. Computational Methods to Study G-Quadruplex–Ligand Complexes. , Journal of the Indian Institute of Science. (2018)

23 Haider, S. and Neidle, S. (2010) Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes. In Methods in molecular biology (Clifton, N.J.) 608pp. 17–37

24 Li, J. et al. (2011) Identification of nonplanar small molecule for G-quadruplex grooves: Molecular docking and molecular dynamic study. Bioorganic Med. Chem. Lett. 21, 6969–6972

25 Aviñó, A. et al. (2017) Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. Proc. Natl. Acad. Sci. 114, E2136–E2145

26 Arba, M. et al. (2015) Molecular Docking and Dynamics Simulations on the Interaction of Cationic Porphyrin-Anthraquinone Hybrids with DNA G-quadruplexes. J. Biomol. Struct. Dyn. 1102, 1–36

27 Pronk, S. et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854

28 Pettersen, E.F. et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–12

29 Morris, G.M. et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–91

30 Sousa Da Silva, A.W. and Vranken, W.F. (2012) ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res. Notes 5, 367

31 Kollman, P.A. et al. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. Acc. Chem. Res. 33, 889–897

32 Kumari, R. et al. (2014) G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. J. Chem. Inf. Model. 54, 1951–1962

33 Phan, A.T. et al. (2006) Structural Diversity of G-Quadruplex Scaffolds. In Quadruplex Nucleic Acids First Edit. (Neidle, S. and Balasubramanian, S., eds), pp. 81–98, Royal Society of Chemistry

34 Miller, M.C. et al. (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG 3(TTAG3)3). J. Am. Chem. Soc. 132, 17105–17107

35 Forli, S. and Botta, M. (2007) Lennard-jones potential and dummy atom settings to overcome the AUTODOCK limitation in treating flexible ring systems. J. Chem. Inf. Model. 47, 1481–1492

36 Allen, S.E. et al. (2016) Dynamic Docking of Conformationally Constrained Macrocycles: Methods and Applications. ACS Chem. Biol. 11, 10–24

37 Chung, W.J. et al. (2014) Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. Angew. Chemie - Int. Ed. 53, 999–1002

38 Ruggiero, E. and Richter, S.N. (2018) Survey and summary G-quadruplexes and G-quadruplex ligands: Targets and tools in antiviral therapy. Nucleic Acids Res. 46, 3270–3283

39 Sargsyan, K. et al. (2017) How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. J. Chem. Theory Comput. 13, 1518–1524

40 Miyoshi, D. et al. (2006) Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. J. Am. Chem. Soc. 128, 7957–7963

41 Agrawal, S. et al. (2008) Energetics of the human tel-22 quadruplex - telomestatin interaction: A molecular dynamics study. J. Phys. Chem. B 112, 6828–6836

42 Dhamodharan, V. et al. (2015) Topology Specific Stabilization of Promoter over Telomeric G-Quadruplex DNAs by Bisbenzimidazole Carboxamide Derivatives. ACS Chem. Biol. 10, 812–833

43 Carvalho, J. et al. (2019) Aptamer-guided acridine derivatives for cervical cancer. Org. Biomol. Chem. 17, 2992–3002

44 Zhou, R. (2015) Modeling of nanotoxicity: Molecular interactions of nanomaterials with bionanomachines, Springer.

Supplementary Information

Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics

*Carvalho, J.; **Santos, T.**; Carrilho, R.; Sousa, F.; Salgado, G.F.; Queiroz, J.A. and Cruz, C. (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. J. Biomol. Struct. Dyn. 38, 2276–2286. DOI: 10.1080/07391102.2019.1632743 – IF: 3.392*

Results

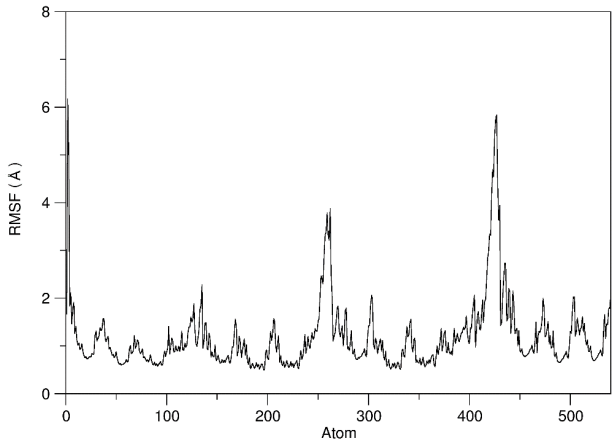

Figure S7.1. Root mean square fluctuation (RMSF) values of all atoms for pre-miRNA 149 G4 during the full 20 ns MD simulation.

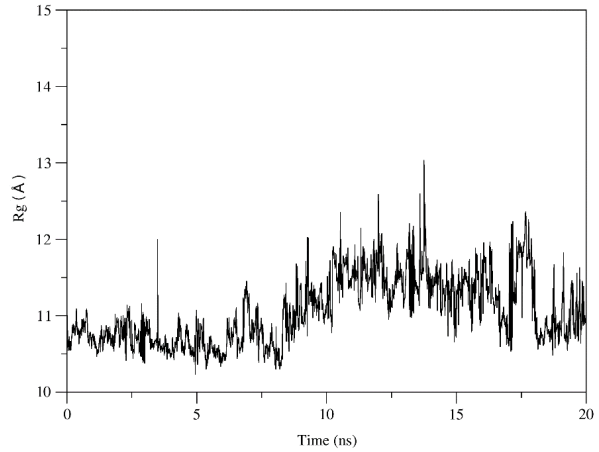

Figure S7.2. Radius of gyration variation for pre-miRNA 149 G4 during the full 20 ns MD simulation.

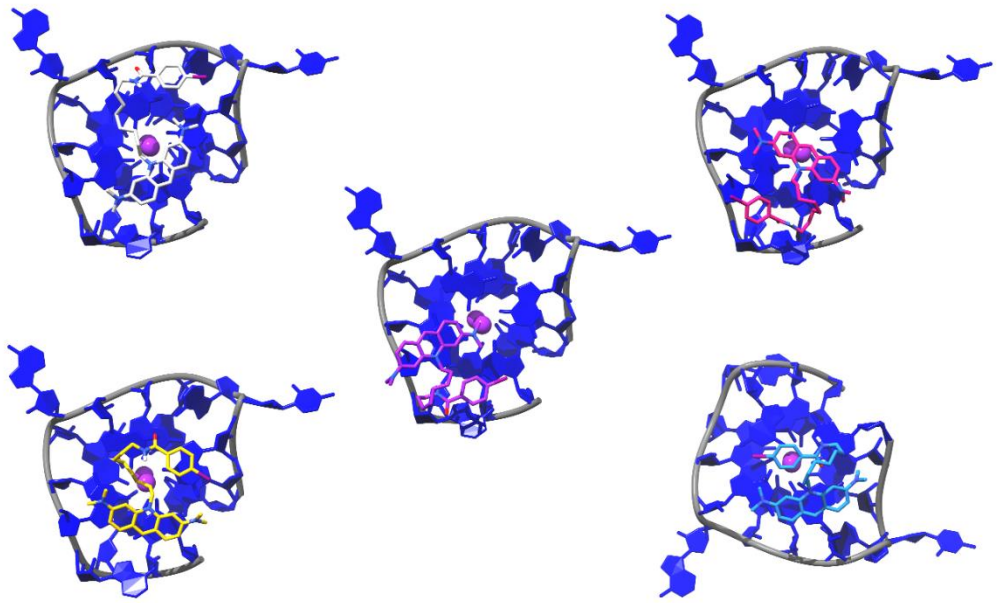

Figure S7.3. Best docking binding poses in terms of binding free energy of ligand C_8. G4 structure is depicted as a blue ribbon with K^+ cations depicted as purple spheres. The ligand different binding conformers are represented in different colors. The obtained ΔG values were -5.99, -4.52, -4.32, -3.64 and -3.59 kcal mol^{-1}.

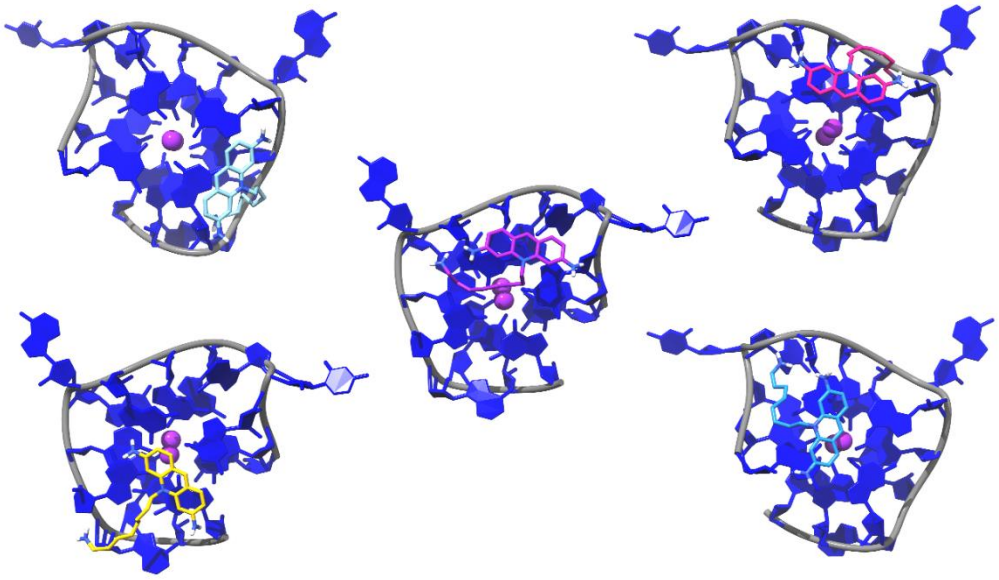

Figure S7.4. Best docking binding poses in terms of binding free energy of ligand C_8-NH_2. G4 structure is depicted as a blue ribbon with K^+ cations depicted as purple spheres. The ligand different binding conformers are represented in different colors. The obtained ΔG values were -4.88, -4.61, -4.57, -4.48 and -4.44 kcal mol^{-1}.

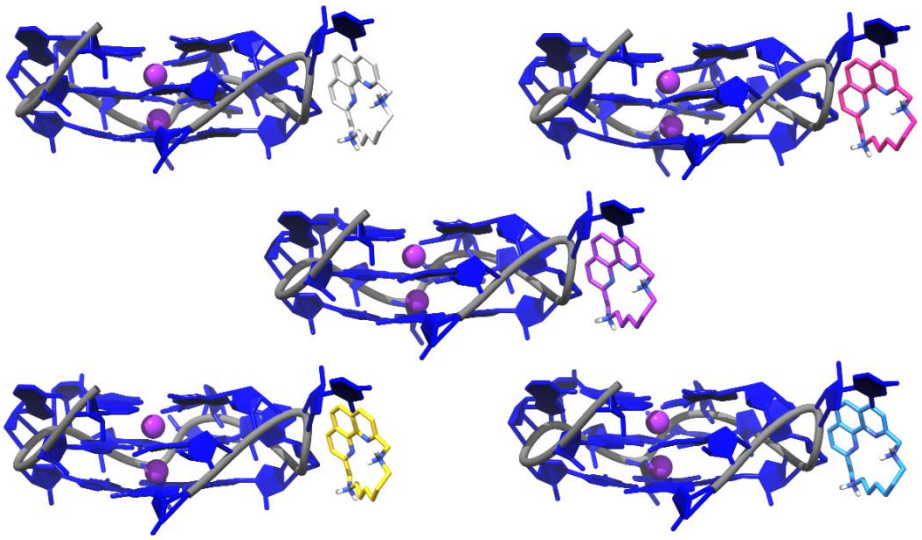

Figure S7.5. Best docking binding poses in terms of binding free energy of ligand [16]phenN_2. G4 structure is depicted as a blue ribbon with K^+ cations depicted as purple spheres. The ligand different binding conformers are represented in different colors. The obtained ΔG values were -7.81 kcal mol^{-1} for all conformations.

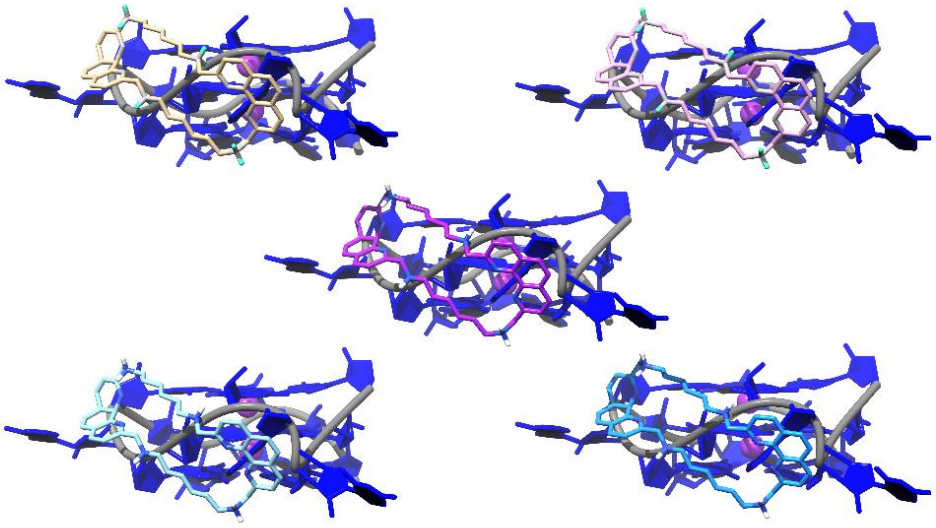

Figure S7.6. Best docking binding poses in terms of binding free energy of ligand [32]phen$_2N_4$. G4 structure is depicted as a blue ribbon with K^+ cations depicted as purple spheres. The ligand different binding conformers are represented in different colors. The obtained ΔG values were -11.07 kcal mol^{-1} for all conformations.

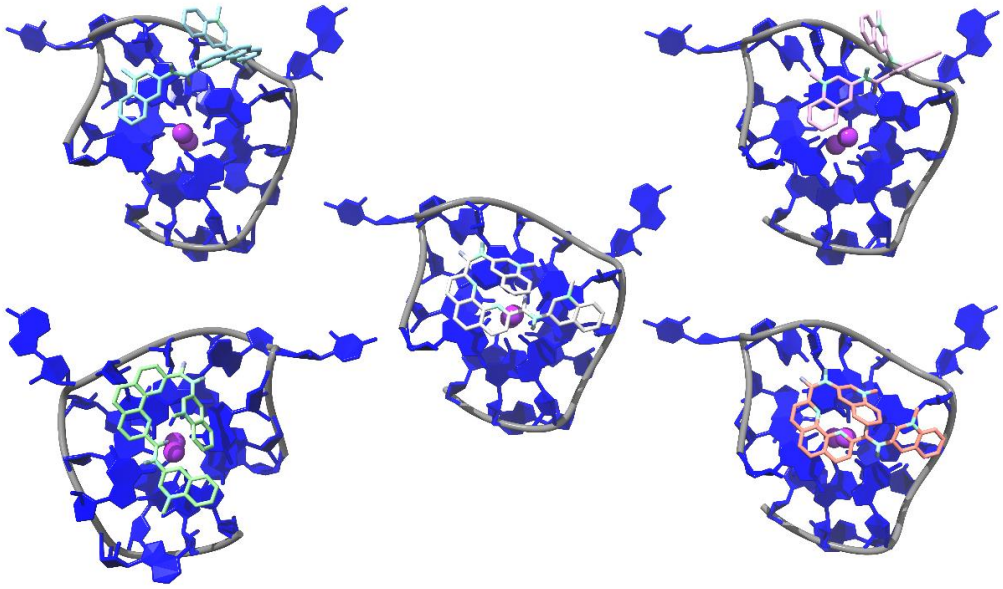

Figure S7.7. Best docking binding poses in terms of binding free energy of ligand $PhenDC_3$. G4 structure is depicted as a blue ribbon with K^+ cations depicted as purple spheres. The ligand different binding conformers are represented in different colors. The obtained ΔG values were -8.21, -8.14, -7.40, -7.28 and -7.26 kcal mol^{-1}.

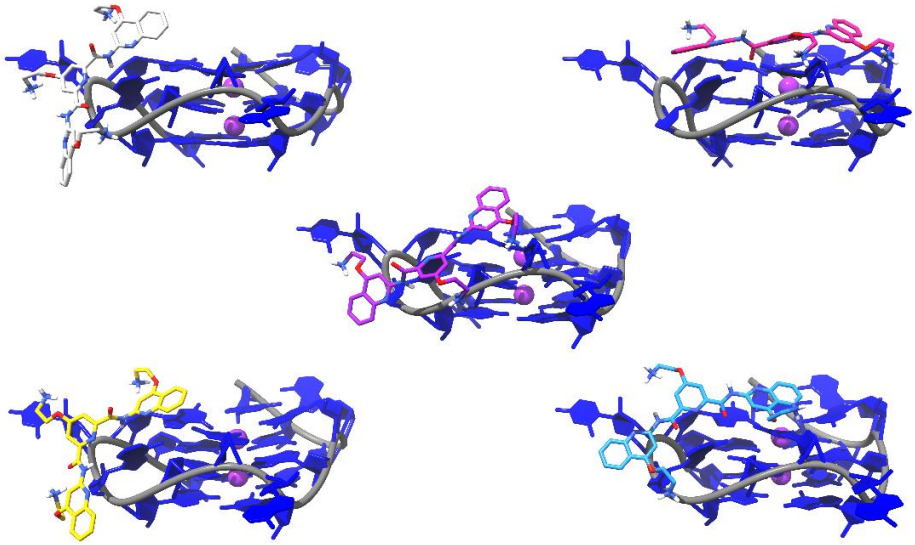

Figure S7.8. Best docking binding poses in terms of binding free energy of ligand PDS. G4 structure is depicted as a blue ribbon with K^+ cations depicted as purple spheres. The ligand different binding conformers are represented in different colors. The obtained ΔG values were -9.53, -8.30, -8.18, -8.01 and -7.82 kcal mol^{-1}.

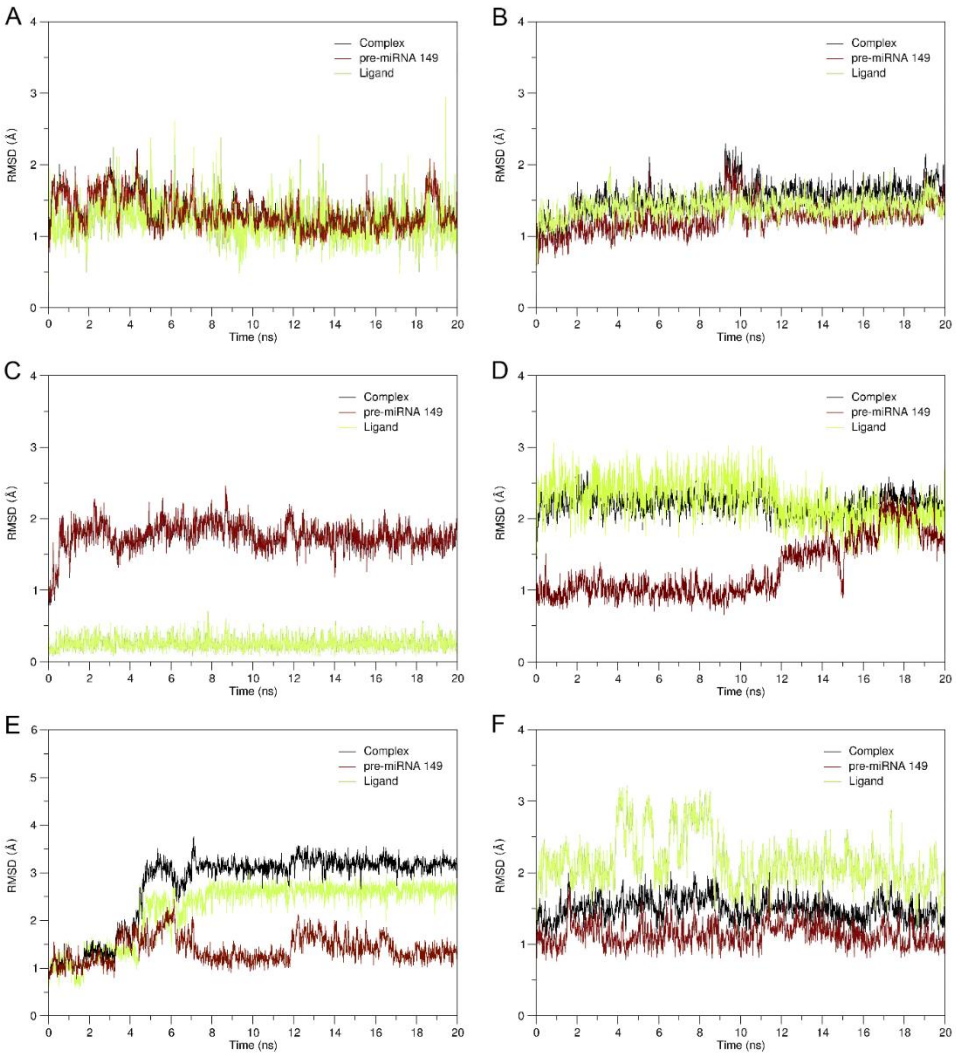

Figure S7.9. RMSD plots of the MD simulation of the complexes formed by pre-miRNA 149 and (A) C_8, (B) C_8-NH_2, (C) [16]phenN_2, (D) [32]phen$_2N_4$, (E) PhenDC_3 and (F) PDS. The black, red and green lines indicate the RMSDs for the rG4-ligand complex, pre-miRNA 149 and ligands, respectively.

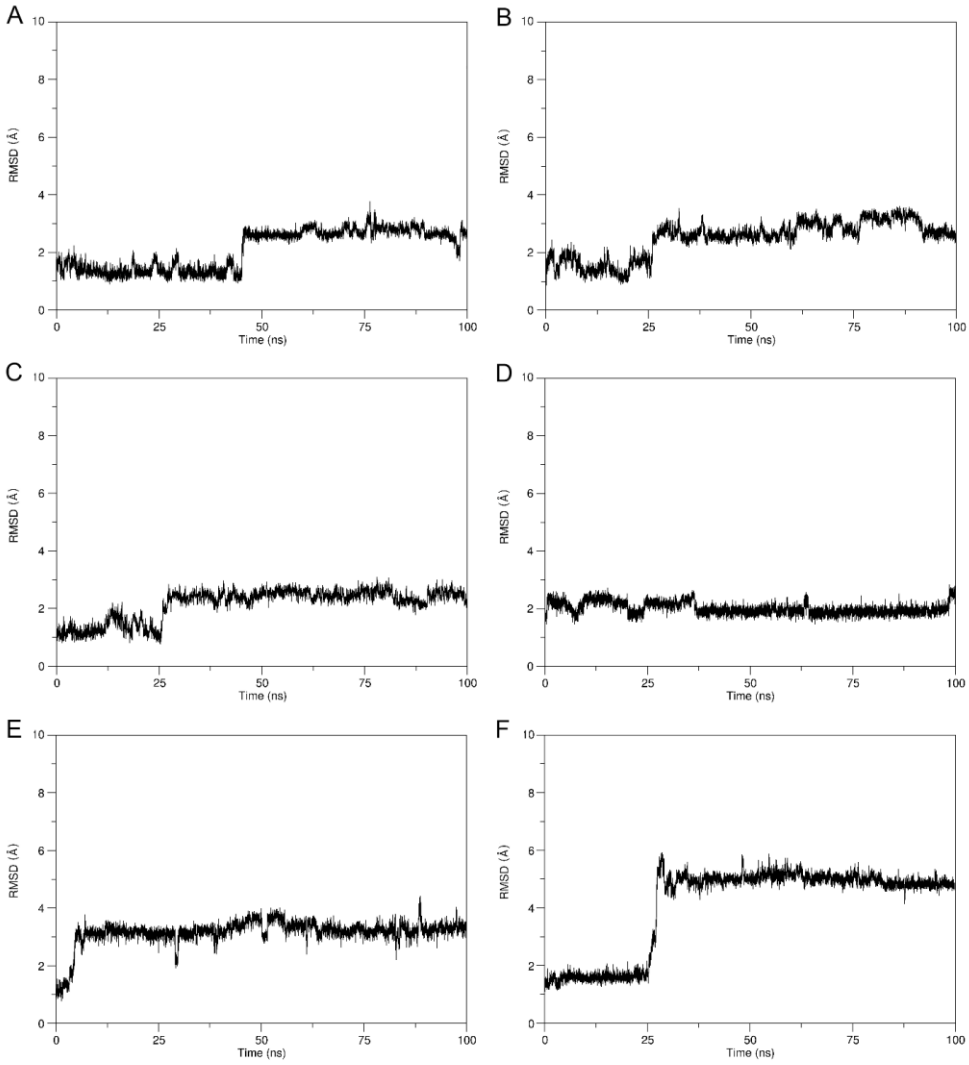

Figure S7.10. RMSD plots of the extended 100 ns MD simulations of the complexes formed by pre-miRNA 149 and (A) C_8, (B) C_8-NH_2, (C) [16]phenN_2, (D) [32]phen$_2N_4$, (E) PhenDC_3 and (F) PDS.

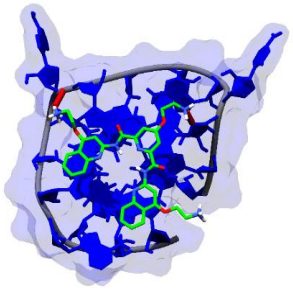

Figure S7.11. Alternative binding pose observed in the course of the 100 ns MD simulation of the complex formed by pre-miRNA 149 and PDS. Contrariwise to the conformation showed in Figure 7.11, the ligand is found stacking in the lower tetrad with additional H-bonds with the loop and G-tetrad residues. G4 structure is depicted as a blue surface, while the ligand is represented in green. Hydrogen bonds are depicted as red lines.

Chapter 8

RNA G-quadruplex as supramolecular carrier for cancer-selective delivery

This chapter was published in:

***Santos, T.**; Pereira, P.; Campello, M.P.C.; Paulo, A.; Queiroz, J.A.; Cabrita, E. and Cruz, C. (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. Eur. J. Pharm. Biopharm. 142, 473–479. DOI: 10.1016/j.ejpb.2019.07.017 – IF: 5.571*

Chapter overview

In chapter 7, an *in silico* approach was applied to understand the binding interaction of six different ligands with the G4 in pre-miRNA 149. In this chapter, C_8 efficiently stabilizes the G4 structure of pre-miRNA 149. Bearing this in mind, we move forward to the characterization of the G4 in pre-miRNA-149 as a molecular carrier for delivering C_8 to cancer cells via nucleolin binding. Previous studies demonstrated that nucleolin is an important multifunctional protein overexpressed on the surface of cancer cells but not in healthy ones. Furthermore, we take advantage of the known preference of nucleolin by G4 parallel structures.

RNA G-quadruplex as supramolecular carrier for cancer-selective delivery

Abstract

Nucleic acid aptamers have emerged as an attractive class of carrier molecules due to their ability to bind with high affinity to specific ligands; their high chemical flexibility; as well as tissue penetration capability. RNA G-quadruplex (rG4) sequences have been described as structures with high stability and selectivity towards cancer cells. Recently, precursor microRNAs (pre-miRNAs) have been described as new G4 forming molecules. Surface nucleolin (NCL) is a known target of aptamer G4 AS1411 and is overexpressed on prostate cancer cells when compared with normal cells. We have shown that the sequence 5′ GGGAGGGAGGGACGGG 3′ found in pre-miR-149 forms a rG4 parallel structure, which can bind NCL. Also, another rG4 sequence with a longer loop was evaluated in terms of G4 formation, stabilization and binding affinity to NCL.

Both rG4s sequences were studied as supramolecular carriers for the cancer-selective delivery of acridine ligand C_8. The rG4s-C_8 complexes showed high affinity ($K_D = 10^{-6}$ M) and stabilization ($T_m > 30$ °C). The affinity of the rG4s-C_8 complexes against NCL was in the low nanomolar range, indicating that C_8 did not affect NCL binding. Noteworthy, the short loop rG4-C_8 complex showed selective antiproliferative effects in prostate cancer cells when compared with normal prostatic cells. The stability and nuclease resistance of rG4 and rG4-C_8 complex were evaluated in biological conditions and revealed the maintenance of G4 structure and complex stability. Furthermore, confocal microscopy studies confirmed the potential of rG4s-C_8 complexes in the targeting of prostate cancer cells.

Overall, it is here demonstrated that the rG4 found in pre-miR-149 can be used as a cancer-selective delivery carrier of C_8 to prostate cancer cells.

Keywords

Precursor microRNA 149; G-quadruplex; Nucleolin; Drug delivery; Prostate cancer

Introduction

In recent years, the use of oligonucleotides as therapeutic agents has received the interest of researchers [1], [2], [3]. However, the efficacy of oligonucleotide-based therapies is highly impaired due to their susceptibility to degradation by nucleases and poor cellular internalization [4], [5]. A way to potentially overcome these limitations is by using G-quadruplex (G4) structures. These structures are assembled through the interactions between four guanines that are organized in a cyclic Hoogsteen hydrogen-bonding arrangement [6], and have been shown to display favorable intrinsic features such as high stability and resistance to nucleases, and enhanced cellular uptake [7].

RNA G-quadruplexes (rG4s) are much more stable than DNA G4s. In addition to important structural advantages related to the thermodynamical stability and less hydration [8], the presence of the 2′-OH group of the ribose sugar locks the RNA in an anti-conformation, thereby favoring the parallel topology [9]. Consequently, rG4s have less topological diversity than DNA homologous [9]. Together, the high stability displayed by rG4s and the conserved parallel topology makes them very interesting molecules for drug development.

rG4s play an important regulatory role in a wide range of biological events and they are distributed in important regions of the genome such as messenger RNAs (mRNAs) and non-coding RNAs [such as microRNAs (miRNAs)] [10]. In mRNAs, the presence of G4s mainly regulates the translation and the G4s found in non-coding RNAs have functional regulatory roles in pre- and post-transcriptional gene expression [10]. Taking this into account, several precursor microRNAs presenting G4 structures (pre-miRNAs-G4s) emerged as regulatory biologically relevant structures in RNA. The ability of pre-miR-149 [11], pre-miR-92b [12], [13] , and pre-miR-let-7e [14] to form G4 structures and regulate Dicer-maturation levels was recently studied. The presence of the G4 structure in pre-miRNAs has influence in miRNA biogenesis at the Dicer-maturation levels by competing with double-stranded stem-loop formation [13], [15]. Under certain conditions, such as the presence of salts and/or ligands, the pre-miRNA sequences can adopt the G4 structure avoiding the Dicer-mediated maturation and leading to subexpression of miRNA levels [11], [13].

The G4 sequence 5′ GGGAGGGAGGGACGGG 3′ (rG4) found in pre-miR-149 was reported as a G4 parallel sequence [16]. Previously it was demonstrated that pre-miR-149 produces miR-149-5p and miR-149-3p, which act as antitumor miRNAs through the targeting of several oncogenic genes in several cancers, namely prostate

cancer [17], [18]. The rG4 sequence partially overlaps with miR-149-3p, a well-known tumor suppressive miRNA [11]. Sequences with the potential to form G4s are attractive NCL binders since NCL has high affinity to G4s [16] and is overexpressed in prostate cancer cells [19], [20]. Recently, it was described that the loop length of G4 structures is a fundamental feature recognized by NCL [16], [21]. Via binding to cell surface NCL, rG4 can gain intracellular access by endocytosis and benefit of NCL shuttling to the nucleus. Based on that, rG4 can bind NCL while carrying ligands for an intracellular targeted delivery with anticancer effect.

Acridine derivatives have been extensively studied due to its anticancer properties [22], namely BRACO-19 that is one of the most studied G4 ligands and the first to prove anticancer activity *in vivo* [23]. Recently the binding of acridine orange derivatives to DNA G4s has been reported using a variety of biophysical and biological experiments [22]. However, their potential to bind rG4s has not been yet unraveled.

Considering these evidences, we performed biophysical and biological studies to evaluate the formation and stabilization of rG4 found in pre-miR-149, and to assess the formation of its supramolecular complex with the acridine orange derivative C_8 and its percursor C_8-NH_2 **(Figure S8.1)**. A long loop rG4 sequence (5′-GGGAUUGGGAUUUUUGGGAUCGGG-3′) was also studied to assess the influence of loop length to NCL binding. The rG4s-C_8 complexes were evaluated in terms of *in vitro* stability, cellular uptake and localization, and cytotoxic activity to prostate cancer *versus* normal cell lines.

Materials and methods

Oligonucleotides and ligands

All oligonucleotides were obtained from Eurogentec (Belgium) or STAB VIDA Genomics (Portugal) with HPLC-grade purification. The rG4 sequences used are 5′-GGGAGGGAGGGACGGG-3′ with loop length 112 and 5′-GGGAUUGGGAUUUUUGGGAUCGGG-3′ with loop length 363. Stock solutions of approximately 1 mM were prepared using nuclease free water and stored at −80 °C until used. The concentration of oligonucleotide samples was determined from the absorbance at 260 nm by using the molar extinction coefficient. Annealing of oligonucleotide sequences was performed by heating the samples for 10 min at 95 °C and slowly cooling on ice for 30 min before the experiments. Synthesis and purification of C_8 and C_8-

NH_2 ligands were performed as previously described [24]. Stock solutions of the compounds were prepared as 10 mM solutions in DMSO and their subsequent dilution was done using nuclease free water. Recombinant NCL peptide (partially RBD 2 and 3) was purchased from Cloud-Clone Corp. (Texas, USA). Stock solutions of approximately 1 mM were prepared using nuclease free water and stored at −80 °C.

Thermal difference spectroscopy (TDS) experiments

The TDS experiments were performed in an Evolution™ 220 UV/visible spectrophotometer (Thermo Scientific, USA). The TDS spectrum was obtained by subtracting the absorbance spectrum at 20 °C from the one at 90 °C. The spectrum at 90 °C was recorded after heating the sample at 90 °C for a few minutes, whereas the spectrum at 20 °C was recorded after annealing. TDS was carried out at 10 μM oligonucleotide strand concentration, in a lithium cacodylate buffer (10 mM) at pH 7.2, containing 0.1 or 5 mM KCl, using 1 cm path-length quartz cells. The UV TDS factors were determined according to the ratios ΔA240nm/ΔA295nm, ΔA255nm/ΔA295nm and ΔA275nm/ΔA295nm, where ΔAλ is the difference between the absorbance at 90 °C and at 20 °C at a given λ.

Circular dichroism spectroscopy

CD spectra were acquired in Jasco J-815 spectrometer (Jasco, USA), using a Peltier temperature controller (model CDF-426S/15). rG4s were annealed by heating at 95 °C for 10 min, following by slowly ice cooling in ice for 30 min. Otherwise stated, a 1 mm path-length quartz cuvette was used with an rG4 concentration of 10 μM in 10 mM lithium cacodylate at pH 7.2 containing 0.1 or 5 mM KCl. The required volume for the titrations was added directly in the quartz cell.

The CD melting experiments were performed in the temperature range 20–100 °C, with a heating rate of 2 °C/minute by monitoring the ellipticity at 263 nm. Spectra acquisition was performed in the absence and presence of 2 M equivalents of ligands. The ligands used were C_8 and C_8-NH_2. Data was converted into fraction folded (θ) plots using the following equation:

$$f = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \qquad \text{(Equation 8.1)}$$

where CD is the ellipticity of the monitored wavelength at each temperature and CD^{min} and CD^{max} are the lowest and highest ellipticity, respectively. Data points were

then fitted to a Boltzmann distribution (OriginPro 2016) and the melting temperatures were determined from the two-state transition model using the first derivative method.

Fluorescence spectroscopy titrations

All steady-state fluorescence measurements were collected at 25 °C on a FluoroMax4 (Horiba, Japan) equipped with a Peltier-type temperature control system. Samples and references were scanned using 1 cm path length quartz microvolume cuvettes with an optimal volume of 700 µL. All spectra were scanned with an integration time of 0.5 s, an emission and excitation slit width of 2 nm and step size of 1 nm. The association between rG4 sequences and ligands/NCL was followed by titrating the oligonucleotides/NCL, respectively, and measuring the change in fluorescence. The titration was performed by adding previously annealed rG4 sequences or NCL stock solutions, followed by 3 min for equilibration. The obtained data was converted into fraction of bound ligand (α) plots using the following equation:

$$\alpha = \frac{I - I_{\lambda}^{free}}{I_{\lambda}^{bound} - I_{\lambda}^{free}} \qquad \text{(Equation 8.2)}$$

where I is the fluorescence intensity of each ligand/rG4 or rG4/NCL ratio and I^{free} and I^{bound} are the fluorescence intensity of the free and fully bound ligand, respectively. Data points were then fitted to a hyperbolic function (OriginPro 2016) and K_D values were determined from the following saturation binding model.

$$\alpha = \frac{[rG4\ or\ nucleolin]}{K_D + [rG4\ or\ nucleolin]} \qquad \text{(Equation 8.3)}$$

where α is the fraction of ligand bound and [rG4 or nucleolin] is the oligonucleotide or NCL concentration. The saturation binding Hill slope model was used.

$$\alpha = \frac{[rG4\ or\ nucleolin]^h}{K_D + [rG4\ or\ nucleolin]^h} \qquad \text{(Equation 8.4)}$$

Where h is the Hill constant which describes cooperativity of ligand binding.

Viability/Proliferation assays

Human prostatic adenocarcinoma (PC-3) and human prostatic (PNT1A) cell lines were grown in Roswell Park Memorial Institute medium (RPMI) supplemented with 10% (m/v) fetal bovine serum (FBS), and 1% (m/v) penicillin-streptomycin. Cultures were

maintained at 37 °C in a humidified atmosphere containing 5% CO_2. Cells were harvested, counted using the trypan blue exclusion method and seeded in 48-well plates (1×10^4 cells/well and 1×10^3 cells/well for compounds and rG4/rG4-C_8 complex, respectively) and after 24 h these were incubated with the compounds (C_8 and C_8-NH_2), rG4 or rG4-C_8 complex. The concentrations tested for the compounds ranged from 100 nM to 10 μM for incubation times of 12, 24, 48 and 72 h. The concentration used for rG4 and rG4-C_8 was 15 μM for an incubation time of 7 days. Wells containing untreated cells were used as a control. At the end of incubation, the media was replaced with fresh serum-free medium containing 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide salt (MTT) and further incubated at 37 °C for 2 h. Finally, MTT containing medium was removed, formazan crystals were dissolved in DMSO and absorbance was recorded in a Bio-Rad xMark™ microplate reader (Bio-Rad, USA) at 570 nm. Cell viability data were expressed as mean ± SEM from at least three different experiments in comparison with untreated cells.

Nuclease stability assay

Nuclease stability assays of rG4 (20 μM) and rG4-C_8 complex were conducted firstly in RNase H solution (0.3-U/μL) (NZYtech, Portugal) and then in cell culture medium RPMI supplemented with 10% (m/v) FBS and 1% (m/v) penicillin-streptomycin for 1, 3, 12, 24 and 48 h at 37 °C. A stock solution of each sequence was prepared in 10 mM lithium cacodylate buffer and 0.1 mM KCl, pH 7.2. For nuclease stability assays a molar ratio of 1:1 of rG4:C_8 was prepared. Thereafter, rG4 and rG4–C_8 complex were annealed by heating to 95 °C for 10 min, following by a slowly cooling in ice for 30 min. 10 μL of each mixture was used for agarose gel electrophoresis, which was carried out using 1% agarose gel stained with 0.01% Green Safe in 1× TAE buffer (Tris-acetate-EDTA). The degradation patterns on the gel were visualized through UV light.

Confocal microscopy studies

PC-3 and PNT1A cell lines were grown in RPMI medium supplemented with 10% (v/v) FBS and 1% (v/v) penicillin-streptomycin. Cultures were maintained in a humidified chamber at 37 °C and 5% CO_2. The cell lines were subsequently harvested, counted using the trypan blue exclusion method and seeded in μ-Slide 8-well flat bottom imaging plates (Ibidi GmbH, Germany) at a plating density of 5×10^4 cells/well and incubated for cell attachment for 24 h at 37 °C and 5% CO_2. Thereafter, cells were incubated with the primary anti-NCL polyclonal antibody (Thermo Scientific, USA) for 2 h at 37 °C. Following primary antibody incubation, cells were washed 3× with fresh serum-free

medium and incubated with secondary antibody anti-rabbit IgG conjugated with Alexa Fluor® 647 (Thermo Scientific, USA) for 1 h at 37 °C. Thereafter, cells were washed 3 × with fresh serum-free medium and stained with Hoechst 33342® nuclear probe (1 μM) for 10 min. Subsequently, the cells were incubated with C_8 (1 μM), rG4-Cy3.5 (1 μM) and rG4-C_8 complexes (1:1 ratio). Thereafter, the cells were transferred to a Zeiss LSM 710 confocal laser scanning microscope (CLSM; Carl Zeiss SMT Inc., USA) equipped with a plane-apochromat 63×/DIC objective and processed in Zeiss Zen (SP2, 2010), in order to evaluate the cellular uptake. The fluorescence images were obtained at 63× magnification.

Results and discussion

Binding and stabilization of the rG4s

The human precursor microRNA 149 (pre-miR-149) was identified by the SHALiPE approach, revealing the formation of a rG4 structure in this non-coding RNA with parallel topology, thermally stable, and conserved in mammals [11]. Herein, we studied the formation and topology of the rG4 found in pre-miR-149 (5′-GGGAGGGAGGGACGGG-3′) by circular dichroism (CD) and UV/Vis spectroscopy, namely by the use of thermal difference spectroscopy (TDS). The TDS spectrum confirms that a G4 structure is adopted displaying two positive peaks at 245 and 275 nm and a negative peak at 295 nm **(Figure S8.2A)** [25]. The TDS factors $\Delta A_{240}/\Delta A_{295}$, $\Delta A_{255}/\Delta A_{295}$, $\Delta A_{275}/\Delta A_{295}$ were determined and revealed values of 4.4, 7.1 and 3.8, respectively. These magnitudes are characteristic of a parallel G4 topology [26]. The CD spectrum shows typical CD signature of parallel G4 topology **(Figure S8.2B)**, with a positive peak at around 260 nm and a negative peak at around 240 nm. Altogether, these data show that the pre-miR-149 sequence adopts a parallel G4 structure in 0.1 mM KCl.

The stabilization through ligand binding is required to decrease the unfolded or partially folded intermediate states. The ligand C_8 was previously described as DNA intercalator [24] and DNA G4 binder/stabilizer of promoter and telomeric G4s with high affinity ($K_D \approx 10^{-7}$ M) and stabilizing ability ($\Delta T_m > 30$ °C) [22]. Herein, we used a non-covalent strategy to associate the ligand C_8 and its precursor C_8-NH_2 with rG4s of pre-miR-149 in an attempt to bind and stabilize the G4 structure. The formation of stable complexes was assessed and characterized by CD and fluorescence spectroscopy. CD titrations were first performed to evaluate complex formation and/or structural modifications of the G4 structure upon ligand binding. The CD spectra of rG4

upon addition of C_8 and C_8-NH_2, showed no relevant variations on RNA ellipticity, suggesting an overall conservation of the parallel G4 topology **(Figure 8.1A and 8.1B)**.

The influence of ligands C_8 and C_8-NH_2 on the thermal stability of the rG4 was investigated by thermal denaturation using CD melting experiments. The ellipticity of parallel rG4 was monitored at 260 nm and distinct temperatures in the presence of KCl, allowing to determine the effect of the acridine orange derivatives in the melting temperature (T_m) of rG4 (51.8 ± 0.1) **(Figure 8.1C)**. In general, we observed that C_8 and C_8-NH_2 at 2 M equivalents led to an increase of more than 30 °C and 18.5 ± 0.4 in rG4 T_m, respectively.

The effect of the loop length in terms of rG4-C_8 complex formation and stability was also evaluated by TDS and CD experiments. The selected sequence was 5′-GGGAUUGGGAUUUUUGGGAUCGGG-3′ with loop length 363. The TDS and CD spectra of the long loop showed that sequence adopts a parallel rG4 in presence of 5 mM KCl **(Figure S8.2C and S8.2D)**. The stabilization through C_8 and C_8-NH_2 at 2 M equivalents is similar to the obtained with the short loop, with $T_m > 30$ °C and 20.0 ± 0.5 °C for C_8 and C_8-NH_2, respectively **(Figure S8.3A, S8.3B and S8.3C)**.

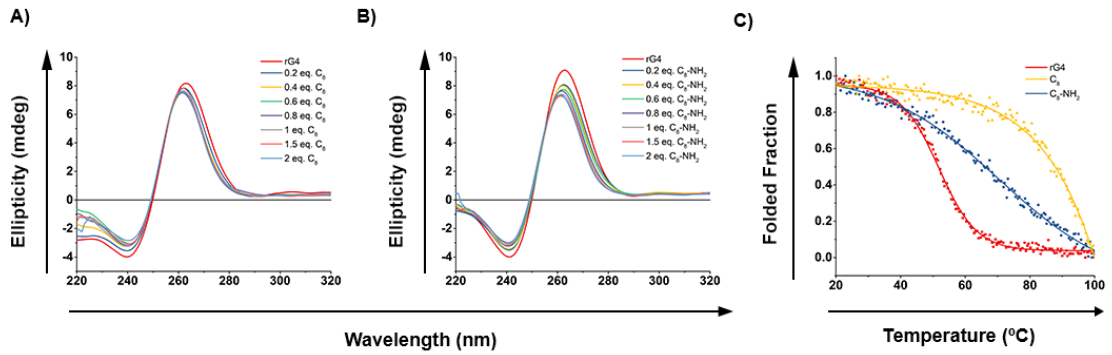

Figure 8.1. CD titration spectra of short loop (112) rG4 at 10 µM with increasing concentrations of **(A)** C_8 and **(B)** C_8-NH_2. Spectra acquisitions were performed at 20 °C in 10 mM lithium cacodylate and 0.1 mM KCl; **(C)** CD melting curves at 10 µM in the absence and presence of 2 M equivalents of ligands in 10 mM lithium cacodylate and 0.1 mM KCl, respectively. Data points were recorded at 263 nm. Temperature corresponds to the temperature set by the Peltier system.

Fluorescence binding studies with rG4 sequences

The binding strength of C_8 and C_8-NH_2 to rG4 sequences with loops 112 and 363 in the parallel conformation was investigated by fluorescence titrations. The fluorescence spectra of rG4 with loop 112 are presented in **Figure 8.2A and 8.2B** and rG4 with loop 363 in **Figure S8.4A and S8.4B**. The fraction of bound ligand (α) at each point of the

titration was calculated following fluorescence changes at the maximum of intensity and was plotted as a function of the rG4 concentration to obtain an isotherm binding curve. The curve was fitted to a saturation binding model (Eq. (4), Section 2) and the apparent dissociation constants (K_D) were determined. The K_D values of rG4 with loop 112 with C_8 and C_8-NH_2 are 1.0 ± 0.1 μM and 0.9 ± 0.1 μM, respectively, and for rG4 with loop 363 are 0.8 ± 0.4 μM and 0.7 ± 0.1 μM, for C_8 and C_8-NH_2, respectively. For both sequences the K_D values are in the low micromolar range, indicating high affinity, being similar to those previously reported for other G4 ligands, such as BRACO-19 [27] or pyridostatin [28].

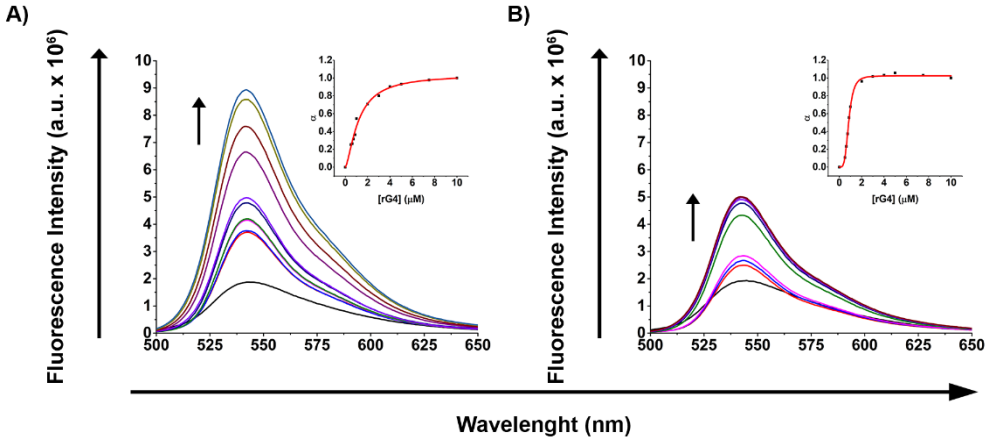

Figure 8.2. Fluorescence emission spectra of **(A)** C_8 and **(B)** C_8-NH_2 at 5 μM with different concentrations of short loop (112) rG4 in 10 mM lithium cacodylate buffer and 0.1 mM KCl. Insets: fraction of ligand bound plots fitted to the saturation binding equation.

Fluorescence binding studies with NCL

Then, we assessed the binding strength of NCL to rG4s (short and long loop rG4) and rG4s-C_8 complexes by fluorescence titrations. The fluorescence experiments were performed with rG4s labelled with Cy3.5. Fluorescence emission spectra of rG4s-Cy3.5 and rG4s-Cy3.5-C_8 complexes were recorded at 577 nm, in the absence and presence of increasing amounts of NCL RBD 2,3. The saturation binding plots were obtained by non-linear regression analysis and are presented in **Figure 8.3 and S8.5** for short and long loop sequences, respectively. The K_D values for the binding interaction between the short loop (112) rG4 and its rG4-C_8 complex with NCL were 309 ± 45 nM and 30 ± 22 nM, respectively. For the NCL interaction with the long loop (363) rG4 and its rG4-C_8 complex, the K_D values were 35 ± 14 nM and 2 ± 1 nM, respectively, indicating an increase in NCL binding upon loop extension. For both cases, the K_D values of complexes is in the low nanomolar range, in agreement with those previously reported

for aptamer AS141, and thereby binds specifically to NCL [29]. This indicates that C_8 does not prevent the recognition of the NCL by the rG4.

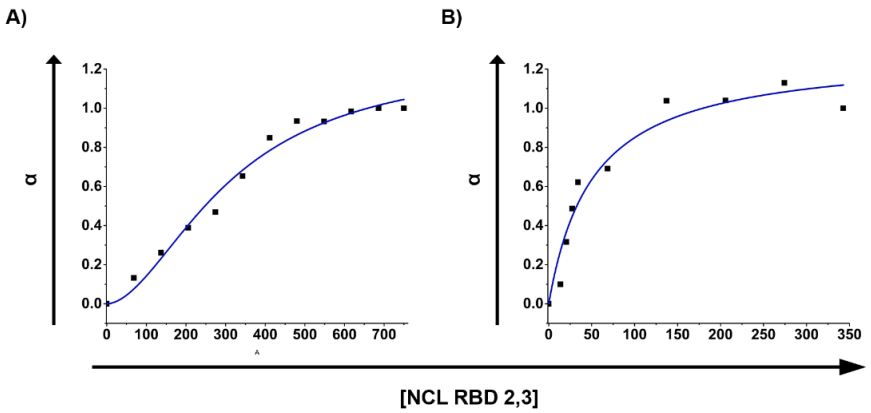

Figure 8.3. Plots of the fraction of bound NCL to **(A)** short loop rG4-Cy3.5 and **(B)** short loop rG4-Cy3.5-C_8 complex at 1 μM in 10 mM lithium cacodylate buffer and 0.1 mM KCl.

Cytotoxic profile and antiproliferative effect of the complexes

The cytotoxicity effect of ligands C_8 and C_8-NH_2 in prostate cancer cell line PC-3 and an epithelial cell line derived from normal prostatic epithelium, PNT1A, was determined by the MTT assay. The results showed that the cytotoxicity of the C_8 and C_8-NH_2 increased with concentration and incubation time **(Figure S8.6)**. The relative cell viability of PC-3 and PNT1A cells, incubated for 72 h with free C_8 and C_8-NH_2 **(Figure S8.6)**, showed indistinct cytotoxic effect for these compounds. In fact, they induced a similar antiproliferative effect in PC-3 and PNT1A cells with IC_{50} values in the ranges 0.3–0.5 μM and 4.3–9.1 μM for C_8 and C_8-NH_2, respectively **(Figure S8.7)**. These IC_{50} values are lower than that obtained with G4 ligand BRACO-19 in DU145 cells after 96 h, with an IC_{50} value of 22.3 ± 0.8 μM [30]. In light of these results, we proceeded with the biological evaluation of C_8 that showed the highest antiproliferative activity and studied the possibility to enhance its cancer selectivity by exploring its supramolecular delivery by rG4 structures. Due to similar T_m and K_D values for long and short loop rG4 sequences, the antiproliferative effect of complex rG4-C_8 was evaluated only using short loop rG4 by MTT assay against PC-3 and PNT1A, to study if the supramolecular delivery strategy would lead to tumour-selective properties. The results are presented in **Figure 8.4**. The rG4 showed reduced and similar cytotoxicity towards the PC-3 and the non-malignant PNT1A cells (75% mean viability). However, after mixing 15 μM rG4 with 1 μM C_8 to allow complex formation, the C_8 toxicity remained quite pronounced against the PC-3 cells (25% mean viability) but was strongly reduced in the non-malignant cells (75% mean viability). This result suggests a cancer-selective antiproliferative effect of complex

rG4-C_8. The lower cytotoxicity of C_8 towards non-malignant cells when complexed with rG4 may be due to the mechanism involved in the selective accumulation of oligonucleotides in cells. It has been proposed that DNA aptamer AS1411 is gradually cleared from normal cells by efflux or exocytosis 24–72 h post-treatment [31]. Additionally, normal cells have increased lysosomal activity in comparison with AS1411-treated cancer cells. Therefore, if the rG4-C_8 complex is capable of resisting cellular trafficking it may be cleared from normal cells by lysosomal degradation, thus promoting C_8 efflux and reducing its effects. In the case of PC-3 cells, the NCL-mediated disruption of rG4-C_8 trafficking and efflux might lead to C_8 entrapping inside the cancer cells [31].

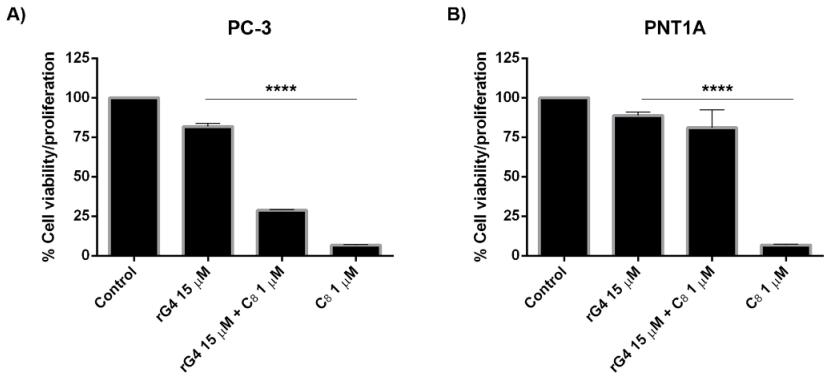

Figure 8.4. Relative cell viability of **(A)** PC-3 and **(B)** PNT1A cell lines incubated for 7 days with short loop (112) rG4, and rG4-C_8 complex at C_8 and rG4 concentrations of 1 μM and 15 μM, respectively. Untreated cells were used as negative controls for cytotoxicity. Mean percentage values relative to untreated cells and standard error of the mean in 3 independent experiments are shown. The bars represent the mean and the lines represent the SEM associated. $^{****}P < 0.0001$ (one-way ANOVA).

Stability of rG4-C_8 complex

Since the rG4-C_8 complex (short loop rG4) exhibited a noteworthy antiproliferative activity, its resistance in biological environments was also evaluated through a degradation assay by incubation of the complexes, firstly with RNase H solution (0.3 U/μL) and then with culture medium supplemented with 10% FBS. Thus, the stability and resistance of the rG4 and rG4-C_8 complex were evaluated through gel agarose electrophoresis analysis **(Figure S8.8)**, after 1, 3, 12, 24 and 48 h of incubation, which is a crucial issue in its development as a potential therapeutic agent. The electrophoretic profile of rG4 and rG4-C_8 showed that both structures were completely resistant after 48 h in the presence of culture medium supplemented with 10% FBS. The electrophoretic profile of the rG4 in presence of RNase H (0.3 U/μL) showed the degradation of rG4. However, an increase in band intensities was observed when C_8 is complexed with rG4, suggesting a stabilizing effect of C_8.

We have also checked by CD spectroscopy during the 7 days if the rG4 parallel topology **(Figure S8.9A)** was maintained in cellular medium and in the presence of C_8 **(Figure S8.9B)**. No significant changes in the ellipticity were noticeable in both cases.

In vitro cellular uptake and intracellular distribution of the rG4-C_8 complex

The subcellular localization and cellular uptake of the rG4s-C_8 complexes (short and long loop rG4) in PC-3 and PNT1A cells were evaluated by fluorescence confocal microscopy. We intend to evaluate if the loop length affects the uptake, intracellular distribution and NCL binding. The rG4s were labeled with Cy3.5 and, taking advantage of the intrinsic fluorescence of C_8, the distribution of the rG4s-C_8 complexes were visualized in the mentioned cell lines. The primary anti-NCL antibody conjugated with the secondary antibody AlexaFluor 647® was used to localize cell surface NCL. We have hypothesized that NCL is the cell surface target of rG4 and rG4-C_8 and may be a key transporter from the membrane through to the nucleus. As seen in **Figure 8.5 and S8.10**, the complexes rG4-C_8 are able to penetrate cell membrane and localize in the cytoplasm of PC-3 cells. C_8 can be seen in the nucleoli in a free state which may suggest partial decomplexation and localization of the nucleolus as described for the ligand [22]. After 1 h of rG4-C_8 complexes incubation, changes in cell morphology consistent with the cytoplasmic vacuolation are observed, but less pronounced for long loop rG4-C_8 complex. The cytoplasmic vacuolation observed is similar to that seen for higher concentrations of the AS1411 in several cancer cells [32], [33], [34]. A more modest uptake of both rG4-C_8 complexes were detected in the cytoplasm of PNT1A cells in agreement with the cytotoxic results **(Figure 8.5)**.

Overall, our study indicates that the rG4-C_8 complexes have tumour-selective properties, which are favourable features for its use as a potential cytotoxic agent [35]. Moreover, the observed cell specificity for PC-3 cells together with the enhanced antiproliferative activity exhibited by rG4-C_8 on cultured cells, in the absence of any transfection agent, suggest a mechanism of internalization mediated by NCL, which is more expressed in prostate cancer cells than in normal prostatic cells.

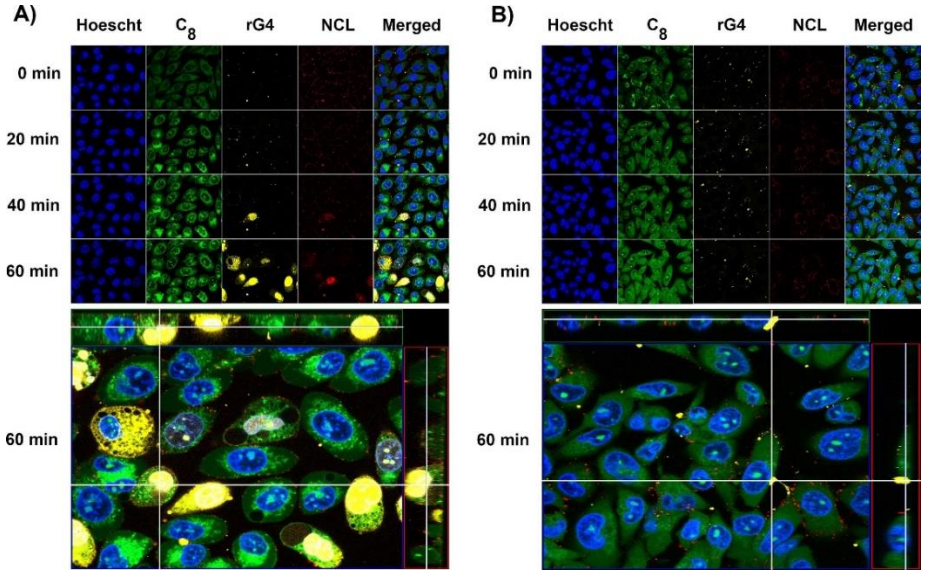

Figure 8.5. Confocal laser scanning microscopy images showing the intracellular uptake of short loop rG4-C_8 complex from 0 to 60 min. For each panel, images showed the cells with nuclear staining by Hoechst 33342® (1 μM, blue); rG4 Cy3.5-C_8 complex (1:1 ratio, yellow); C_8 (1 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:1000). **(A)** Representative z-stacks data showing PC-3 cell line and **(B)** Representative z-stacks data showing PNT1A cell line.

Conclusions

Herein, we have studied the rG4 found in pre-miR-149 as a drug delivery carrier of an acridine-based G4 ligand, C_8, with known anticancer properties. We demonstrated the formation of the rG4-C_8 complex, showing that C_8 has a high affinity to the rG4 and stabilizes the rG4 structure in more than 30 °C. We also evaluated a long loop of rG4 sequence and its C_8 complex in terms of stability and binding to NCL. Both rG4s-C_8 complexes bound NCL with high affinity ($K_D \approx 10^{-9}$ M), suggesting that C_8 complexation did not affect the recognition of NCL, which is necessary to obtain selective biological effects. Consistently, it was observed that rG4-C_8 has an augmented cytotoxic activity in PC-3 cells when compared with non-malignant cells. The nuclease stability assay performed in biological conditions showed that the structure of rG4 and rG4-C_8 remain stable after incubation at 37 °C for 48 h in the presence of cell culture medium. This is a very favorable feature as many therapeutic oligonucleotides are highly unstable and thus easily degraded by nucleases. Finally, the experiments performed with confocal microscopy indicated that complexes rG4s-C_8 internalized cancer cells and were partially maintained during cell internalization and trafficking to the nucleus; C_8 localized *nucleoli* in a free state

which may suggest partial decomplexation. Overall, the results suggest that rG4-C_8 complex can be used for the development of delivery system to prostate cancer cells.

Acknowledgments

Tiago Santos acknowledges Fundação para a Ciência e Tecnologia (FCT), Portugal for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013). C. Cruz acknowledges the FCT project reference IF/00959/2015, FLAD-Healthcare 2020, Portugal project ref. 45/2018 entitled "Development of drug delivery nanocarrier for HPV infection", MIT-EXPL/BIO/0008/2017 entitled "Biomedical device development based on pre-miRNA G-quadruplex" and UTAP-EXPL/NTec/0015/2017 entitled "DRug dElivery nAnosysteM for HPV infection therapy". This work was also supported by the FCT project ref. UID/Multi/04349/2019 and by POCI – COMPETE 2020 – Operational Programme Competitiveness and Internationalisation in Axis I – Strengthening research, technological development and innovation (project POCI-01-0145-FEDER-007491).

References

1 Lundin, K.E. et al. (2015) Oligonucleotide Therapies: The Past and the Present. Hum. Gene Ther. 26, 475–485

2 Ng, E.W.M. et al. (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. Nat. Rev. Drug Discov. 5, 123–132

3 Zhao, N. et al. (2015) Oligonucleotide aptamer-drug conjugates for targeted therapy of acute myeloid leukemia. Biomaterials 67, 42–51

4 Juliano, R. et al. (2008) Mechanisms and strategies for effective delivery of antisense and siRNA oligonucleotides. Nucleic Acids Res. 36, 4158–4171

5 Juliano, R.L. (2016) The delivery of therapeutic oligonucleotides. Nucleic Acids Res. 44, 6518–6548

6 Huang, H. et al. (2017) RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. Genes Dev 31, 2296–2309

7 Bates, P.J. et al. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. Exp Mol Pathol. 86, 151–164

8 Cammas, A. and Millevoi, S. (2017) RNA G-quadruplexes: emerging mechanisms in disease. Nucleic Acids Res. 45, 1584–1595

9 Zaccaria, F. and Fonseca Guerra, C. (2018) RNA versus DNA G-Quadruplex: The Origin of Increased Stability. Chem. Eur. J. 24, 16315–16322

10 Fay, M.M. et al. (2017) RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. J Mol Biol 429, 2127–2147

11 Kwok, C.K. et al. (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. Angew Chem Int Ed Engl 55, 8958–8961

12 Mirihana Arachchilage, G. et al. (2018) Targeting of G-Quadruplex Harboring Pre-miRNA 92b by LNA Rescues PTEN Expression in NSCL Cancer Cells. ACS Chem. Biol. 13, 909–914

13 Mirihana Arachchilage, G. et al. (2015) A Potassium Ion-Dependent RNA Structural Switch Regulates Human Pre-miRNA 92b Maturation. Chem. Biol. 22, 262–272

14 Pandey, S. et al. (2015) The RNA Stem–Loop to G-Quadruplex Equilibrium Controls Mature MicroRNA Production inside the Cell. Biochemistry 54, 7067–7078

15 MacRae, I.J. et al. (2007) Structural determinants of RNA recognition and cleavage by Dicer. Nat Struct Mol Biol 14, 934–940

16 González, V. et al. (2009) Identification and Characterization of Nucleolin as a c-myc G-quadruplex-binding Protein. J. Biol. Chem. 284, 23622–23635

17 Okato, A. et al. (2017) Dual Strands of Pre-miR-149 Inhibit Cancer Cell Migration and Invasion through Targeting FOXM1 in Renal Cell Carcinoma. Int. J. Mol. Sci 18, 1969

18 He, Y. et al. (2018) miR-149 in Human Cancer: A Systemic Review. J. Cancer 9, 375–388

19 Tate, A. et al. (2006) Met-Independent Hepatocyte Growth Factor-mediated regulation of cell adhesion in human prostate cancer cells. BMC cancer 6, 197–197

20 Hovanessian, A.G. et al. (2010) Surface Expressed Nucleolin Is Constantly Induced in Tumor Cells to Mediate Calcium-Dependent Ligand Internalization. PLOS ONE 5, e15787

21 Lago, S. et al. (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. Biochim Biophys Acta Gen Subj 1861, 1371–1381

22 Carvalho, J. et al. (2018) Fluorescent light-up acridine orange derivatives bind and stabilize KRAS-22RT G-quadruplex. Biochimie 144, 144–152

23 Burger, A.M. et al. (2005) The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. Cancer Res 65, 1489

24 Pereira, E. et al. (2017) Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA. Sci. Rep. 7, 42544

25 Mergny, J.-L. et al. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res. 33, e138–e138

26 Karsisiotis, A.I. et al. (2011) Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism. Angew Chem Int Ed Engl 50, 10645–10648

27 Gilbert-Girard, S. et al. (2017) Stabilization of Telomere G-Quadruplexes Interferes with Human Herpesvirus 6A Chromosomal Integration. J. Virol. 91,

28 Koirala, D. et al. (2011) A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. Nat Chem 3, 782–787

29 Park, J.Y. et al. (2018) Gemcitabine-Incorporated G-Quadruplex Aptamer for Targeted Drug Delivery into Pancreas Cancer. Mol Ther Nucleic Acids 12, 543–553

30 Incles, C.M. et al. (2004) A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. Mol Cancer Ther 3, 1201

31 Bates, P.J. et al. (2017) G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. Biochim Biophys Acta Gen Subj. 1861, 1414–1428

32 Shieh, Y.-A. et al. (2010) Aptamer-Based Tumor-Targeted Drug Delivery for Photodynamic Therapy. ACS Nano 4, 1433–1442

33 Bishop, J.S. et al. (1996) Intramolecular G-quartet Motifs Confer Nuclease Resistance to a Potent Anti-HIV Oligonucleotide. J Biol Chem. 271, 5698–5703

34 Reyes-Reyes, E.M. et al. (2015) Mechanistic studies of anticancer aptamer AS1411 reveal a novel role for nucleolin in regulating Rac1 activation. Mol Oncol 9, 1392–1405

35 Chang, T. et al. (2013) General cell-binding activity of intramolecular G-quadruplexes with parallel structure. PLOS ONE 8, e62348–e62348

Supplementary Information

RNA G-quadruplex as supramolecular carrier for cancer-selective delivery

***Santos, T.**; Pereira, P.; Campello, M.P.C.; Paulo, A.; Queiroz, J.A.; Cabrita, E. and Cruz, C. (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. Eur. J. Pharm. Biopharm. 142, 473–479. DOI: 10.1016/j.ejpb.2019.07.017 – IF: 5.571*

Results

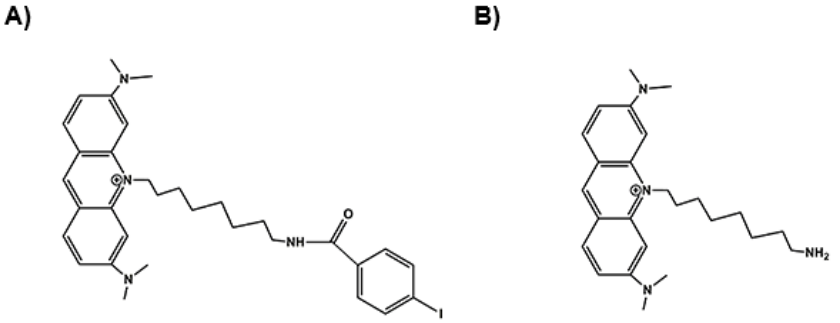

Figure S8.1. Chemical structure of **(A)** C_8 and **(B)** C_8-NH_2.

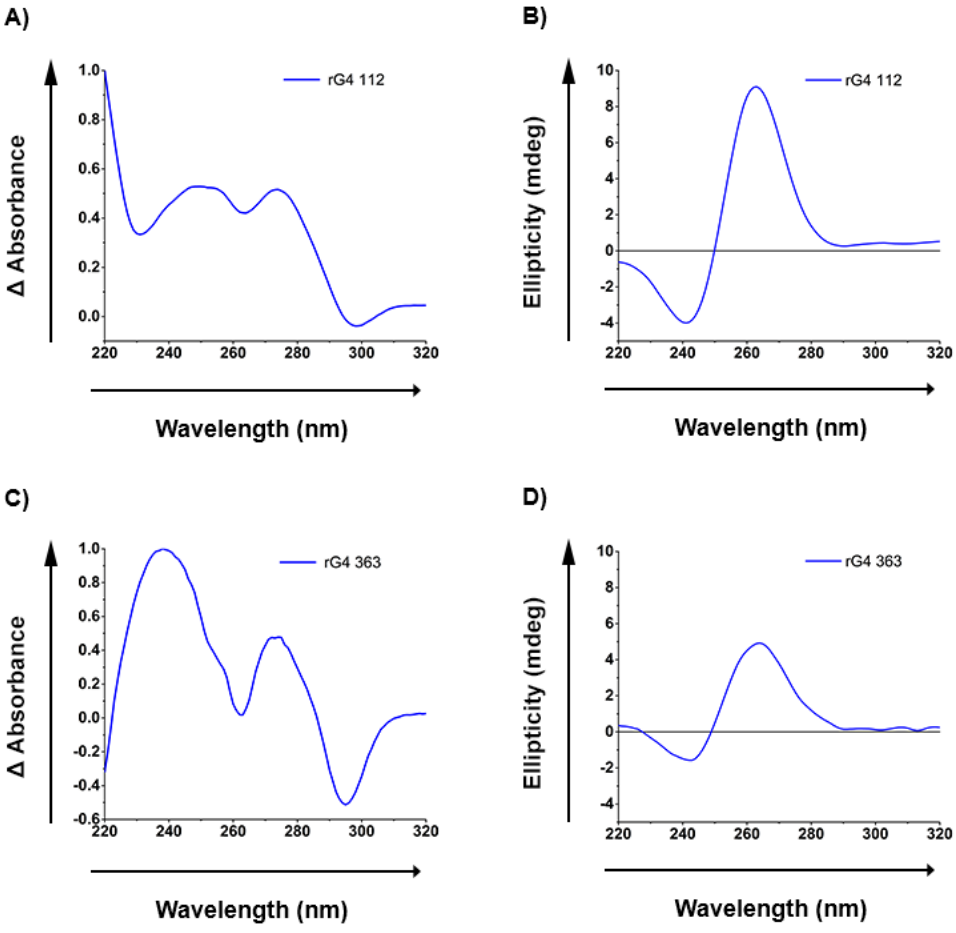

Figure S8.2. A) Thermal difference spectra (TDS) of **(A)** rG4 112 and **(C)** rG4 363 Loop and circular dichroism (CD) spectra of **(B)** rG4 112 and **(D)** rG4 363 loop at 20 °C and 10 μM. Spectra acquisition of rG4 112 and rG4 363 was performed in 10 mM lithium cacodylate and 0.1 or 5 mM KCl, respectively.

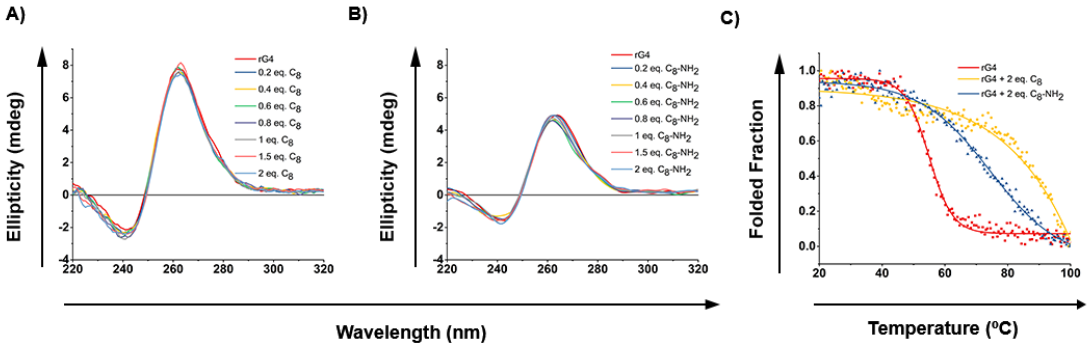

Figure S8.3. CD titration spectra of long loop (363) rG4 at 10 µM with increasing concentrations of **(A)** C_8 and **(B)** C_8-NH_2. **(C)** CD melting curves of long loop rG4 at 10 µM in the absence and presence of 2 molar equivalents of ligands. Data points were recorded at 263 nm. Temperature corresponds to the temperature set by the Peltier system. Spectra acquisitions were performed in 10 mM lithium cacodylate and 5 mM KCl.

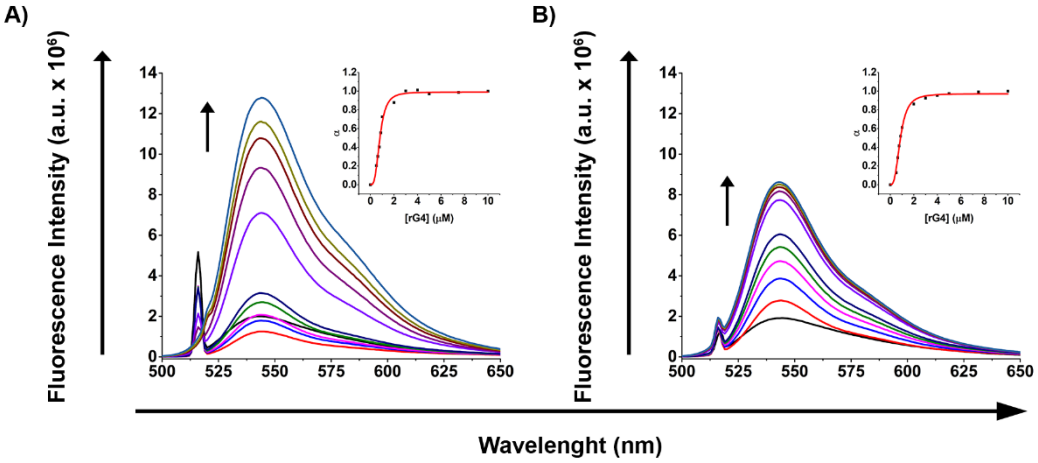

Figure S8.4. Fluorescence emission spectra of **(A)** C_8 and **(B)** C_8-NH_2 at 5 µM with different concentrations of long loop (363) rG4 in 10 mM lithium cacodylate buffer and 5 mM KCl. Insets: fraction of ligand bound plots fitted to the saturation binding equation.

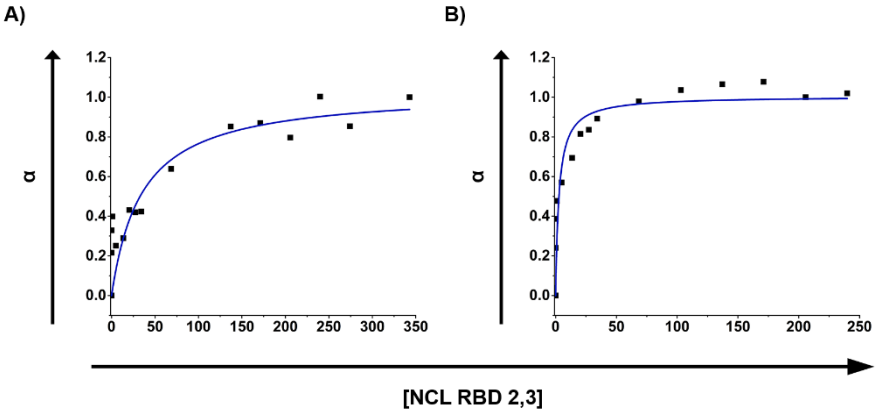

Figure S8.5. Fluorescence emission spectra of **(A)** long loop rG4-Cy3.5 and **(B)** rG4-Cy3.5-C_8 complex at 1 µM with different concentrations of NCL in 10 mM lithium cacodylate buffer and 0.1 mM KCl Insets: Plots of the fraction of bound nucleolin (NCL), fitted to the saturation binding equation.

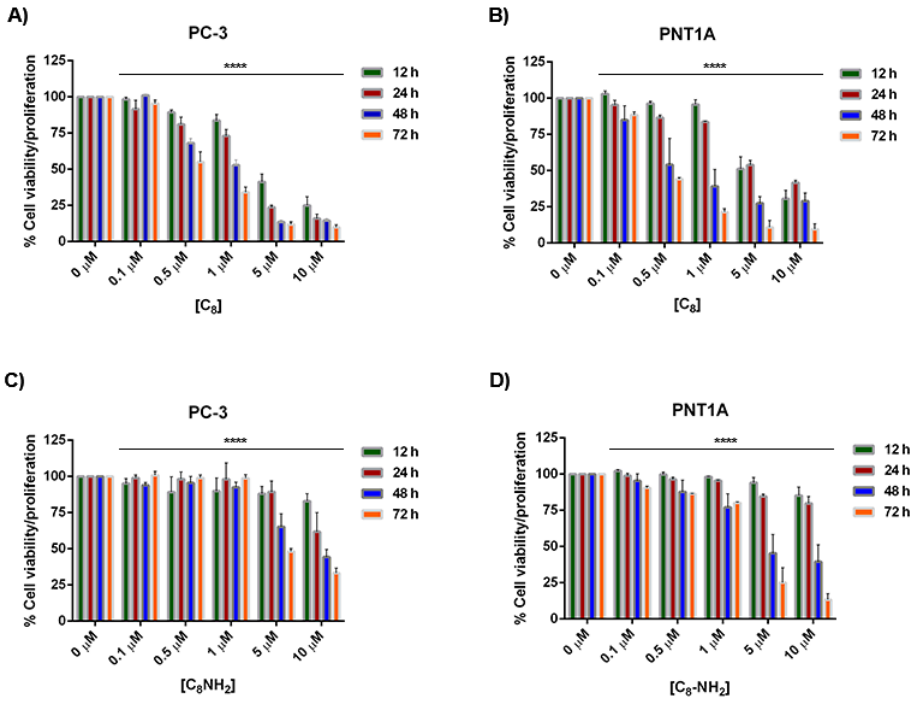

Figure S8.6. MTT assays conducted in PC-3 and PNT1A cell lines, which were incubated in the presence of increasing concentrations of C_8 **(A and B)** and C_8-NH_2 **(C and D)** for 12, 24, 48 and 72 h. Untreated cells were used as negative controls for cytotoxicity. Mean percentage values relative to untreated cells and standard error of the mean in 4 independent experiments are shown. The bars represent the mean and the lines represent the SEM associated. ****P<0.0001 (one-way ANOVA).

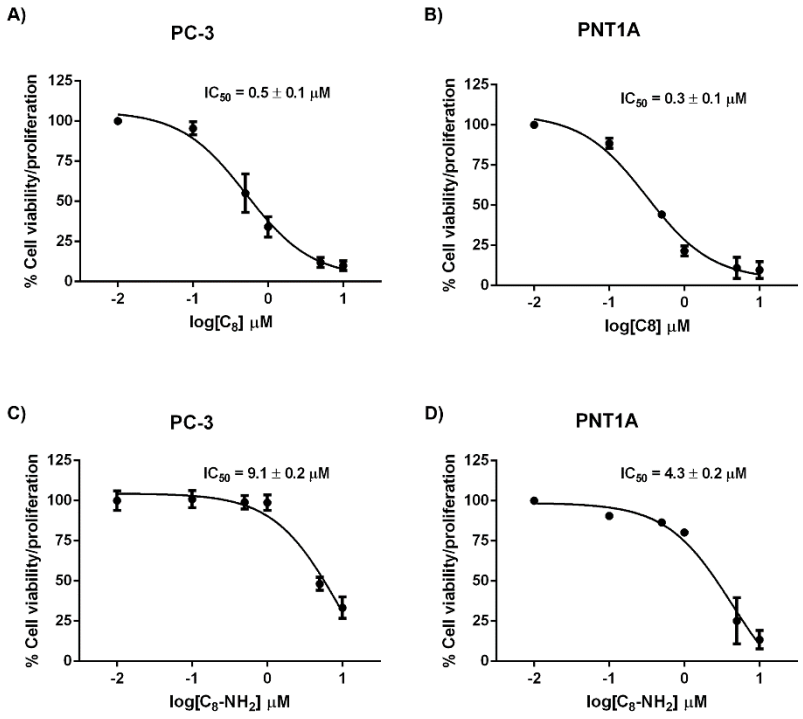

Figure S8.7. Dose-response data of cell viability measured by MTT assay after incubation of PC-3 and PNT1A cells with C_8 **(A and B)** and C_8-NH_2 **(C and D)** for 72 h.

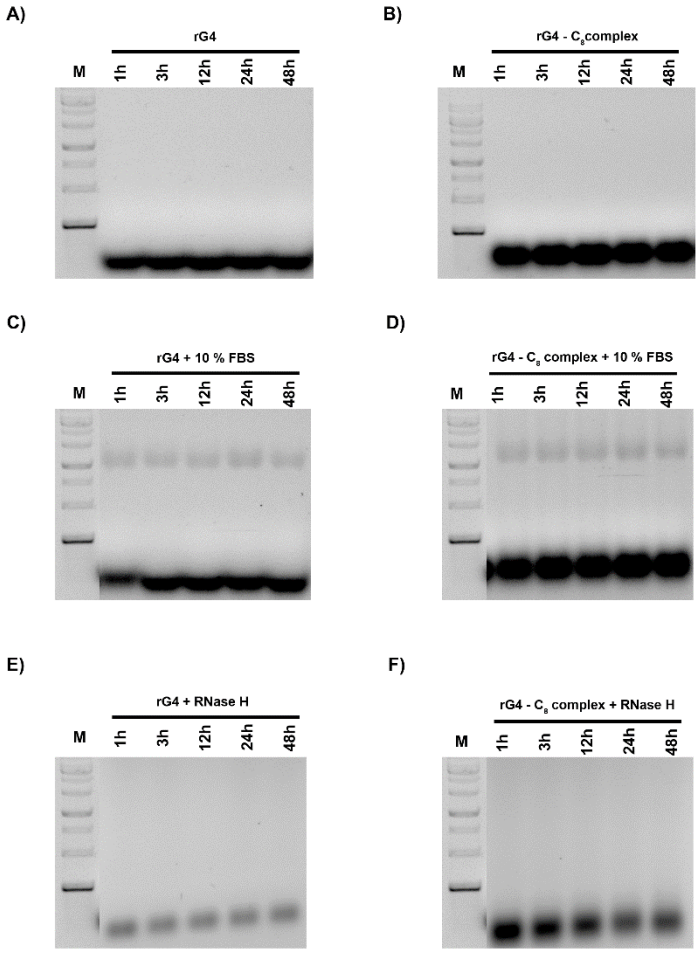

Figure S8.8. Thermostability of short loop rG4 and rG4-C_8 complex in 10 mM lithium cacodylate buffer and 0.1 mM KCl at 1, 3, 12, 24 and 48 h. The experiments were performed at 37 °C in the **(A)** and **(B)** absence, **(C)** and **(D)** presence of RPMI medium containing 10 % FBS and **(E)** and **(F)** presence of RNase H (0.3 U/µL) monitored by 1 % gel agarose electrophoresis.

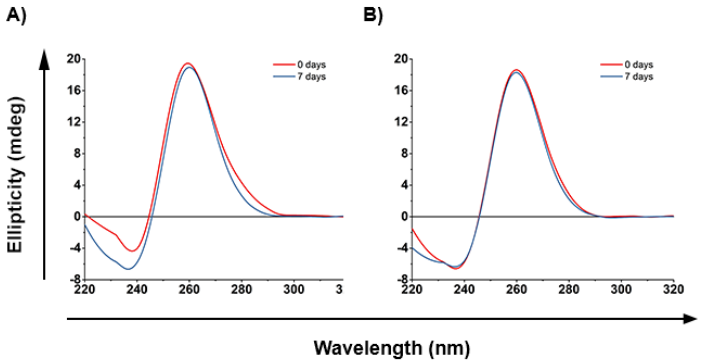

Figure S8.9. Circular dichroism spectra of **(A)** rG4 at 15 µM and **(B)** rG4-C_8 at 15 µM, respectively, after dissolution and incubation at 37 °C in RPMI medium with 10 % FBS without phenol red and 0.1 mM KCl for 7 days.

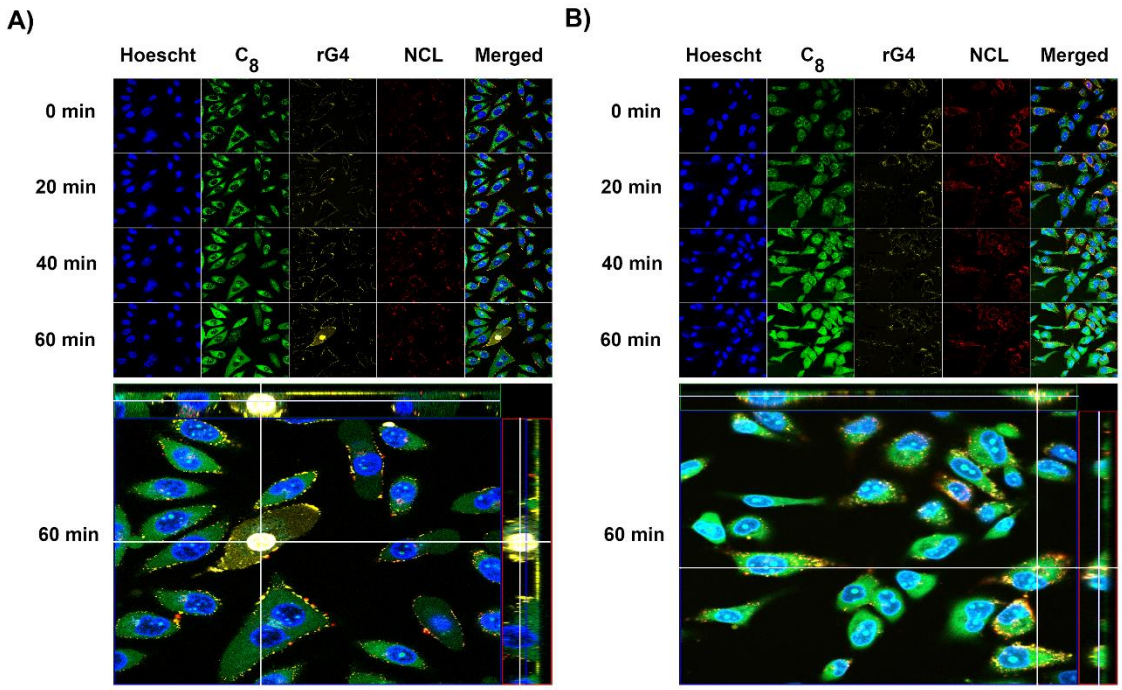

Figure S8.10. Confocal laser scanning microscopy images showing the intracellular uptake of the long loop rG4-C_8 complex from 0 to 60 min. For each panel, images showed the cells with nuclear staining by Hoechst 33342® (1 μM, blue); rG4 Cy3.5-C_8 complex (1:1 ratio, yellow); C_8 (1 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:1000). **(A)** Representative z-stacks data showing PC-3 cell line and **(B)** Representative z-stacks data showing PNT1A cell line.

Chapter 9

Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin

This chapter was published in:

***Santos, T.**; Miranda, A.; Imbert, L.; Jardim, A.; Caneira, C.R.F.; Chu, V.; Conde, J.P.; Campello, M.P.C.; Paulo, A.; Salgado, G.; Cabrita, E.J. and Cruz. C. (2022) Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin. Eur. J. Pharm. Sci. 169, 106093. DOI: 10.1016/J.EJPS.2021.106093 – IF: 4.384*

Chapter overview

In order to unravel the biological potential of G4s in pre-miRNAs, it is important to explore their potential in other relevant biomedical applications. Therefore, this chapter aims to characterize the full-length pre-miRNA 149 sequence. The binding to ligands and the ability to capture nucleolin was also investigated. To this end, a microfluidic approach was developed to detect and capture nucleolin. Altogether, the results revealed the detection of nucleolin in a concentration-dependent manner and could pave the way for future applications of pre-miRNA 149 G4.

Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin

Abstract

One of the most significant challenges in capturing and detecting biomarkers is the choice of an appropriate biomolecular receptor. Recently, RNA G-quadruplexes emerged as plausible receptors due to their ability to recognize with high-affinity proteins. Herein, we have unveiled and characterized the capability of the precursor microRNA 149 to form a G-quadruplex structure and determined the role that some ligands may have in its folding and binding capacity to nucleolin. The G-quadruplex formation was induced by K^+ ions and stabilized by ligands, as demonstrated by nuclear magnetic resonance and circular dichroism experiments. Surface plasmon resonance measurements showed a binding affinity of precursor microRNA 149 towards ligands in the micromolar range (10^{-5}–10^{-6} M) and a strong binding affinity to nucleolin RNA-binding domains 1 and 2 (8.38×10^{-10} M). Even in the presence of the ligand PhenDC3, the binding remains almost identical and in the same order of magnitude (4.46×10^{-10} M). The molecular interactions of the RNA G-quadruplex motif found in precursor miRNA 149 (5′-GGGAGGGAGGGACGGG- 3′) and nucleolin RNA-binding domains 1 and 2 were explored by means of molecular docking and molecular dynamics studies. The results showed that RNA G-quadruplex binds to a cavity between domains 1 and 2 of the protein. Then, complex formation was also evaluated through polyacrylamide gel electrophoresis. The results suggest that precursor microRNA 149/ligands and precursor microRNA 149/nucleolin RNA-binding domains 1 and 2 form stable molecular complexes. The *in vitro* co-localization of precursor microRNA 149 and nucleolin in PC3 cells was demonstrated using confocal microscopy. Finally, a rapid and straightforward microfluidic strategy was employed to check the ability of precursor microRNA 149 to capture nucleolin RNA-binding domains 1 and 2. The results revealed that precursor microRNA 149 can capture nucleolin RNA-binding domains 1 and 2 labeled with Fluorescein 5-isothiocyanate in a concentration-dependent manner, but PhenDC3 complexation seems to decrease the ability of precursor microRNA 149 to capture the protein. Overall, our results proved the formation of the G-quadruplex structure in the precursor microRNA 149 and the ability to recognize and detect nucleolin. This proof-of-concept study could open up a new framework for developing new strategies to design

improved molecular receptors for capture and detection of nucleolin in complex biological samples.

Keywords

pre-miRNA-149, RNA G-quadruplex, nucleolin, G-quadruplex ligands, molecular receptors

Introduction

In the last few years, several studies have investigated the ability of a significant number of biomolecular agents to selectively recognize cancer biomarkers [1,2]. Among the investigated bioreceptors for proteins in cancer, G-quadruplex (G4) molecules seem to be promising biosensing agents [3]. The G4s found in RNA, can adopt a parallel topology and exhibit greater stability, which makes them more suitable to be used as biosensing agents [4].

Precursor microRNAs (pre-miRNAs) adopting G4 structure have been described as structures with high stability, as previously showed for pre-miRNA-149 [5], pre-miRNA-92b [6] and pre-miRNA-26a [7]. The folding of the RNA into a G4 is dependent on the presence of specific cations in solution and could be modulated by the addition of stabilizing/destabilizing small molecules [8]. In the presence of K^+ ions and/or ligands that stabilize G4 formation, like pyridostatin (PDS), the pre-miRNA-149 adopts a G4 form, which prevents miRNA-149 biogenesis, and in absence of ions and ligands the biogenesis of miRNA-149 remains unaffected [5]. On the other hand, in the presence of tetra-(N-methyl-4-pyridyl)porphyrin (TMPyP4), the pre-miRNA-149 G4 is disrupted leading to a restoration of miRNA-149 biogenesis [9]. Recently, we have characterized the RNA G4 (rG4) region of pre-miRNA 149 (5' – GGGAGGGAGGGACGGG - 3') and pre-miRNA 92b (5' – GGGCGGGCGGGAGGG – 3') and proved its high affinity and specific interaction with nucleolin (NCL) Please see Chapters 6 and 8) [10,11].

NCL is a key protein involved in several cellular functions controlling different RNA and DNA metabolism components, including ribosome biogenesis, ribosomal RNA maturation, ribosomal DNA transcription, and chromatin structure [12,13]. Also, it has been reported that NCL is an accessory protein in miRNA biogenesis and controls the induction of apoptosis [14]. NCL is mainly located at the cell nucleolus and nucleoplasm [12], and it is also found in the cytoplasm and on the cell membrane, affecting many biological functions [15]. The overexpression of NCL and increased localization at the cell

membrane has been shown in several cancer types [16,17]. Its ability to bind molecules in the surface of cancer cells associated with its capability to internalize molecules, make NCL a promisor target for the development of novel specific agents.

The G4 aptamer AS1411 has been developed for targeting nucleolin (NCL) and showed promising activity against metastatic renal cell carcinoma and acute myeloid leukemia [18,19]. However, in the last years has also been employed as a biosensing and bioimaging agent to selectively detect nucleolin [20,21]. Ultrasensitive detection of NCL is of utmost importance in the diagnosis and prognosis of cancer [22]. In the past years, several proof-of-concept tools have been designed for the detection of NCL through simple, rapid, sensitive and specific applications [22]. Indeed, recent studies have documented the use of microfluidics to detect NCL in cancer cells based on anti-nucleolin AS1411 or locked nucleic acid (LNA) aptamers [3,21,23].

In this work, we have confirmed the formation of the RNA G4 in pre-miRNA-149 and its ability to recognize NCL by different computational and experimental methods. Overall, this study could foster the development of new molecular recognition agents that specifically detect NCL.

Materials and Methods

Bioinformatic analysis

pre-miRNA-149 sequences were obtained from miRBase version 22.1 [24]. The human pre-miRNA-149 sequence was analysed using QGRS-mapper algorithm [25] and G4RNA screener tool [26], which includes the scoring systems of G4Hunter [27], cGcC [28] and G4NN [29]. The default parameters were applied for QGRS-mapper analysis. For G4RNA screener tool analysis, the default parameters were applied, except for the size of the window that was defined to 89 nt. RNA structure prediction server (https://rna.urmc.rochester.edu/RNAstructureWeb/) was used to determine the predicted hairpin secondary structure of pre-miRNA-149. The pairwise alignment of pre-miRNA-149 sequences from several species was performed using Multiple Alignment Fast Fourier Transform (MAFFT v7 server - https://mafft.cbrc.jp/alignment/server/) with default options. The graphical alignment and conservation of the rG4 region were visualized with Jalview v2.11.1.3 software.

Oligonucleotides and ligands

All oligonucleotides were obtained from Eurofins (USA) with HPLC-grade purification. Oligonucleotide stock solutions of approximately 100 μM were prepared using DEPC water (Sigma-Aldrich, USA) and stored at -80 ºC until used. Annealing of oligonucleotide sequences was performed in 20 mM phosphate buffer (10 mM KH_2PO_4, 10 mM K_2HPO_4, pH 7.1) supplemented with 100 mM KCl by heating the samples for 10 min at 95 ºC and cooling on ice for 30 min before the experiments. Synthesis and purification of the ligands 10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium iodide (C_8) and 10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium (C_8-NH_2) were performed as previously described [30]. 3,3′-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate (PhenDC3) was obtained from Sigma-Aldrich (USA). The chemical structures of each ligand are depicted in **Figure 9.1A**. Stock solutions of the compounds were prepared as 10 mM solutions in DMSO (Thermo Fisher Scientific, USA) and their subsequent dilution was done using nuclease-free water. Fluorescein 5-isothiocyanate (FITC) was also purchased from Sigma-Aldrich (USA).

Cloning, cell-free expression, and purification

The sequence corresponding to NCL RNA binding domains 1 and 2 (NCL RBD1,2) was cloned into a pIVEX 2.4D vector. NCL RBD1,2 was synthesized *in vitro* using a cell-free expression system [31]. Briefly, NCL RBD1,2 was expressed in a volume of 9 mL under RNAse-free conditions in dialysis mode, with a 1/10 ratio of the reaction mixture to the feeding mixture for 16 h at 23 ºC under gentle agitation. The cell-free mixture contained 16 μg/mL of pIVEX 2.4D plasmid encoding the NCL RBD1,2 sequence, 1 mM of each essential amino acid, 0.8 mM of each rNTPs (guanosine-, uracil-, and cytidine-5'-triphosphate), 55 mM HEPES (pH 7.5), 68 μM folinic acid, 0.64 mM cyclic adenosine monophosphate, 3.4 mM dithiothreitol, 27.5 mM ammonium acetate, 2 mM spermidine, 80 mM creatine phosphate, 208 mM potassium glutamate, 16 mM magnesium acetate, 250 μg/mL creatine kinase, 27 μg/mL T7 RNA polymerase, 0.175 μg/mL tRNA, and 400 μL/mL S30 *E. coli* bacterial extract. After incubation, the reaction mixture was diluted in binding buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, and 10 mM imidazole) to a final volume of 45 mL and centrifuged for 45 min at 36,000 g at 4 ºC. Thereafter, the supernatant was applied onto a 5 mL Ni-NTA column that had been previously equilibrated in buffer binding at 4 ºC. The column was washed with 5 % of elution buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, and 500 mM imidazole) to eliminate residual contaminants and the protein was eluted with 50 % of elution buffer. Fraction containing

the NCL RBD1,2 was pooled and concentrated on a 10-kDa cut-off membrane. The purity of each fraction was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and the protein was identified through western blot analysis using the primary anti-NCL antibody (Thermo Fisher, ref. PA3-16875).

NMR spectroscopy

Standard ^{1}H NMR spectra with water suppression using excitation sculpting (acquired with zgesgp pulse program from the Bruker pulse program library) were acquired on a 600 MHz Bruker Avance III spectrometer with a QCI cryoprobe. Pre-miRNA-149 was used at a concentration of approximately 20 μM with a total volume of 200 μL in a 3 mm NMR tube, annealed as described above and supplemented with 10% D_2O (Eurisotop, France). The spectrum was acquired and processed with the software Topspin 3.1. Figures were prepared using TopSpin 4.0.6. Chemical shifts (δ) are reported in ppm.

Circular dichroism

CD spectra were acquired in a Jasco J-815 spectrometer (Jasco, USA), using a Peltier temperature controller (model CDF-426S/15). pre-miRNA-149 sequence was annealed as previously described. Unless otherwise stated, a 1 mm path-length quartz cuvette (Hellma, Germany) was used with pre-miRNA-149 at 10 μM in 10 mM lithium cacodylate buffer (Sigma-Aldrich, USA) at pH 7.2, supplemented with KCl (Thermo Fisher Scientific, USA). The required volume for the titrations was added directly in the quartz cell. The CD melting experiments were performed in the temperature range of 20-100 °C, with a heating rate of 2 °C/min by monitoring the ellipticity at 262 nm. Spectra acquisition was performed in the absence and presence of increasing concentrations of KCl or ligands. Data was converted into fraction folded (f) plots.

$$f = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \qquad \text{(Equation 9.1)}$$

Data points were then fitted to a Boltzmann or bi-dose response distribution equations (OriginPro 2016) and the melting temperatures were determined.

Molecular docking and molecular dynamics simulations

The minimized model of the rG4 found in pre-miRNA-149 (5'-GGGAGGGAGGGACGGG-3) was recently built and described by some of us [32], was used as a target for molecular docking studies. The rG4 structure and the NCL RBD1,2 (PDB: 2KRR) solution structure

were optimized for docking using the Dock Prep tool of Chimera 1.15. After assigning polar hydrogens and Gasteiger charges, docking simulations were carried with AutoDock 4.2.6 program using Lamarckian genetic algorithm. The size of the box was constrained to 100 × 100 × 100 Å along the ×, y, and z axes, respectively, with a grid spacing of 0.375 Å. A total of 10 runs were performed with an initial population of 150 random individuals, a maximum number of evaluations set to 2.5×10^7, the rate of mutation and crossover set to 0.02 and 0.8, respectively, and an elitism value of 1. The most representative structures were selected based on the binding free energy and further processed with MD simulations the all-atom force field AMBER99SB of GROMACS 2020.6 package.

For MD simulations, the rG4/NCL RBD1,2 complex was centered in an octahedral box and solvated with TIP3P water molecules and K atoms to neutralize the system. After an energy minimization of 1000 steps using the steepest descent algorithm, the system was gradually heated to 300 K in 100 ps under the control of Berendsen thermostat followed by 100 ps isobaric simulation under the control of the Parrinello-Rahman barostat. Finally, explicit solvent MD runs of 100 ns were performed for rG4/NCL RBD1,2 complex at constant temperature (300 K) and pressure (1 atm). The Particle mesh Ewald (PME) method was used for calculating long-range electrostatic interactions. A 1 nm cut-off was applied to short-range Lennard-Jones interactions. Coordinates were collected in trajectory files every 10 ps. All molecular images were rendered using UCSF ChimeraX. 2D diagrams of G4/NCL RBD1,2 complex were determined by using LigPlot+ v.2.2.4.

Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel (15%) electrophoresis was used to visualize the oligonucleotides. pre-miRNA-149 samples were prepared at a concentration of approximately 1 μM. Sucrose (Sigma-Aldrich, USA) was added to the samples at a final concentration of 23%. The oligonucleotide marker was loaded in parallel on the gel. The pre-miRNA-149 sequence was injected with and without KCl. A molar ratio of 1:1 of pre-miRNA-149/ligand and pre-miRNA-149/NCL RBD1,2 was prepared, and the mixture was incubated for 30 min. The supramolecular complexes pre-miRNA-149/ligand/NCL RBD1,2 were prepared at a molar ratio of 1:1:1. Electrophoresis was performed at 2 W per gel to reach a temperature close to 20°C. After electrophoresis, the gel was stained with Stains-All solution (Sigma-Aldrich, USA) for 30 min under gentle agitation. Thereafter, the gel was incubated for 2 h in water for discoloration under the light before visualization.

Surface Plasmon Resonance (SPR)

All SPR measurements were carried out on a Biacore T200 (Biacore, GE Healthcare, Sweden) with a SA (streptavidin-coated sensor chip) sensor chip (GE Healthcare, Sweden). The biotin-labeled pre-miRNA-149 sequence (25 nM) was dissolved in 20 mM phosphate buffer (10 mM KH_2PO_4, 10 mM K_2HPO_4, pH 7.1) supplemented with 100 mM KCl and annealed as above described. Meanwhile, the sensor chip was equilibrated with running buffer at 25 µL/min during 1 h. Activation buffer (1 M NaCl, 50 mM NaOH) was injected for 3 min seven times to remove unbound streptavidin from the sensor chip. After that, in order to ensure surface stability, two primes with running buffer were performed and the buffer was flowed for 10 min at flow rate of 1 µL/min. Manual inject was used to immobilize biotin-labeled pre-miRNA-149 on flow cell 2 (100 µL of a 25 nM pre-miRNA-149) and the injection was stopped after the desired level is reached (~ 250 RU to minimize mass transport effects). Flow cell 1 was used as a control and is left blank for subtraction.

For kinetic/affinity analysis, each ligand and NCL RBD1,2 were serially diluted in running buffer. All experiments were performed in triplicate at 25°C. Each ligand and NCL RBD1,2 were injected from low to high concentrations during 75 s with a flow rate of 50 µL/min, followed by dissociation of 600 s. Surface regeneration was achieved by injecting two pulses of 30 s of 10 mM glycine/HCl pH 2.5, and next three 60 s injections of running buffer to remove any trace of regeneration solution.

BiaEvaluation Software was used for data analysis and the likelihood of fittings was assessed through the statistical parameters of Chi^2 and U-value. All sensorgrams were double corrected for non-specific binding and refractive index changes (bulk effect) by subtracting the signals of an equivalent injection across the reference flow cell 1. The equilibrium dissociation constant K_D was obtained by plotting the response at equilibrium (Req) versus the analyte concentration and then fitted as a Langmuir isotherm:

$$Req = \frac{[Analyte]Rmax}{K_D + [Analyte]} + RI \qquad \text{(Equation 9.2)}$$

Rmax is the analyte binding capacity and RI is the bulk refractive index contribution of the sample, which is assumed to be the same for all the injections and used as the Response-axis offset.

Fluorescein isothiocyanate (FITC) labelling

FITC was used to attach a fluorescent label to NCL RBD1,2 via the amine group. The isothiocyanate group reacts with amino terminal and primary amines in NCL RBD1,2. Briefly, 2 mg/mL of NCL RBD1,2 was dissolved in 0.1 M sodium bicarbonate buffer (pH 8.3). Then, FITC dissolved in DMSO (10 mg/mL) was added to the solution. The reaction mixture was incubated for 4 h with continuous stirring. To remove the unbounded FITC, the mixture was washed on a 10-kDa cut-off membrane. The labelling of pre-miRNA-149 with the FITC dye was performed using 5 µg of pre-miRNA-149. The mixture was added to 71µL of labelling buffer (0.020 g of sodium (di)tetraborate in 1 mL of H_2O) and 2 µL of FITC (100 mg of FITC in 200 µL of sterile DMSO). Thereafter, the solution was stirred for 4h at RT, protected from light. Finally, pre-miRNA-149 was incubated with 3 M NaCl and 2.5 volumes of absolute ethanol and incubated overnight at -20 ºC. Excess FITC was removed by centrifugation at 12000g for 30 min at 4ºC. After centrifuging, the pellet was washed with 75% ethanol twice, followed by a 5 min centrifugation at 12000 g (4ºC).

Confocal Microscopy

PC3 cell line was grown in RPMI medium supplemented with 10% (v/v) FBS and 1% (v/v) penicillin-streptomycin. Cultures were maintained in a humidified chamber at 37 °C and 5% CO2. Cells were subsequently harvested and seeded in µ-Slide 8-well flat bottom imaging plates (Ibidi GmbH, Germany) at a plating density of 5×10^4 cells/well and incubated for cell attachment for 24 h at 37 °C and 5% CO_2. Thereafter, cells were incubated with the primary anti-NCL polyclonal antibody (Thermo Scientific, USA) for 2 h at 37 °C. Following primary antibody incubation, cells were washed 3× with fresh serum-free medium and incubated with secondary antibody anti-rabbit IgG conjugated with Alexa Fluor® 647 (Thermo Scientific, USA) for 1 h at 37 °C. Thereafter, cells were washed 3 × with fresh serum-free medium and stained with Hoechst 33342® nuclear probe (1 µM) for 10 min. Subsequently, the cells were incubated with pre-miRNA-149 (10 µM) labeled with FITC and incubated for 12h at 37 °C and 5% CO_2. Thereafter, the cells were transferred to a Zeiss LSM 710 confocal laser scanning microscope (CLSM; Carl Zeiss SMT Inc., USA) equipped with a plane-apochromat 63×/DIC objective and processed in Zeiss Zen (SP2, 2010), in order to evaluate the cellular uptake. The fluorescence images were obtained at 63× magnification.

Microfluidic experiments

Polydimethylsiloxane (PDMS) microfluidic devices were fabricated using standard mold replication techniques, as previously described [33]. The device was designed to have two-channel heights to allow the holding of the beads inside the channel. The packing of streptavidin beads was carried out as previously detailed [33]. Briefly, the liquids were driven inside the microchannel using a syringe pump exerting a negative pressure at the outlet (NE-1002X, New Era Pump System Inc., USA). The beads were first suspended in a solution of 20% (w/w) polyethylene glycol 8000 (PEG) (Sigma-Aldrich, USA) and packed inside the channel at a flow rate of 7 µL/min, followed by a washing step with 20 mM phosphate buffer (10 mM KH_2PO_4, 10 mM K_2HPO_4, pH 7.1) supplemented with 100 mM KCl at a flow rate of 7 µL/min.

Thereafter, the pre-miRNA-149 (5 µM) was incubated with PhenDC3 at 1 molar equivalents during 10 min. After that, different concentrations of NCL RBD1,2 labeled with FITC were added and incubated for 30 min. The experiment was carried out at a flow rate of 1 µL/min for 30 min. Finally, the channels were washed with 20 mM phosphate buffer (10 mM KH_2PO_4, 10 mM K_2HPO_4, pH 7.1) supplemented with 100 mM KCl at a flow rate of 5 µL/min for 2 min to remove non-specifically bound molecules. All fluorescence images were acquired with an exposure time of 2 s in an Axio Imager Z2 microscope using 579 nm as the excitation wavelength (BP 565/30) and collecting the emission at 591 nm (BP 620/60) and processed in Zeiss Zen Software (SP2, 2010) and ImageJ (National Institute of Health, USA). Fluorescence emission values were obtained by averaging the entire end-section of the micro-columns. For every individual experiment, new structures with fresh functionalization steps were performed. ImageJ was used to analyze the fluorescence images.

Results

Bioinformatic analysis of pre-miRNA-149 G4

The ability of pre-miRNA-149 to adopt a G4 structure was investigated by means of an *in silico* approach using the QGRS-mapper algorithm [25] and the G4RNA screener tool [26], which includes the scoring systems of G4Hunter [27], cGcC [28] and G4NN [29]. The results highlighted G4 scores of 41, 0.5506, 2.5333 and 0.2407 for QGRS-mapper, G4Hunter, cGcC and G4NN, respectively (**Figure 9.1B**). The G4 sequence in pre-miRNA-149 (**Figure 9.1C**) identified by the QGRS-mapper algorithm was 5'-GGGAGGGAGGGACGGG-3'. Next, the conservation of the G4 sequence in several species was evaluated by a pairwise alignment of pre-miRNA-149 sequences using

Multiple Alignment Fast Fourier Transform (MAFFT v7 server - https://mafft.cbrc.jp/alignment/server/). The results showed that the rG4 region of pre-miRNA-149 is highly conserved among the different species (**Figure 9.1D**).

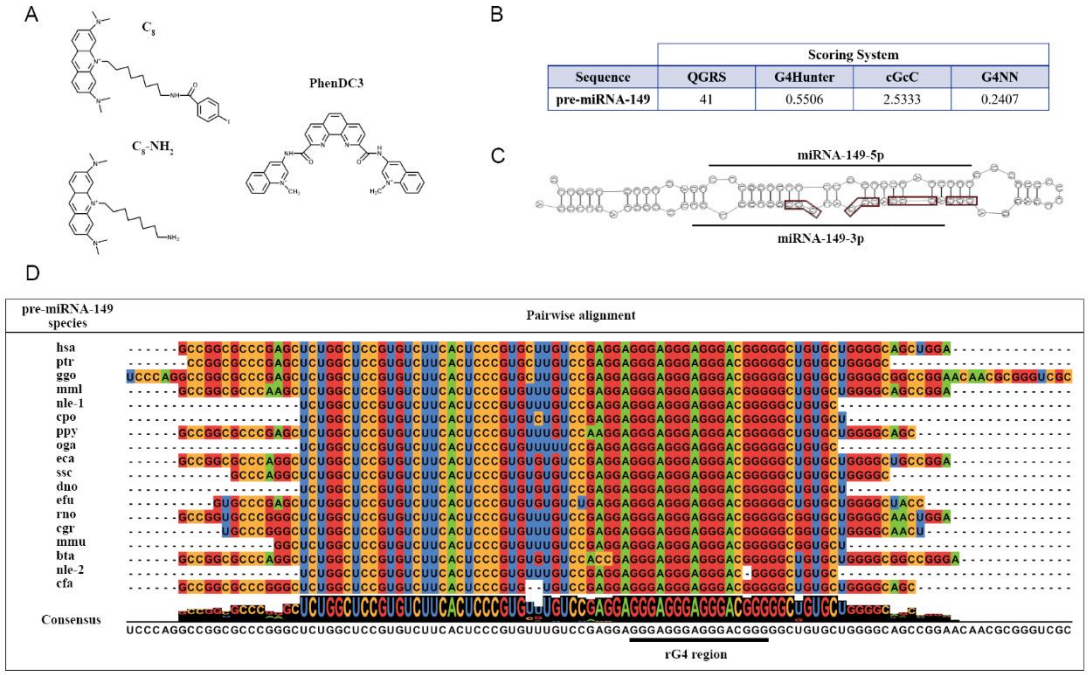

Figure 9.1. (A) Chemical structures of C_8, C_8-NH_2 and PhenDC3. **(B)** G4 scores of pre-miRNA-149 predicted by using QGRS-Mapper, G4Hunter, cGcC and G4NN. **(C)** Predicted secondary structure of pre-miRNA-149 showing miRNA-149-5p, miRNA-149-3p and the rG4 region. **(D)** Pairwise alignment and conservation of pre-miRNA-149 sequences found in several species.

pre-miRNA-149 G4 formation

The ability of pre-miRNA-149 to form a G4 structure in the presence of K^+ was assessed by means of 1D ^{1}H NMR spectroscopy and CD measurements (**Figure 9.2**). As seen in **Figure 9.2A**, the 1D ^{1}H NMR spectrum revealed proton resonances in both the G4 (10-12 ppm) and Watson–Crick (12-14.5 ppm) regions. The resonances in the Watson-Crick region are sharp and well resolved, while we observe intense overlapped signals in the G4 imino proton region. Therefore, in order to assess the formation and stability of the G4 form by pre-miRNA-149, we performed CD measurements to optimize the annealing of the G4 structure. The results depicted in **Figure 9.2B** showed that upon increasing temperature to 95 ºC and cooling the sample in ice, the positive band with maximum ellipticity at 265 nm exhibits a slight shift to 261 nm. The structure also remains stable throughout 24h and 48h of incubation at 4 ºC.

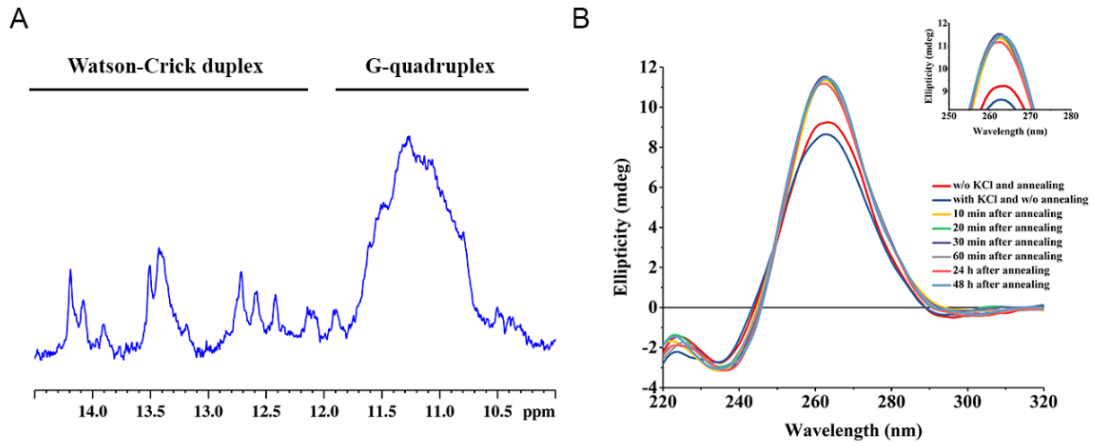

Figure 9.2. (A) 1D ^{1}H spectra of pre-miRNA-149 in the presence of K^+ showing the spectrum region of the Watson-Crick and G4 structures. **(B)** CD spectra of pre-miRNA-149 before and after annealing and in the absence and presence of K^+. Inset: Zoom of the maximum positive band.

Evaluation of K^+ dependence in the formation of pre-miRNA-149 G4

CD measurements of the pre-miRNA-149 in the absence and presence of increasing amounts of KCl were performed to further check the influence of K^+ ions in the formation of the G4 structure by pre-miRNA-149. As seen in **Figure 9.3A**, upon titration with KCl of the pre-miRNA-149 solution, a noticeable increase in the band intensity was observed after addition of 10 mM KCl. The CD profile was conserved and is characteristic of a parallel G4 topology. In fact, it is possible to notice that the band intensity is K^+ dependent. These K^+-dependent amplitude changes in CD bands are indicative of G4 formation. In order to investigate the thermal stability of pre-miRNA-149 in the presence of K^+, CD melting experiments were carried out. As expected and highlighted in **Figure 9.3B** and **Table S9.1,** upon titration with increasing amounts of KCl, the thermal stability of pre-miRNA-149 increases from 57.6 ± 0.2 ºC to 76.4 ± 1.3 ºC in the absence and the presence of 20 mM KCl, respectively. In the presence of 50 mM and 100 mM of KCl, the melting curves could not be fitted by the Boltzmann equation due to high thermal stabilization.

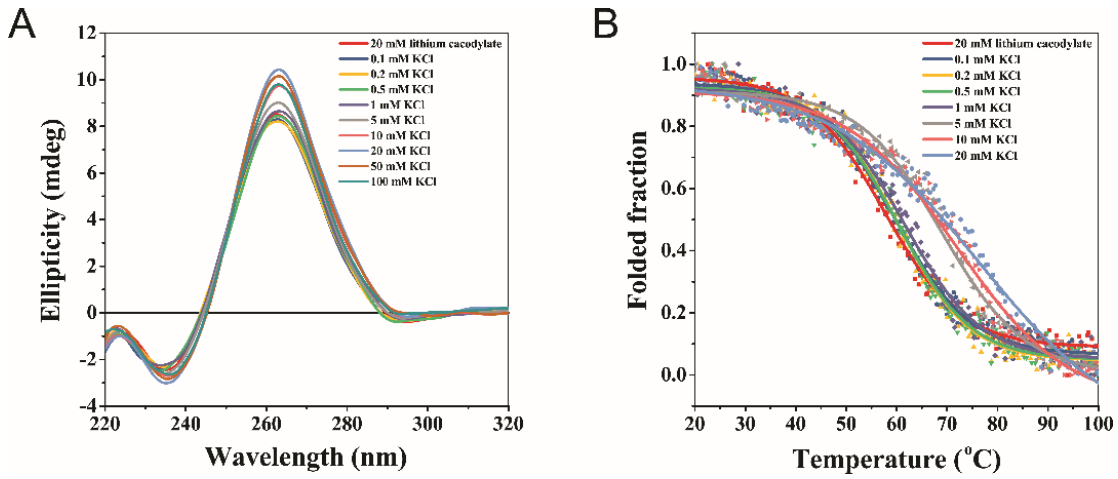

Figure 9.3. (A) CD spectra of pre-miRNA-149 in 20 mM lithium cacodylate buffer in the absence and presence of increasing concentrations of KCl. **(B)** CD melting curves at 10 μM in the absence and presence of increasing concentrations of KCl. Data points were recorded at 262 nm.

Effect of G4 ligands on pre-miRNA-149 stability

The (de)stabilization of the G4 formed by pre-miRNA-149 could be significantly modulated by using ligands that decrease or increase the partially folded states of the G4 structure. Herein, we studied the effect of acridine orange (C_8 and C_8-NH_2) and phenanthroline (PhenDC3) derivatives on pre-miRNA-149. The influence of these ligands on the G4 structure and stability was investigated using CD titrations and CD melting experiments, respectively. The results revealed important findings and are highlighted in **Figure 9.4**. CD titrations demonstrated 371 in the band intensity upon ligand addition. The increase in the band intensity was more prominent in C_8 (**Figure 9.4A**) and PhenDC3 (**Figure 9.4C**) ligands. C_8-NH_2 (**Figure 9.4B**) also demonstrated an increase in the band intensity but less prominent.

CD melting experiments were conducted in order to quantify the thermal stabilization of pre-miRNA-149 by ligands (**Figure 9.4D-E** and **Table S9.2**). The results agree with those observed in CD titration experiments. The initial melting temperature of pre-miRNA-149 in the presence of 20 mM lithium cacodylate buffer supplemented with 1 mM KCl was 61.9 ± 0.2 °C. Upon addition of 2 molar equivalents of C_8-NH_2, the melting temperature was only 63.4 ± 0.4 °C. In the presence of 2 molar equivalents of C_8, the signature of melting curves suggests the presence of two pre-miRNA-149 species, since two distinct melting points were obtained from data (T_m^1 = 64.5 ± 1.9 °C and T_m^2 = 96.3 ± 2.5 °C). In the presence of 2 molar equivalents of PhenDC3, the melting temperature decreased to 58.9 ± 0.6 °C.

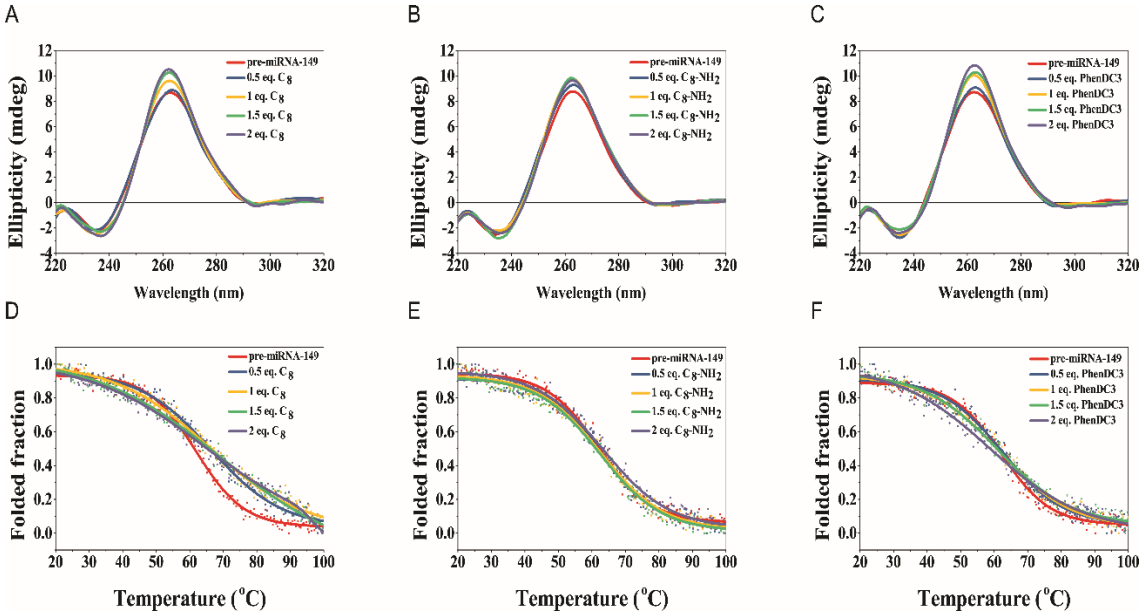

Figure 9.4. (A) CD spectra of pre-miRNA-149 in 20 mM lithium cacodylate and 1 mM KCl in the absence and presence of increasing concentrations of **(A)** C_8, **(B)** C_8-NH_2 and **(C)** PhenDC3. CD melting normalized curves at 10 μM in the absence and presence of increasing concentrations of **(D)** C_8 and (E) C_8-NH_2. CD melting curve without normalization at 10 μM in the absence and presence of increasing concentrations of (F) PhenDC3. Data points were recorded at 262 nm.

In silico analysis of the interaction between NCL RBD1,2 and G4 motif found in pre-miRNA-149

The interaction of the rG4 found in pre-miRNA-149 (5'-GGGAGGGAGGGACGGG- 3') with NCL RBD1,2 was assessed by molecular docking and molecular dynamics (MD) simulations. Some of us described the predicted model of the G4 structure recently (Please see Chapter 7) [32]. Considering that the solution structure of the full-length NCL remains unknown, we have used the available solution structure of NCL RBD1,2 (PDB: 2KRR) for the *in silico* analysis. The best docking conformer in terms of the lowest free binding energy was used to perform the molecular dynamics studies. The complex G4/NCL RBD1,2 after 100 ns simulation was depicted in **Figure 9.5A** (side and top view). MD simulations showed that rG4 structure binds to NCL RBD1,2 in the binding cavity between RBD1 and RBD2. **Figure 9.5A** shows the RMSD fluctuations of the G4/protein, protein, and G4 for complex G4/NCL RBD1,2. The visual analysis of the trajectories proved that the structures were stable throughout the simulation (100 ns). Although an initially noticeable RMSD fluctuation of the G4/protein and protein, the RMSD value seems to stabilize after 40 ns of simulation. Additionally, intermolecular hydrogen bonding is one of the main driving forces contributing to macromolecule and ligand binding. Hydrogen bonds observed between NCL RBD1,2 and G4 structure are shown in **Table S9.3**. Notably, most G4 residues act as hydrogen acceptors in the

intermolecular hydrogen bonds except A4 and G5 residues, which act as hydrogen donors (**Figure 9.5C**).

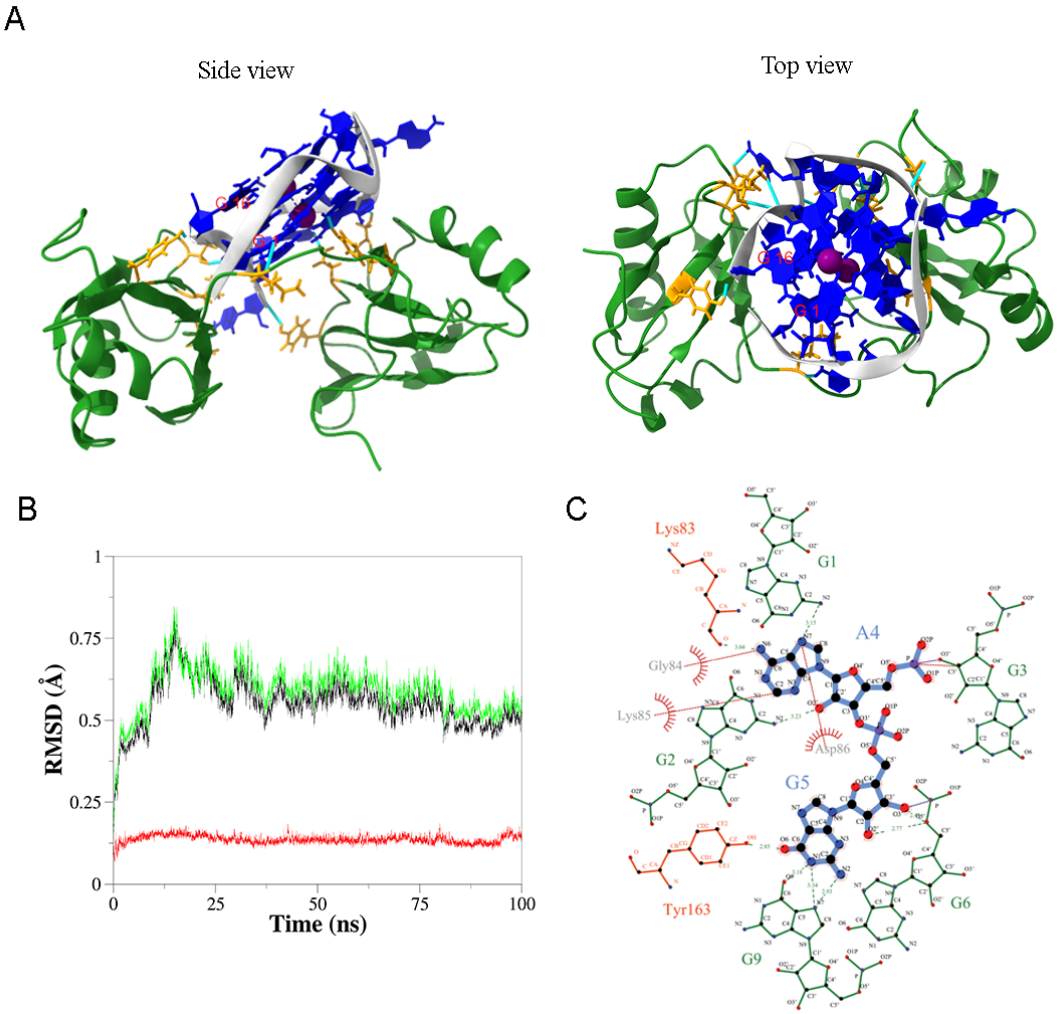

Figure 9.5. *In silico* analysis of G4/NCL RBD1,2 complex. **(A)** Final snapshots of side and top view of the G4/NCL RBD1,2 complex after 100 ns MD production. G4 structure is depicted in blue with the backbone in light gray. K^+ cations are depicted as purple spheres. NCL RBD1,2 is represented in green and contact residues of the protein are highlighted in orange. H-bonds are shown in cyan, while labels are shown in red. **(B)** RMSD plot of the G4/protein (green), protein (black), and G4 (red) for complex G4/NCL RBD1,2. **(C)** Ligplot + image showing site-specific hydrogen bonding and hydrophobic interactions of NCL RBD1,2 residues and rG4 residues A4 and G5. Hydrogen bonds are depicted in dashed green lines and hydrophobic interactions in red.

Evaluation of the formation of the pre-miRNA-149/ligand, pre-miRNA-149/NCL and pre-miRNA-149/ligand/NCL molecular complexes

The complex formation of the pre-miRNA-149 and pre-miRNA-149/ligands with NCL was determined by PAGE (**Figure 9.6**). The oligonucleotides were diluted to 1 µM in 20 mM phosphate buffer, in the presence and absence of 100 mM KCl. The complex with

ligands/NCL was prepared in a 1:1 ratio. The gel showed an intense band in the presence of 20 mM phosphate buffer, suggesting a major dominant folding form in solution (**Figure 9.6A, lane 1**). Since the samples were annealed before their loading on the gel, it is possible that the annealing could have promoted the formation of the G4 structure with 20 mM phosphate buffer. In the presence of 100 mM KCl, the pre-miRNA-149 seems to adopt the same electrophoretic profile as in the presence of 20 mM phosphate buffer (**Figure 9.6A, lane 2**). Considering these results, we proposed that the dominant folding form in solution could be the G4 form of pre-miRNA-149. Thereafter, to evaluate the effect of the ligands in pre-miRNA-149 molecularity, we incubated the pre-miRNA-149 sequence with ligands. The results revealed that pre-miRNA-149 in the presence of the ligands retains its initial molecularity (**Figure 9.6A, lanes 3,4 and 5**), which corroborates that the dominant folding form in solution is the G4 form. Upon addition of NCL RBD1,2, it is worth noticing that the band corresponding to pre-miRNA-149 was less intense (**Figure 9.6B, lane 2**), suggesting the binding of almost all pre-miRNA-149 G4 to NCL RBD1,2. An intense band corresponding to the complex pre-miRNA-149/NCL RBD1,2 was visible at the high molecular weight. A similar electrophoretic profile of pre-miRNA-149 was observed in the presence of G4 ligands and NCL RBD1,2 (**Figure 9.6B, lanes 3,4 and 5**).

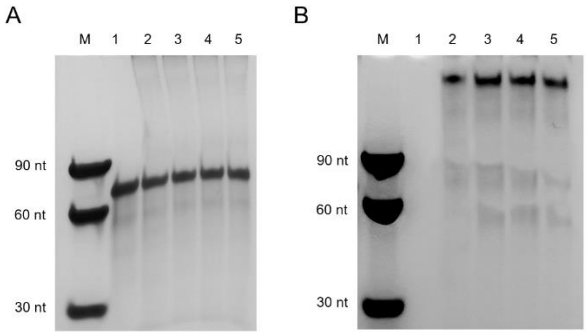

Figure. 9.6. PAGE under native conditions of pre-miRNA-149 in **(A)** absence (Lane M - migration marker; lane 1 – pre-miRNA-149 in the presence of 20 mM K_2HPO_4/KH_2PO_4; lane 2 – pre-miRNA-149 in the presence of 20 mM K_2HPO_4/KH_2PO_4 supplemented with 100 mM KCl; lanes 3 to 5 – pre-miRNA-149/ligand complexes (with 1 molar eq. of C_8, C_8-NH_2 and PhenDC3, respectively) and in the **(B)** presence of NCL RBD1,2 (1 μM) (Lane M - migration marker; lane 1 – NCL RBD1,2 (1 μM); lane 2 – pre-miRNA-149 in the presence of 20 mM K_2HPO_4/KH_2PO_4 supplemented with 100 mM KCl; lanes 3 to 5 – pre-miRNA-149/ligand complexes (with 1 molar eq. of C_8, C_8-NH_2 and PhenDC3, respectively). The gel was supplemented with 10 mM KCl. Migration markers were oligothymidylate single-stranded DNA (n = 30, 60 and 90).

Binding affinity of pre-miRNA-149 and pre-miRNA-149/ligand complexes towards NCL RBD1,2

SPR biosensor was used to investigate the interaction of pre-miRNA-149 and ligands and the binding affinity of pre-miRNA-149 or pre-miRNA-149/ligands towards NCL RBD1,2. Streptavidin sensor chips were used to capture biotin-labeled pre-miRNA-149. The SPR signal responses related to the specific interaction with the pre-miRNA-149 were obtained after subtraction of the signals recorded on the reference flow-cell as well as the running buffer injection by applying a double referencing procedure. The binding curves demonstrated a fast association phase and a slow dissociation, which made it necessary to include glycine pH 2.5 regeneration steps. Thereafter, the SPR signals returned to the baseline. The K_D was obtained by fitting the steady-state response *vs*. the analyte concentration by Langmuir isotherm according to a 1:1 binding stoichiometry (**Figure S9.1**). The values are depicted in **Table 9.1**. For all ligands, the binding affinities were in the concentration range 10^{-5}–10^{-6} M. On the other hand, the binding affinities of pre-miRNA-149 and pre-miRNA-149/NCL RBD1,2 were 8.38×10^{-10} M and 4.46×10^{-10} M, respectively.

Table 9.1. K_D (M) constant values of pre-miRNA-149 measured by SPR biosensor and obtained by the Langmuir isotherm.

Ligand	K_D **(M)**
C_8	9.29×10^{-6}
C_8-NH_2	1.12×10^{-5}
PhenDC3	7.94×10^{-6}
NCL RBD1,2	8.38×10^{-10}
PhenDC3/NCL RBD1,2	4.46×10^{-10}

Co-localization of NCL with pre-miRNA-149 in cancer cells

Once the binding affinity and influence of the G4 ligands have been characterized, the recognition of NCL by pre-miRNA-149, labeled with FITC, was performed in a cellular context via confocal microscopy. Nuclei are stained with Hoechst 33,342® and are displayed in blue. The primary anti-NCL antibody conjugated with the secondary antibody AlexaFluor 647® (red) was used to localize cell surface NCL. Pre-miRNA-149 was labeled with FITC (green) to allow its location in cancer cells. As depicted in **Figure 9.7**, pre-miRNA-149 was able to co-localize with NCL-positive regions. The

cytoplasm and surface of the cells in the merged image were depicted in orange, which means the merge of red and green colors from anti-nucleolin antibody and pre-miRNA-149-FITC, respectively. In addition, inside some cells, we found free pre-miRNA-149, which probably means the internalization of pre-miRNA-149 and its release from nucleolin inside the cell. These promisor results confirmed the potential of pre-miRNA-149 G4 as a molecular recognition agent to detect cell surface NCL in cancer cells.

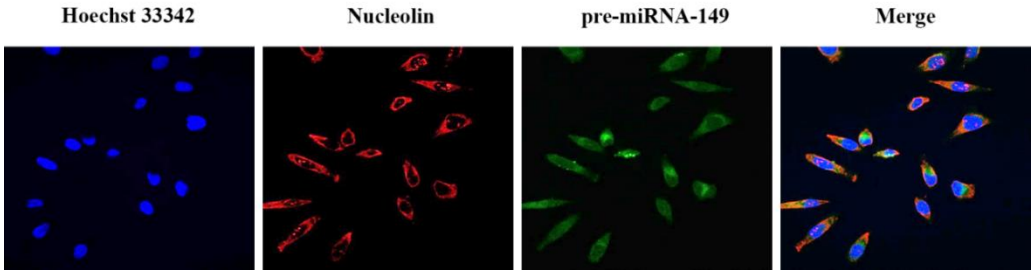

Figure 9.7. Confocal microscopy images showing co-localization of NCL and pre-miRNA-149, labeled with FITC. The images showed the cells with nuclear staining by Hoechst 33342® (1 μM, blue); pre-miRNA-149 labeled with FITC (10 μM, green); and NCL (red). NCL was labeled with the primary anti-NCL polyclonal antibody (1:100) and detected with the secondary antibody against IgG conjugated with Alexa Fluor® 647 (1:1000).

Capture of NCL RBD1,2 by a pre-miRNA-149-based microfluidic platform

In order to confirm the ability of pre-miRNA-149 to capture NCL RBD1,2, we have used a microfluidic platform that allows the capture and fluorescence detection of NCL RBD1,2. The microfluidic assay was carried out with the pre-miRNA-149 G4, modified in 5′ end with biotin, previously annealed, and immobilized in streptavidin sepharose beads. The approach is similar to that performed in SPR biosensor. The streptavidin beads with pre-miRNA-149 G4 immobilized were flowed through the microfluidic platform and due to a biophysical entrapment were confined into the microchannel. In order to detect the NCL RBD1,2 captured by pre-miRNA-149, the protein was labeled with FITC and was injected into the microchannel at different concentrations (0.25, 0.5, 1.25, 2.5 and 5 μM). The results shown in **Figure 9.8** demonstrated that NCL RBD1,2 was detected in a concentration-dependent manner.

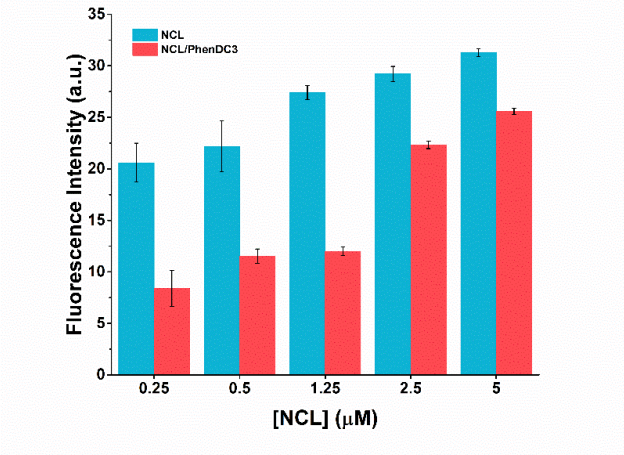

Figure 9.8. Fluorescence intensity of NCL RBD1,2 labeled with FITC and captured by pre-miRNA-149 or by pre-miRNA-149/PhenDC3 complex immobilized in streptavidin beads. Different concentrations of protein were flowed through the microfluidic channel (0.25, 0.5, 1.25, 2.5 and 5 μM).

Discussion

Human pre-miRNA-149 was previously characterized as a G4-forming molecule and since then has been taking the attention of the researchers due to its capability to regulate miRNA biogenesis and to target G4-binding proteins [5,9]. Recently, using different experimental techniques we have characterized the G4 sequence (5'-GGGAGGGAGGGACGGG-3') found in pre-miRNA-149 [10]. However, in biological context, the adjacent nucleotides of a given G4 sequence can have a tremendous impact in G4 formation, since G-richness (the fraction of guanines in the sequence) and G-skewness (G/C asymmetry between the complementary nucleic acid strands) are important parameters for the *in vivo* formation of G4 structures (Please see Chapter 3) [34,35]. Therefore, in this work, we have characterized the pre-miRNA-149 sequence and the binding and capture of NCL.

Firstly, using an *in silico* approach, we demonstrated the formation of a G4 structure by pre-miRNA-149. The results of QGRS-mapper algorithm revealed the formation of a G4 structure in the sequence 5'-GGGAGGGAGGGACGGG-3', which was previously characterized experimentally as a G4-forming sequence [10]. This G4 oligonucleotide showed important biological capabilities and its location in the pre-miRNA-149 sequence suggests a significant post-transcriptional regulatory role in gene expression, since overlaps with miRNA-149-3p, an important antitumor miRNA in several types of cancer [36–38]. On the other hand, the results from computational algorithms G4Hunter, cGcC and G4NN showed low susceptibility of the pre-miRNA-149 sequence to adopt a G4 structure. Despite these computational predictions, an initial study of Kwok et. al experimentally demonstrated the formation of the G4 structure by pre-

miRNA-149 sequence through a method called SHALiPE, which is derived from the SHAPE method and consists of the coupling of selective 2′-hydroxyl acylation with a lithium ion-based primer extension [5]. Therefore, it is important to perform a deeply experimental characterization using different techniques to undoubtedly prove the G4 formation by pre-miRNA-149.

The results from conservation studies also corroborate the importance of the rG4 region to recognize molecules involved in miRNA biogenesis. Among the species analyzed in this study, the rG4 region was shown to be conserved in all of them, which can confirm the biological relevance of the rG4 region [14]. In fact, it has been previously reported that NCL is involved in miRNA biogenesis of microRNAs 15a and 16 [14]. Recently, a study developed by Koralewska et. al also demonstrated that G4s bind to human dicer and inhibit its activity [39]. These results are of greatest importance because indicated that G4s can play an important role in the control of miRNA biogenesis.

Once *in silico* approaches demonstrated the relevance of the G4 structure in the biological context, we employed CD and ^{1}H NMR spectroscopy to deeply investigate the formation of the G4 structure by pre-miRNA-149. The results from ^{1}H NMR spectroscopy suggested the formation of the G4 structure and probably, heterogeneity in the G4 molecularity, as observed by the intense overlapped signals in the G4 imino proton region (10-12 ppm) [40]. The existence of Watson-Crick and G4 resonances in the presence of K^{+}, suggests that the formation of the G4 in pre-miRNA-149 is locked by a stem structure [20].

The results obtained after addition of 100 mM KCl and annealing, revealed a slight shift of the positive band, which suggests a reorganization of the secondary structure. Moreover, the results obtained after cooling the solution at 4 ºC for 24 h or 48 h showed that, after annealing the solution in each step, pre-miRNA-149 retains its ellipticity. In our work with the rG4 region (5'-GGGAGGGAGGGACGGG-3'), we demonstrated that the G4 structure is stable for 7 days without changes in the ellipticity signal intensity, even in the presence of a cellular medium, supplemented with 10% fetal bovine serum [10]. The G4 formation is also dependent from K^{+} concentration as highlighted by CD titrations and CD melting experiments.

The stabilization or destabilization through ligand binding is frequently used to maintain the folded state or to disrupt the G4 structure, respectively. Until now, some small molecules able to interact with G4 pre-miRNA-149 structure have been identified [5,9,10]. For instance, the well-known G4 ligand PDS was found to induce and stabilize

the G4 structure [5], while TMPyP4 was found to disrupt/aggregate the G4 structure [9]. The destabilization of G4 structures by G4 ligands is less commonly described, but some examples have been reported in the last few years [9,41–43]. For instance, a study developed by Zamiri et. al described the disruption of the RNA G4 structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene [41]. The disruption of those structures hinders the binding of proteins to the RNA G4 structures [41].

Acridine orange (C_8 and C_8-NH_2) and phenanthroline (PhenDC3) derivatives were previously described as G4 binders and stabilizers [10,11,44]. Upon addition of 2 molar equivalents, the CD melting results revealed a stabilizer effect of C_8 and C_8-NH_2 ligands and a destabilizer effect of PhenDC3. However, in the presence of 4 molar equivalents of C_8-NH_2, the melting temperature also decreased to 61.7 ± 0.5 ºC (data not shown). The presence of the iodobenzene group at the end of the alkylamide chain seems to have a pronounced effect on the stability of the pre-miRNA-149. Moreover, at 2 molar equivalents C_8 have the best stabilizer effect with two melting points. Comparable results were obtained by Arachchilage et al. for pre-miRNA-92b in the presence of 5 mM KCl [6]. They observed two melting points attributed to the stem-loop and the G4 structures [6]. In a previous report, we analyzed the rG4 region of pre-miRNA-92b (5'-GGGCGGGCGGGAGGG-3') and observed two melting temperatures [11]. We hypothesized that the two melting temperatures could be derived from multiple rG4 conformations with different degrees of stabilization.

Among G4 interacting proteins, NCL is the protein most often reported for its functions upon G4 recognition. It is generally believed that NCL preferentially recognizes G4 structures and helps stabilize them [45]. Therefore, the proper folding of the G4 structure and its stability in the presence K^+ and ligands are fundamental issues for the successful recognition of NCL in the cellular context, whereas K^+ is known to be present in a concentration around 140-150 mM [46]. Our results indicated that in the presence of K^+ concentrations higher than 50 mM, pre-miRNA-149 melting temperature surpasses 90º C, which could significantly contribute to NCL recognition.

Molecular docking and molecular dynamics studies provided important structural insights on the binding of NCL RBD1,2 to the rG4 region of pre-miRNA-149 (5'-GGGAGGGAGGGACGGG- 3'). Recently, we have built the 3D model of rG4 region of pre-miRNA-149 and deeply investigated the interaction of the G4 structure found in pre-miRNA-149 with six distinct well-known G4 ligands. Among the ligands studied were C_8, C_8-NH_2 and PhenDC3. Here, we investigated the binding of NCL RBD1,2 to the parallel rG4 structure. The results revealed the binding of the rG4 structure to a cavity between

domains 1 and 2. The atomic interaction mainly occurs by intermolecular hydrogen bonds and hydrophobic interactions. The results are similar to those reported by some of us with the rG4 region of pre-miRNA-92b [11]. Previously, this binding pocket was also described as the preferential binding site of the NCL aptamer AS1411 and pre-ribosomal RNA [47,48].

In order to evaluate the molecularity of pre-miRNA-149 in the presence of ligands and NCL RBD1,2, we performed PAGE experiments. The results agree with CD data and suggest a total binding of pre-miRNA-149 to NCL RBD1,2 in the presence of C_8 but a partial binding in the presence of C_8-NH_2 and PhenDC3.

One of the key parameters to measure the affinity between molecules in a given G4/ligand or G4/protein interaction is the dissociation constant (K_D). The K_D of ligands (C_8, C_8-NH_2 and PhenDC3) towards pre-miRNA-149 was in the micromolar range (10^{-5}-10^{-6} M), which is in agreement with previous reports of ligands targeting G4 structures [49]. The affinity of NCL RBD1,2 towards pre-miRNA-149 or pre-miRNA-149/PhenDC3 was in the nanomolar range (10^{-10}), which suggests a powerful binding of NCL RBD1,2 to pre-miRNA-149. These results agree with K_D values previously reported for the interaction of AS1411 or AS1411 derivative sequences and NCL [20,50].

Taking into account the intrinsic features of pre-miRNA-149 its potential applicability for therapeutic and diagnostic purposes is of immense interest. Considering that cell surface NCL is overexpressed in the surface of PC3 prostate cancer cells [51], we have used this cell line to evaluate the co-localization of NCL and pre-miRNA-149 labeled with FITC. The results obtained from confocal microscopy studies showed co-localization of pre-miRNA-149 and NCL in the surface PC3 cells. Comparing these results with those previously reported for the rG4 region (5'-GGGAGGGAGGGACGGG-3), where the sequence also colocalizes with cell surface NCL [10], we hypothesized that NCL recognition by pre-miRNA-149 could occur through the rG4 region.

Finally, by covalently immobilizing the pre-miRNA-149 sequence in a microfluidic device, we have successfully captured NCL RBD1,2 in a concentration-dependent manner. The results revealed that even at lower protein concentrations (0.25 μM), pre-miRNA-149 was able to capture NCL RBD1,2. Furthermore, it is possible to observe that in the presence of PhenDC3 the fluorescence intensity of NCL RBD1,2 was lower compared to pre-miRNA-149 without stabilization of PhenDC3, indicating that PhenDC3 could compete with NCL RBD1,2 by the binding to pre-miRNA-149. Indeed, it has been previously described that an excess of PhenDC3 disrupts the interaction of NCL with G4s

[52]. Since pre-miRNA-149 sequence can form a stable and resistant parallel G4, its applicability as biosensing agent to recognize NCL in microfluidic devices is of utmost relevance considering the advantages of those systems. When compared to other methods to recognize and detect NCL, such as western blot and ELISA, the microfluidics platform based on pre-miRNA-149 offers several advantages such as, better stability and reproducibility, shorter assay time (minutes vs hours of incubation), lower cost (oligonucleotides are cheaper than antibodies), less manipulation and procedures required (easier modification and immobilization), and potential to detect multiple targets [53,54].

Compared to other previously described devices and methods for NCL recognition and detection [3,22,23], our approach also revealed to be able to detect NCL. Li et. al. described a microcantilever biosensor system able to detect nucleolin over the range from 10 nM to 250 nM, and with a detection limit of 1.0 nM [3]. The microfluidic platform based on pre-miRNA-149 was able to detect NCL RBD1,2 in a range between 0.25-5 μM with a minimum increase in concentration of 0.25 μM. Moreover, our findings show that exploiting the unexplored functions of existing G4 structures such as, pre-miRNA-149 will extend their potential applications in NCL recognition and detection in more complex samples.

Conclusion

In summary, we characterized the formation of a G4 structure by pre-miRNA-149 and the ability to recognize and capture NCL RBD1,2. Firstly, we have demonstrated the formation of G4 in the presence K^+ ions using NMR and CD experiments. Then, we showed the capacity of C_8 to stabilize the pre-miRNA-149 G4 structure. The complex formation of pre-miRNA-149 G4 with ligands and NCL RBD1,2 was proved using polyacrylamide gel electrophoresis. Using SPR, we determined binding affinities of pre-miRNA-149 G4 towards ligands and NCL RBD1,2. In the presence of PhenDC3, the binding to NCL RBD1,2 remains in the same order of magnitude (10^{-10} M). Finally, a microfluidic approach was applied to detect NCL RBD1,2 labeled with FITC. The results showed that fluorescence intensity depended on the NCL RBD1,2 flowed through the microfluidic channel. Altogether, these results could pave the way for future applications of pre-miRNA-149 G4 as a biosensing agent to detect NCL in cancer.

Acknowledgments

This work was supported by MIT Portugal FCT project BIODEVICE ref. MIT-EXPL/BIO/0008/2017, PESSOA program ref. 5079 and project "Projeto de Investigação Exploratória" ref. IF/00959/2015 entitled "NCL targeting by G-quadruplex aptamers for cervical cancer therapy" financed by Fundo Social Europeu e Programa Operacional Potencial Humano.

This work benefited from access to the Cell-Free platform of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL), an Instruct-ERIC centre, within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). Financial support was provided by Instruct-ERIC (PID: 10168 "Production of the full-length nucleolin for structural studies"). Thanks are due to FCT/MCT for the financial support to CICS-UBI UIDB/00709/2020 research unit, POCI-01-0145-FEDER-022122 research unit PPBI-Portuguese Platform of BioImaging and to the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016), through national funds and, where applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and PIDDAC. Tiago Santos acknowledges Fundação para a Ciência e Tecnologia (FCT) for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013) co-financed by Fundo Social Europeu. André Miranda acknowledges to the research fellowship "Rede Nacional de Ressonância Magnética Nuclear" ref. PINFRA/22161/2016-B4 funded by "Programa Operacional Competitividade e Internacionalização", "Programa Operacional Regional de Lisboa", FEDER, and FCT. C. Cruz acknowledges the grant from FCT ref. UIDP/00709/2020. The authors also acknowledge Jérôme Boisbouvier for kindly revising the manuscript.

References

1 Sun, H. *et al.* (2016) Aptamers: Versatile molecular recognition probes for cancer detection. *Analyst* 141, 403–415

2 Huang, G. *et al.* (2021) The Application of Nucleic Acid Probe–Based Fluorescent Sensing and Imaging in Cancer Diagnosis and Therapy. *Front. Chem.* 9, 410

3 Li, H. *et al.* (2016) Aptamer-based microcantilever biosensor for ultrasensitive detection of tumor marker nucleolin. *Talanta* 146, 727–731

4 Kharel, P. *et al.* (2020) Properties and biological impact of RNA G-quadruplexes: From order to turmoil and back. *Nucleic Acids Res.* 48, 12534–12555

5 Kwok, C.K. *et al.* (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. *Angew. Chemie - Int. Ed.* 55, 8958–8961

6 Mirihana Arachchilage, G. *et al.* (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* 22, 262–272

7 Liu, G. *et al.* (2020) RNA G-quadruplex regulates microRNA-26a biogenesis and function. *J. Hepatol.* 73, 371–382

8 Tao, Y. *et al.* (2021) Recent advances in the development of small molecules targeting RNA G-quadruplexes for drug discovery. *Bioorg. Chem.* 110, 104804

9 Ghosh, A. *et al.* (2019) Restoration of miRNA-149 Expression by TmPyP4 Induced Unfolding of Quadruplex within Its Precursor. *Biochemistry* 58, 514–525

10 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

11 Santos, T. *et al.* (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* 189, 114208

12 Ugrinova, I. *et al.* (2018) Multifaceted nucleolin protein and its molecular partners in oncogenesis. In *Advances in Protein Chemistry and Structural Biology* (Donev, R., ed), pp. 133–164, Elsevier

13 Jia, W. *et al.* (2017) New perspectives of physiological and pathological functions of nucleolin (NCL). *Life Sci.* 186, 1–10

14 Pickering, B.F. *et al.* (2011) Nucleolin protein interacts with microprocessor complex to affect biogenesis of microRNAs 15a and 16. *J. Biol. Chem.* 286, 44095–44103

15 Koutsioumpa, M. and Papadimitriou, E. (2014) Cell Surface Nucleolin as a Target for Anti-Cancer Therapies. *Recent Pat. Anticancer. Drug Discov.* 9, 137–152

16 Berger, C.M. *et al.* (2015) The roles of nucleolin subcellular localization in cancer. *Biochimie* 113, 78–85

17 Hovanessian, A.G. *et al.* (2000) The cell-surface-expressed nucleolin is associated with the actin cytoskeleton. *Exp. Cell Res.* 261, 312–328

18 Rosenberg, J.E. *et al.* (2014) A phase II trial of AS1411 (a novel nucleolin-targeted DNA aptamer) in metastatic renal cell carcinoma. *Invest. New Drugs* 32, 178–187

19 Bates, P.J. *et al.* (2017) G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1414–1428

20 Miranda, A. *et al.* (2021) Locking up the as1411 aptamer with a flanking duplex: Towards an improved nucleolin-targeting. *Pharmaceuticals* 14, 1–20

21 Miranda, A. *et al.* (2021) Aptamer-based approaches to detect nucleolin in prostate cancer. *Talanta* 226, 122037

22 Chalfin, H.J. *et al.* (2017) Nucleolin Staining May Aid in the Identification of Circulating Prostate Cancer Cells. *Clin. Genitourin. Cancer* 15, e477–e481

23 Maremanda, N.G. *et al.* (2015) Quick chip assay using locked nucleic acid modified epithelial cell adhesion molecule and nucleolin aptamers for the capture of circulating tumor cells. *Biomicrofluidics* 9, 054110

24 Kozomara, A. and Griffiths-Jones, S. (2014) MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73

25 Kikin, O. *et al.* (2006) QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34, W676–W682

26 Garant, J.M. *et al.* (2018) G4RNA screener web server: User focused interface for RNA G-quadruplex prediction. *Biochimie* 151, 115–118

27 Brázda, V. *et al.* (2019) G4Hunter web application: A web server for G-quadruplex prediction. *Bioinformatics* 35, 3493–3495

28 Beaudoin, J.D. *et al.* (2014) New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.* 42, 1209–1223

29 Garant, J.M. *et al.* (2017) Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics* 33, 3532–3537

30 Pereira, E. *et al.* (2017) Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA. *Sci. Rep.* 7, 42544

31 Imbert, L. *et al.* (2021) *In vitro* Production of Perdeuterated Proteins in H_2O for Biomolecular NMR Studies. In *Methods in Molecular Biology* 2199pp. 127–149

32 Carvalho, J. *et al.* (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286

33 Pinto, I.F. *et al.* (2017) The application of microbeads to microfluidic systems for enhanced detection and purification of biomolecules. *Methods* 116, 112–124

34 Santos, T. *et al.* (2021) G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* 14, 769

35 Klimentova, E. *et al.* (2020) PENGUINN: Precise Exploration of Nuclear G-Quadruplexes Using Interpretable Neural Networks. *Front. Genet.* 11, 1287

36 Yang, D. *et al.* (2017) Expression of miR-149-3p inhibits proliferation, migration, and invasion of bladder cancer by targeting S100A4. *Am. J. Cancer Res.* 7, 2209–2219

37 Shen, Q. *et al.* (2020) MicroRNA-149-3p inhibits cell proliferation by targeting AKT2 in oral squamous cell carcinoma. *Mol. Med. Rep.* 23, 172

38 He, Y. *et al.* (2018) miR-149 in human cancer: A systemic review. *J. Cancer* 9, 375–388

39 Koralewska, N. *et al.* (2021) RNA and DNA G-quadruplexes bind to human dicer and inhibit its activity. *Cell. Mol. Life Sci.* 78, 3709–3724

40 Živković, M.L. *et al.* (2020) Structure of a DNA G-Quadruplex related to osteoporosis with a G-A bulge forming a pseudo-loop. *Molecules* 25, 4867

41 Zamiri, B. *et al.* (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNAbinding proteins. *J. Biol. Chem.* 289, 4653–4659

42 Morris, M.J. *et al.* (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.* 40, 4137–4145

43 Mitteaux, J. *et al.* (2021) Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J. Am. Chem. Soc.* 143, 12567–12577

44 Figueiredo, J. *et al.* (2021) Targeting nucleolin by RNA G-quadruplex-forming motif. *Biochem. Pharmacol.* 189, 114418

45 Tosoni, E. *et al.* (2015) Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription. *Nucleic Acids Res.* 43, 8884–8897

46 Zacchia, M. *et al.* (2016) Potassium: From Physiology to Clinical Implications. *Kidney Dis.* 2, 72–79

47 Fan, X. *et al.* (2016) Bioactivity of 2′-deoxyinosine-incorporated aptamer AS1411. *Sci. Rep.* 6, 25799

48 Johansson, C. *et al.* (2004) Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *J. Mol. Biol.* 337, 799–816

49 Sun, Z.-Y. *et al.* (2019) Developing Novel G-Quadruplex Ligands: from Interaction with Nucleic Acids to Interfering with Nucleic Acid–Protein Interaction. *Molecules* 24, 396

50 Carvalho, J. *et al.* (2019) Aptamer-based Targeted Delivery of a G-quadruplex Ligand in Cervical Cancer Cells. *Sci. Rep.* 9, 7945

51 Tate, A. *et al.* (2006) Met-independent Hepatocyte Growth Factor-mediated regulation of cell adhesion in human prostate cancer cells. *BMC Cancer* 6, 197

52 Saha, A. *et al.* (2020) Nucleolin discriminates drastically between long-loop and short-loop quadruplexes. *Biochemistry* 59, 1261–1272

53 Jolly, P. *et al.* (2016) DNA aptamer-based sandwich microfluidic assays for dual quantification and multi-glycan profiling of cancer biomarkers. *Biosens. Bioelectron.* 79, 313–319

54 Song, K.M. *et al.* (2012) Aptamers and their biological applications. *Sensors* 12, 612–631

Supplementary Information

Pre-mirna-149 G-quadruplex as a molecular agent to capture nucleolin

***Santos, T.**; Miranda, A.; Imbert, L.; Jardim, A.; Caneira, C.R.F.; Chu, V.; Conde, J.P.; Campello, M.P.C.; Paulo, A.; Salgado, G.; Cabrita, E.J. and Cruz. C. (2022) Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin. Eur. J. Pharm. Sci. 169, 106093. DOI: 10.1016/J.EJPS.2021.106093 – IF: 4.384*

Results

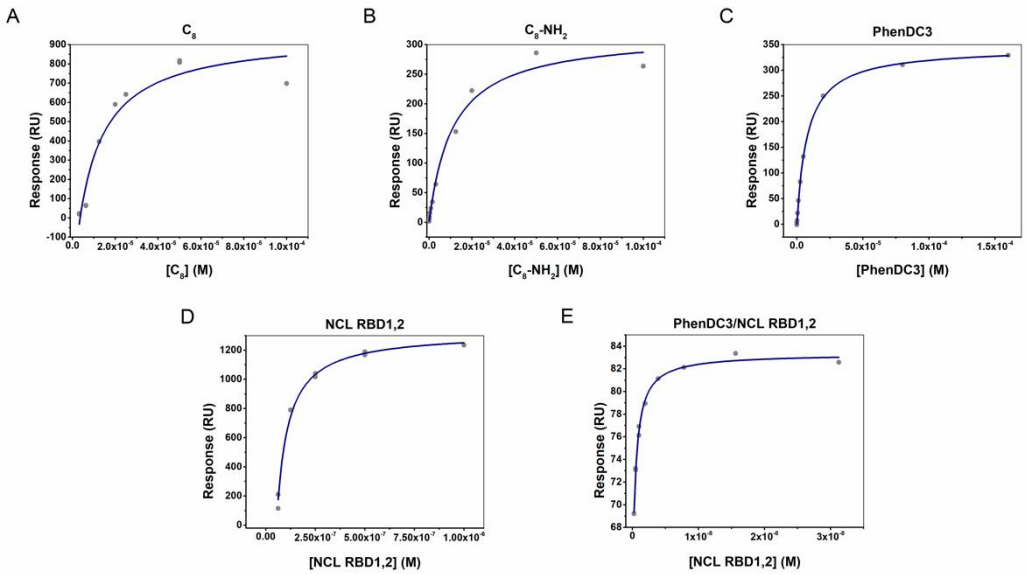

Figure S9.1. Equilibrium binding curves of pre-miRNA-149 G4 upon increasing concentrations of each ligand and/or NCL RBD1,2. K_D values were determined by fitting equilibrium binding data using Langmuir isotherm.

Table S9.1. Melting temperature variations (ΔT_m) values for pre-miRNA-149 G4 upon increasing KCl concentration.

KCl (mM)	**ΔT_m (°C)**
0.1	2.1
0.2	2.6
0.5	2.7
1	3.9
5	11.4
10	14.4
20	18.8

Table S9.2. Melting temperature variations (ΔT_m) values for pre-miRNA-149 G4 upon titration of 2 molar equivalents of ligands.

	ΔT_m (°C) at 2 molar eq.	
Ligand	**ΔT_{m1} (°C)**	**ΔT_{m2} (°C)**
C_8	2.6	34.4
C_8-NH_2	1.5	-
PhenDC3	-	-

Table S9.3. Intermolecular hydrogen bonds between G4 and NCL RBD1,2.

Donor	**Acceptor**	**Distance (D-A) (Å)**
A4-N6	Lys83-O	3.037
C13-N4	Asn16-OD1	3.038
Thr47-OG1	G15-O1P	2.621
Lys49-NZ	G15-O2P	2.691
Tyr52-OH	G16-O3'	2.811
Asn100-ND2	A12-N1	3.201
Tyr103-OH	G14-O2P	2.648
Ser159-OG	G10-O1P	2.621
Tyr163-OH	G5-O6	2.850

Chapter 10

Insights into the interaction of an RNA G-quadruplex with a ligand and nucleolin: A structural perspective

This chapter is presented as a short communication.

Chapter overview

In this chapter, we explored NMR spectroscopy for gathering structural information on the interaction of the G4 in pre-miRNA-149 with C_8 and/or nucleolin. The results of 1D ^{1}H NMR spectroscopy revealed the influence of temperature, concentration, and ionic strength in the formation of the G4 structure. Next, by using STD-NMR, we disclose the contacts at the atomic level between the G4 structure and C_8. Next, ITC provides important information on the affinity and thermodynamic parameters governing the G4/nucleolin interaction. Finally, using protein-based experiments, namely 2D ^{1}H-^{15}N HSQC, we unveil the amino acid residues of nucleolin that are involved in the interaction with the G4 structure, and with the complex G4/C_8. The results revealed a binding pocket in the region between RBD 1 and 2 of nucleolin as the main location for the interaction with the G4 structure and the complex. Overall, this study establishes the driving mode of interaction of nucleolin with the G4 in pre-miRNA 149 and fosters new investigations in the field.

Insights into the interaction of an RNA G-quadruplex with a ligand and nucleolin: A structural perspective

Abstract

Herein, we describe the structural determinants of the interaction of the G-quadruplex (G4) motif found in precursor miRNA 149 (rG4) with the acridine orange derivative C_8 and/or the protein nucleolin by NMR spectroscopy. The results revealed a strong interaction between C_8's aromatic core and iodinated ring with the rG4 structure. We also investigated, by protein observation-based NMR techniques, the complex between rG4 and rG4/C_8 and the protein nucleolin. Analysis of the complexes revealed which residues of the protein are involved in the interaction with the rG4 and rG4/C_8. Furthermore, a binding pocket for both complexes is observed between domains 1 and 2 of the protein, which suggests that the rG4 and the rG4/C_8 complex bind in this region. This complementary structural study opens a new framework to study rG4/ligand/nucleolin complexes and to control the biogenesis of miRNA 149.

Keywords

G-quadruplex; pre-miRNA 149; Nucleolin; C_8; Interaction

Introduction

RNA G-quadruplexes (rG4s) formed by precursor microRNAs (pre-miRNAs) are attractive targets of proteins and small molecules commonly called G4 ligands [1]. Some G4 motifs of pre-miRNAs have been recently identified as rG4-forming sequences. Among those sequences, pre-miRNA-92b (Please see Chapter 6) [2,3], -149 (Please see Chapter 8) [4], -1229 [5], -26a-1 [6] and pre-let-7e (Please see Chapter 5) [7,8] were experimentally found to be able to adopt a rG4 structure. Most of those G4 motifs were characterized, and their molecular interactions with proteins and G4 ligands were investigated by computational methods, several biophysical methods including low resolution structural techniques, and *in vitro* biochemical and biological assays (Please see Chapter 3) [9]. Therefore, understanding, predicting, and modulating rG4s in pre-

miRNAs is exceptionally relevant due to their potential to regulate miRNA biogenesis. Although in the past few years, some rG4 motifs in pre-miRNAs have been identified and modulated, there is a lack of high-resolution structural studies in the field. Among pre-miRNAs containing rG4-forming sequences, pre-miRNA-149 is the most noticeable since it produces miRNA-149-3p and miRNA-149-5p, which play an essential role in oncogenesis [10–12]. Moreover, Balasubramanian's research group found evidence that conformational changes in pre-miRNA 149 could control miRNA-149 biogenesis by switching its structure between G4 and stem-loop [13]. Recently, we have demonstrated that the rG4-forming sequence found in pre-miRNA-149 (5'-GGGAGGGAGGGACGGG-3') (**Figure 10.1a**) could be applied as a supramolecular carrier for the cancer-selective delivery of C_8 ligand (**Figure 10.2a**), in light of the ability of such parallel rG4 to bind to nucleolin (NCL), a protein involved in miRNA biogenesis and overexpressed in cancer cells (Please see Chapter 4) [4,14]. Herein, we unveil for the first time the structural determinants of the complex of the rG4 motif found in pre-miRNA-149 with C_8 and NCL.

Materials and methods

Oligonucleotides and ligands

rG4 sequence (5'-GGGAGGGAGGGACGGG-3') was obtained from Eurogentec (Belgium) with HPLC-grade purification. Stock solutions of approximately 1 mM were prepared and stored at −80 °C until used. The concentration of oligonucleotide samples was determined from the absorbance at 260 nm by using the molar extinction coefficient. Annealing of oligonucleotide sequences was carried out by heating the samples for 10 min at 95 °C and cooling them on ice for 20 min before the experiments. Synthesis and purification of the ligand C_8 (10-(8-(4-iodobenzamide)octyl))-3,6-bis(dimethylamine) acridinium iodide) was performed as previously described [15]. A 4 mM stock solution of C_8 was prepared in deuterated dimethyl sulfoxide (DMSO-d6).

Protein production and purification

The sequence corresponding to NCL RNA binding domains 1 and 2 (NCL RBD1,2) was cloned into a pIVEX 2.4D vector. NCL RBD1,2 was synthesized *in vitro* using a cell-free expression system [16]. Briefly, NCL RBD1,2 was expressed in a volume of 9 mL under RNAse-free conditions in dialysis mode, with 1/10 ratio of the reaction mixture to feeding mixture for 16 h at 23 °C under gentle agitation. The cell-free mixture contained 16 μg/mL of pIVEX 2.4D plasmid encoding the NCL RBD1,2 sequence, 1 mM of each essential amino acid, 0.8 mM of each rNTPs (guanosine-, uracil-, and cytidine-5′-

triphosphate), 55 mM HEPES (pH 7.5), 68 µM folinic acid, 0.64 mM cyclic adenosine monophosphate, 3.4 mM dithiothreitol, 27.5 mM ammonium acetate, 2 mM spermidine, 80 mM creatine phosphate, 208 mM potassium glutamate, 16 mM magnesium acetate, 250 µg/mL creatine kinase, 27 µg/mL T7 RNA polymerase, 0.175 µg/mL tRNA, and 400 µL/mL S30 *E. coli* bacterial extract. After incubation, the reaction mixture was diluted in binding buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, and 10 mM imidazole) to a final volume of 45 mL and centrifuged for 45 min at 36,000 g at 4 °C. Thereafter, the supernatant was applied onto a 5 mL Ni-NTA column that had been previously equilibrated in binding buffer at 4 °C. The column was washed with 5% of elution buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, and 500 mM imidazole) to eliminate residual contaminants and the protein was eluted with 50% of elution buffer. Fraction containing the NCL RBD1,2 was pooled and concentrated on a 10-kDa cut-off membrane. The purity of each fraction was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and the protein was identified through western blot analysis using the primary anti-NCL antibody (Thermo Fisher, ref. PA3–16875).

Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel (15%) electrophoresis was used to assess the molecularity of the rG4 structure in H_2O, 100 µM, and 100 mM KCl. Samples were prepared at a concentration of 30 µM. Sucrose was added to the samples at a final concentration of 23%. The oligonucleotide marker was loaded in parallel on the gel. Electrophoresis was performed at 2 W per gel to reach a temperature close to 20 °C. After electrophoresis, the gel was stained with Stains-All solution for 30 min under gentle agitation. Thereafter, the gel was incubated for 2 h in water for discoloration under the light before visualization. The images were processed using ImageJ.

Isothermal Titration Calorimetry (ITC)

Isothermal titrations were conducted at 25 °C using an ITC200 Microcalorimeter from Microcal, (Northhampton, USA). The protein and rG4 sequence were dialyzed extensively against 50 mM potassium phosphate (pH 6.2) and 100 mM KCl. Buffer from the dialysis was used to provide a baseline for the experiments. The data were analyzed using Origin version 5.0 provided by Microcal using a one-site model. The concentrations used in the ITC cell are 3.9 µM for the rG4 sequence. The concentration of the protein inside the syringe was 395 µM. Following an initial injection of 0.5 µl, a total of 15 injections of 2.49 µl were performed. The initial delay was set to 180 s, and the

delay between each titration point was 120 s. Sensitivity was set to normal and a value of 6, and the reaction mixture was constantly stirred at 600 r.p.m.

NMR Experiments

The samples were recorded on a Bruker Avance III 600 MHz spectrometer equipped with a QCI cryogenic probe head. The water NMR samples were prepared in a D_2O/H_2O ratio of 10%/90%. For the analysis of the rG4 sequence in the presence of potassium and C_8, samples were prepared in the presence of 100 μM KCl. For the analysis of the protein and complexes, samples were prepared between 0.1 to 0.2 mM protein in 50 mM potassium phosphate (6.2), 100-400 mM KCl and either 10% D_2O. The resonance of 2,2,3,3-tetradeutero-3-trimethylsilylpropionic acid (TSP) was used as a chemical shift reference in the experiments (δ TSP = 0 ppm). Two-dimensional ^{1}H-^{15}N HSQC spectra were acquired for NCL RBD1,2-labeled samples with and without rG4 or rG4/C_8 complex. The NCL RBD1,2 ^{15}N resonances of NCL RBD1,2 were assigned using ^{1}H, ^{13}C and ^{15}N backbone and side-chain chemical shifts of the NCL RBD1,2 deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 16646. Spectra were processed using NMRPipe and analyzed using CcpNMR analysis. STD-NMR experiments were performed with a molar ratio of approximately 1:100, with concentrations of 4 and 400 μM for the mixture rG4/C_8. The STD-NMR spectra were acquired with 512 scans in a matrix with 32k data points in t2 in a spectral window of 12 019.230 Hz. An excitation sculpting module with gradients was employed to suppress the water proton signals. Selective saturation of the rG4 resonances (on-resonance spectrum) was performed using a series of 40 Eburp2.1000-shaped 90° pulses (50 ms) for a total saturation time of 2.0 s. For the reference spectrum (off-resonance), the samples were irradiated at -1 ppm. The STD spectra were obtained by subtracting the on-resonance spectrum from the off-resonance spectrum. Then, the percentages of STD were calculated by comparing the intensities of the signals in the STD spectrum with the signal intensities of the reference spectrum (off-resonance spectrum). Control experiments were performed without the rG4 sequence in solution. To determine the STD-derived epitope mapping of C_8, the STD intensities were normalized with respect to that with the highest response.

Results and discussion

We first characterized the conformational behavior of the rG4 motif in the presence of K^+. It has long been believed that, due to their intrinsic features, some rG4s show a tendency to associate/aggregate through stacking interactions between the external G-tetrads [17]. Furthermore, parallel rG4s are especially prone to form high order structures since their external G-tetrads are almost entirely exposed to the solvent [17]. As expected, ionic strength plays an essential role in the shift between monomeric and other high-order structures (dimers and tetramers). Therefore, we investigated the effect of K^+ ions in the formation of the rG4 structure. As seen in **Figure 10.1b**, in the absence and presence of 100 μM KCl (lanes 1 and 2), the sequence adopts a monomeric form, while in the presence of 100 mM KCl (lane 3), the sequence adopts multimeric species (dimers and tetramers).

1D ^{1}H NMR spectroscopy may provide additional valuable insights into the molecularity of the G4 structure. rG4 imino protons presented a specific 1D ^{1}H NMR signature derived from Hoogsteen base pairing. The 1D ^{1}H NMR spectrum of the rG4 in K^+ solution (**Figure 10.1c**) shows a major intramolecular G4 structure with 12 resolved imino proton resonances between 11.2 and 12 ppm. These results indicated the presence of a unique monomeric form and an rG4 with three G-tetrads in the potassium concentration range 0.05-0.25 mM. Above 0.25 mM K^+, the spectra showed peak broadening and the appearance of new resonances which is in agreement with the formation of high-order G4 structures.

Besides investigating the effect of K^+ in rG4 formation, we also assessed the impact of temperature on rG4 structure by recording 1D ^{1}H NMR spectra. The results are shown in **Figure S10.1** and indicate that at temperatures below 300.15 K, a minor secondary G4 conformation can be noted. Since rG4 concentration is an important parameter that induces rG4 multimerization, we have analyzed a more concentrated sample (500 μM). The results seen in **Figure S10.1** showed that conformational preference strongly depends on oligonucleotide concentration. Indeed, the structural requirements of G4 structures to multimerize are not well understood; however, in the last few years, some attempts have been conducted to unveil such determinants. A study developed by Curtis's research group provided new insights into the multimerization of DNA G4s [18], and recently Frasson et al. reviewed the biological effects of multimeric G4s [19]. The major challenge remains in the methods employed to characterize multimers, which cannot provide detailed structural information, especially in the cases where there are

several species in solution. Therefore, the association/aggregation of the rG4 has hindered more detailed structural studies.

Next, we employed 1D ^{1}H NMR spectroscopy to evaluate the formation of the complex rG4/C_8 monitoring the rG4 resonances (**Figure 10.1d**). Upon addition of C_8 into the rG4 solution, a new set of distinct imino proton resonances appeared at a ligand equivalence in the range 0.1-0.75. The imino proton resonances of free rG4 remained visible, indicating a slow-exchange equilibrium in the NMR timescale between free and complexed rG4, as expected for a high-affinity binding between C_8 and rG4. In fact, it was previously determined by fluorescence spectroscopy that C_8 binds the rG4 structure in the micromolar range (10^{-6} M) [4]. Additionally, the data revealed specific binding, as observed by the complex's sharp and well-resolved imino proton resonances in a ratio of 1/2 of rG4/C_8.

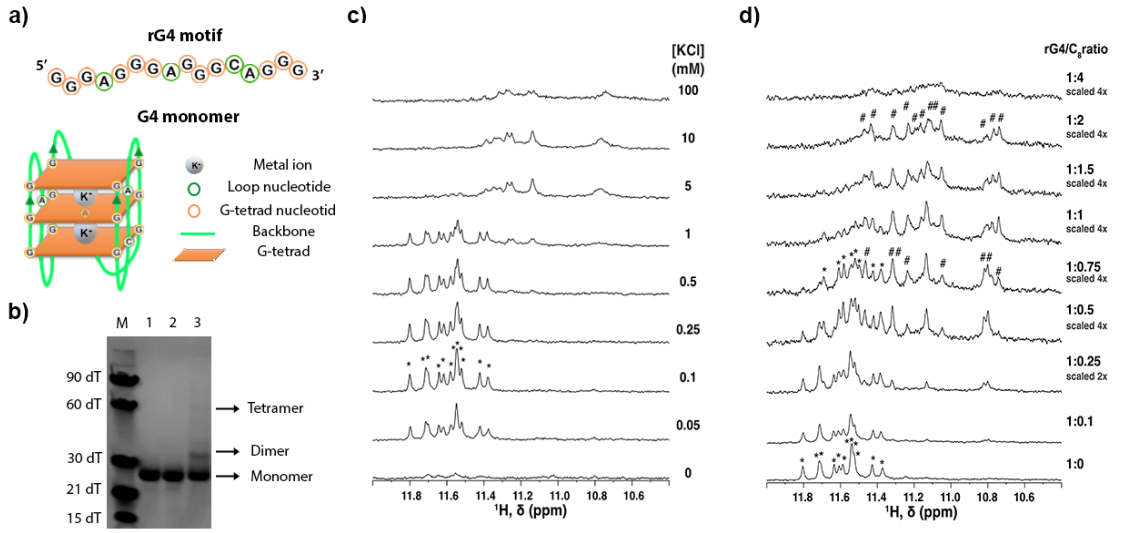

Figure 10.1. a) Schematic illustration showing the rG4 sequence in pre-miRNA 149. **b)** Electrophoretic profile of rG4 structure. M – Marker; 1 – rG4 in the presence of H_2O; 2 - rG4 in the presence of 100 µM KCl and 3 – rG4 in the presence of 100 mM KCl **c)** Effect of KCl salt on pre-miR-149 as monitored by 1D ^{1}H NMR spectroscopy. Imino protons of rG4 solution (100 µM) with increasing concentrations of KCl (as indicated on the right of the spectrum). The spectra were recorded in H_2O Milli-Q, 10 µM TSP-d_4, and 10 % D_2O at 293.15 K. **d)** ^{1}H NMR spectra titration of the rG4 structure with C_8. Imino region of ^{1}H spectra of rG4 solution (100 µM) with increasing concentrations of C_8 (rG4/C_8 ratio and intensity scale are indicated on the right of the spectrum). Imino protons of the rG4 are indicated with asterisks (*), and those in the complex are indicated with #. The spectra were recorded in water containing 100 µM KCl, 10 µM TSP-*d4*, 10 % D_2O at 293.2 K. The annealing procedure was performed for each sample before measurement.

We also followed the formation of the complex rG4/C_8 monitoring the ligand C_8 resonances. As seen in **Figure 10.2b**, the addition of rG4 to a solution of C_8 indicates a strong binding since the aromatic protons of the ligand disappear in a low molar ratio of

0.05/1 rG4/ligand, suggesting again an intermediate to slow-exchange equilibrium between free and complexed C_8 due to a tight binding to the rG4. As expected, these results agree with those from the rG4 perspective study. The addition of 4 µM of rG4 to a solution of 400 µM C_8 led to an overall decrease in the intensity of aromatic protons. Still, its proton shifts (**Figure 10.2c**) demonstrated that the two C_8 moieties bind the rG4 differently. While shifts in the chemical shifts are observed for the acridine protons, the protons from the iodinated ring do not shift substantially.

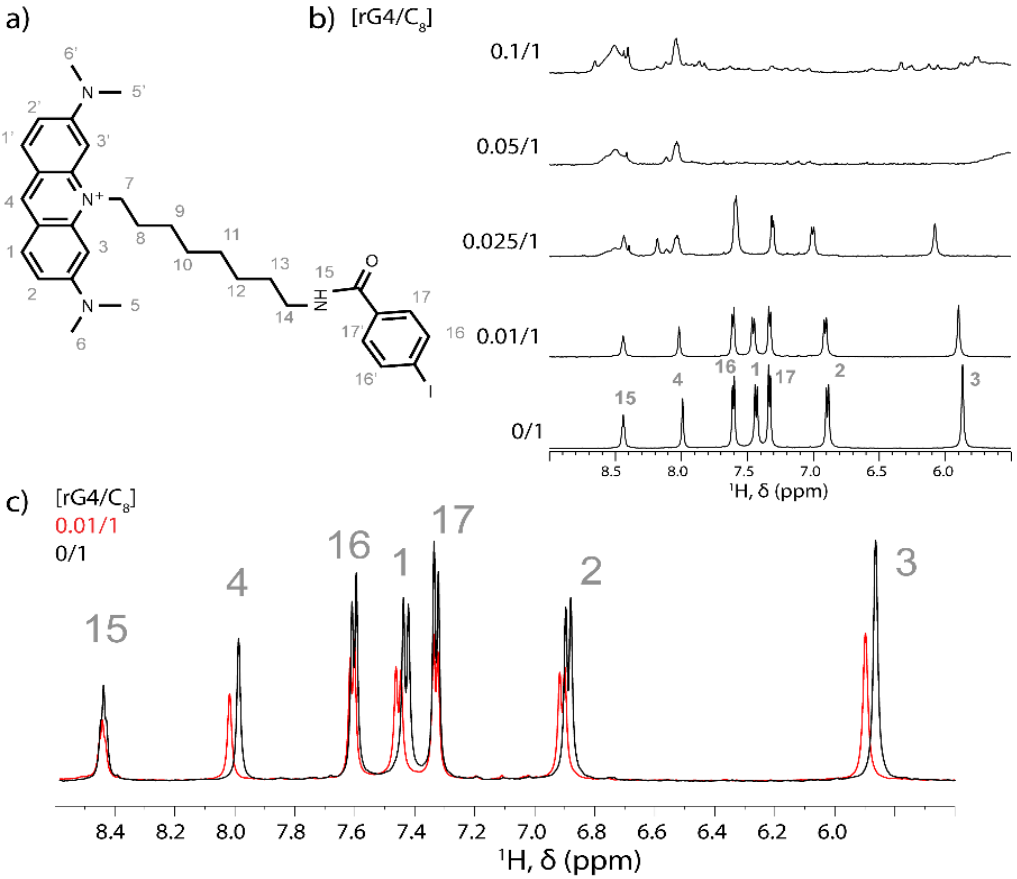

Figure 10.2. ^{1}H NMR spectra titration of C_8 with rG4. **a)** Chemical structure of C_8. **b)** Aromatic region of ^{1}H spectra of C_8 solution (400 µM) with increasing concentrations of rG4 (rG4/ligand ratio). **c)** ^{1}H NMR spectra overlay of rG4 without and with 0.01 molar eq. of C_8. The spectra were recorded in water containing 100 µM KCl, 10 µM TSP-*d4*, 10 % D_2O and 4% DMSO-*d6* at 293.15 K. The annealing procedure was performed for each sample before measurement.

STD-NMR is one of the most used methods to investigate ligand-receptor interactions from the ligand perspective, where only the NMR assignment of the ligand is required. [20,21]. The molecular weight of the rG4 motif (5.3 KDa) is rather low for efficient spin diffusion but bearing in mind its propensity to multimerize and that the presence of low percentages of oligomers could help to achieve an efficient saturation of the rG4, we used STD-NMR technique to unveil the interaction rG4/C_8 at the atomic level. To this end, we

carried out an STD-NMR experiment for the ratio rG4/ligand of 1/100. The off-resonance and STD spectra (irradiating at 12 ppm) are shown in **Figure 10.3**. An additional STD experiment was performed with irradiation at -1 ppm (**Figure S10.2**). To ensure that ligand signals were not affected, proper control experiments were performed, shown in **Figure S10.3** which were used to correct the STD responses in the presence of rG4 in spite of the negligible response.

As expected, the epitope mapping revealed a strong contribution of the aromatic protons from the acridine moiety to the binding, which again proves the strong capability of the ligand to establish π-π stacking interactions with the outer G-tetrads of the rG4 (**Figure 10.3c**). Remarkably, the aromatic protons from the iodine ring demonstrated an STD effect higher than those from the acridine moiety. The large flexibility of the alkylamide side chain permits the iodinated ring to establish other types of interactions, not only with the outer G-tetrads but also with grooves and loops of the rG4 structure.

Again, these results agree with previous reports, in which we described that π-π stacking interactions probably govern the binding of C_8 ligand to DNA and RNA G4 structures despite the presence of other types of interactions (electrostatic, hydrogen bonding, van der Waals) [22]. Altogether, these results pointed out that π-π stacking interactions mainly govern the binding.

These results and STD percentages align with our previous MD simulations of rG4/C_8 complex that showed that C_8 stacks over the top G-tetrad, being the aromatic moieties crucial for the interaction (Please see Chapter 7, page 184, figure 7.6) [22]. In our previous studies, the ligand also demonstrated the ability to establish two additional H-bonds through its oxygen atom and the alkylamide side arm position is in agreement with the STD interactions previously discussed. Moreover, the alkylamide side chain seems to be away from the G-tetrad with a negligible interaction and also has the optimal size to allow the contact of the acridine and iodobenzene moieties with the top G-tetrad of the rG4. These results are strongly in accordance with the data obtained from STD experiments and reveal the importance of validating computational results with experimental techniques.

In summary, we unveiled the contacts governing the rG4/C_8 complex formation at the atomic level. This study will pave the way for future design and development of ligands that may specifically target rG4s in pre-miRNAs (Please see Chapter 2). To the best of our knowledge, it is the first study that unveils such structural determinants.

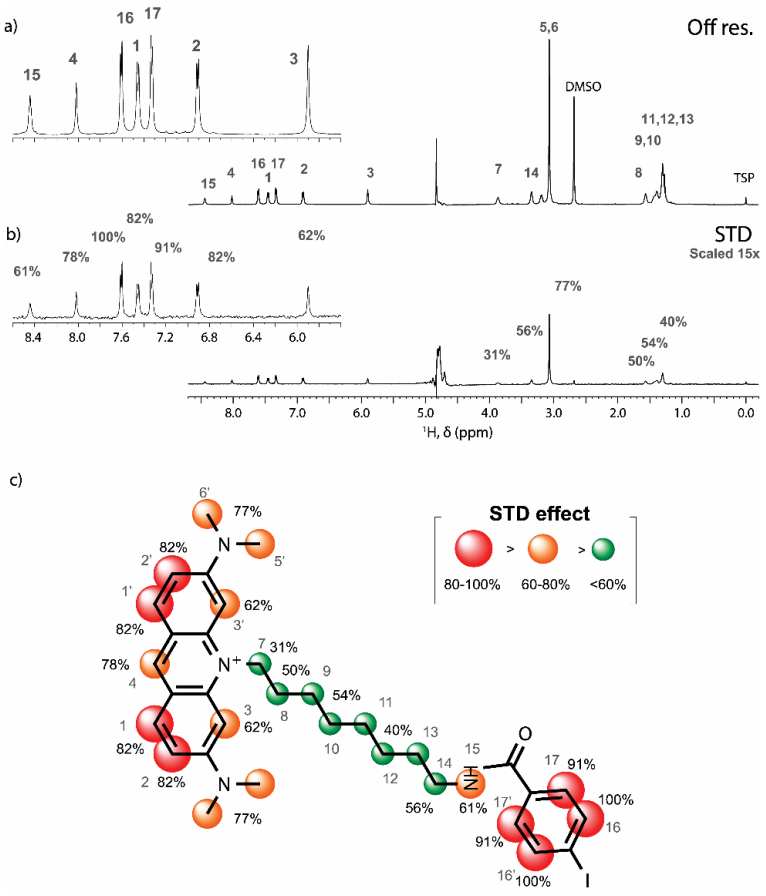

Figure 10.3. STD-NMR of C_8 in the presence of the rG4. **a)** Top - reference spectrum (off-resonance) with resonance assignments; **b)** Bottom – STD NMR spectrum of 400 μM C_8 with 4 μM rG4 (irradiation at 12 ppm) and relative STD NMR intensities in percentage. **c)** Epitope mapping was obtained for the C_8 ligand in the presence of the rG4.

Next, by using protein-based NMR methods, we sought to identify the binding sites of the rG4 on the NCL RNA-binding domains 1 and 2 (RBD1,2). NCL contains four RNA-binding domains (RBDs), which mediate the interaction with several types of nucleic acids. However, the most relevant domains are 1 and 2, which strongly interact with G4 structures. It is widely recognized that NCL RBD1,2 binds with high-affinity DNA G4s [23,24]. This rule-of-thumb seems to be also applied to RNA G4s, but only a few studies have been conducted [25,26]. Furthermore, it is worth noting that NCL interacts with rG4 structures involved in miRNA biogenesis, but until now, no structural studies of the respective complexes are available, which enhances the impact and priority of this study.

Therefore, in order to obtain preliminary information of the rG4/NCL RBD1,2 interaction, we have determined the binding affinity and thermodynamics of the rG4/NCL RBD1,2 complex by means of isothermal titration calorimetry (ITC). The ITC titrations were conducted by placing the rG4 in the sample cell and NCL RBD1,2 in the injection syringe. The results in **Figure 10.4** demonstrated that the rG4/NCL RBD1,2

interaction is enthalpically driven, exhibiting a large negative enthalpy change for the formation of the complex (ΔH value of −18.25 kcal/mol). Moreover, the dissociation constant (K_D) is 1.6 × 10^6 M^{-1}. The overall stoichiometry for binding of NCL RBD1,2 to the rG4 was determined to be 1 protein unit per one rG4.

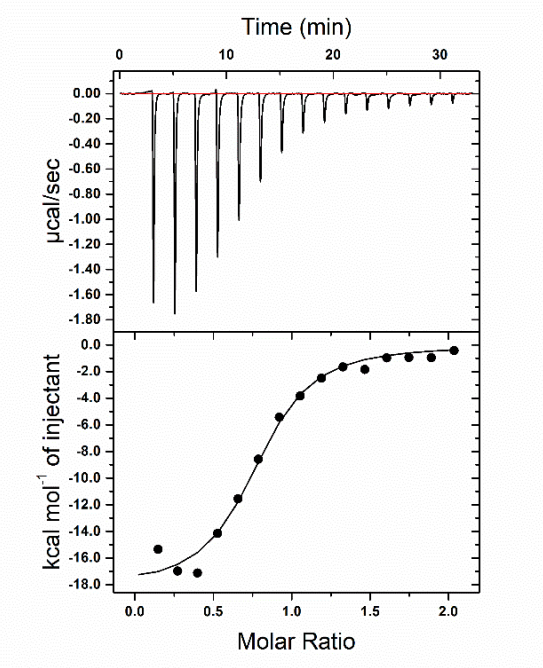

Figure 10.4. Affinity and thermodynamic parameters were measured by isothermal titration calorimetry (ITC) using the rG4 sequence (5'-GGGAGGGAGGGACGGG-3') as a ligand of NCL RBD1,2. The ΔH value was -18.25 kcal/mol, whereas the $T\Delta S$ value was -0.75 kcal/mol.

Next, we investigated the rG4/NCL interaction by titrating the rG4 into ^{15}N-^{13}C labelled NCL RBD1,2 and monitoring the titrations by acquiring 2D ^{1}H-^{15}N HSQC spectra. The assignment of the NCL RBD1,2 structure was based on the data collected by Lane et al., using similar experimental conditions [27], which allowed to assign about 85% of all protein residues (**Figure S10.4**) in the 2D ^{1}H-^{15}N HSQC spectra. Moreover, NMR titrations showed a concentration-dependent decrease in intensities of many protein signals, indicating complex formation to be predominantly in the intermediate to slow exchange NMR time scale. In addition, some NCL residues showed changes in the chemical shift by increasing the concentration of the rG4, which indicated that those residues are probably in a fast exchange in the NMR time scale (**Figure S10.5**).

First, the titrations were performed with an ionic strength of 150 mM K^+; however, the complex seems to aggregate in an NCL/rG4 ratio above 1/0.3. Bearing this result in mind, we increased the solution ionic strength to 450 mM K^+ to increase the complex's stability. In fact, the stability of the complex was enhanced, and we were able to titrate

the protein solution with 0.5 molar eq. of the rG4 (**Figure S10.5**). Even though that rG4 multimerize at concentrations of K^+ higher than 0.25 mM, the specificity of the interaction NCL/rG4 seems not affected at high concentrations of salt since the high order rG4s remain structured. Moreover, considering that NCL presented a significant preference for the rG4 monomer form, we suppose that the equilibrium probably shifts towards rG4/NCL in a ratio of 1/1, and assume for all experiments the binding of one molecule of protein to one molecule of rG4.

Since the majority of the resonances display variations in intensity upon complex formation (**Figure 10.5a**), in order to identify the NCL RBD1,2 residues that were most perturbed by the rG4 structure, peak intensity ratios (I/I_0) were calculated for each residue at ratios of 1/0.25 of NCL RBD1,2/rG4 complex [28] (**Figure 10.5b**). Residues with peak intensity ratios lower than the average intensity minus one standard deviation were mapped onto the structure of NCL RBD1,2 (**Figure 10.5c**). The protein residues that were most affected in the presence of 0.25 molar eq. rG4 structure were polar (D30, R43, K83, R91, R94, K104, D126, K138) and hydrophobic (M46, V53, L101, Y103, I132, A133) residues (**Figure 10.5**). Interestingly, the most affected residues at both molar ratios were clustered and pointed to the linker and the region between the two RBDs, indicating that this protein region is probably the most involved in the interaction with rG4. In addition, some of the residues are positively charged, suggesting that both electrostatic and hydrophobic contacts drive the interaction.

Finally, we assessed the interaction of the protein with the rG4/C_8 complex. 2D ^{1}H-^{15}N HSQC spectra demonstrated a substantial decrease in intensity by increasing the rG4/C_8 complex concentration in solution, therefore suggesting a strong molecular interaction (**Figure 10.5a**). Taking this into account, we determined the peak intensity ratio (I/I_0) at a protein/rG4/C_8 complex molar ratio of 1/0.25 (**Figure 10.5b**). As previously undertaken for NCL RBD1,2/rG4 complex, residues with peak intensity ratios lower than the average intensity minus one standard deviation were mapped onto the structure of NCL RBD1,2 (**Figure 10.5c**). The results showed that the most affected residues of the protein are L10, L15, F32, V40, R43, I44, M46, K83, K89, L96, N100, P102, Q107, A133, I135 and G166 (**Figure 10.5**). In this case, the residues are mainly hydrophobic, indicating that hydrophobic contacts should drive the ternary complex formation. Notably, the binding pocked between the two domains of the protein is even more prominent than that previously observed for protein/rG4 interaction, suggesting that C_8 may have a strong stabilizer effect on the rG4 structure, making the rG4 more prone to interact with the protein.

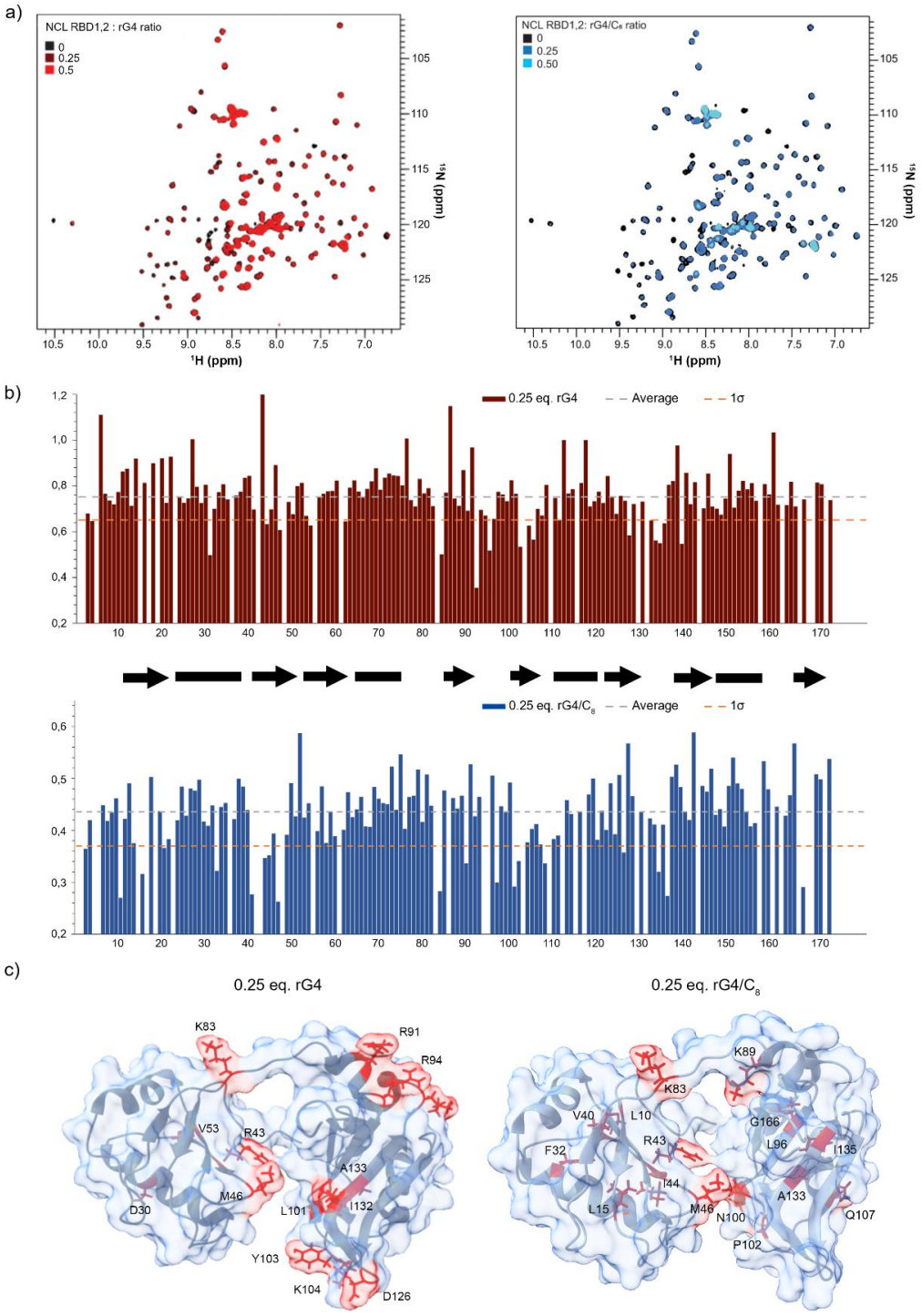

Figure 10.5. NMR titrations of NCL RBD1,2 with unlabeled rG4 and rG4/C8 complex. **a)** Overlay of three ^{1}H-^{15}N HSQC spectra of ^{15}N ^{13}C NCL RBD1,2 with increasing concentrations of unlabeled rG4 and rG4/C8 complex. **b)** Plot of ratio of peak intensities of complex (I) vs free (I_0) for the NCL RBD1,2/rG4 and NCL RBD1,2/rG4/C8 complex at 1/0.25 molar ratio shown with the average peak intensity ratio (dashed gray line) and one standard deviation from the average (1σ, dashed orange line)**. c)** Most affected residues in the presence of 0.25 molar eq. rG4 and rG4/C8 (colored red) were mapped onto the solution NMR structure of NCL RBD1,2 (colored gray). Data were acquired with 200 μM 15N, 13C-labelled NCL in buffer 50 mM potassium phosphate pH 6.2, 0.1 % NaN3, 20 μM TSP-d4, 10 % D2O with 400 mM KCl.

These results are in agreement with our previous results of MD simulation studies of the rG4/NCL RBD1,2 complex that showed that the rG4 structure binds to NCL RBD1,2 in the binding cavity between RBD1 and RBD2 (Please see Chapter 9, page 244, figure 9.5) [29]. In the MD simulation studies with rG4/NCL RBD1,2 complex, the interaction revealed that hydrogen bonds were established between rG4 and NCL RBD1,2 amino acid residues N16, T47 K49, Y52, N100, Y103, S159, and Y163. These amino acid residues are the same or contiguous to those detected by NMR, suggesting that predicted and experimentally determined contact residues are correlated, which validates both studies.

Some differences are observed in the the mapping of the protein upon interaction with the rG4 and with the rG4/C_8 complex. As expected, those differences could be attributed to the presence of C_8. In our previous study, we evaluated the binding of C_8 towards the rG4 via molecular dynamics studies, and found that C_8 strongly interacts with the top G-tetrad (G1-G5-G9-G14) by stacking interactions (Please see Chapter 7, page 184, figure 7.6) [22]. Interestingly, the top G-tetrad of the rG4 is pointed inward in its interaction with the protein, establishing more contacts with RBD2 and fewer with the RBD1. Therefore, the stacking of C_8 on top of G-tetrad may significantly affect the main driven interactions in that region of the protein. By analyzing the results, we can observe that for rG4/NCL RBD1,2 complex, the protein amino acid residues R91, R94, and K104, located in RBD2, strongly interact with the rG4. These polar residues may establish ionic interactions with the negatively charged rG4. However, in the presence of C_8, these polar residues do not interact strongly with the complex rG4/C_8. In this way, we can assume that the presence of C_8 in this region possibly hinders the interaction with these residues of the protein with the rG4. On the other hand, the hydrophobic nature of C_8 seems to favor the interaction with amino acids that have hydrophobic side chains like L96, A133, I135, and G166.

Overall, our results agree also with those observed by Allain et al. concerning the interaction of NCL RBD1,2 with an RNA stem-loop [30]. Their results also unveiled that the RNA binds to the cavity shaped between the two domains, similar to that observed by us. Mechanisms driving NCL/rG4/ligand interactions in the context of human diseases are currently unknown, and this gap in knowledge is accentuated due to the lack of detailed structural information. Our study, together with the recent findings that establish NCL/rG4 as a significant partnership associated with several diseases, will encourage new fruitful structural investigations that could advance the field (Please see Chapter 4).

Acknowledgments

This work was supported by PESSOA program ref. 5079 and project “Projeto de Investigação Exploratória” ref. IF/00959/2015 entitled "NCL targeting by G-quadruplex aptamers for cervical cancer therapy" financed by Fundo Social Europeu e Programa Operacional Potencial Humano. This work benefited from access to the Cell-Free platform of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL), an Instruct-ERIC centre, within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). Financial support was provided by Instruct-ERIC (PID: 10168 "Production of the full-length nucleolin for structural studies"). Thanks are due to FCT/MCT for the financial support to CICS-UBI UIDB/00709/2020 research unit, POCI-01-0145-FEDER-022122 research unit PPBI-Portuguese Platform of BioImaging, C2TN-IST UID/Multi/04349/2019 research unit and to the Portuguese NMR Network (ROTEIRO/0031/2013-PINFRA/22161/2016), through national funds and, where applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and PIDDAC. Tiago Santos acknowledges Fundação para a Ciência e Tecnologia (FCT) for the doctoral fellowship PD/BD/142851/2018 integrated in the Ph.D. Programme in NMR applied to chemistry, materials and biosciences (PD/00065/2013) co-financed by Fundo Social Europeu and Programa Operacional Regional Centro. C. Cruz acknowledges the grant from FCT ref. UIDP/00709/2020.

References

1 Tassinari, M. *et al.* (2021) Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* 49, 3617–3633

2 Santos, T. *et al.* (2020) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem. Pharmacol.* DOI: 10.1016/j.bcp.2020.114208

3 Mirihana Arachchilage, G. *et al.* (2015) A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem. Biol.* 22, 262–272

4 Santos, T. *et al.* (2019) RNA G-quadruplex as supramolecular carrier for cancer-selective delivery. *Eur. J. Pharm. Biopharm.* 142, 473–479

5 Imperatore, J.A. *et al.* (2020) Characterization of a G-Quadruplex Structure in Pre-miRNA-1229 and in Its Alzheimer’s Disease-Associated Variant rs2291418: Implications for miRNA-1229 Maturation. *Int. J. Mol. Sci. 2020, Vol. 21, Page 767* 21, 767

6 Liu, G. *et al.* (2020) RNA G-quadruplex regulates microRNA-26a biogenesis and function. *J. Hepatol.* 73, 371–382

7 Pandey, S. *et al.* (2015) The RNA Stem-Loop to G-Quadruplex Equilibrium Controls

Mature MicroRNA Production inside the Cell. *Biochemistry* 54, 7067–7078

8 Santos, T. *et al.* (2022) Targeting a G-quadruplex from let-7e pre-miRNA with small molecules and nucleolin. *J. Pharm. Biomed. Anal.* 215, 114757

9 Santos, T. *et al.* (2021) G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* 14, 769

10 He, Y. *et al.* (2018) miR-149 in human cancer: A systemic review. *J. Cancer* 9, 375–388

11 Chen, Y. *et al.* (2016) Downregulated expression of miRNA-149 promotes apoptosis in side population cells sorted from the TSU prostate cancer cell line. *Oncol. Rep.* 36, 2587–2600

12 Okato, A. *et al.* (2017) Dual strands of Pre-miR-149 inhibit cancer cell migration and invasion through targeting FOXM1 in renal cell carcinoma. *Int. J. Mol. Sci.* DOI: 10.3390/ijms18091969

13 Kwok, C.K. *et al.* (2016) Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA. *Angew. Chemie - Int. Ed.* 55, 8958–8961

14 Santos, T. *et al.* (2022) Nucleolin: a binding partner of G-quadruplex structures. *Trends Cell Biol.* DOI: 10.1016/j.tcb.2022.03.003

15 Pereira, E. *et al.* (2017) Evaluation of Acridine Orange Derivatives as DNA-Targeted Radiopharmaceuticals for Auger Therapy: Influence of the Radionuclide and Distance to DNA. *Sci. Rep.* 7, 42544

16 Imbert, L. *et al.* (2021) *In Vitro* Production of Perdeuterated Proteins in H_2O for Biomolecular NMR Studies. In *Methods in Molecular Biology* 2199pp. 127–149

17 Varizhuk, A.M. *et al.* (2018) Polymorphism of G4 associates: From stacks to wires via interlocks. *Nucleic Acids Res.* 46, 8978–8992

18 Kolesnikova, S. *et al.* (2017) Multimerization rules for G-quadruplexes. *Nucleic Acids Res.* 45, 8684–8696

19 Frasson, I. *et al.* (2022) Multimeric G-quadruplexes: A review on their biological roles and targeting. *Int. J. Biol. Macromol.* 204, 89–102

20 Mayer, M. and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* 38, 1784–1788

21 Viegas, A. *et al.* (2011) Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.* 88, 990–994

22 Carvalho, J. *et al.* (2020) Ligand screening to pre-miRNA 149 G-quadruplex investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* 38, 2276–2286

23 Lago, S. *et al.* (2017) The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochim. Biophys. Acta - Gen. Subj.* 1861, 1371–1381

24 Saha, A. *et al.* (2020) Nucleolin discriminates drastically between long-loop and short-loop quadruplexes. *Biochemistry* 59, 1261–1272

25 Bian, W.X. *et al.* (2019) Binding of cellular nucleolin with the viral core RNA G-quadruplex structure suppresses HCV replication. *Nucleic Acids Res.* 47, 56–68

26 Lista, M.J. *et al.* (2017) Nucleolin directly mediates Epstein-Barr virus immune evasion through binding to G-quadruplexes of EBNA1 mRNA. *Nat. Commun.* 8, 16043

27 Arumugam, S. *et al.* (2010) Solution structure of the RBD1,2 domains from human nucleolin. *J. Biomol. NMR* 47, 79–83

28 McShan, A.C. *et al.* (2016) NMR identification of the binding surfaces involved in the Salmonella and Shigella Type III secretion tip-translocon protein–protein interactions. *Proteins Struct. Funct. Bioinforma.* 84, 1097–1107

29 Santos, T. *et al.* (2022) Pre-miRNA-149 G-quadruplex as a molecular agent to capture nucleolin. *Eur. J. Pharm. Sci.* 169, 106093

30 Allain, F.H.T. *et al.* (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* 19, 6870–6881

Supplementary Information

Insights into the interaction of an RNA G-quadruplex with a ligand and nucleolin: A structural perspective

Results

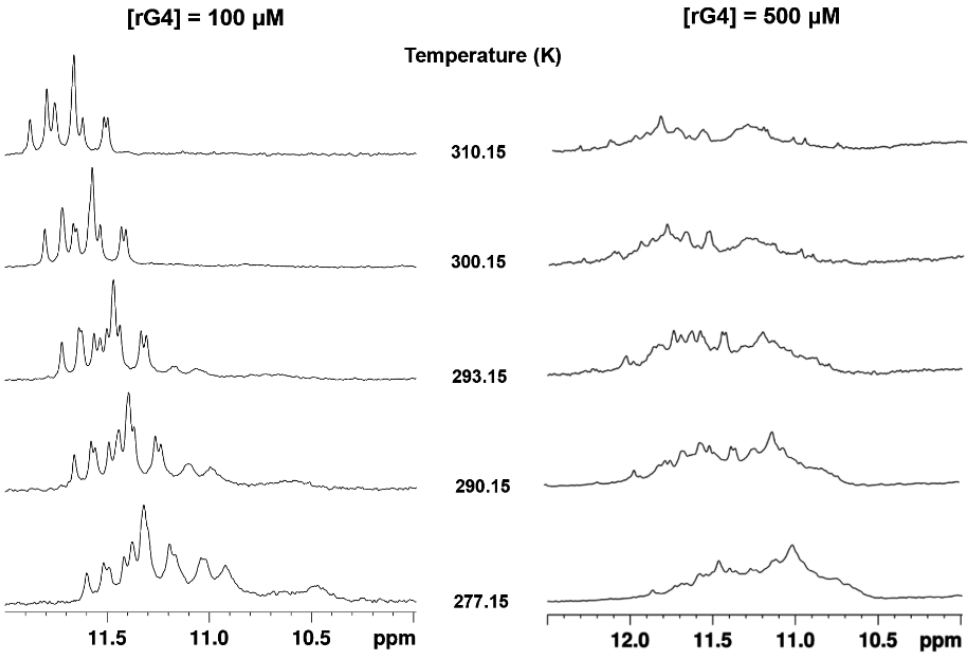

Figure S10.1. ^{1}H NMR spectra titration of the rG4 structure at different concentration and temperature. The spectra were recorded in water containing 100 µM KCl and 10 % D_2O at 293.2 K. The annealing procedure was performed for each sample before measurement.

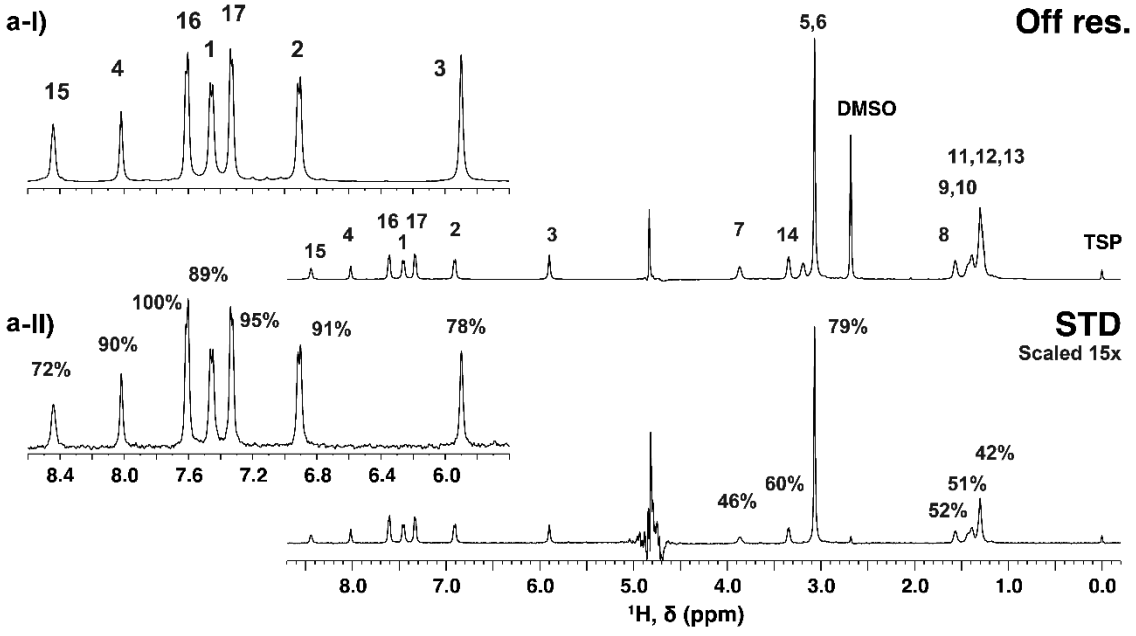

Figure S10.2. STD-NMR of C_8/rG4 with irradiation at -1 ppm. **a-I)** Top - reference spectrum (off-resonance) with resonance assignments; **a-II)** Bottom – STD NMR spectrum of 400 µM C_8 with 4 µM rG4 and relative STD NMR intensities in percentage. The inlay shows the aromatic proton region of C_8.

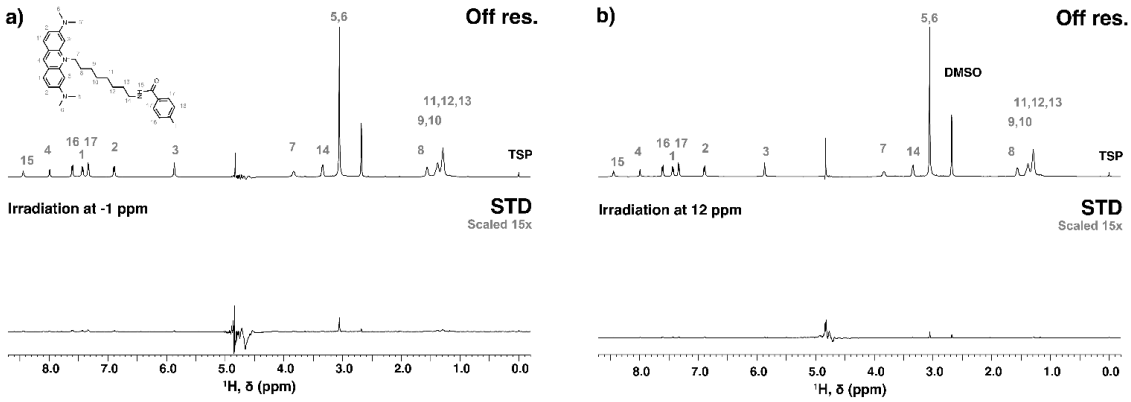

Figure S10.3. Control STD-NMR experiments. **a)** Top - reference spectrum (off-resonance) with resonance assignments; Bottom – STD NMR spectrum of 400 μM C_8 with irradiation at -1 ppm. **b)** Top - reference spectrum (off-resonance) with resonance assignments; Bottom – STD NMR spectrum of 400 μM C_8 with irradiation at 12 ppm.

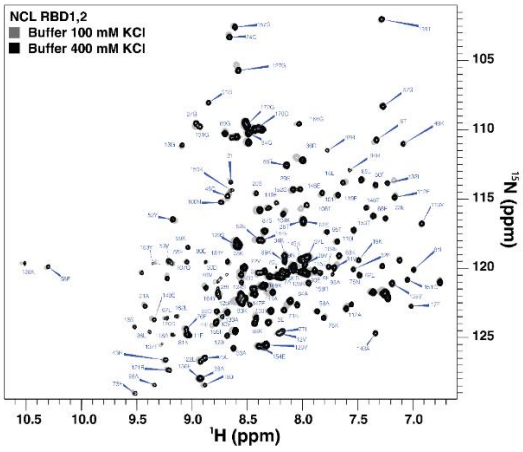

Figure S10.4. ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) spectra of protein NCL RBD1,2 in 100 mM KCl or 400 mM KCl buffered solutions. Data were acquired with 200 μM ^{15}N, ^{13}C-labelled NCL in buffer 50 mM potassium phosphate pH 6.2, 0.1 % NaN_3, 20 μM TSP-d4, 10 % D_2O with 100 mM KCl (gray) or 400 mM KCl (black), at 293.1 K, 600.10 MHz. The spectra were recorded using a standard Bruker HSQC pulse sequence with a sensitivity-enhanced pulsed-field gradient. The assignment based on BMRB 16646 was possible for 85% of all residues.

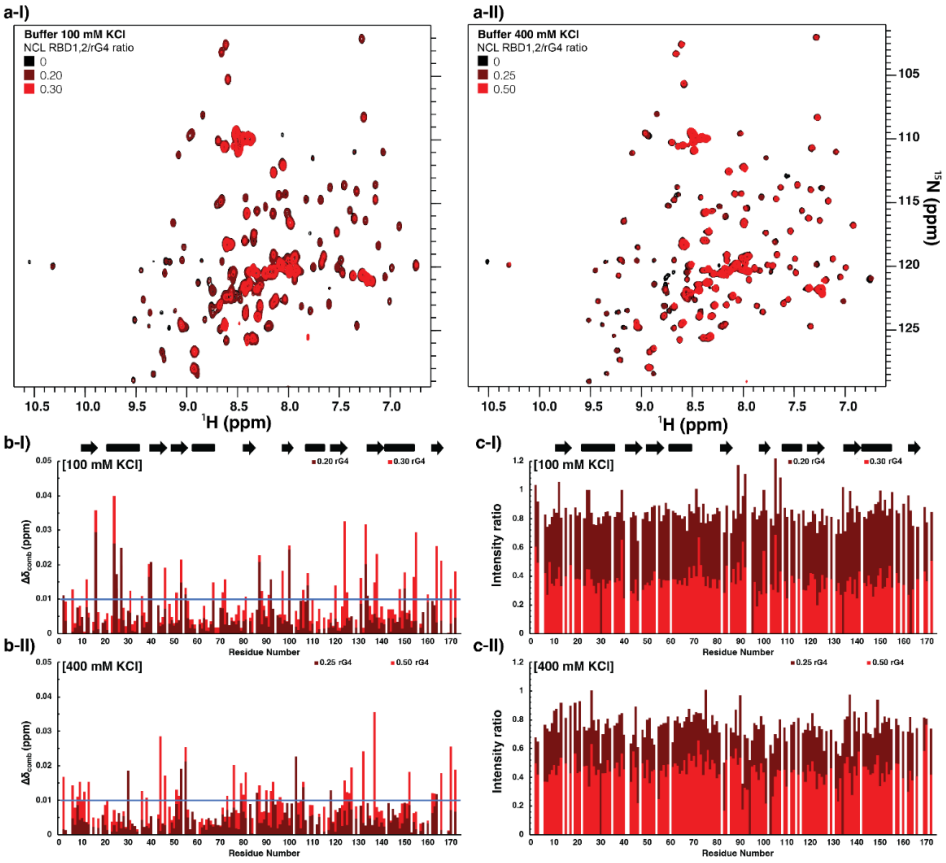

Figure S10.5. NCL RBD1,2/rG4 interactions. **a**) ^{1}H-^{15}N HSQC of NCL RBD 1,2 alone (black), and upon addition of 0.2/0.25 fold excess (maroon) or 0.3/0.5-fold-excess of rG4, in the presence of **a-I)** 100 mM or **a-II)** 400 mM KCl. **b**) Combined chemical shift perturbations plot of NCL RBD1,2 upon addition of rG4, in the presence of **b-I)** 100 mM or **b-II)** 400 mM KCl. **c**) Amide peak height intensity ratios (I/I_0) of NCL RBD1,2 upon addition of rG4, in the presence of **c-I)** 100 mM or **c-II)** 400 mM KCl.

SECTION III

Chapter 11

Conclusions and future perspectives

General conclusions and Future Perspectives

General Conclusions

Since early, the interaction of nucleolin with DNA G4s has been studied in detail, but its partnership with RNA G4s has received less attention. However, further studies evidenced that RNA G4s play essential roles *in vivo*. Consequently, it has been described that RNA G4s are implicated in many biological processes such as transcriptional regulation, mRNA processing, regulation of translation, RNA translocation, and miRNA biogenesis.

In chapter 2, we have reviewed the literature to demonstrate that in the cellular context, RNA G4s are highly transient and switch dynamically over time according to the biological environment in which they are embedded. Furthermore, we also discussed the polymorphic nature of RNA molecules, and how this feature could be used to modulate cellular functions. In a myriad of situations, they need additional stabilization, which can be provided either by ligands or proteins. We have shown several examples of the implications of G4 stabilization using small molecules, also known as G4 ligands, in the regulation of gene expression, miRNA biogenesis, and miRNA–mRNA interactions.

Having recognized the importance of stabilizing G4 structures with ligands, in **chapter 3**, the molecular interactions between G4s and ligands, as well as the most common biophysical methods used to characterize G4/ligand interactions were reviewed to conclude that a complete framework of the interaction between G4s and ligands can only be achieved by combining the data obtained from different experimental methods.

Nucleolin presents a remarkable capacity to bind parallel G4s. Furthermore, nucleolin is overexpressed on the surface of many cancer cells and could be explored as a potential biomarker for developing novel diagnostic and therapeutic strategies. In **chapter 4**, we reviewed the involvement of nucleolin/G-quadruplex partnership in different pathologies to conclude that the availability of accurate models to describe the interactions between nucleolin and G4 structures could provide valuable information for the development of novel diagnostic and therapeutic strategies in neurodegenerative disease, cancer, and viral infection.

In section II of this thesis the binding and stabilization of G4 in pre-miRNAs by several ligands was investigated experimentally, with a particular spotlight on the acridine

orange derivative C_8 and the interaction of the G4s in pre-miRNAs with nucleolin. Therefore, we characterized G4 structures in pre-miRNAs let7e, 92b, and 149, studied nucleolin/G4 complex formation with and without ligands, and investigated cell surface nucleolin as a biomarker of cancer

In **Chapter 5**, the sequence of pre-miRNA let 7e (5'-GGGCUGAGGUAGGAGG-3') was described as a G4 structure. In the presence of K^+, the sequence adopts a G4 of two G-tetrads. The ability of eight well-known ligands to bind and stabilize the G4 structure was unveiled, as well as the influence that the ligands can exert in the binding of the G4 to nucleolin. CD- and FRET-melting assays highlighted the attractive stabilizing properties of 360A and PDS, while the puzzling interacting properties of PhenDC3 and TMPyP4 were demonstrated. Furthermore, the binding affinities between the G4 structure and ligands ranged between 10^{-6} to 10^{-9} M. Lastly, using CD, surface plasmon resonance, PAGE, and confocal microscopy, we reported the binding of the G4 structure to nucleolin, in the presence and absence of each ligand. Overall, the results presented in this chapter revealed the binding of the G4 in pre-miRNA let 7e to ligands and nucleolin, both *in vitro* and *in cells*. Moreover, the results open new opportunities to modulate miRNA biogenesis.

In **Chapter 6**, we establish the G4 sequence in pre-miRNA 92b (5'-GGGCGGGCGGGAGGG-3') as a molecular recognition agent of nucleolin in complex samples. In the presence of 0.1 mM K^+, the G4 structure of pre-miRNA 92b adopts a single dominant major conformation, while in concentrations of K^+ above 0.1 mM, the sequence seems to form multimeric species. Notably, in the presence of C_8, the melting temperature of the G4 structure substantially increases. Moreover, the dissociation constants of the interaction between the G4 or G4/C_8 complex are in the low nanomolar range. A high nuclease resistance was observed, which increases when the G4 structure is complexed with C_8. These findings led us to conclude that G4/C_8 could be used as a recognition agent to sense nucleolin in complex samples. Therefore, we developed a microfluidic platform with the G4 structure as a biosensing agent for the fluorescent detection of nucleolin in biological samples. This proof-of-concept study revealed the suitability of the microfluidic platform to detect nucleolin in a concentration-dependent manner in human plasma samples.

In **Chapter 7**, we carried out an *in silico* investigation to unveil the interaction between six known ligands and the predicted 3D G4 structure of pre-miRNA 149 (5'-GGGAGGGAGGGACGGG-3'). The collected data provided valuable insights for subsequent experimental approaches. First, the modeling of the structure revealed a

compact parallel G4 structure with short loops and three G-tetrads. Moreover, the MD simulations of the interaction between the ligands and the G4 structure indicated that ligands constituted by large aromatic moieties with smaller positively charged side chains are preferred. Indeed, the results showed that ligands C_8 and PhenDC3 interact mainly with the G4 structure via stacking interactions, although some interactions with the side chains were also observed. Altogether, these results predicted a strong capability of the ligands C_8, PhenDC3, and PDS to bind and stabilize the G4 structure of pre-miRNA 149.

In **Chapter 8**, encouraged by the findings described in chapter 7, we evaluated the G4 sequence of pre-miRNA 149 as a carrier of C_8 for cancer cells. C_8 depicted a high affinity (1.0 ± 0.1 μM) for the G4 structure and stabilized it in more than 30 ºC. Furthermore, a strong nucleolin binding to G4 and G4/C_8 complex was observed. Nucleolin binds tightly towards DNA G4s with long loops, but less is known for their binding to RNA G4s with long loops. Encouraged by this hypothesis, we evaluated the binding to a RNA G4 with long loops, and as expected, the trend to bind to RNA G4s with long loops was retained. Next, we evaluated the cytotoxic effect of the G4 and G4/C_8 complex on healthy and cancer cells, and found an interesting synergistic effect of the G4/C_8 complex. Finally, supported by these results, we evaluated the cellular uptake of the complex in both cancer and healthy cells. The results agree with the data from cytotoxicity assays and indicate the potential of the G4/C_8 complex to bind nucleolin on the surface of prostate cancer cells. Moreover, after the cellular uptake mediated by nucleolin, C_8 was partially decomplexed and localized the nucleoli. Overall, this study could pave the way for future research to develop more specific delivery systems for cancer therapeutics.

In **Chapter 9**, inspired by the findings and results obtained in the previous chapters, we unveiled the potential of the full-length pre-miRNA 149 sequence to capture nucleolin in a microfluidic system. In the presence of PhenDC3, the binding affinity of the G4 structure to nucleolin was in the low nanomolar range. The results proved the capability of the sequence to be used as a recognition agent in a microfluidic device and capture nucleolin in a concentration-dependent manner. In a global analysis, the data collected could pave the way to explore potential diagnostic applications based on nucleolin detection by pre-miRNA 149.

In **Chapter 10**, taking into account the key idea that function is closely related to structure and vice-versa, and bringing together all the knowledge and expertise acquired in the previous chapters, we investigated the structural determinants of the interaction of G4 structure with C_8 and nucleolin. The results provided valuable insights into the

binding mode and interactions. First, by using 1D ^{1}H NMR spectroscopy, we characterized the formation of the G4 structure and the effect of concentration, salt, and temperature. The results revealed the presence of multimeric species/aggregation of the G4 structure in the presence of concentrations of G4 and salt above 0.1 mM K^{+}, which hindered the determination of the 3D structure of the G4. To surpass this drawback, we carried out STD-NMR, which permits unveiling the ligand atoms in close contact with the receptor and does not need any assignment of the receptor molecule to provide a reliable epitope mapping. The results revealed that the aromatic core and the iodine ring of C_8 strongly bind the G4. The amino acid residues of the protein that are mostly affected by the binding of the G4 and the complex G4/C_8 were mapped on the protein 3D structure and revealed a binding pocket between RBD 1 and 2.

Overall, the defined goals of this thesis were successfully achieved. Furthermore, the work herein presented significantly impacts the field and encourages reflections while providing new insights and answers to long-standing questions.

Future Perspectives

In the past few years, RNA G4s have been increasingly investigated. As a result, solid evidence has been provided to support their existence in living cells as well as their role in critical pathological and physiological processes. As research evolves, new computational and experimental methods are supposed to be made available, thus allowing the design and development of unprecedented G4-mediated diagnostic and therapeutic applications. The development of novel artificial intelligence algorithms, such as the AlphaFold algorithm for predicting 3D structures of proteins, represents one of the most significant breakthroughs and could boost future investigations of structure-function relationship.

Notwithstanding, some long-standing experimental questions remain to be answered, which may be the beginning of future investigations. For example, the potential of ligands for interfering with the miRNA biogenesis pathway via G4 stabilization or destabilization with ligands could be a good starting point. It may offer new therapeutic possibilities for various human disorders, such as cancer, neurodegenerative diseases, and infectious disease. However, the bottleneck stage seems to be the design and development of ligands that display remarkable affinity and selectivity towards RNA G4s. Furthermore, it is worth noting that in some particular cases, conflicting results have been observed on whether G4-mediated folding is favoured or inhibited. Such conflicting results may be due to the different experimental methods used to characterize G4s and other pathological or physiological circumstances that can alter G4 formation. Additionally, G4/nucleolin interaction received significant attention in the last few years with the development of the G4 ligand CX-3543, the first-in-class to reach Phase II clinical trials to treat several tumour types. However, despite the growing body of evidence for such a strategy, a great deal remains to be discovered.

Overall, the described hypotheses and models could open new frameworks on the diagnostic and therapeutic of the G4 structures involved in miRNA biogenesis. Additionally, we raised the question that nucleolin could be involved in several steps of miRNA biogenesis, which could be exploited to unravel important cellular mechanisms in the near future. Furthermore, our findings and achievements on the interaction of nucleolin RBD1,2 with the G4 structure of pre-miRNA 149 may also foster other investigations at the atomic level of the interaction of the protein with other pre-miRNAs.