UNIVERSIDADE
BEIRA INTERIOR

G-quadruplex targeting by ligands as a lung cancer therapeutic strategy

Joana Patrícia Rodrigues Figueiredo

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

Orientador: Prof. Doutora Carla Patrícia Freire Madeira Alves da Cruz
Co-orientador: Doutor Jean-Louis Mergny

Júri:
Prof. Doutor Paulo Jorge da Silva Almeida
Prof. Doutor Samuel Martins Silvestre
Profª. Doutora Fani Pereira de Sousa
Prof. Doutor Nuno Filipe Ribeiro Pinto de Oliveira Azevedo
Profª. Doutora Maria Alexandra da Silva Paulo
Profª. Doutora Maria do Amparo Ferreira Faustino
Profª. Doutora Maria Eduarda Romãozinho de Almeida Esteves Mendes
Profª. Doutora Carla Patrícia Alves Freire Madeira da Cruz

20 de junho de 2024

Declaração de Integridade

Eu, Joana Patrícia Rodrigues Figueiredo, que abaixo assino, estudante com o número de inscrição D2202 do 3º ciclo em Química da Faculdade de Ciências, 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.

Assinado por: **JOANA PATRÍCIA RODRIGUES FIGUEIREDO**
Num. de Identificação: 13940961
Data: 2024.07.21 22:05:50+01'00'

Universidade da Beira Interior, Covilhã 21/07 /2024

"O que é preciso é criar desassossego. Quando começamos a criar álibis para justificar o nosso conformismo, então está tudo lixado!"

Zeca Afonso

Dedico esta tese aos meus Pais, à minha Avó, ao meu Irmão e ao Carlos, pelo incansável apoio e incentivo. À minha querida afilhada, Mia

Acknowledgments

Em primeiro lugar, quero agradecer à minha orientadora, a Professora Carla Cruz, pela oportunidade que me deu para realizar este trabalho de investigação e pelo voto de confiança. Obrigada por toda a orientação científica, pela disponibilidade e pelo incessante apoio e compreensão. Os desafios que me lançou durante estes anos, assim como todos os conselhos que me deu, contribuíram em muito para o meu desenvolvimento académico e pessoal. Obrigada pela amizade e respeito com que sempre me tratou.

To my co-supervisor Doctor Jean-Louis Mergny, for his invaluable scientific guidance and constant support. Thank you for hosting me in your research group and sharing your expertise. It was an honor to learn with you. Also, to your research group, who assisted me during my LOB visit, special thanks to Anne Cucchiarini for continuous support and collaboration.

To Professor Roberto Quesado for accepting work with us and hosting me in your research group. Also, to Israel Carreira-Barral, who assisted me during my visit to Burgos, thank you for your constant support.

À Universidade da Beira Interior, particularmente à Faculdade de Ciências e ao Centro de Investigação em Ciências da saúde (CICS-UBI) e também ao Laboratoire d'Optique et Biosciences - Ecole Polytechnique, por me darem todas as condições necessárias ao desenvolvimento deste trabalho. Um agradecimento especial à Margarida Carrilho, Sofia Duarte e Ana Raquel Costa por todo o suporte técnico, ajuda e boa disposição.

To Beatriz and Marie-France for hosting me in your homes during my stay in Burgus and Palaiseau. Thank you for your hospitality. It was a pleasure sharing these days with you.

A todos os colegas e amigos de laboratório, que me acompanharam durante este percurso. Josué e Tiago obrigada por me receberem no grupo, por toda a ajuda e ensinamentos. Aos que ainda estão e me acompanham todos os dias, Daniela, Izamara, André e David, obrigada pela ajuda, discussão científica e por todos os momentos de boa disposição. Um agradecimento especial à Jéssica, que para além de colega de laboratório foi amiga e confidente durante estes anos. Ao Eurico, que me acompanha desde o mestrado e comigo partilhou angústias e sucessos, obrigada pela amizade e boa disposição. Ao Pedro, por toda a ajuda e disponibilidade. A todos os alunos de mestrado

que foram passando, obrigada pelos bons momentos. O bom ambiente no laboratório, o espírito de equipa e a entreajuda foram fundamentais durante estes anos.

A todos os meus amigos, em especial à Daniela e à Joana, obrigada pelo constante apoio.

Um profundo e especial agradecimento ao Carlos, pelo apoio incondicional e constante incentivo. Obrigada pelo companheirismo, pela paciência infinita e por teres compreendido sempre tudo.

Um agradecimento especial à minha família. Aos meus tios, às minhas avós e aos meus primos por todo o apoio e carinho. Aos meus pais, exemplos de trabalho e humildade. Obrigada por todas as oportunidades que me deram, pela estabilidade, educação e valores. Isto não seria possível sem a vossa confiança e apoio. Sou-vos eternamente grata! Ao meu irmão, que sendo bastante mais novo, é o meu maior exemplo, crítico e impulsionador.

Por último gostaria de agradecer à Fundação para a Ciência e Tecnologia (FCT), por financiar a minha Bolsa de Doutoramento (SFRH/BD/145106/2019).

Preface

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

Sir Aaron Klug

The discovery and elucidation of G-quadruplexes (G4s) structures represent an extraordinary history marked by transformative insights. The folding of DNA and RNA nucleic acid sequences containing runs of guanines into G4s *in vitro* was the starting point for subsequent research into their significance. From elucidating their unique structural features to developing innovative techniques for their detection and characterization, each milestone has contributed to the understanding of these molecular structures. Over the years, solid evidence of G4 formation in key regulatory regions of the human genome has been reported and has provided compelling support for employing the structural properties of G4 for regulating several biological processes *in vivo*. Nowadays, G4s are considered attractive targets and several efforts have been made to develop small compounds (G4 ligands) able to bind and stabilize these structures, providing an attractive anti-tumor strategy against several cancer types. Although no G4 ligand has yet been approved in the clinic, some compounds are in (pre)clinical evaluation. It is anticipated that in the near future the approval of a new drug-mediated by G4 interaction could be a new benchmark in G4 history.

Considering lung cancer (LC) prevalence worldwide and its molecular heterogeneity, new therapeutic options are needed to enhance treatment sensitivity. The prevalence of G4 in hard-to-treat tumors such as LC leads to the development of G4 ligands to harness G4-forming regions in LC-related genes.

This thesis describes the work developed under the supervision of Dr. Carla Cruz and the co-supervision of Dr. Jean-Louis Mergny and it is the result of a collaborative effort of the Faculty of Sciences of the University of Beira Interior (FC-UBI), Healthy Sciences Research Centre (CICS-UBI, Portugal) and the Laboratoire d'Optique et Biosciences - Ecole Polytechnique / Institut Polytechnique de Paris (LOB, France). The contents resulted from the work developed in the medicinal chemistry field and addressed the synthesis and evaluation of new G4 ligands as a potential LC therapeutic strategy. The thesis is subdivided into three main sections: the introduction and the aims section (**Chapters 1-4**), the experimental section (**Chapters 5-8**), and the conclusions and future perspectives section (**Chapter 9**).

The **first chapter** presents a literature review of promising G4 ligands in cancer therapy and covers their characteristics, targeting challenges, structure-activity relationships, and clinical outcomes.

The **second chapter** provides a literature review on the use of G4 ligands for stabilizing G4s in non-coding RNAs. This review discusses the biological role of G4s within non-coding RNAs focusing on their interaction with small organic molecules for targeted stabilization.

The **third chapter** describes the use of G4 ligands as potential modulators of cellular processes involved in the development and progression of LC. It provides insights into LC disease and a summary of recent advances in the development of G4 ligands capable of targeting LC-related genes.

The **fourth chapter** establishes the specific aims of the thesis.

The **fifth chapter** reports the synthesis of heterocyclic compounds featuring different aromatic scaffolds, describes their induced-stabilization effect towards different DNAs G4 and RNA G4, and discusses their cytotoxicity effects in LC cells.

The **sixth and seventh chapters** outline the development of different series of 1,10-phenanthroline-2,9-disubstituted derivatives. These chapters describe the synthetic routes employed and biophysical assays to study the ability of novel compounds to bind/stabilize G4s motifs within telomeric, oncogenes, and RNA sequences. Additionally, it reports the anti-tumor activity of these compounds in LC cells and provides insights from a structure-activity relationship analysis.

The **eighth chapter** unveils the formation of a G4 structure within the human MIR150 precursor sequence (pre-MIR150). The chapter also investigates the induced stabilization of this G4 structure by a commercial phenanthroline derivative and explores its interaction with nucleolin in LC cells.

Lastly, the **ninth chapter** presents the concluding remarks of the thesis, as well as their impact and future perspectives in the context of G4 targeting by ligands namely concerning LC therapy.

Resumo

Os G-quadruplexes (G4s) são estruturas secundárias não-canónicas de ácidos nucleícos que se podem formar em regiões ricas em guanina no genoma e transcriptoma humanos. Os G4s de DNA são encontrados em regiões regulatórias importantes, como no final dos telómeros e nas regiões promotoras de vários oncogenes. Estas estruturas estão envolvidas no controlo de inúmeros processos celulares, incluindo na replicação, transcrição, estabilidade do genoma, e regulação epigenética. Por outro lado, os G4s de RNA são encontrados em regiões não codificantes estando envolvidos em processos essenciais do metabolismo do RNA, incluindo na regulação do processamento de RNA e na tradução. Devido à sua relevância biológica e características estruturais, os G4s são considerados alvos terapêuticos relevantes. A formação e estabilização de estruturas de G4, no contexto oncológico, pode ser utilizada como estratégia anti-tumoral. Várias moléculas de baixo peso molecular com elevada especificidade para G4s de DNA e RNA (ligandos de G4) foram desenvolvidas e avaliadas quanto ao seu potencial terapêutico. Ao longo dos últimos anos, inúmeros ligandos de G4 evidenciaram atividade anti-tumoral *in vitro* e/ou *in vivo*. Frequentemente, mas nem sempre, os ligandos de G4 possuem um núcleo aromático planar que permite o empilhamento π-π com o quarteto de guanina, e grupo(s) básico(s) ou carregado(s) positivamente para interação com o grupo fosfato do ácido nucleico. Diferentes compostos com estas características químicas, como quinolinas, acridinas, naftalenodiimidas e fenantrolinas, demonstraram ter propriedades de ligação a G4s e atividade anti-tumoral. Estas moléculas orgânicas, ao interagirem com o G4, podem afetar o crescimento tumoral de diversas formas: através da inibição da telomerase ou interferindo na função telomérica; da modulação da expressão de oncogenes pela estabilização de G4s nos seus promotores; e da regulação da biogénese de microRNAs (miRNAs) pela estabilização de G4s em precursores de microRNAs (pre-miRNAs).

O cancro do pulmão (CP) é a principal causa de morte relacionada com o cancro em todo o mundo e divide-se em dois principais subtipos: o cancro do pulmão de pequenas células e o cancro do pulmão de não-pequenas células (CPNPC). Embora as opções terapêuticas convencionais, tais como a cirurgia e quimio/radioterapia, demonstrem eficácia no estadio inicial da doença, a sua eficácia é limitada no CP avançado, fase na qual a maioria dos doentes são diagnosticados. Apesar de os inibidores de controlo imunológico serem reportados na literatura como fármacos clinicamente promissores, o elevado nível de heterogeneidade molecular no CP torna o seu tratamento difícil, sendo necessário o desenvolvimento de novas estratégias terapêuticas. A prevalência da formação de G4 em

importantes regiões regulatórias do CP, incluindo proto-oncogenes (por exemplo, MYC, BCL-2, KRAS, KIT e VEGF) e regiões teloméricas, torna-os alvos terapêuticos atrativos. À semelhança do que acontece noutros tipos de cancro, também no CP a estabilização induzida por ligandos de G4 promove alterações na manutenção dos telómeros e diminui os níveis de expressão de oncogenes. A investigação atual tem evidenciado esforços contínuos de forma a utilizar ligandos para induzir a estabilização de G4 no contexto do CP, tendo diferentes classes de ligandos apresentado atividade antiproliferativa *in vitro* e, em alguns casos, efeitos anti-tumorais *in vivo*.

O objetivo desta tese foi desenvolver novos ligandos de G4 derivados de 1,10-fenantrolina, capazes de ligar/estabilizar estruturas de G4 presentes em regiões regulatórias importantes no CP, atuando como agentes antitumorais.

Neste trabalho, compostos heterocíclicos com núcleos derivados de fenilo, quinolina, naftaleno, acridina e fenantrolina foram avaliados relativamente ao seu efeito estabilizador de estruturas de G4 nos promotores de c-MYC, KRAS e VEGF no motivo telomérico humano e no G4 de RNA pre-MIR150. Um dos derivados de acridina avaliados foi o que apresentou a maior capacidade de estabilização nos diferentes G4s, seguindo-se de um derivado de fenantrolina. A maioria dos compostos demonstrou uma maior eficácia anti-tumoral em células de CP quando comparado com células não-malignas.

Posteriormente, uma nova classe de derivados de fenantrolina foi sintetizada. O seu potencial para ligação/estabilização de sequências formadoras de G4 foi avaliado através de ensaios de transferência de energia de ressonância de Förster (FRET), estudos de dicroísmo circular e de titulações de fluorescência. A atividade antitumoral *in vitro* destes compostos foi avaliada em linhas celulares de CP através do ensaio de MTT.

Dezasseis derivados de 1,10-fenantrolina, contendo substituições nas posições 2 e 9 com cadeias laterais de amina ou amida, foram sintetizados por reações de condensação direta. Entre estes derivados, os com cadeias laterais metoxianilino e etanodiamínio demonstraram capacidade de estabilizar várias topologias de G4 derivadas de sequências teloméricas humanas, com preferência pela topologia híbrida da sequência F21T (ΔT_m =7,2 – 12 °C, determinada por experiências de FRET). Os ligandos apresentaram ainda perfis citotóxicos promissores e exibiram maior eficácia em células de CP A549 comparativamente a células H1299. Além disso, foram também sintetizados dez novos derivados de 1,10-fenantrolina-2,9-bistriazóis através de reações de cicloadição azida/alcino catalisadas por cobre. Através de avaliação biofísica, foram identificados

três ligandos promissores em sequências G4 de KRAS, promovendo o aumento na estabilização térmica (ΔT_m =4,7 – 11,2 °C, determinada por experiências de FRET) e afinidades de ligação variando entre 10^{-6} e 10^{-9} M. A avaliação da atividade anti-tumoral revelou que o composto contendo um anel fenilo ligado à unidade de triazole exibiu um potente efeito inibitório no crescimento das células tumorais A549 e H1299 (IC_{50} = 14,6 e 10,9 μM, respetivamente). Os resultados experimentais demonstraram que os compostos sintetizados têm potencial para se ligar/estabilizar diferentes sequências formadoras de G4, e que podem ser úteis na otimização posterior de ligandos. Além disso, este estudo forneceu informações importantes sobre a relação estrutura-atividade para esta classe de compostos, com o objetivo de maximizar a sua atividade.

Por último, e até onde sabemos, identificámos, pela primeira vez, uma região formadora de G4 dentro do precursor humano do MIR150 (pre-MIR150). O G4 formado apresentou uma topologia paralela com potencial de interagir com a nucleolina (NCL), proteína sobreexpressa na superfície celular de células cancerígenas envolvida em vários processos celulares como a tumorigénese, angiogénese e vias de sinalização. A estabilidade térmica do G4 aumentou na presença do ligando comercial de fenantrolina PhenDC3, tendo sido observada a formação de um complexo ternário G4/PhenDC3/NCL. A estrutura G4 no pré-MIR150 reconheceu células de CP positivas para a NCL que pode ser potencialmente utilizada como sonda em biópsias líquidas.

Palavras-chave

G-quadruplex; ligandos de G-quadruplex; cancro do pulmão; derivados de fenantrolina; síntese química; técnicas biofísicas; atividade anti-tumoral

Resumo alargado

As estruturas de G-quadruplex (G4s) são estruturas secundárias não-canónicas de ácidos nucleicos que se podem formar em regiões do genoma ou transcriptoma ricas em guanina. Estas estruturas são formadas pelo empilhamento planar de dois ou mais arranjos de quatro guaninas, unidas por ligações de hidrogénio de *Hoogsteen* (G-quartetos). A estabilidade do empilhamento dos G-quartetos é favorecida pela presença de catiões metálicos, predominantemente catiões monovalentes como o sódio (K^+) e potássio (Na^+). Em termos de estereoquímica, os G4s podem ser classificados como unimoleculares ou intermoleculares, adotando topologias paralela, antiparalela ou hibrida. Esta classificação é determinada pela orientação das cadeias de DNA ou RNA, bem como do comprimento e composição do *loop*.

Sequências capazes de formar G4s de DNA são maioritariamente encontradas em regiões teloméricas e promotores de oncogenes de células tumorais, enquanto os G4s de RNA são encontrados em regiões não codificantes. A prevalência destas estruturas em regiões importantes de regulação biológica tornou-as alvos terapêuticos atrativos. Neste contexto, a utilização de moléculas de baixo peso molecular (<500Da) capazes de se ligarem ou estabilizarem G4s contribuiu para o desenvolvimento de novas estratégias terapêuticas e de diagnóstico. Ao longo dos anos, várias moléculas têm sido descritas como ligandos de G4, apresentando atividade anti-tumoral *in vitro* e/ou *in vivo*. Normalmente, os ligandos de G4 podem interagir com o G-quarteto por empilhamento π-π e interações electroestáticas. Considerando a grande área do G-quarteto, um ligando eficiente deve apresentar uma superfície heteroaromática planar e grupos básicos ou carregados positivamente capazes de interagir com os grupos fosfato dos ácidos nucleicos. Considerando estes requisitos, várias classes de compostos pertencentes a diferentes famílias químicas, incluindo acridinas, quindolinas, quinazolonas, quinolonas, naftalenodiimidas, perilenodiimidas e fenantrolinas têm vindo a ser desenvolvidos como ligandos de G4. Estas moléculas são capazes de alterar o crescimento tumoral através da: i) inibição da atividade da telomerase ou interferindo com a função dos telómeros; ii) modulação da expressão de oncogenes através da estabilização de G4s nos seus promotores; iii) regulação génica pós-tradução através de estabilização de G4 em regiões não transcritas de mRNA; iv) regulação por impedimento do desenrolamento da helicase ou v) indução de mutações.

O cancro do pulmão (CP) é a principal causa de morte em todo o mundo, com uma estimativa de 1,8 milhões de mortes em 2021. Clinicamente, é classificado em dois

subtipos principais, o CP de pequenas células e CP de não-pequenas células, representando aproximadamente 15% e 85% de todos os casos, respetivamente. Histologicamente, o CPNPC é dividido em adenocarcinoma, carcinoma de células escamosas, carcinoma de células grandes e CPNPC sem especificação, que inclui o CPNPC que não possui as características dos outros tipos. O fumo do tabaco continua a ser o principal fator de risco para o desenvolvimento do CP, no entanto, outros fatores, como a poluição do ar, mutações genéticas e polimorfismo de nucleótido único, também estão associados ao CP. Embora a cirurgia e a quimio/radioterapia sejam eficazes numa fase inicial da doença, a sua eficácia é reduzia em estadios avançados de CP, onde a maioria dos casos são diagnosticados. O avanço da medicina de precisão, através da utilização de variações genéticas associadas ao CP, incluindo mutações em proto-oncogenes como os membros de recetor do fator de crescimento epidérmico (EGFR), MYC e a família RAS têm melhorado significativamente os resultados do tratamento. Adicionalmente, a utilização de bloqueadores de pontos de controlo imunológico que visam o ligando de morte programada (PD-L1) ou a proteína de morte programada (PD-1), tem mostrado resultados promissores e respostas duradouras no tratamento. Apesar destes avanços, a resistência aos medicamentos e o elevado grau de heterogeneidade dos tumores dificultam o tratamento, sendo necessária a implementação de novas estratégias terapêuticas.

A prevalência de estruturas de G4 em regiões regulatórias importantes para o desenvolvimento do CP, incluindo proto-oncogenes (MYC, BCL-2, KRAS, KIT, EGFR, HIFα, VEGF e WNT) e regiões teloméricas, têm sido reportadas como sendo alvos terapêuticos promissores para o tratamento do CP. Como observado para outros tipos de cancro, a estabilização das estruturas de G4 induzida pelos ligandos promove alterações dos telómeros e reduz os níveis de expressão de oncogenes associados ao CP. Esforços têm sido realizados para compreender o efeito induzido pelos ligandos na estabilização de estruturas de G4 envolvidas na progressão e desenvolvimento do CP, sendo que diferentes ligandos de G4s, utilizados isoladamente ou em combinação, demonstraram atividade anti-proliferativa *in vitro* e, em alguns casos, efeitos anti-tumorais *in vivo*.

Neste contexto, o principal objetivo desta tese de doutoramento consistiu no desenvolvimento de novos derivados de fenantrolina capazes de ligar e estabilizar diferentes sequências de G4 relacionadas com a progressão do CP, com o intuito de alcançar efeitos anti-tumorais. Adicionalmente, este trabalho teve como objetivo explorar o potencial de uma sequência de G4 de RNA para reconhecer e ligar à proteína nucleolina (NCL), sobreexpressa em células tumorais do CP.

Numa primeira fase, procedeu-se à avaliação de diferentes compostos contendo núcleos heteroaromáticos planares como oxazole, acridina, naftaleno e fenantrolina com o objetivo de explorar o seu efeito na estabilização de diversas estruturas de G4 presentes em oncogenes humanos (c-MYC, KRAS e VEGF), sequências teloméricas e G4 de RNA. Os derivados contendo os núcleos de naftaleno e fenantrolina foram sintetizados com rendimento que variam entre baixos e elevados. O efeito na estabilização das estruturas de G4 induzida pelos diferentes compostos foi avaliado por experiências de FRET. Os derivados de acridina demonstraram uma maior capacidade de estabilização em todas as sequências de G4 testadas, sendo que o derivado C_8NH_2 induzido uma estabilização mais acentuada na sequência telomérica humana com topologia híbrida F21T (ΔT_m = +26.8 °C). O composto $PhenCOOCH_3$ provou ser o derivado de fenantrolina com maior capacidade de estabilização de G4, induzindo maior estabilização térmica nas sequências FKrasT e F21T (ΔT_m = +3,4 °C), demonstrando seletividade para células do CP, com valor de IC_{50} de 6,4 μM.

Após o sucesso na avaliação dos diferentes compostos heteroaromáticos, e com o intuito de aumentar as suas propriedades de estabilização, procedeu-se à funcionalização do núcleo de 1,10-fenantrolina através da introdução de cadeias laterais de amina e amida nas posições 2 e 9. Desta forma, foi sintetizada e estruturalmente caraterizada uma biblioteca de dezasseis derivados 2,9-dissubstituidos de 1,10-fenantrolina. Os compostos foram obtidos com rendimentos moderados a bons. Entre os derivados sintetizados, aqueles que apresentam cadeias laterais de metoxianilino (**2b**) e etanodiamínio (**3a** e **5b**) demonstraram induzir estabilização térmica em todas as sequências teloméricas testadas (F21T, F21CTAT e F25CebT), com preferência para a topologia híbrida da sequência F21T (ΔT_m = + 7.2 - 12 °C, determinada por experiências de FRET). Os derivados sintetizados demonstraram atividade antiproliferativa moderada em linhas celulares de CP (A549 e H1299), sendo mais eficazes na linha celular A549. O ligando **5b** provou ser o derivado com maior atividade antitumoral, induzindo uma viabilidade celular de 24.5 % em células A549, sugerindo uma atividade mediada pela estabilização de G4.

De forma a aumentar a diversidade química dos derivados de 1,10-fenantrolina, foram sintetizados e caracterizados dez novos derivados de 1,10-fenantrolina-2,9-bistriazóis. Estes derivados foram sintetizados por reação de cicloadição de azida/alcino catalisadas por cobre, com rendimentos moderados a bons. Embora os diferentes derivados não tenham demonstrado capacidade de induzir estabilização térmica nas sequências teloméricas testadas, demonstraram efeitos estabilizadores promissores em sequências de G4 presentes no KRAS. Três derivados contendo grupos terminais de *N´N*-

dimetilmetanamina (**4a**), *N´N*- dimetilmetanamínio (**4b**), e piperazínio (**5b**) provaram ter a capacidade de se ligar e estabilizar a estrutura de G4, com valores de ΔT_m entre 4.7 e 11.2 °C e valores de K_D variando entre 10^{-6} a 10^{-9} M. Apesar da capacidade de estabilização de estruturas de G4 presentes no oncogene KRAS, os derivados sintetizados não demonstraram atividade antitumoral promissora, com exceção do derivado 2a, que apresentou um valor de IC_{50} de 14.6 e 10.9 μM em células A549 e H1299, respetivamente.

Por último, e, pela primeira vez, a estrutura de G4 presente no precursor humano do MIR150 (pre-MIR150) foi identificada e estudada. Esta estrutura foi utilizada como molécula de reconhecimento da nucleolina (NCL) expressa na superfície celular de células de CP e nas células mononucleares do sangue periférico humano (PBMCs) de doentes diagnosticados com CP. Na presença de iões de K^+ ou quando complexada com o ligando comercial PhenDC3, a estrutura de G4 apresentou uma topologia paralela. A interação G4/NCL ou G4-PhenDC3/NCL apresentou um valor de K_D na ordem do nanomolar. Adicionalmente, e através da técnica de microscopia de confocal, foi observada maior co-localização do complexo G4/PhenDC3 com a NCL em células A549 e PBMCs. Estes resultados indicam que o complexo G4/PhenDC3 tem potencial para ser utilizado como agente de reconhecimento da NCL.

Os resultados obtidos durante esta tese de doutoramento permitiram obter informações importantes sobre a relação estrutura-atividade deste tipo de ligandos, que serão necessárias para otimizações posteriores. Alguns compostos sintetizados mostraram capacidade de se ligar e estabilizar diferentes estruturas de G4 envolvidas na progressão do CP, evidenciando, em alguns casos, atividade antitumoral promissora, destacando-se a necessidade de estudar o seu mecanismo de ação. Adicionalmente, este trabalho permitiu a caracterização de uma estrutura de G4 no pre-MIR150 com alta afinidade para a NCL, estabelecendo uma potencial aplicação para o diagnóstico do CP.

Palavras-chave

G-quadruplex; ligandos de G-quadruplex; cancro do pulmão; derivados de fenantrolina; síntese química; técnicas biofísicas; atividade anti-tumoral

Abstract

G-quadruplexes (G4s) are non-canonical four-stranded nucleic acid secondary structures that can form in guanine-rich regions in the human genome and transcriptome. DNAs G4s are found in critical regulatory regions of the human genome such as the end of telomers and the promoter regions of several oncogenes. These structures have been implicated in the control of key cellular events including replication, transcription, genome stability, and epigenetic regulation. On the other hand, RNAs G4 are found within non-coding regions and are implicated in crucial RNA metabolism events, including the regulation of RNA processing and translation. Due to their biological relevance and structural features, G4s are considered suitable therapeutic targets. The prevalence of G4 folding in the cancer context and its stabilization can be used as an anti-tumor treatment strategy in different cancer types. Several synthetic small molecules with high specificity for G4s DNAs and RNAs relative to double-stranded DNA (G4 ligands) have been developed and evaluated for their therapeutic potential. Over the years, thousands of small molecules have been reported as G4 ligands with *in vitro* and/or *in vivo* anti-tumor activity. Often, but not always, G4 ligands have a planar aromatic core for π-π stacking with G-quartet and a positively charged or basic group(s) to interact with the phosphate backbone of nucleic acid. Different scaffolds containing these chemical features, such as quindolines, acridines, naphthalene diimides, and phenanthrolines have proved to have G4 binding properties and anti-tumor activities. These small organic molecules interacting with G4s may affect cancer cell growth in different ways, including inhibition of telomerase or interference with telomere function, modulation of oncogenes expression by stabilizing G4 in their promoter and regulating microRNA biogenesis (miRNAs) by stabilizing G4 in microRNA precursors (pre-miRNAs).

Lung cancer (LC) is the leading cause of cancer-related death worldwide and is divided into two major subtypes, small-cell lung cancer (SCLC), and non-small cell lung cancer (NSCLC). Although therapeutic options such as surgery and chemoradiotherapy prove efficacy in the early stage of the disease, their effectiveness is limited in advanced LC where most patients are diagnosed. Although immune checkpoint inhibitors have shown promising clinical outcomes, the high level of molecular heterogeneity in LC makes treatment difficult and new therapeutic strategies are needed. The prevalence of G4 formation in important regulatory systems in LC including proto-oncogenes (e.g. MYC, BCL-2, KRAS, KIT, and VEGF) and telomere regions make then attractive targets. Similarly to other cancer types, G4 ligand-induced stabilization promotes changes in

telomere maintenance and decreases oncogene expression levels. Ongoing efforts have been made to harness G4 ligands for inducing G4 stabilization in the LC context. Different classes of G4s ligands may lead to *in vitro* antiproliferative activity and, in some cases, *in vivo* anti-tumor effects.

Overall, this thesis aims to develop novel G4 ligands derivatived from 1,10-phenanthroline, which can bind/stabilize G4 structures present in important regulatory regions in LC acting as anti-tumor agents.

The chemical synthesis and screening of heterocyclic compounds featuring phenyl, quinoline, naphthalene, acridines, and phenanthroline scaffolds were assessed to explore their stabilization effect towards G4 structures in c-MYC, KRAS and VEGF promoters, human telomeric motif, and pre-MIR150 RNA G4. The acridine moiety exhibited the highest G4 stabilization, followed by phenanthroline. Additionally, most compounds proved greater anti-tumor efficacy in LC cells than in non-malignant cells.

Subsequently, a novel class of phenanthroline derivatives was synthesized and structurally characterized. In addition, to assess their potential to bind/stabilize G4-forming sequences, several biophysical techniques, including förster resonance energy transfer (FRET) melting assay, circular dichroism (CD) studies and fluorescence titrations were employed. The *in vitro* anti-tumor activity of these compounds has been evaluated against LC cell lines *via* MTT assay.

Sixteen derivatives of 1,10-phenanthroline were synthesized, featuring substituents at positions 2 and 9 with either amine or amide side chains, by direct condensation reactions. Among these derivatives, those bearing methoxyaniline and ethane-diaminium side chains have been shown to stabilize various G4 topologies derived from human telomere sequences. These included hybrid, parallel, and antiparallel G4 structures, with a marked preference for the hybrid topology of the F21T sequence (ΔT_m =7.2 – 12 °C, by FRET melting). Additionally, the ligands displayed promising cytotoxic profiles and exhibited greater efficacy against A549 than H1299 LC cells. Additionally, novel ten 1,10-phenanthroline-2,9-bistriazoles derivatives were synthesized *via* cooper catalyze azide/alkyne cycloaddition reactions. Through biophysical assessment, three promising G4 ligands against KRAS G4 sequences were identified, exhibiting increases in melting temperature (ΔT_m =4.7 – 11.2 °C, by FRET melting) and binding affinities ranging between 10^{-6} and 10^{-9} M. Evaluation of anti-tumor activity revealed that the compound bearing a phenyl ring linked to the triazole moiety exhibited a potent inhibitory effect on A549 and H1299 cancer cell growth (IC_{50} = 14.6 and 10.9 μM,

respectively). Overall, the experimental results demonstrated that the synthesized compounds have the potential to bind/stabilize different G4-forming sequences and may serve as promising scaffolds for the development of G4 ligands. Additionally, this study provides invaluable insights into the structure-activity relationships for this class of compounds to maximize their activity.

Finally, to the best of our knowledge, we have identified, for the first time, a G4-forming region within the human precursor of MIR150 (pre-MIR150). The G4-forming region folds into a parallel RNA G4 and has the potential to interact with nucleolin (NCL), which is a protein overexpressed on the cell surface of cancer cells involved in several cellular processes, including tumorigenesis, angiogenesis, and signaling pathways. The thermal stability of G4 increases in the presence of a commercial phenanthroline ligand (PhenDC3), and the formation of ternary complex G4/PhenDC3/NCL was observed. Moreover, the G4 structure in pre-MIR150 recognized NCL-positive LC cancer cells, and liquid biopsies when fluorescently labeled, can be used as a probe.

Keywords

G-quadruplex; G-quadruplex ligands; lung cancer; phenanthroline derivatives; chemical synthesis; biophysical techniques; anti-tumor activity.

Table of Contents

List of Figures

List of Tables

List of Schemes

List of Abbreviations

ε	Molar extinction coefficient
δ	Chemical shift
Δ	Reflux
ΔT_m	Change in melting temperature
A2780	Ovarian carcinoma cell line
A375	Melanoma cell line
A431	Epidermoid carcinoma cell line
A549	Non-small cell lung carcinoma cell line
ADC	Adenocarcinoma
AGS	Gastric adenocarcinoma cell line
ALT	Alternative lengthening of telomeres
ATCC	American type culture collection
ATP	Adenosine Triphosphate
Auoxo6	Binuclear gold(III)-complex
BCL-2	B-cell Lymphoma-2
BG4	The monoclonal antibody that targets DNA G-quadruplex
Boc	*tert*-butyloxycarbonyl protecting group
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 1 susceptibility protein
CD	Circular Dichroism
cex-NDI	Core-extended naphthalene diimides
CLEAR	Coordinated lysosomal expression and regulation
cPDS	carboxy-pyridostatin
CTLA-4	Cytotoxic T lymphocyte antigen-4
CuAAC	Copper-catalyzed azide/alkyne cycloaddition reaction
d	Duplet
DDR	DNA damage response
DEPT-135	Distortionless enhancement by polarization transfer-135
DMSO	Dimethyl sulfoxide
DMSO-d_6	Dimethyl sulfoxide-d_6
DU145	Prostate carcinoma cell line
EGFR	Epidermal growth factor receptor
ELISA	Enzyme linked immunosorbent assay
EMSA	Electrophoresis mobility shift assay

EPI	Epiberberine
ERCC1	Excision repair cross-complementation group 1
ESI	Electrospray ionization
ESI-MS	Electrospray ionization Mass Spectrometry
FAM	Fluorescein
FBS	Fetal bovine serum
FDA	Food and drug administration
FRET	Fluorescence resonance energy transfer
G	Guanine
G4	G-quadruplex
G4RP	RNA G-quadruplex-specific precipitation
GRO	Guanine-rich oligonucleotides
H1299	Non-small cell lung carcinoma cell line
H2AX	H2A histone family member X
H460	Lung carcinoma cell line
HeLa	Uterus adenocarcinoma cell line
Hep2B 2.17	Liver cancer cell line
HepG2	Hepatocellular carcinoma cell line
hERG	human Ether-a-go-go Related Gene
HL-60	Acute promyelocytic leukemia cell line
HOP-62	Lung adenocarcinoma cell line
HR	Homologous recombination
HRD	Homologous recombination deficient
HR-MS	High-resolution mass spectrometry
HT-29	Colorectal adenocarcinoma cell line
hTelo	Human telomeric G-quadruplex
h-TERT	Human telomerase reverse transcriptase
ITC	Isothermal titration calorimetry
J	Coupling constant
K_A	Association constant
K_D	Dissociation constant
LC	Lung cancer
LCC	Large cell carcinoma
LMP	Lysosomal membrane permeability
LNA	Locked nucleic acid
lncRNA	Long non-coding RNA
m	Multiplet

MCF-7	Breast adenocarcinoma cell line
MDA-MB-231	Breast adenocarcinoma cell line
MIA-PaCa2	Pancreas carcinoma cell line
miRNA	Micro RNA
MRC-5	Lung fibroblast cell line
mRNA	Messenger RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NCI-H226	Lung squamous cell carcinoma cell line
NCI-H23	Non-small cell lung adenocarcinoma cell line
NCI-H322M	Bronchioalveolar carcinoma cell line
NCI-H460	Lung carcinoma cell line
NCI-H522	Non-small cell lung cancer adenocarcinoma cell line
NCL	Nucleolin
ncRNA	Non-coding RNA
NDIs	Naphthalenediimides
NER	Nucleotide excision repair
NGS	Next-generation sequencing
NHDF	Normal human dermal fibroblasts cell line
NHE III	Nuclease hypersensitivity element
NMM	*N*-methyl mesoporphyrin IX
NMR	Nuclear magnetic resonance
NMR ^{1}H	Nuclear magnetic resonance of proton ^{1}H
NMR ^{13}C	Nuclear magnetic resonance of carbon ^{13}C
NOS	Non-small cell lung cancer otherwise specified
n-ROTB	Number of rotatable bonds
NSCLC	Non-small cell lung cancer
PALB2	Gene that encodes a BRCA2-interacting protein
PANC-1	Pancreas carcinoma cell line
PBMCs	Peripheral blood mononuclear cells
PC-3	Prostate adenocarcinoma
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
PDS	Pyridostatin
PhenDC3	3,3′-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate
piRNA	Piwi-interacting RNA
POT1	Protection Of Telomeres 1

pre-miRNA	Precursor-micro RNA
pri-miRNA	Primary-microRNA
PTEN	phosphatase and tensin homologue
q	Quartet
rDNA	ribosomal DNA
rG4s	RNA G-quadruplex
rRNA	ribosomal RNA
RISC	RNA-induced silencing complex
rt	Room temperature
s	Singlet
SAR	Structure-activity relationship
sb	Large singlet
SCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
SGC7901	Gastric cancer cell line
SK-RC-42	Renal carcinoma cell line
SPR	Surface plasmon resonance
sRNAs	Small RNA
SW620	Colon adenocarcinoma cell line
t	triplet
TAMRA	Tetramethylrhodamine
TERRA	Telomeric repeat-containing RNA
TFEB	Transcription factor EB
TLC	Thin layer chromatography
T_m	Melting temperature
TMPyP4	5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine
TNM	Classification of malignant tumors
TPSA	Topological polar surface area
TRF1	Telomeric repeat binding factor 1
TRF2	Telomeric repeat binding factor 2
TSS	Transcription start site
UTR	Untranslated regions
UV-Vis	Ultraviolet-visible
UXF1138L	Uterine carcinoma cell line
VEGF	Vascular endothelial growth factor
WTN1	Wingless-type MMTV integration site family, member 1

List of Scientific Publications

Publications related to this doctoral thesis:

Figueiredo, J.; Mergny, J-L. and Cruz, C. (2024) G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives. *Life Sciences* 340, 122481. DOI: 10.1016/j.lfs.2024.122481. IF: 6.1

Figueiredo, J.; Djavaheri-Mergny, M.; Ferret, L.; Mergny, J-L. and Cruz, C. (2023) Harnessing G-quadruplex ligands for lung cancer treatment: A comprehensive overview. *Drug Discovery Today* 28, 103808. DOI: 10.1016/j.drudis.2023.103808. IF: 7.4

Figueiredo, J.; Carreira-Barral, I.; Quesada, R.; Mergny, J-L. and Cruz, C. (2022) Synthesis and evaluation of 2,9-disubstituted-1,10-phenanthroline derivatives as G-quadruplex binders. *Bioorganic & Medicinal Chemistry* 73, 116971. DOI: 10.1016/j.bmc.2022.116971. IF: 3.5

Figueiredo, J.; Peitinho, D.; Campello, M.P.C.; Oliveira, M.C.; Paulo, A.; Mergny, J-L. and Cruz, C. (2022) Screening of scaffolds for the design of G-quadruplex ligands. *Applied Sciences* 12, 2170. DOI: 10.3390/app12042170. IF: 2.7

Figueiredo, J.[‡]; Santos, T.[‡]; Miranda, A.[‡]; Alexandre, D.; Teixeira, B.; Simões, P.; Lopes-Nunes, J. and Cruz, C. (2021) Ligands as stabilizers of G-quadruplex in non-coding RNAs. *Molecules* 26, 6164. DOI: 10.3390/molecules26206164. IF: 4.927

Figueiredo, J.; Miranda, A.; Lopes-Nunes, J.; Carvalho, J.; Alexandre, D.; Valente, S.; Mergny, J-L. and Cruz, C. (2021) Targeting nucleolin by RNA G-Quadruplex forming motif. *Biochemical Pharmacology* 189, 114418. DOI: 10.1016/j.bcp.2021.114418. IF: 6.100

[‡] Equal Contribution

Publications unrelated to this doctoral thesis:

Moreira, D.; Leitão, Daniela.; Lopes-Nunes, J.; Santos, T.; **Figueiredo, J**.; Miranda, A.; Alexandre, D.; Tomaz, C.; Mergny, J-L. and Cruz, C. (2022) G-quadruplex Aptamer-Ligand Characterization. *Molecules* 27, 6781. DOI:10.3390/molecules27206781. IF: 4.927

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: 3.9

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: 3.847

Lopes-Nunes, J.[‡]; Carvalho, J.[‡]; **Figueiredo, J.**[‡]; Ramos, C. I. V.; Lourenço, L. M. O.; Tomé, J. P. C.; G.P.M.S. Neves, M.; Mergny, J.-L.; Queiroz, J. A.; Salgado, G. F. and Cruz, C. (2020) Phthalocyanines for G-quadruplex aptamers binding. *Bioorganic Chemistry* 100, 103920. DOI: 10.1016/j.bioorg.2020.103920. IF: 5.275

Figueiredo, J.[‡]; Lopes-Nunes, J.[‡]; Carvalho, J.[‡]; Antunes, F.; Ribeiro, M.; Campello, M. P. C.; Paulo, A.; Paiva, A.; Salgado, G. F.; Queiroz, J. A.; Mergny, J.-L. and Cruz, C. (2019) AS1411 derivatives as carriers of G-quadruplex ligands for cervical cancer cells. *International Journal of Pharmaceutics* 568, 118511. DOI:10.1016/j.ijpharm.2019.118511. IF: 4.845

Eddahmi, M.; Moura, N. M. M.; Bouissane, L.; Faustino, M. A. F.; Cavaleiro, J. A. S.; Paz, F. A. A.; Mendes, R. F.; **Figueiredo, J.**; Carvalho, J.; Cruz, C.; Neves, M. G. P. M. S. and Rakib, E. M. (2019) Synthesis and Biological Evaluation of New Functionalized Nitroindazolylacetonitrile Derivatives. *ChemistrySelect* 4, 14335–14342. DOI: 10.1002/slct.201904344. IF: 1.811

[‡] Equal Contribution

1st SECTION

Introduction and aims

1st Chapter

G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives

This chapter was published in:

***Figueiredo, J.**; Mergny, J-L. and Cruz, C. (2024) G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives. Life Sciences 340, 122481. DOI: 10.1016/j.lfs.2024.122481*

Chapter overview

This chapter provides an updated and comprehensive overview of the role of G4 ligands in cancer therapy. Additionally, it addresses the use of some G4 ligands that may lack the properties needed to reach (pre)clinical evaluation but still be useful as probes, as well as some methods to access G4 interactions. Over the years, several small molecule chemotypes, particularly heteroaromatic molecules, have been reported for their ability to bind and stabilize G4. The structural properties of heteroaromatic compounds contribute to their strong binding affinity for G4s structures. However, some challenges in G4 targeting are frequently reported and discussed within this chapter. The ability to modify functional groups and substitution patterns on heteroaromatic rings allows for the optimization of G4 ligands binding specificity, selectivity, and "drug-like" character. Throughout the chapter, promising heterocyclic-based G4 ligands with *in vivo* efficacy are presented and discussed. Overall, the comparative insights and structure-activity relationship covered in this chapter emphasize the importance of developing G4 ligand-based drugs for clinical therapeutic transformation, considering pharmacological issues, and insights from clinical trials.

G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives

Abstract

Guanine-rich sequences can form G-quadruplexes (G4) in living cells, making these structures promising anticancer targets. Compounds able to recognize these structures have been investigated as potential anticancer drugs; however, no G4 binder has yet been approved in the clinic. Here, we describe G4 ligands structure-activity relationships, *in vivo* effects as well as clinical trials. Addressing G4 ligand characteristics, targeting challenges, and structure-activity relationships, this review provides insights into the development of potent and selective G4-targeting molecules for therapeutic applications.

Keywords

G-quadruplex ligand; ligand design; heterocycle compounds; structure-activity relationship; biological activity; drug-DNA interactions

Introduction

G-quadruplexes (G4s) are non-canonical four-stranded nucleic acid secondary structures formed by the stacking of two or more guanine quartets (G-quartets). Each G-quartet is composed of four guanines arranged in a square co-planar configuration, connected by Hoogsteen hydrogen bonds (**Figure 1.1**) [1]. G4 formation is strongly dependent on metal cations (most often monovalent cations such as Na^+ or K^+), coordinated with the guanine carbonyl oxygen O6 atoms pointing towards the inner channel formed at the center of G-quartets [2]. G4s are highly polymorphic and can be formed intermolecularly or intramolecularly. These structures can adopt different conformations, including parallel, antiparallel, or hybrid topologies, depending on the relative strand orientation, number or quartets, length and composition of the loops, and additional elements such as capping base pairs or triads and flanking nucleotides (**Figure 1.1**) [1]. While DNA G4s exhibit greater topological diversity, RNA G4s predominantly adopt a so-called parallel conformation, in which all guanines adopt an

anti-glycosidic bond angle and all four strands run in the same direction [3]. The characterization of G4 structures can be validated through the use of multiple spectroscopic and biophysical techniques, including NMR spectroscopy [4] X-ray crystallography [5], and circular dichroism (CD) [6].

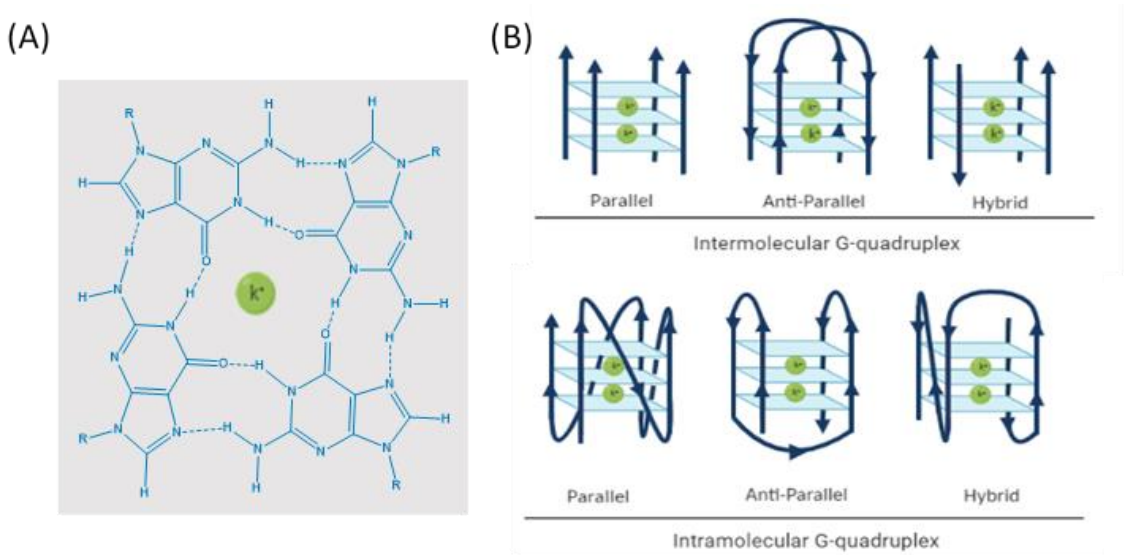


Figure 1.1. G4 structures. **(A)** Illustration of a G-quartet formed by four Hydrogen-bonded guanines and a central potassium ion (K^+). **(B)** Different topologies of G4 structures [1].

Accumulating evidence for the formation of G4s in living cells [7–9] and their involvement in regulatory functions and diverse gene expression processes make these structures attractive targets for drug design [10]. In this sense, developing ligands capable of binding and stabilizing G4 structures may not only allow the characterization of these structures *in vitro* and in cells, but also contribute to the development of novel therapeutic and diagnostic strategies. Many of these ligands interact with G4s in a relatively indiscriminate manner, meaning that they recognize G4s of different topologies, proving their potential as multi-targeting agents. These molecules interacting with G4 may affect cancer cell growth in different ways: i) inhibition of telomerase or, in a broader sense, interference with telomere function [11]; ii) modulation of oncogene expression by stabilizing G4 in their promoters [12]; iii) post-translational gene regulation by targeting mRNA 5′-untranslated regions G4 [13]; iv) regulation by impediment of helicase unwinding [14] or v) induction of mutations [15]. G4 polymorphism offers the possibility to design compounds able to recognize one topology, rather than simply aiming at differentiating duplex and G4 [16]. Some high resolution structures of G4s complexed with small molecules unveil that G4s can contain unique capping and loop structures that are specifically recognized by ligands, offering a structural basis for rational drug design [17]. Recent efforts in this direction are worth

noting. For example, Granzhan and colleagues designed a library of several cationic bis(acylhydrazones) [18]. Most ligands exhibited consistent behavior when interacting with different G4-DNA targets, as shown by "*a strong mutual correlation of ΔT_m values obtained with different G4 structures*". There was one interesting outlier though, which, "*in addition to anomalously high stabilization of all four G4-DNA, demonstrated a significant bias towards the antiparallel G4 22CTA*." Another attractive strategy to improve specificity among G4 would be to target a "double-G4", i.e., a region capable of forming two juxtaposed G4s. A recent report illustrates the design of a "double ligand" susceptible to recognize these double G4s [19]. Previous studies also reported the use of berberine-bisquinolinium conjugates as selective, sensitive, and fluorescent sensors for human telomere double-G4 [20]. Moreover, new strategies to exploit more ligands targeting telomeric multimeric G4 were reported [21]. Innovative studies focus on unraveling the mechanistic details of interactions of small molecules to specifically bind a parallel topology. A novel G4 ligand, CD-34, was identified thanks to small molecule microarray screening to selectively target the parallel conformation of c-MYC G4 and exhibits smaller stabilization for BCL-2 and minimal for VEGF, KRAS, MYB, and HIF1α. CD-34 led to a downregulation of MYC expression levels with a weaker effect on other genes [22,23]. NMR experiments revealed that phenyl-ethenyl-quinoline (PEQ) ligand specifically binds to the major c-MYC G4 [24].

Despite these efforts, few ligands have entered pre-clinical evaluation or early phases of clinical trials and none of them has progressed beyond phase II clinical trials so far. Some ligands, such as BRACO-19 (compound **1**), RHPS4 (compound **3**) quarfloxin (compound **9**), and APTO-253 (compound **16**) were withdrawn due to lack of efficacy or toxicity issues (see below) [25–28]. However, CX-5461 (compound **10**, see below) has achieved significant development for safety, tolerability and pharmacokinetics and remains in clinical evaluation [29] (**Figure 1.2**). Since no targeted therapy has received clinical approval in pancreatic ductal adenocarcinoma (PDAC) an important milestone was reached with the recent introduction of QN-302 (compound **14**) in phase I clinical trial in patients with advanced or metastatic solid tumors such as PDAC, sarcomas, prostate cancer and gastrointestinal stromal tumors (NCT06086522) [30] (**Figure 1.2**).

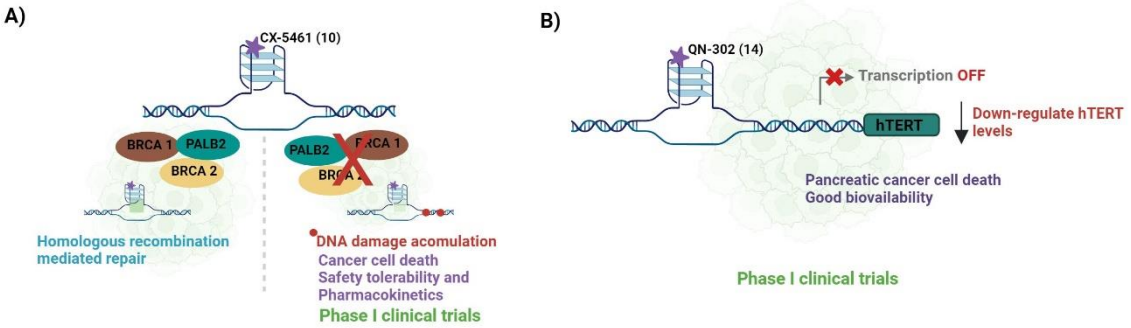


Figure 1.2. Effects of G-quadruplex (G4) ligands. CX-5461(**10**) and QN-302 (**14**) ligands are in phase I clinical trials evaluation in patients with **(A)** solid tumors and BRCA1/2, PALB2, or homologous recombination deficiency mutation [85-89], and **(B)** solid tumors (metastatic, or advanced cancer), including pancreatic cancer [99-100]. CX-5461 entered phase Ib clinical trials and QN-302 recently entered phase I.

We believe that comparative insights and structure-activity relationship (SAR) approaches can help the research community for the development of more potent, specific, and selective G4-targeting molecules. Thus, we focus on some promising heteroaromatic G4 binders, particularly the most active *in vivo*. We will start to emphasize some G4 ligand characteristics and frequent challenges in G4 targeting followed by an overview of some SAR approaches, and their significance in the development of G4 ligands for therapeutic purposes.

G4 ligands as probes and/or drugs

While a large number of organic small molecules recognizing G4 structures have been developed as drugs, some compounds may lack the properties needed to reach (pre)clinical phase. Yet, these molecules may still be useful as probes to track G4 formation in vitro or in cells. The requirements of a perfect G4 probe may be significantly different from those needed to develop a G4-based drug. In the former case, the probe should have minimal cytotoxicity or be used as sub-cytotoxic doses. Ideally, these probes should not actively promote G4 formation but only bind to pre-existing targets (a few G4 ligands have been reported to act as chaperones and should be avoided; see below). Based on their action mode small-molecule probes can be divided into several major classes [31]. The first is “light-up” probes which are molecules that exhibit fluorescence enhancement upon G4 binding. The second is “light-off” probes that exhibit a decrease in fluorescence upon G4 binding. The third class is “permanent probes”, in which molecules exhibit no alteration fluorescence upon G4 interaction. The fourth class would

encompass compounds having different emission maxima when free and bound to G4: binding to a quadruplex would lead to a change in fluorescence color. Finally, nonfluorescent compounds may also be useful, for example when radioactively labeled [32] or by acting as competitors for fluorescent compounds. Several promising probes have been developed which include molecules that can recognize both RNA and DNA structures, display a preference for a given G4 topology (often, but not always, parallel G4s), with the ultimate goal of recognizing a specific G4 in cells [33]. Optimized core-extended naphthalene diimides (cex-NDI) may also be exploited as G4 light-up sensors. Interestingly, these cexNDI compounds were able to differentiate among various G4 topologies, showing a preference for parallel G4 [34]. However, the quest for ideal probes remains challenging since their affinity and specificity is generally weaker than antibodies.

In addition, we demonstrated that some, but not all, G4 ligands not only bind to and lock preformed G4 structures but also act as G4 chaperones and actively promote their formation by increasing their kinetics of association, as shown for a bisquinolinium pyridine dicarboxamide derivative [35]. G4 chaperones, while attractive as drugs, may pose problems as probes, as they would reveal G4 formation in sequences that would have remained unfolded - or differently folded - in the absence of the ligand. Unfortunately, this possibility is rarely investigated.

Assessing G4-ligands interactions

Although several classes of chemical compounds have played a crucial role in probing the biology of G4, many G4 ligands do not have "*sufficient affinity or selectivity for interrogating G4 cellular functions with confidence*" [36]. It is therefore important not to overlook an early and seemingly trivial analysis of the binding properties of candidates. The most appropriate parameter to compare ligands would be to determine dissociation constants for various nucleic acid structures at physiological temperature and under near-physiological conditions. Unfortunately, reliable dissociation constants (K_D) values cannot always be determined, due to ligand self-stacking interactions, multiple binding sites and various stoichiometries of binding. In addition, this determination is somewhat time-consuming, and difficult to perform when comparing dozens of compounds or during high-throughput screening. For this reason, semi-quantitative methods based on melting assays using changes in absorbance, Förster Resonance Energy Transfer (FRET), or ellipticity are often used. While the FRET melting

assay can easily be adapted to 96- or 384-well format, it only provides a crude estimate of binding strength at 37 °C. Marchand et al. [37] found that, in most cases, ligand-G4 binding is entropically driven, and this can be due when ligand potency is deduced from melting techniques. It would therefore be advisable to design a screening strategy based on an isothermal assay as the one proposed by Lacroix et al [38]. This test, initially designed to identify compounds binding to the G4-prone 5' end of the human telomerase RNA (hTR), can be easily adapted to any G4 forming sequence and changed into human telomere sequence. Additionally, the use of commercially available G4 ligands with negligible duplex activity such as pyridostatin (PDS) [39], carboxy-pyridostatin (cPDS) [40], BRACO-19 [41] and PhenDC3 [42] can be useful to study the G4-interactions. The employment of these pioneering G4 ligands can be important for the development of other G4-targeting small molecules; they may also be used as reference compounds when new molecules are tested. Unfortunately, none of these reference compounds is highly fluorescent, arguing for the need for novel attractive fluorescent derivatives.

Main challenges in G4 targeting

The design of effective G4 binders is dependent on the available high-resolution G4 structure or G4-drug complexes (obtained by NMR [43], and/or X-ray crystallography [44]). Unfortunately, only a few have been completely characterized.

The general binding feature of G4 ligands typically includes a heteroaromatic moiety. These ligands are commonly stacked on the outer surface of the quartet, forming additional hydrogen bonds or/and electrostatic interactions with substituents and/or the core within the grooves and loops. For example, charged amide links can improve rigidity within a molecule, and alkyl links can increase the flexibility of the ligand. Typical examples of terminal stacking (5′ and/or 3′) of G-quartet are a macrocyclic hexaoxazole and naphthalene diimides (NDI) [44,45]. Some cases of ligands like binuclear gold(III)-complex (Auoxo6), Epiberberine (EPI), or a tripodal cationic fluorescent probe, NBTE, are examples of the 5′-end cap or loop binders [46–48]. These ligands are selective for G4 over duplex DNA and EPI also converts other telomeric G4 to the hybrid-2 G4 topology. The probe NBTE also converts the antiparallel form of telomeric G4 into a hybrid one. More surprisingly, clear exceptions to these guidelines are either non-planar molecules such as steroid derivatives that selectively stabilize the G4 [49] or neutral or even negatively charged ligands such as *N*-Methyl Mesoporphyrin IX (NMM), which is

selective for G4 versus duplex DNA and tend to prefer parallel folds over antiparallel ones [50].

Several efforts have been made to introduce chemical functionalities onto the lead/scaffold ligand to enhance G4 ligand selectivity and/or specificity, keeping in mind that these modifications should not compromise the drug-like properties of the ligand itself [51]. Most of G4 ligands failed in terms of "drug-likeness" properties [52] and/or exhibited little or no correlation between their in vitro activity and in vivo efficacy. For example, some G4s ligands pose aqueous solubility issues, which can be overcome by incorporating cationic or ionizable groups or ring heteroatoms. In addition, large planar aromatic ligands tend to self-aggregate, which may affect membrane permeability, cellular uptake, and chemical and enzymatic stability. Overall, most compounds do not abide to Lipinski's rule [53]. Although Lipinski's Rule of Five does not need to be "slavishly followed" [54], it cannot be completely ignored. Considerations about physico-chemical properties underlying the Rule of Five such as molecular weight and logP, can help to avoid cellular uptake and membrane permeability problems [54]. Despite several ligands exhibiting promising properties in terms of G4 in vitro stabilization the in vivo efficacy has often been disappointing. One should also remember that specific G4 interaction with proteins, epigenetic modifications in promoters, cell proliferation-dependent transcriptional activity, presence of single nucleotide polymorphisms, and telomere protein composition may interfere with the ligand-G4 interaction. As mentioned before, several studies have been done to determine the specificity of ligands towards a specific G4 sequence and encouraging results have been obtained [23,24]. However, these observations are mostly supported by biophysical data. Considering that there are thousands of putative G4 sequences in cells, more cell-based assays to measure ligand binding affinity are needed. For instance, some ligands show reduced MYC protein expression levels. However, MYC protein levels can be controlled by several pathways, and the development of assays to determine G4-driven transcription is needed to avoid off-target effects. Transcriptome-wide studies of the effects of G4 ligands often reveal pleiotropic effects (as observed with the **20A** compound, for example [55]) demonstrating that many genes are affected, either directly or indirectly. Establishing which genes are selectively affected would require short treatments coupled with nascent RNA analysis.

Moreover, even if a fair number of G4 compounds reached the preclinical phase, only a few entered clinical trials. Interestingly, the phase I clinical trial results revealed 14 % rate of partial responses among CX5461 treated female patients with cancers exhibiting mutations in the BRCA1/2 or PALB2 genes [29]. Encouraged by the identification of

synthetic lethal interactions between CX-5461 and BRCA1/2 mediated by homologous recombination (HR) genetic interaction screens with CX-5461 and other G4-targeting were recently performed [56]. PDS and PhenDC3 reveal gene sensitivity and novel genetic interactions. This opens an opportunity to develop drug combination therapies for G4- targeting drugs, aiming to enhance clinical outcomes [56].

Some evidence indicates that G4 binders are able to modulate innate immunity genes and autophagy in cancer cells at non-toxic doses. In a recent review, Capranico and colleagues proposed the use of G4 binders at low concentrations to boost antitumor immunity, in combination with standard therapeutics [57]. This scenario requires specific attention in the choice of appropriate immune proficient animal models.

Additionally, accurate models such as 3D, organotypic cultures, and patient-derived xenografts will be valuable to obtain a more realistic demonstration of the therapeutic potential of G4 ligands [51]. To avoid problems related to off-target toxicity such as that induced by BRACO19 and AS1410 an initial toxicological assessment should be considered, namely in terms of hERG (human *Ether-a-go-go Related Gene*) inhibition.

Efforts to develop G4 ligands that go beyond mere G4 binding, aiming to enhance their in vitro stability or increase their in vivo lifespan are needed. This requires a substantial investment in novel synthetic methodologies [58].

G4 ligands: targeting telomeres and telomerase: a case study?

Historically, G4 ligands were designed to target human telomeric repeats [59]. Telomeres are the ends of linear chromosomes. In humans and all vertebrates, telomeric sequences are composed of the repetition of the hexanucleotide TTAGGG. This sequence is prone to G4 formation, and initial efforts in G4 ligand development were aimed at the design of indirect telomerase inhibitors, acting on the DNA substrate and locking it into an inactive conformation. Since then, it was found that quadruplexes were not always inhibitory and could recruit telomerase [60], and most G4 ligands behave as inhibitors of telomeric DNA amplification rather than telomerase extension [61]. G4 ligands cannot therefore be considered “simple” telomerase inhibitors, and their anti-telomerase activity has often been overestimated due to caveats in the classical TRAP assay. For example, while telomestatin has an IC_{50} in the low nM range in this test, its activity drops

by three orders of magnitude when using a direct method to measure telomerase activity [61]. In addition, several G4 ligands are active in telomerase-negative alternative lengthening of telomeres (ALT) cancer cells. This does not mean that G4 ligands do not act on telomeres, as they may inhibit DNA polymerases and helicases activity on telomeric substrates causing telomere uncapping by delocalization of proteins from telomeres. For this reason, a number of compounds are tested for their affinity for the human telomeric quadruplex composed of (GGGTTA) repeats, and this family of G-sequences is by far the most studied G4 alone or for its interaction with G4 ligands. The extreme polymorphism of the telomeric quadruplex is both an asset and a nuisance: while it complicates structural analyses, it allows multiple conformations to coexist, offering different kinds of binding epitopes to the ligands tested. Given that the energetic difference between these conformers is relatively low, preference for a given conformation may be deduced by a possible shift towards a given topology.

Promising heterocyclic-based G4 ligand classes

A few classes of heterocyclic compounds have been synthesized with different aromatic cores, alkylation warheads, or hydrophilic spacers to improve solubility, binding or tissue penetration **(Table 1.1)**. In the next sections, we present examples of these ligands.

Table 1.1. List of G4 ligands discussed in this chapter.

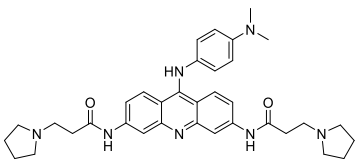
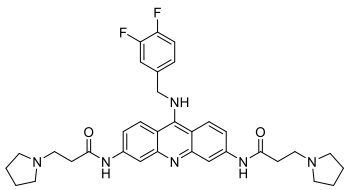
Ligand	Target G4	Tumor type	Effect	Ref
 BRACO-19 (**1**)	Telomere	Uterine cancer	Reduced hTERT expression *In vivo* tumor growth inhibition in early stage	[41,63,64]
 AS1410	Telomere	Non-small cell lung cancer (in combination with cisplatin)	-	[66,67]

Table 1.1 (*continued*)

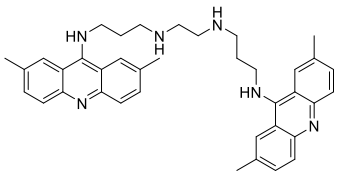
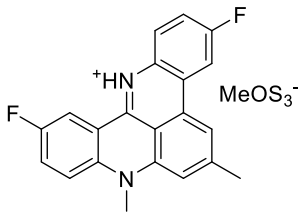
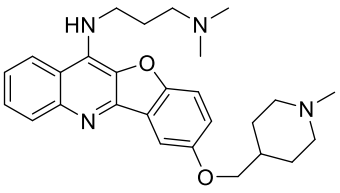
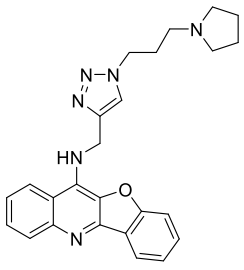
Ligand	Target G4	Tumor type	Effect	Ref
 **2**	c-MYC	Cervical cancer	Downregulation of c-MYC transcription Cell-cycle arrest and apoptosis *In vivo* tumor growth inhibition	[69]
 RHPS4 (**3**)	Telomere	Breast cancer Melanoma cancer Prostate cancer Lung cancer Colorectal cancer	Telomerase inhibition Reduction in telomere length DNA damage *In vivo* tumor growth inhibition	[70-74]
 **4**	c-MYC	Burkitt's lymphoma	Downregulation of c-MYC transcription *In vivo* tumor growth inhibition	[77]
 **5**	c-MYC	Lung cancer	Downregulation of c-MYC transcription *In vivo* tumor growth inhibition	[78]

Table 1.1 (*continued*)

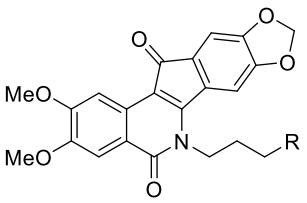
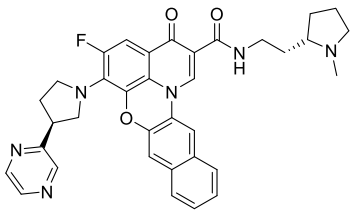
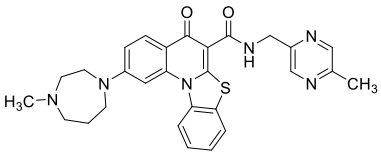
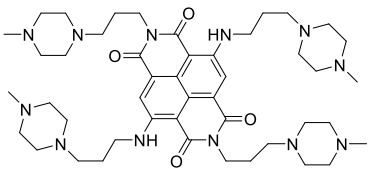
Ligand	Target G4	Tumor type	Effect	Ref
 **6** R=4-methylmorpholine **7** R= 1-methyl-1*H*-imidazole **8** R= 2-(methylamino)ethan-1-ol	c-MYC	Solid Tumors Lymphoma	Downregulation of c-MYC Topoisomerase I inhibition *In vivo* tumor growth inhibition	[79]
 Quarfloxin (**9**)	rDNA	Carcinoid/ neuroendocrine tumors Breast cancer Pancreatic cancer	Disrupt rDNA G4/nucleolin interaction Inhibition of RNA Polymerase I transcription Cell apoptosis and DNA damage *In vivo* tumor growth inhibition	[82,83]
 CX-5461(**10**)	Multitarget	BRCA1/2-deficient tumors Pancreatic cancer Melanoma	inhibition of RNA-polymerase I Cell apoptosis and DNA damage *In vivo* tumor growth inhibition	[85-89]
 **11**	HSP90 Telomere	Pancreatic Cancer	Decrease in telomerase activity Reduce HSP90 and hTERT protein expression levels *In vivo* tumor growth inhibition	[93-94]

Table 1.1 (*continued*)

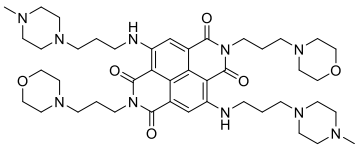
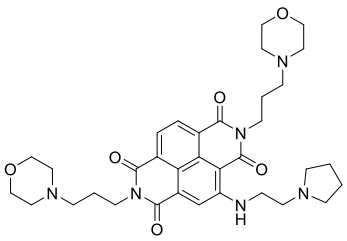
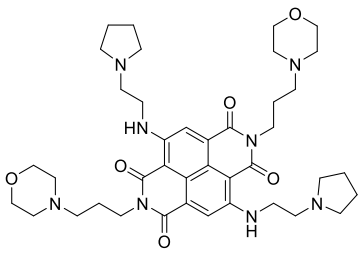
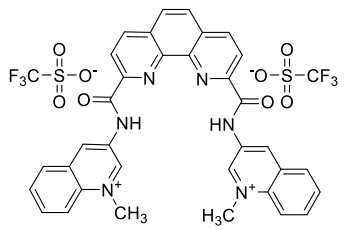
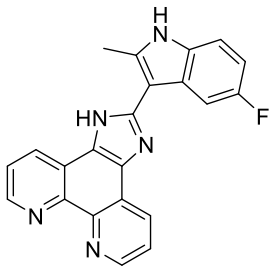
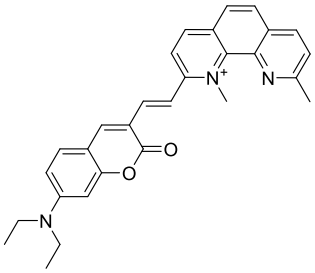
Ligand	Target G4	Tumor type	Effect	Ref
 **12**	Multitarget	Pancreatic Cancer	Reduces BCL-2 levels *In vivo* tumor growth inhibition	[95-97]
 **13**	Multitarget	Pancreatic cancer	Downregulation of multiple gene targets *In vivo* tumor growth inhibition Promotion *in in vivo* survival rate in the KPC model	[98]
 **14**	Telomere	Pancreatic cancer	Downregulation of multiple gene targets *In vivo* tumor growth inhibition/regression	[99,100]
 **15**	Multitarget	-	G4-unfolding inhibition by helicases *In vitro* and *in vivo* displace of the heme bound to G4	[76-81]
 **16**	Multitarget	Myeloid leukemia Pulmonary Fibrosis	Cell-cycle arrest and DNA damage Induction of the KLF4 tumor suppressor	[82,83]

Table 1.1 (*continued*)

Ligand	Target G4	Tumor type	Effect	Ref
 **17**	BCL-2	Breast cancer Lung Cancer	Reduction of BCL-2 mRNA level Cell apoptosis *In vivo* tumor growth inhibition	[84]

Acridines

The 3,6,9-trisubstituted acridine derivative 1, called BRACO-19, was developed by Neidle and colleagues in 2001, to bind human DNA G4, inhibiting telomerase in the elongation of single-stranded telomere overhang [62]. The binding affinity to human telomeric G4 determined through Surface Plasmon Resonance (SPR) is high (K_A = 1.6 × 107 M^{-1}), and BRACO-19 showed potent telomerase inhibition with an IC_{50} of 0.06 μM in A2780 human ovarian carcinoma cells, inducing senescence or selective cell death [63]. The *in vitro* activity of **1** was evaluated in human uterus carcinoma UXF1138L cells, which exhibit short telomeres; an IC_{50} of 2.5 μM was reported. Treatment at 1 μM for 24 h led to a dramatic decrease in human telomerase reverse transcriptase (hTERT) nuclear expression, but residual cytoplasmic hTERT staining was observed accompanied by atypical mitoses [41]. Intraperitoneal administration of compound 1 to early-stage UXF1138L xenograft models inhibits tumor growth, displaying single-agent antitumor activity efficacy, with complete regressions observed in some animals but not advanced-stage xenografts. In contrast, oral administration regimen was found to be inactive [64]. The combination of **1** (at a nontoxic dose) with the anticancer drug Paclitaxel displayed a synergistic antitumor activity in A431 epidermal carcinoma [65] (**Figure 1.3**). Despite **1** representing one the most potent cell-free inhibitors of human telomerase with high affinity for G4, some biopharmaceutical limitations became apparent. Its sub-optimal pharmacological profile (half-life of 1.1 h), limited bioavailability after an hour, lack of membrane permeability, and small therapeutic window hampered further development [25,64]. AS1410 (not to be confused with AS1411, a G-rich oligonucleotide) is derived from BRACO-19 with a difluorobenzylamino group at the 9-position [66]. AS1410 showed increased hydrophobicity, improved binding affinity, and extended plasma half-

life (2h) compared with BRACO-19 [67]. AS1410 also demonstrated synergy with cisplatin, leading to a significant reduction in tumor volume in an A549 lung cancer xenograft model when administered in combination [68] (**Figure 1.3**). However, it showed modest activity alone in the DU145 xenograft models and several other tumor types at a tolerated dose. Since the maximum tolerated doses for both AS1410 and BRACO-19 remain relatively low (2 mg/kg) the program was abandoned.

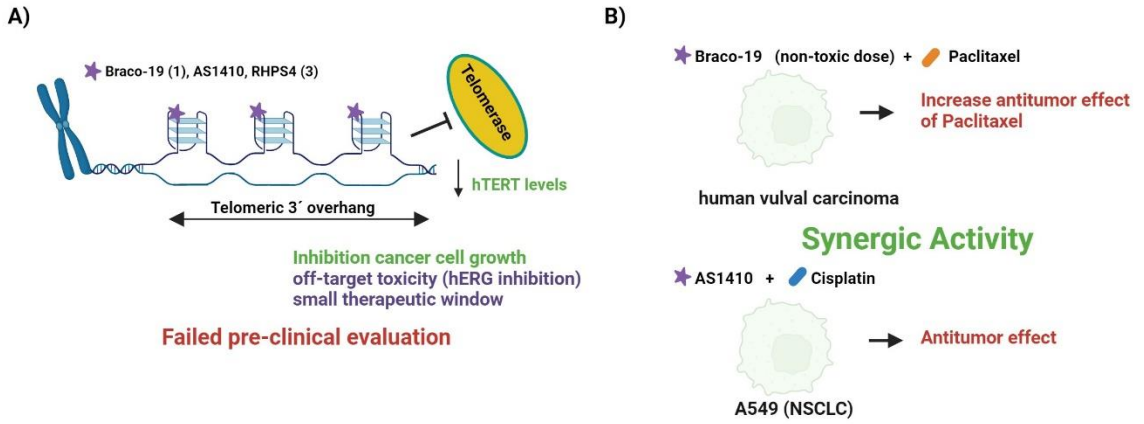


Figure 1.3. Effects of G-quadruplex (G4) ligands administered as **(A)** standalone agents or **(B)** in combination with anti-cancer drugs. **(A)** BRACO-19 (1), AS1410, and RHPS4 (**3**) ligands can modulate the hTERT expression levels by stabilization of G4 in the 3′-overhang of chromosome ends. These ligands have pharmacological limitations and failed pre-clinical evacuation. **(B)** BRACO-19 and AS1410 showed synergistic *in vivo* antitumor effects with Paclitaxel [63,64] and Cisplatin [66,67], respectively.

A series of disubstituted bisacridine derivatives was synthesized to bind both G4 and i-motif (a different class of DNA quadruplexes formed by C-rich sequences) on the c-MYC gene [69]. The compounds containing a long alkyl amine linker stabilized both c-MYC G4 and i-motif. Compound 2 was the most potent c-MYC G4 stabilizer according to FRET-melting results (Fpy33T: ΔT_m = +11 °C and Fpu22T: ΔT_m = +9.9 °C at 1 μM (5 M equivalents)). Moreover, **2** inhibited cervical SiHa cell short-term proliferation with an IC_{50} value of 1.22 μM, induced cellular apoptosis, and led to cell-cycle arrest cells by suppressing c-MYC transcription (**Figure 1.4**). These effects were associated with the formation of G4s in the S phase of the cell cycle and the i-motifs found in the late G1 phase. The derivative 2 also suppressed tumor growth in a SiHa cervical xenograft model, inducing a reduction in tumor weight of 41.6 % [69].

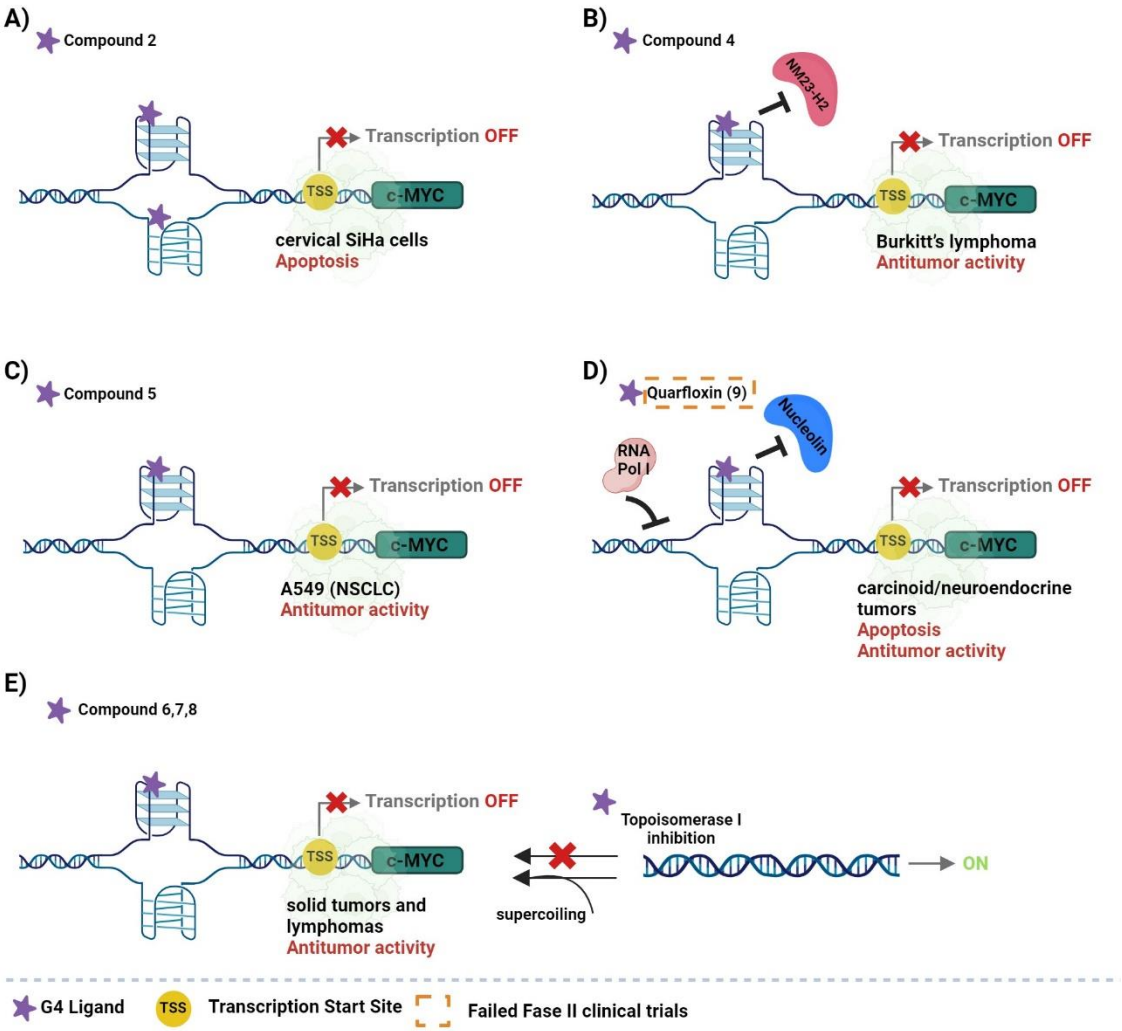


Figure 1.4. Examples of effects of G-quadruplex (G4) ligands in the promoter region of c-MYC promoter gene. **(A)** Compound **2** stabilizes c-MYC G4 and i-motif with *in vivo* activity in the cervical SiHa cancer model [69]. **(B)** Compound **4** downregulates c-MYC transcription by stabilizing the c-MYC G4 and disrupting the binding of the protein NM23-H2 to DNA. In a human Burkitt's lymphoma xenograft model presented *in vivo* activity [77]. **(C)** Compound **5** shows *in vivo* activity in lung cancer xenograft model, possibly caused by c-MYC downregulation [78]. **(D)** Quarfloxin (**9**) stabilizes c-MYC G4 [82-83]. disrupts the interaction with nucleolin, consequently inhibiting Pol I transcription, and triggering apoptotic cell death in cancer cell. Quarfloxin did not proceed past Phase II clinical trials in patients with carcinoid/neuroendocrine. **(E)** Compounds **6**, **7** and **8** are human topoisomerase I inhibitors that induce DNA damage and may overcome drug resistance, possibly through binding to c-MYC G4 [79].

To increase π-π stacking, the tricycle system of the acridine was modified to a tetracyclic or pentacyclic system. Compound **3**, better known as RHPS4, is a pentacyclic acridinium derivative that promotes G4 formation in the telomeric overhang. RHPS4 is a potent telomerase inhibitor (IC_{50} = 0.33 μM) and a DNA damage inducer [70]. 3 induces a

senescent-like growth arrest in MCF-7 cells (0.5-1 μM, 15 days) and a reduction in telomere length. This activity was observed at noncytotoxic drug concentrations that are suitable for extended drug administration [71]. Human breast tumor xenografts showed a correlation between telomere length and growth Inhibitory potency of this compound [71]. These results align with the disruption of the protective telomere capping status achieved through the stabilization of G4 structure by compound **3**, leading to greater susceptibility of cells with shorter telomeres [71]. In addition, **3** induced a rapid and potent DNA damage response at telomeres, characterized by the phosphorylation of H2AX in human fibroblasts expressing hTERT and melanoma cells [72]. Different tumor xenografts were used to evaluate the antitumor effect of 3 [72]. Notably, CG5 breast xenografts displayed high sensitivity with marked inhibition (about 80 %) after treatments with compound **3**, and this effect persisted for at least 30 days and 40 % were cured.

The anticancer effects of **3** were linked to telomere dysfunction and were countered in tumors that overexpress TRF2 or POT1 [72]. The suppression of tumor growth was accompanied by a strong DNA damage response which resulted in telomere damage and apoptosis within the xenografted tumors. [72]. Despite its attributes as a telomere-targeting agent and desirable pharmaceutical profile [73], RHPS4 has notable off-target cardiotoxicity effects in a guinea pig model [26]. This was related to a potent inhibition of the hERG (human Ether-a-go-go Related Gene) receptor. In addition, interaction with other targets involved in abnormal cardiac function as β2 adrenergic and muscarinic receptors was observed. Two RHPS4 acetylamine derivatives were developed and showed a reduction in hERG activity while maintaining the on-target features [26]. Additional developments resulted in structurally-related compounds. The most promising showed 40 % of hERG inhibition but retained strong interaction with the M2 muscarinic receptor (99 % inhibition) [74]. Despite its promising antitumor activity, RHPS4 and related compounds did not advance to clinical trials because of these offtarget effects, highlighting the importance of initial pharmacological screening.

Overall, whether tricyclic compounds or extended acridines such as **3** are optimal for selective recognition remains an open question: while a pentacyclic scaffold offers a larger stacking area with a terminal quartet, up to three lateral chains may be appended to a tricyclic core. Increasing the length of side chains on acridine derivatives and including terminal groups such as pyrrolidine and piperazine may result in a significant loss in affinity for the telomeric G4. These ligands may rather be designed to target the G-quartet and the duplex stem-loop of a promoter quadruplex, enabling stabilization of the stem-loop resulting in the downregulation of transcription and induction of cell

death. In addition, fusing these acridines to carbazoles or thiazole orange scaffolds may be appropriate for cellular investigations through fluorescence lifetime imaging microscopy.

Quindolines

SYUIQ-5 is an indoloquinoline ligand substituted at C-11 by a positively charged nitrogen atom, allowing π–π stacking and electrostatic interactions between the indoloquinoline moiety and the G4 [75,76]. Based on this scaffold a series of disubstituted quindolines derivatives were synthesized with the introduction of a second alkylamino side chain at the 7-, 8-, or 9-positions of the quindoline ring [77]. This allowed to add positive charges through methylation at the nitrogen atom at the end of the side chain. The best ligand was substituted by alkylamines side chains at 7-,11- positions, with ΔT_m values ranging from +13.5 to +36.7 °C at 2 μM (10 M equivalents) to c-MYC G4 determined by FRET-melting and a K_D of 10-7 M, determined by microscale thermophoresis [77]. In Burkitt's lymphoma cells, this disubstituted quindoline derivative **4** downregulated c-MYC transcription by stabilizing the c-MYC G4 and disrupting binding of the protein NM23-H2 to DNA, revealing itself as a promising anticancer agent. However, in a human Burkitt's lymphoma xenograft model, this derivative presented in vivo activity with a reduction in tumor volume of only 27.4 % compared to the negative control [77] (**Figure 1.4**).

To improve the selectivity/affinity of quindolines to G4s, a series of novel triazole-containing benzofuroquinoline derivatives was developed using a 1,3-dipolar cycloaddition of a series of alkyne and azide building blocks [78]. The ΔTm values for c-MYC G4 obtained by FRET melting experiments varied between +16 and + 22 °C at 2 μM (5 M equivalents); triazole functionalization enhanced the interaction between ligand and G4 [78]. The derivatives bearing long *N*-side chains evidenced better activity against c-MYC G4; they were able to downregulate c-MYC gene transcription and its expression in Raji cells. Compared to the used positive control, doxorubicin, the propylpyrrolidine-based derivative **5** showed similar tumor growth inhibition rates (about 38 %) in lung cancer xenograft model, possibly caused by c-MYC downregulation [78] (**Figure 1.4**).

Overall, the equilibria between the protonated/deprotonated forms of these quindoline species should be carefully considered to improve binding. Quindoline derivative compounds containing electrondonating amino substituents at the aromatic skeleton increase the in vivo activity mediated by G4 binding by enhancing the basicity of the pyridine nitrogen atom and its protonation at physiologic pH. One should keep in mind

that these modifications may reduce drug-like properties, increase molecular weight and cationic charge.

Quinazolones and quinolones

The indenoisoquinoline derivatives indotecan (**6**, NCT01051635), indimitecan (**7**, NCT01051635), and **8** (NCT03030417) progressed to phase I clinical trials at the National Cancer Institute for treatment of solid tumors and lymphomas. These compounds are human topoisomerase I inhibitors that induce DNA damage and may overcome drug resistance, possibly through binding to c-MYC G4 [79]. Structure–activity of c-MYC G4 binding by these derivatives showed strong KD values in the nanomolar range and moderate stabilization (ΔT_m > 6 °C at 1.5 μM (10 M equivalents) by FRET-melting). These compounds cause both c-MYC downregulation and topoisomerase I inhibition, offering a dual approach to drug development [79] (**Figure 1.4**).

The fluoroquinolone derivative **9**, also known as Quarfloxin, was the first G4 stabilizer to enter clinical trials (NCT00955786) [80]. This compound was rationally designed from norfloxacin and fluoroquinolone QQ5863 G4 ligand by optimization of the planar core and addition of presumable groove-binding arms to increase the selectivity for G4 over duplex structures and eliminating a topoisomerase II poisoning effect [81]. **9** exhibited *in vivo* antitumor activity against mice breast xenografts models. Daily treatment was well-tolerated, suppressing tumor growth. The treatment in MIA PaCa-2 pancreatic xenografts resulted in significant tumor growth inhibition with minimal impact on animal body weight [82]. Quarfloxin (**9**) was initially recognized as a specific binder for the c-MYC G4. However, further studies revealed that the interaction between **9** and G4 structures in ribosomal DNA (rDNA) quadruplexes disrupted the interaction with nucleolin, consequently inhibiting Pol I transcription, and triggering apoptotic cell death in cancer cells [82] (Fig. 4). Quarfloxin (**9**) showed robust evidence of biological activity in Phase II clinical trials (NCT00485966) and proved to be well tolerated with a satisfactory toxicity profile. However unusual pharmacokinetic properties suggest an extensive biodistribution profile. Quarfloxin accumulates in tissues derived from the neural crest at a significantly higher level than those observed in plasma. Due to bioavailability issues, quarfloxin did not progress to phase II clinical trial [27,83]. Another quarfloxin derivative, CX-5461 (**10**), was developed by the reduction of the 9 heteroaromatic core and side arms size [84]. Theoretically, these structural modifications should increase the “drug-like” properties of the 10 but reduce the G4 interaction properties. However, also this fluoroquinolone derivative proved to be a good

G4 stabilizer (c-MYC, ckit1, and h-Telo; ΔT_m > 15 °C at 1 μM (5 M equivalents) by FRET-melting). 10 induced G4-mediated DNA damage and an increase in in vivo G4 structures [84]. After oral administration (50 mg/kg), 10 showed in vivo antitumor activity in pancreatic xenograft models and melanoma (A375) with tumor growth inhibition of 69 % on day 31 and 79 % on day 32, respectively [85]. The anticancer activity of 10 was also related to the inhibition of RNA-polymerase I [86] and further evaluated in a phase I clinical trial in advanced hematological malignancies showing modest activity [87] and good pharmacokinetics properties. Recent evidence suggests that the primary action mechanism is topoisomerase II poisoning in regions containing G4s [88,89]. In a phase I clinical trial (NCT02719977) CX-5461 exhibited synthetic lethality in homologous recombination deficient (HRD) models, establishing a clinical proof-of-concept for this Gquadruplex stabilizer [29]. A dose-limiting photosensitivity was noted. CX-5461 is now in a phase Ib study (NCT04890613) to determine a tolerable dose administered by IV infusion in patients with solid tumors and BRCA1/2, PALB2, or HRD mutation for future Phase II trials. Unfortunately, CX-5461 also causes extensive, nonselective and collateral mutagenesis in different cell lines, to "magnitudes that exceed known environmental carcinogens" [15]. This mutagenic effect could potentially contribute to future cancer risk. For an overview of CX-drugs, we refer readers to the recent review of L.H. Hurley [90].

Our vision regarding the design of quinazolones and quinolones derivatives is that the incorporation of electron-withdrawing groups, such as halogens, can improve the G4 binding properties. Conversely, the reduction of aromatic core and in amine side chains may decrease this activity. An equilibrium between the length of the amine side chains must be reached to improve bioavailability.

Naphthalenediimides and perylenediimides

Naphthalene diimides (NDIs) provide a prospect for improving both G4 binding selectivity and potency, using aromatic cores with various pendant groups, and controlling their affinity towards different G4- forming sequences [91,92].

The Neidle group reported a tetra-substituted NDI derivative, compound **11** [93]. This derivative stabilized two putative G4 forming sequences in the HSP90 promoter (FHSP90aT: ΔT_m = +34.9 °C; FHSP90bT: ΔT_m = +31.0 °C at 0.5 μM (2.5 M equivalents)) [94] and human intramolecular telomeric G4 (F21T: ΔT_m = +28.3 ◦C at 0.5 μM (2.5 M equivalents)) [93]. Moreover, the murine MIA-Pa-Ca2 xenograft model treated with BMSG-SH-3 showed a 50 % reduction in tumor volume after intraperitoneal

administration, and a 50 % decrease in telomerase activity, related with a 30 % reduction of HSP90 and hTERT protein expression levels [94]. **11** was quickly taken into cancer cell nuclei [95] and in vivo tests showed no significant off-target effects. Nevertheless, this compound was not considered a drug candidate molecule because of its high molecular weight, four positive charges (methyl-piperazine), and deviation from Lipinski's rule [94]. To improve 11 pharmacological properties, the same group studied another series of tetra-substituted NDIs derivatives using a structure-based design [44]. The tetrasubstituted NDI ligand, **12** containing two side chains with N-methyl-piperazine end groups and two side chains with morpholine end-groups emerged as a lead compound of this series. The N-methylpiperazine groups were changed to morpholine ones, and this did not affect binding to human telomeric G4 and stabilization of HSP90 promoter (FHSP90aT: ΔT_m = +33.1 °C; FHSP90bT: ΔT_m = +28.6 °C; F21T: ΔT_m = +26.6 °C at 1 μM (5 M equivalents) by FRET melting) [96]. **12** stabilized several distinct intramolecular G4s structures, e.g., human telomeric, BCL-2 promoter (ΔT_m = +26.4 °C) and two KRAS sequences (KRAS1: ΔT_m = +22.5 °C and KRAS2 ΔT_m = +19.8 °C) [96]. It was found to be more potent than **11** against several pancreatic cell lines, with an IC_{50} value of 10 nM [96]. **12** also regulated BCL-2 expression at the translational level through the targeting of the G4-forming sequences in the BCL-2 promoter region and displayed in vivo anti-tumor activity in a MIA PaCa-2 xenograft model with an 80 % growth tumor decrease after intravenous administration and no tumor regrowth seen after >200 days post-treatment [97]. The crystal structure of compound **12** within an intramolecular human telomeric parallel G4 (PDB code 3UYH) showed that the morpholine groups substituted by *N*-methyl-piperazine did not promote the loss of intermolecular interactions with the grooves of the G4; indeed the morpholine ring favored the interactions with G4 grooves and watermediated G4 side-chain hydrogen-bonding [96]. A new tri-substituted NDI termed **13** was designed containing two side chains ending in morpholino groups, and a third side chain terminating in a protonated pyrrolidino group. This compound has a reduced molecular weight and is also less basic than the protonated *N*-methyl-piperazine ring on **12**: it was anticipated that this alteration would impact the overall polarity and pharmacological availability of the compound [98]. The decrease in side chains number and in cationic charge of the 13 led to a decreased stabilization of G4 compared to **12**. Nevertheless, **13** downregulated more G4-containing genes and elevated the occurrence of double-strand break as a result of torsional strain on DNA and chromatin structure [98]. In combination with SAHA (HY-10221) compound **13** increased the level of γ-H2AX protein in MIA PaCa-2 and PANC-1 cells [98] and displayed potent antiproliferative effects against pancreatic cancer cell lines [98]. Treatment with **13** displayed in vivo antitumor activity MIA PaCa-2 xenograft

model and prolonged survival in the KPC mouse model with a superior profile than gemcitabine [98]. Transcriptome analysis indicated that the effectiveness and potency of 13 stemmed from its capability to downregulate multiple gene targets that are dysregulated in pancreatic cancer cells, most containing G4 elements within their regulatory regions [98]. In brief, tetra-substituted NDI compounds, like **12**, have one cationic acyclic side chain located within a G4 groove; each three side chains of **13** were designed to interact with a groove and the absence of a fourth substituent was offset by the tight fit of the three others. Neidle et al hypothesized that the introduction of a fourth substituent group, such as a pyrrolidine-phenyl ring closely linked to the NDI core, could enhance the affinity to G4s, giving the tetrasubstituted NDI, **14** [99]. The phenyl group is stacked on the terminal G-quartet, adopting a ruffled arrangement with the G-quartet, and the cationic tail of the phenyl residing is positioned within the fourth G4 groove, as observed in crystal structures [99]. **14** shows great affinity for hTERT G4 and HuTel21 G4 with KD values of 4.9 and 28.4 nM, respectively, and 10 times stronger inhibitory effect on pancreatic cancer cells ($GI_{50} \approx 1.5$ nM) and *in vivo* activity compared with previous NDI compounds [100]. The pharmacological properties of 14 demonstrated favorable bioavailability, in line with its *in vivo* efficacy in xenograft and genetic models for pancreatic cancer [99,100]. Transcriptome analysis revealed that 14 downregulated numerous cancer-related gene pathways, including Wnt/β-catenin signaling [100]. The development of NDI-based ligands as promising therapeutic agents is an example of the success of rational drug design for anticancer compounds mediated by G4 activity. The structure-activity relationship studies provided invaluable information for the optimization of these compounds, which were improved over the years. Among them, the most promising compounds for the treatment of pancreatic cancer were the NDI **13** and **14** [100]. NDI **14** recently entered phase Ia clinical trial in patients with solid tumors (metastatic, or advanced cancer), including pancreatic cancer (NCT06086522). In our perspective, the preparation of tetrasubstituted NDI-core compounds is encouraged since, compared to disubstituted analogs, these derivatives generally showed greater *in vivo* activity. NDI-based scaffolds, especially when functionalized with *N*-alkylmorpholine units, may exhibit enhanced anticancer properties. Therefore, *N*-alkylmorpholine-bearing cex-NDI may be attractive in theranostic approaches by combining the morpholines anti-cancer features [44] with the increased fluorescence of cex-NDI compounds [101]. Finally, as illustrated by Zuffo et al [34], mitigating the affinity of a cex-NDI may result in an increased selectivity: reducing unspecific electrostatic interactions may be key to allow a good core selectivity.

Phenanthrolines

Phenanthroline-based ligands comprise an additional category known for their strong π-π interactions with a G-quartet. De Cian et al described a 6,6'-disubstituted-2,2'-bipyridine and 2,9-disubstituted1,10-phenanthroline ligand, **15**, which is one of the most popular G4 ligands, thanks to its outstanding selectivity [102]. **15** exhibits a highaffinity binding to G4s with nanomolar (K_D) affinity and can inhibit G4-unwinding by specialized helicases [102,103]. **15** displays a marked preference for G4s over duplex and single-stranded DNAs and was previously shown to stack on the 5′-quartet of a parallel G4 structure found in the human c-MYC promoter [104]. Very recently, an unexpected mode of interaction was revealed: **15** intercalated into the antiparallel chair-type G4 adopted by the human telomeric sequence. It layed between two G-quartets on one side and a dynamic yet fully formed Gquartet capped by a T-A-A triad on the other side. **15** inhibited FANCJ and DinG helicases unwinding of an unimolecular G4 substrate in a drug concentration-dependent manner [105]. Another report showed that **15** could also displace heme bound to G4s in vitro and *in vivo*, altering transcription in cultured human cells, upregulating genes and heme oxidase that support heme degradation and iron homeostasis [106]. After that, innumerous modifications in phenanthroline core have been explored in order to improve ligand potency, selectivity and functionality, as detailed by Verga and colleagues [25]. Compound **16**, also called APTO-253, is a derivative of 2-indolyl imidazole [4,5-d] phenanthroline, that entered phase I clinical stage (NCT02267863) for treating acute myeloid leukemia [107]; however, its development was discontinued. Intracellularly, this molecule undergoes a transformation from a monomer into a ferrous complex [Fe(253)3] and both exhibit the ability to stabilize G4 structures found in telomeres, MYC, and KIT promoters but do not bind double-stranded DNA [107]. The effect of **16** on the expression of MYC and its downstream target genes, on cell-cycle arrest, DNA damage, and stress responses was proposed to result from an action on G4 DNA motifs [107]. Moreover, this compound represents a pioneering MYC inhibitor without inducing myelosuppression, rendering it particularly suitable for the treatment of acute myeloid leukemia patients with compromised bone marrow function [107]. It was recently found that 16 mediated anticancer efficacy by triggering the expression of the Krüppel-like factor 4 (KLF4) tumor suppressor and this ability was attributed to MYC inhibition [108]. Indeed, the administration of 16 in 2 different pulmonary fibrosis mouse models, was able to reverse the persistent fibrosis and idiopathic pulmonary fibrosis, which are the two most prevalent fibrotic lung diseases, with a poor prognosis and currently no available cure [108]. **17** is another phenanthroline derivative, which has a donor-acceptor neocuproine

conjugate and binds BCL-2 G4 in sub-micromolar affinity through a hybrid binding mode involving loop-stacking and groove interactions [109]. **17** reduced the mRNA level of BCL-2, inducing apoptosis of the cancer cells. The compound also demonstrated in vivo anti-tumor efficacy in mouse xenograft models of breast cancer and human lung carcinoma (A549), resulting in an average of tumor growth inhibition of 56 % and 64 %, respectively [109]. Overall, regarding the sub-class of phenanthroline compounds derived from PhenDC3, this scaffold can be considered as nearly-optimal in terms of G4 recognition and selectivity [36]. Further modifications of these compounds may therefore rather be aimed at improving cellular/ tumor uptake, sub-cellular localization, and in vivo properties. In addition, the variety of derivatives with additional desirable functionalities (e.g., fluorescence, biotin, crosslinking) designed within this family should be inspiring for other series.

Conclusion and perspectives

This review provides an overview of a range of ligands that target G4s and discuss G4-mediated in vivo effects and their limitations. Few G4 ligands with suitable pharmacokinetic properties have been discovered to date. This can be attributed to a lack of a systematic evaluation in vitro and in vivo, which represents one of the main bottlenecks for the translation of the lead ligands into the clinic. Moreover, the correlation between biophysical and biological/cellular/*in vivo* properties of G4 ligands is often relatively weak. In other words, the best ligands (in terms of affinity or selectivity in vitro) are seldom the most active compounds in cells, and some of the compounds which reached the clinic are relatively weak G4 ligands. A structure-based approach should contribute to optimize the scaffolds and confer topological preferences. For some series, further optimization may be futile (as discussed above for PhenDC3 derivatives) while a significant room for improvement exists for other families. Many questions need to be answered in order to improve the therapeutic outcome of G4 ligands. We strongly suggest carrying out a systematic and robust comparison of new compounds with commercially available benchmark G4 ligands, such as PhenDC3, PDS, PDC, 360 A, BRACO19, or RHPS4. Finally, G4 compounds are often assumed to act in the cytoplasm (for RNA targets) or in the nucleus (for both RNA and DNA targets). However, G4 ligands may target other compartments such as lysosomes and mitochondria. Recent findings suggest that G4 ligands from different chemical families have the potential to control the lysosomal pathway in different cell types (for a review, [110]). Depending on the cellular context or the compound, inhibiting or stimulating the lysosomal pathway in synergy

with a G4 ligand may impact cell fate decisions: both strategies could be of interest depending on the context and the specific disease. This example illustrates that one cannot limit the analysis of these molecules to nucleic acid targets and that other cellular components need to be considered as well. With these limitations and caveats in mind, some G4 binders available for clinical trials exhibit remarkable *in vivo* activity and, in some cases, a synergetic activity that can be combined with conventional anticancer therapies. We advocate a proper assessment of the cellular properties of these compounds to accelerate the development of G4- based therapies.

Funding

J. Figueiredo acknowledges a PhD fellowship from FCT ref. SFRH/ BD/145106/2019. J.L.M. acknowledges funding from Institut National de la Sant´e et de la Recherche M´edicale (INSERM), as well as ANR G4Access (ANR-20-CE12-0023) and INCa G4Access grants. C. Cruz acknowledges project UIDP/00709/2020, and project Instruct-ERIC Pilot R&D application ID 2473. Thanks are due to FCT/MCT for the financial support to 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, co-financed by the FEDER through COMPETE 2020, POCI, PORL, and PIDDAC.

Acknowledgments

We thank our colleagues (L. Guittat, M. Djavaheri-Mergny) for helpful discussions and all reviewers for helpful suggestions and constructive criticism.

References

[1] Burge, S. *et al.* (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* 34, 5402–5415.

[2] Bhattacharyya, D. *et al.* (2016) Metal Cations in G-Quadruplex Folding and Stability. *Front Chem* 4, 38.

[3] Joachimi, A. *et al.* (2009) A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg Med Chem* 17, 6811–6815.

[4] Papp, C. *et al.* (2023) Stable bulged G-quadruplexes in the human genome: identification, experimental validation and functionalization. *Nucleic Acids Res* 51, 4148–4177.

[5] Parkinson, G.N. and Collie, G.W. (2019) X-Ray Crystallographic Studies of G-Quadruplex Structures. *G-Quadruplex structures Nucleic Acids: Methods and Protocols*, 131–155.

[6] Vorlíčková, M. *et al.* (2012) Circular dichroism and guanine quadruplexes. *Methods* 57, 64–75.

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

[8] Summers, P.A. *et al.* (2021) Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat Commun* 12, 162.

[9] Esnault, C. *et al.* (2023) G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nat Genet* 55, 1359–1369.

[10] Tian, T. *et al.* (2018) G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem* 4, 1314–1344.

[11] de Cian, A. *et al.* (2008) Targeting telomeres and telomerase. *Biochimie* 90, 131–155.

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

[13] Gomez, D. *et al.* (2010) A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. *Nucleic Acids Res* 38, 7187–7198.

[14] Gray, L.T. *et al.* (2014) G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol* 10, 313–318.

[15] Koh, G.C.C. *et al.* (2024) The chemotherapeutic drug CX-5461 is a potent mutagen in cultured human cells. *Nat Genet* 56, 23–26.

[16] Ou, T.M. *et al.* (2008) G-quadruplexes: Targets in anticancer drug design. *ChemMedChem* 3, 690–713.

[17] Chen, L. *et al.* (2022) DNA G-Quadruplex in Human Telomeres and Oncogene Promoters: Structures, Functions, and Small Molecule Targeting. *Acc Chem Res* 55, 2628–2646.

[18] Reznichenko, O. *et al.* (2023) Optimization of G-Quadruplex Ligands through a SAR Study Combining Parallel Synthesis and Screening of Cationic Bis(acylhydrazones). *Chem Eur J* 29, e202202427.

[19] Kench, T. *et al.* (2023) Dimeric Metal-Salphen Complexes Which Target Multimeric G-Quadruplex DNA. *Bioconjug Chem* 34, 911-921.

[20] Liao, T.-C. *et al.* (2020) Human telomere double G-quadruplex recognition by berberine-bisquinolinium imaging conjugates in vitro and in cells. *Int J Biol Macromol* 158, 1299–1309.

[21] Zhao, J. and Zhai, Q. (2020) Recent advances in the development of ligands specifically targeting telomeric multimeric G-quadruplexes. *Bioorg Chem* 103, 104229.

[22] 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, 139–48.

[23] 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.

[24] 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.

[25] Taetz, S. *et al.* (2006) Biopharmaceutical Characterization of the Telomerase Inhibitor BRACO19. *Pharm Res* 23, 1031–1037.

[26] Iachettini, S. *et al.* (2013) On and off-target effects of telomere uncapping G-quadruplex selective ligands based on pentacyclic acridinium salts. *J Exp Clin Cancer Res* 32, 68.

[27] Lim, J. *et al.* (2009) Quarfloxin phase I clinical data and scientific findings supporting the selection of carcinoid/neuroendocrine tumors as the phase II indication | Cancer Research | American Association for Cancer Research. in *100th AACR Annual Meeting 2009 Proceedings*, pp. 868–869.

[28] Ohanian, M. *et al.* (2021) A Phase 1a/b Dose Escalation Study of the MYC Repressor Apto-253 in Patients with Relapsed or Refractory AML or High-Risk MDS. *Blood* 138, 3411–3411.

[29] Hilton, J. *et al.* (2022) Results of the phase I CCTG IND.231 trial of CX-5461 in patients with advanced solid tumors enriched for DNA-repair deficiencies. *Nat Commun* 13, 3607.

[30] Ahmed, A.A. *et al.* (2023) The Potent G-Quadruplex-Binding Compound QN-302 Downregulates S100P Gene Expression in Cells and in an In Vivo Model of Pancreatic Cancer. *Molecules* 28, 2452.

[31] Largy, E. *et al.* (2012) Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes. *Top Curr Chem*, 111–177.

[32] Granotier, C. *et al.* (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res* 33, 4182–4190.

[33] Yuan, J.-H. *et al.* (2020) Recent advances in fluorescent probes for G-quadruplex nucleic acids. *Biochem Biophys Res Commun* 531, 18–24.

[34] 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–e115.

[35] de Cian, A. and Mergny, J.-L. (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res* 35, 2483–2493.

[36] Verga, D. *et al.* (2023) Targeting Quadruplex Nucleic Acids: The Bisquinolinium Saga. *Handbook of Chemical Biology of Nucleic Acids*, 1–57, Springer Nature Singapore.

[37] 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.

[38] Lacroix, L. *et al.* (2011) Fluorescence-based duplex–quadruplex competition test to screen for telomerase RNA quadruplex ligands. *Nucleic Acids Res* 39, e21–e21.

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

[40] 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.

[41] 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.

[42] Chung, W.J. *et al.* (2014) Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angew Chem Int Ed* 53, 999–1002.

[43] Dai, J. *et al.* (2011) Solution Structure of a 2:1 Quindoline–c-MYC G-Quadruplex: Insights into G-Quadruplex-Interactive Small Molecule Drug Design. *J Am Chem Soc* 133, 17673–17680.

[44] 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.

[45] Chung, W.J. *et al.* (2013) Solution Structure of an Intramolecular (3 + 1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative. *J Am Chem Soc* 135, 13495–13501.

[46] Wirmer-Bartoschek, J. *et al.* (2017) Solution NMR Structure of a Ligand/Hybrid-2-G-Quadruplex Complex Reveals Rearrangements that Affect Ligand Binding. *Angew Chem Int Ed* 56, 7102–7106.

[47] Lin, C. *et al.* (2018) Molecular Recognition of the Hybrid-2 Human Telomeric G-Quadruplex by Epiberberine: Insights into Conversion of Telomeric G-Quadruplex Structures. *Angew Chem Int Ed Engl* 57, 10888–10893.

[48] Liu, L. *et al.* (2020) Quantitative Detection of G-Quadruplex DNA in Live Cells Based on Photon Counts and Complex Structure Discrimination. *Angew Chem* 132, 9806–9813.

[49] Brassart, B. *et al.* (2007) A New Steroid Derivative Stabilizes G-Quadruplexes and Induces Telomere Uncapping in Human Tumor Cells. *Mol Pharmacol* 72, 631–640.

[50] Nicoludis, J.M. *et al.* (2012) Optimized End-Stacking Provides Specificity of *N* -Methyl Mesoporphyrin IX for Human Telomeric G-Quadruplex DNA. *J Am Chem Soc* 134, 20446–20456.

[51] Alessandrini, I. *et al.* (2021) On the Road to Fight Cancer: The Potential of G-Quadruplex Ligands as Novel Therapeutic Agents. *Int J Mol Sci* 22, 5947.

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

[53] Savva, L. and Georgiades, S.N. (2021) Recent Developments in Small-Molecule Ligands of Medicinal Relevance for Harnessing the Anticancer Potential of G-Quadruplexes. *Molecules* 26, 841.

[54] Neidle, S. (2012) Design Principles for Quadruplex-binding Small Molecules. *Therapeutic Applications of Quadruplex Nucleic Acids*, pp. 151–174, Elsevier.

[55] Beauvarlet, J. *et al.* (2019) Modulation of the ATM/autophagy pathway by a G-quadruplex ligand tips the balance between senescence and apoptosis in cancer cells. *Nucleic Acids Res* 47, 2739–2756.

[56] Jin, M. *et al.* (2023) A synthetic lethal approach to drug targeting of G-quadruplexes based on CX-5461. *Bioorg Med Chem Lett* 91, 129384.

[57] Miglietta, G. *et al.* (2022) Ligands stimulating antitumour immunity as the next G-quadruplex challenge. *Mol Cancer* 21, 180.

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

[59] Sun, D. *et al.* (1997) Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J Med Chem* 40, 2113–2116.

[60] Oganesian, L. *et al.* (2007) Telomerase Recognizes G-Quadruplex and Linear DNA as Distinct Substrates. *Biochemistry* 46, 11279–11290.

[61] de Cian, A. *et al.* (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc Natl Acad Sci* 104, 17347–17352.

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

[63] Harrison, R.J. *et al.* (2003) Trisubstituted Acridine Derivatives as Potent and Selective Telomerase Inhibitors. *J Med Chem* 46, 4463–4476.

[64] 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.

[65] Gowan, S.M. *et al.* (2002) A G-Quadruplex-Interactive Potent Small-Molecule Inhibitor of Telomerase Exhibiting in Vitro and in Vivo Antitumor Activity. *Mol Pharmacol* 61, 1154–1162.

[66] Martins, C. *et al.* (2007) Structure-based design of benzylamino-acridine compounds as G-quadruplex DNA telomere targeting agents. *Bioorg Med Chem Lett* 17, 2293–2298.

[67] 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.

[68] 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.

[69] Kuang, G. *et al.* (2020) Syntheses and Evaluation of New Bisacridine Derivatives for Dual Binding of G-Quadruplex and i-Motif in Regulating Oncogene c-myc Expression. *J Med Chem* 63, 9136–9153.

[70] Phatak, P. *et al.* (2007) Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Br J Cancer* 96, 1223–1233.

[71] Cookson, J.C. *et al.* (2005) Pharmacodynamics of the G-Quadruplex-Stabilizing Telomerase Inhibitor 3,11-Difluoro-6,8,13-trimethyl-8 *H* -quino[4,3,2- *kl*]acridinium methosulfate (RHPS4) in Vitro: Activity in Human Tumor Cells Correlates with Telomere Length and Can Be Enhanced, or Antagonized, with Cytotoxic Agents. *Mol Pharmacol* 68, 1551–1558.

[72] Salvati, E. *et al.* (2007) Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J Clin Investig* 117, 3236–3247.

[73] Cookson, J.C. *et al.* (2005) Antitumor Polycyclic Acridines. 17. Synthesis and Pharmaceutical Profiles of Pentacyclic Acridinium Salts Designed To Destabilize Telomeric Integrity. *J Med Chem* 48, 7198–7207.

[74] Rizzo, A. *et al.* (2014) Identification of novel RHPS4-derivative ligands with improved toxicological profiles and telomere-targeting activities. *J Exp Cancer Res* 33, 81.

[75] Zhou, J.-M. *et al.* (2006) Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines. *Oncogene* 25, 503–511.

[76] Ou, T.-M. *et al.* (2011) Inhibition of Cell Proliferation by Quindoline Derivative (SYUIQ-05) through its Preferential Interaction with c-myc Promoter G-Quadruplex. *J Med Chem* 54, 5671–5679.

[77] Liu, H.-Y. *et al.* (2017) New Disubstituted Quindoline Derivatives Inhibiting Burkitt's Lymphoma Cell Proliferation by Impeding *c-MYC* Transcription. *J Med Chem* 60, 5438–5454.

[78] Zeng, D.-Y. *et al.* (2017) Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-b]quinolone Derivatives as c-myc G-Quadruplex Specific Stabilizers via Click Chemistry. *J Med Chem* 60, 5407–5423.

[79] Wang, K.B. *et al.* (2019) Indenoisoquinoline Topoisomerase Inhibitors Strongly Bind and Stabilize the MYC Promoter G-Quadruplex and Downregulate MYC. *J Am Chem Soc* 141, 11059–11070.

[80] Quarfloxin in Patients With Low to Intermediate Grade Neuroendocrine Carcinoma-ClinicalTrials.gov, (n.d.) https://clinicaltrials.gov/ct2/show/NCT00780663 (accessed December 20,2022).

[81] Brooks, T.A. and Hurley, L.H. (2010) Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* 1, 641–649.

[82] Drygin, D. *et al.* (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res* 69, 7653–7661.

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

[84] 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.

[85] Drygin, D. *et al.* (2011) Targeting RNA Polymerase I with an Oral Small Molecule CX-5461 Inhibits Ribosomal RNA Synthesis and Solid Tumor Growth. *Cancer Res* 71, 1418–1430.

[86] Haddach, M. *et al.* (2012) Discovery of CX-5461, the first direct and selective inhibitor of RNA polymerase I, for cancer therapeutics. *ACS Med Chem Lett* 3, 602–606.

[87] Khot, A. *et al.* (2019) First-in-Human RNA Polymerase I Transcription Inhibitor CX-5461 in Patients with Advanced Hematologic Cancers: Results of a Phase I Dose-Escalation Study. *Cancer Discov* 9, 1036–1049.

[88] Bruno, P.M. *et al.* (2020) The primary mechanism of cytotoxicity of the chemotherapeutic agent CX-5461 is topoisomerase II poisoning. *Proc Natl Acad Sci U S A* 117, 4053–4060.

[89] Bossaert, M. *et al.* (2021) Transcription-associated topoisomerase 2α (TOP2A) activity is a major effector of cytotoxicity induced by G-quadruplex ligands. *Elife* 10, e65184.

[90] Xu, H. and Hurley, L.H. (2022) A first-in-class clinical G-quadruplex-targeting drug. The bench-to-bedside translation of the fluoroquinolone QQ58 to CX-5461 (Pidnarulex). *Bioorg Med Chem Lett* 77, 129016.

[91] Platella, C. *et al.* (2021) Disentangling the Structure–Activity Relationships of Naphthalene Diimides as Anticancer G-Quadruplex-Targeting Drugs. *J Med Chem* 64, 3578–3603.

[92] Al Kobaisi, M. *et al.* (2016) Functional Naphthalene Diimides: Synthesis, Properties, and Applications. *Chem Rev* 116, 11685–11796.

[93] Hampel, S.M. *et al.* (2010) Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells. *Bioorg Med Chem Lett* 20, 6459–6463.

[94] Gunaratnam, M. *et al.* (2011) Targeting pancreatic cancer with a G-quadruplex ligand. *Bioorg Med Chem* 19, 7151–7157.

[95] Hampel, S.M. *et al.* (2013) Mechanism of the Antiproliferative Activity of Some Naphthalene Diimide G-Quadruplex Ligands. *Mol Pharmacol* 83, 470–480.

[96] 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.

[97] Ohnmacht, S.A. *et al.* (2015) A G-quadruplex-binding compound showing anti-tumour activity in an in vivo model for pancreatic cancer. *Sci Rep* 5, 11385.

[98] 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.

[99] Vo, T. *et al.* (2020) Substituted Naphthalenediimide Compounds Bind Selectively to Two Human Quadruplex Structures with Parallel Topology. *ACS Med Chem Lett* 11, 991–999.

[100] Ahmed, A.A. *et al.* (2020) Asymmetrically Substituted Quadruplex-Binding Naphthalene Diimide Showing Potent Activity in Pancreatic Cancer Models. *ACS Med Chem Lett* 11, 1634–1644.

[101] Platella, C. *et al.* (2021) DNA Binding Mode Analysis of a Core-Extended Naphthalene Diimide as a Conformation-Sensitive Fluorescent Probe of G-Quadruplex Structures. *Int J Mol Sci* 22, 10624.

[102] De Cian, A. *et al.* (2007) Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J Am Chem Soc* 129, 1856–1857.

[103] Bharti, S.K. *et al.* (2013) Specialization among Iron-Sulfur Cluster Helicases to Resolve G-quadruplex DNA Structures That Threaten Genomic Stability. *J Biol Chem* 288, 28217–28229.

[104] 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.

[105] Bharti, S.K. *et al.* (2013) Specialization among Iron-Sulfur Cluster Helicases to Resolve G-quadruplex DNA Structures That Threaten Genomic Stability. *J Biol Chem* 288, 28217–28229.

[106] Gray, L.T. *et al.* (2019) G-quadruplexes Sequester Free Heme in Living Cells. *Cell Chem Biol* 26, 1681-1691.e5.

[107] Local, A. *et al.* (2018) APTO-253 Stabilizes G-quadruplex DNA, Inhibits MYC Expression, and Induces DNA Damage in Acute Myeloid Leukemia Cells. *Mol Cancer Ther* 17, 1177–1186.

[108] Penke, L.R. *et al.* (2022) KLF4 is a therapeutically tractable brake on fibroblast activation that promotes resolution of pulmonary fibrosis. *JCI Insight* 7, e160688.

[109] Suseela, Y.V. *et al.* (2020) Recognition of G-quadruplex topology through hybrid binding with implications in cancer theranostics. *Theranostics* 10, 10394–10414.

[110] Ferret, L. *et al.* (2023) G-quadruplex ligands as potent regulators of lysosomes. *Autophagy* 19, 1901-1915.

2nd Chapter

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

This work 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-quadruplex in non-coding RNAs. Molecules 26, 6164. DOI: 10.3390/molecules26206164.*

Chapter overview

This chapter provides an overview of the use of different G4 ligands employed to bind and stabilize the G-quadruplex structures found in different classes of non-coding RNAs. Non-coding RNAs are a group of transcripts that do not translate into proteins but have important multiple biological regulatory functions. The prevalence and regulation in the stability of G-quadruplex structures within these RNA sequences, such as long non-coding RNAS, pre-miRNAs emerged as a potential therapeutic strategy. G4 ligands can regulate the RNA G4 folding equilibrium and prevent important cellular events such as miRNA biogenesis and miRNA-mRNA interactions. Taken together, this chapter summarizes and discusses the prevalence of RNAs G4s within non-coding RNAs, focusing on the molecular mechanism underlying its function and highlighting the use of G4-stabilizing ligands to modulate its biological functions.

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].

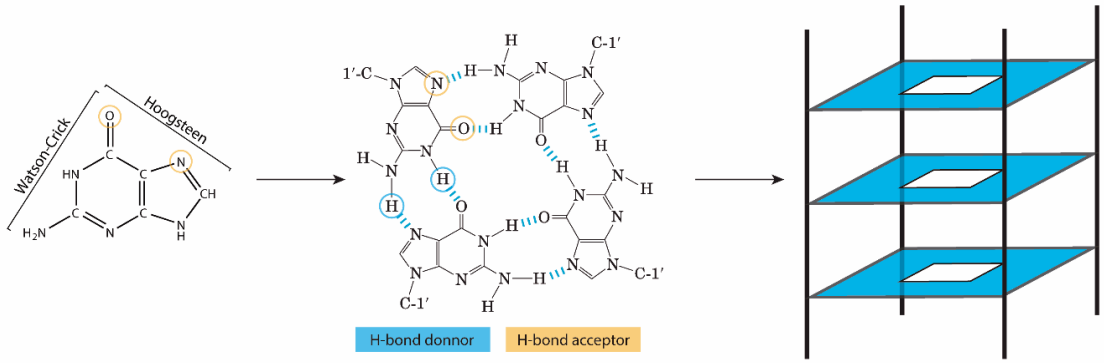


Figure 2.1. Illustration of G-tetrad formation through Hoogsteen bonds and G-tetrads stacking.

The stability of the negatively charged core of the G-quartet composed of O6 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.

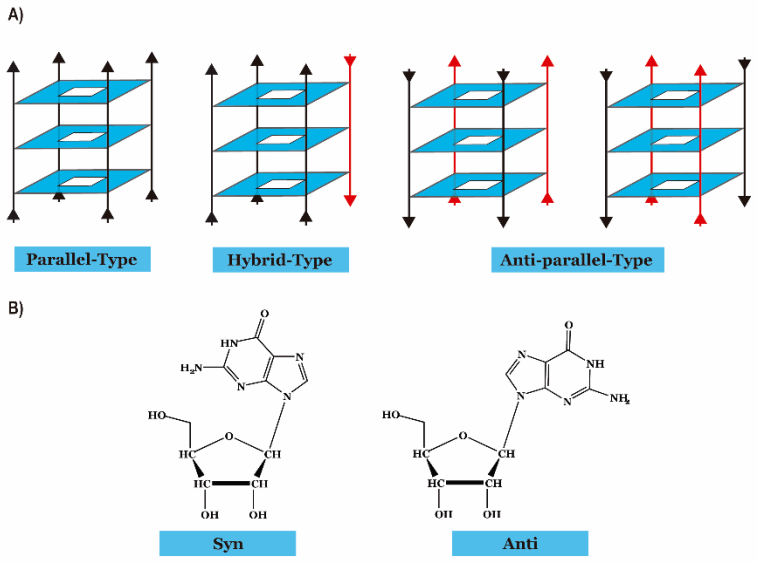


Figure 2.2. (A) Schematic representation of different G4 topologies (parallel, hybrid, and anti-parallel). **(B)** Conformations of G4s according to strands polarity when polarity differences are associated with an angle between the G-bases and the pentose, namely to the 2' hydroxyl group.

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

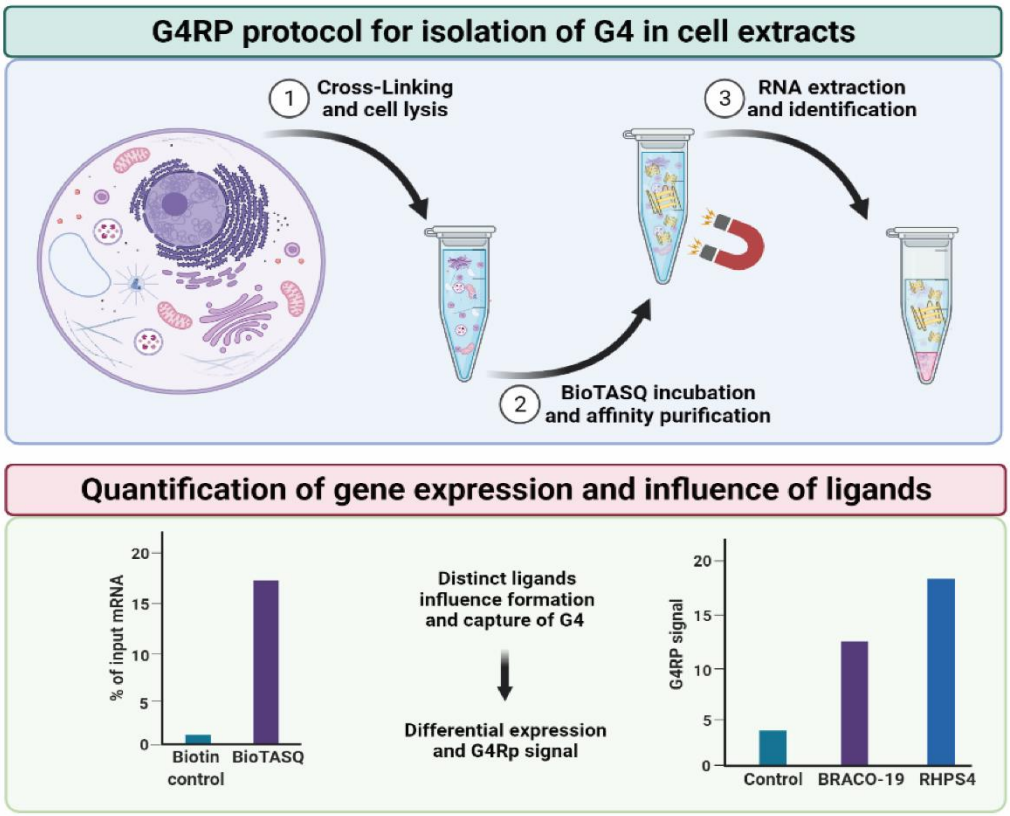


Figure 2.3. Illustration of G4RP method protocol for isolation of G4 targets from human cell extracts and the influence of the ligands in gene expression. Adapted from ref. [25]. Copyright (2018), with permission from Springer Nature.

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].

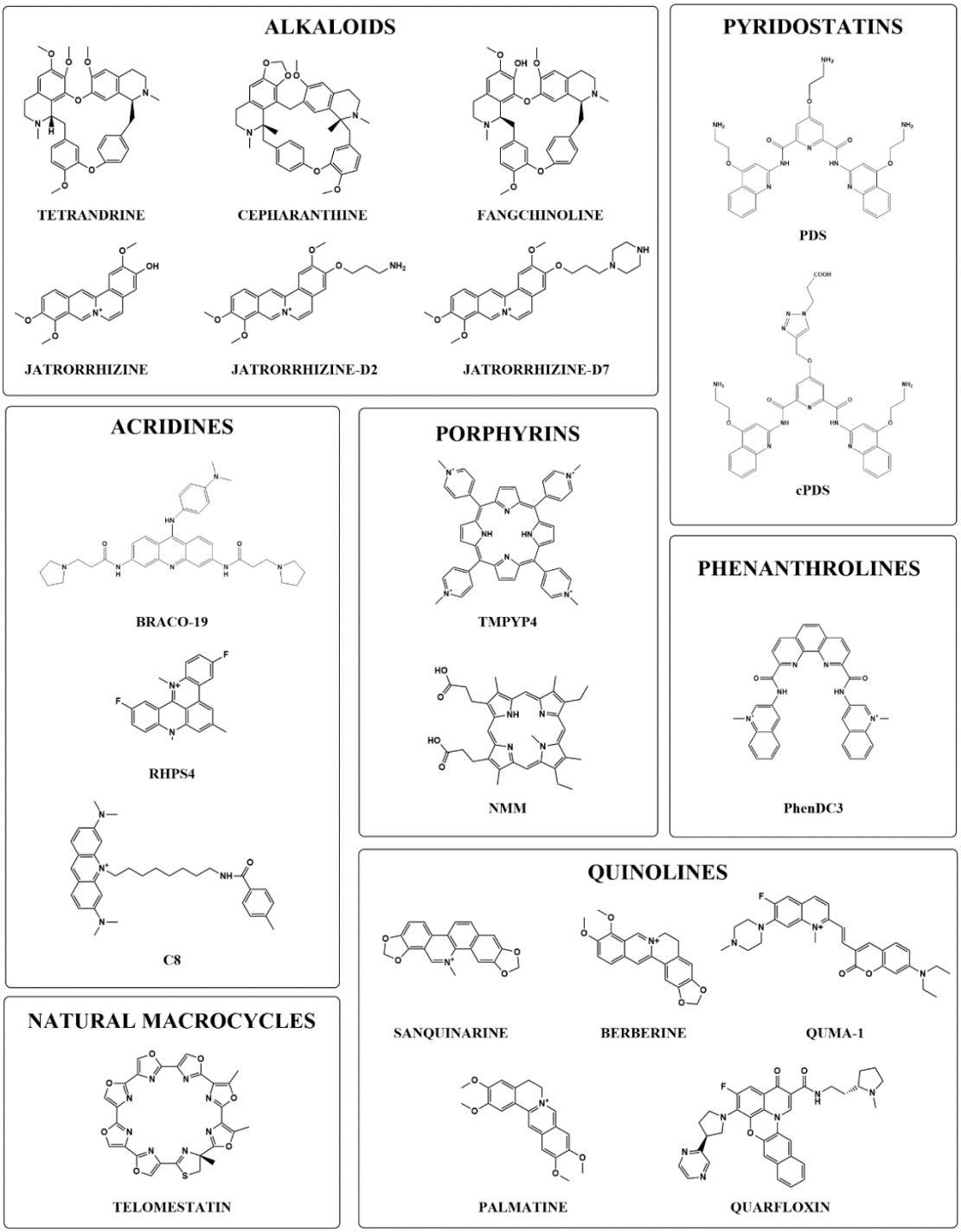


Figure 2.4. Chemical structures of the G4 ligands presented in this review. Small molecules are divided according to their chemical families. Structures were designed with ChemDraw 20.0 ®—PerkinElmer.

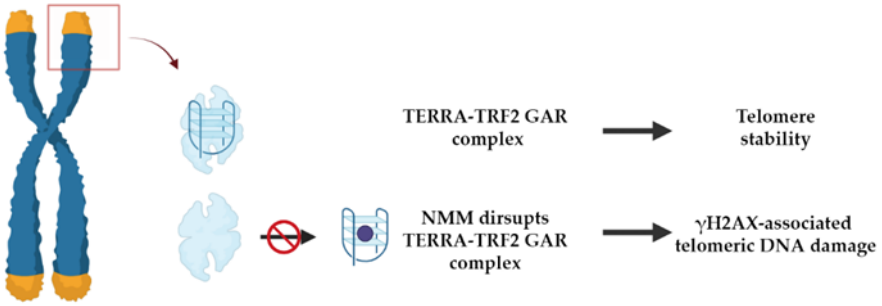


Figure 2.5. Schematic representation of TERRA rG4 interaction with TRF2 GAR domain that is required for telomere stability and integrity. Treatment of human melanoma cells with N-methyl mesoporphyrin IX (NMM) leads to disruption of TERRA and the induction of γH2AX associated telomeric DNA damage.

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.

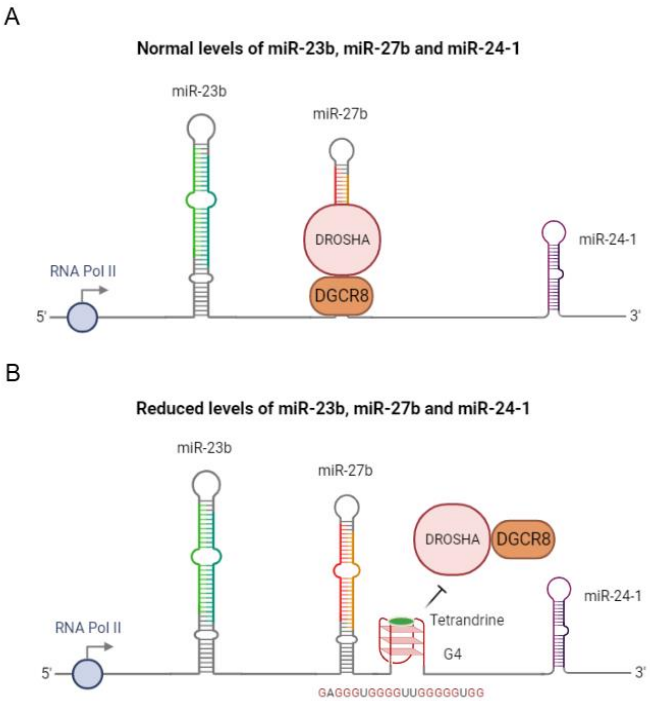


Figure 2.6. Representation of the influence of G-rich sequence in normal **(A)** and reduced **(B)** levels of miR-23b, miR 27b, and miR-24-1. Adapted from ref. [42]. Copyright (2021), with permission from Oxford University Press.

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 [45-48]. 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].

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**) [50]. C_8 is an acridine orange derivative [51] that has been reported as a promising G4 binder [52-55].

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].

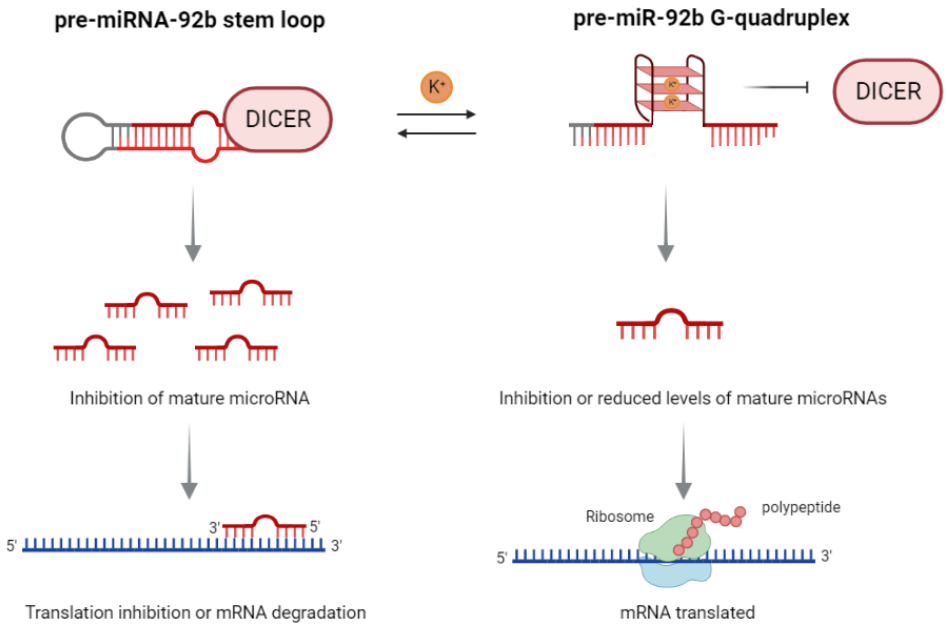


Figure 2.7. Representation of the ionic profiling to induce G4 formation by unwinding the stem–loop of pre-miR-92b Adapted from [44] copyright (2015), with permission from Elsevier.

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].

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 C8 derivative with more effectiveness than C_8-NH_2. These results were anticipated by molecular dynamics in an *in silico* study [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].

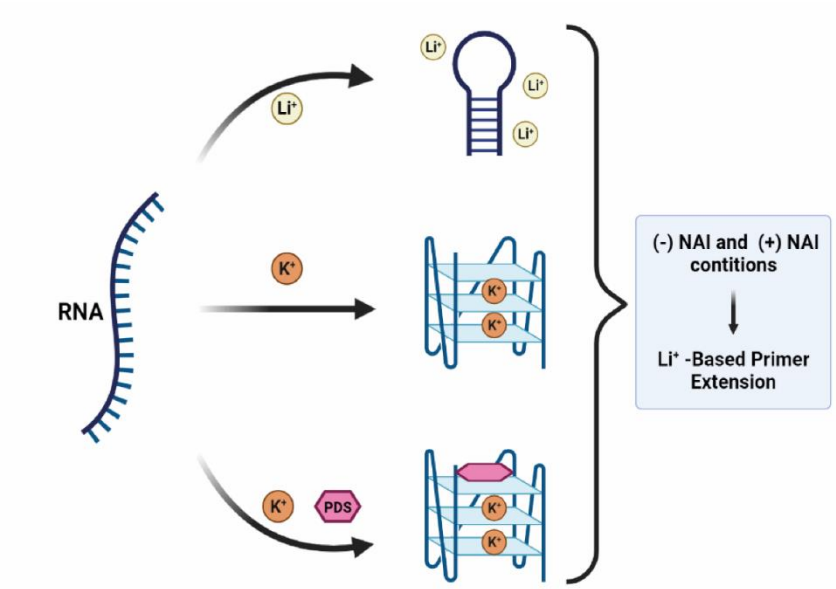


Figure 2.8. Representation of SHALiPE method [57]. Copyright (2016), with permission from Wiley.

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].

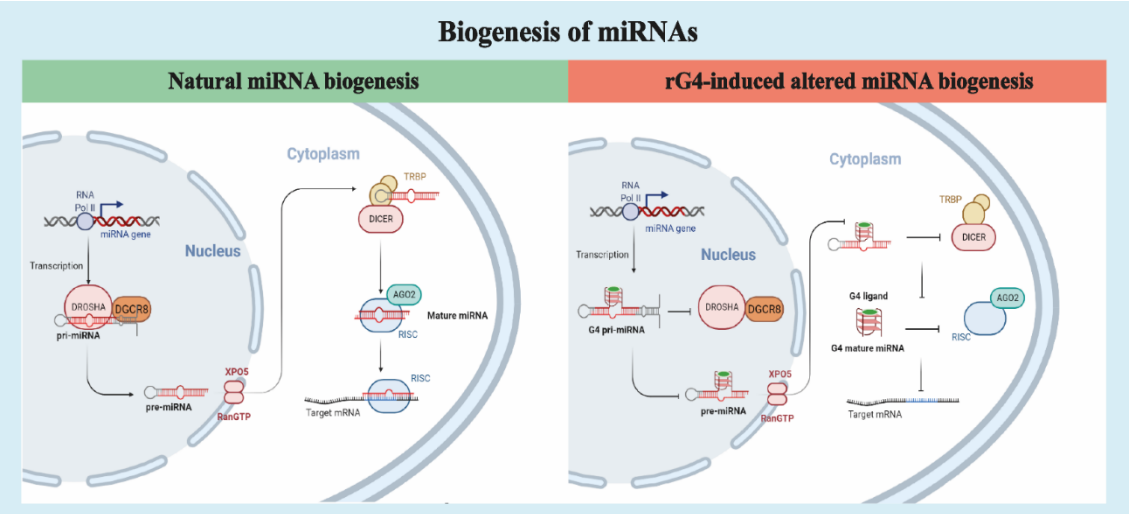


Figure 2.9. Biogenesis of miRNAs. Comparison of natural biogenesis versus interference of the rG4s in the biogenesis and function of miRNAs.

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].

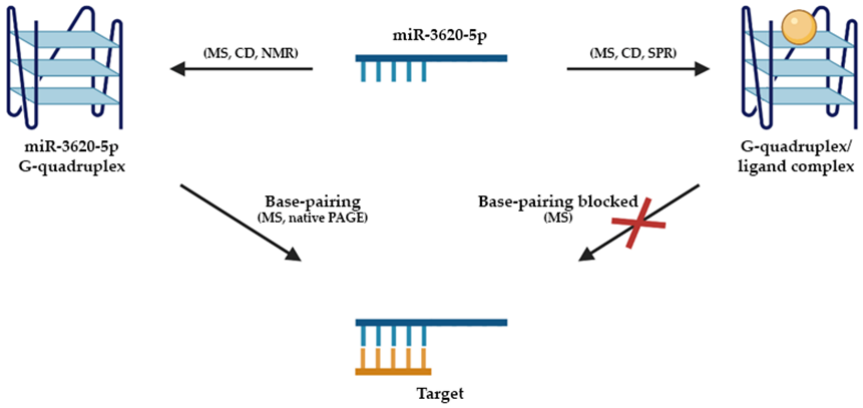


Figure 2.10. Representation of G4 formation in miR-3620-5p and the inhibition effect of sanguinarine on the base pair formation of miR-3620-5p with its target sequence. Adapted from ref. [73]. Copyright (2016), with permission from Elsevier.

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 chapter.

		rG4 Forming Sequence (5′-3′)	Ligands	Biological Target	Effect of G4 Formation	Refs.
Gene Cluster	mir-23b mir-27b mir-24-1	GAGGGTGGGGGGTTGGGGGTGG	Tetrandrine	Drosha-DGCR8	Decreased of pri-miR levels	[42]
	miR-K12–1-9	GGGTGGGAGGAAGGATGTGGGGGTGGG	$TmPyP_4$, PDS	promoter activity	Transcriptional regulation of virus encode miRNA.	[89]
Pri-miRNA	Pri-mir200c	GGUGGGCGGGCUGGGCGGGGG	cPDS, PhenDC3	Drosha-DGCR8	Reverse Transcription Stalling	[41]
	Pri-mir451a	GGGCACUUGGGAAUGGCAAGG				
	Pri-mir497	GGAGGGGGUGGG				
Pre-miRNA	pre-miR-92b	GGGCGGGCGGGAGGG	LNA, C_8	Dicer	Decrease mature levels of miRNA	[49,50]
	pre-miR-let-7e		$TmPyP_4$		Increase mature levels of miRNA	[56]
	pre-miR-149	GGGAGGGAGGGACGGGGG	PDS, C_8		Decrease mature levels of miRNA	[57-58]
	pre-miR-1229	GGGUAGGGUUUGGGGGAG AGCGUGGGCUGGGGUUCAGGG ACA		Dicer		[61]
	pre-miR-26a-1	GGAUAGGCUGUGCAGGUCCCAAUGGG	PDS, cPDS		Decrease mature levels of miRNA	[66]
	pre-miR-150	GGCCUGGGGGACAGGGACCUGGG	PhenDC3			[67]
miRNA	miR-3620-5p	GUGGGCUGGGCUGGGCUGGGCC	Sanguinarine	Blocked the base pairing with target sequence	Depression of miRNAs	[73]
	miRNA-1587	UUGGGCUGGGCUGGGUUGGG	Jatrorrhizine derivatives, pseudopalmatine			[80,81]
	miRNA-149	AGGGAGGGACGGGGGCUGUGC	NMM			[75]
	miRNA-197	CGGGUAGAGAGGGCAGUGGGAGG				
	miRNA-432	UCUUGGAGUAGGUCAUUGGGUGG				
	miRNA-765	UGGAGGAGAAGGAAGGUGAUG				
	miR-92a	TAATGGGGTGGGGGCTGGGAA	Palmatine			[85]

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/00709/202

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 Physio* 9, 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. *Nutritional Epigenomics*, Academic Press: Salt Lake City, UT, USA, 119-132.

[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 Rep* 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 Chem 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 ed 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] Agarwala, P. *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* 13, 3156.

[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 Genom Bioinform* 2, 1-14.

[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] Arachchilage, G.M. *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] Carvalho, J. *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, 83-92.

[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.

3rd Chapter

Harnessing G-quadruplex ligands for lung cancer treatment: A comprehensive overview

This chapter was published in:

***Figueiredo, J.**; Djavaheri-Mergny, M.; Ferret, L.; Mergny, J-L. and Curz, C. (2023) Harnessing G-quadruplex ligands for lung cancer treatment: A comprehensive overview. Drug Discovery Today 28, 103808. DOI: 10.1016/j.drudis.2023.103808.*

Chapter overview

This chapter presents a comprehensive review of advancements in G4 ligands as a therapeutic strategy for LC. Despite an extensive study of G4 ligands as anti-tumor agents in literature, to the best of our knowledge, there has been no comprehensive summary of their use as G4 modulators for lung cancer therapy. The present chapter furnishes an up-to-date overview of the significance of G4 as a valuable therapeutic target in LC treatment and outlines the G4 ligands that target the genes involved in its onset and progression. The synergistic effects between G4 ligands and other LC treatments were also explored. Additionally, a brief discussion about the use of G4 as a therapeutic agent in LC and some considerations regarding the rational design of G4 ligands are provided.

Harnessing G-quadruplex ligands for lung cancer treatment: A comprehensive overview

Abstract

Lung cancer (LC) remains a leading cause of mortality worldwide, and new therapeutic strategies are urgently needed. One such approach revolves around the utilization of four-stranded nucleic acid secondary structures, known as G-quadruplexes (G4), which are formed by G-rich sequences. Ligands that bind selectively to G4 structures present a promising strategy for regulating crucial cellular processes involved in the progression of LC, rendering them potent agents for lung cancer treatment. In this review, we offer a summary of recent advancements in the development of G4 ligands capable of targeting specific genes associated with the development and progression of lung cancer.

Keywords

lung cancer; G-quadruplex structures; G-quadruplex ligands; anticancer activity; drug discovery; synergistic effects

Introduction

Lung Cancer (LC) is the leading cause of cancer-related death, with an estimated 1.8 million deaths in 2020 [1]. Clinically, LC is classified into two main subtypes, small cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC), accounting for ~15% and ~85% of all LC cases, respectively. They differ from each other depending on the origin, histological characteristics, genetic mutation profile, level of aggressiveness of the tumor and the effectiveness of responses to anti-tumor therapies. In the USA, NSCLC claims more lives each year than all other major cancers combined. It is further divided into histological subtypes: adenocarcinoma (ADC), squamous cell carcinoma (SCC), large cell carcinoma (LCC), and NSCLC-not otherwise specified (NOS), which includes NSCLCs that do not have the characteristics of the other types [2]. Tobacco smoke remains the major risk factor for LC development; however, other factors such as air pollution,

environmental exposure, mutations, and singlenucleotide polymorphisms, are known to be associated with LC [3]. Current LC therapeutic options (surgery, chemotherapy, radiotherapy, or immunotherapy) are chosen based on the cancer stage, and histologic and genetic characteristics of the tumor [4]. These treatments can be used alone or in combination [5] and an immunotherapy-based regimen alone or in combination with chemotherapy is often the preferred option for patients for whom the status of tumors can predict the benefit derived from immune checkpoint inhibitors [6]. Although surgery and chemoradiotherapy are effective during the early stage of the disease, they show limited efficacy in advanced LC, when most patients are diagnosed.

In this field, the emergence of precision medicine has greatly enhanced treatment outcomes [7] Genetic variations associated with LC, more specifically in NSCLC, involve mutations in proto-oncogenes, such as the members of epidermal growth factor receptor (EGFR), MYC, and RAS families as well as PIK3CA, NKX2-1, ALK, RET MET, LKB1, BRAF, and ROS1 [8] To target specific oncogenic abnormalities, the US Food and Drug Administration (FDA) has approved numerous drugs, including gefitinib, crizotinib, entrectinib, selpercatinib, tepotinib, and villafinil. Additionally, immune checkpoint blockers, such as antibodies against cytotoxic T lymphocyte antigen-4 (CTLA-4) and monoclonal antibodies (mAbs) that target programmed death ligand1 (PD-L1) or programmed death (PD-1), have produced long lasting responses [9]. FDA-approved immune checkpoint inhibitors, such as ipilimumab, durvalumab, and nivolumab, are also available [10-12]. Despite these advances, drug resistance remains a challenging issue for LC patients [13]. A better understanding of the molecular profile of tumors and identification of target gene alterations could lead to the development of novel targets and new therapeutic strategies.

Over the past few decades, G4 structures have emerged as attractive targets for cancer therapy [14] The evidence that G4 containing-genes are involved in hallmarks of cancer (sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, and modulating anti-tumor immunity [15] and the immune response [16,17]) makes these structures attractive targets for therapeutics, in particular in LC [18], because many of the oncogenes involved in LC are likely to be regulated by G4s. In this study, we provide examples of LC related G4s and an overview of the use of G4 ligands with *in vitro* or *in vivo* activity against these targets (**Table 3.1**). Several LC-related G4s are also found in a variety of cancers, making it challenging to predict their selectivity. The results achieved in this area, which could assist researchers in developing more

potent compounds against LC, can be applied either as standalone treatments or in combination with other anti-cancer agents.

G-quadruplex as a therapeutic target

G4 structures are non-canonical nucleic acid secondary structures formed by the stacking of two or more square arrangements of guanines (G-quartets) held together by Hoogsteen hydrogen bonds and formed in the presence of central metal cations (typically K^+ or Na^+) [19,20]. G4s can be intramolecular or intermolecular and can adopt different topologies, including parallel, antiparallel, or hybrid, arising from strand orientation, and variations in loop size, sequence, and length [21].

Computational approaches have identified over 700 000 putative G4-forming sites in the human genome [22]. These structures are commonly present in specific regions of the genome, such as DNA replication origins, chromosome ends (telomeres), promoters, and gene transcriptional regulatory regions. This nonrandom distribution determines their roles in several biological processes, including DNA replication, telomere maintenance, transcription, homologous recombination (HR), and epigenetics [23]. G4s are associated with regulating the expression of several oncogenes and evidence demonstrates that alterations in G4 structure formation can modulate gene expression of, for example, MYC [24], Kit transmembrane tyrosine kinase receptor (KIT) [25], B cell lymphoma-2 (BCL-2) [26], KRAS [27], and vascular endothelial growth factor (VEGF) [28].

Over 3000 small molecules targeting G4 structure are currently described in the G4 ligands database (https://www. g4ldb.com/) and new chemotypes are frequently reported [29].

Several molecules that exhibit G4-binding properties reached clinical trials in patients with cancer. Well-known examples include CX-3543 (Quarfloxin), which reached Phase II clinical trials against carcinoid/neuroendocrine tumors and was shown to inhibit c-MYC expression, and CX-5461, described as an inhibitor of RNA polymerase I transcription affecting ribosomal RNA genes (rRNA) in particular [30,31] Another interesting compound is APTO-253, a small molecule that inhibits c-MYC expression, stabilizes G4 DNA, and induces cell cycle arrest and apoptosis in acute myeloid leukemia cells [32]. Interestingly, treatment with these ligands leads to DNA damage. For instance,

APTO-253 treatment induced early DNA damage assessed by induction of cH2AX accumulation and cellular stress pathways [32]. Treatment with CX-3543 and CX-5461 also results in increased levels of DNA damage, which, when combined with ionizing radiation and hydroxyurea, enhances their cytotoxic effect [33] Similar effects were observed for a well-known G4 ligand, pyridostatin (PDS), which induces DNA damage as shown by the formation of cH2AX foci [34]. Evidence suggested that DNA damage accumulation induced by CX-3543, CX-5461, and other G4 ligands is more effective in BRCA1/2-deficient cancer cells [34,35]. Capranico and colleagues [36] discovered that G4 ligands are capable of inducing DNA damage through a mechanism dependent on R-loops. Altogether, these observations have important implications for the discovery of new G4-based anticancer agents.

In addition, other FDA-approved compounds were later shown to recognize G4 structures. Among them, prochlorperazine, promazine, and chlorpromazine, all FDA-approved phenothiazine antipsychotic drugs, were shown to bind to G4s in vitro [37]. Indenoisoquinoline derivatives are human topoisomerase I inhibitors in clinical testing that bind to the c-MYC G4 [38]. Despite the clinical importance of these molecules, it remains to be determined whether their clinical effects are mediated, at least in part, by G4 stabilization. The rational design for new G4-interact small molecules comprises several criteria. These ligands exhibit variations in their specificity, binding mode, and cell permeability, but are characterized as molecules that can bind and stabilize G4 structures with high affinity and selectivity over duplex nucleic acid sequences to avoid off-target toxicity.

G-quadruplex as therapeutic agents in lung cancer

Beyond the potential as a therapeutic target, the G4 structures formed by G-rich oligonucleotides (GRO) can act as therapeutic tools with clinical applicability. In a recent report Carvalho *et al.* classified G4 nucleic acids as "*friends, as healing anticancer drugs"* and "*foes, within the harmful cancer cell"* [39]. G4-forming oligonucleotide sequences aimed at cancer-related targets have been reported as therapeutic agents. In this field, aptamers containing G-rich sequences present additional advantages over unstructured sequences, such as chemical and thermodynamic stability, non-immunogenicity, serum nuclease resistance, and enhance cellular uptake. Moreover, the structural stability of G4-motif and its negatively charged density enhanced electrostatic interactions with positively charged binding targets [40].

One of the most studied aptamers is AS1411, which went into clinical trials and showed promising activity in acute myeloid leukemia and metastatic renal cell carcinoma [41]. AS1411 is a 26-mer G-rich DNA sequence targeting nucleolin that tends to form a polymorphic G4 structure. AS144 inhibits *in vitro* A549 cell proliferation (86%) after 5-day treatment at 10 μM concentration [42]. AS1411 administration in human xenografts derived from A549 NSCLC cells resulted in delayed in tumor growth [43]. Moreover, the DTNs-AS1411 delivery system induced A549 cell apoptosis regulating BCL-2/Bax ration.[44] *in vivo* administration showed no acute biotoxicity during the first day post-injection which can lead to a novel potential strategy for LC therapy [44].

The aptamer T40214 and its analog T40231 are also able to form an intramolecular G4 structure [45]. These aptamers strongly suppressed the growth of NSCLC tumors *in vivo* blocking p-STAT 3, and downstream the proteins BCL-2, BCL-x_L, MCl-1, surviving, VEGF, Cyclin D1, and c-MYC [46].

Guidelines for G4 ligands rational design

G4 are dynamic and highly polymorphic structures and drug design for G4-interact molecules should consider these structural features and should be guided by a rational process (**Figure 3.1**). Most G4 ligands have an aromatic core for π-π stacking with G-quartet, a positively charged or basic group to interact with loops or grooves, and a steric bulk to avoid intercalation with duplex DNA [47]. However, some reported ligands do not obey these rules and show an interesting interaction profile. In this field, the resolution of G4 structures by high-resolution methods, such as NMR or X-ray crystallography which provides structural information about the G4 behavior can help the researchers to predict the G4-ligand binding mode against a specific G4 target [48]. This prediction can be made using molecular dynamic simulations, guiding efforts in the synthetic route or helping in the discovery of new chemotypes that can occur naturally. The initial *in vitro* screening against G4 targets is also fundamental to characterize ligand-G4 interaction and validate their specificity and selectivity. In this field, we strongly encourage the use of *in vitro* methods (well characterized in a recent overview [49]) to discover potential hits and compare their activity with well-known G4 ligands. The comparison of biophysical data on ligands-G4 interactions would be potentially interesting for SAR analysis. However, this analysis requires careful attention, and a fair comparison is not always possible. Usually, biophysical studies are collected using different experimental methods, on different quadruplex targets using different

experimental conditions which hamper a realistic comparison of these effects. Later analyses would require standardized protocols, which could be discussed in future "Guidelines" as the ones that have been proposed to study G4 *in vitro* [49]. In any case, these *in vitro* activities are often poorly correlated with *in vivo* efficacies. These discrepancies can occur due to the differences between the *in vitro* and *in vivo* behavior of G4 and the lack of "drug-like" properties commonly reported in G4 ligands. In general, incorporating chemical groups or implementing structural modifications to improve G4 interactions often leads to a reduction in drug-like properties. Therefore, finding a balance between a molecule's potency as a G4 ligand and its suitability as a drug-like compound requires careful consideration. Some informatics tools can help by predicting some properties such as solubility, lipophilicity, and pharmacokinetics. A computational tool for identifying multi-target-directed anticancer lead molecules against G4 DNA called G4-QuadScreen was recently reported [50], which predicts G4 selectivity, interaction, stabilization, and cytotoxicity in multiple experimental conditions and biological responses against different G4 targets. The validation parameters are confirmed by FRET melting, FID, and cell viability assays and are indicative of the good discriminatory for the four models, being a promising tool to accelerate G4 ligands design.

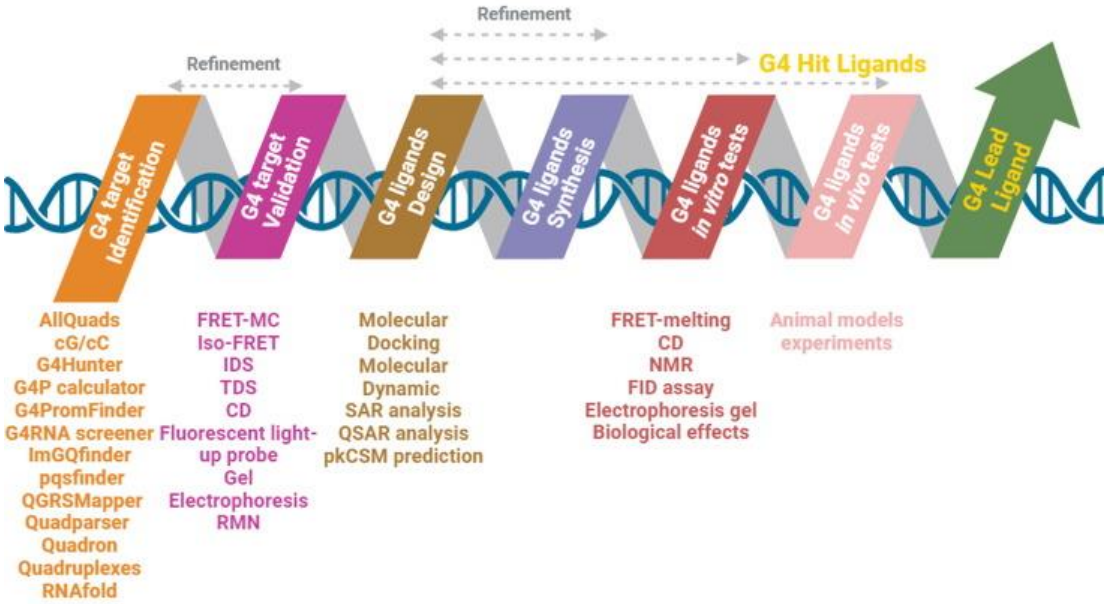


Figure 3.1. Steps of rational G-quadruplex (G4) ligand design, from G4 target identification to the selection of a potential G4 lead ligand. Abbreviations: CD, circular dichroism; FID, fluorescent intercalator displacement; FRET-MC, fluorescence resonance energy transfer-melting competition; IDS, isothermal difference spectra; (Q)SAR, (quantitative) structure–activity relationship; NMR, nuclear magnetic resonance; TDS, thermal difference spectra.

Cancer-related genes in lung cancer and their interaction with small molecules

c-MYC

In LC, c-MYC transcription levels are frequently dysregulated and contribute to malignancy by regulating cancer cell growth and survival, by activating cycle-promoting proteins, increasing antiapoptotic protein levels, and regulating cellular metabolism [51]. High-level amplification in MYC has been identified in primary SCLC samples [52]. MYC dramatically accelerates tumorigenesis and metastases in Rb1/Trp53-null SCLC.41 Given that the c-MYC protein has traditionally been deemed undruggable, finding approaches to inhibit c-MYC transcription could serve as significant tools in the fight against LC progression [54].

The c-MYC promotor region contains a nuclease hypersensitivity element (NHE III1) that controls 90% of total c-MYC transcription levels. The NHE III1 element, located upstream of the P1 promoter of the c-MYC gene, contains a G-rich region that can adopt intramolecular G4 conformations (**Figure 3.2**) [55]. These G4 structures can act as c-MYC transcriptional repressor elements and G4-interact molecules have been shown to suppress c-MYC transcription and expression [56].

Piperine, a natural alkaloid, presented a high affinity for c-myc G4 (K_{D1} = 2.5 nM; K_{D2} = 1.1 mM; determined by fluorescence microscopy) compared with tel22, c-kit21 G4 sequences, and double-strand DNA [57]. It showed a cytotoxic effect on various cancer cell lines (HeLa, PC3, HepG2, and MCF-7 cell lines) and its exposure to A549 cells resulted in apoptosis-mediated cell death [57]. **PIP-2**, a piperine analog, also showed significant selectivity (over bcl-2, tel22, c-kit21 G4, and duplex sequences) and affinity for c-myc G4 (K_D = 2 nM; determined by fluorescence microscopy), downregulating the transcription and translation of c-MYC. PIP-2 exhibits the lowest IC_{50} for A549 cancer cells (IC_{50} = 6 mM) over other cancer cells (HeLa, DU, A431, and MCF-7), inducing cell apoptosis through its accumulation in the nucleus [58]. The benzoxaziones **1–3** display promising in vitro anticancer properties against cancer cells (SK-RC-42 and SGC7901). These compounds inhibit cell migration and display notable cytotoxicity (IC_{50} = 4.1, 7.3, and 5.4 μM, respectively) in A549 cells, with good selectivity compared with normal cells [59]. EMSA assays demonstrated that these compounds could bind the G4 in the c-myc gene promoter, downregulating c-MYC expression in a dose-dependent manner [60].

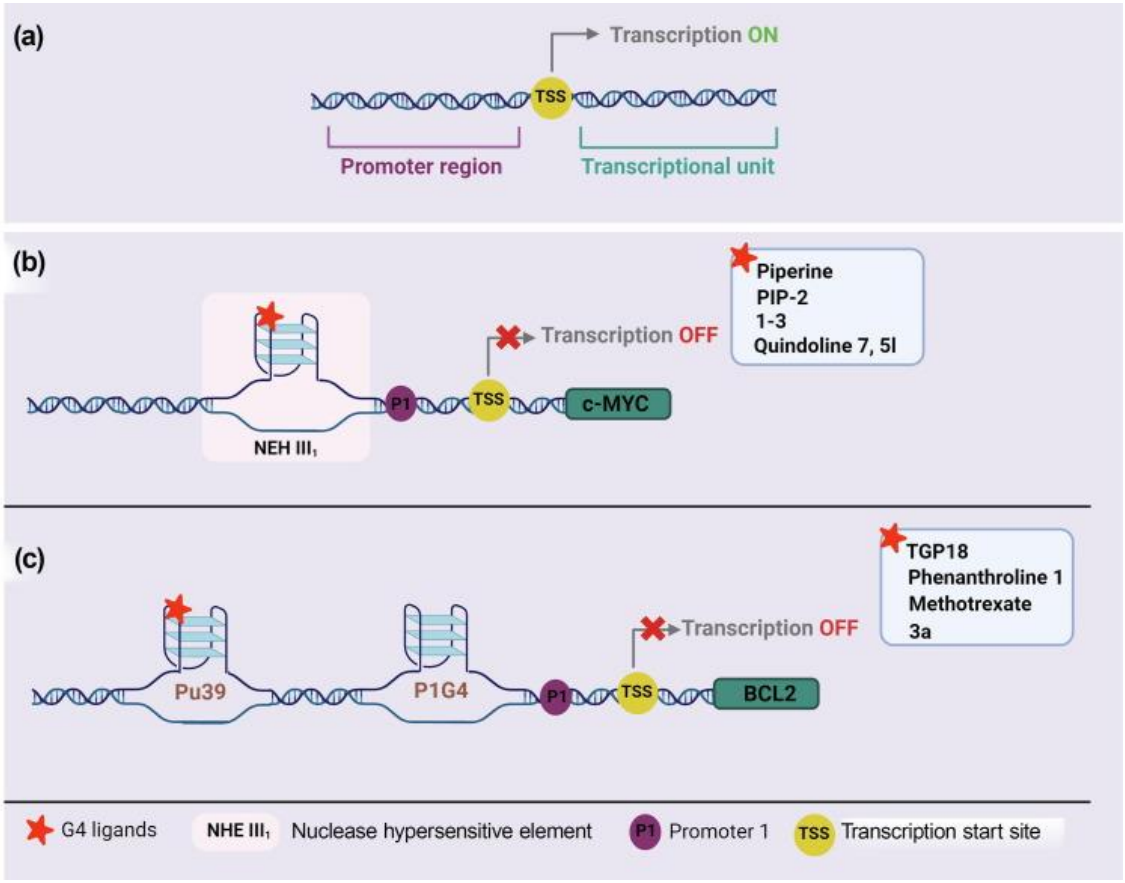


Figure 3.2. Formation of G-quadruplex (G4) structures in the promoter region affecting gene transcription. **(a)** Gene transcription process in the absence of any G4 structures. **(b, c)** G4 ligands can enhance G4 stability, and G4 formation proximal to the transcription start site (TSS) could inhibit the transcription process of **(b)** c-myc and **(c)** BCL-2.

Quindoline has been characterized as an appropriate scaffold for the development of G4 ligands. In this field, quindolines containing different terminal groups on their side chains proved to be good c-myc G4 stabilizers [60]. **Quindoline 7** increased the melting temperature (T_m) value of the FPu22T sequence [ΔT_m = +16 C; determined by fluorescence resonance energy transfer (FRET) melting] with a K_D value of 4.6 μM (determined by SPR) showing similar affinity to duplex DNA [60,61]. A series of novel derivatives containing a 1,4-triazole moiety were synthesized to improve their selectivity and stabilizer properties [60,61]. Among the synthesized ligands, derivative 5l exhibited higher binding with c-myc G4, with a K_D value of 2.7 μM (determined by SPR) and weaker binding with other kinds of DNA structure. This compound was also more potent against A549 cells (IC_{50} = 0.3 μM) compared with **7** (IC_{50} = 2.1 μM), through the downregulation of c-myc transcription. **Quindoline 5l** inhibited tumor growth in a human lung cancer xenograft model (A549). Treatment with **5l** (12 days at a dose of 6.3 mg/kg per day) or **7** (7.5 mg/kg) led to *in vivo* antitumor effects (inhibition rates of 38.1 and 29.2%, respectively) [61].

BCL-2

BCL2 encodes a mitochondrial membrane protein that protects cells against apoptosis [62]. The BCL-2 protein family members can be classified as proapoptotic (BH123 and BH3-only protein subfamilies) and antiapoptotic (BCL-2, BCL-XL, MCL1, BCL-W, BCL-B, and BCL-2A1) [63]. Elevated BCL-2 levels were reported in several cancers, and increased BCL-2 expression has been found in NSCLC [64]. This overexpression promotes tumor growth and contributes to chemotherapeutic resistance in xenograft models of NSCLC [65,66]. Several strategies aimed at inhibiting BCL-2 have been developed, including antisense nucleotides, peptides, and small molecules to disrupt protein-protein interactions [67]. BCL2 amplification and translation are the most common mechanisms that cause its overexpression [68]. The region upstream of the P1 promoter (−59 to −19) of BCL-2 contains a GC-rich sequence that forms a stable G4, called Pu39 (Figure 2) [69]. In K^+ solution Pu39 can adopt three distinct intramolecular G4 structures [70]. Immediately upstream of the BCL-2 P1 promoter, there is another 28- mer G4-forming region, called P1G4, which can act as a transcriptional repressor, modulating BCL-2 transcription (**Figure 3.2**) [71]. Thus, a wide range of small molecules have been synthesized as BCL-2-targeting G4 ligands [72].

TGP18 is a 1,10-phenanthroline derivative that interacts specifically with BCL-2 G4 with high affinity (K_D = 728 nM determined by fluorescence titrations) [73]. This derivative had an antiproliferative effect on MDA-MB-231 and A549 cells (IC_{50} = 4 and 6 μM, respectively) downregulating BCL-2 transcription and promoting cell cycle arrest and inducing apoptosis in A549 cells. The compound also demonstrated in vivo antitumor activity in xenograft models (A549) showing a dosedependent antitumor effect with the minimal dose (0.5 mg/kg) producing significant tumor growth inhibition compared with gemcitabine (100 mg/kg) [73]. Another 1,10-phenanthroline derivative (**1**), containing an imidazole ring stabilizes BCL-2 G4 (K_a = 0.9 μM^{-1}; determined by electronic absorption) and displays in vitro antitumor activity against cancer cells (SMMC7721, SW620, and A549 cell lines), with a more potent antiproliferative effect in A549 cells (IC_{50} ~15 μM) [74]. This derivative has the potential to trigger apoptosis through a mitochondria-mediated pathway by binding to the BCL-2 G4 [74]. **Methotrexate** is an FDA-approved anti-cancer drug that inhibits dihydrofolate reductase activity. Interestingly, a recent study reported that methotrexate shows higher affinity and selectivity toward BCL-2 G4 (K_D = 9 nM; determined by steady-state fluorescence titration) over other G4s and duplex DNA [75]. **Methotrexate** was cytotoxic in different cancer cell lines (A549, HeLa, PC3, and HepG2) with maximum toxicity for A549 cells (IC50 = 5 μM). It downregulated the transcript level of BCL2 and inhibited the

expression of BCL-2 in a dose-dependent manner [75]. A naphthalimide-benzotriazole **3a** displayed antiproliferative activity against cancer cells (A549, HT-29, PC-3, HL-60, MDAMB-231, and HepG2) and proved to be more effective against A549 cells (IC_{50} = 6.7 μM). **3a** interacts strongly and selectively with the G4 motif within the BCL-2 promoter region, downregulating the expression levels of BCL-2, promoting apoptosis and DNA damage, and inducing autophagy in A549 cells [76].

Telomere and human telomerase reverse transcriptase

Human telomeric DNA comprises tandem repeats of the TTAGGG hexanucleotide located at the end of the chromosomes. Telomeres are essential to maintain chromosomal stability and cell proliferation. Telomerase activity plays a pivotal role in preserving telomere length homeostasis and can serve as a facilitator of tumorigenesis [77], because increased levels of human telomerase reverse transcriptase (hTERT) are found in ~90% of human cancers and are associated with unfavorable outcomes in solid tumors, including LC [77,78].

Significant research has focused on the development of G4-interact molecules as modulators of telomerase activity. This activity can occur in two ways, by inducing G4 formation at the 3´-single-stranded overhang of telomeres or by directly targeting the G-rich region of the hTERT promoter [79].

Historically, the first strategy was to indirectly target telomerase activity by locking its telomeric substrate into a 'locked' inactive conformation [80, 81]. The initial idea was that, provided telomerase is inhibited over a long period, the telomeres would progressively shorten and ultimately reach a critical length, leading to proliferation arrest or cell death. Unfortunately, such inhibition would only be efficient after multiple divisions, making this strategy viable only for tumors with very short telomeres. Nevertheless, this inhibition might also have effects not mediated by progressive telomerase shortening. Two 6- fluoroisoquinoline derivatives, **IQ1** and **IQ2**, showed telomerase inhibition activity by a selective interaction with human telomeric G4 (hTELO) [82]. Both derivatives promoted an antiproliferative effect in cancer cell lines (A549, HeLa, and Jurkat E6.1) after 72 h, with more activity against A549 cells, with **IQ1** being more efficient than **IQ2** (IC_{50} = 1.7 and 13 μM, respectively). The TRAP assay revealed that **IQ1** is also more potent than **IQ2** in inhibiting telomerase activity. Both derivatives arrest cancer cell growth at the G0/G1 phase, initiate mitochondrial depolarization, and activate caspases 3/7 to trigger cell apoptosis [82]. Compounds **12a-c**, of the series of 2-(1*H*-pyrazol-4-yl)-1*H*-imidazo [4,5-f] [1,10]phenanthrolines

derivatives, showed stabilization of the telomeric G4 with ΔT_m values between +12 and +21 °C (determined by FRET melting) [83]. Derivative **12c** showed an antiproliferative effect in A549 cells (IC_{50} = 1.5 μM) and strongly inhibited cell migration [83]. Based on that, the same group synthesized a novel series of 4,5-diazafluorene derivatives [84]. Between the derivatives, compound **14c**, bearing a phenylpyrazole group, had the best antiproliferative activity against cancer cell lines (A549, AGS, HepG2, and MCF-7), especially A549 cells (IC_{50} = 1.1 μM), against which it showed a strong migration inhibitory activity. This activity can be related to the interaction with 26- nucleotide telomeric G4 (K_a = 0.7 μM^{-1}; determined by SPR) and inhibition of telomerase activity [84].

Three derivatives of 1,4,5,8-naphthalenetetracarboxylic dimide (NDI) were synthesized and evaluated against two telomeric sequences [85]. Among them, **BBZ-ARO**, containing bisbenzimidazole side chains was the most potent derivative in terms of affinity toward *h-telo* 22-mer and *h-telo* 12-mer quadruplexes (K_a = 8.5 μM^{-1} and 6.6 μM^{-1}; determined by ITC). **BBZ-ARO** displayed telomerase inhibition of 50% at 4.6 μM and a toxic effect in A549 and HeLa cells, with preference for A549 cells (IC_{50} = 19.7 μM) after 72 h of exposure [85]. Four naphthylamide– thiourea conjugates (3a–d) exhibited significant selectivity for telomeric G4 over duplexes with high affinity for hTELO G4 (K_a > 0.1 μM^{-1}; determined by fluorescence titrations) [86]. These derivatives display preferential localization in the nuclei, inhibiting the growth of A549 cells (IC_{50} in the 3–4 μM range) [86]. A series of 1,10-phenanthroline disubstituted at positions 2,9 showed inhibitory activity on several cancer cell lines, including in a panel of lung cancer cells [87]. The authors hypothesized that this activity is related to the ability of the compounds to stabilize telomeric G4 and possible inhibition of telomerase activity. Among these derivatives, compound **6b** with two methyl residues on pyridine-3yl induced a higher stabilizing effect against the telomeric G4 sequence (ΔT_m = 24 °C; determined by CD melting) [87]. It also displayed anticancer activity on various cancer cells with more efficacy against A549, HOP-62, NCI-H226, NCIH23, NCI-H322M, NCI-H460, and NCI-H522 LC cell lines (GI_{50} between 0.2 and 2.3 μM). The triarylpyridine **20A**, reported as a telomeric G4 binder [88], promotes senescence and apoptotic cell death in A549 lung cancer cells [89,90]. **20A** treatment promoted global DNA damage but no specific telomeric damage [89]. Moreover, in combination with chloroquine (an inhibitor of lysosome functions), **20A** collaboratively induced lysosomal membrane permeabilization and led to human lung A549 cancer cell death [91].

hTERT, encoding for the telomerase catalytic subunit, has been the focus of several studies with G4 ligands. The hTERT core promoter contains a G-rich sequence that

includes three Sp1 binding sites and folds into a compact stacked three-G4 conformation [92]. The effect of a platinum-derived G4-ligand, **Pt-ctpy**, in association with radiation, was studied on NSCLC (A549 and H1299), and glioblastoma cells [93]. **Pt-ctpy** treatment modified the expression of the telomere-related genes *hTERT* and *TRF1*.

Other target genes (c-KIT, KRAS, VEGF, EGFR, HIFα, and WNT)

Additional G4s have been identified in other potential regulatory regions within the promoters of cancer-related genes (**Figure 3.3**). These G4 are also considered potential targets in LC and can be targetable by specific G4 ligands.

c-KIT, encoded by the c-kit proto-oncogene, is overexpressed in SCLC cells and nude mouse model xenografts [94]. G4s regions were identified in an NHE region of the c-KIT promoter and found to be involved in the regulation of c-KIT [95].

K-RAS is a frequently mutated oncogene prevalent not only in pancreatic cancer, but also in lung carcinomas (25% of NSCLC) [96]. Several mutations in the coding sequence of K-RAS result in amino acid substitution and yield a protein with altered functionalities: mutant K-RAS is locked in the activated state, transmitting continuous signals for cell proliferation to the nucleus. Somatic activation of the K-RAS oncogene promotes early-onset LC in mice [97]. The K-RAS proto-oncogene contains a nuclease hypersensitive element (NHE) with repetitive guanine residues in the promoter region. This region can adopt a quadruplex conformation, the topology of which has been solved [98]. Multiple transcription factors interact with the G4 structure formed in this region [99].

VEGF is commonly overexpressed in a variety of tumor types, including LC. It has an important role in sustaining the development and progression of LC [100]. Its expression is primarily regulated at the transcriptional level and can be induced during tumor hypoxia, hypoglycemia, loss of tumor suppressor genes, or activation of growth factor signaling cascades. Clinical studies have correlated increased VEGF mRNA and protein levels with tumor progression, leading to poorer prognosis and postoperative outcomes in NSCLC. The VEGF promoter contains a quadruplex-prone G-rich sequence near the transcription start site (TSS) [101] which is essential for basal or inducible VEGF transcription [28]. Its negative regulatory role in transcription has been demonstrated in vitro by the marked decrease in VEGF expression in the presence of quadruplex-stabilizing agents [102].

EGFR is a cell surface-receptor tyrosine kinase receptor that is frequently mutated and overexpressed in NSCLC [103]. For this reason, novel G4 ligands were developed as

alternative EGFR signaling inhibitors in LC therapy. A 30 nucleotide-long sequence located at position -272 in the EGFR promoter can adopt a G4 conformation, and G4 ligands, such as naphthalene diimide derivatives, are efficient inducers of this conformation, leading to decreased gene expression [104]. The major G4 in the human EGFR oncogene promoter can be targeted by the well-known G4 ligand, PDS [105].

HIFa and other genes involved in gene activation are often overexpressed in cancer and have a high density of G4-forming motifs in their promoters [106,107].

The WNT signal transduction cascade regulates several cellular processes, such as cell proliferation, survival, migration, and development, by initiating a series of signaling cascades downstream. Its abnormal activation is an independent poor prognostic marker of NSCLC [108,109]. A G-rich sequence within the WNT1 promoter can adopt both hairpin and G4 structures that can be used to modulate gene regulation [110]. **SJ26**, a tetracyclic azafluorenone compound, showed a moderate stabilization profile against G4s. It inhibits the WNT1-mediated downstream signaling pathway and suppressed the migration activity of H1299, MCF-7, and Hep2B 2.17 cancer cells [111].

Several compounds have been tested for their activity against two or more of the targets listed above. Triazole (**5b**) and morpholine (**3a** and **3b**) bisnaphthalimide-substituted derivatives proved to be multitarget G4 binders [112]. Compounds **3a**, **3b**, and **5b** showed high affinity toward hTELO, c-MYC, and c-KIT G4 (K_a > 4.4 μM^{-1}, determined by UV-Vis spectroscopy. These compounds inhibited the growth of A549 lung cancer cells (IC_{50} = 0.2-1 μM) with less cytotoxicity against non-cancerous MRC-5 cells [112]. Another multitarget compound is the indolonaphthyridine derivative **5** [113]. This compound stabilized the cMYC, BCL-2, HTS, and c-KIT1 G4s (ΔT_m = 3.4-6.3 °C; determined by FRET melting) with a relative preference against c-KIT1 and selectivity over duplex DNA. The A549 cell viability after treatment with 20 μM of this compound for 48 h was 69% [113]. Despite the limited cytotoxicity of **5**, this scaffold holds potential for further structural modifications and enhancement of its activities. Berberine and its derivative coptisine stabilize the KRAS quadruplex with two dominant binding modes [114]. These derivatives also showed thermal stabilization toward other G4 structures, such as c-MYC, VEGF, and two hybrid hTELO sequences, showing a preference for parallel c-MYC and VEGF. Moreover, berberine and coptisine significantly stall the Taq DNA polymerase synthesis of DNAs and lower KRAS mRNA levels, inhibiting the cell proliferation of H460 and A549 lung cancer cells [114].

Synergistic effects

Synergetic anticancer effects have previously been described with G4 ligands and antineoplastic treatments (see examples in **Figure 3.3b**). Neidle and colleagues [115] showed that a synergetic behavior on cell growth arrest was observed in lung adenocarcinoma cells with two unrelated G4 ligands, the trisubstituted acridine compound **AS1410** and triazole-derivative **CL1** in combination with cisplatin. **AS1410** showed only a moderate effect on tumor growth when administered alone but produced rapid and highly significant decreases in tumor volume when utilized in combination with cisplatin. Merle et al [93] reported that noncytotoxic concentrations of a G4 ligand (**Pt-ctpy**) showed a strong radiosensitizing effect on NSCLC cells. A seminal study [116] allowed the identification of G4 sensitizer genes, which behave as genetic vulnerabilities to G4 ligands when repressed, opening new therapeutic perspectives for G4 ligands in combinatorial treatments for patients underexpressing these proteins. At least two of these genes are overexpressed in NSCLC (*WEE1* and *USP1*), both proteins being targeted by specific inhibitors (pimozide is a potent USP1-targeting drug, whereas MK1775/AZD1775 is a WEE1 kinase inhibitor that has been evaluated in the clinic).

Loss of excision repair cross-complementation group 1 (ERCC1) is frequently found in LC [117] and is the most frequent DNA damage response (DDR) deficiency in NSCLC. XPF forms a heterodimer with ERCC1, and this structure-specific endonuclease has a crucial role in DNA damage excision, a ratelimiting step in nucleotide excision repair (NER). ERCC1 expression and cisplatin sensitivity are correlated: deficiencies in XPF and ERCC1 cause hypersensitivity to interstrand crosslinks. Interestingly, XPF is required for homologous recombination repair of double-strand breaks at DNA sequences forming G4s [118]. In line with this observation, XPF-deficient cells are exquisitely sensitive to G4 ligands, such as PDS, which are known to create replication stress [119] and alter the option of replication origins in human cells [120]. For a review on G4 and replication, we refer readers to that by Tracy M. Bryan [121].

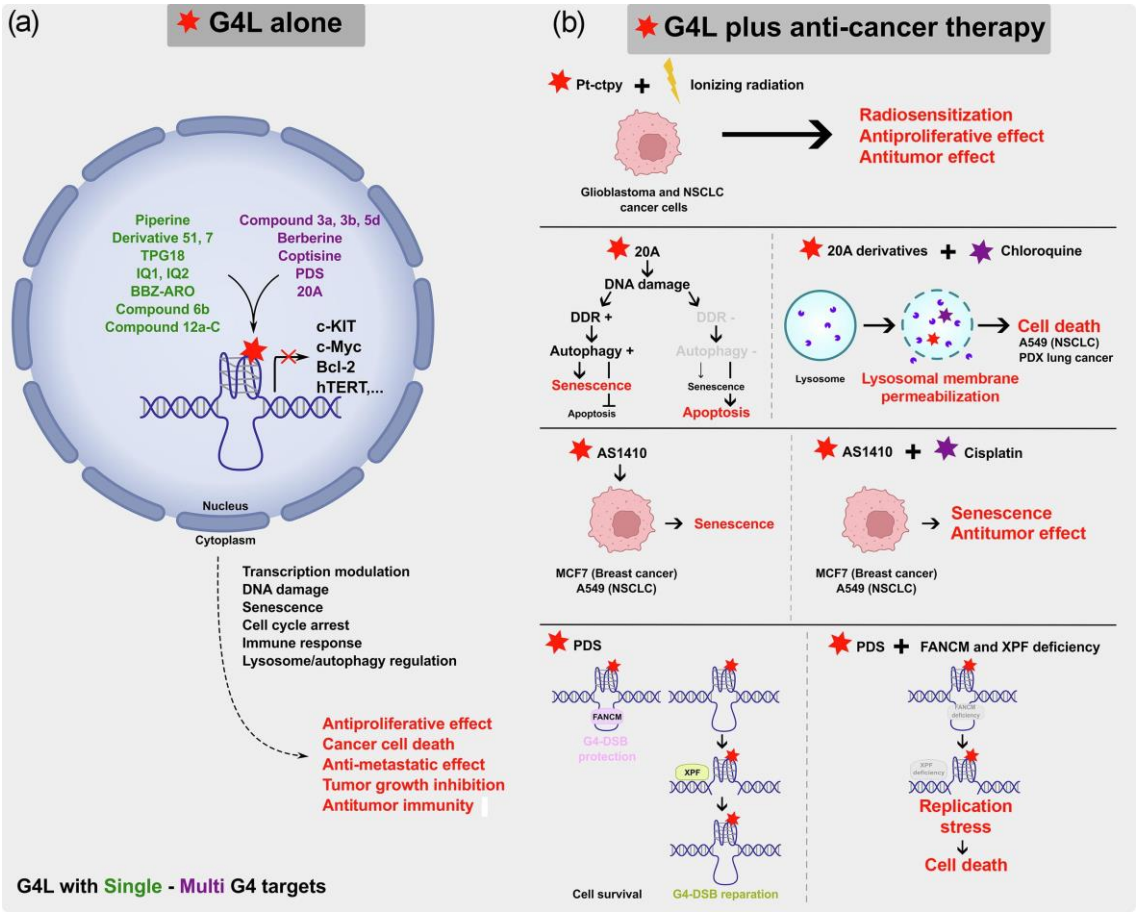


Figure 3.3. Examples of the effect of G-quadruplex (G4) ligands, either administered as standalone agents or in combination with anti-cancer treatments, on lung cancer (LC). **(a)** Mode of action of G4 ligands. G4 ligands can modulate the transcription of certain pivotal genes implicated in the development and progression of LCs [24-28] and can target multiple genes, whereas others have a more limited target scope. G4 ligands can influence transcription, cell cycle, senescence, DNA damage response (DDR), and immune response. Some G4 ligands also have the ability to accumulate within the lysosome compartment, which can lead to lysosomal membrane permeabilization and cell death [91, 122-128]. As a result, some G4 ligands display anti-cancer effects (in red). **(b)** Synergistic anti-cancer effects of G4 ligands. The G4-ligand **Pt-ctpy**, hampers tumor cell proliferation, disrupts cell cycle, and triggers apoptotic cell death in human glioblastoma and LC cells [93]. **Pt-ctpy** synergistically enhanced the efficacy of ionizing radiation, leading to the elimination of cancer cells *in vitro* and also reduced tumor growth in an *in vivo* cancer model [93] The **20A** promotes senescence in cancer cells through activation of both DDR and autophagy [91]. Disruption of either DDR or autophagy not only reduces senescence onset, but also directs cancer cells apoptosis, thus establishing DDR/autophagy as a pivotal link connecting senescence and apoptosis [76,89,90]. The combination of **20A** (including its derivatives) and chloroquine (an inhibitor of lysosomal functions) results in a marked increase in lysosomal membrane permeabilization, accompanied by a substantial induction of cell death [91]. **AS1410**, manifests a modest impact on tumor growth when administered as a single agent, its potency significantly increased when utilized in combination with cisplatin [115]. ERCC1/XPF is required for homologous recombination (HR)-mediated repair at DNA sequences derived from common fragile sites and G4s. The DNA translocase FANCM has a significant role in preventing DSB formation. Deficiency in XPF and FANCM causes cell sensitivity to the **pyridostatin** (**PDS**) [119]. Depletion of XPF within cells lacking FANCM significantly hampers cell proliferation following **PDS** treatment, suggesting a synthetic lethal interplay between XPF and FANCM [118].

Lysosomes have essential roles in the degradation and recycling of intra- and extracellular material, which stem from the autophagy and endocytosis pathways, respectively. They regulate many biological processes, including nutrient sensing, cellular signal transduction to the nucleus, cell death pathways, and immunosurveillance [122]. A body of evidence suggests that lysosomes undergo many alterations during tumor transformation (alteration of volume, composition, and distribution of lysosomes) [123]. These changes appear to be important for tumor progression, especially favoring tumor invasion, angiogenesis, or drug resistance. Paradoxically, these alterations make cancer cells more vulnerable to the permeabilization of their membrane (lysosomal membrane permeability; LMP), ultimately driving cells towards death [124]. LMP has been proposed as a new strategy to combat cancer drug resistance, particularly the acquired resistance to cisplatin in NSCLC [125]. LMP-induced cell death occurs through both caspase-dependent and -independent mechanisms. This observation could have important implications for cancer, particularly for the treatment of tumors that are resistant to therapy because of failure of the apoptotic machinery. Of note, some anticancer agents (including DNA ligands) lose their efficacity as a result of their accumulation within the lysosome, which causes a reduction in their accessibility to their cellular targets [126]. Interestingly, lysosomes appear to be involved in the fate of G4 ligands in lung cancer cells, given that Bierbach and colleagues [127] found a high degree of colocalization in lung adenocarcinoma cells of a platinum-acridine hybrid G4 ligand with LysoTracker, indicating accumulation of the hybrids in the lysosome but not in mitochondria. It was also shown that **20A**, a G4 ligand from the triarylpyridine family, accumulates within lysosomes. Inhibiting lysosomal functions with chloroquine significantly amplified LMP and cell death induced by this ligand. This phenomenon was observed not only in human lung carcinoma A549, but also in two patient-derived xenograft cell lines from lung adenocarcinomas and HeLa cells [91]. Similar effects were observed among three other G4 ligands from the triarylpyridine family, reinforcing the notion that targeting lysosomes holds significant promise as a strategy to augment cancer cell death induced by this class of G4 ligands [91,128].

Macroautophagy (also referred to as autophagy) represents a principal lysosomal catabolic process that has a role in fostering tumor resistance [129]. Substantial evidence also indicates that targeting autophagy can render cancer cells, including lung cancer cells, more responsive to some G4 ligand treatments [76,89,130]. This implies the potential involvement of autophagy in conferring resistance to certain G4 ligands. A master transcription factor involved in the regulation of lysosome is TFEB, which belongs to the MIT/TFE family of transcription factors. TFEB drives the expression of

several genes associated with lysosomal biogenesis and function (including those involved in autophagy), which all are a part of the coordinated lysosomal expression and regulation (CLEAR) network [131,132]. Interestingly, TFEB contains multiple regions compatible with G4s formation both at the DNA and RNA levels [128]. This suggests a direct modulation of lysosome function through the regulation of TFEB transcription or mRNA fate by quadruplex formation.

Altogether, these observations suggest that lung tumors can be targeted by G4 ligands, alone or in combination with other treatments. Recent evidence also indicates that G4 ligands can target other cellular components and compartments, such as lysosomes (for a review, see [128]) and mitochondria (for a review, see [133]), opening new opportunities for synergistic effects.

Table 3.1. G4 ligands discussed in this chapter.

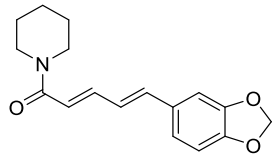
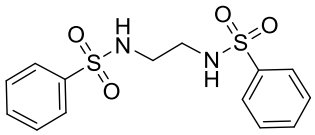
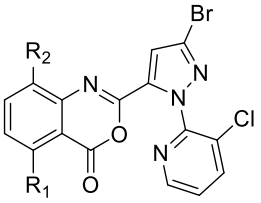
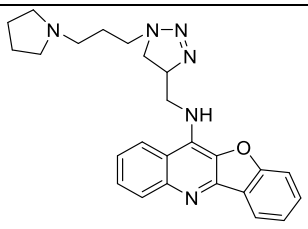
Ligand	G4 target	Affinity	Activity	Ref.
 **Piperine**	c-MYC	K_{D1} = 2.5 nM K_{D2} = 1.1μM	*In vitro* cytotoxicity A549 cells apoptosis	[57]
 **PIP-2**	c-MYC	K_D = 2 nM	Downregulation of c-MYC *In vitro* cytotoxicity A549 cells apoptosis	[58]
 **Benzoxaziones 1-3** **1** R_1=R_2=H **2** R_1=Br; R_2=H **3** R_1=H; R_2=OCH_3	c-MYC	-	Downregulation of c-MYC *In vitro* cytotoxicity Inhibition of A549 cell migration	[59]
 **Quindoline 5l**	c-MYC	K_D = 2.7 nM	*In vivo* antitumor growth inhibition (A549)	[61]

Table 3.1. (*continued*)

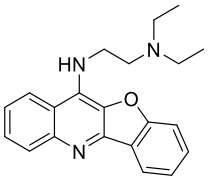
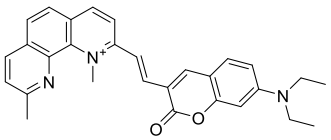
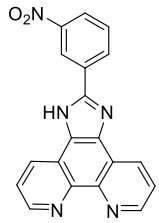
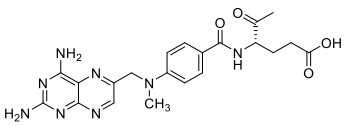
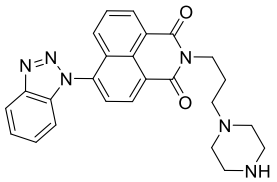
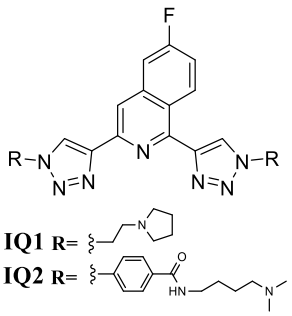
Ligand	G4 target	Affinity	Activity	Ref.
 **Quindoline 7**	c-MYC	K_D = 4.6 μM	*In vivo* antitumor growth inhibition (A549)	[60,61]
 **TGP18**	BCL-2	K_D = 728 nM	Downregulation of BCl-2 Cell cycle arrest A549 Cell apoptosis *In vitro* and *in vivo* antitumor effect	[73]
 **Phenanthroline 1**	BCL-2	K_a= 0.9 μM^{-1}	*In vitro* cytotoxicity on A549 cells Cell apoptosis	[74]
 **Methotrexate**	BCL-2	K_D = 9 nM	Downregulation of BCl-2 Decrease expression of bcl-2 protein *In vitro* cytotoxicity on A549 cells	[75]
 **Naphthalimide-benzotriazole 3a**	BCL-2	-	Downregulation of BCl-2 A549 cell apoptosis and autophagy induction DNA damage	[76]
 **IQ1** R= **IQ2** R=	hTELO	-	*In vitro* A549 antiproliferative effect Telomerase activity inhibition Arrest cancer cell growth Mitochondrial depolarization Cell apoptosis	[82]

Table 3.1. (*continued*)

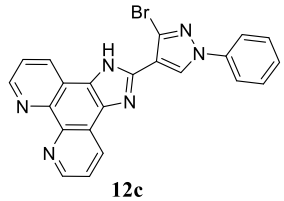
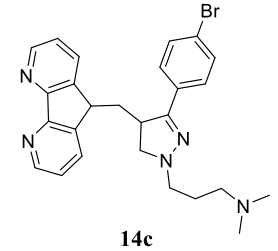
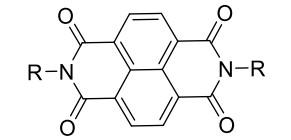
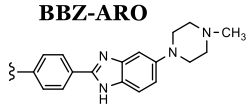
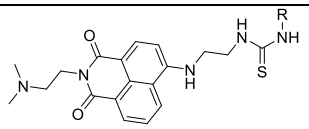
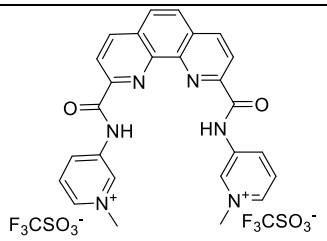
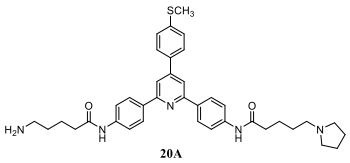
Ligand	G4 target	Affinity	Activity	Ref.
 **12c**	hTELO	-	*In vitro* A549 antiproliferative effect Inhibition of cell migration	[83]
 **14c**	hTELO	*K*a= 0.7 μM^{-1}	*In vitro* A549 antiproliferative effect Inhibition of cell migration Telomerase activity inhibition	[84]
 **BBZ-ARO** **R=** 	hTELO	*K*a = 6.6 μM^{-1}	Telomerase activity inhibition Cytotoxic effect on A549 cells	[85]
 **a** R=CH_3 **b** R=CH_2CH_3 **c** R=$(CH_2)_5CH_3$ **d** R=Ph	hTELO	*K*a > 0.1 μM^{-1}	A549 cell growth inhibition	[86]
 **6b**	hTELO	-	Antiproliferative effect in several LC cell lines	[87]
 **20A**	hTELO	-	A549 cell senescence and apoptosis DNA damage Lysosomal membrane permeabilization and A549 cell death (in combination with chloroquine)	[88-91]

Table 3.1. (*continued*)

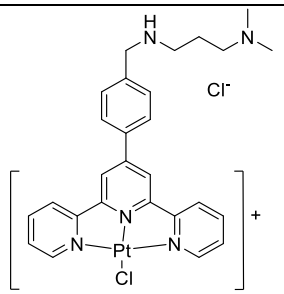
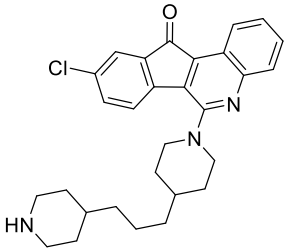
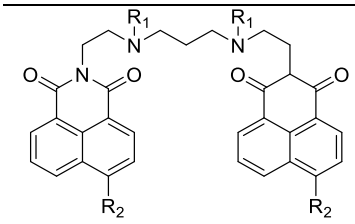
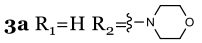
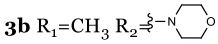
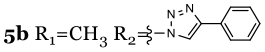
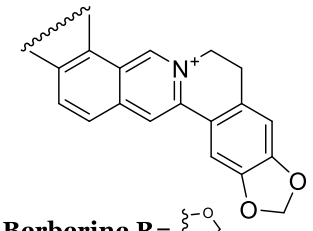
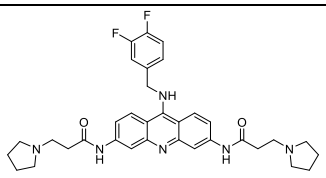
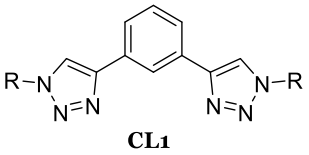
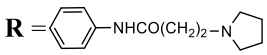
Ligand	G4 target	Affinity	Activity	Ref.
 **Pt-ctpy**	hTERT	-	Modified expression of *hTERT* and *TRF1* Radiosensibilization of A549 and H1299 cells	[93]
 **SJ26**	WNT1	-	Inhibition of WNT1-mediated signaling pathway Inhibition of H1299 cells migration	[111]
 **3a** R_1=H R_2=  **3b** R_1=CH_3 R_2=  **5b** R_1=CH_3 R_2= 	Multitarget	*K*a > 4.4 μM^{-1}	Inhibition of A549 cell growth	[112]
 **Berberine R=**  **Coptisine R=** 	Multitarget	-	Block Taq DNA polymerase synthesis Downregulation of KRAS mRNA Inhibition of H460 and A549 cells	[114]
 **AS1410**	hTELO	-	Cell senescence In vivo tumor growth inhibition (In combination with cis-platin)	[115]

Table 3.1. (*continued*)

Ligand	G4 target	Affinity	Activity	Ref.
 **CL1** 	hTELO	-	Cell senescence Inhibition of A549 cells growth (In combination with cis-platin)	115

Concluding remarks

In this review, we highlighted the importance of G4-forming sequences and their stabilization as a therapeutic approach for LC. Different ligands can stabilize G4s, leading to decreased proliferative activity of LC cells or, in some cases, decreased in vivo tumor growth. Unfortunately, even when showing interesting activity in cell culture, these ligands often exhibit limited efficacy in in vivo models. Therefore, one needs to better identify and characterize the targets of these ligands, which could result in uncovering novel strategies or synergies that enhance their efficacy in treating LC. Depending on the ligand, one can observe changes in telomere maintenance and/or alterations in gene expression. Supporting their potential as a therapeutic tool further experimental data confirmed that treatment with some G4 ligands leads to elevated levels of DNA damage, suggesting this as a therapeutic tool.

Another interesting issue is the accumulation of G4 ligands in other cellular organelles, such as lysosomes and mitochondria. Some G4 ligands, such as BMVC, have been demonstrated to localize within mitochondria, eventually leading to cell death in H1299 cells [134]. RHPS was observed to bind to mitochondrial DNA G4 at low doses, resulting in the acute inhibition of mitochondrial transcription elongation, ultimately leading to inhibition of the respiratory complex [135]. Despite these interesting results, their mechanisms of action remain poorly understood and further studies are required to identify target mitochondrial G4 and their impact on respiratory function. Most ligands reported here were targeted to a particular G4 structure; however, in most cases, these ligands can bind to other G4s, and it is difficult to predict their mode of action. The efficacy of ligands is not only related to their *in vitro* G4 stabilization ability, but also to

other issues, such as cellular uptake, release, their potential to be sequestrated into the lysosome, exposure time to cancer cells, and concentration used. At this point, comparing the effectiveness of ligands presents challenges because of variations in experimental conditions across studies. Some G4 ligands reported here demonstrated remarkable in vivo activity, and, in some instances, a synergetic effect was evident through the combination of a G4 ligand with drugs, as observed in the case of the G4 ligand **20A** and chloroquine.

Acknowledgments

J.F. acknowledges a doctoral fellowship grant from the Foundation for Science and Technology (FCT) ref. SFRH/ BD/145106/2019. C.C. acknowledges grants from project PAPILOMA ref. CENTRO-01-0145-FEDER-181235, NRC-LPCC Bolsa Dr. Rocha Alves 2022, and Instruct-ERIC Pilot R&D application ID 2473. This work received financial support from national funds FCT/MCTES through CICS-UBI ref. UIDB/00709/2020 and UIDP/00709/2020, was supported by funds from the Institut National de la Sante´ et de la Recherche Médicale (INSERM), and was funded by ANR G4Access (ANR-20-CE12-0023) and INCa (PL-BIO 2020-117) G4Access grants to J.L.M. M.D.M. is supported by funds from the Institut National de la Sante´ et de la Recherche Médicale (INSERM) and grants from Agence National de la Recherche (ANR-21-CE44-0016 (CISCO) and ANR-23-CE13 (JANUS), INCa (PLBIO23-216). L.F is supported by a doctoral fellowship from the Ecole Doctorale Paris-Saclay.

References

[1] Sung, H. *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71, 209–249.

[2] Nicholson, A.G. *et al.* (2022) The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *J Thorac Oncol* 17, 362–387.

[3] Akhtar, N. and Bansal, J.G. (2017) Risk factors of Lung Cancer in nonsmoker. *Curr Probl Cancer* 41, 328–339.

[4] Travis, W.D. *et al.* (2015) The 2015 World Health Organization Classification of Lung Tumors. *J Thorac Oncol* 10, 1243–1260.

[5] Herbst, R.S. *et al.* (2018) The biology and management of non-small cell lung cancer. *Nature* 553, 446–454.

[6] Martinez, P. *et al.* (2019) Immunotherapy for the First-Line Treatment of Patients with Metastatic Non–Small Cell Lung Cancer. *Clin Cancer Res* 25, 2691–2698.

[7] Politi, K. and Herbst, R.S. (2015) Lung Cancer in the Era of Precision Medicine. *Clin Cancer Res* 21, 2213–2220.

[8] El-Telbany, A. and Ma, P.C. (2012) Cancer Genes in Lung Cancer: Racial Disparities: Are There Any? *Genes Cancer* 3, 467–480.

[9] Wojtukiewicz, M.Z. *et al.* (2021) Inhibitors of immune checkpoints—PD-1, PD-L1, CTLA-4—new opportunities for cancer patients and a new challenge for internists and general practitioners. *Cancer Metastasis Rev* 40, 949–982.

[10] Tomasini, P. *et al.* (2012) Ipilimumab: its potential in non-small cell lung cancer. *Ther Adv Med Oncol* 4, 43–50.

[11] Antonia, S.J. *et al.* (2017) Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. *N Engl J Med* 377, 1919–1929.

[12] Forde, P.M. *et al.* (2022) Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. *N Engl J Med* 386, 1973–1985.

[13] Thai, A.A. *et al.* (2021) Lung cancer. *Lancet* 398, 535–554.

[14] Kosiol, N. *et al.* (2021) G-quadruplexes: a promising target for cancer therapy. *Mol Cancer* 20, 40.

[15] Miglietta, G. *et al.* (2021) G-quadruplex binders as cytostatic modulators of innate immune genes in cancer cells. *Nucleic Acids Res* 49, 6673–6686.

[16] Chung, S.-Y. *et al.* (2023) A G-quadruplex stabilizer, CX-5461 combined with two immune checkpoint inhibitors enhances in vivo therapeutic efficacy by increasing PD-L1 expression in colorectal cancer. *Neoplasia* 35, 100856.

[17] Wang, X.-D. *et al.* (2023) Novel phenanthrene imidazoles as telomeric G-quadruplex ligands trigger potent immunogenic cell death in triple-negative breast cancer. *Int J Biol Macromol* 249, 126068.

[18] Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.

[19] Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* 344, 410–414.

[20] Gellert, M. *et al.* (1962) HELIX FORMATION BY GUANYLIC ACID. *Proc Natl Acad Sci USA* 48, 2013–2018.

[21] Burge, S. *et al.* (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* 34, 5402–5415.

[22] Hänsel-Hertsch, R. *et al.* (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* 18, 279–284.

[23] Teng, F.-Y. *et al.* (2021) G-quadruplex DNA: a novel target for drug design. *Cell Mol Life Sci* 78, 6557–6583.

[24] Siddiqui-Jain, A. *et al.* (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c- *MYC* transcription. *Proc Natl Acad Sci USA* 99, 11593–11598.

[25] Rankin, S. *et al.* (2005) Putative DNA Quadruplex Formation within the Human c-kit Oncogene. *J Am Chem Soc* 127, 10584–10589.

[26] Dexheimer, T.S. *et al.* (2006) Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. *J Am Chem Soc* 128, 5404–5415.

[27] Cogoi, S. and Xodo, L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* 34, 2536–2549.

[28] Sun, D. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res* 33, 6070–6080

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

[30] Drygin, D. *et al.* (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res* 69, 7653–7661.

[31] Xu, H. and Hurley, L.H. (2022) A first-in-class clinical G-quadruplex-targeting drug. The bench-to-bedside translation of the fluoroquinolone QQ58 to CX-5461 (Pidnarulex). *Bioorg Med Chem Lett* 77, 129016.

[32] Local, A. *et al.* (2018) APTO-253 Stabilizes G-quadruplex DNA, Inhibits MYC Expression, and Induces DNA Damage in Acute Myeloid Leukemia Cells. *Mol Cancer Ther* 17, 1177–1186.

[33] Wang, Y. *et al.* (2019) G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat Commun* 10, 943.

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

[35] Zimmer, J. *et al.* (2016) Targeting BRCA1 and BRCA2 Deficiencies with G-Quadruplex-Interacting Compounds. *Mol Cell* 61, 449–460.

[36] 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.

[37] Castillo-Gonzalez, D. *et al.* (2013) FDA-approved Drugs Selected Using Virtual Screening Bind Specifically to G-quadruplex DNA. *Curr Pharm Des* 19, 2164–2173.

[38] Wang, K.-B. *et al.* (2019) Indenoisoquinoline Topoisomerase Inhibitors Strongly Bind and Stabilize the MYC Promoter G-Quadruplex and Downregulate MYC. *J Am Chem Soc* 141, 11059–11070.

[39] 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.

[40] Roxo, C. *et al.* (2019) G-Quadruplex-Forming Aptamers—Characteristics, Applications, and Perspectives. *Molecules* 24, 3781.

[41] 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.

[42] 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.

[43] Ireson, C.R. and Kelland, L.R. (2006) Discovery and development of anticancer aptamers. *Mol Cancer Ther* 5, 2957–2962.

[44] Xu, X. *et al.* (2016) G4-Tetra DNA Duplex Induce Lung Cancer Cell Apoptosis in A549 Cells. *Nanoscale Res Lett* 11, 437.

[45] Jing, N. *et al.* (2004) G-Quartet Oligonucleotides:A New Class of Signal Transducer and Activator of Transcription 3 Inhibitors That Suppresses Growth of Prostate and Breast Tumors through Induction of Apoptosis. *Cancer Res* 64, 6603–6609.

[46] Weerasinghe, P. *et al.* (2007) Inhibition of Stat3 activation and tumor growth suppression of non-small cell lung cancer by G-quartet oligonucleotides. *Int J Oncol* 31, 129–136.

[47] Monchaud, D. and Teulade-Fichou, M.-P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org Biomol Chem* 6, 627–636.

[48] Criscuolo, A. *et al.* (2022) Insights into the Small Molecule Targeting of Biologically Relevant G-Quadruplexes: An Overview of NMR and Crystal Structures. *Pharmaceutics* 14, 2361.

[49] Luo, Y. *et al.* (2022) Guidelines for G-quadruplexes: I. In vitro characterization. *Biochimie* 214, 5-23.

[50] Bhat-Ambure, J. *et al.* (2023) G4-QuadScreen: A Computational Tool for Identifying Multi-Target-Directed Anticancer Leads against G-Quadruplex DNA. *Cancers* 15, 3817.

[51] Chanvorachote, P. *et al.* (2020) C-myc Contributes to Malignancy of Lung Cancer: A Potential Anticancer Drug Target. *Anticancer Res* 40, 609–618.

[52] Sos, M.L. *et al.* (2012) A framework for identification of actionable cancer genome dependencies in small cell lung cancer. *Proc Natl Acad Sci* 109, 17034–17039.

[53] Mollaoglu, G. *et al.* (2017) MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer Cell* 31, 270–285.

[54] Llombart, V. and Mansour, M.R. (2022) Therapeutic targeting of "undruggable" MYC. *EBioMedicine* 75, 103756.

[55] González, V. and Hurley, L.H. (2010) The c-MYC NHE III(1): function and regulation. *Annu Rev Pharmacol Toxicol* 50, 111–129.

[56] Wang, W. *et al.* (2020) Human MYC G-quadruplex: From discovery to a cancer therapeutic target. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1874, 188410.

[57] Tawani, A. *et al.* (2016) Evidences for Piperine inhibiting cancer by targeting human G-quadruplex DNA sequences. *Sci Rep* 6, 39239.

[58] Pandya, N. and Kumar, A. (2021) Piperine analogs arrest c-myc gene leading to downregulation of transcription for targeting cancer. *Sci Rep* 11, 22909.

[59] Jiang, S. *et al.* (2020) Anti-cancer activity of benzoxazinone derivatives via targeting c-Myc G-quadruplex structure. *Life Sci* 258, 118252.

[60] Ou, T.-M. *et al.* (2007) Stabilization of G-Quadruplex DNA and Down-Regulation of Oncogene *c - myc* by Quindoline Derivatives. *J Med Chem* 50, 1465–1474.

[61] Zeng, D.-Y. *et al.* (2017) Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-*b*]quinolone Derivatives as *c-myc* G-Quadruplex Specific Stabilizers via Click Chemistry. *J Med Chem* 60, 5407–5423.

[62] Qian, S. *et al.* (2022) The role of BCL-2 family proteins in regulating apoptosis and cancer therapy. *Front Oncol* 12, 985363.

[63] Taylor, R.C. *et al.* (2008) Apoptosis: controlled demolition at the cellular level. *Nat Rev Mol Cell Biol* 9, 231–241.

[64] Zhu, C.-Q. (2006) Immunohistochemical markers of prognosis in non-small cell lung cancer: a review and proposal for a multiphase approach to marker evaluation. *J Clin Pathol* 59, 790–800.

[65] Choi, J. *et al.* (2005) Bcl-2 Promotes Invasion and Lung Metastasis by Inducing Matrix Metalloproteinase-2. *Cancer Res* 65, 5554–5560.

[66] Hu, Y. *et al.* (2004) Antitumor Efficacy of Oblimersen Bcl-2 Antisense Oligonucleotide Alone and in Combination with Vinorelbine in Xenograft Models of Human Non–Small Cell Lung Cancer. *Clin Cancer Res* 10, 7662–7670.

[67] Radha, G. and Raghavan, S.C. (2017) BCL2: A promising cancer therapeutic target. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1868, 309–314.

[68] Rantanen, S. *et al.* (2001) Causes and Consequences of BCL2 Overexpression in Diffuse Large B-Cell Lymphoma. *Leuk Lymphoma* 42, 1089–1098.

[69] Agrawal, P. *et al.* (2014) The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K in K^+ solution. *J Am Chem Soc* 136, 1750–1753.

[70] Dai, J. *et al.* (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res* 34, 5133–5144.

[71] Onel, B. *et al.* (2016) A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription. *J Am Chem Soc* 138, 2563–2570.

[72] Singh, M. *et al.* (2022) BCL2 G quadruplex-binding small molecules: Current status and prospects for the development of next-generation anticancer therapeutics. *Drug Discov Today* 27, 2551–2561.

[73] Suseela, Y.V. *et al.* (2020) Recognition of G-quadruplex topology through hybrid binding with implications in cancer theranostics. *Theranostics* 10, 10394–10414.

[74] Li, L. *et al.* (2017) Microwave-Assisted Synthesis of Imidazo[4,5-f][1,10]phenanthroline Derivatives as Apoptosis Inducers in Chemotherapy by Stabilizing Bcl-2 G-quadruplex DNA. *Molecules* 22, 829.

[75] Pandya, N. *et al.* (2021) Interaction analysis of anti-cancer drug Methotrexate with bcl-2 promoter stabilization and its transcription regulation. *Gene Rep* 23, 101155.

[76] Wang, X. *et al.* (2022) Design, synthesis and bioactivity of novel naphthalimide-benzotriazole conjugates against A549 cells via targeting BCL2 G-quadruplex and inducing autophagy. *Life Sci* 302, 120651.

[77] Jafri, M.A. *et al.* (2016) Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med* 8, 69.

[78] Lu, C. *et al.* (2004) Prognostic Factors in Resected Stage I Non–Small-Cell Lung Cancer: A Multivariate Analysis of Six Molecular Markers. *J Clin Oncol* 22, 4575–4583.

[79] Taka, T. *et al.* (2013) Telomere shortening and cell senescence induced by perylene derivatives in A549 human lung cancer cells. *Bioorg Med Chem* 21, 883–890.

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

[81] Mergny, J.-L. and Hélène, C. (1998) G-quadruplex DNA: A target for drug design. *Nat Med* 4, 1366–1367.

[82] Maiti, S. *et al.* (2018) Human Telomeric G-Quadruplex Selective Fluoro-Isoquinolines Induce Apoptosis in Cancer Cells. *Bioconjug Chem* 29, 1141–1154.

[83] Liu, J. *et al.* (2017) Synthesis and the interaction of 2-(1 H -pyrazol-4-yl)-1 H -imidazo[4,5-f][1,10]phenanthrolines with telomeric DNA as lung cancer inhibitors. *Eur J Med Chem* 133, 36–49.

[84] Zhou, K. *et al.* (2019) Design, synthesis of 4,5-diazafluorene derivatives and their anticancer activity via targeting telomeric DNA G-quadruplex. *Eur J Med Chem* 178, 484–499.

[85] Sur, S. *et al.* (2017) Naphthalenediimide-Linked Bisbenzimidazole Derivatives as Telomeric G-Quadruplex-Stabilizing Ligands with Improved Anticancer Activity. *ACS Omega* 2, 966–980.

[86] Ou, Z. *et al.* (2017) Synthesis, G-quadruplex binding properties and cytotoxicity of naphthalimide–thiourea conjugates. *New J Chem* 41, 9397–9405.

[87] Craciun, A.-M. *et al.* (2021) New 2,9-disubstituted-1,10-phenanthroline derivatives with anticancer activity by selective targeting of telomeric G-quadruplex DNA. *Spectrochim Acta A Mol Biomol Spectrosc* 249, 119318.

[88] Smith, N.M. *et al.* (2011) Unraveling the relationship between structure and stabilization of triarylpyridines as G-quadruplex binding ligands. *Org Biomol Chem* 9, 6154.

[89] Beauvarlet, J. *et al.* (2019) Modulation of the ATM/autophagy pathway by a G-quadruplex ligand tips the balance between senescence and apoptosis in cancer cells. *Nucleic Acids Res* 47, 2739–2756.

[90] Beauvarlet, J. *et al.* (2019) Activation of the Ataxia Telangiectasia Mutated/Autophagy pathway by a G-quadruplex ligand links senescence with apoptosis. *Mol Cell Oncol* 6, 1604047.

[91] Beauvarlet, J. *et al.* (2020) Triarylpyridine Compounds and Chloroquine Act in Concert to Trigger Lysosomal Membrane Permeabilization and Cell Death in Cancer Cells. *Cancers (Basel)* 12, 1621.

[92] Monsen, R.C. *et al.* (2020) The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res* 48, 5720–5734.

[93] Merle, P. *et al.* (2015) Highly efficient radiosensitization of human glioblastoma and lung cancer cells by a G-quadruplex DNA binding compound. *Sci Rep* 5, 16255.

[94] Blackhall, F.H. *et al.* (2003) Expression and Prognostic Significance of Kit, Protein Kinase B, and Mitogen-activated Protein Kinase in Patients with Small Cell Lung Cancer 1. *Clin Cancer Res* 9, 2241–2247.

[95] 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.

[96] Blasco, R.B. *et al.* (2011) c-Raf, but Not B-Raf, Is Essential for Development of K-Ras Oncogene-Driven Non-Small Cell Lung Carcinoma. *Cancer Cell* 19, 652–663.

[97] Johnson, L. *et al.* (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* 410, 1111–1116.

[98] 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.

[99] Cogoi, S. *et al.* (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res* 36, 3765–3780.

[100] Frezzetti, D. *et al.* (2017) VEGF as a potential target in lung cancer. *Expert Opin Ther Targets* 21, 959–966.

[101] Muench, D. *et al.* (2019) Quadruplex-forming oligonucleotide targeted to the VEGF promoter inhibits growth of non-small cell lung cancer cells. *PLoS One* 14, e0211046.

[102] Taka, T. *et al.* (2012) Down-regulation of the human VEGF gene expression by perylene monoimide derivatives. *Bioorg Med Chem Lett* 22, 518–522.

[103] Passaro, A. *et al.* (2021) Overcoming therapy resistance in EGFR-mutant lung cancer. *Nat Cancer* 2, 377–391.

[104] Recagni, M. *et al.* (2019) Distinct biological responses of metastatic castration resistant prostate cancer cells upon exposure to G-quadruplex interacting naphthalenediimide derivatives. *Eur J Med Chem* 177, 401–413.

[105] Liu, Y. *et al.* (2023) Structure of the Major G-Quadruplex in the Human EGFR Oncogene Promoter Adopts a Unique Folding Topology with a Distinctive Snap-Back Loop. *J Am Chem Soc* 145, 16228–16237.

[106] Fleming, A.M. *et al.* (2015) A Role for the Fifth G-Track in G-Quadruplex Forming Oncogene Promoter Sequences during Oxidative Stress: Do These "Spare Tires" Have an Evolved Function? *ACS Cent Sci* 1, 226–233.

[107] Fleming, A.M. *et al.* (2019) Location dependence of the transcriptional response of a potential G-quadruplex in gene promoters under oxidative stress. *Nucleic Acids Res* 47, 5049–5060.

[108] Xu, X. *et al.* (2011) Aberrant Wnt1/β-Catenin Expression is an Independent Poor Prognostic Marker of Non-small Cell Lung Cancer After Surgery. *J Thorac Oncol* 6, 716–724.

[109] Clevers, H. and Nusse, R. (2012) Wnt/β-Catenin Signaling and Disease. *Cell* 149, 1192–1205.

[110] Kuo, M.H.-J. *et al.* (2015) Conformational Transition of a Hairpin Structure to G-Quadruplex within the WNT1 Gene Promoter. *J Am Chem Soc* 137, 210–218.

[111] Chang, L.-C. *et al.* (2016) Identification of a new class of WNT1 inhibitor: Cancer cells migration, G-quadruplex stabilization and target validation. *Oncotarget* 7, 67986–68001.

[112] Ou, Z. *et al.* (2019) Novel triazole and morpholine substituted bisnaphthalimide: Synthesis, photophysical and G-quadruplex binding properties. *J Mol Struct* 1185, 27–37.

[113] Catalano, R. *et al.* (2019) Targeting multiple G-quadruplex–forming DNA sequences: Design, biophysical and biological evaluations of indolo-naphthyridine scaffold derivatives. *Eur J Med Chem* 182, 111627.

[114] Wang, K.-B. *et al.* (2022) Structural insight into the bulge-containing KRAS oncogene promoter G-quadruplex bound to berberine and coptisine. *Nat Commun* 13, 6016.

[115] 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.

[116] Zyner, K.G. *et al.* (2019) Genetic interactions of G-quadruplexes in humans. *Elife* 8, e46793.

[117] Olaussen, K.A. *et al.* (2007) ERCC1 as a risk stratifier in platinum-based chemotherapy for nonsmall-cell lung cancer. *Curr Opin Pulm Med* 13, 284–289.

[118] Li, S. *et al.* (2019) ERCC1/XPF Is Important for Repair of DNA Double-Strand Breaks Containing Secondary Structures. *iScience* 16, 63–78.

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

[120] Prorok, P. *et al.* (2019) Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat Commun* 10, 3274.

[121] Bryan, T.M. (2019) Mechanisms of DNA Replication and Repair: Insights from the Study of G-Quadruplexes. *Molecules* 24, 3439.

[122] Lawrence, R.E. and Zoncu, R. (2019) The lysosome as a cellular centre for signalling, metabolism and quality control. *Nat Cell Biol* 21, 133–142.

[123] Kallunki, T. *et al.* (2013) Cancer-associated lysosomal changes: friends or foes? *Oncogene* 32, 1995–2004.

[124] Aits, S. and Jäättelä, M. (2013) Lysosomal cell death at a glance. *J Cell Sci* 126, 1905–1912.

[125] Circu, M. *et al.* (2017) Modulating lysosomal function through lysosome membrane permeabilization or autophagy suppression restores sensitivity to cisplatin in refractory non-small-cell lung cancer cells. *PLoS One* 12, e0184922.

[126] Kang, C.-C. *et al.* (2013) Chemical principles for the design of a novel fluorescent probe with high cancer-targeting selectivity and sensitivity. *Integr Biol* 5, 1217–1228.

[127] Pickard, A.J. *et al.* (2014) Redesigning the DNA-Targeted Chromophore in Platinum-Acridine Anticancer Agents: A Structure-Activity Relationship Study. *Chem A Eur J* 20, 16174–16187.

[128] Ferret, L. *et al.* (2023) G-quadruplex ligands as potent regulators of lysosomes. *Autophagy* 19, 1901–1915.

[129] Ishaq, M. *et al.* (2020) Autophagy in cancer: Recent advances and future directions. *Semin Cancer Biol* 66, 171–181.

[130] Orlotti, N.I. *et al.* (2012) Autophagy acts as a safeguard mechanism against G-quadruplex ligand-mediated DNA damage. *Autophagy* 8, 1185–1196.

[131] Napolitano, G. and Ballabio, A. (2016) TFEB at a glance. *J Cell Sci* 129, 2475–2481.

[132] Settembre, C. and Ballabio, A. (2014) Lysosomal Adaptation: How the Lysosome Responds to External Cues. *Cold Spring Harb Perspect Biol* 6, a016907–a016907.

[133] Sahayasheela, V.J. *et al.* (2023) Mitochondria and G-quadruplex evolution: an intertwined relationship. *Trends Genet* 39, 15–30.

[134] Huang, W.-C. *et al.* (2015) Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res* 43, 10102–10113.

[135] Falabella, M. *et al.* (2019) G-quadruplex dynamics contribute to regulation of mitochondrial gene expression. *Sci Rep* 9, 5605.

4th Chapter

Aims of the thesis

Aims of the thesis

Despite all efforts, LC remains the leading cause of worldwide cancer-related mortality. Late-stage diagnosis and difficult treatment in advanced stages remain difficult to overcome and novel therapeutic strategies are urgently needed. Among these, the use of small organic compounds to target G4 within LC-related genes offers a promising avenue for modulating cellular processes involved in LC progression.

Therefore, the global aim of this doctoral thesis consisted in the development of novel phenanthroline-based derivatives capable of binding and stabilizing different G4 sequences related to LC progression, aiming at potential anti-tumor effects. Furthermore, this work has the objective to explore the potential of one RNA G4 sequence to recognize and bind to nucleolin. Based on the obtained findings, this project intends to furnish potential drug candidates to be used in LC treatment highlighting the structure-activity relationship of this class of compounds.

To achieve this main goal, the present work is divided into the following specific objectives:

- To evaluate different heteroaromatic scaffolds in terms of its G4s stabilizer properties and anti-tumor effects.
- To synthesize novel 1,10-phenanthroline-based derivatives and to evaluate their potential to target G4 structures with specificity over duplex DNA.
- To biophysically characterize those synthesized compounds in terms of their G4 binding and/or stabilizer properties through FRET-melting, circular dichroism, and fluorescence techniques.
- To evaluate i*n vitro* the antiproliferative activity induced by compounds in LC cancer and healthy cell lines.
- To characterize the formation of the G4 structure in the full-length human pre-miRNA 150 sequence and to determine its affinity for NCL expressed in LC cells.

The next section is divided into four chapters (5^{th} to 8^{th} chapter) and will present in detail all experimental procedures employed to achieve these aims, as well as the exposition and discussion of all results.

2nd SECTION

Experimental section

5th Chapter

Screening of Scaffolds for the Design of G-Quadruplex Ligands

This work was published in:

Figueiredo, J.[‡]*; Peitinho, D.; Campello, M.P.C.; Oliveira, M.C.; Paulo, A.; Mergny, J-L. and Cruz, C. (2022) Screening of scaffolds for the design of G-quadruplex ligands. Appl. Sci 12, 2170. DOI: 10.3390/app12042170*

Chapter overview

The research for small organic molecules capable of interacting with G4 structures comprises conventional medicinal chemistry methodologies. These methodologies may include rational drug design of *hit* or *lead* molecules, which can involve the design of G4-interactive small molecules based on G4 structure and structure-activity relationships studies. Additionally, optimization strategies are employed to enhance the potency, selectivity, and pharmacokinetic properties of the promising scaffolds to discover new *lead* candidates. This chapter describes the screening of several scaffolds featuring different planar heteroaromatic cores, such as oxazole, acridine, naphthalene, and phenanthroline moieties towards different G4 structures found within c-MYC, KRAS, and VEGF promoters, human telomeric F21T, and pre-MIR150 RNA G4. It describes the chemical synthesis of the naphthalene and phenanthroline scaffolds. The different scaffolds were screened using FRET-melting studies, a high-throughput method, to assess their G4-stabilizers properties. Moreover, their cytotoxic potential in LC and healthy cell lines was assessed, and their drug-like properties were predicted using *in silico* methods. This study aims to provide insights into G4-interacting scaffolds, leading to their further structural optimization.

Screening of Scaffolds for the Design of G-Quadruplex Ligands

Abstract

In the last decade, progress has been made in G-quadruplex (G4) ligands development, but for most compounds, the ligand binding mode is speculative or based on low resolution methods, with its discovery based on structure-based approaches. Herein, we report the synthesis of small (MW < 400 Da) heterocycle compounds, containing different aromatic scaffolds, such as phenyl, quinoline, naphthalene, phenanthroline and acridine moieties, in order to explore their stabilization effect towards different DNA G4s, such as those found in c-MYC, KRAS21 and VEGF promoters, 21G human telomeric motif and pre-MIR150. The fluorescence resonance energy transfer (FRET) melting assay indicates that the acridine moiety is the most active scaffold, followed by phenanthroline. The different scaffolds are promising in terms of drug-like properties and, in general, the IC_{50} values of the respective heterocycle compounds are lower in a cancer cell line, when compared with a normal cell line. The acridine derivative C_5NH_2 has the most favorable cytotoxic profile in terms of cell selectivity.

Keywords

G-quadruplex; heterocycle compounds; scaffolds; drug-design

Introduction

The formation of G4 requires monovalent (Na^+ and K^+) or divalent/multivalent (e.g., Sr^{2+}) cations, or small molecules, known as G4 ligands, which are chemical compounds that specifically bind and stabilize the structure of G4. Many efforts are being made to target G4s as a therapeutic approach, given their implication in carcinogenesis [1] or virology [2]. As such, G4s provide recognition sites for ligands, since different G4 structures adopt specific conformations. These ligands generally have an aromatic surface, allowing π-π stacking interactions with a terminal G-quartet, one or more positive charge(s) or basic groups to selectively bind to the loops or grooves of the G4,

and a geometry/shape preventing intercalation into double-stranded DNA. Many G4 ligands have characteristic cores that can be chemically modified, allowing the synthesis of various analogues whose therapeutic activity in cancer is being investigated. Disubstituted amidoanthraquinones were the first G4 ligands to be reported [3]. Later on, the cationic porphyrin TMPyP4 emerged as a potent telomerase inhibitor upon binding to telomeric G4s [4]. However, TMPyP4 also binds to duplex and triplex DNA [5,6]. Natural compounds, such as telomestatin, berberine and its derivatives, were considered potent telomerase inhibitors [7,8]. However, the structural complexity of telomestatin is a huge obstacle for its efficient large-scale synthesis and production as a drug. Pyridostatin, BRACO-19, naphthalene diimides and bisquinolinium compounds, such as PhenDC3, have been used to broadly promote G4 formation in high-throughput sequencing of G4 structures [9]. Some of these G4 ligands have entered clinical trials, namely the fluoroquinolones CX-3543 and CX-5461, that selectively bind and stabilize a broad spectrum of G4 structures, including those harboured in c-MYC, c-KIT, and telomeres [10]. These fluoroquinolones interact with a G4 found in ribosomal DNA and disrupt the binding between these G4s and protein complexes, thereby inhibiting ribosome biogenesis [11]. Other novel ligands were reported to be selective for a sub-class of quadruplex structures, such as the diquinolinyl-pyridine ligands, that show a preference for the parallel conformation of telomeric, c-MYC and c-KIT G4s [12], or ToxaP, which prefers an antiparallel fold [13]. However, there is still a long way to go for the development of G4 ligands. Indeed, the multiplicity of potential targets, the variety of ligand binding sites and the differences in the ligand effects, in vitro and in vivo, make it difficult to unravel if the stabilization or destabilization of G4s promotes or inhibits gene expression. A major limitation for the clinical application of G4 ligands is related to their selectivity and potential side effects on normal cells. In this regard, it is important to screen different scaffolds and analyze the binding to G4s and the cellular effects on normal cells and cancer cells. In this context, we synthetized and tested some heterocycle compounds that will be able to act as a scaffold for G4 ligand/drug design. Different scaffolds, such as quinoline, phenyl, acridine and phenanthroline (**Figure 5.1**), were investigated to their capacity to stabilize different DNA and RNA G4 structures by the FRET melting assay [14]. Additionally, the in silico molecular properties were predicted and the cytotoxicity effects were also evaluated in a lung cancer cell line (A549) *vs.* normal cell (NHDF).

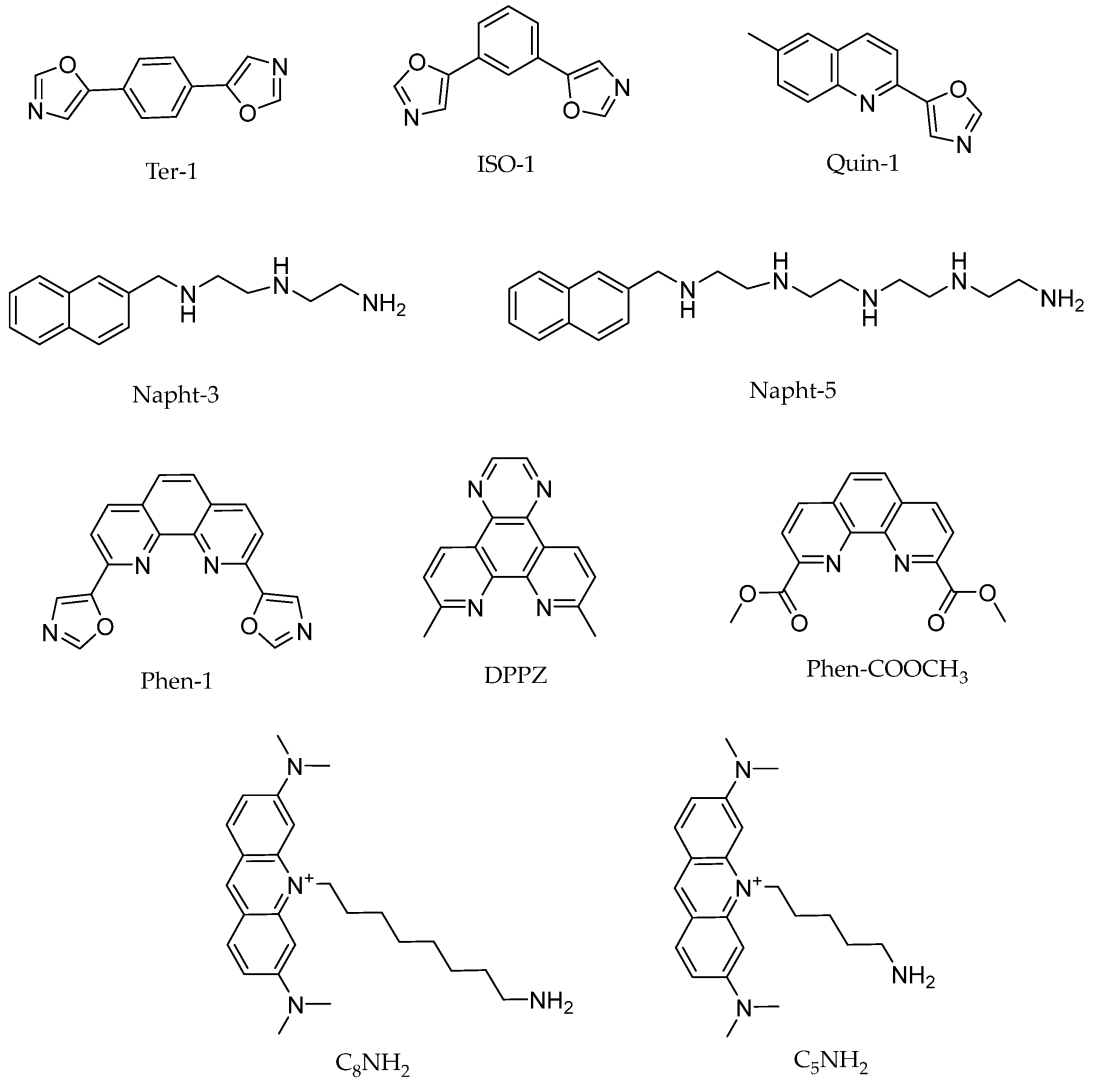


Figure 5.1. Chemical structures of the different scaffolds used in the study.

Materials and Methods

Chemical synthesis

All reagents and solvents were used without additional purification, unless stated otherwise, and were obtained from commercial suppliers such as Sigma-Aldrich (St. Louis, MO, USA) and Across Organics (Geel, Belgium). Chemical reactions were followed by thin-layer chromatography (TLC) using Merck-Nagel (0.2 mm) plates which are visualized by UV detection. Melting points were measured for *in vitro* tested compounds and were recorded on a Büchi B-540 melting point apparatus and are uncorrected.

^{1}H and ^{13}C spectra were performed at 25 °C on a Bruker Avance III 400 MHz spectrometer with a TXI probe. The spectra were processed in the software MestReNova

12.0.0. (Mestrelab Research S.L., Santiago de Compostela, Spain) and are presented in the Appendix A (**Figure A1**).

Synthesis of Phenyl and Quinoline derivatives

The synthesis of 1,4-bis(oxazol-5-yl)benzene (Ter-1), 1,3-bis(oxazol-5-yl)benzene (ISO-1) and 6-methyl-(2-oxazol-5-yl)quinoline (Quin-1) was performed as previously described [15].

Synthesis of Naphthalene derivatives

The naphthalene derivatives ligands were prepared according to methods described previously, with slight modifications [16]. Briefly, to a solution of the desired polyamine (1.0 g, 9.6 mmol) in a mixture (40 mL) of absolute CH_3CN:EtOH (*v*/*v*; 1.7:1) at room temperature, was added dropwise one solution of 1-naphthaldehyde (0.5 g, 3.2 mmol) in dried CH_3CN (10 mL). The solution was stirred at room temperature over 2 h and the solvent was evaporated to dryness. The resulting residue was dissolved in dried EtOH (15 mL) and $NaBH_4$ (1.1 g, 29.5 mmol) was added portion wise. The mixture was stirred at room temperature for 24 h. After that, the mixture was filtered, and the solvent was evaporated under reduced pressure. The residue obtained was extracted with chloroform (3 × 50 mL) and evaporated to dryness obtaining one oil which was dissolved in a mixture of MeOH:EtOH (*v*/*v*; 1:1) and precipitated with some drops of HCl 37% to give the different naphthalene derivatives:

N^1-(2-aminoethyl)-*N*^{2}-(naphtalen-1-ylmethyl)etane-1,2-diamine (Napht-3)

From diethylenetriamine (1.0 g, 9.6 mmol) and 1-naphtaldeyde (0.5 g, 3.2 mmol). Yield: 79% (1.2 g, 4.9 mmol). White solid. mp 203 °C. ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 9.99–9.35 (m, 3H), 8.18 (t, *J*= 10.1 Hz, 1H), 7.98–7.93 (m, 1H), 7.72–7.46 (m, 2H), 4.67 (s, 1H), 3.44 (s, 4H), 3.11 (s, 4H). ^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ 133.7, 131.4, 130.2, 129.3, 129.2, 128.3, 127.4, 126.9, 125.8, 124.2, 47.6, 44.4, 43.7, 43.2, 35.6.

N^1-(2-aminoethyl)-*N*^{2}-(2-((2-((naphtalen-1-ylmethyl) amino) ethyl) amino) etyl) etane-1,2-diamine (Napht-5)

From tetraethylenepentamine (1.8 g, 9.6 mmol) and 1-naphtaldeyde (0.5 g, 3.2 mmol). Yield: 11% (0.2 g, 0.7mmol). White solid. mp 310 °C. ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 10.67–8.64 (m, 2H), 8.50–8.26 (m, 1H), 8.19–7.97 (m, 1H), 7.96–7.44 (m, 2H), 4.83 (d, *J* = 8.0 Hz, 1H), 4.19 (s, 1H), 1.32 (s, 1H). ^{13}C NMR (101 MHz, DMSO-*d*$_6$) 133.8, 131.4,

130.2, 129.3, 129.2, 128.3, 127.4, 126.9, 125.8, 124.2, 70.3, 47.7, 44.6, 44.0, 43.4, 36.0, 31.8, 29.5, 29.2.

Synthesis of Phenanthroline Derivatives

2,9-Bis(oxazole-5-yl)-1,10-phenantroline (Phen-1)

The synthesis was performed as previously described [15].

2,9-Dimethyl-1,10-phenanthroline-5,6-dione

2,9-Dimethyl-1,10-phenanthroline (2.0 g, 9.6 mmol) was added into a solution of 60% H_2SO_4 (23 mL). To this solution, potassium bromate was added in batches (1.8 g, 11.0 mmol) for 1 h and the mixture was stirred at room temperature for 20 h. After that, the solution was poured into water (200 mL) and carefully neutralized to pH 7 using a solution of sodium hydroxide 8 M [17]. The precipitate was filtered and washed with water giving a yellow solid with 70% yield (1.6 g, 6.7 mmol); mp 177 °C. ^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ (ppm) 8.26 (d, 2H, CH), 7.52 (d, 2H, CH) 2.68 (s, 6H, CH_3); ^{13}C NMR (101 MHz, DMSO-*d*$_6$) 178.4, 164.7, 152.3, 136.6, 127.2, 125.2, 25.3.

7,10-Dimethylpyrazino[2,3-*f*][1,10]phenanthroline (DPPZ)

2,9-Dimethy-1,10-phenanthroline-5,6-dione (0.6 g, 2.5 mmol) was dissolved in dried EtOH (50 mL) and ethylenediamine was added (0.2 g, 2.9 mmol). The stirred solution was refluxed overnight. After cooling the solution was concentrated to half of the initial volume [18]. The brown solid which formed was filtered obtaining 60% of yield (0.4 g, 1.5 mmol); mp 265–266 °C. ^{1}H NMR (400 MHz, DMSO-*d*$_6$): δ (ppm) 9.30 (d, J = 8.3 Hz, 2H, CH), 9.10 (s, 1H), 7.80 (d, J = 8.3 Hz, 2H, CH), 2.85 (s, 6H, CH_3) ^{13}C NMR (101 MHz, DMSO-*d*$_6$): δ (ppm) 161.4, 146.6, 145.4, 139.8, 133.2, 124.8, 124.7, 25.5 (CH_3).

2,9-Dicarbaldehyde-1,10-phenanthroline (PhenCOH)

SeO_2 (4.3 g, 38.4 mmol) was taken in dioxane (100 mL) at 60 °C. To this solution, a small increment of 2,9-dimethyl-1,10-phenanthroline (4.0 g, 19.2 mmol) dissolved in dioxane (35 mL) was added over 2 h and refluxed at 100 °C overnight. The mixture was filtered through celite in a hot condition and dioxane was removed under reduced pressure. The residue was purified by column chromatography using silica gel (0.060–0.200 mm, 60 Å and MeOH:$CHCl_3$ (v/v; 95:5) solvent mixture [19]. After complete separation, the solution was evaporated under vacuum conditions to give a pale yellow solid with a 50% yield (2.3 g, 9.7 mmol); ^{1}H NMR (400 MHz, $CDCl_3$-*d*$_6$): δ (ppm) 10.57 (s, 2H, CHO), 8.52

(d, J = 8.2 Hz, 2H, CH), 8.39 (d, J = 8.1 Hz, 2H, CH), 8.06 (s, 2H, CH). ^{13}C NMR (101 MHz, $CDCl_3$-d_6) δ (ppm) 193.4 (COH), 152.7, 146.0, 138.0, 131.6, 129.1, 120.5.

1,10-Phenanthroline-2,9-dicarboxylic Acid (PhenCOOH)

2,9-Dicarbaldehyde-1,10-phenanthroline (0.6 g, 2.4 mmol) was added into HNO_3 (2 mL; 65 wt %). The mixture was refluxed over 6 h, and then the content was poured into cursed ice [19]. The solid was filtered to give a light-yellow final product with an 89% yield (0.6 g, 2.2 mmol); ^{1}H NMR (400 MHz, DMSO-d_6): δ (ppm) 8.74 (d, J = 8.3 Hz, 2H, CH), 8.42 (d, J = 8.2 Hz, 2H, CH), 8.22 (s, 2H, CH). ^{13}C NMR (101 MHz, DMSO-d_6): δ (ppm) 166.7, 148.7, 145.1, 138.7, 131.0, 129.0, 124.0, 40.6, 40.4, 40.2, 40.0, 39.7, 39.5, 39.3.

2,9-Dicarbomethoxy-1,10-phenanthroline (Phen-$COOCH_3$)

1,10-Phenanthroline-2,9-dicarboxylic acid (0.6 g, 2.1 mmol) was suspended in MeOH (150 mL) and conc. H_2SO_4 (15 mL) was added to this solution. The reaction mixture was refluxed overnight [19]. After that, the mixture was poured into crushed ice and allowed to cool. The precipitate was filtered and was obtained as a light yellow product with a 53% yield (0.3 g, 1.0 mmol); mp 184–186 °C. ^{1}H NMR (400 MHz, $CDCl_3$-d_6): δ (ppm) 8.75 (d, J = 8.3 Hz, 2H, CH), 8.43 (d, J = 8.2 Hz, 2H, CH), 8.23 (s, 2H, CH), 4.04 (s, 6H, OCH_3). ^{13}C NMR (101 MHz, $CDCl_3$-d_6): 166.0, 148.1, 145.5, 138.6, 131.0, 129.0, 124.2, 53.2.

Acridine Derivatives

The synthesis and purification of 10-(5-aminopentyl)-3,6-bis(dimethylamino)acridinium iodide (C_5-NH_2), and 10-(8-aminooctyl)-3,6-bis(dimethylamino)acridinium iodide (C_8-NH_2) were performed as described [20].

Oligonucleotides and Ligands

All the oligonucleotide sequences were obtaining from Eurofins Genomics (Munich, Germany) and Eurogentec (Seraing, Belgium) with HPLC-grade purification. The oligonucleotide sequences were double labelled with fluorescein (FAM) and tetramethylrhodamine (TAMRA) at the 5′ and 3′ ends, respectively. The following oligonucleotide sequences were used: c-MYC (5′-TGAGGGTGGGTAGGGTGGGTAA-3′), KRAS21 (5′-AGGGCGGTGTGGGAAGAGGGA-3′), 21G (5′-GGGTTAGGGTTAGGGTTAGGG-3′), VEGF (5′-CGGGGCGGGCCGGGGGCGGGGT-3′) pre-MIR150 (5′-GGCCUGGGGGACAGGGACCUGGG-3′) and DNA duplex (5′-TATAGCTATA-hexaethyleneglycol-TATAGCTATA-3′). Stock solutions were prepared

using Milli-Q water and stored at −20 °C. The oligonucleotide concentrations were measured at 260 nm with a UV–Vis spectrophotometer (Thermo Scientific™ Evolution 220, Thermo Fisher Scientific, Waltham, MA, USA) using the molar extinction coefficients (ε) provided by the manufacturer. The annealing procedure for FmycT, FkrasT, F21T, FvegfT and FdxT consisted of heating to 95 °C for 10 min, followed by cooling in ice for 30 min and for Fpre-MIR150T consisted of heating to 95 °C for 10 min, followed by cooling in ice for more than 10 min. Stock solutions of the compounds were prepared as 10 or 5 mM solutions in dimethyl sulfoxide (DMSO) (Thermo Fisher Scientific, Waltham, MA, USA) and their subsequent dilution was done using Milli-Q water.

In silico Simulations

In silico studies were performed to estimate some molecular properties of the compounds to be tested in vitro. The structures of the different derivatives tested were drawn in ChemDraw 20.0 software (PerkinElmer, Waltham, MA, USA), and the SMILES notation was obtained for all tested compounds. Molecular weight (MW), n-octanol:water partition coefficient (Log *P*), number of rotatable bonds (n-ROTB), number of hydrogen bond acceptors (n-ON acceptors) and number of hydrogen bond donors (n-OHNH donors) were calculated using pKCSM online software [21] and topological polar surface area (TPSA) was determined by Molinspiration property engine v2018.10 on the Molinspiration online server (www.molinspiration.com, accessed on 15 February 2022).

Förster Resonance Energy Transfer (FRET) Melting Assays

FRET melting assays were performed in a CFX Connect™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) equipped with a FAM filter using a 96-well plate. The labelled oligonucleotide sequences were annealed before the experiment as previously described, and the fluorescence intensity was recorded every 1 °C between 25 and 95 °C with a temperature increment of 1 °C min^{-1}. The excitation and detection wavelengths were 492 and 516 nm, respectively. The buffer used in the experiments was 10 mM lithium cacodylate (pH 7.2) supplemented with the necessary amount of KCl and LiCl. The labelled oligonucleotide sequence was used at 0.2 μM and the ligands at 5 or 1 μM final concentration. Each experimental condition was tested in duplicate on two separate assays. The melting temperature was determined from the normalized curves as the mid-transition $T_{1/2}$ temperature (Appendix A, **Figure A2**).

Cytotoxicity in Human Cell Lines

Cell cultures

Non-small human lung adenocarcinoma epithelial cell line (A549) and normal human dermal fibroblasts cell line (NHDF) were purchased at American Type Culture Collection (ATCC), cultured in 75 cm^2 T-flasks and maintained in a humidified atmosphere at 37 °C and 5% CO_2. A549 cells were grown with Ham's-F12 medium supplemented with 10% fetal bovine serum (FBS) and 1% streptomycin-penicillin (SP) antibiotic and NHDF with RPMI medium supplemented with 10% FBS, 1% SP antibiotic, 2 mM L-glutamine, 10 mM HEPES and 1 mM sodium pyruvate.

MTT Assay

The in vitro cell proliferation was determined using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Sigma-Aldrich, St. Louis, MO, USA) assay. The A549 and NHDF cells were plated in 96-well culture plates with density of 1×10^4 cells/mL in the appropriate culture medium. After 48 h of adherence cells were treated with the different compounds in the study for 72 h. At the end of incubation, the medium was removed and replaced with MTT solution and further incubated at 37 °C for 4 h. Then, the MTT solution was removed, and the resulting formazan crystals were dissolved with 100 µL of DMSO. The absorbance was measured at 570 nm using a microplate reader Bio-rad Xmark spectrophotometer (Bio-Rad, California, USA). The cytotoxicity was normalized control conditions (untreated cells). Each experiment was performed in quadruplicate in two independent experiments.

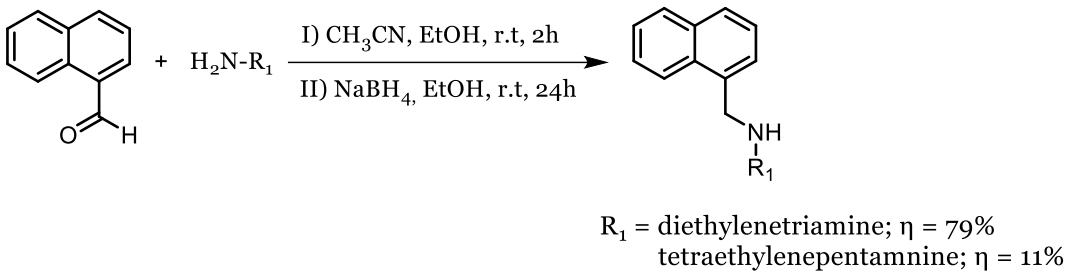
Results and Discussion

Chemistry

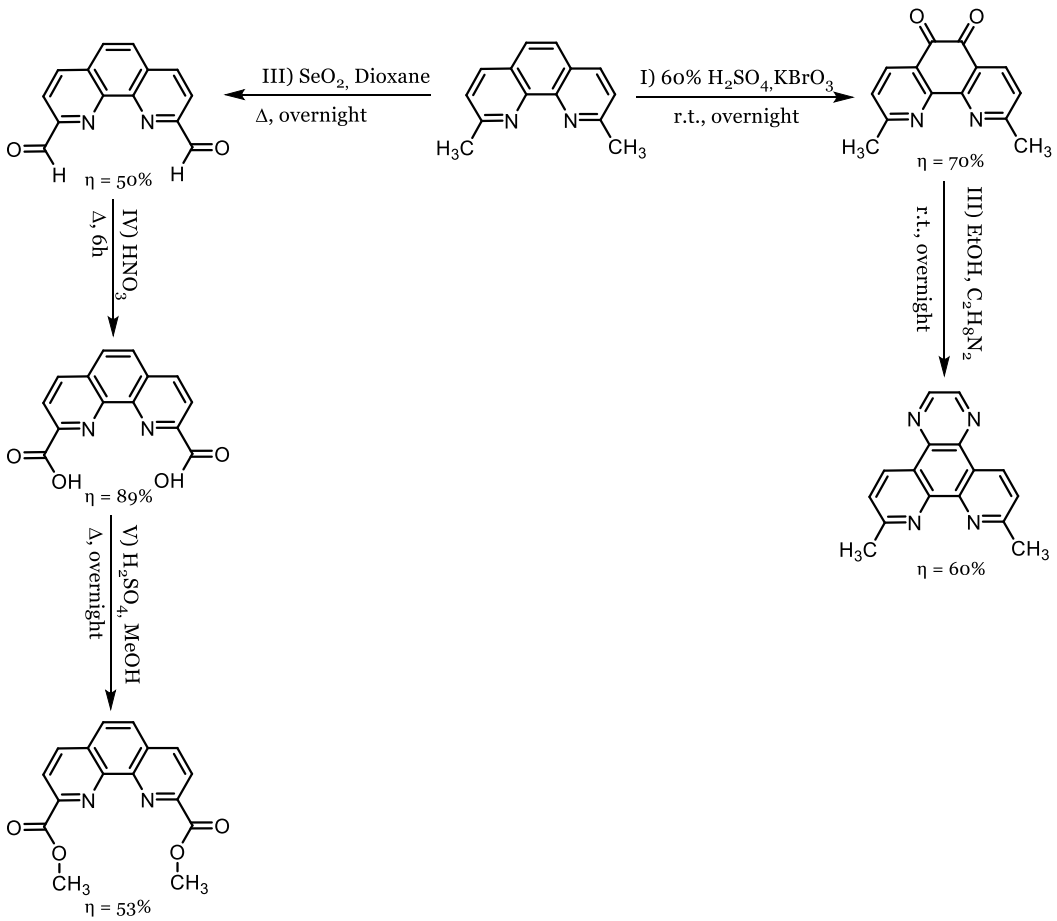
Several small molecules have been reported in the literature as ligands/drugs for the targeting of G4 structures, including new chemotypes [22]. Among the several classes of ligands, synthetic heterocycles have proven to be a promising scaffold for the design of G4 binders and stabilizers, in particular, fused aromatic compounds.

Herein, we describe the chemical synthesis, full characterization and *in vitro* evaluation of four different aromatic scaffolds containing oxazole, acridine, naphthalene and phenanthroline moieties. The naphthalene and phenanthroline derivatives were

synthesized as shown in **Scheme 5.1** and **Scheme 5.2**, with low to high yields, i.e., 11–89%, respectively (experimental section). All starting compounds were commercially obtained and most of these derivatives were easily obtained with a high degree of purity, as confirmed for NMR analysis. Naphthalene derivatives (Napht-3 and Napht-5) were synthetized by following a two-step route, previously reported by Clares et al. [16]. Firstly, a condensation between 1-naphtaldehyde and diethylenetriamine and tetraethylenepentamine, followed by in situ reduction in the formed imine with sodium borohydride, gave derivatives Napht-3 and Napht-5 with 79% and 11% yields, respectively (**Scheme 5.1**). The phenanthroline derivatives were synthetized using procedures already described (**Scheme 5.2**) [17-19]. The synthesis of 7,10-dimethylpyrazino[2,3-*f*][1,10]phenanthroline (DPPZ) was done by initial oxidation of 2,9-dimethyl-1,10-phenanthroline with potassium bromate in sulfuric acid solution, affording 2,9-dimethy-1,l0-phenanthroline-5,6-dione. After that, the dione reacted with ethylenediamine by condensation to give DPPZ with a yield of 60%. The 2,9-dicarbomethoxy-1,10-phenanthroline ($COOCH_3$) was synthetized from the oxidation of 2,9-dimethyl-1,10-phenanthroline to the aldehyde, which was oxidized to a dicarboxylic acid derivative and converted to Phen$COOCH_3$ by Fisher esterification, in a yield of 53%. The phenyl (ISO-1 and Ter-1), quinoline (Quin-1) and acridine derivatives (C_5NH_2 and C_8NH_2) were synthesized as previously described [15,20].



Scheme 5.1. Chemical synthesis of naphthalene derivatives.



Scheme 5.2. Chemical synthesis of phenanthroline derivatives.

In silico Studies

Taking into account that G4 ligands present poor drug-like properties, a computational study using Lipinski's rule and Veber's parameters was performed to predict in vitro permeation and oral absorption [23,24]. The *in silico* parameters are presented in **Table 5.1**. All tested ligands presented MW (210.236–395.615), Log *P* (0.657–4.093), n-ON acceptors (3–6) and n-OHNH donors (0–5), in the range defined by Lipinski's rule [23]. The n-ROTB between 2 and 13 and TPSA from 38.92 to 181.68, revealed that most of the tested scaffolds comply with Veber's parameters (n-ROTB < 10 and TPSA < 10 Å^2) [24]. None of the tested ligands violated Lipinski´s rule of five, indicating that they have favorable properties that achieve the criteria of drug-likeness.

Table 5.1. *In silico* molecular descriptors of *in vitro* evaluated compounds [a].

Ligand	TPSA	MW	Log *P*	n-ROTB	n-ON acceptors	n-OHNH donors	Lipinski's violations[b]
Ter-1	181.68	212.208	2.997	2	4	0	0
ISO-1	52.06	212.208	2.997	2	4	0	0
Quin-1	38.92	210.236	3.198	1	3	0	0
Napht-3	50.08	243.354	1.48	7	3	3	0
Napht-5	74.13	329.492	0.657	13	5	5	0
Phen-1	77.85	314.304	4.093	2	6	0	0
DPPZ	51.57	260.3	3.343	0	4	0	0
$PhenCOOCH_3$	78.39	296.282	2.356	2	6	0	0
C_8NH_2	35.50	395.615	2.3372	10	3	1	0
C_5NH_2	35.50	353.534	1.1669	7	3	1	0

[a]TPSA, topological polar surface area(Å); molecular weight (g/mol); Log *P*, n-octanol:water partition coefficient; n-ROTB; number of rotatable bonds; n-ON acceptors, number of hydrogen bond acceptors; n-OHNH donors, number of hydrogen bond donors.

[b]Lipinski rule of five: MW < 500 Da; Log *P* ≤5; n-OHNH acceptors ≤ 10; n-OHNH donors ≤5. A maximum of 1 violation is generally permitted.

FRET Melting Studies

The ability of the scaffolds to bind and stabilize G4 structures was evaluated by a high-throughput FRET melting assay [25]. This method is commonly used for first screening of G4 binding and stabilization efficiency of a small molecule [19]. The scaffolds were tested against four G4 DNA sequences (c-MYC, KRAS21 and VEGF promoters, and 21G human telomeric repeats) one G4 RNA (pre-MIR150) [26] and a double-stranded DNA control. The ligands were screened at 5 μM, except acridine derivatives, which were tested at 1 μM. The melting temperatures are summarized in **Table 5.2**.

Overall, FRET-melting data showed that screened scaffolds are poor G4 stabilizers, with the notable exception of two acridine precursors, C_8NH_2 and C_5NH_2. This evidence suggests that the alkylamino chain is also important for G4 binding and stabilization beyond the aromatic ring. C_8NH_2 proved to be the most effective binder for all G4 sequences (ΔT_m = 17.4 °C for FmycT, 26.8 °C for F21T, 7.2 °C for FkrasT, 4.5 °C for FvegfT, and 4.6 °C for Fpre-MIR150T at 1 μM concentration), followed by C_5NH_2 (ΔT_m = 16.3 °C for FmycT, 15.6 °C for F21T, 18.0 °C for FkrasT, 2.5 °C for FvegfT, and 0.9 °C for Fpre-MIR150T at 1 μM concentration). The FRET-melting studies corroborate the other series of acridine derivatives, in which the nature of the acridine scaffold and the alkylamide chain influence the G4 stabilization. Previous studies have shown that the acridine moiety interacts with the G4 surface by end stacking interactions and the

alkylamide interacts by groove binding [27]. In addition, acridine derivatives have stronger G4 stabilizing efficiency on pre-miRNA structure [28].

Table 5.2. The thermal stabilization induced by different scaffolds measured by FRET-melting experiments.

		ΔT_m (°C)					
	Ligand	**FmycT**	**F21T**	**FkrasT**	**FvegfT**	**Fpre-MIR150T**	**FdxT**
5μM	Ter-1	0.0 ± 0.2	0.0 ± 0.0	0.0 ± 1.9	0.0 ± 1.6	0.0 ± 0.5	0.0 ± 0.1
	ISO-1	0.0 ± 0.3	0.1 ± 0.0	0.0 ± 1.0	0.5 ± 1.2	0.0 ± 0.0	0.0 ± 0.3
	Quin-1	0.0 ± 1.0	0.2 ± 0.5	0.0 ± 0.9	0.0 ± 2.0	0.0 ± 0.2	0.0 ± 0.0
	Napht-3	0.0 ± 0.1	0.8 ± 0.3	2.0 ± 1.0	0.6 ± 2.1	0.0 ± 0.3	0.0 ± 0.4
	Napht-5	0.0 ± 0.2	0.8 ± 0.1	0.1 ± 1.0	0.0 ± 1.9	0.0 ± 0.3	0.0 ± 0.2
	Phen-1	0.0 ± 0.1	2.5 ± 0.1	0.3 ± 0.6	0.0 ± 2.0	0.0 ± 0.1	0.0 ± 0.2
	DPPZ	0.6 ± 0.3	0.4 ± 0.1	0.6 ± 0.2	0.0 ± 1.4	0.0 ± 0.1	0.0 ± 0.2
	$PhenCOOCH_3$	0.6 ± 0.8	3.4 ± 0.2	3.4 ± 1.8	0.0 ± 2.0	0.0 ± 0.1	0.0 ± 0.3
1μM	C_8NH_2	17.4 ± 0.4	26.8 ± 0.3	7.2 ± 0.4	4.5 ± 0.5	4.6 ± 0.0	0.1 ± 0.1
	C_5NH_2	16.3 ± 0.2	15.6 ± 0.2	18.0 ±2.6	2.5 ± 0.9	0.9 ± 1.3	0.3 ± 0.2

ΔT_m represents the difference in melting temperature [$\Delta T_m = T_m$(G4 sequence + ligand) – T_m(G4 sequence)]. The buffer used was 10 mM lithium cacodylate, pH 7.2. The T_m values for the DNAs are 65.3 ± 0.1 °C [c-MYC promoter DNA in 10 mM KCl and 90 mM LiCl], 52.1 ± 0.2 °C [21G telomeric DNA in 10 mM KCl and 90 mM LiCl], 41.4 ± 2.1 °C [KRAS promoter DNA in 10 mM KCl and 90 mM LiCl], 66.9 ± 2.0 °C [VEGF promoter DNA in 1 mM KCl and 99 mM LiCl], for the RNA is 68.5 ± 0.2 °C [pre-MIR150 in 1 mM KCl and 99 mM LiCl] and 65.3 ± 0.0 °C [duplex DNA in 100 mM KCl]. All experiments were done in duplicate, and the values reported correspond to the average of two measurements with the estimated standard deviation.
Negative ΔT_m values are reported as zero.

The phenyl and quinoline scaffolds displayed null stabilization of the DNA/RNA G4 structures studied, when compared with phenanthroline derivatives. The phenanthroline analogue Phen-2, described previously, which has a π system extension, showed an increase in melting temperature of FmycT and F21T, when compared with its precursor, Phen-1. Compound Phen-1 exhibited a minimal increase in T_m of 2.5 °C to F21T and no stabilization of FmycT even at 5 μM, while the Phen-2 showed a moderate increase in T_m of 8.6 °C for F21T and 11.3 °C for FmycT at the same concentration [15]. The functionalization with ester group in the phenanthroline moiety, when compared with oxazole functionalization, improved the G4 binding capacity against F21T and FkrasT sequences. On the contrary, the introduction of an additional pyrazine ring in phenanthroline moiety reduced stabilization effect on all G4s. Compounds containing naphthalene scaffolds presented poor G4 stabilization, which was slightly better for FkrasT, in the case of derivatives containing a short polyamine chain linked directly to

the ring (Napht-3). All tested scaffolds exhibited negligible stabilization for FdxT (**Table 5.2**).

Cytotoxicity Assay

The scaffold's effect on cell viability was evaluated by an MTT assay, using A549 lung cancer and NHDF healthy cell lines. The results of IC_{50} are presented in **Table 5.3**. Overall, the scaffolds were found to be much more active on the cancer cell line. The acridine precursors showed a high potency against the A549 cell line, with IC_{50} values of 1.3 μM (C_8NH_2) and 3.6 μM (C_5NH_2). Most relevantly, C_5NH_2 shows an IC_{50} of 17.8 μM in normal human fibroblasts, indicating a good selectivity for its cellular growth inhibitory activity in human cancer cells over human normal cells. The anticancer activity of C_8NH_2 is consistent with that previously reported for PC-3 (prostate adenocarcinoma) and PNT1A (normal human prostate cell line) [28]. These results can be correlated with the FRET-melting results, which may indicate that the cytotoxicity effect of the acridine ligands can be related to G4 binding ability. On the contrary, the phenanthroline scaffold containing a pyrazine ring (DPPZ) showed lower IC_{50} values in both cell lines (IC_{50} of 0.4 μM and 0.2 μM for cancer and normal cell line, respectively), suggesting lower cellular selectivity. Most probably, this effect is not mediated by G4 binding, since DPPZ showed weak G4 stabilizer capacity in FRET-melting experiments. The phenanthroline derivative containing an oxazole ring (Phen-1) presented low and no-selective cytotoxicity effect in cancer (IC_{50} = 22.1 μM) and healthy (IC_{50} = 20.4 μM) cell lines. Conversely, the introduction of an ester group in the phenanthroline scaffold conferred potency against A549 cells (IC_{50} = 6.4 μM) and a weak inhibitory effect on healthy cells (IC_{50} = 37.0 μM). Therefore, Phen$COOCH_3$ can be a promising candidate for further cellular studies. Similarly, phenyl (Ter-1 and ISO-1) derivatives were shown to be promising anticancer candidates, particularly Ter-1, with a pronounced cytotoxicity effect in cancer cells (IC_{50} = 1.9 μM) and weak toxicity in healthy cells (IC_{50} = 34.0 μM). These results suggest that the 1,4 disubstituted phenyl ring confers higher anticancer activity, when compared with the 1,3 disubstituted congener (ISO-1; with an IC_{50} value of 9.8 and 32.0 μM for cancer and healthy cells lines, respectively). The Quinoline derivative also presents selectivity and cytotoxicity effects (IC_{50} = 9.0 μM) in cancer cells. Naphthalene scaffolds exhibited a low inhibitory effect on the cancer cell line, but Napht-3 presented low IC_{50} on the fibroblast cell line (NHDF). This trend can be related with the size of the polyamine chain, since Napht-3 has a short polyamine chain and is more toxic in normal cells (IC_{50} = 5.3 μM) than the derivative containing the pentamine chain

(Napht-5; IC_{50} = 24.1 μM). For cancer cells, Napht-5 is more potent (IC_{50} = 12.5 μM) than Napht-3 (IC_{50} = 19.4 μM).

Table 5.3. IC_{50} values induced by different scaffolds in A549 and NHDF cell lines.

Ligand	IC_{50} (μM)		Selectivity factor
	A549	NHDF	
Ter-1	1.9	34.0	17.9
ISO-1	9.8	32.0	3.3
Quin-1	9.0	32.6	3.6
Napht-3	19.4	5.3	0.3
Napht-5	12.1	24.1	2.0
Phen-1	22.0	20.7	0.9
DPPZ	0.40	0.21	0.5
$PhenCOOCH_3$	6.4	37.0	5.8
C_8NH_2	1.3	3.8	2.9
C_5NH_2	3.6	17.8	4.9

IC_{50} values were determined in two separate experiments independently with a 95% confidence interval.

Conclusions

Different scaffolds of phenyl, quinoline, naphthalene, phenanthroline and acridine types have been investigated for their G4 binding potential. Four of these scaffolds (Napht-3, Napht-5, DPPZ and $PhenCOOCH_3$) were synthetized and screened for the first time against a G4 structure. Concerning their synthesis, the derivatives were obtained with moderate to good yields. The FRET-melting experiments showed that the acridine compounds are the most promising scaffolds and, most importantly, C_8NH_2 can discriminate among different G4 structures. Overall, the different molecules did not present a high pronounced toxic effect in normal cell lines, in contrast to cancers cells. Additionally, these scaffolds hit the drug-like criteria. This study could lead to the identification of G4-interactive molecules, deserving further optimization of their structures towards the development of drug candidates.

Funding

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, to CICS-UBI UIDB/00709/2020 research unit, project UID/Multi/04349/2019, 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. J.F. acknowledges the doctoral fellowship grant from FCT ref. SFRH/BD/145106/2019. C.C. acknowledges the grant from FCT ref. UIDP/00709/2020. J.-L.M. acknowledges support from INCa PL-Bio (G4Access) and ANR-20-CE12-0023 grants and SYMBIT project [CZ.02.1.01/0.0/0.0/15 003/0000477] financed by the ERDF.

References

[1] Figueiredo, J. *et al.* (2021) Ligands as Stabilizers of G-Quadruplexes in Non-Coding RNAs. *Molecules* 26, 6164.

[2] Abiri, A. *et al.* (2021) Unlocking G-Quadruplexes as Antiviral Targets. *Pharmacol Rev* 73, 897–923.

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

[4] Wheelhouse, R.T. *et al.* (1998) Cationic porphyrins as telomerase inhibitors: The interaction of tetra- (N-methyl-4-pyridyl)porphine with quadruplex DNA. *J Am Chem Soc* 120, 3261–3262.

[5] Guliaev, A.B. and Leontis, N.B. (1999) Cationic 5,10,15,20-Tetrakis (N -methylpyridinium-4-yl)porphyrin Fully Intercalates at 5‘-CG-3‘ Steps of Duplex DNA in Solution. *Biochemistry* 38, 15425–15437.

[6] Lee, Y.-A. *et al.* (2003) Binding of meso -Tetrakis (N -methylpyridium-4-yl)porphyrin to Triplex Oligonucleotides: Evidence for the Porphyrin Stacking in the Major Groove. *J Am Chem Soc* 125, 8106–8107.

[7] Kim, M.Y. *et al.* (2002) Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J Am Chem Soc* 124, 2098–2099.

[8] Franceschin, M. *et al.* (2006) Natural and synthetic G-quadruplex interactive berberine derivatives. *Bioorg Med Chem Lett* 16, 1707–1711.

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

[10] 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.

[11] Drygin, D. *et al.* (2011) Targeting RNA Polymerase I with an Oral Small Molecule CX-5461 Inhibits Ribosomal RNA Synthesis and Solid Tumor Growth. *Cancer Res* 71, 1418–1430.

[12] Das, R. *et al.* (2017) Design, Synthesis and Biological Evaluation of New Substituted Diquinolinyl-Pyridine Ligands as Anticancer Agents by Targeting G-Quadruplex. *Molecules* 23, 81.

[13] Hamon, F. *et al.* (2011) An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies. *Angew Chem Int Ed* 50, 8745–8749.

[14] De Rache, A. and Mergny, J.-L. (2015) Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* 115, 194–202.

[15] Medeiros-Silva, J. *et al.* (2017) Phenanthroline-bis-oxazole ligands for binding and stabilization of G-quadruplexes. *Biochim Biophys Acta Gen Subj* 1861, 1281–1292.

[16] Clares, M.P. *et al.* (2004) Synthesis and H^+, Cu^{2+}, and Zn^{2+} Coordination Behavior of a Bis(fluorophoric) Bibrachial Lariat Aza-Crown. *Inorg Chem* 43,6114-6122.

[17] Zheng, R.H. *et al.* (2010) A new and convenient synthesis of phendiones oxidated by KBrO3/H2SO4 at room temperature. *Chin Chem Lett* 21, 1270–1272.

[18] Garas, A.M.S. and Vagg, R.S. (2000) Synthesis of some novel derivatives of 1,10-phenanthroline. *J Heterocycl Chem* 37, 151–158.

[19] R, S.K. *et al.* (2018) Development of the Smartphone-Assisted Colorimetric Detection of Thorium by Using New Schiff's Base and Its Applications to Real Time Samples. *Inorg Chem* 57, 15270–15279.

[20] 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.

[21] Pires, D.E.V. *et al.* (2015) pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem* 58, 4066–4072.

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

[23] Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 44, 235–249.

[24] Veber, D.F. *et al.* (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45, 2615–2623.

[25] De Cian, A. *et al.* (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods* 42, 183–195.

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

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

[28] Santos, T. *et al.* (2021) Recognition of nucleolin through interaction with RNA G-quadruplex. *Biochem Pharmacol* 189, 114208.

6th Chapter

Synthesis and evaluation of 2,9-disubstituted-1,10-phenanthroline derivatives as G-quadruplex binders

This work was published in:
Figueiredo, J.; *Carreira-Barral, I.; Quesada, R.; Mergny, J-L. and Cruz, C. (2022) Synthesis and evaluation of 2,9-disubstituted-1,10-phenanthroline derivatives as G-quadruplex binders. Bioorganic & Medicinal Chemistry 73, 116971. DOI: 10.1016/j.bmc.2022.116971.*

Chapter overview

After the successful screening of different scaffolds of G4 ligands, we move on to the structural modification of the 1,10-phenanthroline scaffold. Indeed, phenanthroline-based derivatives have been characterized as potent and selective G4 ligands. Their structural flexibility allows for the introduction of several functional groups, substituents, and stereochemical modifications enabling the development of structurally diverse libraries. This chapter describes the chemical synthesis of a library of 2,9-disubstituted-1,10-phenanthroline derivatives substituted with flexible amine and amide side chains at 2,9-positions. The rationale behind this strategy derived from the potential enhancement of G4-binding properties through the introduction of amine side chains into the phenanthroline core, thereby increasing electrostatic interactions with the phosphate backbone of nucleic acids. The presence of these cationic chains also enhances the water solubility of the compounds. Biophysical techniques were employed to characterize G4-stabilizer properties of the ligands towards three different G4 topologies derived from the human telomeric sequence. Moreover, their antiproliferative activity was assessed in LC and healthy cell lines. The findings presented and discussed in this chapter allowed us to understand the G4-stabilizing properties and anti-tumor effects induced by these chemical structural modifications.

Synthesis and evaluation of 2,9-disubstituted-1,10-phenanthroline derivatives as G-quadruplex binders

Abstract

G-quadruplex (G4) structures are non-canonical DNA/RNA secondary structures able to form within guanine rich nucleic acids sequences. They are present in several regions of the human genome including gene promoters, untranslated sequences, and telomeres. Due to their biological relevance G4 structures are considered important drug targets, in particular for anticancer therapies, leading to the development of G4 stabilizing small molecules. Telomeric regions have received special attention in this field since they can fold into several distinct intramolecular G-quadruplexes topologies. Herein, we report the synthesis of 2,9-disubstituted-1,10-phenanthroline derivatives and their ability to stabilize different intramolecular telomeric G4 sequences. We evaluated ligand-induced stabilization, selectivity and specificity of ligands using Förster Resonance Energy Transfer (FRET) melting experiments and circular dichroism (CD). In addition, we assessed the cytotoxicity of ligands against two cancer cell lines (A549 and H1299) and one healthy cell line (NHDF).

Keywords

2,9-disubstituted-1,10-phenanthroline derivatives; Biophysical evaluation; G-quadruplex structures; In vitro studies.

Introduction

G-quadruplex (G4) are non-canonical DNA or RNA secondary structures formed by the stacking of two or more square arrangements of guanines (G-quartet) [1,2]. Each G-quartet is formed by four guanine bases interacting by Hoogsteen hydrogen bonds and their formation depends on the presence of a central metal cation (most often monovalent cation, such as Na^+ or K^+) [3]. G4 structures can be classified in terms of

stoichiometry (uni-, bi-, tri- and tetramolecular) and exhibit a great variety of topologies, which are a consequence of the DNA/RNA strand orientation (parallel, antiparallel and hybrid), as well as variations in loop size, sequence and length, the presence of small molecules and annealing conditions [4-6].

G4-forming sequences are present in different species and are non-randomly distributed in the human genome within telomeres, oncogenic promoters and RNA 5′-untranslated regions (5′-UTR) [7,8]. Biologically, G4 structures can regulate key cellular processes such as telomere maintenance, transcription, translation, replication, and genomic instability, playing an important role in carcinogenesis and tumor development [9,10]. Due to their biological relevance, G4 are considered potential therapeutic targets, particularly in anticancer therapy [11]. In this sense, small molecules that can bind and stabilize G4 structures could be considered as anticancer agents [12]. Usually, G4 binders can interact *via* π-π stacking and electrostatic interactions with the G-quartet. To be efficient, they should have a heteroaromatic planar chromophore and one or more flexible substituents positively charged [13]. On the basis of these principles, several classes of small molecules belonging to different chemical families have been developed as selective G4 binding ligands which include fused heteroaromatic polycyclic compounds [14,15] such as acridines [16,17] quindolines [18,19] and naphthalenediimides [20,21]. Another important heteroaromatic scaffold that has attracted interest as a G4 binding ligand is 1,10-phenanthroline [22]. Phenanthroline-based ligands present an extended aromatic surface that provides efficient π-π stacking interactions with a terminal G-quartet [23]. Based on that, phenanthroline derivatives have been reported as G4 stabilizing compounds both in the form of small organic molecules [24] or coordination complexes [25,26]. Several 1,10-phenanthroline derivatives have been studied, including those containing 4,7-diamino positively charged substituents bearing in side chains which confer appreciable G4 stabilizing properties and high water solubility but low selectivity over duplex DNA [27]. Novel tetrasubstituted phenanthrolines containing 4,7-diamino substituents and 2,9-carboxamide-linked quinolines were also investigated and have proven to be potent G4 binders with high selectivity [28]. Another notable group of compounds belonging to the 1,10-phenanthroline class displays a 2,9-substitution pattern. Among them, bisquinolinium amides are the most potent and selective G4 binders, PhenDC3 being the most studied derivative [29]. Besides that, other 2,9-disubstituted derivatives including diaminobenzyl moieties have been developed as potent G4 binding ligands through the introduction of additional aromatic surfaces and amine side chains [30,31].
Here, we describe the synthesis of new 1,10-phenanthroline derivatives **1a-e, 2a-d, 3a, 4a-d and 5a-b**, substituted at positions 2 and 9. All these derivatives were evaluated to

determine ligand-induced stabilization and selectivity for various G4 sequences and topologies through FRET melting experiments. We further evaluated the interaction mode of the most potent derivatives by circular dichroism (CD) spectroscopy. Finally, we tested the antiproliferative activity against two different non-small lung cancer cell lines (A549 and H1299) and one non-malignant cell line (NHDF).

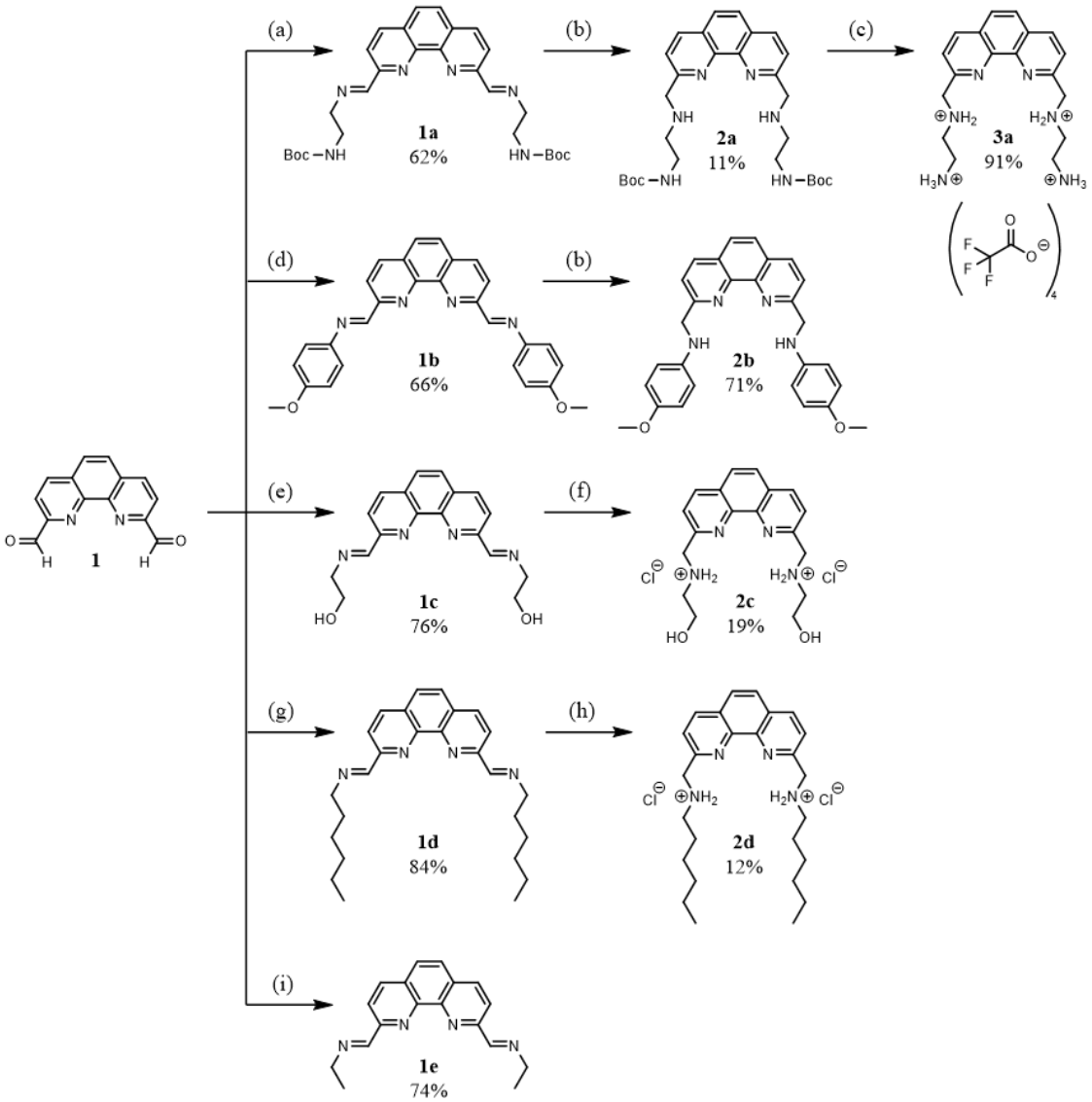
Results and discussion

Chemistry

Considering the evidence that the extended planar aromatic surface of 1,10-phenanthroline scaffolds confers promising properties for the efficient π-π stacking interactions between a G4 ligand and a G-quartet, we hypothesized that the di-substitution at positions 2 and 9 with amine or amide chains could increase the affinity of the neutral phenanthroline core through the introduction of additional interactions with the negatively charged phosphate groups present in the loops and grooves of the G4.

The 2,9-disubstituted-1,10-phenanthroline derivatives **1a-e**, **2a-d** and **3a** were obtained from the 2,9-dicarbaldehyde-1,10-phenanthroline (**1**) previously synthetized by us [22]. The imines **1a-e** were easily prepared by a direct condensation between the aldehyde 1 and primary amines under reflux in chloroform and methanol mixtures (1:1, V/V) (**Scheme 6.1**). The ligand 1a was prepared with a yield of 62 % by condensation between aldehyde 1 and tert-butyl (2-aminoethyl)carbamate, which was synthesized as reported in the literature [32]. For the synthesis of the ligands 1b-e commercial primary amines were used; 4-methoxyaniline to prepare 1b, 2-aminoethanol for 1c, n-hexylamine for 1d, and ethylamine for **1e**. These ligands were obtained with 66, 76, 84, and 74 % yields, respectively. The ligands **1c** and **1e** were synthesized according to procedures described in the literature, however, it should be noted that these compounds were not isolated by the authors [33,34]. After that, the compounds **1a-d** were reduced to the corresponding amines **2a-d** using sodium borohydride as reductive agent in refluxing methanol (**Scheme 6.1**). The compounds **2a-b** were obtained in 11 and 71 % yields, respectively. The compounds **2c-d** were obtained as hydrochloride salts with 19 and 12 % yields, respectively, using concentrated hydrochloric acid in methanol. Solubility issues of the hydrochloride salts resulted in low yields for these compounds. The compound **2c** was obtained according to the procedure described in the literature for the non-protonated

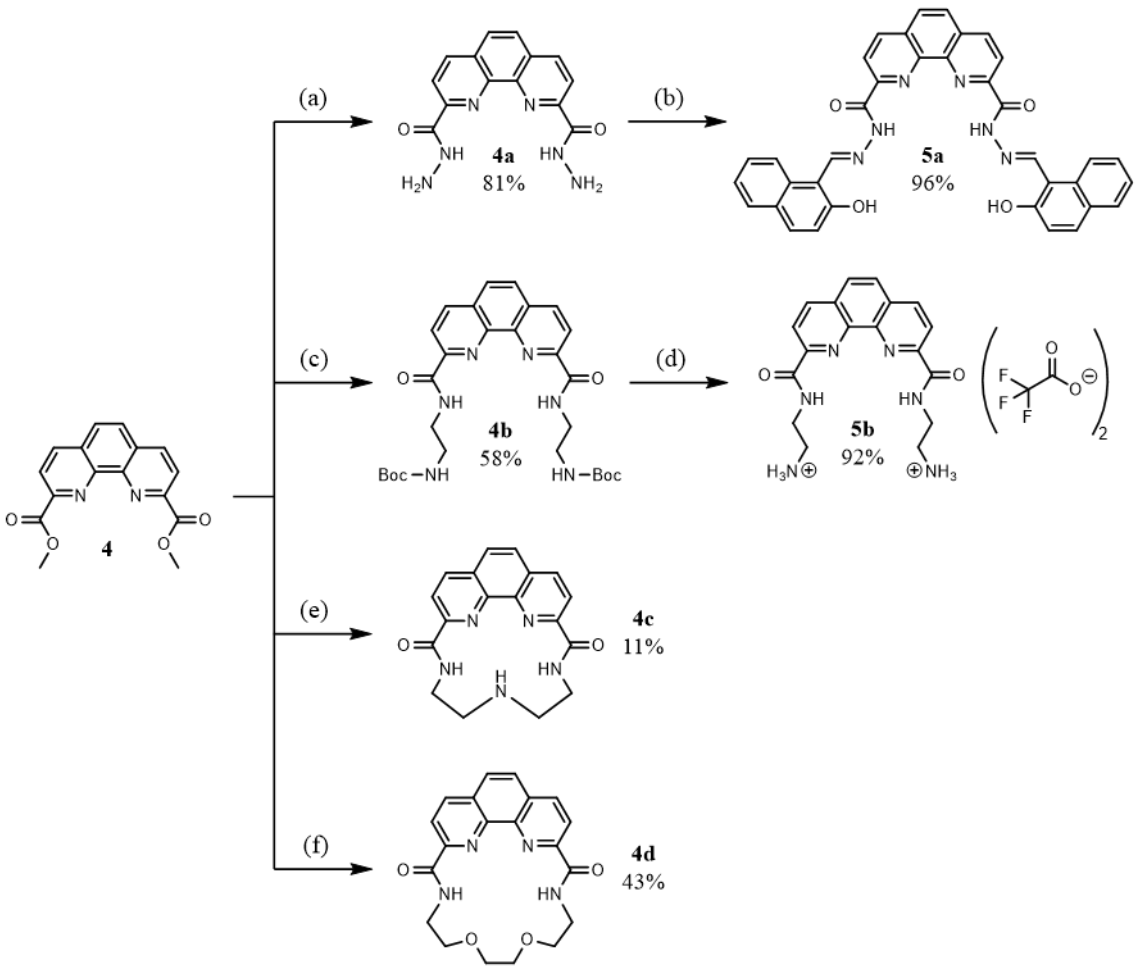
compound [33]. Additionally, the ligand **2a** was converted into the trifluoroacetate salt with 91 % yield by treatment with trifluoroacetic acid in dichloromethane.



Scheme 6.1. Synthesis of compounds **1a-e**, **2a-d** and **3a**. Reaction conditions: **(a)** *tert*-butyl (2-aminoethyl)carbamate, 1:1 (*V*/*V*) $CHCl_3$:MeOH, Δ, 2 h; **(b)** $NaBH_4$, MeOH, Δ, 2.5 h; **(c)** 1:1 (*V*/*V*) TFA:DCM, rt, 24 h; **(d)** 4-methoxyaniline, 1:1 (*V*/*V*) $CHCl_3$:MeOH, Δ, 2.5 h; **(e)** 2-aminoethanol, 1:1 (*V*/*V*) $CHCl_3$:MeOH, Δ, 2.5 h; **(f)** (1) $NaBH_4$, MeOH, Δ, 2.5 h; (2) HCl(aq), MeOH; **(g)** *n*-hexylamine, 1:1 (*V*/*V*) $CHCl_3$:MeOH, Δ, 3 h; **(h)** (1) $NaBH_4$, MeOH, Δ, 3 h; (2) HCl(aq) 1 M; (i) ethylamine, 1:1 (*V*/*V*) $CHCl_3$:MeOH, Δ, 16 h. Yields are indicated below each compound number.

The same strategy was used for the synthesis of 2,9-disubstituted-1,10-phenanthroline derivatives **4a-d.** These compounds were obtained by reacting 2,9-dicarbomethoxy-1,10-phenanthroline (**4**) [22] and different amines (**Scheme 6.2**). The hydrazide **4a** was obtained using hydrazine hydrate by reflux in ethanol and subsequently it was converted into the compound **5a** by condensation with 2-hydroxy-1-naphthaldehyde according to

the procedures previously reported in the literature [35]. The amide **4b** was obtained with 58 % yield by reaction of compound **4** with *tert*-butyl (2-aminoethyl)carbamate [32] in a chloroform and methanol mixture (1:1; *V*/*V*) at room temperature. After that, **4b** was converted into the amide **5b** by treatment with trifluoroacetic acid in dichloromethane at room temperature with 92 % yield. The macrocycles **4c-d** were synthesized under the same reaction conditions utilized for the compound **4b** using bis(2-aminoethyl)amine to prepare **4c** and 1,2-bis(2-aminoethoxy)ethane for **4d.** These compounds were obtained with 11 and 43 % yields, respectively. Full characterization data of all synthesized compounds are provided in the Appendix B (**Figure B1-B51**).



Scheme 6.2. Synthesis of compounds **4a-d** and **5a-b**. Reaction conditions: **(a)** H_2N-NH_2·H_2O, EtOH, Δ, 3 h; **(b)** 2-hydroxy-1-naphthaldehyde, H_2SO_4, EtOH, Δ, overnight; **(c)** *tert*-butyl (2-aminoethyl)carbamate, 1:1 (*V*/*V*) $CHCl_3$:MeOH, rt, 24 h; **(d)** 1:1 (*V*/*V*) TFA:DCM, rt, 1.5 h; **(e)** bis(2-aminoethyl)amine, 1:1 (*V*/*V*) $CHCl_3$:MeOH, rt, 24 h; **(f)** 1,2-bis(2-aminoethoxy)ethane, 1:1 (*V*/*V*) $CHCl_3$:EtOH, rt, 96 h. Yields are indicated below each code. Compounds **4a** and **5a** were prepared as reported in the literature [35].

FRET melting assay

The ability of the synthetized phenanthroline derivatives **1a-e**, **2a-d**, **3a**, **4a-d** and **5a-b** to bind and stabilize G4 structures was evaluated through high throughput FRET melting assay [36] using doubly labelled oligonucleotide sequences. To evaluate the specificity against a particular G4 sequence, the FRET melting experiments were carried out using G-rich oligonucleotides (**Table B1**) that are able to adopt various G4 topologies including antiparallel G4, F21CTAT, a mutant telomeric sequence derived from the DNA human telomeric sequence, and parallel G4s such as F25cebT, derived from the human minisatellite sequence. We also used F21T oligonucleotide, derived from the human telomere sequence that adopts a 3 + 1 hybrid-type topology in K^+ conditions. Then, to elucidate the selectivity of ligands over double-stranded DNA, FRET melting experiments were performed using a duplex control sequence (FdxT). The ligand-induced change in melting temperature of G4 (ΔT_m) was calculated by the difference between the melting temperature (T_m) of G4 sequences in the presence and absence of each ligand. The ΔT_m values are presented in **Table 6.1**. The G4 ligand PhenDC3 was used as positive control.

Table 6.1. Thermal stabilization for the compounds **1a-e**, **2a-d**, **3a**, **4a-d** and **5a-b**, measured by FRET melting.

	ΔT_m (°C)[a] at 5 μM ligand concentration			
Ligand	**F21T**	**F21CTAT**	**F25cebT**	**FdxT**
1a	0.6 ± 0.1	NS	0.6 ± 0.6	NS
1b	0.8 ± 0.4	NS	0.7 ± 0.3	NS
1c	1.1 ± 0.2	NS	0.9 ± 0.5	NS
1d	1.1 ± 0.1	NS	0.7 ± 0.2	NS
1e	1.5 ± 0.4	NS	1.2 ± 0.5	NS
2a	1.2 ± 0.3	NS	0.9 ± 0.2	NS
2b	9.2 ± 1.6	4.1 ± 1.4	3.8 ± 0.3	NS
2c	1.2 ± 0.3	NS	0.4 ± 0.5	NS
2d	5.3 ± 0.5	NS	NS	NS
3a	12.0 ± 0.4	5.2 ± 1.8	7.8 ± 1.0	NS
4a	1.8 ± 0.4	NS	1.2 ± 0.3	NS
4b	0.7 ± 0.1	NS	0.8 ± 0.5	NS
4c	0.4 ± 0.6	NS	1.7 ± 0.2	NS
4d	0.5 ± 0.4	NS	0.4 ± 0.4	NS
5a	1.0 ± 0.8	NS	2.6 ± 0.2	NS
5b	7.2 ± 0.5	2.9 ± 0.4	1.8 ± 0.4	NS
PhenDC3	> 30.0	> 30.0	> 30.0	0.6 ± 0.1

[a]ΔT_m values of F21T, F21CTAT, F25CebT and FdxT (0.2 μM) were performed in buffer containing 10 mM lithium cacodylate (pH 7.2), 10 mM KCl and 90 mM LiCl. T_m values for the DNAs are 53.4 ± 0.2 (F21T), 59.7 ± 0.2 (F21CTAT), 65.7 ± 0.2 (F25cebT) and 68.7 ± 0.0 (FdxT).

Negative ΔT_m values are reported as NS (no stabilization).

Error margins correspond to SD of three replicates.

For the 2,9-disubstituted-1,10-phenanthroline derivatives **1a-e**, **2a-d**, **3a**, **4a-d** and **5a-b,** the ΔT_m values ranged from 0.4 °C to 12.0 °C at 5 μM ligand concentration. As observed in **Table 6.1**, the ligands were found to be more selective for the hybrid G4 topology of the F21T sequence and for the parallel topology of the F25cebT than for the antiparallel topology of the F21CTAT sequence. Ligands **2b**, **3a** and **5b** stabilized all G4 sequences. Among the tested compounds, ligand **3a** is the best G4 stabilizer for all sequences with ΔT_m values of 12.0, 5.2 and 7.8 °C for F21T, FCTAT and F25cebT, respectively, consistent with the highest number of positive charges. The ligand **5b** also stabilizes the G4, in particular sequence F21T with ΔT_m value of 7.2 °C. These results suggest that ligands with an ethane-diaminium on their 2,9-substituents exhibit better activity. These ligands (**3a** and **5b**) are also more potent than the parent **1a**, **2a** and **4b** with a terminal carbamate group. It is important to note that the stabilization values vary between ligands **3a** and **5b**, showing a greater enhancement of the thermal stability of G4 for the amine (**3a**) compared to amide (**5b**). These results suggest that electrostatic interactions between the positive charges of polyamine side chains and the negative charges of phosphate groups increase the thermal stabilization of the G4.

Ligand **2b** with methoxyaniline moieties in the lateral chains is also able to stabilize all G4 sequences, in particular F21T with ΔT_m of 9.2 °C. These results suggest that the presence of an additional aromatic group could lead to improvement of G4 stabilization properties. Again, ligand **2b** is significantly more potent than **1b**, which indicates that amine groups improve the potential of G4 stabilization when compared with the imine groups (**1b**). Interestingly, ligand **2d** with 2,9-substituents with hexan-aminium chains appears to stabilize selectively a hybrid-type topology of F21T sequence with ΔT_m of 5.3 °C. The macrocycles **4c-d** do not present significant G4 binder capacity, however ligand **4c** shows residual G4 stabilization (ΔT_m = 1.7 °C). Similar phenanthroline derivatives have been described in the literature, however **4c-d** present lower G4 stabilization properties than [16]phenN_4 for the F21T sequence [37]. Finally, none of these ligands are able to stabilize the duplex sequence FdxT.

Circular dichroism (CD) studies

After the initial ligand screening by FRET melting assay, CD spectroscopy was employed to determine the conformation of the G4 in the presence of the most promising ligands (**3a, 2b** and **5b**) in terms of thermal stabilization. CD spectra were acquired to characterize the conformations of pre-folded G4 sequence in presence of 5 M equivalents of each ligand (**Figures. B51-B53**) [38]. The unlabelled oligonucleotide sequences used were 22AG, 22CTA and 26ceb (**Table B1**). For the 22AG sequence, upon titration with

3a and **2b** no significant changes in the CD spectra are observed, maintaining the typical antiparallel bands with a large positive band at 295 nm and a negative band at 265 nm. Conversely, upon titration of the ligand **5b**, an increase of ellipticity around 260 nm at 3 and 5 molar equivalents was observed, suggesting the formation of the hybrid topology [39]. Upon addition of 5 molar equivalents of each ligand to 22CTA and 26ceb sequences, no relevant changes were found in the ellipticity and topology. These results suggest an overall conservation of the G4 structures.

After that, we also determined T_m of the G4 sequences by CD in presence of different concentrations of the ligands **2b**, **3a** and **5b** (**Table 6.2**). The thermal denaturation was monitored at the wavelength of maximum CD ellipticity.

Table 6.2. Thermal stabilization induced by **2b**, **3a** and **5b** in the DNA sequences 22CTA, 26ceb and 22AG measured by CD melting experiments.

Ligand	Ligand concentration (molar equivalents)	ΔT_m (°C)[a]		
		22AG	22CTA	26ceb
2b	0.5	0.5	0.9	1.1
	1.0	0.9	1.8	2.3
	3.0	1.5	2.6	5.9
	5.0	1.1	3.1	8.1
3a	0.5	0.7	0.6	0.9
	1.0	2.1	1.3	2.1
	3.0	3.4	1.9	5.4
	5.0	3.5	2.1	7.6
5b	0.5	0.5	0.4	3.4
	1.0	0.3	1.3	4.6
	3.0	2.5	2.9	9.7
	5.0	1.4	4.3	11.7

[a] ΔT_m represents difference in melting temperature [$\Delta T_m = T_m$ (DNA sequence + ligand) – T_m (DNA)]. The T_m values for DNA alone are 66.0 ± 0.3°C (22AG), 62.0 ± 0.2°C (22CTA) and 78.1 ± 0.1 °C (26ceb). The experiments were performed in buffer containing 10 mM lithium cacodylate (pH 7.2) and 100 Mm KCl.

Overall, the CD melting experiments showed a greater stabilization when compared with FRET melting experiments. These differences can be explained by several factors such as the use of the doubly labelled oligonucleotides sequences, the difference in oligonucleotide ligand concentrations and potassium concentrations in FRET melting experiments [30]. As observed in **Table 6.2**, the analyzed ligands showed a preference for parallel topology, namely for the 26ceb sequence, as evidenced by the highest thermal stabilization at 5 molar equivalents. The best ligand for this sequence was **5b** followed by **2b** and **3a**. The same stabilization tendency was verified for 2CTA antiparallel G4. Finally, for 22AG the ligand **3a** proved to be the best stabilizer followed by **5b** and **2b**; the same stabilization tendency was verified for 22CTA antiparallel G4. In the specific case of the ligand **5b**, the variation between 3 and 5 molar equivalents can be explained

by a rearrangement of the secondary structure, as shown in CD spectra (**Figures B53, B54 and B55**), where a new band around 260 nm appears.

Cellular studies

The effect of the ligands on cell viability was evaluated against two human lung cancer cell lines (A549 and H1299) and human dermal fibroblasts (NHDF) by the MTT assay. For this, a screening was performed at 10 μM concentration of each ligand. The results of the cytotoxicity assays are presented in **Figure 6.1** and **Table B2**. Overall, the 2,9-disubstituted-1,10-phenanthroline derivatives synthetized display moderate cytotoxicity against cancer cell lines. The amides **4a-d** and **5b** were shown to be more cytotoxic against A549 cells than H1299 and less cytotoxic towards non-malignant cells. Ligand **5a** displayed more cytotoxicity for H1299 cell line than for A549, conversely to the ligand **4a** that presented more cytotoxicity for A549 than H1299 cells. This evidence indicates that the introduction of 2-hydroxynaphthyl groups confers selectivity for H1299 cells and does not increase the toxicity to non-malignant cells. Ligand **5b** exhibits higher cytotoxicity towards A549 cancer cells (24.5 % of cell viability) and nearly no toxicity towards non-malignant cells (75.1 % cell viability). Interestingly, this ligand has been identified as one of the most potent G4 binder analyzed by FRET melting assay. This result can be related with a better cellular uptake for this compound or due to G4 interaction. Ligand **5b**, containing amide groups in lateral chains, also proved to be more potent against cancer cells than the parent ligand **3a** that bears amine groups in the same positions. Ligands **3a** and **2b**, which were identified as the best G4 binders through the FRET melting analysis, do not have significant effects on the cell viability of both cancer and non-malignant cells. Conversely, the ligand **2b** proved to be more toxic in non-malignant than cancer cells. These results suggest that the cytotoxicity of these derivatives (**2b** and **3a**) may not be related with their G4 binder ability. Ligand **2d** displayed the highest cytotoxicity against both cancer cell lines with 27.4 and 42.9 % of cell viability for A549 and H1299, respectively. However, it also showed high toxicity towards non-malignant cells (22.4 %). This ligand also has moderate stabilization (ΔT_m (5.3 °C)) towards G4 F21T. Additionally, it is more effective than its congener **1d**, which can indicate that the presence of hexylaminium residues in lateral chains confers more toxicity to cancer cells than the hexylmethanimine group. Nonetheless, ligand **1d** with hexylmethanimine fragments in lateral chains is less toxic in NHDF cells. None of the evaluated ligands presents higher cytotoxicity than gemcitabine, a commercial anticancer drug.

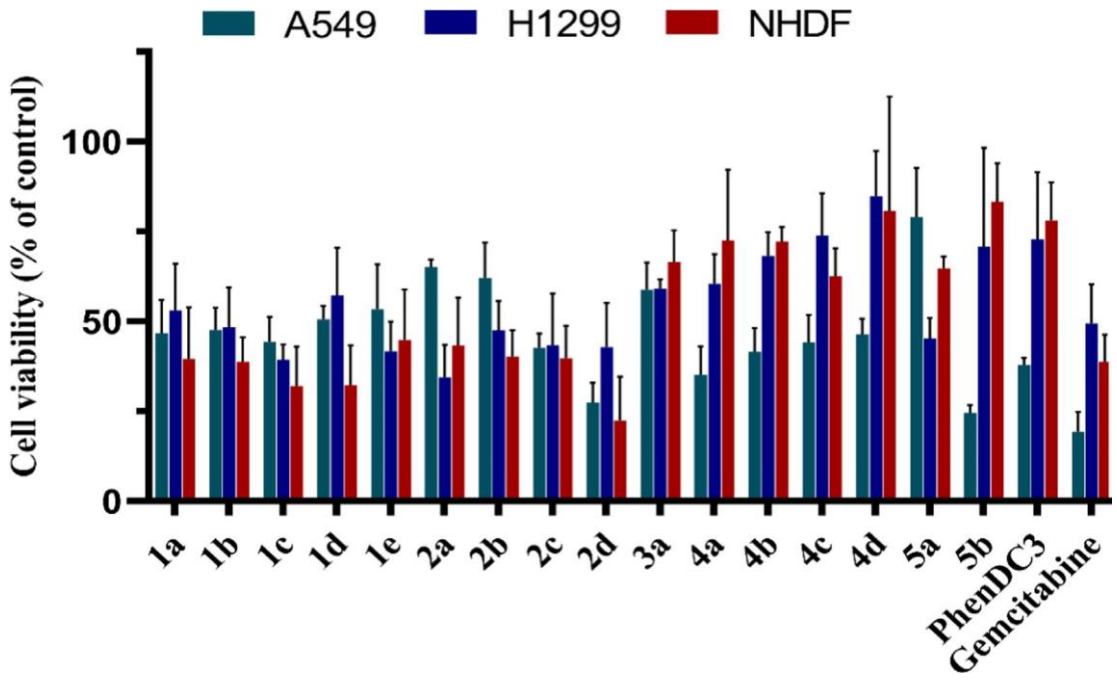


Figure 6.1. Relative cell viability of A549, H1299 and NHDF cells measured by the MTT assay after 48 h incubation with ligands at 10 μM concentration. Bars represent the mean ± SD of cell viability relative to control cells (no treatment). Error margins correspond to SD of three replicates.

Conclusion

In this report we described the synthesis of new 2,9-disubstituted-1,10-phenanthroline derivatives. All synthesized ligands showed selectivity towards G4 over duplex structure. The most promising ligands in terms of thermal stabilization were ligands **2b**, **3a** and **5b**. Compounds **3a** and **5b** display ethane-aminium substituents connected to the phenanthroline core through amine or amide groups whereas ligand 2b is equipped with a methoxyaniline moiety in the lateral chains. Upon ligand titration the CD spectra showed that the overall topology of the antiparallel 22CTA and parallel 26ceb sequences was maintained followed by an increase in ellipticity, indicating ligand association and stabilization of the preformed G4 structure. However, upon addition of 5 molar equivalents of ligand 5b some changes in the hybrid conformation of 22AG were observed. Overall, the ligands presented an adequate cytotoxic profile in cancer cell lines, being more effective against A549 than H1299 cells. The ligand **5b** proved to be the best compound in terms of cytotoxicity against A549 cells.

This study provides invaluable information about the structure–activity relationship for this kind of compounds in order to maximize their activity, however further

investigations are required to reveal the complete mechanism of action for these derivatives.

Experimental section

Chemistry

General procedures and methods

Commercial reagents were used as received without any purification. NMR spectra were recorded at 298 K on Varian MERCURY 300 MHz and Bruker AVANCE NEO 500 MHz spectrometers, employing $CDCl_3$ or DMSO-d_6 as solvents, with their residual signals being used to reference the spectra. High-resolution mass spectra were registered on an Agilent 6545 Q-TOF mass spectrometer coupled to a 1260 Infinity liquid chromatographer from the same brand; the ionisation source employed was electrospray in its positive mode.

Synthesis of precursors

tert-Butyl (2-aminoethyl)carbamate

The experimental procedure was adapted from that found in the literature [32]. Di-*tert*-butyl dicarbonate (2.45 g, 11.2 mmol) was dissolved in dioxane (30 mL) and this solution added, drop by drop, to a solution of freshly distilled ethylenediamine (6 mL, 87.5 mmol, 7.8 equiv.) in dioxane (30 mL) over a period of 2 h. The mixture was stirred for 20 h at room temperature. The solvent was removed by rotary evaporation and water (50 mL) was poured into the crude. The white precipitate formed (the bis-substituted product) was filtered off and the filtrate extracted with dichloromethane (3 × 50 mL). The organic phase was dried over anhydrous sodium sulphate, filtered and concentrated to dryness. The obtained residue was purified by column chromatography, using SiO_2 as the stationary phase and a dichloromethane-methanol mixture (from 0 to 5 % of methanol) as the eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator, yielding a yellow oil which gradually solidified to give a white solid. Yield: 1.13 g, 63 %. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 6.78-6.40 (m, 1H), 3.49 (s, 2H), 2.95-2.89 (m, 2H), 2.60-2.45 (m, 2H), 1.37 (s, 9H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 155.7 (C=O), 77.5 (Cq), 42.5 (CH_2), 41.0 (CH_2), 28.3 (CH_3). HR-MS (+ESI): found *m/z* 161.1287 ($[M+H]^+$), $[C_7H_{17}N_2O_2]^+$ requires *m/z* 161.1285 (monoisotopic mass).

Synthesis of imines and amines

Compound **1a**

1,10-phenanthroline-2,9-dicarbaldehyde (0.18 g, 0.77 mmol), which was synthesised as reported in the literature [35], was dissolved in a 1:1 (*V*/*V*) chloroform:methanol mixture (15 mL) and *tert*-butyl (2-aminoethyl)carbamate (0.26 g, 1.63 mmol, 2.1 equiv.) was added as a solution in the same mixture of solvents (5 mL). The reaction mixture was heated to reflux with stirring until TLC revealed the total consumption of the dialdehyde (2 h). Solvents were removed by rotary evaporation and diethyl ether (15 mL) was added to the obtained foam. The resulting solution was sonicated, which led to the formation of a white precipitate; the content was left in the freezer overnight. The solid was isolated by vacuum filtration, washed with diethyl ether (5 mL) and dried *in vacuo*. Yield: 251 mg, 62 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.89 (s, 2H), 8.44 (d, J = 8.4 Hz, 2H), 8.34 (d, J = 8.4 Hz, 2H), 7.90 (s, 2H), 4.85 (sb, 2H), 3.89–3.86 (m, 4H), 3.67–3.40 (m, 4H), 1.44 (s, 18H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 164.4 (imine CH), 155.9 (C=O), 154.6 (ArC), 145.5 (ArC), 137.0 (ArCH), 129.8 (ArC), 127.5 (ArCH), 120.7 (ArCH), 79.4 (Cq), 61.2 (CH2), 41.1 (CH2), 28.5 (CH3). HR-MS (+ESI): found *m/z* 521.2869 ($[M + H]^+$), $[C28H37N6O4]^+$ requires *m/z* 521.2871 (monoisotopic mass).

Compound **2a**

Compound **1a** (0.19 g, 0.36 mmol) was dissolved in anhydrous methanol (20 mL) and the resulting solution heated to reflux with stirring. Sodium borohydride (0.029 g, 0.77 mmol, 2.1 equiv.) was added in portions and reaction progress monitored by TLC. Once the reaction was completed (2.5 h), the solvent was removed by rotary evaporation. The crude was dissolved in dichloromethane (25 mL) and this solution washed with water (3 × 25 mL). The organic phase was dried over anhydrous sodium sulphate, filtered and concentrated to dryness. The resulting foam was re-crystallised from a dichloromethane-hexane mixture; the supernatant was poured into a round-bottomed flask and the solvent removed by rotary evaporation, and the pale yellow solid washed carefully with hexane and dried *in vacuo*. Yield: 21 mg, 11 % (this yield refers to the isolated solid. The compound is quite soluble in this mixture of solvents; the chemical yield is 94 %). ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.22 (d, J = 8.2 Hz, 2H), 7.77 (s, 2H), 7.69 (d, J = 8.2 Hz, 2H), 5.85-5.26 (m, 2H), 4.29 (s, 4H), 3.44-3.24 (m, 4H), 2.93 (t, J = 5.6 Hz, 4H), 2.23 (s, 2H), 1.41 (s, 18H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 160.4 (ArC), 156.4 (C=O), 145.4 (ArC), 136.8 (ArCH), 127.9 (ArC), 126.1 (ArCH), 122.4 (ArCH), 79.1

(Cq), 55.6 (CH_2), 49.3 (CH_2), 40.5 (CH_2), 28.5 (CH_3). HR-MS (+ESI): found *m/z* 525.3183 ($[M+H]^+$), $[C_{28}H_{41}N_6O_4]^+$ requires *m/z* 525.3184 (monoisotopic mass).

Compound **3a**

Compound **2a** (0.065 g, 0.12 mmol) was dissolved in dichloromethane (3 mL) and trifluoroacetic acid (3 mL) was added. The reaction mixture was stirred at room temperature for 24 h. Both the solvent and the acid were evaporated under an air current, affording a brown oil that was dried *in vacuo* for 30 min. Subsequently, diethyl ether (30 mL) was added, and the content left in the freezer overnight. This was followed by sonication; a powdery solid was formed and the suspension was stirred at room temperature for 30 min. After that, the supernatant was disposed, and diethyl ether (30 mL) added again. This process was performed three times to remove the excess of trifluoroacetic acid. The brown solid was isolated by vacuum filtration, washed with diethyl ether (3 × 5 mL) and dried *in vacuo*. Finally, it was purified by re-crystallisation (diffusion of diethyl ether into a solution of the compound in methanol). ^{19}F NMR of a mixture of the product and 2,2,2-trifluoroethanol (reference) allowed to determine the fluorine content of the sample and, therefore, the number of trifluoroacetic acid molecules per molecule of compound (**4**). Yield: 88 mg, 91 %. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 10.10–6.70 (sb, 10H), 8.65 (d, J = 8.3 Hz, 2H), 8.10 (s, 2H), 7.93 (d, J = 8.3 Hz, 2H), 4.78 (s, 4H), 3.51 (t, J = 6.4 Hz, 4H), 3.37 (t, J = 6.4 Hz, 4H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 159.1 (Cdouble bondO, q, 2J = 31 Hz), 152.7 (ArC), 143.9 (ArC), 137.9 (ArCH), 128.2 (ArC), 126.9 (ArCH), 122.5 (ArCH), 116.9 (Cq, q, 1J = 299 Hz), 51.1 (CH_2), 44.4 (CH_2), 35.5 (CH_2). ^{19}F NMR (300 MHz, DMSO-d6): δ (ppm) = -73.8 (s). HR-MS (+ESI): found *m/z* 325.2141 ($[M + H]^+$), $[C_{18}H_{25}N_6]^+$ requires *m/z* 325.2135 (monoisotopic mass).

Compound **1b**

1,10-phenanthroline-2,9-dicarbaldehyde (0.10 g, 0.42 mmol) was dissolved in a 1:1 (*V*/*V*) chloroform:methanol mixture (15 mL) and 4-methoxyaniline (0.11 g, 0.89 mmol, 2.1 equiv.) was then added. The reaction mixture was heated to reflux with stirring until TLC revealed the total consumption of the dialdehyde (2.5 h). Solvents were removed by rotary evaporation and the product was re-crystallised from a dichloromethane-hexane mixture. The yellow solid was isolated by vacuum filtration, washed with diethyl ether (10 mL) and dried *in vacuo*. Yield: 125 mg, 66 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 9.18 (s, 2H), 8.66 (d, J = 8.4 Hz, 2H), 8.38 (d, J = 8.4 Hz, 2H), 7.91 (s, 2H), 7.48 (d, J = 8.9 Hz, 4H), 7.00 (d, J = 8.9 Hz, 4H), 3.87 (s, 6H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 159.2 (ArC), 158.4 (imine CH), 155.3 (ArC), 145.6 (ArC), 143.3 (ArC),

136.8 (ArCH), 129.6 (ArC), 127.3 (ArCH), 123.1 (ArCH), 120.7 (ArCH), 114.5 (ArCH), 55.5 (CH_3). HR-MS (+ESI): found m/z 447.1813 ($[M + H]^+$), $[C_{28}H_{23}N_4O_2]^+$ requires m/z 447.1816 (monoisotopic mass).

Compound **2b**

Compound **1b** (0.12 g, 0.28 mmol) was dissolved in anhydrous methanol (10 mL) and the resulting solution heated to reflux with stirring. Subsequently, sodium borohydride (0.02 g, 0.55 mmol, 2 equiv.) was added in portions. A precipitate was formed and reaction progress was monitored by TLC. Once the reaction was completed (2.5 h), the mixture was cooled down to room temperature and the white solid isolated by vacuum filtration, washed with methanol (5 mL) and diethyl ether (5 mL) and dried *in vacuo*. Yield: 89 mg, 71 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.21 (d, J = 8.3 Hz, 2H), 7.77 (s, 2H), 7.73 (d, J = 8.3 Hz, 2H), 6.81 (d, J = 9.0 Hz, 4H), 6.75 (d, J = 9.0 Hz, 4H), 4.84 (s + s_b, 6H), 3.75 (s, 6H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 159.8 (ArC), 152.2 (ArC), 145.2 (ArC), 142.4 (ArC), 136.8 (ArCH), 127.9 (ArC), 126.0 (ArCH), 121.4 (ArCH), 115.0 (ArCH), 114.4 (ArCH), 55.8 (CH_3), 51.1 (CH_2). HR-MS (+ESI): found m/z 451.2135 ($[M + H]^+$), $[C_{28}H_{27}N_4O_2]^+$ requires m/z 451.2129 (monoisotopic mass).

Compound **1c**

The procedure followed to synthesize this compound was adapted from that found in the literature (it should be noted that it was not isolated by the authors) [33]. 1,10-phenanthroline-2,9-dicarbaldehyde (0.10 g, 0.42 mmol) was dissolved in a 1:1 (V/V) chloroform:methanol mixture (15 mL) and 2-aminoethanol (54 µL, 0.89 mmol, 2.1 equiv.) was then added. The reaction mixture was heated to reflux for 2.5 h. Solvents were removed by rotary evaporation and the product was re-crystallised from a chloroform-diethyl ether mixture. The white solid was isolated by vacuum filtration, washed with diethyl ether (10 mL) and dried *in vacuo*. Yield: 104 mg, 76 %. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 8.70 (s, 2H), 8.57 (d, J = 8.4 Hz, 2H), 8.34 (d, J = 8.4 Hz, 2H), 8.10 (s, 2H), 4.72 (t, J = 5.4 Hz, 2H), 3.86–3.82 (m, 4H), 3.78–3.73 (m, 4H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 163.3 (imine CH), 154.2 (ArC), 144.8 (ArC), 137.0 (ArCH), 129.3 (ArC), 127.4 (ArCH), 119.9 (ArCH), 63.2 (CH_2), 60.5 (CH_2). HR-MS (+ESI): found m/z 345.1320 ($[M + Na]^+$), $[C_{18}H_{18}N_4O_2Na]^+$ requires m/z 345.1322 (monoisotopic mass).

Compound **2c**

The procedure followed to synthesise the non-protonated compound was adapted from that found in the literature [33]. Compound **1c** (0.10 g, 0.32 mmol) was dissolved in anhydrous methanol (10 mL) and the resulting solution heated to reflux with stirring. Subsequently, sodium borohydride (0.024 g, 0.64 mmol, 2 equiv.) was added in portions and reaction progress was monitored by TLC. Once the reaction was completed (2.5 h), solvents were removed by rotary evaporation and dichloromethane (20 mL) was added to the crude. After its sonication, the mixture was kept in the freezer for an hour and then filtered. The filtrate was concentrated to dryness and the residue dissolved in methanol (5 mL). Addition of some drops of concentrated hydrochloric acid followed by partial evaporation of the solvent led to the formation of a pale pink precipitate that was isolated by vacuum filtration, washed with diethyl ether (5 mL) and dried *in vacuo*. Yield: 23 mg, 19 %. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 9.85 (s, 4H), 8.76 (d, J = 8.3 Hz, 2H), 8.17 (s, 2H), 8.07 (d, J = 8.3 Hz, 2H), 4.89–4.82 (s_b + s, 10H), 3.87 (t, J = 5.3 Hz, 4H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 152.5 (ArC), 140.6 (ArC), 139.8 (ArCH), 128.6 (ArC), 127.1 (ArCH), 124.3 (ArCH), 56.5 (CH_2), 49.6 (CH_2), 49.5 (CH_2). HR-MS (+ESI): found m/z 164.0946 ($[M + 2H]^{2+}$), $[C_{18}H_{24}N_4O_2]^{2+}$ requires m/z 164.0944 (diisotopic mass).

Compound **1d**

1,10-phenanthroline-2,9-dicarbaldehyde (0.10 g, 0.42 mmol) was dissolved in a 1:1 (*V*/*V*) chloroform:methanol mixture (15 mL) and *n*-hexylamine (123 µL, 0.93 mmol, 2.2 equiv.) was subsequently added. The reaction mixture was heated to reflux with stirring; after 3 h no signals of the dialdehyde were observed in the ^{1}H NMR spectrum of an aliquot taken from the mixture, so solvents were removed by rotary evaporation and the pure product, a pale orange solid, dried *in vacuo*. Yield: 144 mg, 84 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.86 (s, 2H), 8.42 (d, J = 8.4 Hz, 2H), 8.31 (d, J = 8.4 Hz, 2H), 7.86 (s, 2H), 3.77 (t, J = 6.8 Hz, 4H), 1.81–1.71 (m, 4H), 1.54–1.22 (m, 12H), 0.91–0.87 (m, 6H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 162.7 (imine CH), 154.9 (ArC), 145.4 (ArC), 136.7 (ArCH), 129.5 (ArC), 127.1 (ArCH), 120.3 (ArCH), 61.5 (CH_2), 31.6 (CH_2), 30.7 (CH_2), 27.1 (CH_2), 22.6 (CH_2), 14.0 (CH_3). HR-MS (+ESI): found m/z 425.2660 ($[M + Na]^+$), $[C_{26}H_{34}N_4Na]^+$ requires m/z 425.2676 (monoisotopic mass).

Compound **2d**

Compound **1d** (0.14 g, 0.36 mmol) was dissolved in anhydrous methanol (10 mL) and the resulting solution heated to reflux with stirring. Subsequently, sodium borohydride (0.027 g, 0.71 mmol, 2 equiv.) was added in portions and reaction progress monitored by TLC. Once the reaction was completed (3 h), solvents were removed by rotary evaporation. The crude was dissolved in dichloromethane (20 mL) and this solution washed with a 1 M HCl aqueous solution (1 × 20 mL). The organic phase was dried over anhydrous sodium sulphate, filtered and concentrated to dryness. The product was purified by column chromatography, using neutral Al_2O_3 as the stationary phase and a dichloromethane-methanol mixture (from 1 to 5 % of methanol) as the eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator. Dichloromethane (5 mL) was added and then, some drops of hexane, which drove to the formation of a white precipitate that was filtrated and dried *in vacuo*. Yield: 20 mg, 12 %. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 9.68 (s, 4H), 8.62 (d, J = 7.8 Hz, 2H), 8.09 (s, 2H), 7.91 (d, J = 7.8 Hz, 2H), 4.66 (s, 4H), 3.21–3.03 (m, 4H), 1.93–1.73 (m, 4H), 1.48–1.20 (m, 12H), 0.96–0.79 (m, 6H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): the record of a reliable spectrum was not possible due to the low solubility of the compound in DMSO-d_6. HR-MS (+ESI): found m/z 204.1623 ($[M + 2H]^{2+}$), $[C_{26}H_{40}N_4]^{2+}$ requires m/z 204.1621 (diisotopic mass).

Compound **1e**

The procedure followed to synthesise this compound was adapted from that found in the literature (it should be noted that it was not isolated by the authors) [34]. 1,10-phenanthroline-2,9-dicarbaldehyde (0.10 g, 0.42 mmol) was dissolved in a 1:1 (V/V) chloroform:methanol mixture (15 mL) and a 2.0 M solution of ethylamine in tetrahydrofuran (423 µL, 0.85 mmol, 2 equiv.) was subsequently added. The reaction mixture was heated to reflux with stirring for 16 h. Solvents were removed by rotary evaporation; the crude was dissolved in dichloromethane (20 mL) and this solution washed with water (1 × 20 mL). The organic phase was dried over anhydrous sodium sulphate, filtered and concentrated to dryness, yielding a brown solid that was dried *in vacuo*. Yield: 91 mg, 74 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.71 (s, 2H), 8.20 (d, J = 8.3 Hz, 2H), 8.03 (d, J = 8.3 Hz, 2H), 7.53 (s, 2H), 3.63 (q, J = 7.2 Hz, 4H), 1.18 (t, J = 7.2 Hz, 6H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 162.4 (imine CH), 154.9 (ArC), 145.3 (ArC), 136.8 (ArCH), 129.5 (ArC), 127.1 (ArCH), 120.3 (ArCH), 55.5 (CH_2), 16.0 (CH_3). HR-MS (+ESI): no peak associated with the compound was found in the spectrum.

Synthesis of amides (including macrocycles)

Compound **4a**

Dimethyl 1,10-phenanthroline-2,9-dicarboxylate and compound **4a** were synthesized as reported in the literature [35]. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 10.77 (s, 2H), 8.69 (dd, J = 8.3, 1.0 Hz, 2H), 8.40 (dd, J = 8.3, 1.0 Hz, 2H), 8.13 (d, J = 1.0 Hz), 5.02 (sb,4H). HR-MS (+ESI): found m/z 297.1083 ([M + H]$^+$), [$C_{14}H_{13}N_6O_2$]$^+$ requires m/z 297.1095 (monoisotopic mass).

Compound **5a**

Compound **5a** was synthesised as reported in the literature [35]. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 12.93 (s, 2H), 12.84 (s, 2H), 9.99 (s, 2H), 8.87 (d, J = 8.3 Hz, 2H), 8.64 (d, J = 8.3 Hz, 2H), 8.29 (s, 2H), 8.13 (d, J = 8.6 Hz, 2H), 7.92 (d, J = 9.0 Hz, 2H), 7.75 (d, J = 8.1 Hz, 2H), 7.23 (d, J = 8.9 Hz, 2H), 7.02 (t, J = 7.5 Hz, 2H), 6.57 (t, J = 7.7 Hz, 2H). HR-MS (+ESI): found m/z 605.1929 ([M + H]$^+$), [$C_{36}H_{25}N_6O_4$]$^+$ requires m/z 605.1932 (monoisotopic mass).

Compound **4b**

Dimethyl-1,10-phenanthroline-2,9-dicarboxylate (0.10 g, 0.34 mmol) was dissolved in a 1:1 (V/V) chloroform:methanol mixture (15 mL) and *tert*-butyl (2-aminoethyl)carbamate (0.54 g, 3.38 mmol, 10 equiv.) was added as a solution in the same mixture of solvents (5 mL). The reaction mixture was allowed to stir at room temperature for 24 h. Solvents were removed by rotary evaporation; the crude was dissolved in dichloromethane (50 mL) and this solution washed with water (3 × 50 mL). The organic phase was dried over anhydrous sodium sulphate, filtered and concentrated to dryness. The residue was purified by column chromatography, using SiO_2 as the stationary phase and a hexane–ethyl acetate mixture (from 1:1 to pure ethyl acetate) as the eluent. The fractions containing the desired product were combined and the solvent removed, yielding a solid that was re-crystallised from a dichloromethane-hexane mixture. The resulting white solid was filtrated, washed with hexane (5 mL) and dried *in vacuo*. Yield: 109 mg, 58 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 9.36 (s, 2H), 8.58 (d, J = 8.3 Hz, 2H), 8.42 (d, J = 8.3 Hz, 2H), 7.91 (s, 2H), 5.28 (sb, 2H), 3.77–3.72 (m, 4H), 3.60 (sb, 4H), 1.23 (s, 18H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 165.5 (amide C=O), 157.1 (Boc C=O), 149.9 (ArC), 144.3 (ArC), 137.6 (ArCH), 130.4 (ArC), 127.8 (ArCH), 121.5 (ArCH), 79.5 (Cq), 41.1 (CH_2), 40.0 (CH_2), 28.3 (CH_3). HR-MS (+ESI): foundm/z553.2781 ([M + H]$^+$), [$C_{28}H_{37}N_6O_6$]$^+$ requires m/z 553.2769 (monoisotopic mass).

Compound **5b**

Compound **4b** (0.072 g, 0.13 mmol) was dissolved in dichloromethane (3 mL) and trifluoroacetic acid (3 mL) was added. The reaction mixture was stirred at room temperature for 1.5 h. Both the solvent and the acid were evaporated under an air current, affording a pale yellow oil that was dried *in vacuo* for 30 min. Subsequently, diethyl ether (25 mL) was added, and the content left in the freezer overnight. This was followed by sonication; a powdery solid was formed and the suspension was stirred at room temperature for 30 min. After that, the supernatant was disposed, and diethyl ether (25 mL) added again. This process was performed five times to remove the excess of trifluoroacetic acid. The white solid was isolated by vacuum filtration, washed with diethyl ether (3 × 5 mL) and dried *in vacuo*. ^{19}F NMR of a mixture of the product and 2,2,2-trifluoroethanol (reference) allowed to determine the fluorine content of the sample and, therefore, the number of trifluoroacetic acid molecules per molecule of compound (2). Yield: 70 mg, 92 %. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 9.52 (t, J = 5.8 Hz, 2H), 8.77 (d, J = 8.3 Hz, 2H), 8.50 (d, J = 8.3 Hz, 2H), 8.22 (s, 2H), 8.00 (s_b, 6H), 3.75 (q, J = 5.8 Hz, 4H), 3.18 (s_b, 4H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 164.8 (amide C=O), 158.7 (TFA C=O, q, ^{2}J = 32 Hz), 149.4 (ArC), 143.7 (ArC), 138.3 (ArCH), 130.4 (ArC), 128.0 (ArCH), 121.2 (ArCH), 38.9 (CH_2), 37.2 (CH_2). ^{19}F NMR (300 MHz, DMSO-d_6): δ (ppm) = -73.6 (s). HR-MS (+ESI): found m/z 177.0901 ($[M + 2H]^{2+}$), $[C_{18}H_{22}N_6O_2]^{2+}$ requires m/z 177.0897 (diisotopic mass).

Compound **4c**

Dimethyl-1,10-phenanthroline-2,9-dicarboxylate (0.10 g, 0.34 mmol) was dissolved in a 1:1 (*V*/*V*) chloroform:methanol mixture (10 mL) and bis(2-aminoethyl)amine (550 µL, 5.09 mmol, 15 equiv.) was subsequently added. The reaction mixture was allowed to stir at room temperature for 24 h. Solvents were removed by rotary evaporation; the raw product was dissolved in chloroform (20 mL) and this solution washed with distilled water (1 × 20 mL). The organic phase was dried over anhydrous sodium sulphate, filtered and concentrated to dryness. Addition of methanol (5 mL) led to the formation of a white precipitate that was isolated by filtration, washed with diethyl ether (5 mL) and dried *in vacuo*. Yield: 13 mg, 11 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 10.48–10.26 (m, 2H), 8.32 (d, J = 8.2 Hz, 2H), 8.27 (d, J = 8.2 Hz, 2H), 7.71 (s, 2H), 3.63 (q, J = 4.9 Hz, 4H), 3.12–3.06 (m, 4H), 1.03 (p, J = 8.0 Hz, 1H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 163.5 (C=O), 149.2 (ArC), 142.8 (ArC), 137.1 (ArCH), 129.1 (ArC), 126.9 (ArCH), 120.1 (ArCH), 49.2 (CH_2), 39.2 (CH_2). HR-MS (+ESI): found m/z 336.1458 ($[M + H]^{+}$), $[C_{18}H_{18}N_5O_2]^{+}$ requires m/z 336.1455 (monoisotopic mass).

Compound **4d**

Dimethyl-1,10-phenanthroline-2,9-dicarboxylate (0.10 g, 0.34 mmol) was dissolved in a 1:1 (*V*/*V*) chloroform:ethanol mixture (10 mL) and 1,2-bis(2-aminoethoxy)ethane (745 µL, 5.08 mmol, 15 equiv.) was subsequently added. The reaction mixture was allowed to stir at room temperature for 96 h. Solvents were removed by rotary evaporation and the resulting oil turned into a solid quickly. Methanol (5 mL) was added and the white precipitate was isolated by filtration, washed with diethyl ether (5 mL) and dried *in vacuo*. Yield: 55 mg, 43 %. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.97 (t, J = 5.2 Hz, 2H), 8.62 (d, J = 8.3 Hz, 2H), 8.43 (d, J = 8.3 Hz, 2H), 7.91 (s, 2H), 3.91–3.86 (m, 4H), 3.78–3.75 (m, 8H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 164.3 (C=O), 149.9 (ArC), 144.4 (ArC), 137.7 (ArCH), 130.6 (ArC), 127.8 (ArCH), 121.8 (ArCH), 70.4 (CH_2), 70.1 (CH_2), 39.3 (CH_2). HR-MS (+ESI): found m/z 381.1562 ($[M + H]^+$), $[C_{20}H_{21}N_4O_4]^+$ requires m/z 381.1557 (monoisotopic mass).

In vitro evaluation

Oligonucleotides sequences and ligands

Oligonucleotide sequences were purchased from Eurogentec (Belgium) with HPLC-grade purification and were used without further purification. Stock solutions were prepared in Milli-Q water at 100 µM concentration and stored at −20 °C. All sequences tested are shown in **Table B1** (Appendix B). The annealing procedure for all nucleotide sequences were 95 °C for 5 min followed by slow cooling to room temperature in the corresponding buffer before measurements. Stock solutions of the compounds were prepared as 10 or 5 mM solutions in dimethyl sulfoxide (DMSO) (Thermo Fisher Scientific, USA).

Förster Resonance Energy Transfer (FRET) melting experiments

FRET melting assay was performed in 96-well plates using a CFXp6 qPCR equipment (Bio-Rad, California, USA). The oligonucleotides sequences were purchased dual labeled with fluorescein (FAM), a fluorescent donor, and tetramethyl rhodamine (TAMRA), a fluorescent acceptor, at the 5′ and 3′ ends, respectively, and are shown in **Table B1**. The oligonucleotides were annealed at 0.23 µM concentration in 10 mM lithium cacodylate (pH 7.2) supplemented with 10 mM KCl and 90 mM LiCl. The sequences and ligands were added in each well (final concentrations of 0.2 µM and 5 µM, respectively). The microplate was incubated at 25 °C and after that the FAM fluorescence (λ(exc) = 495 nm, λ(em) = 518 nm) was recorded every 0.5 °C between 25 and 95 °C with a temperature

increment of 0.5 °C min^{-1}. Each assay was performed in triplicate. The melting temperature (T_m) was determined from the normalized curves as the mid-transition temperature and the stabilization induced by ligands (ΔT_m) was calculated by the difference between T_m of oligonucleotide sequence alone and measured with the ligand.

Circular dichroism spectroscopy

CD spectra were recorded at 25 °C on a Jasco J-815 spectrometer (Jasco, USA), using a Peltier temperature controller (model CDF-426S/15). The oligonucleotide sequences used are shown in **Table B1**. For the titration experiments, the required volume of the ligand solution was added to 10 μM of pre-folded oligonucleotide sequence dissolved in 10 mM lithium cacodylate (pH 7.2) supplemented with 100 mM KCl in the quartz cuvette (Hellma, Germany) of 1 mm path length. Each spectrum was acquired through wavelengths ranging from 220 to 320 nm, with a scan speed of 200 nm/min, 1 nm bandwidth, 1 s integration time over 3 averaged accumulations.

For CD melting studies, the temperature was monitored by the Peltier system. The required amount of ligand (until 5 M equivalents) was added to the cuvette containing pre-folded oligonucleotide solution in 10 mM lithium cacodylate (pH 7.2) supplemented with 100 mM KCl. The denaturation process was analyzed by monitoring the wavelength of maximum ellipticity in 20–100 °C range with a 4 °C/min temperature gradient. Data were converted into fraction folded (θ) plots according to Equation (1), fitted to a Boltzmann distribution using OriginPro2021.

Equation 1: $$\theta = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}}$$

The melting temperature (T_m) was determined from a two-state transition model where CD is the ellipticity at maximum of ellipticity at each temperature and CD^{min} and CD^{max} are the lowest and highest ellipticities, respectively.

Cell cultures

Non-small human lung cancer cell lines (A549 and H1299), and human dermal fibroblast cell line (NHDF) were obtained from American Type Culture Collection (ATCC). Cells were cultured in 75 cm^2 *T*-flasks at 37 °C in a humidified atmosphere containing 5 % CO_2. A549 cells were cultured in Ham's F-12 medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10 % fetal bovine serum (FBS) and 1 % streptomycin-penicillin (SP) antibiotic, and H1299 and NHDF with RPMI medium (Sigma-Aldrich, St. Louis, MO,

USA) supplemented with 10 % FBS, 1 % SP antibiotic, 2 mM L-glutamine, 10 mM HEPES and 1 mM sodium pyruvate.

Cell viability assay

The cell viability was determined using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) (Sigma-Aldrich, St. Louis, MO, USA) assay. Briefly, cells were seeded in 96-well culture plates at a density of 2×10^4 cells / mL. After 24 h of adherence the medium was removed, and cells were incubated with appropriate medium containing different ligands concentration. After 48 h of incubation the medium was replaced with MTT solution and incubated during 4 h at 37 °C. After that, the MTT solution was removed, and formazan crystal formed were dissolved in 100 µL of DMSO. The absorbance was measured at 570 nm in a microplate reader Biorad Xmark spectrophotometer (Bio-Rad, California, USA). Cells cultured in the absence of ligands (untreated cells) were used as control. Each experiment was done in triplicate in three independent experiments. The cytotoxicity was normalized to untreated cells.

Acknowledgements

Joana Figueiredo acknowledges doctoral fellowship grant from FCT ref. SFRH/BD/145106/2019 and COST Action CA17140 Nanomedicine-from the bench to the bedside (NANO2CLINIC) for the Short-Term Scientific Mission (STSM) grant (ECOST-STSM-Request-CA17140-47763).
This work was supported by project ORACLE ref. UIDP/00709/2020, MICN/AEI/10.13039/501100011033 (grant PID2020-117610RB-100), the SYMBIT project (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477) financed by the ERDF and ANR G4Access [ANR-20-CE12-0023]. 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.

References

[1] Gellert, M. *et al.* (1962) HELIX FORMATION BY GUANYLIC ACID. *Proc Natl Acad Sci* 48, 2013–2018.

[2] Kim, J. *et al.* (1991) Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* 351, 331–332.

[3] Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* 344, 410–414.

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

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

[6] Beseiso, D. *et al.* (2022) The first crystal structures of hybrid and parallel four-tetrad intramolecular G-quadruplexes. *Nucleic Acids Res* 50, 2959–2972.

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

[8] Patel, D.J. *et al.* (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res* 35, 7429–7455.

[9] 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.

[10] Kharel, P. *et al.* (2020) The role of RNA G-quadruplexes in human diseases and therapeutic strategies. *Wiley Interdiscip Rev RNA* 11, e1568.

[11] Kosiol, N. *et al.* (2021) G-quadruplexes: a promising target for cancer therapy. *Mol Cancer* 20, 40.

[12] Awadasseid, A. *et al.* (2021) G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. *Biomed Pharmacother* 139, 111550.

[13] Monchaud, D. and Teulade-Fichou, M.-P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org Biomol Chem* 6, 627–636.

[14] Neidle, S. (2016) Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J Med Chem* 59, 5987–6011.

[15] Duarte, A.R. *et al.* (2018) Design of Modular G-quadruplex Ligands. *Chem Med Chem* 13, 869–893.

[16] Harrison, R.J. *et al.* (2003) Trisubstituted Acridine Derivatives as Potent and Selective Telomerase Inhibitors. *J Med Chem* 46, 4463–4476.

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

[18] Ou, T.M. *et al.* (2007) Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J Med Chem* 50, 1465–1474.

[19] Lu, Y.-J. *et al.* (2008) 5- *N* -Methylated Quindoline Derivatives as Telomeric G-Quadruplex Stabilizing Ligands: Effects of 5- N Positive Charge on Quadruplex Binding Affinity and Cell Proliferation. *J Med Chem* 51, 6381–6392.

[20] 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.

[22] Figueiredo, J. *et al.* (2022) Screening of Scaffolds for the Design of G-Quadruplex Ligands. *Appl Sci* 12, 2170.

[23] De Cian, A. *et al.* (2007) Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J Am Chem Soc* 129, 1856–1857.

[24] Wei, C.-Y. *et al.* (2013) 4-(1H-Imidazo[4,5-f]-1,10-phenanthrolin-2-yl)phenol-based G-quadruplex DNA binding agents: Telomerase inhibition, cytotoxicity and DNA-binding studies. *Bioorg Med Chem* 21, 3379–3387.

[25] Wu, Q. *et al.* (2016) Arene Ruthenium(II) Complexes as Low-Toxicity Inhibitor against the Proliferation, Migration, and Invasion of MDA-MB-231 Cells through Binding and Stabilizing c-myc G-Quadruplex DNA. *Organomet* 35, 317–326.

[26] Duskova, K. *et al.* (2016) Human telomeric G-quadruplex DNA interactions of N-phenanthroline glycosylamine copper(II) complexes. *Bioorg Med Chem* 24, 33–41.

[27] Nielsen, M.C. *et al.* (2009) Design, synthesis and evaluation of 4,7-diamino-1,10-phenanthroline G-quadruplex ligands. *Bioorg Med Chem* 17, 8241–8246.

[28] Larsen, A.F. *et al.* (2012) Tetrasubstituted Phenanthrolines as Highly Potent, Water-Soluble, and Selective G-Quadruplex Ligands. *Chem - Eur J* 18, 10892–10902.

[29] Chung, W.J. *et al.* (2014) Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC 3. *Angew Chem Int Ed Engl* 53, 999–1002.

[30] Gueddouda, N.M. *et al.* (2017) Design, Synthesis, and Evaluation of 2,9-Bis[(substituted-aminomethyl)phenyl]-1,10-phenanthroline Derivatives as G-Quadruplex Ligands. *ChemMedChem* 12, 146–160.

[31] Guillon, J. *et al.* (2021) Design, synthesis, and antiproliferative effect of 2,9-bis[4-(pyridinylalkylaminomethyl)phenyl]-1,10-phenanthroline derivatives on human leukemic cells by targeting G-quadruplex. *Arch Pharm* 354, 2000450

[32] Krapcho, A.P. and Kuell, C.S. (1990) Mono-Protected Diamines. N-tert-Butoxycarbonyl-α,ω-Alkanediamines from α,ω-Alkanediamines. *Synth Commun* 20, 2559–2564.

[33] Baysal, A. *et al.* (2011) Synthesis and structural characterization of a novel seven-coordinate cobalt(II) complex: 2,9-Bis(ethanolamine)-1,10-phenanthrolinechlorocobalt(II) chloride. *Inorg Chim Acta* 371, 107–110.

[34] Hernández-Ayala, L.F. *et al.* (2018) Synthesis, characterization, theoretical studies and biological activity of coordination compounds with essential metals containing N4-donor ligand 2,9-di(ethylaminomethyl)-1,10-phenanthroline. *Inorg Chim Acta* 470, 187–196.

[35] R, S.K. *et al.* (2018) Development of the Smartphone-Assisted Colorimetric Detection of Thorium by Using New Schiff's Base and Its Applications to Real Time Samples. *Inorg Chem* 57, 15270–15279.

[36] De Cian, A. *et al.* (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods* 42, 183–195.

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

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

[39] 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

7th Chapter

Synthesis of 1,10-phenanthroline-2,9-bistriazoles: Evaluation as G-quadruplex binders and anti-tumor activity

This work was submitted for publication

Figueiredo, J.; *Carreira-Barral, I.; Quesada, R.; laranjo, M.; Mergny, J-L. and Cruz, C. (2024) Synthesis of 1,10-phenanthroline-2,9-bistriazoles: Evaluation as G-quadruplex binders and anti-tumor activity.*

Chapter overview

Following the development of 2,9-disubstituted-1,10-phenanthroline containing amine and amide side chains as G4 binders, we explore an alternative 2,9-substitutional pattern. This chapter describes the chemical synthesis of a novel library of 1,10-phenanthroline-2,9-bistriazoles derivatives. The introduction of triazole groups can lead to the enhancement of the binding affinities through additional hydrogen bonding or π-π interactions. In addition, we explore the structural diversity of the triazole-linked substituents, including aryl, alkyl, and protonable amino groups. The synthesized compounds were assessed for their stabilizing properties towards different G4 structures within promoters, telomeric regions, and RNA sequences. The antiproliferative activity of the compounds was evaluated against LC cell lines and non-malignant cells. The results obtained in this chapter provide valuable insights into the structure-activity relationship of this class of compounds regarding their G4 stabilizing properties and anti-tumor effects.

Synthesis of 1,10-phenanthroline-2,9-bistriazoles: Evaluation as G-quadruplex binders and anti-tumor activity

Abstract

Novel 1,10-phenanthroline-2,9-bistriazoles derivatives has been synthesized by copper-catalyzed azide/alkyne cycloaddition reactions and assessed for their ability to bind and stabilize G-quadruplex (G4) structures. Ten novel compounds were evaluated using Förster resonance energy transfer (FRET) melting, circular dichroism (CD), and fluorescence spectroscopy on several G4 sequences. Biophysical characterization led to the identification of compounds **4a**, **4b**, and **5b** as good G4 ligands of KRAS G4 sequences. The anticancer activity of all derivatives was also assessed, revealing weak effects. However, compound **2a** exhibited a potent inhibitory effect on A549 and H1299 cancer cell growth with low cytotoxicity towards non-malignant cells MRC-5 apparently not connected with its G4-binding ability; thus, the mechanism of anticancer activity of **2a** will be investigated in future studies.

Keywords

G-quadruplexes; G4 ligands; 1,10-Phenanthroline-2,9-bistriazoles; Biophysical studies; Lung cancer; Antiproliferative activity

Introduction

DNA and RNA nucleic acid sequences containing consecutive runs of guanines can form four-stranded noncanonical secondary structures called G-quadruplexes (G4s) [1]. G4s are formed by stacking two or more quartets of guanines (G-quartets), bound by Hoogsteen-type base-pair hydrogen bonds, and stabilized by a central spine of generally monovalent cations (typically K^+ and Na^+) [2]. These structures are dynamic and structurally diverse, displaying intramolecular (self-folded by a single strand of G-rich oligonucleotide) or intermolecular (resulting from the association of multiple G-rich

strands) conformations, and can adopt multiple topologies (parallel, anti-parallel, and hybrid) depending upon their sequence, strand orientation, loop size, annealing conditions, crowding, and the presence of small molecules [3].

G4s can be formed *in vitro* [4] and in cells [5] and G4-prone motifs are found in key genomic regulatory regions such as gene promoters, telomeres, and RNA 5′-untranslated regions (5′-UTR) [6]. G4 structures are shown to be involved in controlling processes at the DNA and RNA levels, such as DNA replication, telomere maintenance, gene transcription, translation, and genomic stability [7]. Moreover, immunostaining with quadruplex-specific antibody showed that cancer tissues seem to exhibit more G4s than non-malignant cells [8]. G4 levels showed to be notably increased in the single-stranded overhang DNA of telomeres [9], and nuclease-hypersensitive elements (NHE) of a number of proto-oncogenes promoters such as *MYC*, *KIT*, *BCL-2*, and *KRAS*, highlighting them as an attractive target for anticancer approach [10]. The *in vivo* dynamic of endogenous G4s, including folding, unfolding, and topological changes, depends on the environmental conditions. These dynamics can influence biological processes, and the use of small molecules to bind and stabilize G4s (G4 ligands) can modulate their activity [11]. Telomerase activity is upregulated in cancer cells to fight telomere erosion during cell division and allowing unlimited proliferation. Thus, G4 ligands can prevent the unfolding of G4 telomeric DNA, inhibiting telomerase activity and/or altering telomere capping and halting tumor growth [12]. In addition, the overexpression of proto-oncogenes also leads to uncontrolled cancer cell proliferation, and G4 ligands can modulate oncogene expression by stabilizing G4 in their promoters. In this sense, the use of G4 ligands emerged as a promising anticancer therapeutic tool [13].

The possible roles of G4 formation as important regulatory systems in lung cancer (LC) have stimulated efforts to harness G4 ligands that would facilitate the stabilization of these structures, leading to *in vitro* and *in vivo* antiproliferative and anti-tumor effects [14].

The literature reports a great variety of scaffolds such as quindolines [15], acridines [16], naphthalene diimides [17,18], and phenanthrolines [19] which have been synthesized and proved to have G4 binding properties. An often - but not always - shared property of G4 binders is a planar aromatic core that interacts with external G-quartets through π-π stacking. Additionally, cationic side chains are frequently introduced to interact with the phosphate backbone [20].

Several studies have used 1,10-phenanthroline derivatives as G4 stabilizers of the telomeric DNA of several oncogenes exhibiting potential antiproliferative activity towards A549 LC cells [21]. The heteroaromatic core of the phenanthroline moiety defines a planar aromatic scaffold that provides an efficient π-π stacking with a terminal G-quartet and allows for diverse chemical modifications. Within its chemical space, *bisquinolinium* derivatives are the most efficient G4 stabilizers [22]. Among them, PhenDC3 (**Figure 7.1**), a phenanthroline dicarboxidiamide, became the benchmark G4 stabilizing ligand, showing excellent affinity and selectivity towards G4 over duplex [23]. Based on these observations, the G4 binding properties of phenanthroline-based derivatives have been optimized in several ways. An interesting approach is the functionalization of PhenDC3 analogs with triazole-linked positively charged groups. Triazole-linked groups lead to more potent and selective G4 ligands and different scaffolds containing the triazole moiety have been reported [24-26]. Phenanthroline derivatives containing 2,9-bistriazoles were identified as potent and selective G4 ligands towards telomeric and c-KIT G4 DNA [27]. Recently, novel 1,10-phenanthroline compounds disubstituted at positions 2 and 9 with amide groups linked to triazole moieties showed to be G4 stabilizers and to induce moderate anticancer effects against colon cancer cells [28].

In a previous work, we synthesized and characterized several new 2,9-disubstituted-1,10-phenanthroline derivatives containing imine, amine, or amide side chains that present an adequate cytotoxic profile against LC cell lines (**Figure 7.1**) [29]. Here we present the synthesis of novel 2,9-disubstituted-1,10-phenanthroline derivatives bearing 1,2,3-triazole groups. We evaluated the ligand-induced stabilization against several telomeric and non-telomeric oligonucleotide sequences through several biophysical techniques, including Förster resonance energy transfer (FRET), circular dichroism (CD), and fluorescence titrations. The antiproliferative effect of the synthesized ligands towards two different human non-small LC cell lines (A549 and H1299) and one human non-malignant cell line (MRC-5) was evaluated by the MTT assay.

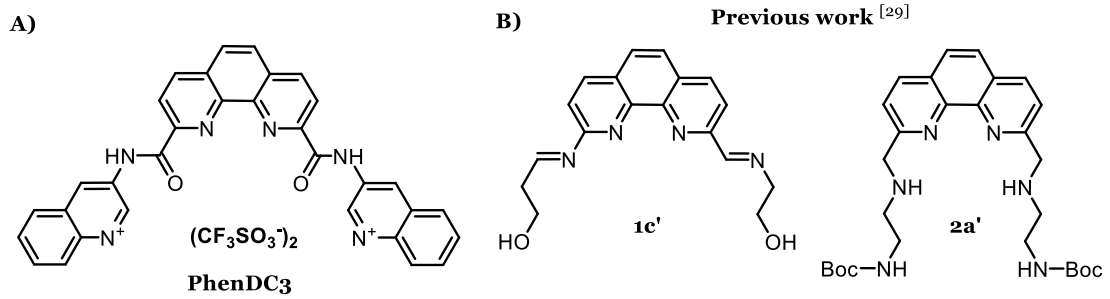
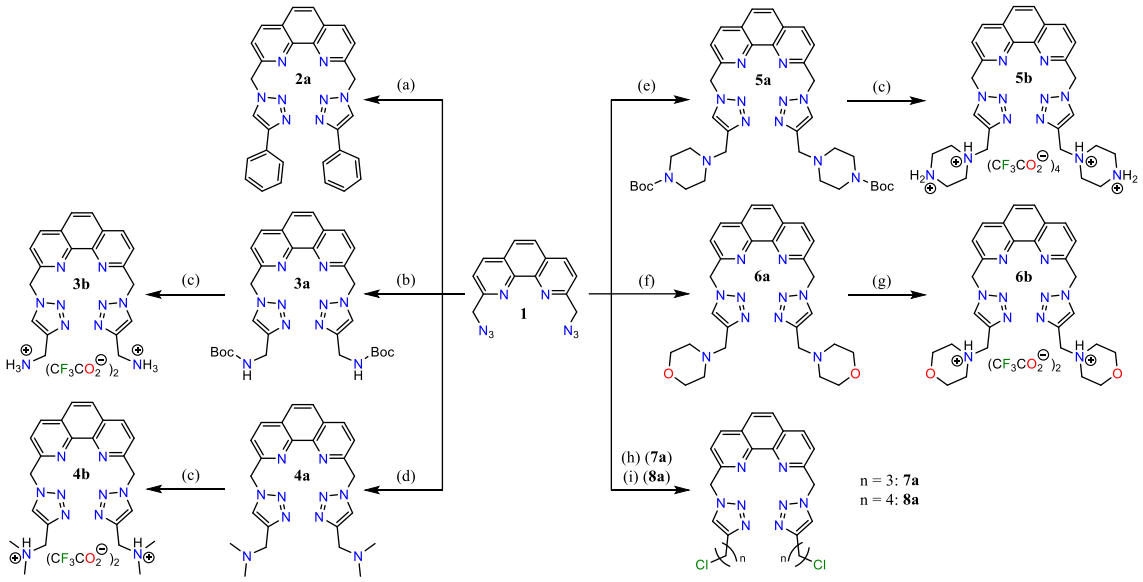


Figure 7.1. Representation of the **A)** compound PhenC3 and **B)** 2,9-disubstituted-1,10-phenanthroline derivatives **1c'** and **2a'** previously synthesized [29].

Results and Discussion

Chemistry

Copper-catalyzed azide/alkyne cycloaddition reactions (CuAAC) represent versatile synthetic tools for the generation of molecular diversity. We envisaged the use of the diazido derivative 2,9-bis(azidomethyl)-1,10-phenanthroline **1** as starting material for the generation of a library of compounds bearing 2,9-bis(methylene-1,2,3-triazoles) with different substitutions. Both aryl and alkyl substituents were targeted, and protonable amino groups were included to further enhance the electrostatic interactions with the negatively charged phosphate groups of DNA structures. All compounds were prepared employing **1** and the appropriate alkyne. A suitable Cu(II) complex precursor to catalyze these reactions was prepared by mixing equimolar amounts of $CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) in a (1:1, v/v) DMSO:H_2O mixture. The use of this Cu(II)-TBTA complex prevented the coordination of the phenanthroline to the copper ions, which would have resulted in poor yields of complex mixtures when $CuSO_4$ was employed as the copper source [27]. Treatment of some compounds with trifluoroacetic acid allowed deprotection and protonation of amino-substituted derivatives. These compounds were studied both as neutral compounds and trifluoroacetate salts (**Scheme 7.1**).



Scheme 7.1. Synthesis of compounds **2a**-**8a** and **3b**-**6b**. Reaction conditions: **(a)** Phenylacetylene, sodium L-ascorbate, Cu(II)-TBTA complex (6 mol%), DIPEA, DCM:H_2O, rt, 72 h; **(b)** *tert*-Butyl prop-2-yn-1-ylcarbamate (compound **9**, see ESI), sodium L-ascorbate, Cu(II)-TBTA complex (6 mol%), DIPEA, DCM:H_2O, rt, 24 h; **(c)** TFA, DCM, rt; **(d)** *N*,*N*-dimethylprop-2-yn-1-amine, sodium L-ascorbate, Cu(II)-TBTA complex (15 mol%), DIPEA, DCM:H_2O, rt, 96 h; **(e)** *tert*-Butyl 4-(prop-2-yn-1-yl)piperazine-1-carboxylate, sodium L-ascorbate, Cu(II)-TBTA complex (15 mol%), DIPEA, DCM:H_2O, rt, 72 h; **(f)** 4-(Prop-2-yn-1-yl)morpholine, sodium L-ascorbate, Cu(II)-TBTA complex (15 mol%), DIPEA, DCM:H_2O, rt, 96 h; **(g)** TFA, rt, 15 h; **(h)** 5-Chloropent-1-yne, sodium L-ascorbate, Cu(II)-TBTA complex (15 mol%), DIPEA, DCM:H_2O, rt, 48 h; **(i)** 6-Chlorohex-1-yne, sodium L-ascorbate, Cu(II)-TBTA complex (15 mol%), DIPEA, DCM:H_2O, rt, 72 h. Cu(II)-TBTA complex was prepared from a 5 mM stock solution containing equimolar amounts of $CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) in a (1:1, *v*/*v*) DMSO:H_2O mixture.

In vitro evaluation

Characterization of phenanthroline derivatives G4 thermal stabilization

Thermal stabilities of G-quadruplex structures in the presence and absence of ligands were initially assessed by FRET melting [30,31]. The labeled oligonucleotide sequences presented in Table S1 were employed in this assay, and ligand PhenDC3 was used as a reference. The variation in melting temperature (ΔT_m) of G4 induced by ligands was calculated by the difference between the melting temperature (T_m) of G4 sequences in the presence and absence of each ligand. The new phenanthroline derivatives **2a**-**8a** and **3b**-**6b** were evaluated towards several DNA G4-forming sequences capable of adopting parallel (FKRAST, F25CebT), antiparallel (F21CTAT), and hybrid (F21T) G4 conformations, as well as a parallel RNA G4 sequence (FpreMIR150T). The selectivity of the ligands towards G4 was also tested against a double-stranded DNA hairpin sequence (FdxT).

None of the tested derivatives induced a significant stabilization of the parallel FpreMIR150T and the antiparallel F21CTAT quadruplexes. A poor stabilization was observed for the parallel F25CebT sequence where the most promising derivative, ligand **5b**, induced a ΔT_m value of only ±2.1 °C at 5μM concentration (Table S2). Despite being modest, the ligand-induced stabilization was more evident for the hybrid F21T sequence. The most stabilizing derivatives were ligands **5b**, **4b**, and **4a**, which induced ΔT_m values of 4.9, 3.1, and 2.4 °C, respectively (**Table C2**). A similar trend was noted for previously synthesized 2,9-disubstituted-1,10-phenanthroline derivatives, which induced a higher thermal stabilization of the hybrid F21T than the FCebT and FCTAT sequences [29]. However, the derivatives previously synthesized exhibited a higher degree of stabilization at the same ligand concentration than the novel 1,10-phenanthroline-2,9-bistriazoles. These results suggest that functionalizing with a triazole moiety does not lead to enhanced stabilization of these telomeric sequences.

Interestingly, higher ΔT_m values were observed on parallel KRAS G4 sequences. As shown in **Figure 7.2** and **Table C3** (Appendix C), 1,10-phenanthroline-2,9-bistriazoles **2a**-**8a** and **3b**-**6b** induced ΔT_m values ranging from 2.3 to 11.2 °C at 5 μM ligand concentration and 0.4 to 5.1 °C at 2 μM ligand concentration. These findings indicate that ligand-induced stabilization occurs in a concentration-dependent manner. The synthesized ligands exhibited a preference for the FKRAS-21RT and FKRAS-22RTT sequences over the FKRAS-32RT sequence. However, the PhenDC3 ligand induced greater stabilization in all KRAS sequences. As observed for the telomeric F21T sequence, ligands **4a-b**, and **5b** exhibited the most stabilizing effect in all KRAS G4 sequences, inducing ΔT_m values between 4.7 and 11.2 °C. Among them, **5b** proved to be the most stabilizing derivative, inducing ΔT_m values of 7.6, 10.6, and 11.2 °C for FKRAS-21RT, FKRAS-22RTT, and FKRAS-32RT sequences, respectively. These results suggest that the presence of piperazine groups (**5b**) could contribute to the enhancement of G4 stabilization, as the replacement of these groups by morpholine ones (**6b**) diminishes G4 stabilization. Similarly, carbamate protecting groups in **5a** strongly decrease stabilization compared to congener **5b**. These findings may be associated with a reduction in the positive charge of the ligand, consistent with the notion that electrostatic interactions between the ligands and the negatively charged phosphate groups of the oligonucleotide sequence increase G4 thermal stabilization. Also, compound **3a**, containing carbamate end groups, showed lower stabilizing effects compared to **3b**. A similar trend was observed for ligands **4a-b**, where an increase in the positive charge of the ligand (**4b**) promoted enhanced stabilization of all KRAS G4 sequences. Interestingly, the increase in the positive charge of **6b** compared with **6a** did not significantly alter the stabilization profile. Additionally, ligand **8a**, containing

chlorobutyl side chains, demonstrated similar stabilization to **7a** with chloropropyl side chains. The introduction of a phenyl ring linked to the triazole moiety (**2a**) did not enhance the stabilization of the sequences.

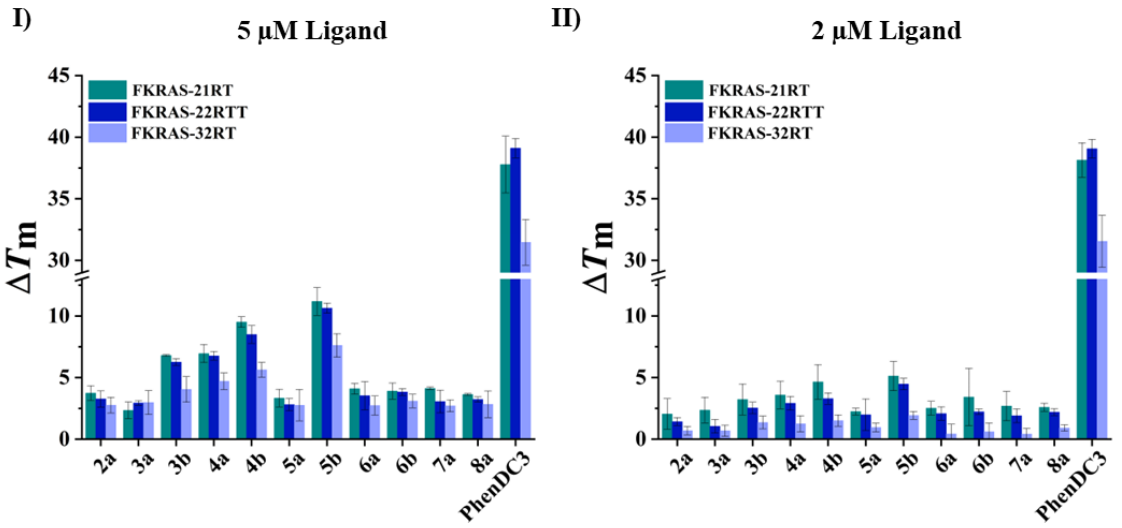


Figure 7.2. FRET-melting results. Ligand-induced changes in thermal stabilization (ΔT_m, °C) of the KRAS sequences (0.2 µM) in the presence of **I)** 5 µM and **II)** 2 µM ligand concentration. PhenDC3 was used as a reference compound at 0.5 µM. The experiments were performed in triplicate in a buffer containing 10 mM lithium cacodylate (pH 7.2), 90 mM LiCl, and 10 mM KCl.

Taken together, the initial screening conducted through FRET-melting experiments indicate that the ligands are more effective against KRAS G4 sequences, with ligands **4a-b**, and **5b** emerging as *hit* compounds for thermal stabilization within this series of 1,10-phenanthroline-2,9-bistriazoles. No stabilization effect was observed towards the hairpin duplex DNA sequence (**Table C2**).

Circular dichroism (CD) studies

CD experiments were performed to characterize the G4 conformation in the presence of the *hit* compounds (**4a-b** and **5b**) identified in FRET-melting experiments. CD spectra were acquired using a pre-folded G4 sequence in the presence of increasing concentrations of each ligand (up to 25 molar equivalents). The unlabeled oligonucleotide sequences used were KRAS-21R, KRAS-22RT, and KRAS-32R (**Table C1**). In the presence of 100 mM KCl, all KRAS sequences displayed the characteristic signature of a typical parallel-G4 topology, evidenced by a positive band at 264 nm and a negative one around 240 nm (**Figure C44**). Upon addition of each ligand, no significant variations in the CD signal were observed, suggesting an overall conservation of G4 topology (**Figure C44**).

Additionally, CD-melting experiments were carried out to confirm further the KRAS G4-stabilizing effect of **4a-b** and **5b**. The stabilization potential of these derivatives was observed following thermal denaturation, monitored at the wavelength of maximum CD ellipticity (**Figure 7.3**).

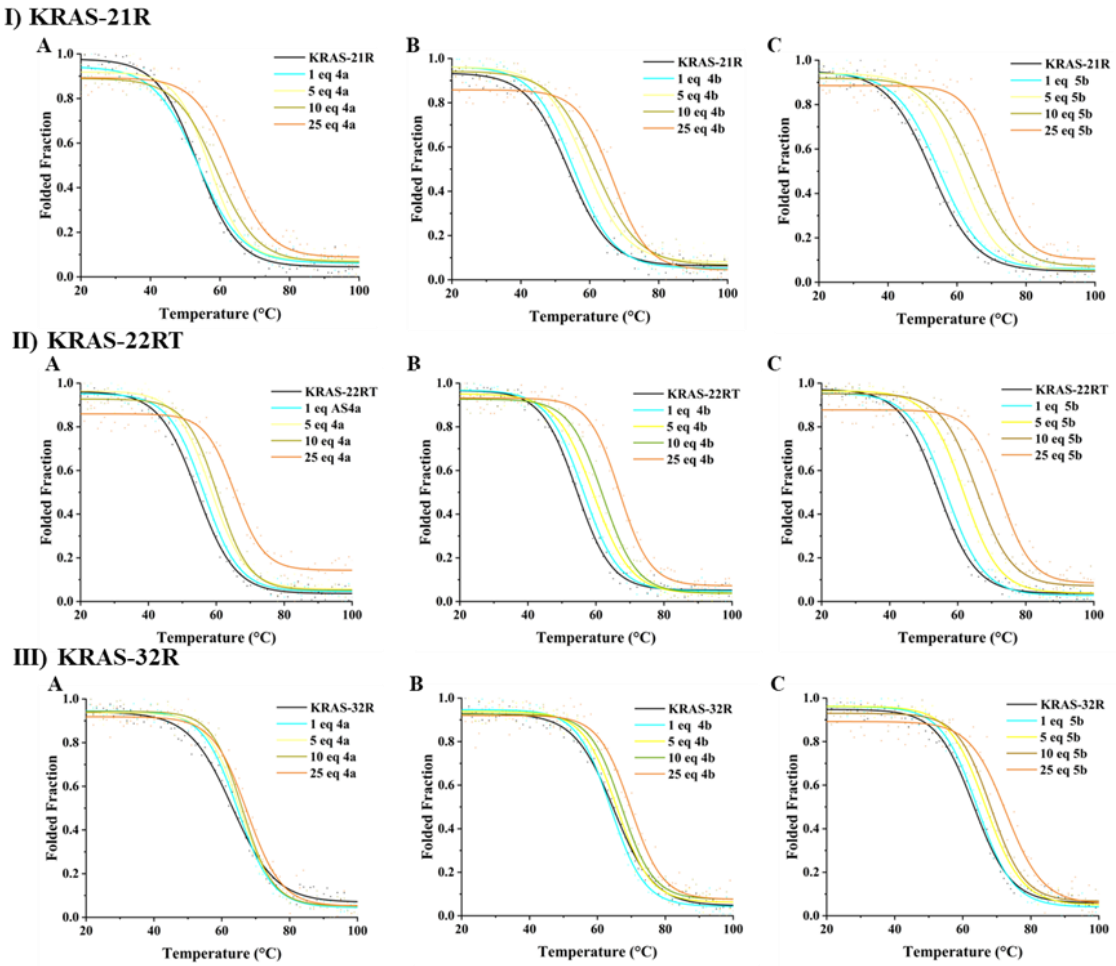


Figure 7.3. CD-melting curves of **I)** KRAS-21R **II)** KRAS-22RT, and **III)** KRAS-32R G-quadruplex in the absence and presence of increasing molar equivalents of the ligands **A) 4a**, **B) 4b**, and **C) 5b**. The experiments were performed in a buffer containing 10 mM lithium cacodylate (pH 7.2) and 100 mM KCl.

As commonly observed, the ligand-induced stabilization determined by CD differed from that obtained by FRET-melting studies, owing to differences in experimental conditions, presence or absence of fluorescent labels and exact oligonucleotide G4 sequences [32]. As depicted in **Table 7.1**, the assessed ligands maintained a slight preference for stabilizing the KRAS-21R and KRAS-22RT over the KRAS-32R sequence. Consistent with the structure-activity relationship (SAR) analyses of the FRET-melting assay, in the CD-melting experiments, ligand **5b**, featuring piperazine end groups, displayed the highest stabilizing activity towards all KRAS G4 sequences, followed by **4b**, and **4a** at same molar concentration (**Table C4**). As observed in FRET-melting experiments, the

thermal stabilization induced by these ligands was shown to be dependent on the ligand concentration. None of the ligands exhibited significant stabilization at 1 molar equivalent (max. +2.5 °C) and the highest stabilization was obtained after adding 25 molar equivalents of each ligand, evidencing a concentration dependent effect.

Table 7.1. CD-melting results. Ligand-induced changes in thermal stabilization (ΔT_m, °C) of the KRAS sequences (10 μM) in the presence of increasing ligand concentrations.

Ligand	Ligand concentration (molar equivalents)	ΔT_m (°C)[a]		
		KRAS-21R	KRAS-22RT	KRAS-32R
4a	1	0.2	1.9	1.5
	5	3.4	4.0	2.3
	10	5.4	6.0	2.9
	25	9.9	11.0	4.3
4b	1	1.4	1.9	0.0
	5	4.9	4.9	0.8
	10	7.7	7.8	2.2
	25	12.9	12.6	4.9
5b	1	2.2	2.5	1.0
	5	8.1	7.4	3.4
	10	12.0	11.2	5.0
	25	18.6	18.4	9.5

[a] ΔT_m represents difference in melting temperature [$\Delta T_m = T_m$ (DNA sequence + ligand) – T_m (DNA)]. The T_m values for DNA alone are 53.4 ± 0.7 °C (KRAS-21R), 54.3 ± 0.1 °C (KRAS-22RT), and 63.7 ± 0.9 °C (KRAS-32R). The experiments were performed in a buffer containing 10 mM lithium cacodylate (pH 7.2) and 100 mM KCl.

Characterization of phenanthroline derivatives G4-binding affinity

Spectrofluorimetric titration assays were conducted using double-labeled DNA sequences (FAM and TAMRA) (**Table C1**), to assess the affinity of phenanthrolines **4a**-**4b**, and **5b** for the FKRAS-21RT and FKRAS-22RTT G4 sequences. These sequences were selected based on their demonstrated highest degree of thermal stabilization induced by the selected ligands in previous experiments. However, it is important to note that ligand-induced thermal stabilization does not provide insights into other types of interactions and may not be directly correlated with binding affinity [33] The ligands were added to a fixed pre-folded G4 sequence in 10 mM lithium cacodylate (pH 7.2) supplemented with 100 mM KCl (consistent with the buffer used in CD experiments). The emission spectra of oligonucleotide sequences, with an excitation wavelength of 495 nm, exhibited a broad band centered at 517 nm. The fluorescence titration data was fitted to the Hill saturation binding mode (Equation 3, experimental section), giving an apparent dissociation constant (K_D) and Hill constant (n) for the different ligands (**Figure 7.4**). After the addition of ligands, the fluorescence spectra showed fluorescence quenching (**Figure C45**), and n values indicate negative cooperativity (n < 1) (**Table C5**). The K_D values of compounds **4a**, **4b**, and **5b** for the FKRAS-21RT sequence were

5.4×10^{-7}, 2.7×10^{-9}, and 3.5×10^{-6}, respectively. These results showed that the interaction between the tested phenanthroline derivatives and the FKRAS-21RT sequence is characterized by a medium to high affinity (K_D values in the 10^{-6} – 10^{-9} M range). Among these compounds, ligand **4b** exhibited the highest affinity for binding to this G4 sequence, followed by **4a** and **5b**. For the FKRAS-22RTT sequence, the K_D values for compounds **4a**, **4b**, and **5b** were 7.5×10^{-7}, 4.1×10^{-8}, and 2.1×10^{-6}, respectively. Ligands also displayed medium to high affinities (with K_D values in the 10^{-6} - 10^{-8} M range) towards the FKRAS-22RTT sequence. Ligand **5b** proved to have the lowest affinity for both sequences. These results suggest that the affinity of these ligands is not directly correlated to the thermal-induced stabilization, as the best G4 stabilizing agent (**5b**) is not the best G4 binder (**4b**).

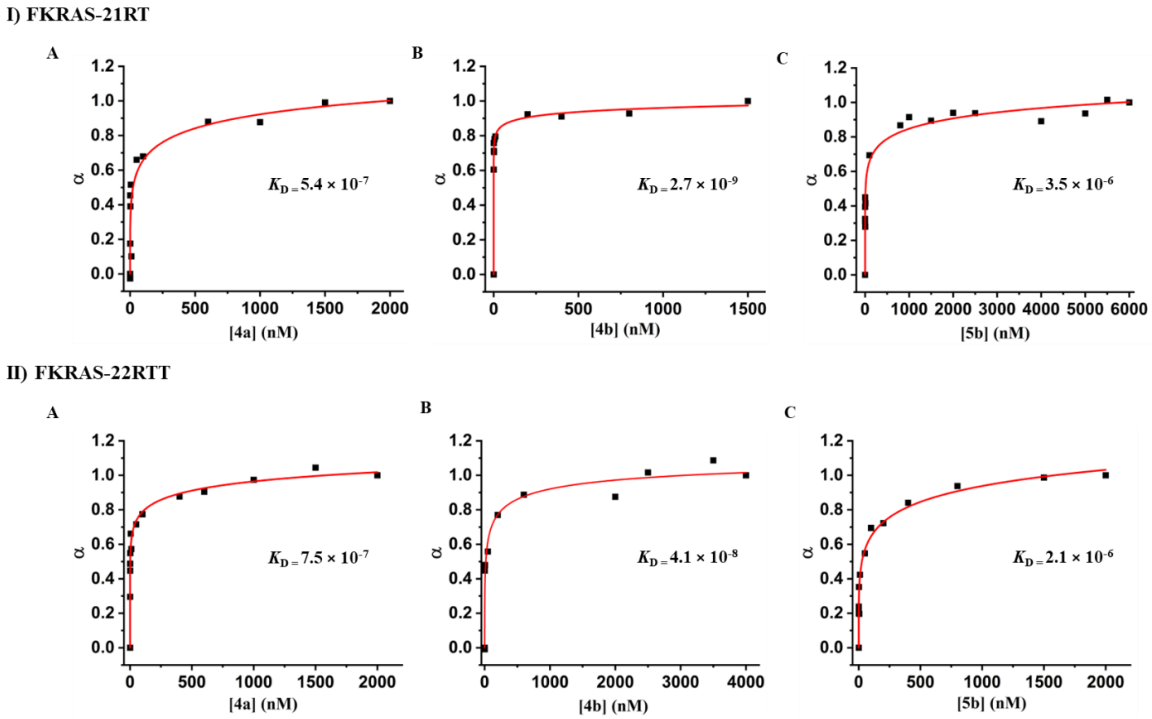


Figure 7.4. Saturation binding plots obtained by fluorescence titrations with compounds **A) 4a**, **B) 4b**, and **C) 5b** and fitted to the Hill saturation binding mode for **I)** KRAS-21RT and **II)** KRAS-22RTT. The experiments were performed in a buffer containing 10 mM lithium cacodylate (pH 7.2) and 100 mM KCl.

Antiproliferative activity of phenanthroline derivatives in malignant and non-malignant cell lines

The MTT assay was performed to investigate the anti-tumor activity of the newly synthesized ligands. Thus, all 1,10-phenanthroline-2,9-bistriazoles derivatives were tested on the human lung adenocarcinoma A549 and H1299 cells and lung non-

malignant MRC-5 fibroblasts. Initially, the screening was performed at a 10 μM concentration of each ligand (**Figure 7.5** and **Table C6**). Gemcitabine (GEM), a chemotherapeutic drug employed in LC treatment, was used as a positive control. Overall, the synthesized 1,10-phenanthroline-2,9-bistriazoles did not exhibit important antiproliferative effects on either cancer cell line and exerted no toxicity in non-malignant cells. The most active compound against cancer cell lines was phenanthroline derivative **2a.** This derivative, containing a phenyl ring linked to a triazole moiety, displayed higher cytotoxicity for the H1299 cell line than for A549 (metabolic activity of 36.6 and 66.0%, respectively, at 10 μM concentration) and an IC_{50} value of 14.6 and 10.9 μM for the A549 and H1299 cell lines, respectively (**Table 7.2**). Compound **2a** also exhibited lower cytotoxicity towards non-malignant cells (metabolic activity of 70.3%). Its activity on H1299 cells was also more pronounced than that induced by PhenDC3, with IC_{50} values of 10.9 μM and 28.0 μM, respectively (**Table 7.2**). However, compound **2a** is less potent than PhenDC3 (IC_{50} = 1.4 μM) in A549 cells (Table 2). This compound (**2a**) exhibited efficacy comparable to the previously synthesized most potent derivatives imine **1c'** and amine **2a'** previously synthesized) against H1299 cells and less toxicity towards non-malignant cells (at 10 μM concentration) [29]. These findings suggest that substituting methylaminoethanol (imine **1c'**) and tert-butyl(2-(methylamino)ethyl)carbamate (amine 2a) by 4-phenyl-4,5-dihydro-1,2,3-triazole (**2a**) side chains maintains the anti-tumor efficacy against H1299 cells and improves selectivity for malignant cells over non-malignant one. Derivative **7a**, featuring chloropropyl side chains, also demonstrated more cytotoxicity against H1299 cells (metabolic activity of 55.8%) compared to A549 cells (metabolic activity of 76.5%). The increase in the length of the chlorine side chain in ligand **8a** resulted in a lower antiproliferative activity in H1299 cells (metabolic activity of 86.1%). Despite being modest stabilizers, ligands **2a** and **7a** induced thermal stabilization of KRAS G4 sequences by 2.8 to 3.7 °C and 2.7 to 4.1 °C, respectively (**Table C3**). These results suggest that their antiproliferative activity may not be linked to their interaction with G4, or through the stabilization of the tested G4 sequences. Likewise, the most effective derivatives in terms of G4 stabilization (**4a**, **4b**, and **5b**) did not exhibit significant anti-tumoral activity in cancer cell lines, with cell metabolic activity no higher than 20% (**Figure 7.5**). Overall, the novel 1,10-phenanthroline-2,9-bistriazoles exhibit less efficacy against cancer cells but also induce less toxicity in non-malignant cells than the 2,9-disubstituted-1,10-phenanthroline derivatives previously reported.[29] However, the novel triazole-containing compound **2a** emerges as a promising anti-tumor agent against A549 and H1299 cells, with a mechanism of action which needed to be subsequently clarified.

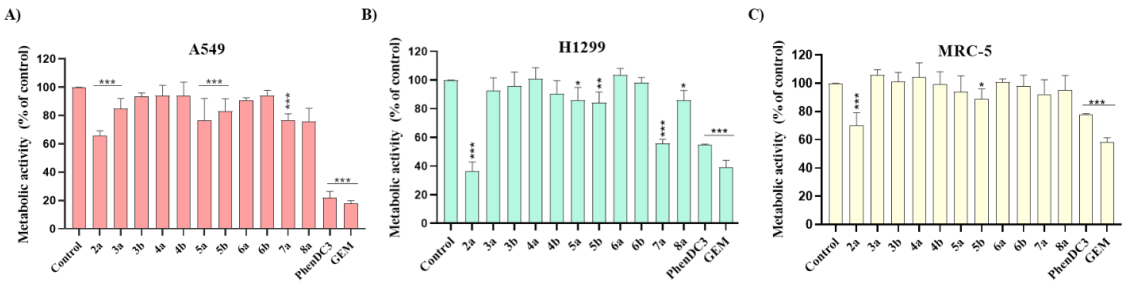


Figure 7.5. Relative metabolic activity of **A)** A549, **B)** H1299, and **C)** MRC-5 cells measured by the MTT assay after 48 h incubation with ligands at 10 µM concentration. Bars represent the mean ± SD of cell viability relative to control cells (no treatment). Error margins correspond to the SD of three replicates. Statistical analysis was performed using One-way ANOVA using Dunnett's post-hoc statistical hypothesis to analyze the statistical significance of the ligands' treatments compared to the control; statistical differences to the control are represented by (*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.001$.

Table 7.2. IC_{50} values of ligand **2a**, PhenDC3, and Gemcitabine in A549 and H1299 cell lines measured by MTT assay after 48 h incubation. The values were determined in three experiments independently with a 95% confidence interval.

	A549		H1299	
Compound	**IC_{50} (µM)**	**R^2**	**IC_{50} (µM)**	**R^2**
Ligand 2a	14.6	0.92	10.9	0.82
PhenDC3	1.4	0.98	28.0	0.94
Gemcitabine	0.03	0.98	nd[a]	nd[a]

[a] not determined.

In silico studies

In silico experiments were carried out to predict Lipinski's rule and Veber's parameters [34,35]. These studies were conducted to gain predictive insights into the cellular activity of the synthesized ligands and are shown in **Table 7.3**. Four parameters are considered in Lipinski's rule of five, comprising an *n*-octanol:water partition coefficient (logP), not exceeding 5, a molecular weight (MW) lower than 500 Da, no more than 10 hydrogen bond acceptors (n-ON acceptors), and no more than 5 hydrogen bond donors (n-OHNH). This rule predicts the oral bioavailability of tested drug candidates [36]. Although compliance is not an essential requirement, certain physicochemical parameters such as MW and logP can help understand issues related to membrane permeability or cellular uptake [37]. These problems are frequently reported in G4 ligands, which commonly have poor "drug-like" properties [38]. The compounds with the most important antiproliferative activity (**2a** and **7a**) did not exhibit more than one Lipinski's rule violation and a maximum of one violation is generally permitted [34]. However, the remaining derivatives, except ligands **4a** and **8a**, exhibited more than one violation to

the rule. Ligand **4a**, which displayed G4 binding/stabilization properties, also exhibits a "drug-like" character. This evidence suggests that the G4 binding/stabilizing activity of this ligand against the tested G4 sequences is not sufficient to induce an antiproliferative effect. Ligands **4b** and **5b** exhibited two violations of the rule. This lack in the physicochemical properties of these ligands may be related to their low anti-tumor activity. Despite these results being predictive, they can help understand the outcome of the cell studies.

Table 7.3. Calculated molecular properties for the tested ligands[a].

Ligand	TPSA[b]	MW	log *P*	n-ROTB[b]	n-ON acceptors	n-OHNH donors	Lipinski's violations[c]
2a	87.22	494.56	5.40	6	8	0	1
3a	163.88	600.68	4.11	8	12	2	2
3b	142.50	628.49	-1.86	6	12	2	2
4a	93.70	456.56	2.19	8	10	0	0
4b	96.09	684.60	-2.05	8	12	2	2
5a	152.78	738.90	4.17	8	14	0	2
5b	129.30	994.75	-6.82	8	16	4	2
6a	112.16	540.63	1.73	8	12	0	2
6b	114.56	768.68	-2.51	8	14	2	2
7a	87.22	495.42	4.41	10	8	0	0
8a	87.22	523.47	5.19	12	8	0	1

[a]TPSA, topological polar surface area (Å); MW, molecular weight (g/mol); Log P, n-octanol:water partition coefficient; n-ROTB, number of rotatable bonds; n-ON acceptors, number of hydrogen bond acceptors; n-OHNH donors, number of hydrogen bond donors.
[b]Veber's parameters: n-ROTB < 10 and TPSA < 10 Å^2.
[c] Lipinski's rule of five: MW < 500 Da; Log *P* ≤ 5; n-OHNH acceptors ≤ 10; n-OHNH donors ≤ 5. A maximum of 1 violation is generally permitted.

Conclusion

Ten 1,10-phenanthroline-2,9-bistriazoles derivatives were synthesized and investigated for their potential G4 binding/stabilizing properties. The compounds exhibited high thermal stabilization for the KRAS G4 sequences over telomeric, RNA G4 sequences, and duplex DNA. FRET melting assays and CD spectroscopy revealed that compounds **4a**, **4b**, and **5b** exhibit the best stabilizing properties against KRAS G4 sequences, ranging between 4.7 and 11.2 °C (according to FRET melting) or 4.3 and 18.6 °C (according to CD-melting). Moreover, these compounds demonstrated good binding affinities for G4 with K_D ranging from 10^{-6} and 10^{-9} M, as determined from fluorescence titrations. All derivatives were also evaluated for their *in vitro* anticancer activity against LC cell lines A549 and H1299, revealing weak antiproliferative activity. However, new triazole-

containing compound **2a** exhibited potent metabolic activity inhibition in both A549 and H1299 cancer cells, with IC_{50} values of 14.6 and 10.9 μM, respectively, therefore constituting a promising anti-tumor agent against LC. Further studies are required to better clarify the mechanism of action of this compound. Overall, the experimental results demonstrated that the synthesized compounds have potential as G4 binding/stabilizing agents and may serve as promising scaffolds for the development of new G4 ligands.

Experimental section

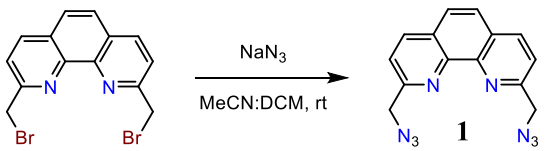
Chemistry

General procedures and methods

Commercial reagents were used as received without any purification. NMR spectra were recorded at 298 K on a Varian MERCURY (300 MHz) or Bruker AVANCE NEO 4500 (500 MHz) spectrometers, employing $CDCl_3$ or DMSO-d_6 as solvents, with their residual signals being used to reference the spectra. High-resolution mass spectra were registered on an Agilent 6545 Q-TOF mass spectrometer coupled to a 1260 Infinity liquid chromatographer from the same brand; the ionization source employed was electrospray in its positive mode.

Precursors

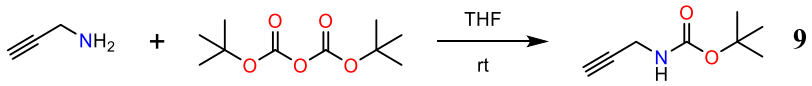
Compound **1**



2,9-bis(bromomethyl)-1,10-phenanthroline (500 mg, 1.36 mmol) was placed in a 50-mL round-bottomed flask and acetonitrile (25 mL) was added. The resulting suspension was stirred at room temperature and dichloromethane was added until achieving the complete dissolution of the starting material. Subsequently, sodium azide (444 mg, 6.83 mmol, 5 equiv.) was added, observing the formation of a white precipitate (sodium bromide) shortly after. Stirring was kept for 5 hours, after which the solvents were removed by rotary evaporation. Dichloromethane (30 mL) was added and the resulting suspension washed with water (2 × 30 mL). The organic phase was dried over anhydrous

sodium sulfate, filtered and concentrated to dryness, affording a yellow oil that solidified under vacuum. Yield: 393 mg, 99%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.33 (d, J = 8.3 Hz, 2H), 7.83 (s, 2H), 7.80 (d, J = 8.3 Hz, 2H), 4.97 (s, 4H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 156.8 (ArC), 145.2 (ArC), 137.5 (ArCH), 128.3 (ArC), 126.5 (ArCH), 121.4 (ArCH), 56.6 (CH_2). HR-MS (+ESI): found m/z 291.1106 ($[M+H]^+$), $[C_{14}H_{11}N_8]^+$ requires m/z 291.1101 (monoisotopic mass); found m/z 313.0927 ($[M+Na])^+$, $[C_{14}H_{10}N_8Na]^+$ requires 313.0921 (monoisotopic mass).

Compound **9**



9

The experimental procedure was adapted from that found in the literature [39]. Prop-2-yn-1-amine (1.16 mL, 18.1 mmol) was dissolved in anhydrous tetrahydrofuran (15 mL) and di-*tert*-butyl dicarbonate (4.36 g, 20.0 mmol, 1.1 equiv.) was added in one portion. The reaction mixture was stirred at room temperature for 24 hours and concentrated *in vacuo*. The crude was dissolved in ethyl acetate (50 mL) and this solution washed with water (3 × 10 mL) and brine (10 mL). The organic phase was dried over anhydrous sodium sulfate, filtered and the solvent removed by rotary evaporation. The residue was purified by column chromatography, using SiO_2 as stationary phase and a 9:1 hexane-ethyl acetate mixture as eluent. The fractions containing the desired product were combined and the solvent was evaporated, yielding a yellow oil which solidified after remaining under vacuum for several hours. Yield: 2.30 g, 82%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 4.83 (s_b, 1H), 4.07-3.71 (m, 2H), 2.19 (t, J = 2.5 Hz, 1H), 1.42 (s, 9H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.4 (C=O), 80.2 (CH), 80.1 (C_q), 71.3 (C_q), 30.5 (CH_2), 28.4 (CH_3). HR-MS (+ESI): found m/z 156.1017 ($[M+H]^+$), $[C_8H_{14}NO_2]^+$ requires m/z 156.1019 (monoisotopic mass);); found m/z 178.0841 ($[M+Na])^+$, $[C_8H_{13}NNaO_2]^+$ requires m/z 178.0838 (monoisotopic mass).

Triazole-containing compounds

Compound **2a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and phenylacetylene (189 µL, 1.72 mmol, 5 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (6 mol%, 4.134 mL of a 5 mM stock solution containing equimolar amounts of

$CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in a DMSO:H_2O (1:1, *v*/*v*) mixture) and *N*,*N*-diisopropylethylamine (300 µL, 1.72 mmol, 5 equiv.) were added, in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by TLC (SiO_2, dichloromethane-methanol mixture with 2% of methanol). Upon completion of the reaction (3 days), dichloromethane (25 mL) was added, and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous sodium sulfate, filtered and concentrated to dryness. The raw product was purified by column chromatography, using SiO_2 as stationary phase and a dichloromethane-methanol mixture (from 0 to 1% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator. Methanol (15 mL) was added to the residue and the pale yellow non-crystalline precipitate formed isolated by vacuum filtration, washed with diethyl ether (10 mL) and dried *in vacuo*. Yield: 84 mg, 49%. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 8.82 (s, 2H), 8.54 (d, J = 8.3 Hz, 2H), 8.02 (s, 2H), 7.91-7.87 (m, 4H), 7.56 (d, J = 8.3 Hz, 2H), 7.48-7.42 (m, 4H), 7.36-7.30 (m, 2H), 6.12 (s, 4H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 155.4 (ArC), 146.8 (ArC), 144.5 (ArC), 137.9 (ArCH), 130.6 (ArC), 128.9 (ArCH), 128.1 (ArC), 127.9 (ArCH), 126.7 (ArCH), 125.2 (ArCH), 122.4 (ArCH), 121.6 (ArCH), 55.6 (CH_2). HR-MS (+ESI): found *m*/*z* 495.2042 ($[M+H]^+$), $[C_{30}H_{23}N_8]^+$ requires *m*/*z* 495.2040 (monoisotopic mass);); found m/z 517.1863 $([M+Na])^+$, $[C_{30}H_{22}N_8Na]^+$ requires *m*/*z* 517.1860 (monoisotopic mass).

Compound **3a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and compound **9** (118 mg, 0.76 mmol, 2.2 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (6 mol%, 4.134 mL of a 5 mM stock solution containing equimolar amounts of $CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in a DMSO:H_2O (1:1, *v*/*v*) mixture) and *N*,*N*-diisopropylethylamine (300 µL, 1.72 mmol, 5 equiv.) were added, in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by TLC (SiO_2, dichloromethane-methanol mixture with 2% of methanol). Upon completion of the reaction (24 hours), dichloromethane (25 mL) was added and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous sodium sulfate, filtered and concentrated to dryness. The raw product was purified by column chromatography, using SiO_2 as stationary phase and a

dichloromethane-methanol mixture (from 0 to 4% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator, yielding a white non-crystalline solid that was dried *in vacuo*. Yield: 129 mg, 62%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.26 (d, J = 8.3 Hz, 2H), 7.83 (s, 2H), 7.79 (s, 2H), 7.42 (d, J = 8.3 Hz, 2H), 6.10 (s, 4H), 5.20 (s_b, 2H), 4.42 (d, J = 5.9 Hz, 4H), 1.39 (s, 18H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.9 (ArC), 155.5 (C=O), 146.4 (ArC), 145.3 (ArC), 138.0 (ArCH), 128.6 (ArC), 127.0 (ArCH), 122.9 (ArCH), 121.7 (ArCH), 79.8 (C_q), 56.5 (CH_2), 36.3 (CH_2), 28.4 (CH_3). HR-MS (+ESI): found m/z 601.3004 ($[M+H]^+$), $[C_{30}H_{37}N_{10}O_4]^+$ requires m/z 601.2994 (monoisotopic mass); found m/z 623.2824 $([M+Na])^+$, $[C_{30}H_{36}N_{10}NaO_4]^+$ requires m/z 623.2813 (monoisotopic mass).

Compound **3b**

Compound **3a** (226 mg, 0.38 mmol) was dissolved in dichloromethane (2.5 mL) and trifluoroacetic acid (2.5 mL) was added. The reaction mixture was stirred at room temperature for 24 hours. Both the solvent and the acid were evaporated under an air current, affording a brown oil that was dried *in vacuo* for 60 minutes. Subsequently, diethyl ether (30 mL) was added and the content left in the freezer overnight. This was followed by sonication; a powdery solid was formed and the suspension was stirred at room temperature for 60 minutes, after which the supernatant was disposed and diethyl ether (30 mL) was added again. This process was performed three times to remove the excess of trifluoroacetic acid. The beige solid was filtrated under vacuum, washed with diethyl ether (3 × 5 mL) and dried *in vacuo*. Finally, it was purified by re-crystallization (diffusion of diethyl ether into a solution of the compound in methanol). ^{19}F NMR of a mixture of the product and 2,2,2-trifluoroethanol (reference) allowed to determine the fluorine content of the sample and, therefore, the number of trifluoroacetic anions per molecule of compound (2). Yield: 205 mg, 87%. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 8.56 (d, J = 8.3 Hz, 2H), 8.39 (s, 2H), 8.34 (s_b, 6H), 8.03 (s, 2H), 7.51 (d, J = 8.3 Hz, 2H), 6.14 (s, 4H), 4.21 (s, 4H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 158.5 (C=O, q, ^{2}J = 31 Hz), 155.4 (ArC), 144.5 (ArC), 140.7 (ArC), 137.9 (ArCH), 128.1 (ArC), 126.8 (ArCH), 125.4 (ArCH), 121.7 (ArCH), 117.2 (C_q, q, ^{1}J = 299 Hz), 55.4 (CH_2), 34.0 (CH_2). ^{19}F NMR (282 MHz, DMSO-d_6): δ (ppm) = -73.5 (s). HR-MS (+ESI): found m/z 201.1014 ($[M+2H]^{2+}$), $[C_{20}H_{22}N_{10}]^{2+}$ requires m/z 201.1009 (diisotopic mass).

Compound **4a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and *N*,*N*-

dimethylprop-2-yn-1-amine (927 µL, 8.61 mmol, 25 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (15 mol%, 10.33 mL of a 5 mM stock solution containing equimolar amounts of $CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in DMSO:H_2O (1:1, *v*/*v*) mixture) and *N*,*N*-diisopropylethylamine (300 µL, 1.72 mmol, 5 equiv.) were added, in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by ^{1}H NMR spectroscopy. Upon completion of the reaction (4 days), dichloromethane (25 mL) was added to the reaction mixture, and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous sodium sulfate, filtered and concentrated to dryness. The raw product was purified by column chromatography, using neutral Al_2O_3 as stationary phase and a dichloromethane-methanol mixture (from 0 to 1.2% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator, yielding a pale yellow non-crystalline solid that was dried *in vacuo*. Yield: 19 mg, 12%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.26 (d, J = 8.3 Hz, 2H), 7.83 (s, 2H), 7.73 (s, 2H), 7.40 (d, J = 8.3 Hz, 2H), 6.12 (s, 4H), 3.64 (s, 4H), 2.28 (s, 12H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.8 (ArC), 146.2 (ArC), 145.3 (ArC), 138.1 (ArCH), 128.7 (ArC), 127.0 (ArCH), 123.4 (ArCH), 121.6 (ArCH), 56.6 (CH_2), 54.5 (CH_2), 45.3 (CH_3). HR-MS (+ESI): found *m/z* 457.2582 ($[M+H]^+$), $[C_{24}H_{29}N_{10}]^+$ requires *m/z* 457.2571 (monoisotopic mass); found *m/z* 479.2396 ($[M+Na])^+$, $[C_{24}H_{28}N_{10}Na]^+$ requires *m/z* 479.2391 (monoisotopic mass).

Compound **4b**

Compound **4a** (16 mg, 0.04 mmol) was dissolved in a mixture of trifluoroacetic acid (4 mL) and dichloromethane (1 mL), and the reaction mixture was stirred at room temperature for 5 hours. The acid and the solvent were evaporated under an air current, affording a yellow oil that was dried *in vacuo* for 30 minutes. Subsequently, diethyl ether (10 mL) was added and the content left in the freezer overnight. This was followed by sonication; a powdery solid was formed and the suspension was stirred at room temperature for 30 minutes, after which the supernatant was disposed and diethyl ether (30 mL) was added again. This process was performed five times to remove the excess of trifluoroacetic acid. Finally, the solvent was disposed and the yellow solid dried under vacuum. ^{19}F NMR of a mixture of the product and 2,2,2-trifluoroethanol (reference) allowed to determine the fluorine content of the sample and, therefore, the number of trifluoroacetic anions per molecule of compound (2). Yield: 17 mg, 70%. ^{1}H NMR (500 MHz, DMSO-d_6): δ (ppm) = 10.21 (s_b, 2H), 8.55 (d, J = 8.3 Hz, 2H), 8.54 (s, 2H), 8.03

(s, 2H), 7.53 (d, J = 8.3 Hz, 2H), 6.15 (s, 4H), 4.48 (s, 4H), 2.80 (s, 12H). ^{13}C NMR {DEPT-135} (126 MHz, DMSO-d_6): δ (ppm) = 158.0 (C=O, q, ^2J = 34 Hz), 155.2 (ArC), 144.5 (ArC), 137.9 (ArCH), 137.1 (ArC), 128.1 (ArC), 127.8 (ArCH), 126.8 (ArCH), 121.6 (ArCH), 117.2 (C_q, q, ^1J = 298 Hz), 55.4 (CH_2), 50.5 (CH_2), 41.8 (CH_3). ^{19}F NMR (471 MHz, DMSO-d_6): δ (ppm) = -73.5 (s). HR-MS (+ESI): found m/z 229.1322 ($[M+2H]^{2+}$), $[C_{24}H_{30}N_{10}]^{2+}$ requires m/z 229.1322 (diisotopic mass).

Compound **5a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and *tert*-butyl 4-(prop-2-yn-1-yl)piperazine-1-carboxylate (773 mg, 3.44 mmol, 10 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (15 mol%, 10.33 mL of a 5 mM stock solution containing equimolar amounts of $CuSO_4{\cdot}5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in a DMSO:H_2O (1:1, *v*/*v*) mixture) and *N*,*N*-diisopropylethylamine (300 µL, 1.72 mmol, 5 equiv.) were added, in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by ^{1}H NMR spectroscopy. Upon completion of the reaction (3 days), dichloromethane (25 mL) was added to the reaction mixture, and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous sodium sulfate, filtered and concentrated to dryness. The crude was purified by column chromatography, using SiO_2 as stationary phase and a dichloromethane-methanol mixture (from 0 to 12% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator, yielding a pale yellow non-crystalline solid that was dried *in vacuo*. Yield: 105 mg, 41%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.28 (d, J = 8.3 Hz, 2H), 7.84 (s, 2H), 7.73 (s, 2H), 7.43 (d, J = 8.3 Hz, 2H), 6.12 (s, 4H), 3.72 (s, 4H), 3.43-3.40 (m, 8H), 2.49-2.46 (m, 8H), 1.43 (s, 18H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.7 (ArC), 154.7 (Boc C=O), 145.2 (ArC), 145.0 (ArC), 138.1 (ArCH), 128.7 (ArC), 127.0 (ArCH), 123.6 (ArCH), 121.6 (ArCH), 79.7 (C_q), 56.6 (CH_2), 53.4 (CH_2), 52.7 (CH_2), 43.4 (CH_2), 28.5 (CH_3). HR-MS (+ESI): found m/z 739.4159 ($[M+H]^+$), $[C_{38}H_{51}N_{12}O_4]^+$ requires m/z 739.4151 (monoisotopic mass); found m/z 761.3978 ($[M+Na]^+$), $[C_{38}H_{50}N_{12}NaO_4]^+$ requires m/z 761.3970 (monoisotopic mass).

Compound **5b**

Compound **5a** (73 mg, 0.10 mmol) was dissolved in dichloromethane (4 mL) and trifluoroacetic acid (4 mL) was added. The reaction mixture was stirred at room

temperature for 3 days. Both the solvent and the acid were evaporated under an air current, affording a brown oil that was dried *in vacuo* for 60 minutes. Subsequently, diethyl ether (15 mL) was added and the content left in the freezer for 5 hours. This was followed by sonication; a powdery solid was formed and the suspension was stirred at room temperature for 60 minutes, after which the supernatant was disposed and diethyl ether (15 mL) was added again. This process was performed three times to remove the excess of trifluoroacetic acid. The resulting pale brown solid was purified by re-crystallization (diffusion of diethyl ether into a solution of the compound in methanol). ^{19}F NMR of a mixture of the product and 2,2,2-trifluoroethanol (reference) allowed to determine the fluorine content of the sample and, therefore, the number of trifluoroacetic anions per molecule of compound (4). Yield: 77 mg, 78%. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 8.70 (s_b, 6H), 8.54 (d, J = 8.3 Hz, 2H), 8.34 (s, 2H), 8.03 (s, 2H), 7.48 (d, J = 8.3 Hz, 2H), 6.08 (s, 4H), 3.88 (s, 4H), 3.19-3.07 (m, 8H), 2.84-2.74 (m, 8H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 158.8 (C=O), 155.5 (ArC), 144.5 (ArC), 140.6 (ArC), 137.9 (ArCH), 128.1 (ArC), 126.8 (ArCH), 126.3 (ArCH), 121.6 (ArCH), 55.4 (CH_2), 51.1 (CH_2), 48.3 (CH_2), 42.0 (CH_2). ^{19}F NMR (282 MHz, DMSO-d_6): δ (ppm) = -73.9 (s). HR-MS (+ESI): found m/z 539.3105 ($[M+H]^+$), $[C_{28}H_{35}N_{12}]^+$ requires m/z 539.3102 (monoisotopic mass).

Compound **6a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and 4-(prop-2-yn-1-yl)morpholine (444 µL, 3.44 mmol, 10 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (15 mol%, 10.33 mL of a 5 mM stock solution containing equimolar amounts of $CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in a DMSO:H_2O (1:1, *v*/*v*) mixture) and *N*,*N*-diisopropylethylamine (300 µL, 1.72 mmol, 5 equiv.) were added, in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by ^{1}H NMR spectroscopy. Upon completion of the reaction (4 days), dichloromethane (25 mL) was added to the reaction mixture, and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous sodium sulfate, filtered and concentrated to dryness. The raw product was purified by column chromatography, using neutral Al_2O_3 as stationary phase and a dichloromethane-methanol mixture (from 0 to 1% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator, yielding a pale yellow non-crystalline solid that was dried *in vacuo*. Yield: 27

mg, 14%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.28 (d, J = 8.3 Hz, 2H), 7.85 (s, 2H), 7.71 (s, 2H), 7.43 (d, J = 8.3 Hz, 2H), 6.12 (s, 4H), 3.70-3.68 (m, 12H), 2.54-2.51 (m, 8H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.7 (ArC), 145.3 (ArC), 144.6 (ArC), 138.2 (ArCH), 128.7 (ArC), 127.0 (ArCH), 123.9 (ArCH), 121.7 (ArCH), 66.8 (CH_2), 56.6 (CH_2), 53.7 (CH_2), 53.4 (CH_2). HR-MS (+ESI): found m/z 541.2775 ($[M+H]^+$), $[C_{28}H_{33}N_{10}O_2]^+$ requires m/z 541.2782 (monoisotopic mass); found m/z 563.2598 ($[M+Na])^+$, $[C_{28}H_{32}N_{10}NaO_2]^+$ requires m/z 563.2602 (monoisotopic mass).

Compound **6b**

Compound **6a** (25 mg, 0.05 mmol) was dissolved in trifluoroacetic acid (4 mL) and the solution was stirred at room temperature for 15 hours. The acid was evaporated under an air current, affording a yellow oil that was dried *in vacuo* for 60 minutes. Subsequently, diethyl ether (10 mL) was added and the content left in the freezer for 2 hours. This was followed by sonication; a powdery solid was formed and the suspension was stirred at room temperature for 60 minutes, after which the supernatant was disposed and diethyl ether (10 mL) was added again. This process was performed three times to remove the excess of trifluoroacetic acid. The resulting yellow solid was purified by re-crystallization (diffusion of diethyl ether into a solution of the compound in methanol). ^{19}F NMR of a mixture of the product and 2,2,2-trifluoroethanol (reference) allowed to determine the fluorine content of the sample and, therefore, the number of trifluoroacetic anions per molecule of compound (2). Yield: 20 mg, 57%. ^{1}H NMR (300 MHz, DMSO-d_6): δ (ppm) = 10.77 (s_b, 2H), 8.56-8.53 (m, 4H), 8.03 (s, 2H), 7.54 (d, J = 8.3 Hz, 2H), 6.14 (s, 4H), 4.50 (s, 4H), 3.80 (s, 8H), 3.25 (s, 8H). ^{13}C NMR {DEPT-135} (75 MHz, DMSO-d_6): δ (ppm) = 155.2 (ArC), 144.5 (ArC), 137.9 (ArCH), 136.9 (ArC), 128.1 (ArC), 127.8 (ArCH), 126.8 (ArCH), 121.7 (ArCH), 63.5 (CH_2), 55.5 (CH_2), 50.9 (CH_2), 50.1 (CH_2). ^{19}F NMR (282 MHz, DMSO-d_6): δ (ppm) = -73.7 (s). HR-MS (+ESI): found m/z 541.2791 ($[M+H]^+$), $[C_{28}H_{33}N_{10}O_2]^+$ requires m/z 541.2782 (monoisotopic mass).

Compound **7a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and 5-chloropent-1-yne (730 µL, 6.89 mmol, 20 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (15 mol%, 10.33 mL of a 5 mM stock solution containing equimolar amounts of $CuSO_4 \cdot 5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in a DMSO:H_2O (1:1, v/v) mixture) and *N*,*N*-diisopropylethylamine (300 µL, 1.72 mmol, 5 equiv.) were added,

in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by means of ^{1}H NMR spectroscopy. Upon completion of the reaction (2 days), dichloromethane (25 mL) was added to the reaction mixture, and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous sodium sulfate, filtered and concentrated to dryness. The raw product was partially purified by column chromatography, using SiO_2 as stationary phase and a dichloromethane-methanol mixture (from 0 to 2% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator. The resulting solid was re-crystallized from a dichloromethane-diethyl ether mixture; crystals were carefully collected, washed with diethyl ether and dried *in vacuo*. The mother liquor was concentrated to dryness and ethyl acetate (10 mL) was added to the residue. After sonication, the white non-crystalline precipitate formed was isolated by vacuum filtration, washed with ethyl acetate (3 × 5 mL) and hexane (2 × 5 mL), and dried *in vacuo*. The purity of both fractions was confirmed by ^{1}H NMR spectroscopy. Yield: 46 mg, 27%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.25 (d, J = 8.3 Hz, 2H), 7.82 (s, 2H), 7.58 (s, 2H), 7.39 (d, J = 8.3 Hz, 2H), 6.08 (s, 4H), 3.57 (t, J = 6.4 Hz, 4H), 2.90 (t, J = 7.3 Hz, 4H), 2.17 (p, J = 6.8 Hz, 4H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.8 (ArC), 147.3 (ArC), 145.3 (ArC), 138.1 (ArCH), 128.6 (ArC), 127.0 (ArCH), 122.1 (ArCH), 121.6 (ArCH), 56.5 (CH_2), 44.3 (CH_2), 31.9 (CH_2), 22.8 (CH_2). HR-MS (+ESI): found m/z 495.1583 ($[M+H]^+$), $[C_{24}H_{25}Cl_2N_8]^+$ requires m/z 495.1574 (monoisotopic mass); found m/z 517.1401 ($[M+Na])^+$, $[C_{24}H_{24}Cl_2N_8Na]^+$ requires m/z 517.1393 (monoisotopic mass).

Compound **8a**

The experimental procedure was adapted from that found in the literature [40]. Compound **1** (100 mg, 0.34 mmol) was dissolved in dichloromethane (6.9 mL) and 6-chlorohex-1-yne (1.044 mL, 8.61 mmol, 25 equiv.) was added, followed by water (6.2 mL). Subsequently, (+)-sodium L-ascorbate (136 mg, 0.69 mmol, 2 equiv.), Cu(II)-TBTA complex (15 mol%, 10.33 mL of a 5 mM stock solution containing equimolar amounts of $CuSO_4{\cdot}5H_2O$ and tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine in a DMSO:H_2O (1:1, *v*/*v*) mixture) and *N*,*N*-diisopropylethylamine (300 μL, 1.72 mmol, 5 equiv.) were added, in this order, to the reaction mixture. The mixture, which turned from yellow to orange upon the addition of the Cu(II) complex, was stirred vigorously at room temperature; reaction progress was monitored by ^{1}H NMR spectroscopy. Upon completion of the reaction (3 days), dichloromethane (25 mL) was added to the reaction mixture, and the organic phase was washed with water (25 mL) and brine (25 mL), dried over anhydrous

sodium sulfate, filtered and concentrated to dryness. The raw product was partially purified by column chromatography, using Al_2O_3 as stationary phase and a dichloromethane-methanol mixture (from 0 to 1% of methanol) as eluent. The fractions containing the desired product were combined and the solvent removed in a rotary evaporator. Ethyl acetate (10 mL) was added to the residue and, after sonication, the white crystalline solid formed was isolated by vacuum filtration, washed with ethyl acetate (3 × 5 mL) and hexane (2 × 5 mL), and dried *in vacuo*. Yield: 38 mg, 21%. ^{1}H NMR (300 MHz, $CDCl_3$): δ (ppm) = 8.27 (d, J = 8.3 Hz, 2H), 7.84 (s, 2H), 7.52 (s, 2H), 7.39 (d, J = 8.3 Hz, 2H), 6.11 (s, 4H), 3.58-3.54 (m, 4H), 2.81-2.77 (m, 4H), 1.88-1.83 (m, 8H). ^{13}C NMR {DEPT-135} (75 MHz, $CDCl_3$): δ (ppm) = 155.8 (ArC), 148.4 (ArC), 145.1 (ArC), 138.0 (ArCH), 128.5 (ArC), 126.9 (ArCH), 121.6 (ArCH), 121.5 (ArCH), 56.4 (CH_2), 44.7 (CH_2), 32.0 (CH_2), 26.6 (CH_2), 25.0 (CH_2). HR-MS (+ESI): found m/z 545.1718 ($[M+Na]^+$), $[C_{26}H_{28}Cl_2N_8Na]^+$ requires m/z 545.1706 (monoisotopic mass).

In vitro evaluation

Oligonucleotide sequences and ligands

KRAS oligonucleotide sequences were purchased from Eurofins and telomeric, and RNA sequences from Eurogentec. All sequences were obtained as dried samples with HPLC-grad purification. Stock solutions were prepared in Milli-Q water and stored at -20 °C. The oligonucleotide concentrations were determined at 260 nm in a UV-Vis spectrophotometer (Thermo Scientific™ Evolution 220) using the molar extinction coefficients (ε) provided by the manufacturer. All sequences are represented in Table S1. Samples were annealed in the corresponding buffer and kept at 95 °C for 5 minutes followed by slow cooling at room temperature before the experiments. Stock solutions of the compounds were prepared at 10 mM concentration in dimethyl sulfoxide (DMSO) (Thermo Fisher Scientific, USA).

Förster Resonance energy transfer (FRET) melting experiments

FRET-melting experiments were run on CFXp6 qPCR equipment (Bio-Rad, California, USA) in 96-well plates on the dual-label oligonucleotide sequences reported in Table S1. The oligonucleotides were annealed at 0.23 μM concentration in a buffer containing lithium cacodylate (10 mM, pH 7.2), KCl (10 mM), and LiCl (90 mM). The sequences were added to each well with a final concentration of 0.2 μM and ligands with 5 μM. The microplate was incubated at 25 °C and the measurements were performed at the FAM fluorescence (λ_{exc} = 495 nm, λ_{em} = 518 nm) and recorded every 0.5 °C between 25 and 95 °C with a temperature increment of 0.5 °C/min. The melting temperature (T_m) was

determined from the normalized curves as the mid-transition temperature and the stabilization induced by ligands (ΔT_m) was calculated by the difference between T_m of the oligonucleotide sequence alone and measured with the ligand. Data are presented as an average of three independent experiments.

Circular dichroism (CD) spectroscopy

CD experiments were run on Jasco J-815 spectrometer (Jasco, USA) with a Peltier temperature controller (model CDF-426S/15) using a quartz cell (Hellma, Germany) of 1 mm path length. Each spectrum was acquired at 20 °C from 220 to 320 nm, with a scan speed of 200 nm/min, 1 nm bandwidth, 1 second integration time over 3 averaged accumulations. The oligonucleotide sequences used are shown in Table S1. For titration experiments, the required volume of the ligand solution was added to the pre-folded oligonucleotide sequence (10 μM) dissolved in a buffer containing lithium cacodylate (10 mM, pH 7.2), KCl (10 mM) and LiCl (90 mM).

For CD-melting experiments, temperature was monitored by the Peltier system. The required amount of ligand (until 25 molar equivalents) was added to the quartz cell containing the pre-folded oligonucleotide solution (10 μM) in lithium cacodylate (10 mM, pH 7.2), KCl (10 mM) and LiCl (90 mM). The denaturation process was analyzed by monitoring the wavelength of maximum ellipticity in the 20-100 °C range with a 2 °C/min temperature gradient. Data were converted into fraction folded (θ) plots according to Equation 7.1, fitted to a Boltzmann distribution using OriginPro 2021.

$$\theta = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \qquad \text{(Equation 7.1)}$$

The melting temperature (T_m) was determined from a two-state transition model where CD is the ellipticity at maximum of ellipticity at each temperature and CD^{min} and CD^{max} are the lowest and highest ellipticities, respectively.

Fluorescence binding studies

Fluorescence titrations were conducted on a Horiba FluoroMax fluorometer (Kyoto, Japan), with a temperature control system using quartz suprasil cuvettes (Hellma, Germany) of 10 mm path length. For the experiments, the tested ligands (**4a-b** and **5b**) were successively added to a pre-folded oligonucleotide solution (1 μM) dissolved in a buffer containing lithium cacodylate (10 mM, pH 7.2) and KCl (100 mM). After each

addition, the mixture was left for 5 min for equilibrium. Each spectrum was acquired at 25 °C with an integration time of 0.5 s, an emission and excitation slit fixed at 1 nm and 2 nm, respectively, and a step size of 1 nm, averaged over three scans. The oligonucleotide sequences labeled with FAM (**Table S1**) were excited at 495 nm and the emission spectra were recorded from 510 to 700 nm.

Data were converted into a fraction of bound ligand (α) plots using Equation 7.2:

$$\alpha = \frac{I - I_{\lambda}^{free}}{I_{\lambda}^{bound} - I_{\lambda}^{free}} \qquad \text{(Equation 7.2)}$$

where I is the fluorescence intensity at maximum wavelength for each DNA:ligand ratio and I^{free} and I^{bound} are the fluorescence intensities of the free and fully bound ligand, respectively. Data points were then fitted into the Hill saturation binding model using OriginPro 2021 and K_D values were determined from Equation 7.3:

$$\alpha = \frac{[ligand]^n}{K_D + [ligand]^n} \qquad \text{(Equation 7.3)}$$

in which K_D is the apparent equilibrium dissociation constant, [DNA] is the concentration of DNA and n is the Hill constant which defines the cooperativity of ligand binding.

Cell cultures

The human non-small cell lung cancer cell lines (A549 and H1299), and the human non-malignant lung fibroblast cell line (MRC-5) were obtained from the American Type Culture Collection (ATCC) and cultured in 75 cm^2 *T*-flasks at 37 °C in a humidified atmosphere containing 5% CO_2. A549 cells were cultured in Ham's F-12 medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% fetal bovine serum (FBS) and 1% streptomycin-penicillin (SP) antibiotic, H1299 in RPMI medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% FBS, 1% SP antibiotic, 2 mM l-glutamine, 10 mM HEPES and 1 mM sodium pyruvate and MRC-5 in EMEM medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% FBS, 1% SP antibiotic and 1% of non-essential amino acids (Sigma-Aldrich, St. Louis, MO, USA).

Metabolic activity assay

Metabolic activity was studied using the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) (Sigma-Aldrich, St. Louis, MO, USA) method. Briefly, 2×10^4 cells/mL were incubated in 100 µL of an appropriate medium in 96-well culture

plates. After 24 h of incubation to allow cell attachment, the medium was removed, and cells were incubated with the appropriate medium containing different ligand concentrations. In initial screening assays, the cells were incubated with 10 μM of each ligand. The most active ligands were treated at concentrations ranging from 0.1 to 100 μM to investigate the concentration effect. After 48 hours of incubation, the medium was replaced with MTT solution and incubated for 2 hours at 37 °C. Then, the MTT solution was removed, and the formazan crystals formed were dissolved in 100 μL of DMSO. Absorbance was measured at 570 nm using a microplate reader Biorad Xmark spectrophotometer (Bio-Rad, California, USA). Cells cultured without ligand concentration (untreated cells) were used as controls. Each experiment was measured in triplicate in three independent experiments. Cytotoxicity was normalized to untreated cells. The IC_{50} of ligands was calculated using GraphPad Prism 8.0 software.

In silico experiments

The structures of all ligands were drawn in ChemBioDraw 20.0 software and the SMILES notation was obtained for each molecule. Lipinski's rule was calculated using the online platform pkCSM [36]. The Molinspiration property engine version 2018.10 was used to calculate the Topological polar surface area (TPSA) (www.molinspiration.com).

Acknowledgments

The authors thank A. Cucchiarini and L. Guittat for helpful discussions.

Funding

Joana Figueiredo acknowledges the doctoral fellowship grant from FCT ref. SFRH/BD/145106/2019. This article/publication is based upon work from COST Action CA 17140 "Cancer Nanomedicine from the Bench to the Bedside" supported by COST (European Cooperation in Science and Technology). Thanks are due to FCT/MCT for the financial support to CICS-UBI ref. UIDB/00709/2020 and UIDP/00709/2020 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. C.C. acknowledges the

grant Instruct-ERIC Pilot R&D application ID 2473. J.L.M., benefited from support from Inserm, CNRS, INCa G4 Access and ANR (ANR-20-CE12-0023) grants.

References

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

[2] Bhattacharyya, D. *et al.* (2016) Metal Cations in G-Quadruplex Folding and Stability. *Front Chem* 4, 38.

[3] Burge, S. *et al.* (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* 34, 5402–5415.

[4] 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.

[5] Biffi, G. *et al.* (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem* 5, 182–186

[6] Patel, D.J. *et al.* (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res* 35, 7429–7455.

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

[8] Biffi, G. *et al.* (2014) Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PLoS One* 9, e102711.

[9] Parkinson, G.N. *et al.* (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417, 876–880.

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

[11] Zhang, Z.-H. *et al.* (2023) In vivo dynamics and regulation of DNA G-quadruplex structures in mammals. *Cell Biosci* 13, 117.

[12] Neidle, S. and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat Rev Drug Discov* 1, 383–393.

[13] Kosiol, N. *et al.* (2021) G-quadruplexes: a promising target for cancer therapy. *Mol Cancer* 20, 40.

[14] Figueiredo, J. *et al.* (2023) Harnessing G-quadruplex ligands for lung cancer treatment: A comprehensive overview. *Drug Discov Today* 28, 103808.

[15] Ou, T.M. *et al.* (2011) Inhibition of cell proliferation by quindoline derivative (SYUIQ-05) through its preferential interaction with c-myc promoter G-quadruplex. *J Med Chem* 54, 5671–5679.

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

[17] Platella, C. *et al.* (2021) Disentangling the Structure–Activity Relationships of Naphthalene Diimides as Anticancer G-Quadruplex-Targeting Drugs. *J Med Chem* 64, 3578–3603.

[18] 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.

[19] Local, A. *et al.* (2018) APTO-253 Stabilizes G-quadruplex DNA, Inhibits MYC Expression, and Induces DNA Damage in Acute Myeloid Leukemia Cells. *Mol Cancer Ther* 17, 1177–1186.

[20] Franceschin, M. (2009) G-Quadruplex DNA Structures and Organic Chemistry: More Than One Connection. *European J Org Chem* 2009, 2225–2238.

[21] Liu, J. *et al.* (2017) Synthesis and the interaction of 2-(1 H -pyrazol-4-yl)-1 H -imidazo[4,5-f][1,10]phenanthrolines with telomeric DNA as lung cancer inhibitors. *Eur J Med Chem* 133, 36–49.

[22] Verga, D. *et al.* (2023) Targeting Quadruplex Nucleic Acids: The Bisquinolinium Saga. In *Handbook of Chemical Biology of Nucleic Acids*, pp. 1–57, Springer Nature Singapore.

[23] De Cian, A. *et al.* (2007) Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J Am Chem Soc* 129, 1856–1857.

[24] Zeng, D.-Y. *et al.* (2017) Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-*b*]quinolone Derivatives as *c-myc* G-Quadruplex Specific Stabilizers via Click Chemistry. *J Med Chem* 60, 5407–5423.

[25] Ou, Z. *et al.* (2019) Novel triazole and morpholine substituted bisnaphthalimide: Synthesis, photophysical and G-quadruplex binding properties. *J Mol Struct* 1185, 27–37.

[26] Moorhouse, A.D. *et al.* (2006) Stabilization of G-Quadruplex DNA by Highly Selective Ligands via Click Chemistry. *J Am Chem Soc* 128, 15972–15973.

[27] Nielsen, M.C. *et al.* (2014) Phenanthroline-2,9-bistriazoles as selective G-quadruplex ligands. *Eur J Med Chem* 72, 119–126.

[28] Craciun, A.-M. *et al.* (2021) New 2,9-disubstituted-1,10-phenanthroline derivatives with anticancer activity by selective targeting of telomeric G-quadruplex DNA. *Spectrochim Acta A Mol Biomol Spectrosc* 249, 119318.

[29] Figueiredo, J. *et al.* (2022) Synthesis and evaluation of 2,9-disubstituted-1,10-phenanthroline derivatives as G-quadruplex binders. *Bioorg Med Chem* 73, 116971.

[30] De Cian, A. *et al.* (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods* 42, 183–195.

[31] De Rache, A. and Mergny, J.-L. (2015) Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* 115, 194–202.

[32] Gueddouda, N.M. *et al.* (2017) Design, Synthesis, and Evaluation of 2,9-Bis[(substituted-aminomethyl)phenyl]-1,10-phenanthroline Derivatives as G-Quadruplex Ligands. *ChemMedChem* 12, 146–160.

[33] Santos, T. *et al.* (2021) G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals* 14, 769.

[34] Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 44, 235–249.

[35] Veber, D.F. *et al.* (2002) Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J Med Chem* 45, 2615–2623.

[36] Pires, D.E. V. *et al.* (2015) pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J Med Chem* 58, 4066–4072.

[37] Figueiredo, J. *et al.* (2024) G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives. *Life Sci* 340, 122481.

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

[39] Arlegui, A. *et al.* (2020) A pH-Switchable Aqueous Organocatalysis with Amphiphilic Secondary Amine–Porphyrin Hybrids. *Eur J Org Chem*, 2020, 4399-4407.

[40] Nielsen, M. C *et al.* (2014) Phenanthroline-2,9-bistriazoles as selective G-quadruplex ligands. *Eur. J. Med. Chem* 72, 119-126.

8th Chapter

Targeting nucleolin by RNA G-quadruplex-forming motif

This chapter was published in:

***Figueiredo, J.**; Miranda, A.; Lopes-Nunes, J.; Carvalho, J.; Alexandre, D.; Valente, S.; Mergny, J-L. and Cruz, C. (2021) Targeting nucleolin by RNA G-Quadruplex forming motif. Biochemical Pharmacology 189, 114418. DOI: 10.1016/j.bcp.2021.114418.*

Chapter overview

Some G4 sequences have the potential to recognize, with high affinity, proteins overexpressed on the cancer cell surface such as the NCL. Therefore, this chapter aims to characterize the pre-MIR150 RNA G4 sequence and evaluate its potential for detecting/recognizing nucleolin expressed in LC and peripheral blood mononuclear cells of LC patients. As far as we know, we have identified and characterize for the first time the presence of G4 structure within human pre-MIR150. We explore its interaction with a well-known G4 ligand, PhenDC3, and determine the affinity of the complex G4/PhenDC3 to NCL.

Targeting nucleolin by RNA G-quadruplex-forming motif

Abstract

A high level of nucleolin (NCL) expression is often associated with a poor prognosis of patients with lung cancer (LC), suggesting that NCL can be used as a possible biomarker. NCL has been shown to display a marked preference for the binding to G-quadruplexes (G4). Here, we investigate the formation of an RNA quadruplex structure in a sequence found in the human precursor pre-MIR150 with the potential to recognize NCL. Circular dichroism (CD) spectra of pre-MIR150 G4-forming sequence (designated by rG4) indicate the formation of a parallel quadruplex structure in KCl or when complexed with the well-known G4 ligand PhenDC3. The thermal stability of rG4 is very high, and further increases in the presence of PhenDC3. The binding affinities of rG4 to PhenDC3 and NCL RBD1,2 are similar with KD values in the nanomolar range. PAGE results suggest the formation of a ternary quadruplex-ligand-protein complex (rG4-PhenDC3-NCL RBD1,2), indicative that PhenDC3 does not prevent the binding of rG4 to NCL RBD1,2. Finally, rG4 can recognize NCL-positive cells and, when fluorescently labeled, can be used as a probe for this protein. ELISA experiments indicate altered NCL expression patterns in liquid biopsies of LC patients in a non-invasive manner, potentially helping the diagnosis, prognosis, and patient response to treatment. Hence, labeled rG4 could be used as a detection probe of LC in liquid biopsies.

Keywords

pre-MIR150 G-quadruplex forming sequence; Nucleolin; Molecular recognition; Lung Cancer; Liquid biopsies

Introduction

Lung Cancer (LC) is the most common cause of cancer death worldwide [1]. LC can be divided into small-cell lung cancer (SCLC) and non–small cell lung cancer (NSCLC),

which is the most common type of LC and the leading cause of cancer-related mortality [2]. Current treatments have relatively low response rates and significant toxicity [3], [4]. Therefore, identification of new targets and biomarkers are needed to improve the early diagnosis and treatment in LC [5].

NCL is the protein most often reported for its biological functions upon nucleic acid recognition and it is mainly present in the nucleolus, but it can be found also in nucleoplasm, cytoplasm and cell membranes [6]. Cell surface NCL has been described as overexpressed in cancer cells as well as a tumor angiogenic marker, including in LC [7]. NCL expression levels were shown to be increased in LC and can positively regulate the autonomous and highly mutagenic genetic element LINE1, promoting genomic instability and LC progression [8]. The phosphorylation of NCL was also reported as a key factor in proliferation and migration of LC cells and the increase of the NCL levels in LC tissues when compared to normal lung tissues is also significantly associated with pathologic state and T status [9, 10].

NCL has been described as an RNA binding protein, but may also recognize DNA structures [11]. Nucleic acid sequences able to form G4 structures have been reported as molecular recognition probes to detect NCL at the cell surface of LC cells [12]. In particular, RNA sequences able to form G4 have been reported as recognition molecular probes to detect NCL in cancer cells [13].

Recent data from multiple studies strongly support the potential of miRNA as potential biomarkers and therapeutic targets for LC [14, 15]. Altered miRNA expression is also associated with tumor progression and survival in LC cancer patients. Several LC miRNA signatures have been proposed to improve molecular staging and classification of LC [16], as in other cancer types: higher expression of MIR150 is found in many solid cancer types, such as gastric cancer, breast cancer and LC [17]. MIR150 promotes the proliferation of LC cells through negative regulation of the pro-apoptotic gene early growth response factor 2 (EGR2) and by targeting p53, respectively, suggesting its pro-tumorigenic function [18]. We identify the guanine-rich motif which can form a G4 structure in the human precursor MIR150 (pre-MIR150). The RNA sequence (5′-GGCCUGGGGGACAGGGACCUGGG-3′) can form a stable G4, allowing recognition by specific molecules such as proteins or small ligands [19]. With respect to miRNA processing, NCL can interact with the microprocessor complex in the nucleus to promote the processing of miRNAs that contain GU-enriched sequences [20]. To date, studies of NCL in LC have focused on p53 translation through binding to both the 5′- and 3′-UTRs of p53 mRNA acting as a therapeutic strategy [21, 22]. The present study aimed to exploit

NCL detection/recognition by the pre-MIR150 G4-forming sequence (designated as "rG4") [23]. To improve NCL detection by rG4, structural studies are required. Herein, we evaluate the formation and stabilization of pre-MIR150 G4-forming sequence and its interaction with NCL was characterized by CD spectroscopy, fluorescence resonance energy transfer (FRET) melting assay, fluorescence titrations and non-denaturing polyacrylamide gel electrophoresis. In parallel, we assessed the co-localization of NCL with rG4 in a LC cell line and in peripheral blood mononuclear cells (PBMCs) of LC patients by confocal microscopy. Also, NCL levels in liquid biopsies of LC patients were determined by an ELISA assay.

Material and methods

Oligonucleotides and ligand

The longer nucleotide sequence (80 nucleotides) representative of the pre-MIR150 (5′-CCAUGGCCCUGUCUCCCAACCCUUGUACCAGUGCUGGGCUCAGACCCUGGUACAG GCCUGGGGGACAGGGACCUGGGGAC-3′) was synthesized by Eurofins Genomics (Germany) with HPLC-grade purification. The G-rich motif (5′-GGCCUGGGGGACAGGGACCUGGG-3′) found in pre-MIR150 unlabelled and labelled (5′-FAM – GGCCUGGGGGACAGGGACCUGGG – TAMRA-3′) were purchased from Eurofins Genomics (Germany) with HPLC-grade purification. The labelled control duplex FdxT (5′-FAM-TATAGCTATA-hexaethyleneglycol-TATAGCTATA-TAMRA-3′) was purchased from Eurofins Genomics (Germany) with HPLC-grade purification. Stock solutions were prepared using nuclease-free water and stored at −20 °C until being used. The oligonucleotide concentrations were measured at 260 nm with a UV–Vis spectrophotometer (Thermo Scientific™ Evolution 220) using the molar extinction coefficients (ε) provided by the manufacturer. For all experiments, the oligonucleotide was annealed by heating to 95 °C for ten more min and cooling on ice for more 10 min.

Recombinant NCL RBD1,2 (20.3 kDa) was purchased from NZYtech (Portugal/project number: S181101_1). 3,3′-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate (PhenDC3) was obtained from Sigma-Aldrich (USA) with HPLC grade purification and purity ≥ 97%. Stock solutions of PhenDC3 were prepared as 10 mM solutions in DMSO (Thermo Fisher Scientific, USA) and its subsequent dilution was done using nuclease-free water.

In silico analysis of *pre-MIR150* G4-forming sequence

To identify the G4-forming sequence, we have performed a bioinformatic analysis in pre-MIR150. The pre-MIR150 sequence was obtained from miRBase version 22.1 [24]. The analysis was carried using GQRS-mapper [25]. The motif found in pre-MIR150 that forms the G4 structure matches this consensus.

CD spectroscopy

Circular dichroism spectra were acquired in a Jasco J-815 spectrometer (Jasco, USA), using a Peltier temperature controller (model CDF-426S/15). The longer nucleotide sequence (80 nucleotides) representative of the pre-MIR150 at 5 μM was dissolved in 0.5 mM of lithium cacodylate (Sigma-Aldrich, USA), annealed as previously described and titrated with increase KCl concentrations (PanReac, Spain). The motif (5′-GGCCUGGGGGACAGGGACCUGGG-3′) found in pre-MIR150 (rG4) at 10 μM was dissolved in 0.5 mM lithium cacodylate, annealed and titrated with KCl until 100 mM.

For the titration experiments with PhenDC3, the required volume of PhenDC3 (until 1 M equivalent) was added directly to 10 μM of pre-folded rG4 solution dissolved in 0.5 mM lithium cacodylate and 100 mM KCl in the quartz cuvette (Hellma, Germany) of 1 mm path-length. Spectra were acquired through wavelengths ranging from 200 to 340 nm, with a scan speed of 100 nm/min, 1 nm bandwidth, 1 s integration time over 3 averaged accumulations.

For CD melting experiments, rG4 solutions were prepared in 0.5 mM lithium cacodylate with increasing KCl concentrations (10 μM to 100 mM) in order to obtain a melting temperature around 50 °C. For ligand titration, the required amount of PhenDC3 solution (until 1:1 molar ratio) was added to the cuvette containing pre-annealed rG4. The denaturation process was analyzed by monitoring the wavelength of maximum ellipticity (264 nm) through temperatures ranging from 20 to 100 °C (temperature set by the Peltier system) with a heating rate of 2 °C/min. Data were converted into folded fraction (θ) plots according to Equation (1), fitted to a Boltzmann distribution using OriginPro2016.

$$\theta = \frac{CD - CD_{\lambda}^{min}}{CD_{\lambda}^{max} - CD_{\lambda}^{min}} \quad (1)$$

The melting temperature (*T*m) was determined from a two-state transition model where CD is the ellipticity at 264 nm at each temperature and CD^{min} and CD^{max} are the lowest and highest ellipticities, respectively.

Fluorescence titrations experiments

Fluorescence titrations were performed on a FluoroMax4 spectrofluorimeter (Horiba, Japan) using high-precision quartz suprasil cuvettes with light paths of 10 × 4 mm (Hellma, Germany). For these experiments, rG4 at 1 μM was pre-annealed in 20 mM phosphate buffer with 100 mM of KCl (PanReac, Spain). The titration was performed by adding 2 M equivalents of PhenDC3 at 2 μM and NCL RBD1,2 until 4 M equivalents (4 μM), followed by 10 min for equilibration. All spectra were acquired with an integration time of 0.5 s, an emission and excitation slit fixed at 2 nm and step size of 1 nm, averaged over three scans. The labeled rG4 was excited at 495 nm and the emission spectra were collected between 510 and 700 nm. The obtained fluorescence data were converted into a fraction of ligand-bound (α) plots according to Equation (2), where I is the fluorescence intensity at maximum wavelength of rG4 at each NCL:RNA ratio, and I_{free} and I_{bound} are the fluorescence intensities of the free and fully bound ligand, respectively.

$$\alpha = \frac{I - I_{\lambda}^{free}}{I_{\lambda}^{bound} - I_{\lambda}^{free}} \quad (2)$$

Additionally, the data points were then fitted to a two-site binding model function, using the OriginPro 2016 software, and K_D values were determined from the saturation binding model described in Equation (3), where α is the fraction of ligand-bound and [rG4] is the concentration of rG4:

$$\alpha = \frac{[\mathrm{rG4}]}{K_{D1} + [\mathrm{rG4}]} + \frac{[\mathrm{rG4}]}{K_{D2} + [\mathrm{rG4}]} \quad (3)$$

Förster resonance energy transfer (FRET) melting assays

FRET melting experiments were performed in a 96-well plate using a CFX Connect™ Real-Time PCR Detection System (Bio-Rad, USA). The rG4 was purchased labeled with fluorescein (FAM) and rhodamine (TAMRA) at the 5′ and 3′ ends, respectively. The labeled rG4 at 0.2 μM was dissolved and annealed in 10 mM lithium cacodylate pH 7.2 (Sigma-Aldrich, USA) supplemented with 1 mM of KCl (PanReac, Spain) and 99 mM LiCl (Sigma-Aldrich, USA). The samples were prepared by aliquoting 15 μL of rG4 solution into each strip and 5 μL of PhenDC3 solution at 0.2 μM. Subsequently, samples were incubated for 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 excitation and emission were set to 495 and 518 nm, respectively. Each experimental

condition was tested in triplicate on at least two separate plates. The melting temperature was determined from normalized curves as the mid-transition $T_{1/2}$ temperature.

Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gel (20%) was prepared by diluting 40% acrylamide-bisacrylamide (19:1, GRiSP Lda, Portugal) in water and adding 10 mM KCl (PanReac, Spain). TBE 1× (89 mM Tris-base, Fisher Scientific, USA, 89 mM boric acid, ChemLAB, Belgium and 1 mM EDTA, PanReac, Spain), 0.1% ammonium persulfate (LabChem, USA) and 0.1% tetramethylenediamine (Acros, USA).

rG4 at 20 μM in 20 mM phosphate buffer with 100 mM of KCl was incubated at 2:1 PhenDC3:rG4 ratio for complex formation. Different concentrations of NCL RBD1,2 (0.002, 4, 12, 20, 50 and 60 μM) were incubated during 30 min with rG4 before electrophoresis. Before loading on the gel, the samples were mixed with 8 μL of 50% sucrose (Sigma-Aldrich, USA). Oligonucleotide markers were purchased from NZYTech (9, 15, 21, 30, 60 and 90 nucleotides) and loaded in parallel on the gel. Electrophoresis was performed in a vertical electrophoretic cell Mini-Protean II (Bio-Rad, USA) with a Biometra P25 power supply machine at 25 °C in TBE buffer (Sigma-Aldrich, USA). Gels were run at 150 V and 22 mA for 2 h. After electrophoresis, the gels were stained by Stains-All solution (Sigma-Aldrich, USA) overnight followed by discoloration in water under the light before visualization on GE Typhoon Trio Imager Scanner (GE Healthcare, USA).

Confocal microscopy

To understand the ability of the rG4 to bind NCL in vitro, the A549 lung cancer cell line (ref. 86012804-1VL, Sigma Aldrich, USA) and PBMCs isolated from LC patients were used in confocal microscopy studies. The PBMCs were seeded at a density of 1 × 106 cell/well in treated μ-slide 8 well plates (IBIDI, Germany). The passage number of the A549 lung cancer cell line was 3 and the cells were seeded at a density of 105 cell/mL in treated μ-slide 8 well (IBIDI, Germany). Cells were grown at 37 °C in a 5% CO2 humidified atmosphere. After 24 h, cells were treated with 1 μM of pre-annealed rG4, labeled with fluorescein (FAM) and rhodamine (TAMRA) at the 5′ and 3′, respectively, and with 2 μM of PhenDC3 for 1 h. Next, NCL antibody (PA3-16875, Invitrogen, USA; dilution of 1: 100) was added during 2 h followed by a secondary antibody (Alexa Fluor 647®, A21244, Invitrogen™, USA; dilution of 1: 1000) for 1 h. The nuclei of the cells were stained with the 1 μM nuclear probe Hoechst 33,342 (H3570, Invitrogen™ Thermo Fisher Scientific, Portugal) for 15 min. The probe was washed by rinsing with PBS

(Sigma-Aldrich, USA) three times and the cells were imaged using a Zeiss AxioObserver LSM 710 microscope (Zeiss, Germany). Images were processed with the ImageJ software.

Peripheral blood mononuclear cells isolation

The study included 31 patients previously diagnosed with LC in different stages with and without treatment. The control group included 32 healthy male and female volunteers under the age of 35 with no history of cancer disease. All tested subjects gave written informed consent. Approximately 10 mL of resh human blood samples were collected in ethylenediaminetetraacetic acid (EDTA)-coated tubes (Sigma-Aldrich, USA) following the protocol approved by the Ethical Committee of University Hospital Center Cova da Beira (study number 35/2019). The isolation of PBMCs was performed as follows. Briefly, the blood samples were centrifuged at 3000 rpm (Beckman Coulter Allegra X-22R, SX4250 rotor) for 15 min, at room temperature. After centrifugation, the plasma fraction was removed and an equal volume of PBS (Sigma-Aldrich, USA) was added to the tube, followed by gentle mixing by inverting the tube several times. The mixture was then gently transferred into a 50 mL polypropylene tube containing an equal volume of Biocoll (VWR, Portugal) separating solution and centrifuged at 2200 rpm for 30 min at room temperature. The PBMCs (ring) formed in the interphase between Biocoll and blood fraction was aspirated using a plastic Pasteur pipette and transferred to a new 50 mL polypropylene tube. Then, PBS was added to the 50 mL polypropylene tube and the mixture was centrifuged at 1800 rpm for 10 min at 4 °C. After discarding the supernatant, the pellet was resuspended in 10 mL of PBS and centrifuged again at 1500 rpm for 10 min at 4 °C. The supernatant was discarded, the pellet resuspended in 10 mL of pre-warmed RBC lysis solution and gently stirred for 10 min at 37 °C. The sample was then centrifuged at 1500 rpm for 10 min at room temperature. The resulting pellet was then washed two times with PBS by centrifugation at 1500 rpm for 10 min at room temperature. Finally, the PBMCs were resuspended in 1 mL of PBS containing 10% DMSO and stored at −80 °C until further use.

Enzyme-linked immunosorbent assay (ELISA)

ELISA was performed to quantify NCL protein in the PBMCs of both LC patients and control subjects. Prior to the assay, cell lysates from 1 × 106 human PBMC cells/sample were obtained by centrifugation at 13000 rpm for 5 min at 4 ˚C followed by resuspension in 20 µL of RIPA lysis buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS; ref. 89900, Sigma Aldrich, USA) and incubation on ice

for 10 min. The NCL levels in the cell lysates were determined with a commercially available ELISA kit (ref. CSB-EL015535HU; CUSABIO, USA) according to the manufacturer's protocol. The sensitivity of the assay for NCL was 7.81 pg/mL as per manufacturer's product data-sheet. All samples and standards were tested in triplicate in four different plates. The plate was measured spectrophotometrically at a wavelength of 450 nm, with the correction wavelength set at 540 nm. The absorbance of the RIPA lysis buffer was subtracted to the readings of the test samples. The concentration of NCL in the samples was determined by comparing the absorbance of the samples to the standard curve. The standard curve was obtained by plotting NCL concentration on the x-axis versus absorbance on the y-axis and the data points fitted to a nonlinear regression.

Results

pre-MIR150 G4-forming sequence

rG4 formation was investigated using the online QGRS Mapper tool (https://bioinformatics.ramapo.edu/QGRS/analyze.php). The predicted rG4 consists of 4 guanylic acid (G) stretches, which are located 59–82 nucleotides (nt) from the 5′end (**Figure 8.1**).

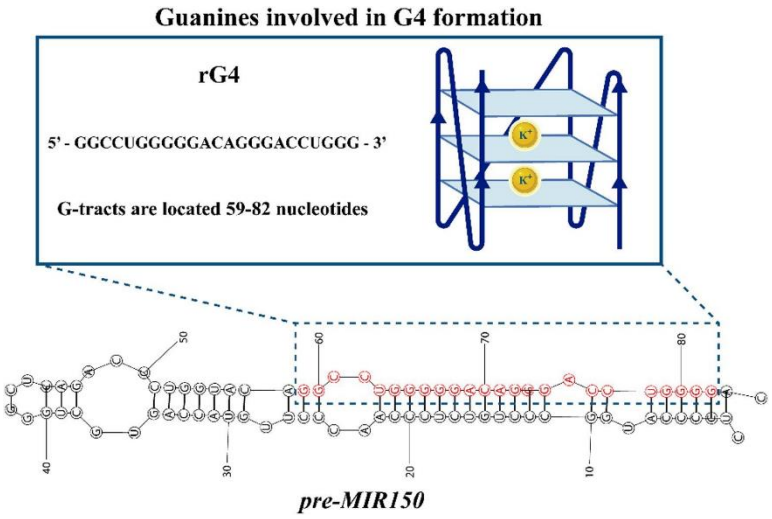


Figure 8.1. Secondary structure model for the human *pre-MIR150* using in silico analysis.

CD spectroscopy

CD spectroscopy was first performed using a longer nucleotide sequence (80 nucleotides) representative of the *pre-MIR150* to confirm G4 formation. The CD spectra presented in **Figure 8.2** are typical for a parallel G4 topology, displaying a positive band at 264 nm and a negative band at 240 nm.

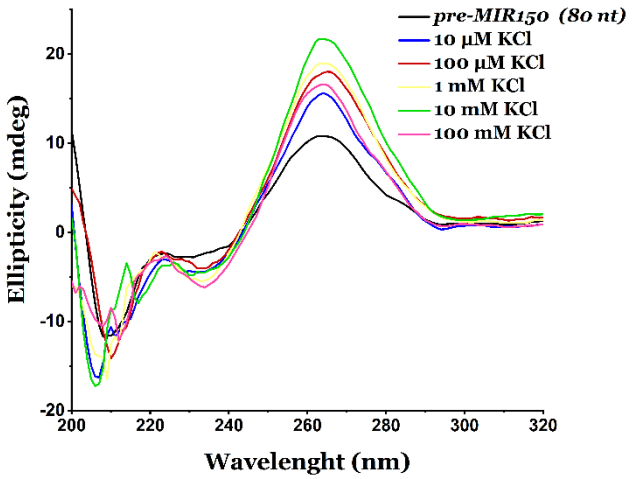


Figure 8.2. CD titration of longer nucleotide sequence (80 nucleotides) representative of the pre-MIR150 in 0.5 mM lithium cacodylate upon increasing KCl concentration from 10 µM to 100 mM.

The G4-prone motif in pre-MIR150, identified by in silico analysis, was then assessed by CD spectroscopy to confirm G4 formation. This motif was designated as rG4 (or *pre-MIR150* G4-forming sequence). The CD spectra of the rG4 (**Figure 8.3**) indicate also the formation of a parallel quadruplex, whether it was annealed in 0.5 mM lithium cacodylate containing variable concentrations of KCl or complexed with 10 µM PhenDC3. Of note, the CD spectra remain nearly identical at 100 µM and 100 mM KCl, showing that minimal amounts of K^+ are required to support G4 formation, suggesting a very stable RNA G4.

The effect of salt concentration and PhenDC3 binding on the thermal stability of rG4 was determined by CD melting and presented in **Figure 8.4**. The T_m of rG4 in presence of 10 µM and 100 mM of KCl is 54.2 °C and 66.3 °C, respectively. Upon addition of 1 equivalent of PhenDC3, a clear increase in T_m value (ΔT_m = 17.6 °C) was observed, indicating ligand stabilization of the G4 structure.

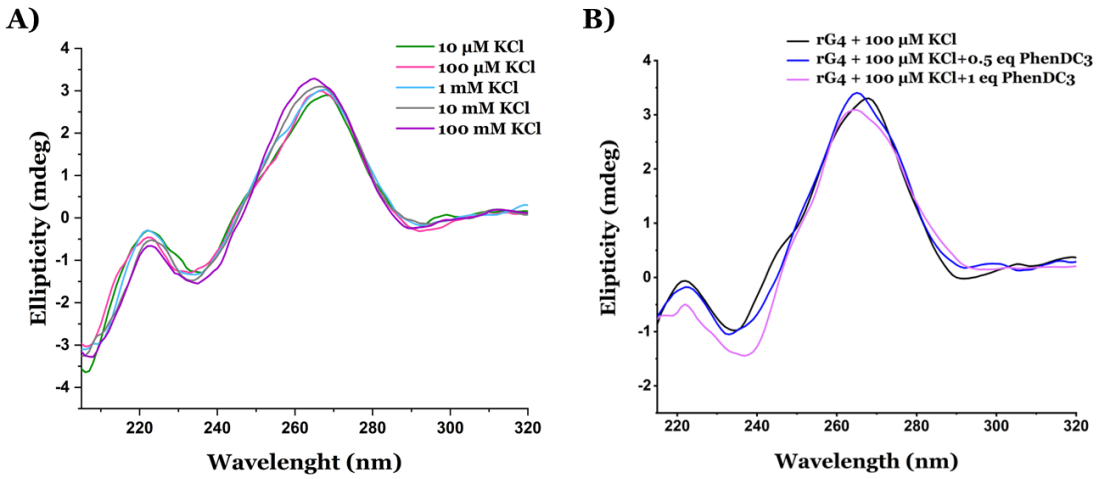


Figure 8.3. **A)** CD titration of rG4 in 0.5 mM lithium cacodylate with increasing concentrations of KCl from to 10 µM to 100 mM. **B)** CD titration spectra of rG4 in 0.5 mM lithium cacodylate containing 100 µM of KCl with increasing concentrations of PhenDC3 until 1 M equivalent.

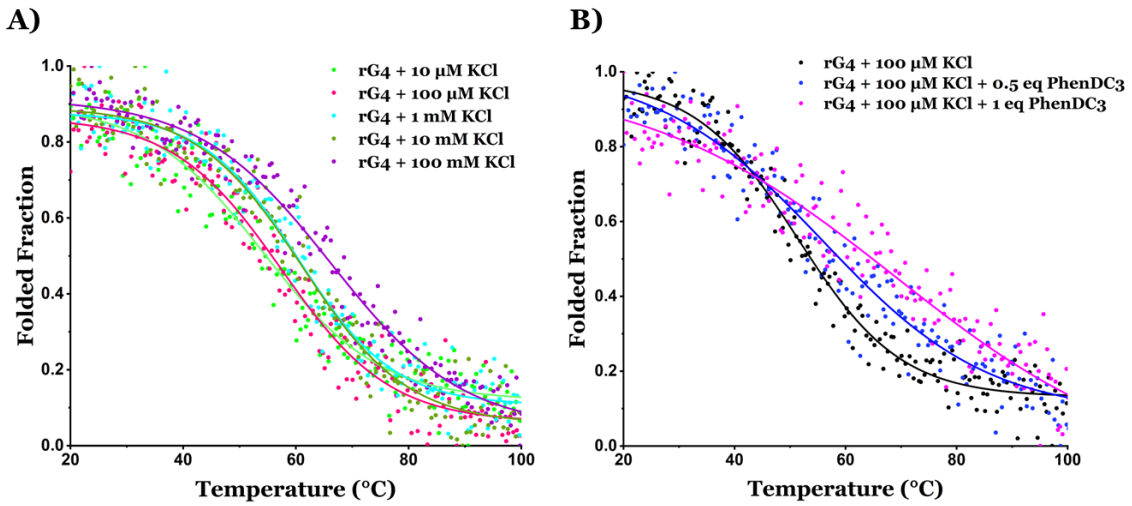


Figure 8.4. **A)** CD melting curves of rG4 in 0.5 mM lithium cacodylate with increasing concentrations of KCl from 10 µM to 100 mM. **B)** CD melting curves of rG4 in 0.5 mM lithium cacodylate in the presence of increasing molar eq. of PhenDc3 until 1M equivalent. Data points were recorded at 264 nm. The Boltzman curve fitting of the data points is shown.

Fluorescence titrations experiments

Following CD characterization, fluorescence spectroscopy was used to determine the K_D values of PhenDC3 for the pre-folded rG4. K_D constants were determined from the variation of the fluorescence intensities. Fluorescence titration curves and K_D values are presented in **Figure 8.5**.

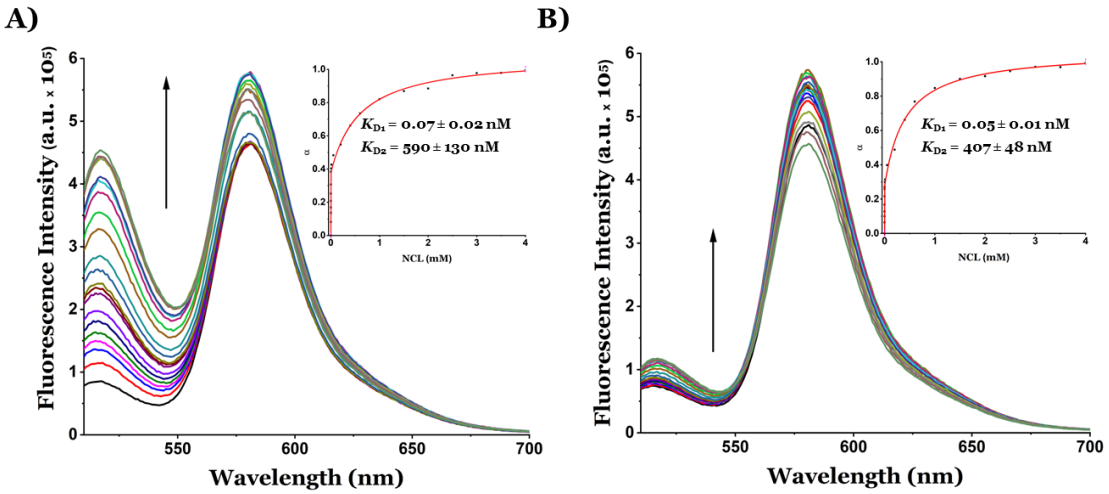


Figure 8.5. **A)** Fluorescence emission spectra of rG4 at 1 µM with increasing concentration of NCL RBD1,2 until 4 M equivalents; **B)** Fluorescence emission spectra of rG4-PhenDC3 (1:2) with increasing concentration of NCL RBD1,2 until 4 M equivalents. Spectra were acquired in 20 mM potassium phosphate buffer containing 100 mM KCL pH 6.9.

For this experiment, fluorescence emission spectra were recorded using the rG4 sequence labeled with 5′-FAM and 3′-TAMRA in 20 mM potassium phosphate buffer supplemented with 100 mM KCl. Upon addition of 2 equivalents of PhenDC3 and NCL RBD1,2, we observed an enhancement in the fluorescence emission. These data allowed calculating the fold increase in fluorescence units at NCL RBD1,2 saturation, which were 2-fold and 5-fold increase in fluorescence units in the presence and absence of PhenDC3, respectively. The data were fitted by a non-linear regression analysis using a two-site bind equation (Equation (3)) to obtain the saturation binding plots. The analysis (**Figure 8.5**) revealed that NCL RBD1,2 binds to rG4 with K_{D1} = 0.07 ± 0.02 nM and K_{D2} = 590 ± 130 nM and to rG4-PhenDC3 conjugate with roughly similar values (K_{D1} = 0.05 ± 0.01 nM and K_{D2} = 407 ± 48 nM). These results show that PhenDC3 does not affect the binding between rG4 and NCL RBD1,2. Similarly, the experiments were also performed with a duplex-forming FdxT: 5′-FAM-TATAGCTATA-hexaethyleneglycol-TATAGCTATA-TAMRA-3′ labeled sequence. The linear relationship for the association of NCL with the duplex sequence, suggesting weak binding with association constant in the 10^{-4} nM range in the absence and presence of PhenDC3, respectively (data not shown).

Evaluation of rG4-PhenDC3 complex stability by FRET-melting assay

FRET melting of labeled rG4 with a fluorophore (5′-FAM) and a quencher (3′-TAMRA) was initially studied through fluorescence-monitored thermal denaturation

experiments. Interestingly, and in agreement with the CD measurements, rG4 forms a highly stable structure with a melting temperature of 69.1 °C in the presence of 1 mM KCl. This G4 becomes even more thermally stable in the presence of PhenDC3 (2 μM), as the T_m increases from 69.1 to at least 87.5 °C ($\Delta T_{1/2} \geq 18.4$ °C; the exact value being hard to determine as melting is only partial at the maximal temperature) (**Figure 8.6**).

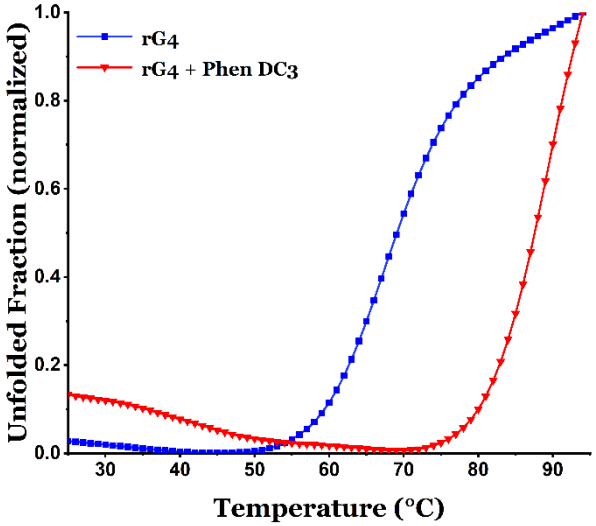


Figure 8.6. Normalized FRET melting curves of rG4 (1 mM of KCl and 99 mM LiCl) in the presence and absence of PhenDC3.

Gel electrophoresis analysis

Next, we used native polyacrylamide gel electrophoresis to characterize rG4 formation in the absence and presence of PhenDC3 and NCL RBD1,2 and to discriminate molecularity of rG4 and complexes present in solution.

Appropriate reference oligonucleotides were used as a ladder for comparison. The rG4 was diluted to 20 μM and annealed in 20 mM phosphate buffer with 100 mM of KCl, and complexes with PhenDC3 were prepared at 1:2 mM ratio. NCL RBD1,2 was added to each rG4-PhenDC3 sample at different concentrations (0.002, 4, 12, 20, 50 and 60 μM). The native 20% polyacrylamide gel is presented in **Figure 8.7**.

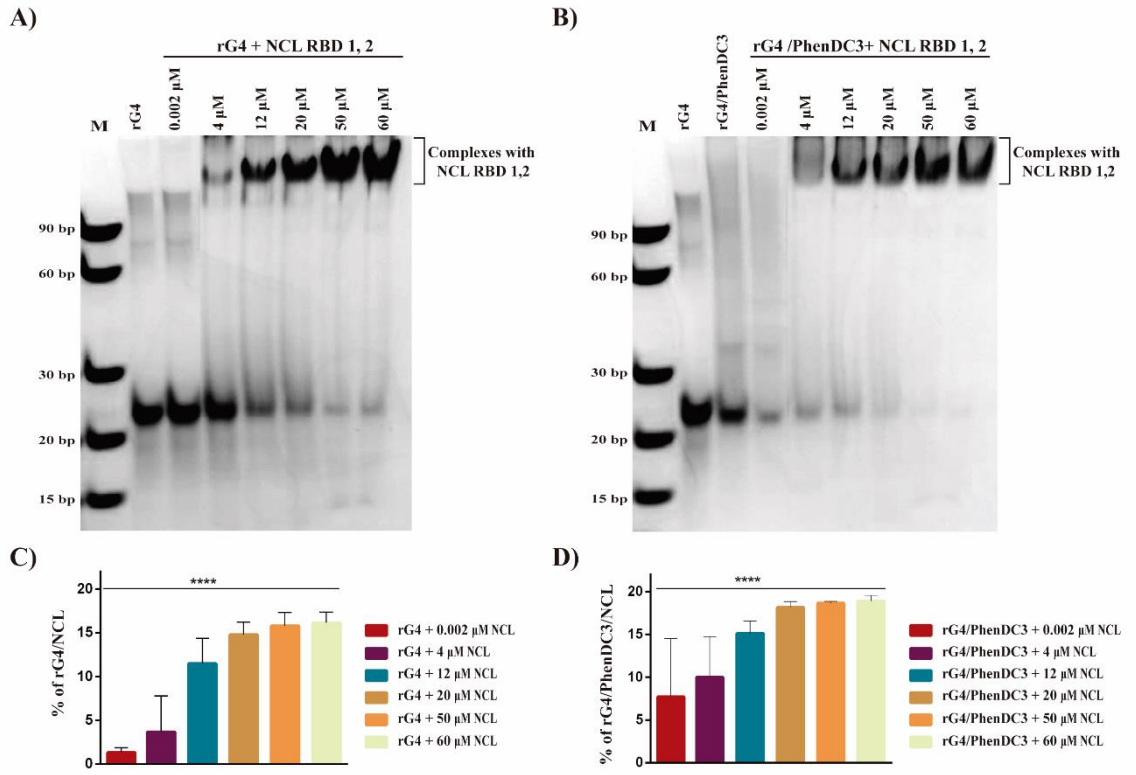


Figure 8.7. Native gel electrophoresis containing rG4 complexes formed between **A)** NCL RBD1,2 and **B)** PhenDC3-NCL RBD1,2. The samples were loaded on a non-denaturing 20% acrylamide gel supplemented with 10 mM KCl. Migration was performed at 20 °C. Oligothymidylates (15, 21, 30, 60 and 90 nucleotides) were used as markers. After electrophoresis, the gels were stained and visualized on GE Typhoon Trio Imager Scanner (GE Healthcare, USA). **C)** Percentage of rG4-NCL RBD1,2 complexes, formed in the presence of different concentrations of protein (0.002, 4, 12, 20, 50 and 60 µM). **D)** Percentage of rG4-PhenDC3-NCL RBD1,2 complexes, formed in the presence of PhenDC3 (40 µM) and different concentrations of NCL RBD1,2 (0.002, 4, 12, 20, 50 and 60 µM). Three independent experiments were performed. Data are shown as mean with a 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 PhenDC3 and between the NCL RBD1,2 concentrations (the asterisks indicate $p < 0.0001$).

The lane 2 of **Figure 8.7A** displayed a single band which migrated at a position equivalent to 23 nucleotides, corresponding to rG4, compatible with a monomolecular structure. Now, focusing on the upper bands of lanes 4–8 of **Figure 8.7A**, we observe that rG4 got stuck up at the top of the well in these lanes, as shown by a darker region at the edge of the well. This suggests the formation of higher-order structures involving the RNA and the protein. We also observed a band that migrated close to 23 nucleotides, indicating the presence of free rG4, and its intensity decreased with increasing concentrations of NCL RBD1,2. Next, we checked the migration of rG4 in the presence of PhenDC3. The lane 3 of **Figure 8.7B** shows two bands with a marked difference in electrophoretic mobility indicative of a different structuring. This observation indicates

the formation of a rG4-PhenDC3 complex, in which the PhenDC3 stabilizes the rG4 structure and appeared in the form of a higher-order complex which is consistent with the CD and FRET-melting data. Lanes 5–9 of **Figure 8.7** showed a retarded band at the edge of the well, more pronounced with increasing concentrations of NCL RBD1,2, attributable to the formation of a ternary complex, involving rG4, PhenDC3 and NCL RBD1,2. The lower band intensity corresponding to the rG4 form of the oligonucleotide decreased when the concentration of NCL RBD1,2 increased.

Confocal microscopy analysis

To investigate the *in vitro* binding potential of the rG4 to NCL, lung cancer A549 cells were analyzed by confocal imaging (**Figure 8.8**).

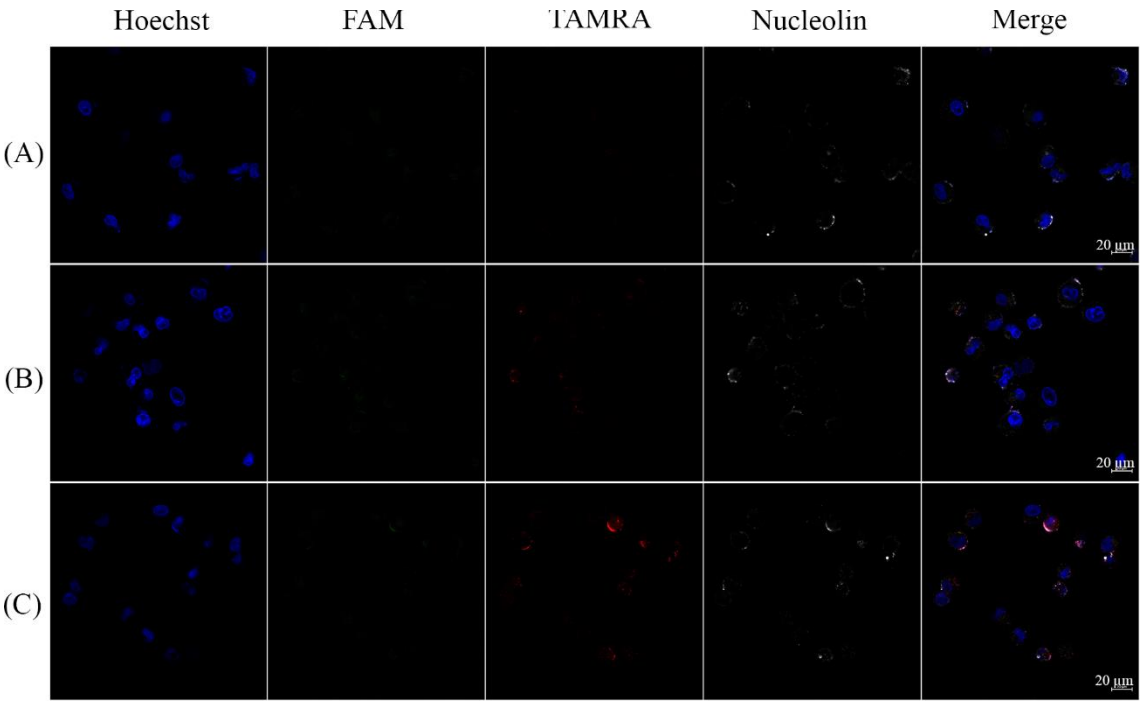


Figure 8.8. NCL immunofluorescence in A549 cell line **(A)** without rG4, **(B)** with rG4, labeled with fluorescein (5′-FAM, green) and rhodamine (3′-TAMRA, red) and **(C)** rG4 labeled conjugate with PhenDC3. The anti-NCL primary antibody was conjugated with an Alexa Fluor 647 secondary antibody (grey). Scale bar: 20 µm.

The A549 cell line presents high levels of NCL expression [26]; if NCL acts as a receptor for rG4 on A549 cells, then we would expect the rG4 to be colocalized at the cell surface. For that purpose, the cells were incubated with annealed rG4 labeled with 5′-FAM and 3′-TAMRA, and its localization in the cell line was analyzed through the intrinsic fluorescence of FAM. The primary anti-NCL antibody conjugated with the secondary antibody AlexaFluor 647® was used to localize cell surface NCL. As observed in **Figure**

8.8, NCL was located at the surface of the cancer cells. Moreover, a fluorescent signal was observed with the incubation of annealed rG4 and PhenDC3. Additionally, confocal microscopy indicated that rG4 and complex rG4-PhenDC3 colocalize with NCL, suggesting the ability of this rG4 to recognize and bind NCL on the surface of A549 cells, even if bound to the PhenDC3 G4 ligand. The results also confirmed the internalization of rG4 and its location in the cytoplasm.Relatively to the cells incubated with the rG4 (without PhenDC3) low fluorescence was observed on the surface of the cells, suggesting that PhenDC3 is important for the rG4 stabilization and maintenance to improve the in vitro NCL recognition.

Similar experiments were performed with PBMCs isolated from blood samples of LC patients. As previously observed with A549 cells, NCL was expressed at the surface of these cells as observed in **Figure 8.9**.

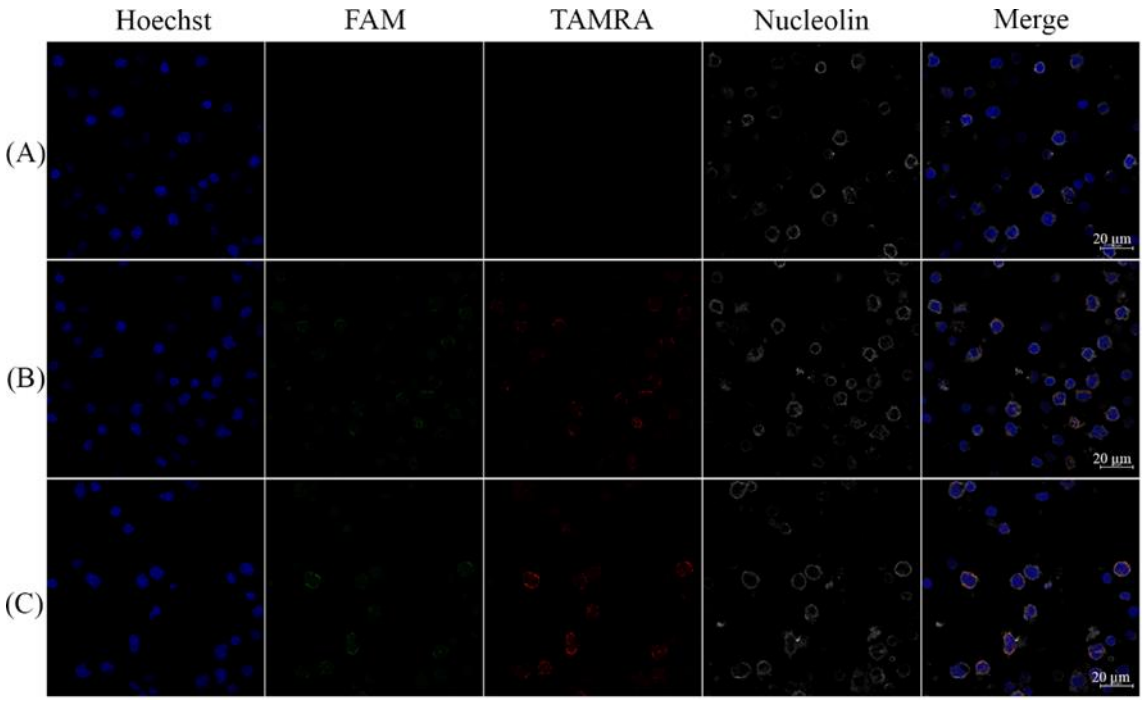


Figure 8.9. NCL immunofluorescence in PBMCs from LC patient **(A)** without rG4, **(B)** with rG4, labeled with fluorescein (5′-FAM, green) and rhodamine (3′-TAMRA, red) and **(C)** rG4 labeled conjugate with PhenDC3. The anti-NCL primary antibody was conjugated with an Alexa Fluor 647 secondary antibody (grey). Scale bar: 20 µm.

After incubation with rG4 and complex rG4-PhenDC3, the rG4 was able to colocalize with NCL, suggesting that it is able to recognize this protein in these cells. Moreover, for the cells incubated with the complex rG4-PhenDC3, the rG4 fluorescence was higher, suggesting that PhenDC3 is important for rG4 stabilization and improved NCL recognition.

NCL expression in LC biopsies by ELISA experiments

To evaluate the expression levels of NCL in patients with LC and the potential correlation of different NCL expression levels with the histological group of LC patients, we conducted ELISA experiments using PBMCs isolated from blood samples of both LC patients and healthy subjects. When compared to tissue biopsies, liquid biopsies are attractive as these are less invasive, easier to obtain and with less pain and risk associated for the patient and allow stratification and monitoring of ongoing therapies. The blood samples were obtained from 31 LC patients (19 men and 12 women, mean age 60 years), who fall into the different histological categories of this disease, together with samples from 32 healthy subjects (18 men and 14 women, mean age 35 years), following CHUCB Ethics Committee recommendations. This study comprised 27 patients (87.1%) with NSCLC, the most common malignant tumors, and 4 patients (12.9%) with Small Cell Lung Cancer (SCLC). The results of the ELISA experiments are shown in **Figure 8.10**.

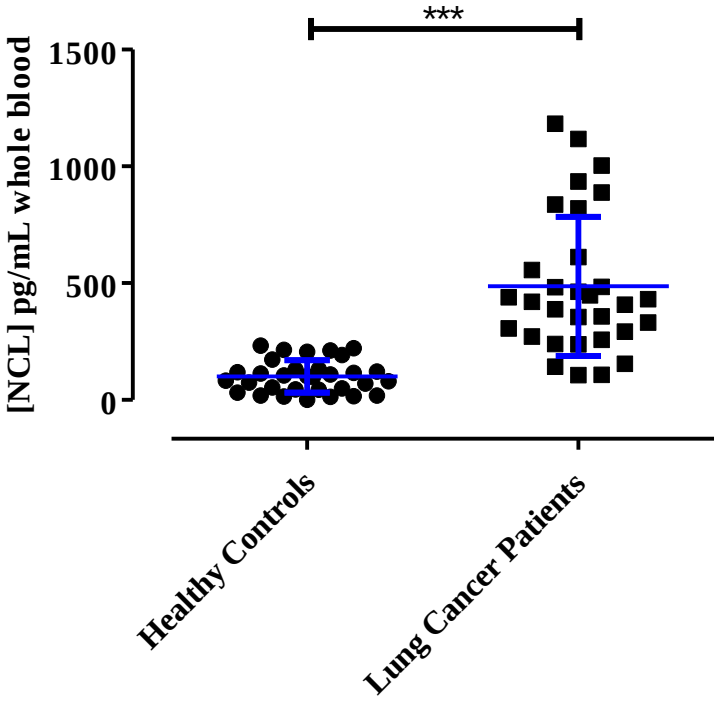


Figure 8.10. Scatter plot representation of ELISA values in LC patients (n = 31) compared to healthy controls (n = 32). Blue bars indicate the corresponding mean ± SD; *P-value* < 0.0001.

The analysis of the scatter plot suggests that there is a significant increase in NCL expression in the PBMCs of LC patients (mean expression of 486 ± 297 pg/mL) when compared to healthy controls (mean expression of 100 ± 69 pg/mL) that is around 5-fold mean expression increase of NCL in the PBMCs of LC patients showing the potential of using NCL as a biomarker for diagnosis. The same LC patients were then divided into

categories taking into account the LC histological type, namely adenocarcinoma, squamous cell carcinoma, large-cell undifferentiated carcinoma and SCLC. This was done as an attempt to determine whether there was any phenotypic variation in terms of NCL expression in PMBCs of different types of LC, which could be used as specific biomarker for a given LC type. The scatter plot showing the NCL levels by histological category is shown in **Figure 8.11**.

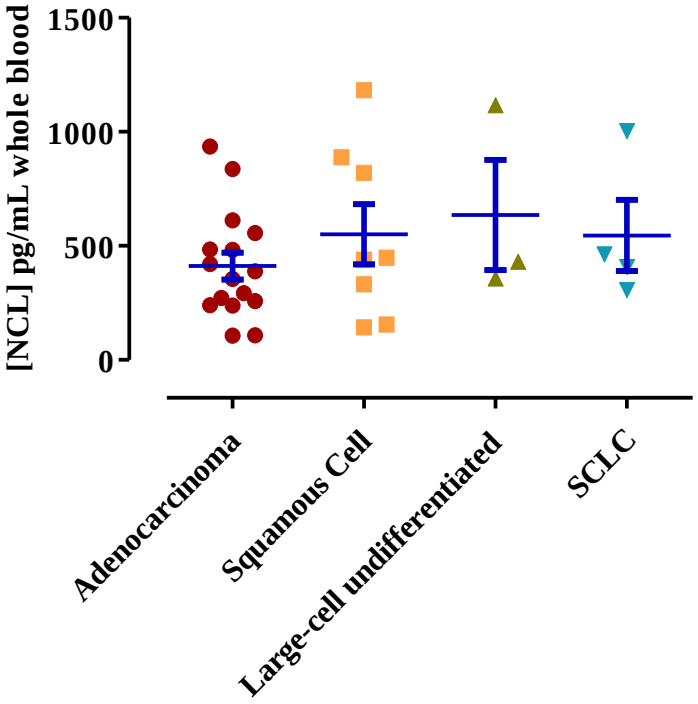


Figure 8.11. Scatter plot representation of ELISA values in the histological category of LC patients namely adenocarcinoma (n = 16), squamous cell carcinoma (n = 8), large-cell undifferentiated carcinoma (n = 3) and SCLC (n = 4). Blue bars indicate the corresponding mean ± SD.

Discussion

Previous studies have shown the presence of G4 structures in precursor pre-miRNA [27] and these structures can interact with specific proteins, acting as sensing probes. The G4 structures presented in pre-miRNAs coexist in equilibrium with the stem-loop structures, however, the transition between a hairpin and G4 structures can be influenced by salt addition or small molecules [28-30].

The present study exploits the used of pre-MIR150 G4-forming sequence as a probe for detecting NCL, which is overexpressed in LC cells and contains RNA-binding domains

that mediate the interaction with the G4 structures [11]. Using in silico analysis, we found a G-rich sequence in pre-MIR150 that can fold into G4. After that, CD spectroscopy confirmed G4 formation with a parallel topology. This motif of 23 nucleotides (rG4) was selected to study the NCL recognition. We demonstrated that G4 formation can be also mediated by the well-known G4 binder PhenDC3, which stabilizes rG4 by 17.6 °C (**Figure 8.4B**) and does not affect the parallel topology after ligand binding (**Figure 8.3B**). These results were confirmed by FRET melting studies that indicated a significant G4 stabilization ($\Delta T_{1/2}$ = 18.4 °C) at a PhenDC3: rG4 ratio of 10: 1 (**Figure 8.6**).

The assessment of binding characteristics of rG4 towards PhenDC3 and NCL RBD1,2 was determined by fluorescence titrations, and complex rG4-PhenDC3 formation with NCL RBD1,2 is consistent with a two-site binding model [31].

The binding constants obtained from fitting are in the nanomolar range and showed a slight increase in affinity for NCL RBD1,2 binding to the rG4-PhenDC3 complex comparing with free rG4, suggesting that PhenDC3 binding does not prevent NCL RBD1,2 recognition.

Focusing on fluorescence emission spectra (**Figure 8.5**), we observed an increase in fluorescence, resulting from the interaction of rG4 with NCL RBD1,2 in the absence of PhenDC3. The highest fluorescences may be due to a conformational change of rG4 in order to recognize NCL RBD1,2 which leads to an increasing in the distance among fluorophores disrupting the FRET-phenomenon. To the contrary, in the presence of PhenDC3, a stabilization effect on G4 is observed, resulting in the maintenance of the secondary structure and, consequently, conservation of the fluorophores distance and low fluorescence emission. Overall and considering the K_D values, it is verified that the rG4-PhenDC3 complex promotes the interaction with NCL RBD1,2. The experiments using a non-G4 sequence demonstrated that the binding of NCL is indeed G4-mediated as the binding constants were suggestive of weak binding. Indeed, NCL has been described before as a potent G4-binding protein [32].

The PAGE experiment presented in **Figure 8.7** shows the molecularity of rG4, rG4-PhenDC3 and rG4-PhenDC3-NCL RBD1,2 complexes, consistent with CD, FRET-melting and fluorescence titrations. In addition, the rG4 and complexes formed additional bands suggesting that are stable under non-denaturing conditions. The upper band found in the gel corresponding to higher molecular weight is consistent with the complex rG4-PhenDC3-NCL RBD1,2 and its dependency on NCL RBD1,2 concentration appears to be more pronounced in the presence of PhenDC3. Indeed, the highest affinity of rG4 and stability towards NCL RBD1,2 was demonstrated in the presence of PhenDC3.

These evidences also suggest that PhenDC3 can be used to stabilize the G4 without affecting the binding to NCL RBD1,2 and improving its recognition through G4 structure. We showed by confocal microscopy that NCL is expressed at the surface of A549 cells and in PBMCs of LC patients. Additionally, annealed rG4 labeled with 5′-FAM and 3′-TAMRA with PhenDC3 localize with NCL observed by the overlapping of 3′-TAMRA (green) and 5′-FAM (red) suggesting that the complex is still formed intracellularly (purple) (**Figure 8.8**, **Figure 8.9**). We conclude from these observations that rG4 can recognize NCL-positive cells and, when fluorescently labeled, can be used as a probe for the detection of this protein.

ELISA experiments were performed using blood samples of both LC patients and healthy cancer-free volunteers to monitor NCL expression in biopsy liquids and ascertain the potential of using rG4 as a molecular recognition probe to detect NCL in biopsy liquids. NCL levels were already shown to be increased in NSCLC tissues when compared to normal lung tissues, and this overexpression was significantly associated with pathologic stage and T status [10]. However, to the best of our knowledge, NCL expression and its clinical relevance in liquid biopsies has been overlooked until now. The results of the ELISA experiments with 31 LC patients belonging to different histological groups presented in **Figure 8.10** indicated an increase in the expression of NCL when compared to the healthy controls (around 5-fold mean expression increase). These preliminary results suggest that liquid biopsies may be used to detect altered NCL expression patterns in LC patients in a non-invasive manner. Indeed, many of these patients were diagnosed with advanced-stage disease, who could benefit from an easily detected and quantified circulating biomarker such as NCL expressed by PBMCs. This detection could be performed using labeled rG4 as we demonstrated previously. Regarding the NCL expression levels for the different histological categories of LC, the results suggest that there was no significant difference between the tested types of LC, namely adenocarcinoma, squamous cell carcinoma, large-cell undifferentiated carcinoma and SCLC (**Figure 8.11**), contrarily to what was observed in the tissue biopsies by Xu and collaborators [10]. Nevertheless, as already shown by Showe and collaborators, the PBMC-associated gene signatures (such as NCL increased expression as herein demonstrated) can help predict the outcome in LC patients independent of demographic data or TNM staging [33]. More importantly, this information may persist after tumor resection which could aid patient long-term follow-up. Our results show that NCL expression levels have the potential to be used in LC patients in a non-invasive manner, potentially helping the diagnosis. The results obtained also showed the potential used of rG4 as a molecular recognition probe to detect NCL.

Acknowledgements

This work was supported by MIT Portugal FCT project BIODEVICE ref. MIT-EXPL/BIO/0008/2017, UTAustin FCT project DREAM ref. UTAP-EXPL/NTec/0015/2017, PESSOA 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 funds and, where applicable, co-financed by the FEDER through COMPETE 2020, POCI, PORL and PIDDAC. J. Figueiredo acknowledges a doctoral fellowship grant from the FCT – Foundation for Science and Technology ref. SFRH/BD/145106/2019. J. Carvalho acknowledges a doctoral fellowship grant from Foundation for Science and Technology (FCT) ref. SFRH/BD/122953/2016. D. Alexandre and A. Miranda acknowledge the fellowship grant from PTNMR project, ref. PINFRA/22161/2016-B1 and PINFRA/22161/2016-B4, respectively. Jéssica Lopes-Nunes acknowledges the fellowship reference UTAP-EXPL/NTec/0015/2017-B1. This work was also supported by the SYMBIT project reg. no. CZ.02.1.01/0.0/0.0/15_003/0000477 financed from the ERDF and by INCa PL-Bio and ANR grants G4Access (to J.L.M.).

References

[1] Bray, F. *et al.* (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394–424.

[2] Zappa, C. and Mousa, S.A. (2016) Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res* 5, 288–300.

[3] Baize, N. *et al.* (2020) Carboplatin plus etoposide versus topotecan as second-line treatment for patients with sensitive relapsed small-cell lung cancer: an open-label, multicentre, randomised, phase 3 trial. *Lancet Oncol* 21, 1224–1233.

[4] Benson, Z. *et al.* (2017) Conditioning neoadjuvant therapies for improved immunotherapy of cancer. *Biochem Pharmacol* 145, 12–17.

[5] Friedlaender, A. et al. (2019) Identifying successful biomarkers for patients with non-small-cell lung cancer. *Lung Cancer Manag* 8.

[6] Abdelmohsen, K. and Gorospe, M. (2012) RNA-binding protein nucleolin in disease. *RNA Biol* 9, 799-808.

[7] Berger, C.M. *et al.* (2015) The roles of nucleolin subcellular localization in cancer. *Biochimie* 113, 78–85.

[8] Ramos, K.S. *et al.* (2020) The nucleolin antagonist N6L inhibits LINE1 retrotransposon activity in non-small cell lung carcinoma cells. *J Cancer* 11, 733–740.

[9] Huang, F. *et al.* (2019) Phosphorylation of nucleolin is indispensable to its involvement in the proliferation and migration of non-small cell lung cancer cells. *Oncol Rep* 41, 590–598.

[10] Xu, J. yu *et al.* (2016) Prognostic significance of nuclear or cytoplasmic nucleolin expression in human non-small cell lung cancer and its relationship with DNA-PKcs. *Tumor Biol* 37, 10349–10356.

[11] 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.

[12] Li, J. *et al.* (2014) Aptamer imaging with Cu-64 labeled AS1411: Preliminary assessment in lung cancer. *Nucl Med Biol* 41, 179–185.

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

[14] Roncarati, R. *et al.* (2019) The importance of microRNAs in RAS oncogenic activation in human cancer. *Front Oncol* 9, 988.

[15] Heng, W. *Sen et al.* (2019) Lung cancer stem cells: origin, features, maintenance mechanisms and therapeutic targeting. *Biochem Pharmacol* 160, 121–133.

[16] Grenda, A. *et al.* (2019) MicroRNAs aid the assessment of programmed death ligand 1 expression in patients with non-small cell lung cancer. *Oncol Lett* 17, 5193–5200.

[17] Zhang, L. *et al.* (2018) Serum MicroRNA-150 Predicts Prognosis for Early-Stage Non- Small Cell Lung Cancer and Promotes Tumor Cell Proliferation by Targeting Tumor Suppressor Gene SRCIN1. *Clin Pharmacol Ther* 103, 1061–1073.

[18] Zhang, N. *et al.* (2013) miR-150 promotes the proliferation of lung cancer cells by targeting P53. *FEBS Lett* 587, 2346–2351.

[19] 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.

[20] Gongol, B. *et al.* (2019) Shear stress regulation of miR-93 and miR-484 maturation through nucleolin. *Proc Natl Acad Sci USA* 116, 12974–12979.

[21] Lucchesi, C. *et al.* (2016) Modulation of the p53 family network by RNA-binding proteins. *Trans Cancer Res* 5, 676–684.

[22] Mazan-Mamczarz, K. *et al.* (2003) RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. *Proc Natl Acad Sci U S A* 100, 8354–8359.

[23] Desvignes, T. *et al.* (2015) miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. *Trends Genet* 31, 613–626.

[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] Joo, E.J. *et al.* (2010) Induction of nucleolin translocation by acharan sulfate in A549 human lung adenocarcinoma. *J Cell Biochem* 110, 1272–1278.

[27] Gambari, R. *et al.* (2011) Targeting microRNAs involved in human diseases: A novel approach for modification of gene expression and drug development. *Biochem Pharmacol* 82, 1416–1429.

[28] Mirihana Arachchilage, G. *et al.* (2015) A Potassium Ion-Dependent RNA Structural Switch Regulates Human Pre-miRNA 92b Maturation. *Chem Biol* 22, 262–272.

[29] 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* 21, 767.

[30] Bugaut, A. *et al.* (2012) An RNA Hairpin to G-Quadruplex Conformational Transition. *J Am Chem Soc* 134, 19953–19956.

[31] Mongelard, F. and Bouvet, P. (2007) Nucleolin: a multiFACeTed protein. *Trends Cell Biol* 17, 80–86.

[32] Bates, P.J. *et al.* (1999) Antiproliferative Activity of G-rich Oligonucleotides Correlates with Protein Binding. *J Biol Chem* 274, 26369–26377.

[34]. Suzuki, K. *et al.* (2011) Prognostic immune markers in non-small cell lung cancer. *Clin Cancer Res* 17, 5247-5256.

3rd SECTION

Conclusions section

9th Chapter

Conclusions and Future Perspectives

Conclusions and future perspectives

General conclusions

Over the years, DNA and RNA G4s have emerged as promising drug targets due to their involvement in key biological processes. The use of G4 ligands to modulate G4 folding within biological systems has been explored as a potential therapeutic strategy, particularly in the context of cancer. The first section of this work provides an important overview of this field, highlighting promising G4 ligands in particular, heterocyclic-based compounds that act as anticancer agents by targeting G4 structures within telomeric regions, oncogenes promoters, and non-coding RNAs.

Some classes of compounds containing different chemical moieties (acridines, quindolines, quinazolones, quinolones, naphthalenediimides, perylenediimides, and phenanthrolines) have demonstrated significant effects against cancer cell growth with remarkable *in vitro* and/or *in vivo* activity and some of them reached (pre)clinical evaluation. However, some issues such as lack of efficacy, poor pharmacokinetics, biodistribution problems, and off-target toxicity often compromise their development as G4-targeting drugs. To try to solve this problem, we advocate that the integration of approaches combining SAR with other techniques, such as high-throughput screening, or system biology which should help to better understand ligand behavior and design better ligands.

Additionally, we reviewed the use of G4 ligands as stabilizing agents in non-coding RNAs, which have recently emerged as promising anticancer targets. The formation of G4 can control the regulation of RNA metabolism, including miRNA biogenesis. For instance, controlling the equilibrium between G4 and stem-loop in pre-miRNAs by G4 ligands has been suggested as a therapeutic approach in several human cancers. Moreover, the G4 structures within these sequences demonstrate the potential to interact with NCL.

LC is considered a hard-to-treat cancer due, in part, to its high level of molecular heterogeneity. Considering the late diagnosis and drug resistance the current therapeutic options are often disappointing. To overcome these issues a new therapeutic approach mediated by targeting G4 LC-related genes has emerged as a promising option. Several G4 ligands demonstrated remarkable activity against LC standalone or in combination with commercial drugs.

Thus, the main goal of this thesis was to develop novel G4 ligands based on a phenanthroline scaffold, more specifically, to study the influence of different substitutions at 2,9 positions of the 1,10-phenanthroline core as G4 ligands and to access their anti-tumor activity in LC cell lines. Additionally, we studied the formation of G4 structure in the human pre-MIR150 and evaluated their ability to recognize NCL, which is overexpressed in LC cells and PBMCs derived from LC patients, offering a potential diagnostic tool.

In the **fifth chapter** we investigated different scaffolds featuring five heteroaromatic cores (oxazole, naphthalene, acridine, and phenanthroline) for their G4 stabilizing potential. Compounds containing naphthalene (Napht-3 and Napht-5) and phenanthroline (DPPZ and $PhenCOOCH_3$) cores were synthesized with low to high yields. All compounds obey Lipinski´s Rule of Five, suggesting favorable properties to achieve the drug-like criteria. Specifically, none of the compounds exhibited a molecular weight higher than 500 Da, and their log *P* values were smaller or equal to 5, indicating potential for cellular uptake and membrane permeability. Regarding G4 stabilizing properties assessed through FRET-melting experiments, acridine-containing compounds (C_5NH_2 and C_8NH_2) proved greater stabilization in all tested G4. Among them, C_8NH_2 featuring a longer flexible alkylamine chain, exhibited enhanced stabilization, notably in the hybrid F21T sequence (ΔT_m = +26.8 °C). The compound $PhenCOOCH_3$ emerged as the most effective stabilizer among phenanthroline-containing compounds, inducing greater thermal stabilization in FKrasT and F21T sequences at 5 μM concentration (ΔT_m = +3.4 °C). Acridine compounds exhibited potent activity against A549 cells, with IC_{50} values of 1.3 and 3.6 μM, respectively. $PhenCOOCH_3$ exhibited a good selectivity factor for LC cells, with an IC_{50} of 6.4 μM, suggesting a promising scaffold for further optimization.

In the **sixth chapter**, efforts were made to enhance the G4 binder properties by functionalizing the 1,10-phenanthroline scaffold through the introduction of amine and amide side chains at 2,9 positions. A library of sixteen 2,9-disubstituted-1,10-phenanthroline compounds was successfully synthesized with moderate to good yields. Based on the G4 stabilizer properties and antiproliferative activity of compounds, some SAR analysis was made. Among the synthesized ligands, three novel compounds induced thermal stabilization in all tested telomeric sequences (F21T, F21CTAT, and F25cebT). Two of these compounds feature ethane-aminium in lateral side chains (**3a** and **5b**), while one features methoxyaniline moieties (**2b**). These findings reflect the evidence that the introduction of positive charges via polyamine side chains enhances electrostatic interactions with the negative charges of phosphate groups in oligonucleotides.

Additionally, the presence of an aromatic core in the lateral side chain may further improve G4 stabilizer properties. Based on cell viability, the compounds exhibited moderate cytotoxicity against LC cells. Notably, ligand 5b proved the best anti-tumor activity against A549 cells resulting in 24.5% cell viability, suggesting that their activity may be mediated by G4 interaction.

To increase the chemical diversity of 1,10-phenanthroline derivatives, the **seventh chapter** describes the development of 2,9-disubstituted-1,10-phenanthroline compounds bearing 1,2,3-triazole groups. A library of ten novel derivatives was synthesized via CuAAC reactions with moderate to good yields. When tested as G4 ligands, these compounds did not present better stabilizer activity of telomeric sequences than derivatives described in the previous chapter. However, promising stabilizer effects were found against KRAS G4 sequences. Three compounds were identified as promising G4 stabilizers and binding agents with ΔT_m values ranging between 4.7 and 11.2 °C and K_D from 10^{-6} and 10^{-9} M. These compounds bearing *N´N*- dimethylmethanamine (**4a**), *N´N*- dimethylmethanaminium (**4b**), and positive piperazine (**5b**), end groups. The antiproliferative activity of synthesized compounds did not appear to be mediated from its action as G4 stabilizers since the best G4 ligands did not exhibit a pronounced antiproliferative effect on LC cells and near toxicity in non-malignant cells. However, the novel triazole-containing compound 2a exhibited an IC_{50} of 14.6 and 10.9 μM in A549 and H1299 cells, respectively. These findings might be related to the poor "drug-like" properties predicted for the majority of compounds.

Finally, in the **ninth chapter,** the folding of a G4 structure within human pre-MIR150 (5′-GGCCUGGGGGACAGGGACCUGGG-3′) was described for the first time and studied as a molecular recognition agent of NCL in LC cells and PBMCs derived from LC patients. In the presence of K^+ or when complexed with PhenDC3, the sequence adopts a parallel G4, and upon the addition of PhenDC3, the melting temperature of the G4 structure substantially increases. Moreover, the K_D of the interaction between G4/NCL or G4-PhenDC3/NCL is in the nanomolar range. The PAGE assay suggests that the highest affinity and stability of the G4 towards NCL were in the presence of PhenDC3. Additionally, confocal microscopy indicated the highest co-localization of the complex G4/PhenDC3 with NCL expressed on the surface of A549 LC cells and PBMCs isolated from blood samples of LC patients. These findings allow us to conclude that the G4/PhenDC3 complex has the potential to be used as a recognition agent of NCL in these LC samples.

Overall, the work developed in this thesis has not only introduced new perspectives but has also reaffirmed structural features of G4 ligands. Indeed, some structural

modifications introduced at the 1,10-phenanthroline core failed to enhance the G4 stabilizing potential, and in some cases, as we argue in the first chapter "*may be futile*". Besides that, it is essential to highlight the G4 stabilizing effect of derivatives 2a,3a, and 5b (detailed in the 6th chapter) and the anti-tumor activity induced by compound 5b against LC cells. Additionally, the stabilization induced by compounds 4a,4b, and 5b (detailed in the 7th chapter) against KRAS G4 sequences deserves to be highlighted. Furthermore, and even without apparent G4-mediated activity, derivative 2a (also detailed in the 7th chapter) presented promising anti-tumor activity in LC cells. From the medicinal chemistry standpoint, this work also provides valuable SAR insights into these compounds, paving the opportunity for further investigations. Moreover, this work allows the elucidation of a G4 structure within pre-miR150 opening the opportunity for further studies and establishing the potential use as a diagnostic tool in the LC.

Future Perspectives

The G4s are promising targets for drug design against the development and progression of LC. The findings presented in this work provide insights into this field and stimulate reflections on future work presented herein:

- Despite the important structural features and SAR elucidation described in the sixth and seventh chapters, the synthesized 2,9-disubstituted-1,10-phenanthroline derivatives did not reach the stabilizing/binding ability reported for other phenanthroline-based compounds, highlighting the need for further optimization.

- In order to investigate the discrimination among different DNA or RNA G4, it would be interesting to evaluate the binding/stabilizing potential of all compounds outlined in the sixth chapter in other biologically relevant LC-related G4 sequences.

- Regarding the biological effects of the tested compounds, these should be further studied, with particular attention to compound 5b described in the sixth chapter. Studies to explore their telomere targeting activity, such as *in vitro* TRAP or telomere uncapping assays to assess their inhibitory effects on human telomerase activity, should be employed. Additionally, experiments to investigate the effect of the ligand on different phases of cell cycle distribution

in A549 cells, for instance using FACS analysis, would provide valuable insights.

- Considering the synergetic anti-tumor effects observed between some G4 ligands and antineoplastic treatments, exploring the synergic activity of synthesized compounds with commercially available drugs such as gemcitabine or cisplatin would be of considerable interest.

- Considering the possible role of *MIR150* in the progression of LC, as described in the eighth chapter, the investigation of its precursor (*pre-MIR150*) as a therapeutic target in LC should be further explored. Characterization of the G4 structure within pre-MIR150 presents an opportunity to explore its role in the biogenesis of MIR150. Studies aiming at confirming G4 folding *in vivo,* using techniques such as NMR and/or confocal microscopy, hold significant interest. Additionally, determining how stabilization induced by PhenDC3 on the pre-MIR G4 sequence affects MIR150 expression levels in A549 cells through Luciferase reporter assay would provide valuable insights.

In addition, some outstanding questions in the field of targeting G4 nucleic acids should be addressed:

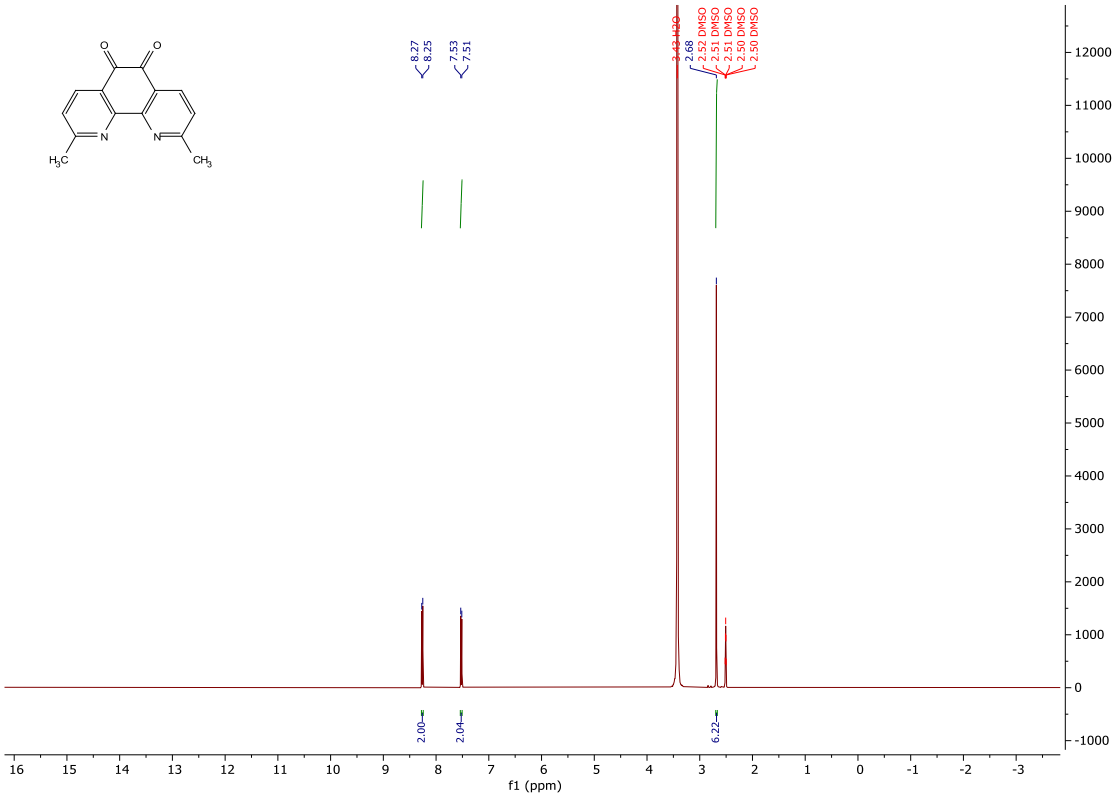
- **Affinity versus Selectivity:** Is it possible to design a ligand for a specific sequence? Or G4 topology? If not, what other genes may be affected? Is the multi-targeting a real problem? Is it possible to find a trade-off between selectivity and affinity, as tight binders may recognize the most negatively charged nucleic acids? Which level of selectivity between G4 and duplex is necessary? Is selectivity absolutely necessary for a therapeutic effect? If not, what type of undesired unspecific effects can we expect?

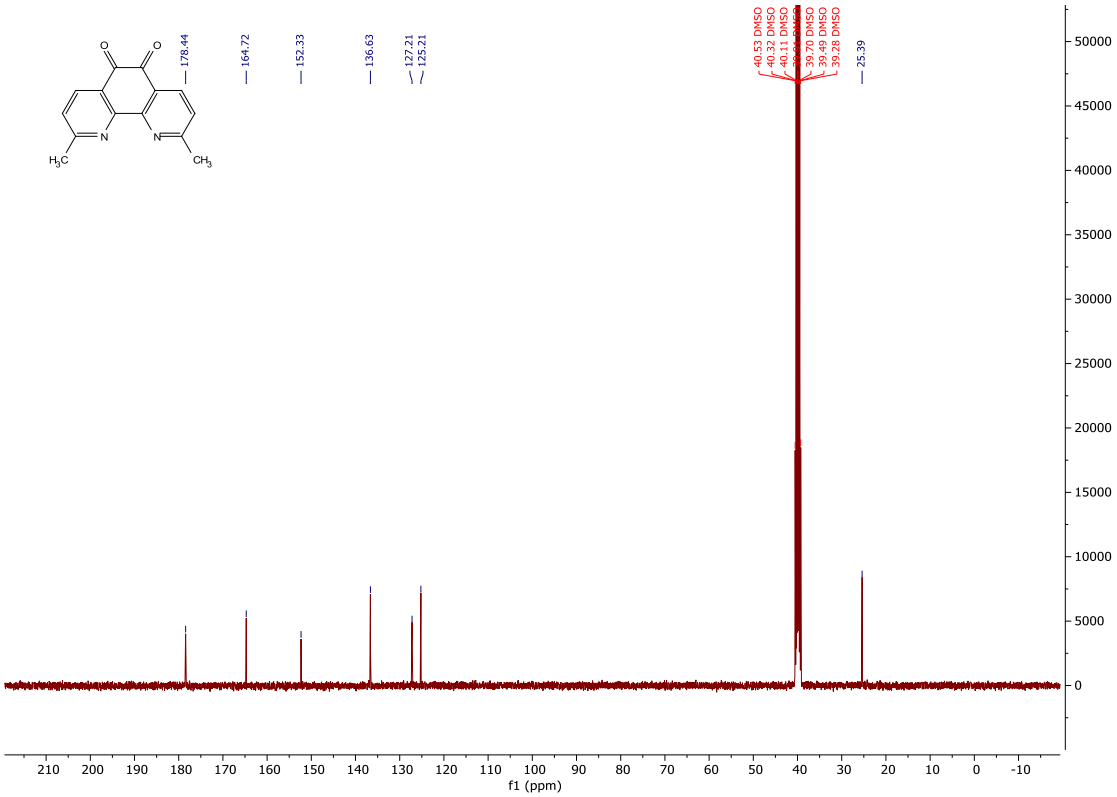
- **G-quadruplex ligands in the clinic:** Is it possible to improve the poor pharmacokinetics and target selectivity of the G4 ligands? Would decreasing the cationic character of G4-binding ligands improve cellular penetration? Do all G4 ligands stimulate autophagy in cancer cells?

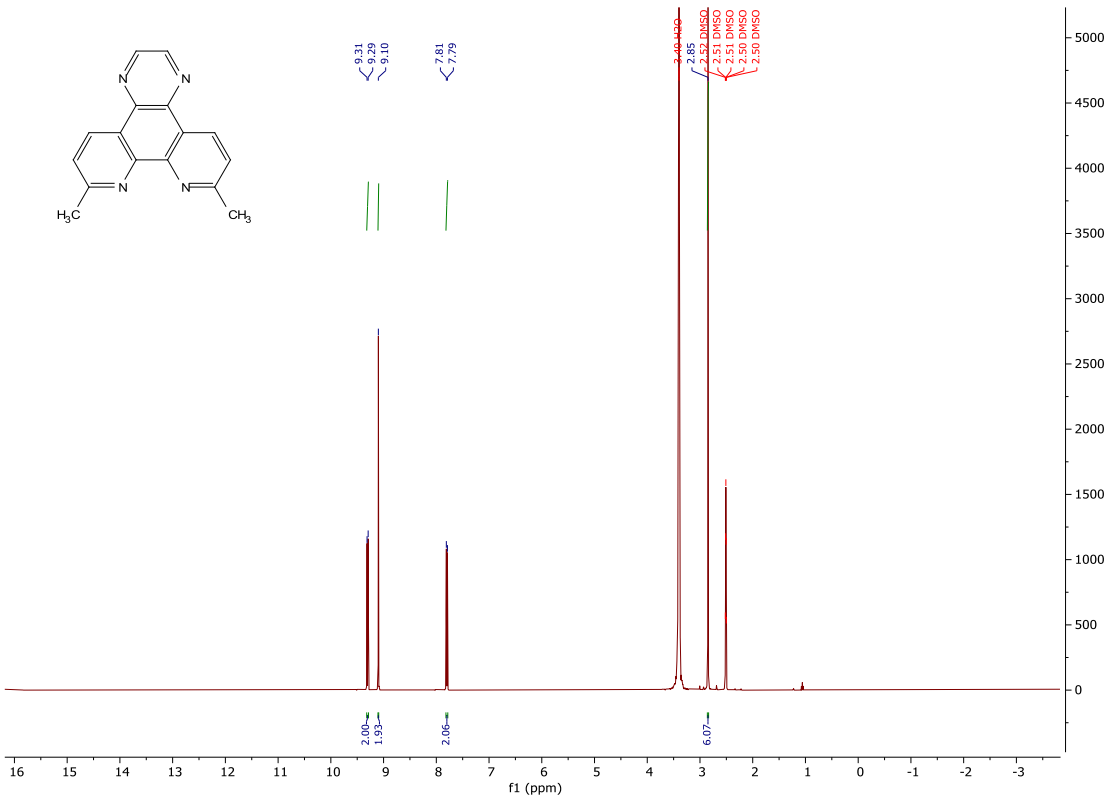
- **G-quadruplex ligands against LC:** Which LC cells or tumor type would be more sensitive to treatment with G4 ligands? Could mitochondrial DNA G4 be a future target in LC? Can the similarity between LC-related genes with other cancer types difficult the therapeutic efficacy?

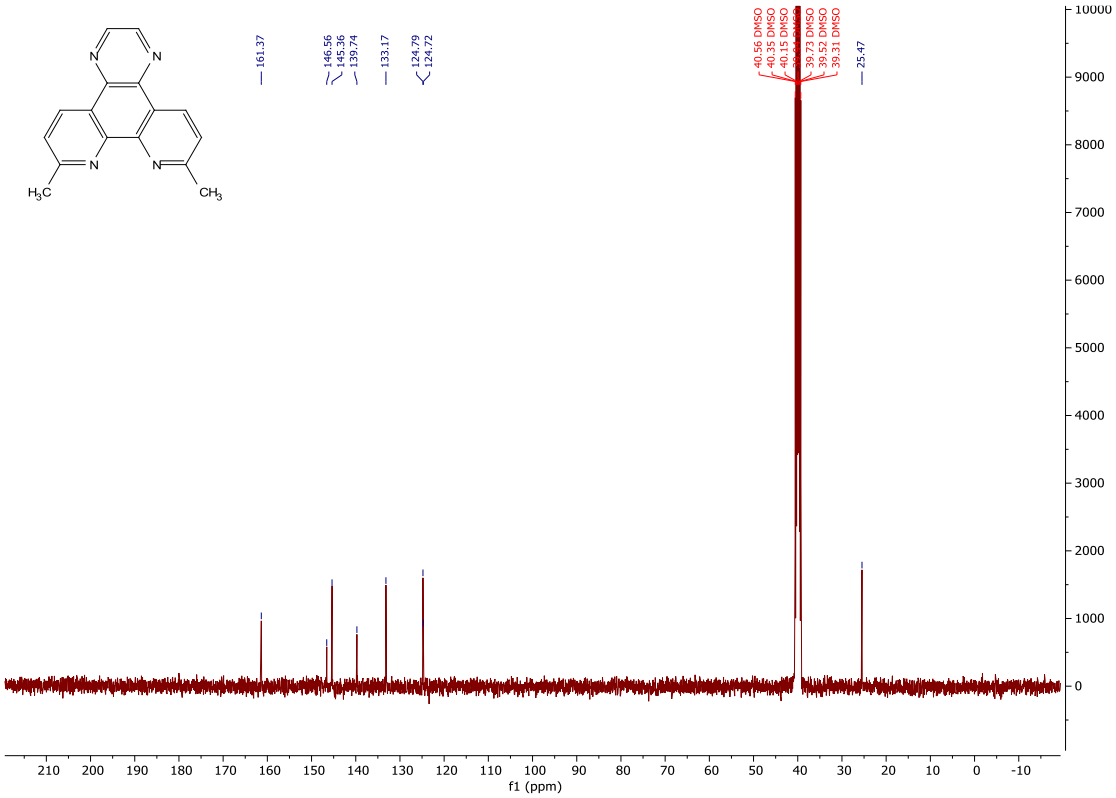
Appendix A

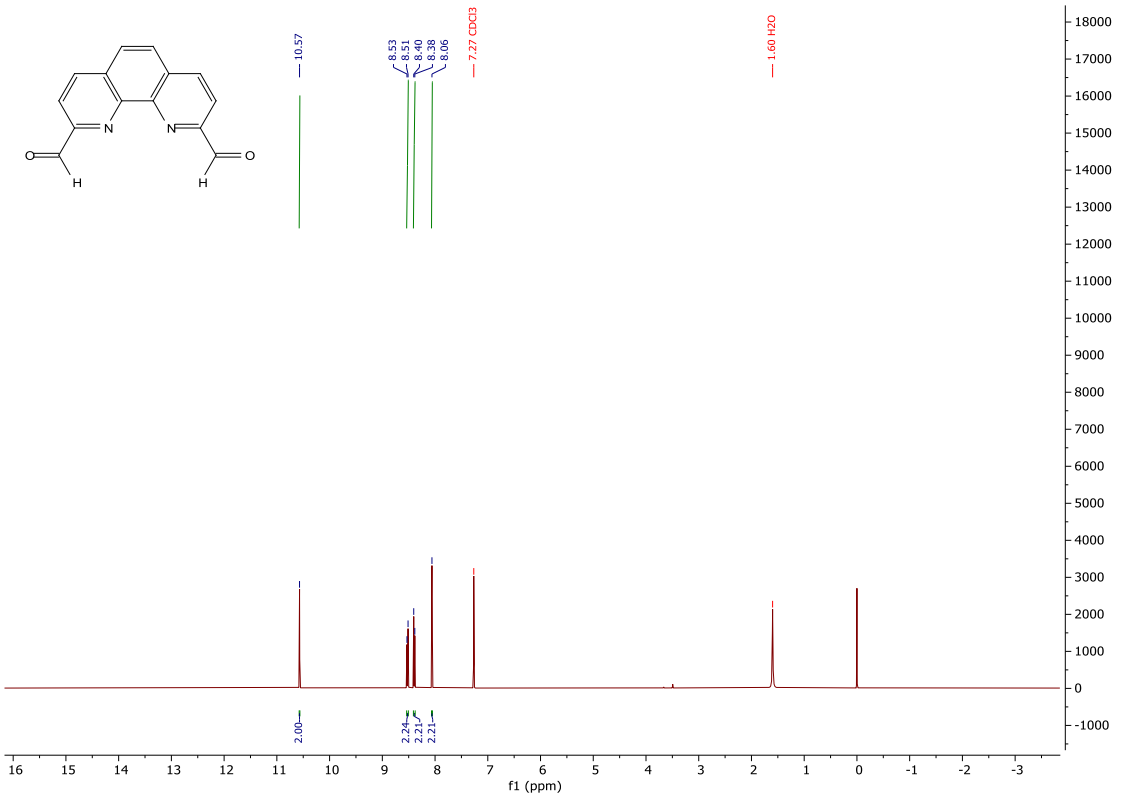
Supporting information for 5th chapter

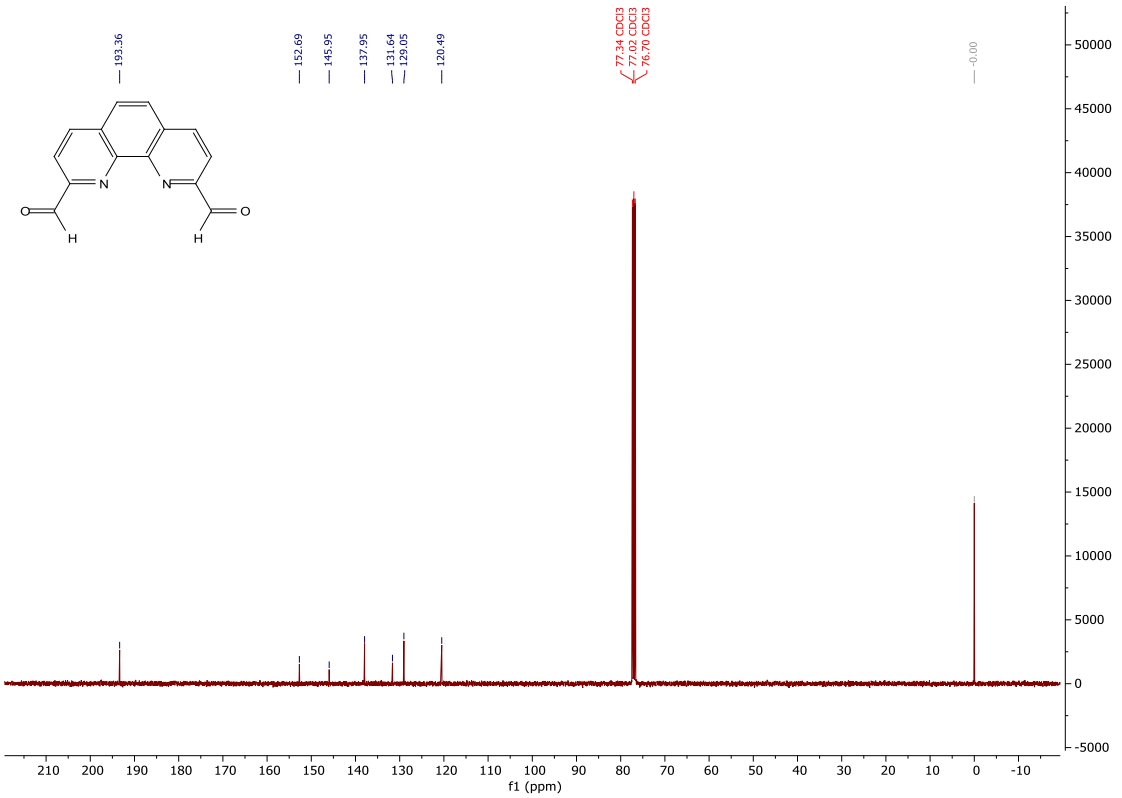


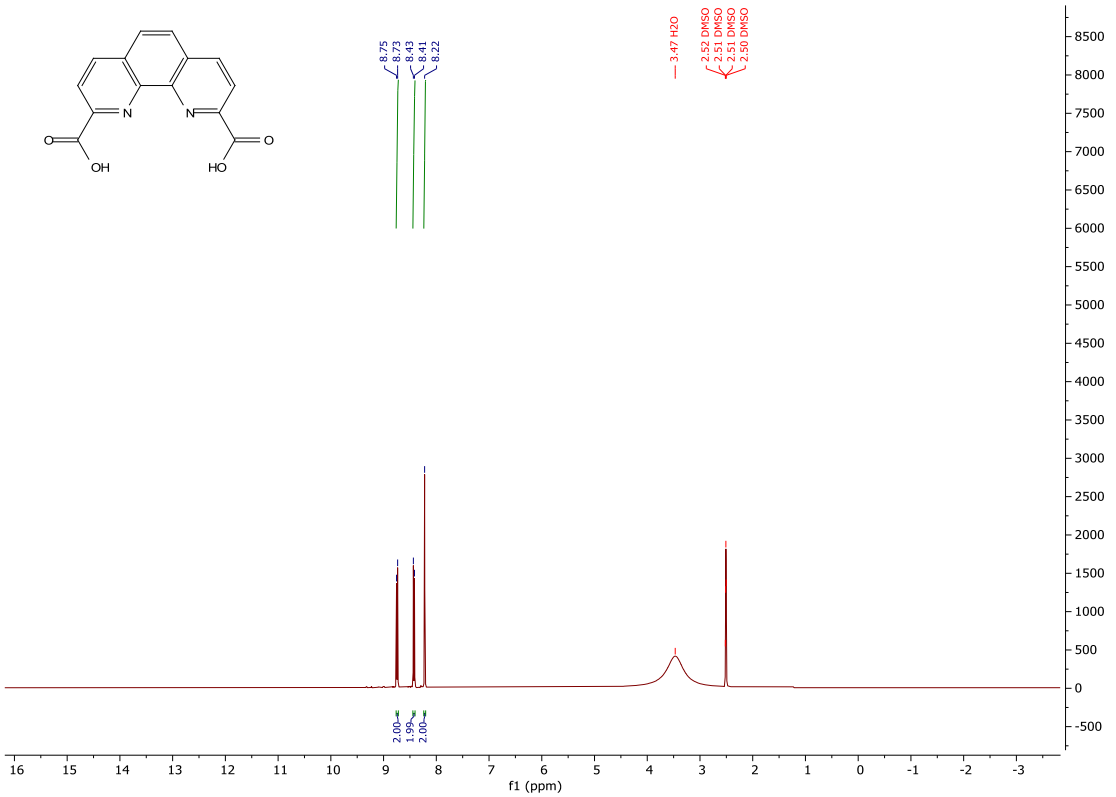


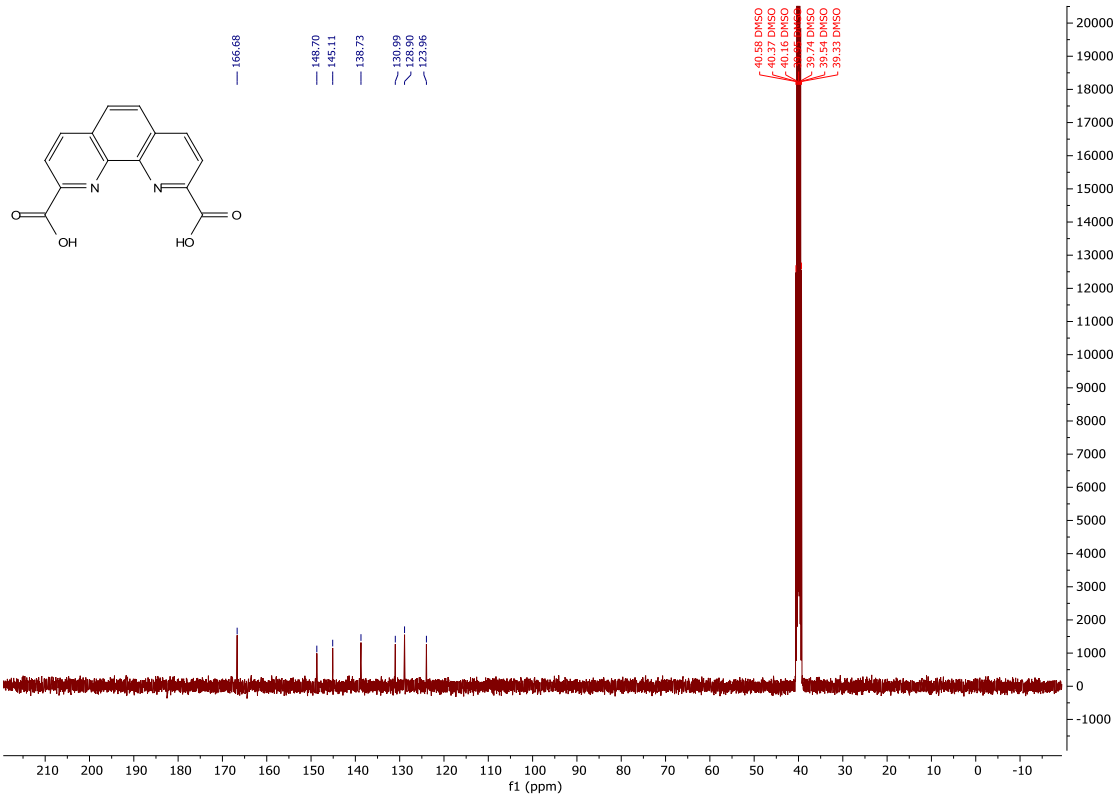

9.31
9.29
9.10
7.81
7.79
3.40 H2O
2.65
2.51 DMSO
2.51 DMSO
2.51 DMSO
2.50 DMSO
2.50 DMSO
2.00
1.93
2.06
6.07
f1 (ppm)

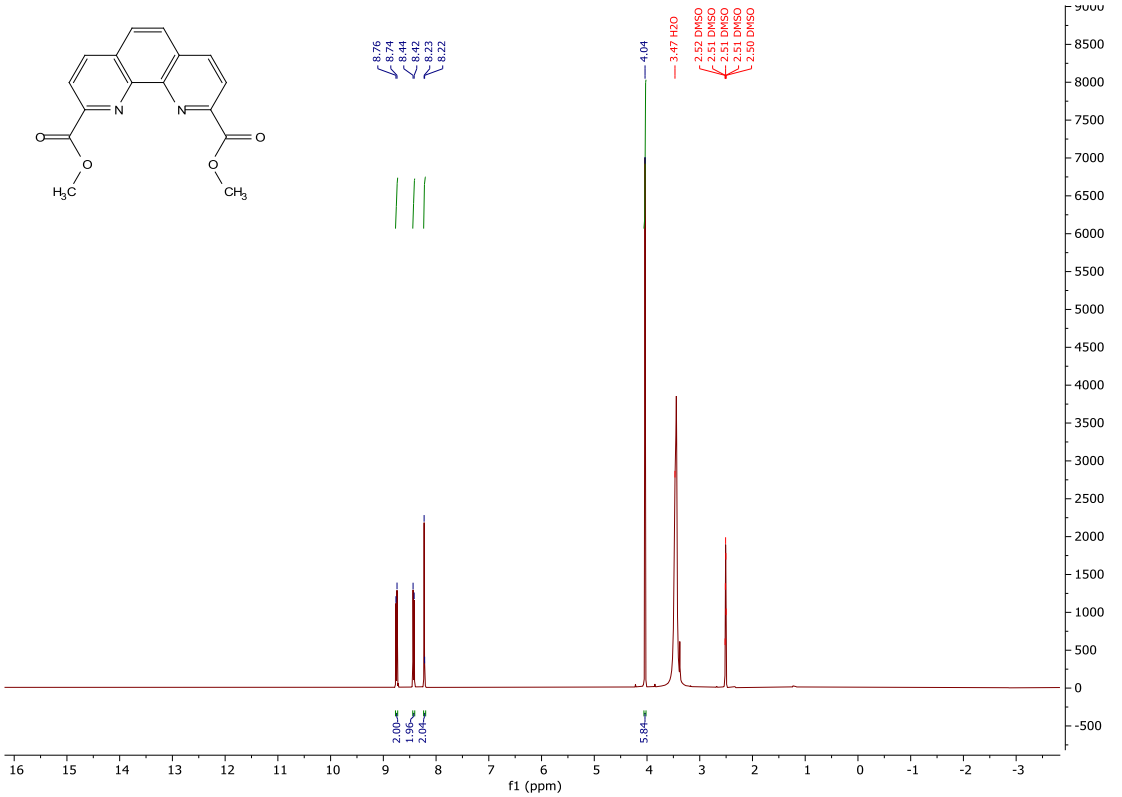

161.37
146.56
145.36
139.74
133.17
124.79
124.72
40.56 DMSO
40.35 DMSO
40.15 DMSO
39.73 DMSO
39.52 DMSO
39.31 DMSO
25.47
f1 (ppm)

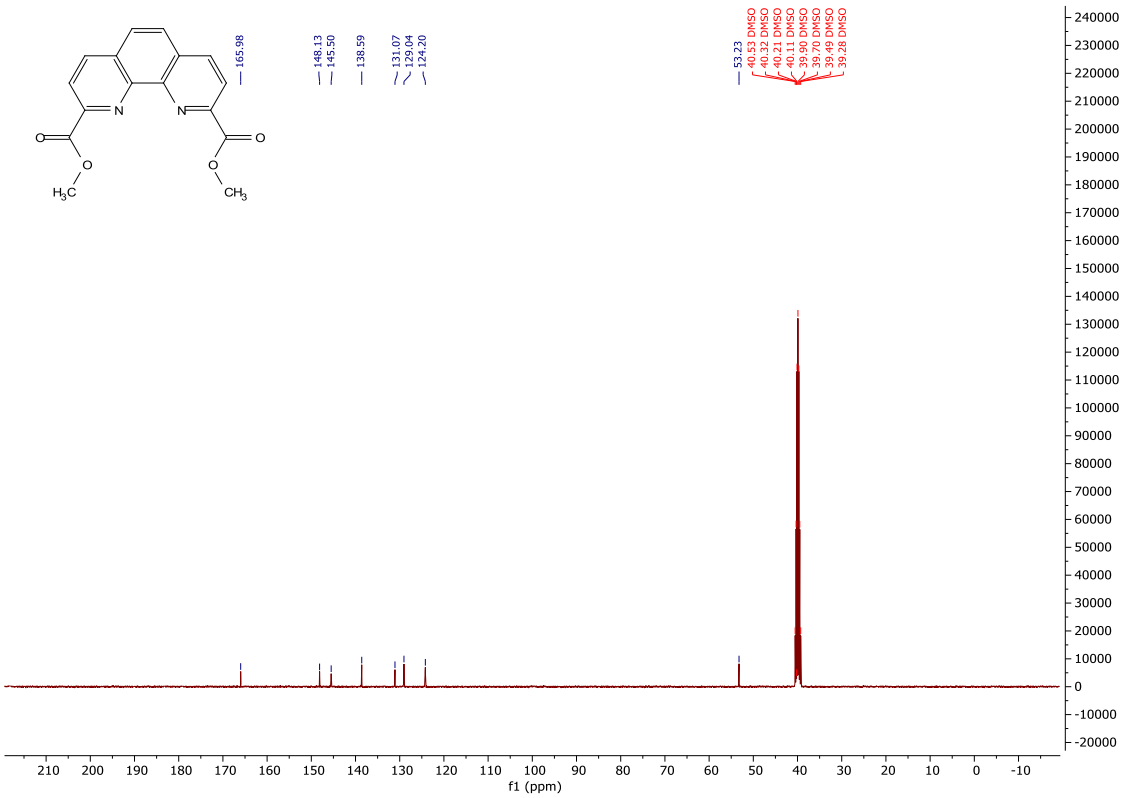

10.57
8.53
8.51
8.40
8.38
8.06
7.27 CDCl3
1.60 H2O
2.00
2.24
2.21
2.21
f1 (ppm)

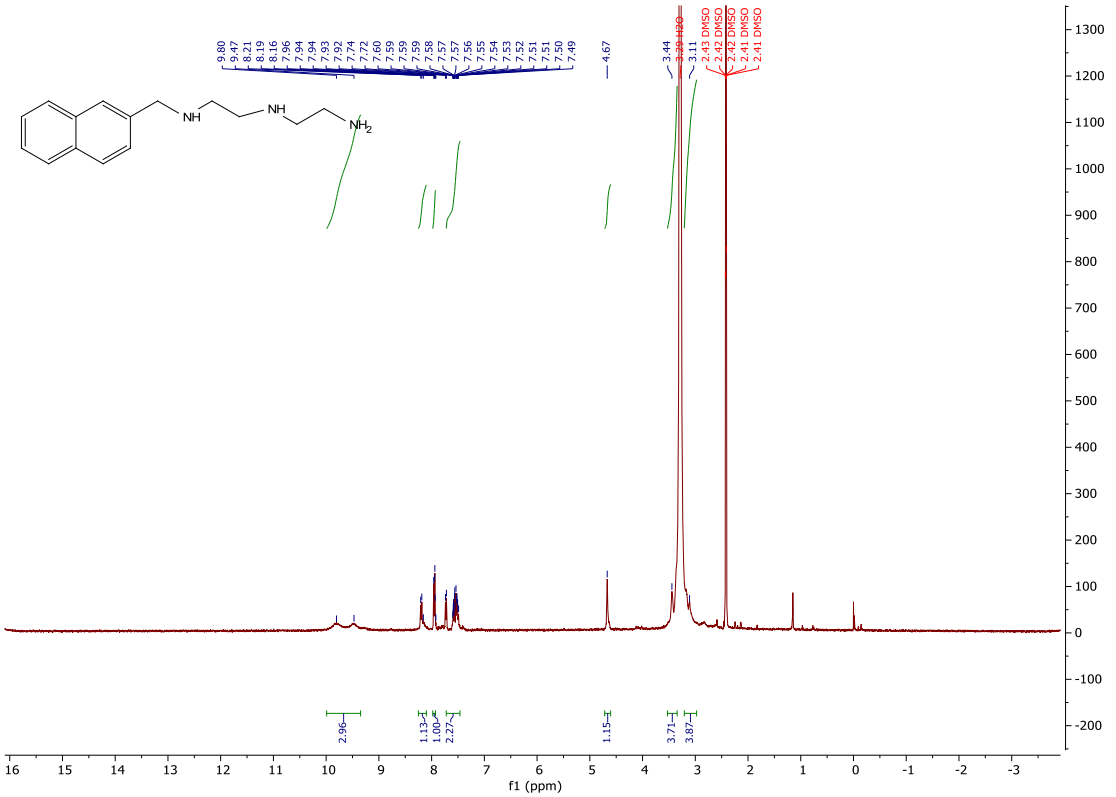

193.36
152.69
145.95
137.95
131.64
129.05
120.49
77.34 CDCl3
77.02 CDCl3
76.70 CDCl3
0.00
f1 (ppm)

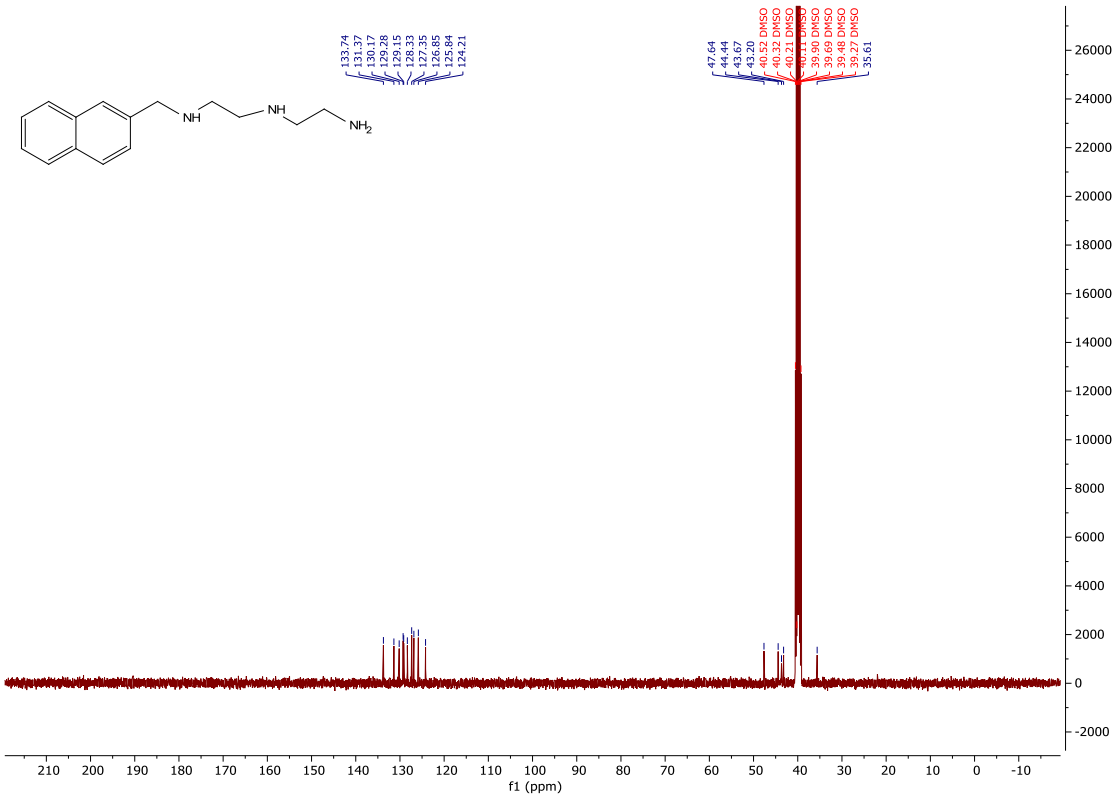

8.75
8.73
8.43
8.41
8.22
3.47 H2O
2.52 DMSO
2.51 DMSO
2.51 DMSO
2.50 DMSO
2.00
1.98
1.00
2.00
f1 (ppm)

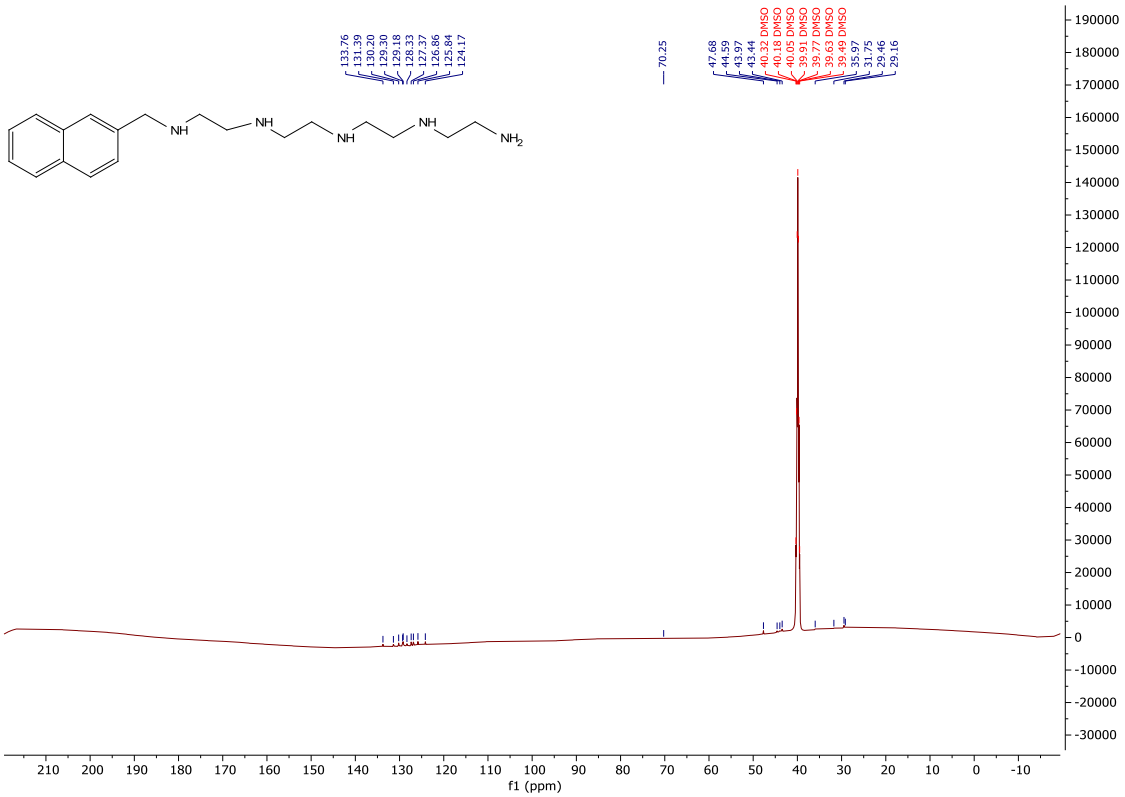

166.68
148.70
145.11
138.73
130.99
128.86
123.96
40.58 DMSO
40.37 DMSO
40.16 DMSO
39.74 DMSO
39.54 DMSO
39.33 DMSO
f1 (ppm)


8.76
8.74
8.44
8.42
8.23
8.22
4.04
3.47 H2O
2.52 DMSO
2.51 DMSO
2.51 DMSO
2.51 DMSO
2.50 DMSO
2.00
1.96
2.04
5.84
f1 (ppm)


165.98
148.13
145.50
138.59
131.07
129.04
124.20
53.23
40.53 DMSO
40.32 DMSO
40.21 DMSO
40.11 DMSO
39.90 DMSO
39.70 DMSO
39.49 DMSO
39.28 DMSO
f1 (ppm)


9.80
9.47
8.21
8.19
8.16
7.96
7.94
7.94
7.93
7.92
7.74
7.72
7.60
7.59
7.59
7.59
7.58
7.57
7.57
7.56
7.55
7.54
7.53
7.52
7.51
7.51
7.50
7.49
4.67
3.44
3.39 H2O
3.11
2.43 DMSO
2.42 DMSO
2.42 DMSO
2.41 DMSO
2.41 DMSO
NH
NH
NH2
2.96
1.13
1.00
2.27
1.15
3.71
3.87
f1 (ppm)


133.74
131.37
130.17
129.28
129.15
128.33
127.35
126.85
125.84
124.21
47.64
44.44
43.67
43.20
40.52 DMSO
40.32 DMSO
40.21 DMSO
39.90 DMSO
39.69 DMSO
39.48 DMSO
39.27 DMSO
35.61
NH
NH
NH2
f1 (ppm)



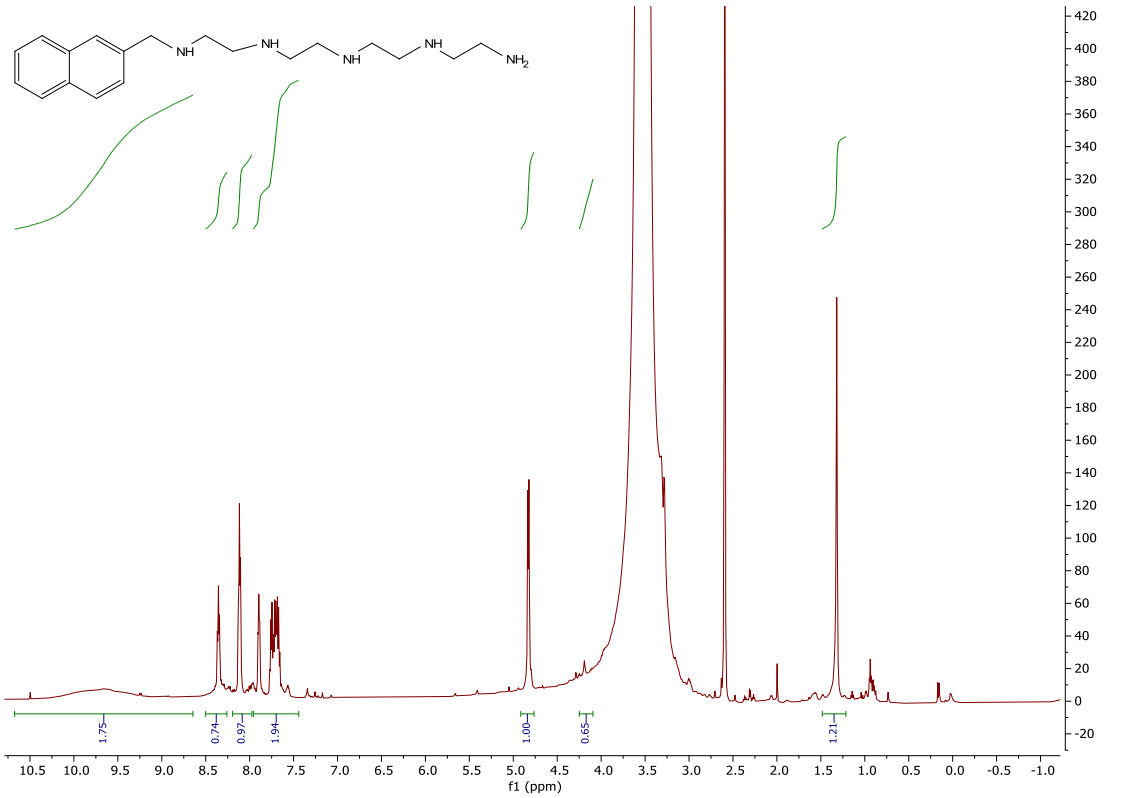


Figure A1. [1]HNMR and [13]C NMR spectra of synthesized compounds.

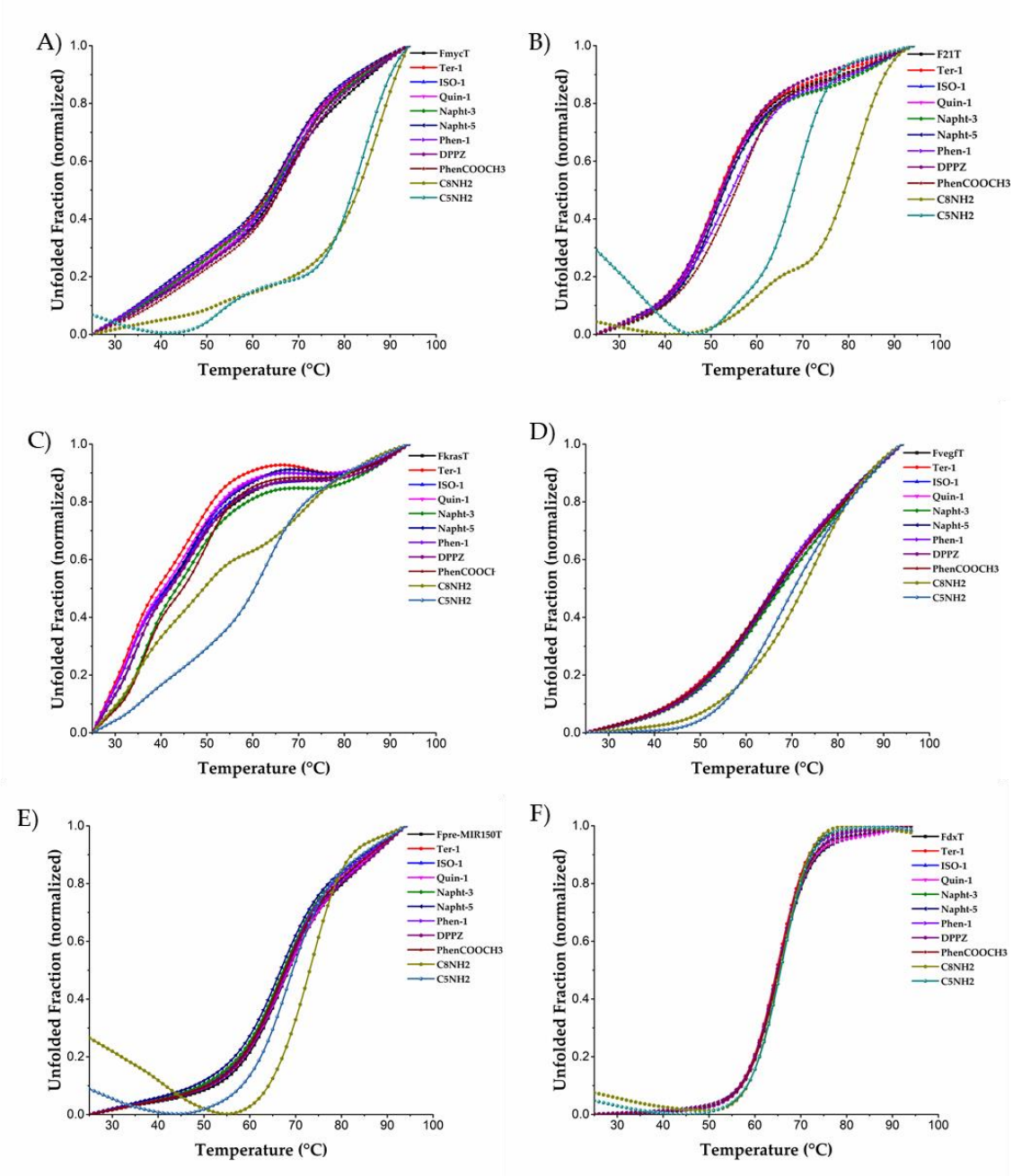


Figure A2. Normalized FRET melting curves of (A) FmycT, (B) F21T, (C) FkrasT, (D) FvegfT, (E) Fpre-MIR150T and (F) FdxT in the presence and absence of ligands. The buffer used was 10 mM lithium cacodylate, pH 7.2. FmycT, F21T and FkrasT was supplemented with 10 mM KCl and 90 mM LiCl, FvegfT and Fpre-MIR150T with 1 mM KCl and 99 mM LiCl and FdxT with 100 mM KCl.

Appendix B

Supporting information for 6th chapter

^{1}H, ^{13}C, DEPT-135, NOESY and HR-MS spectra

Tert-Butyl(2-aminoethyl)carbamate

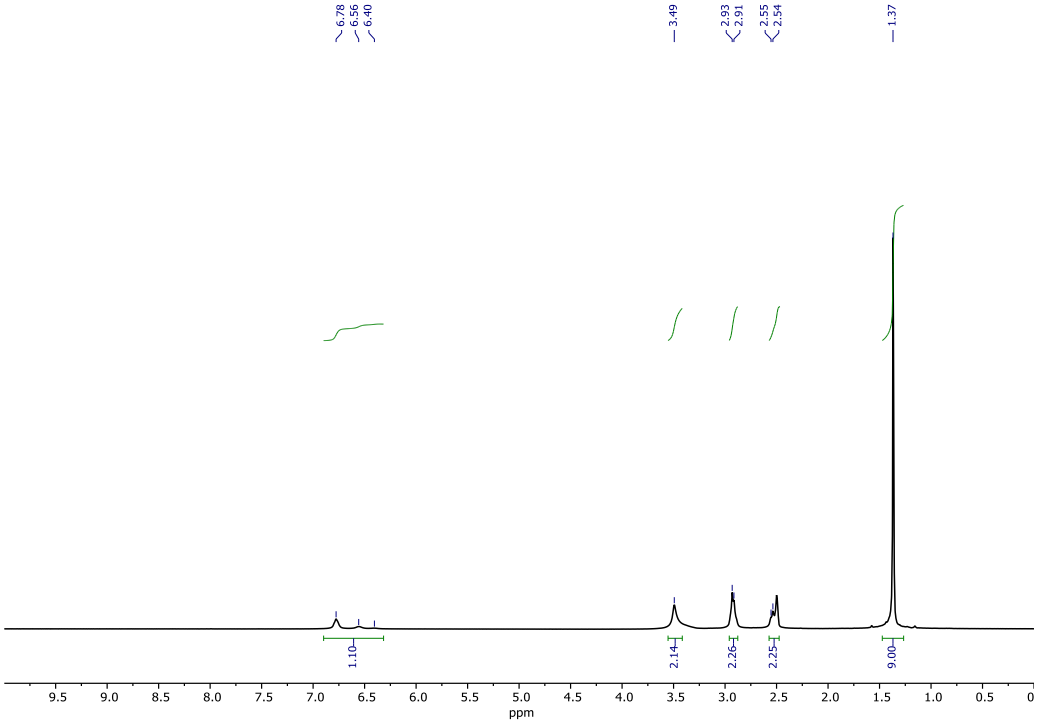


Figure B1. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for *tert*-butyl (2-aminoethyl)carbamate.

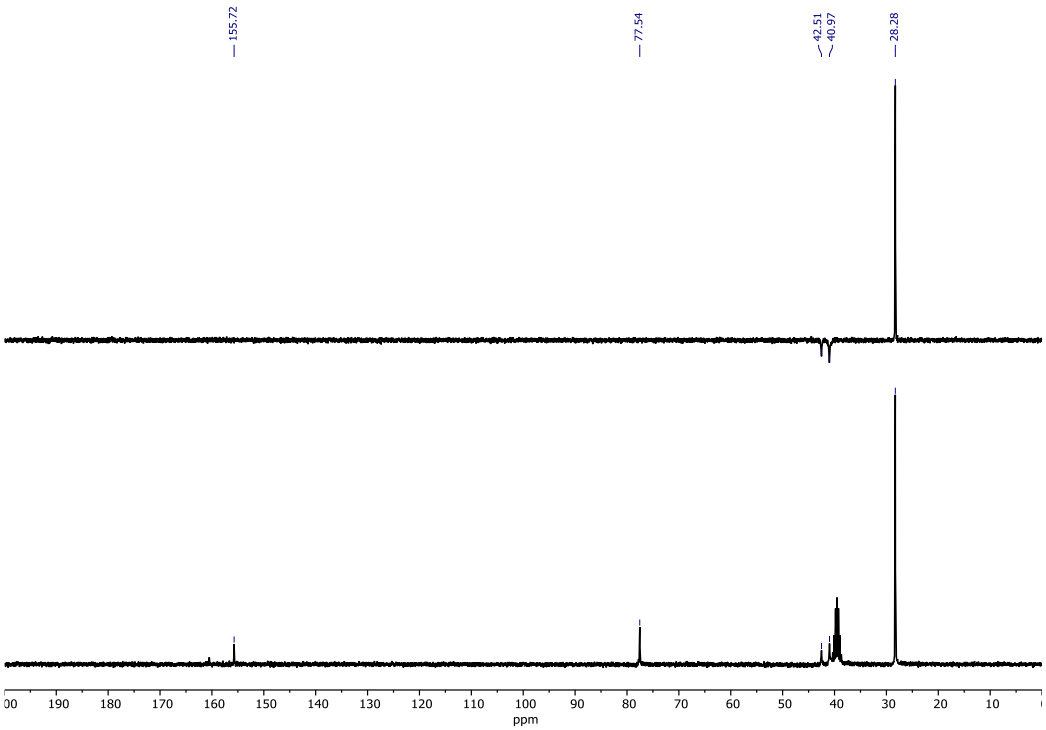


Figure B2. ^{13}C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for *tert*-butyl (2-aminoethyl)carbamate.

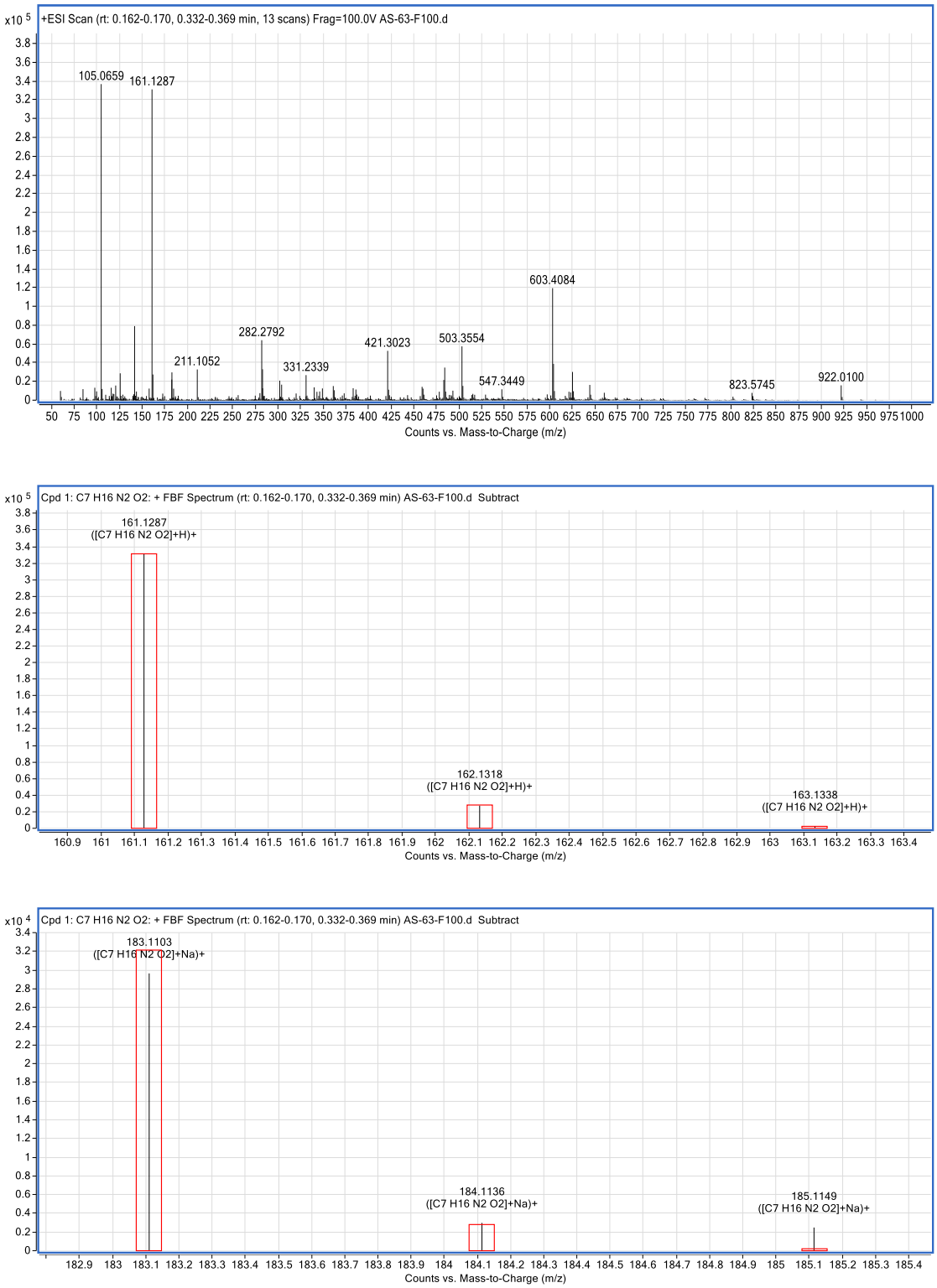


Figure B3. HR-MS (+ESI) spectra for *tert*-butyl (2-aminoethyl)carbamate.

Imines and amines

Compound **1a**

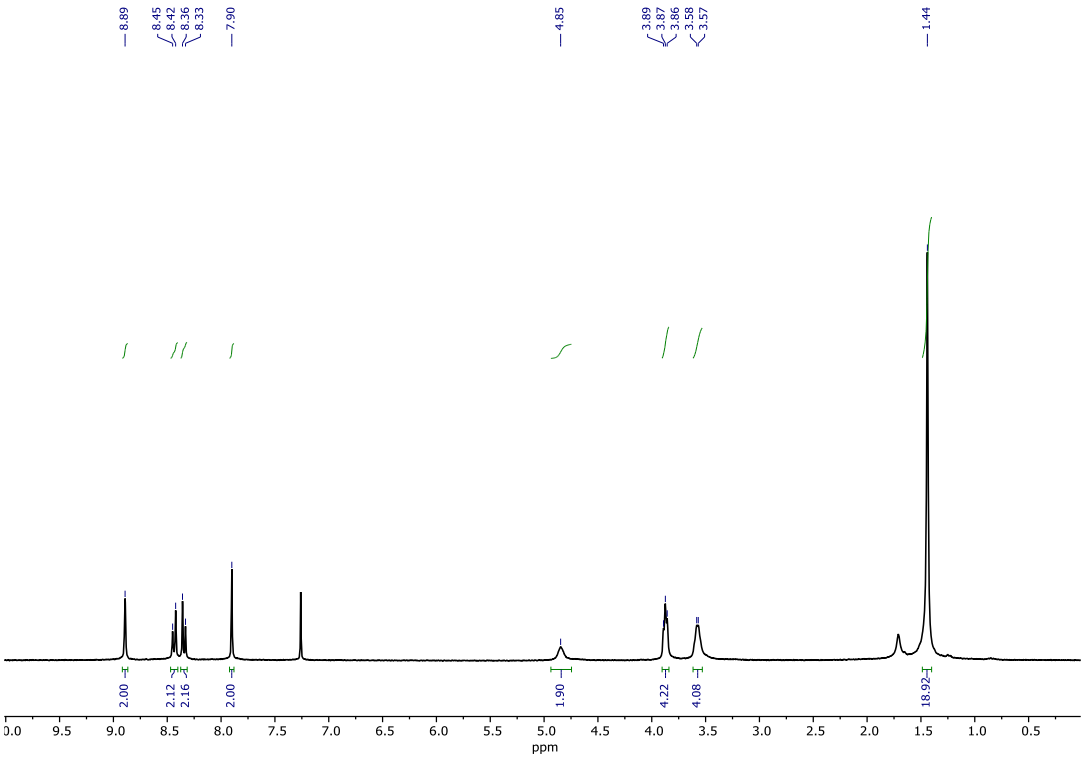


Figure B4. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **1a**.

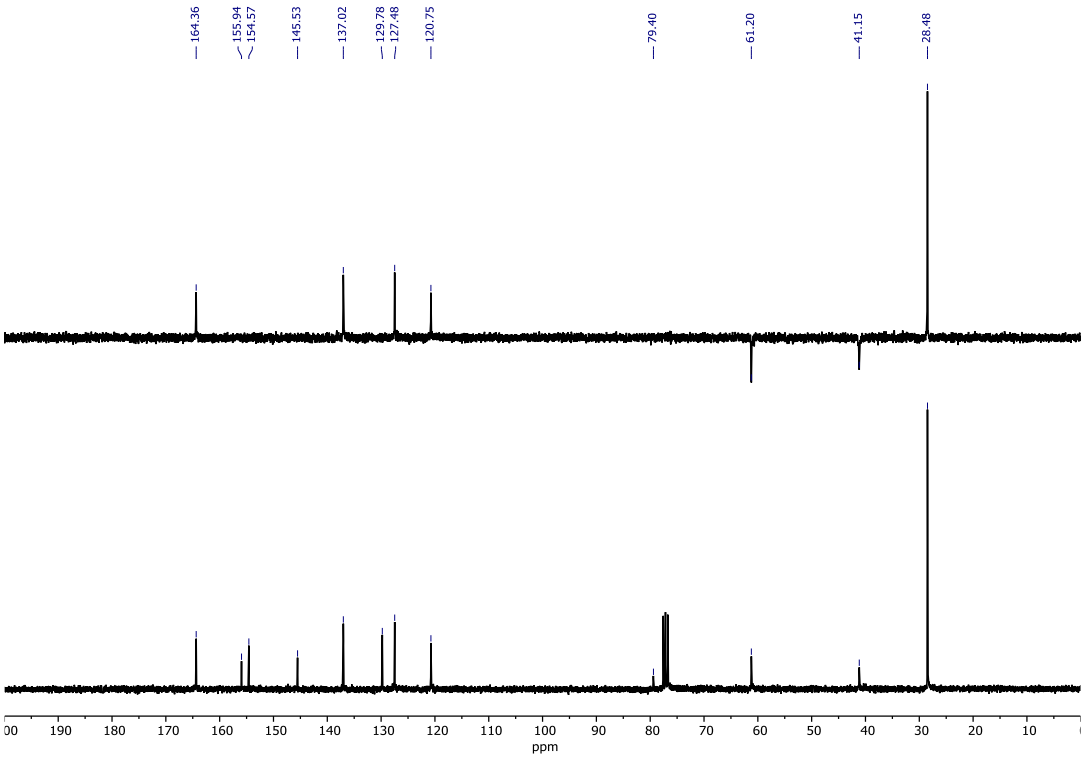


Figure B5. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **1a**.

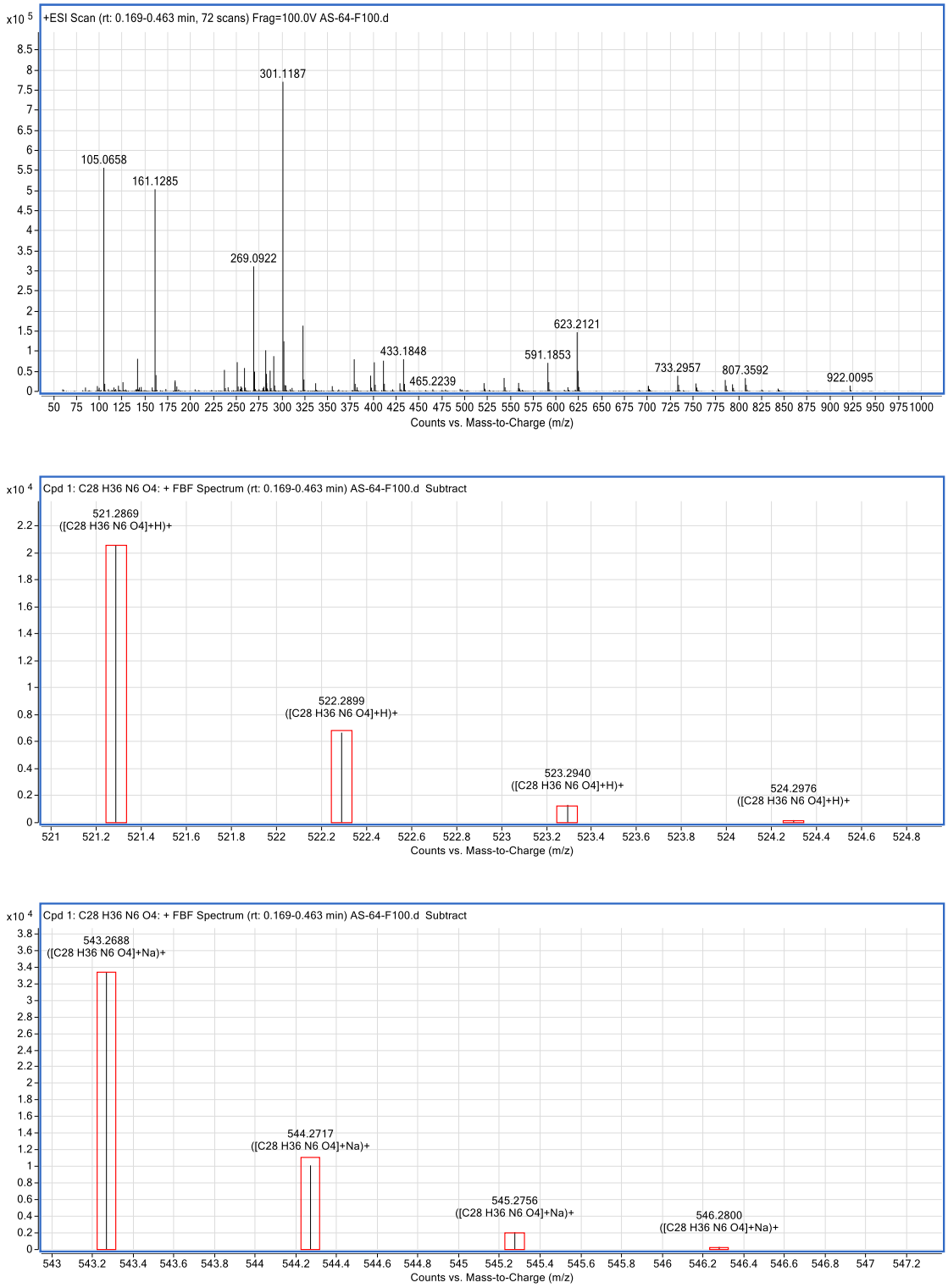


Figure B6. HR-MS (+ESI) spectra for compound **1a**.

Compound **2a**

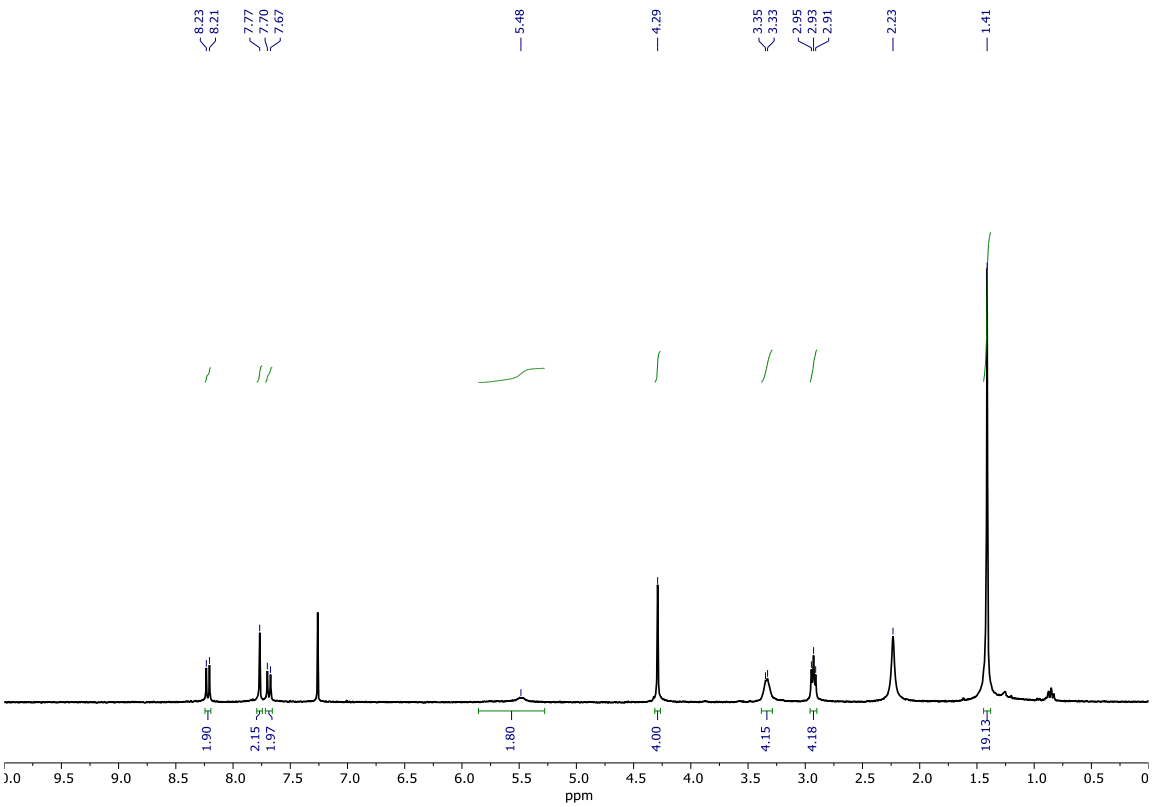


Figure B7. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **2a**.

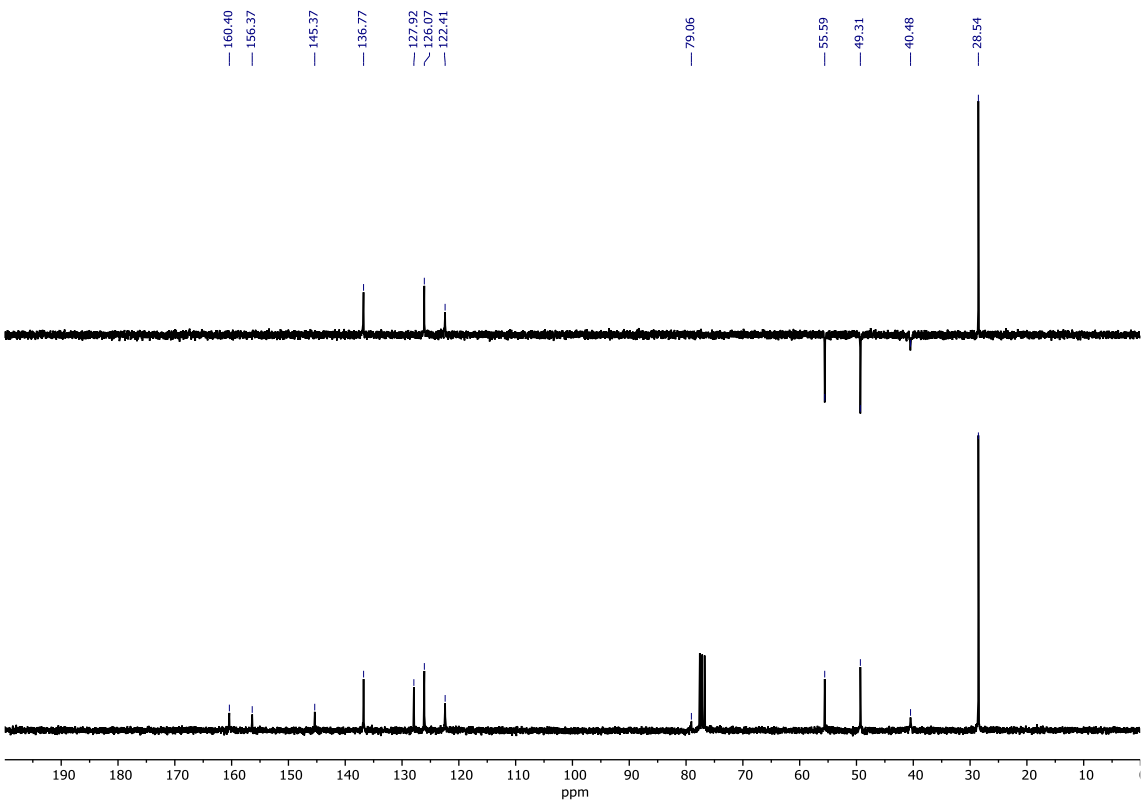


Figure B8. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **2a**.

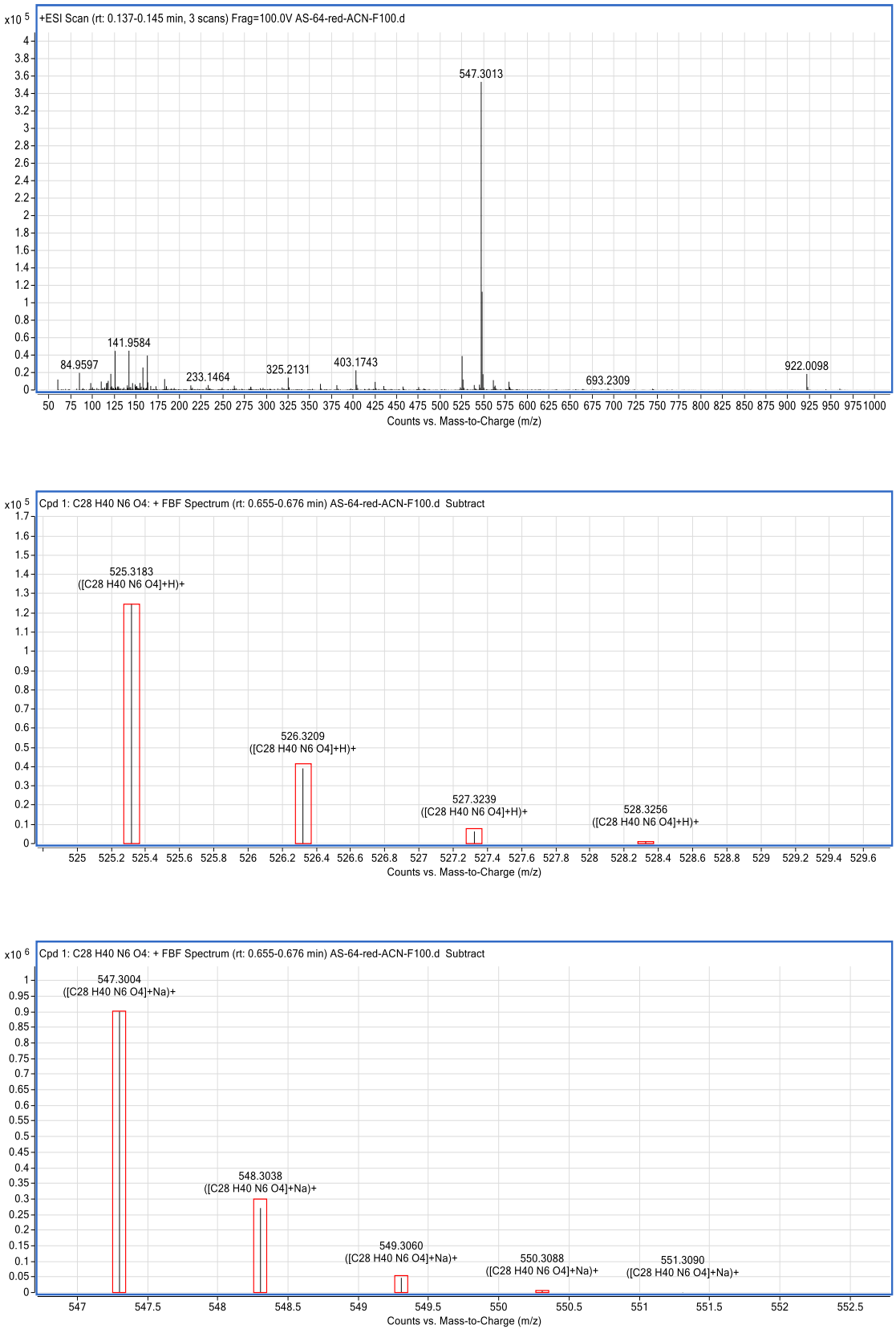


Figure B9. HR-MS (+ESI) spectra for compound **2a**.

Compound **3a**

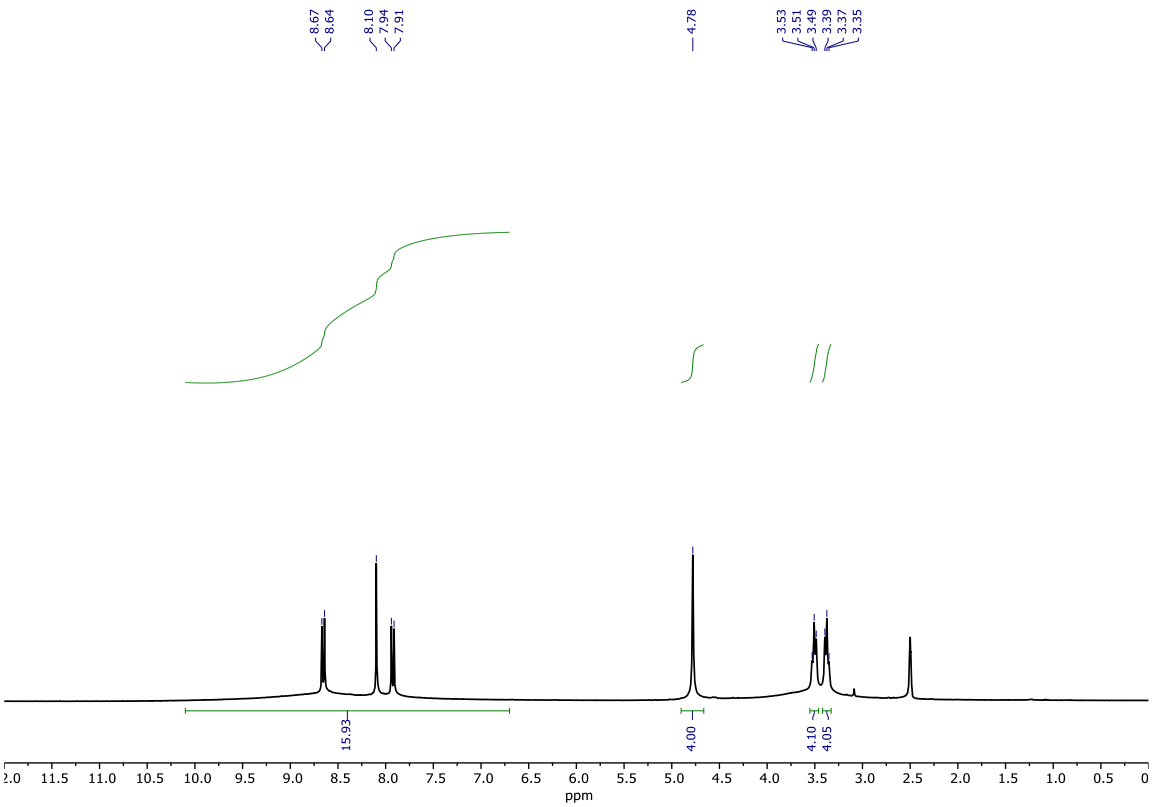


Figure B10. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **3a**.

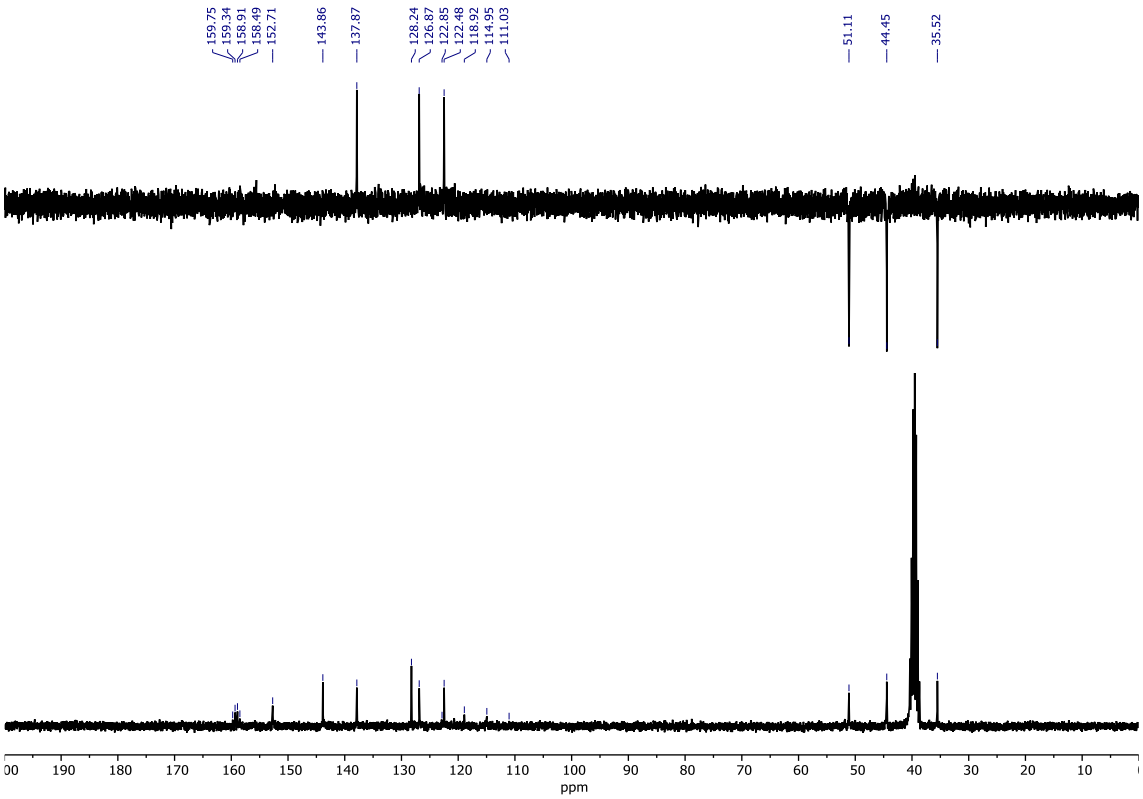


Figure B11. ^{13}C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **3a**.

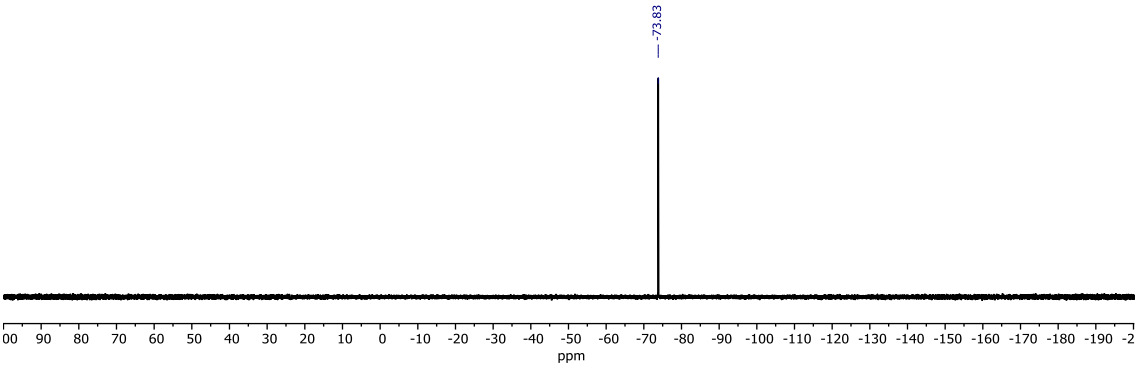


Figure B12. ^{19}F NMR spectrum (300 MHz, DMSO-d_6) for compound **3a.**

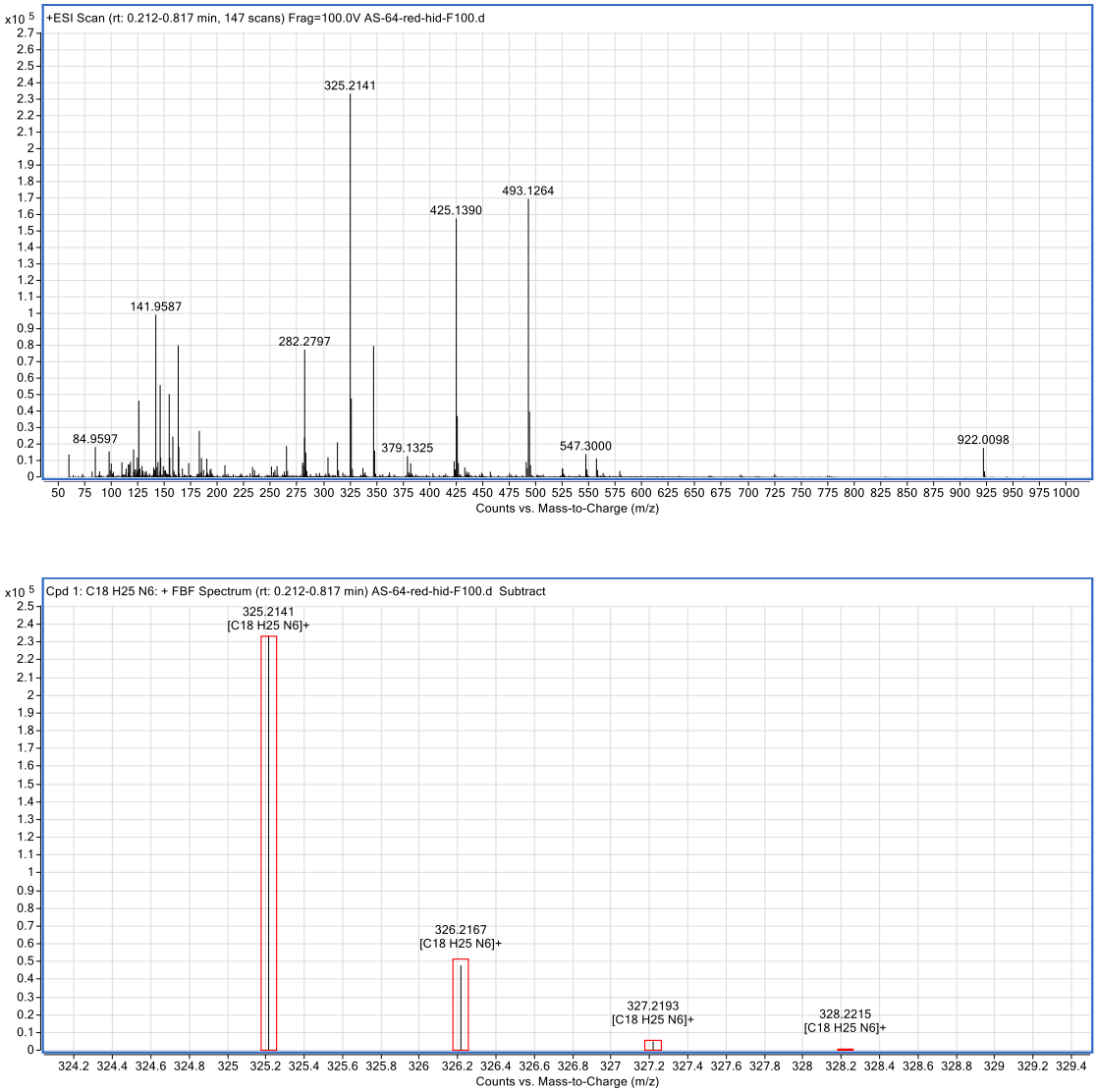


Figure B13. HR-MS (+ESI) spectra for compound **3a**.

Compound **1b**

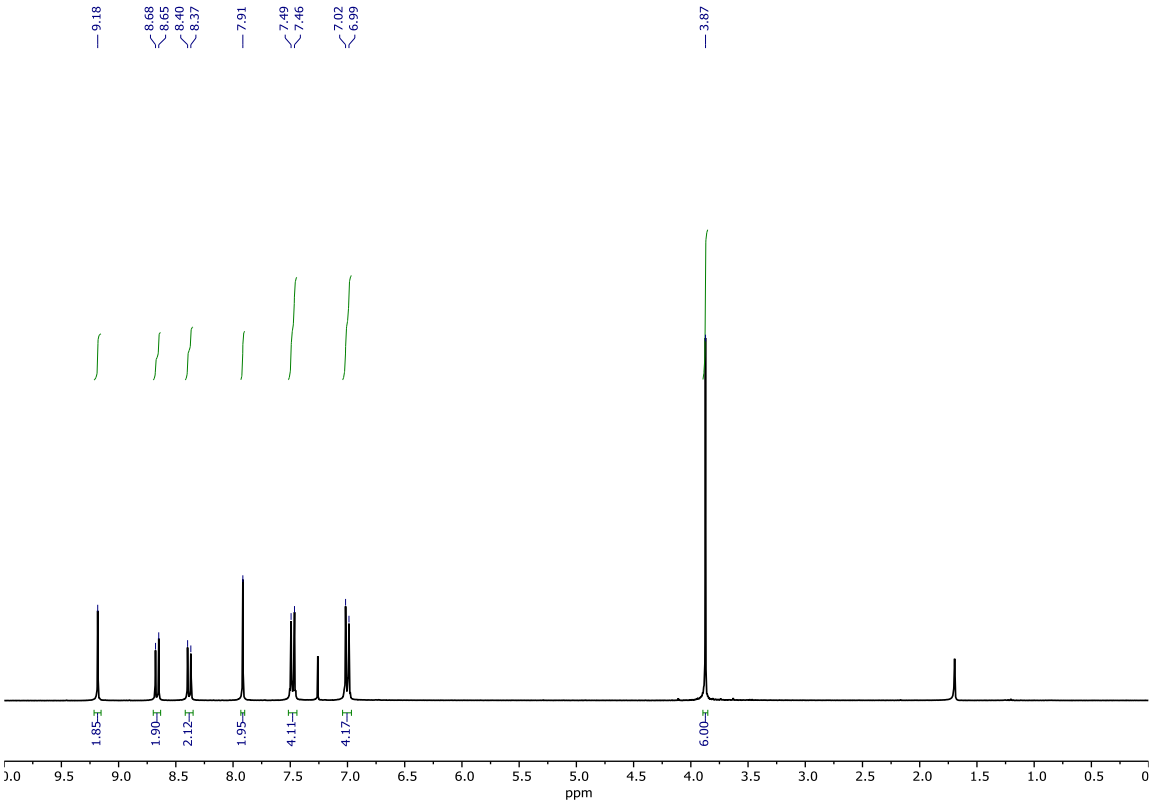


Figure B14. [1]H NMR spectrum (300 MHz, CDCl3) for compound **1b**.

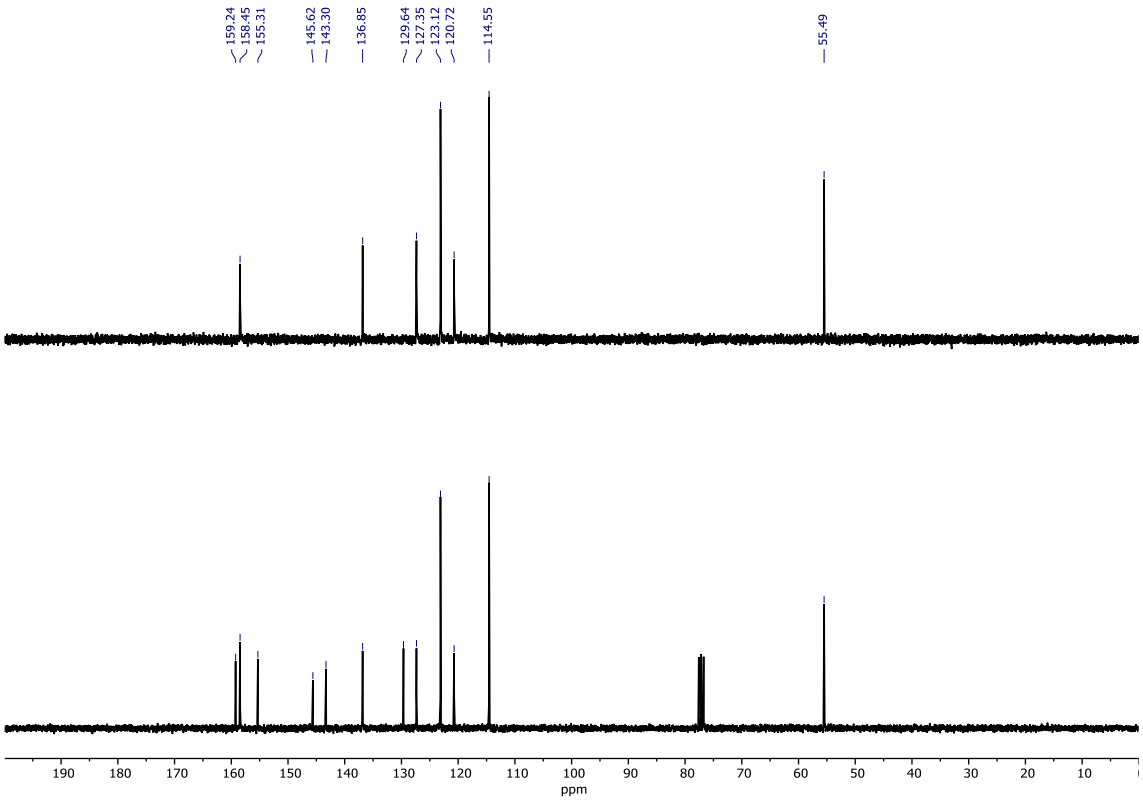


Figure B15. [13]C and DEPT-135 NMR spectra (75 MHz, CDCl3) for compound **1b**.

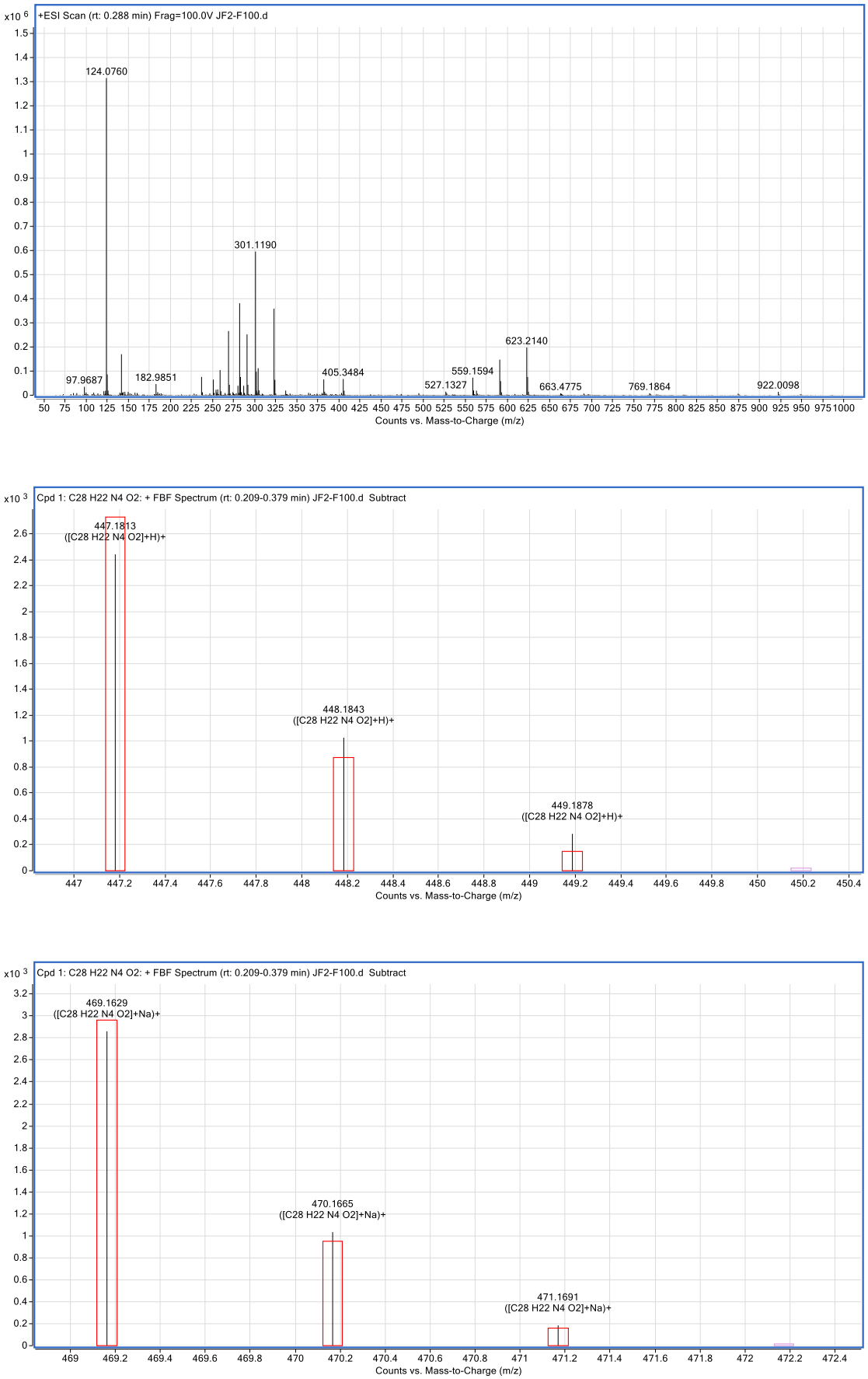


Figure B16. HR-MS (+ESI) spectra for compound **1b**.

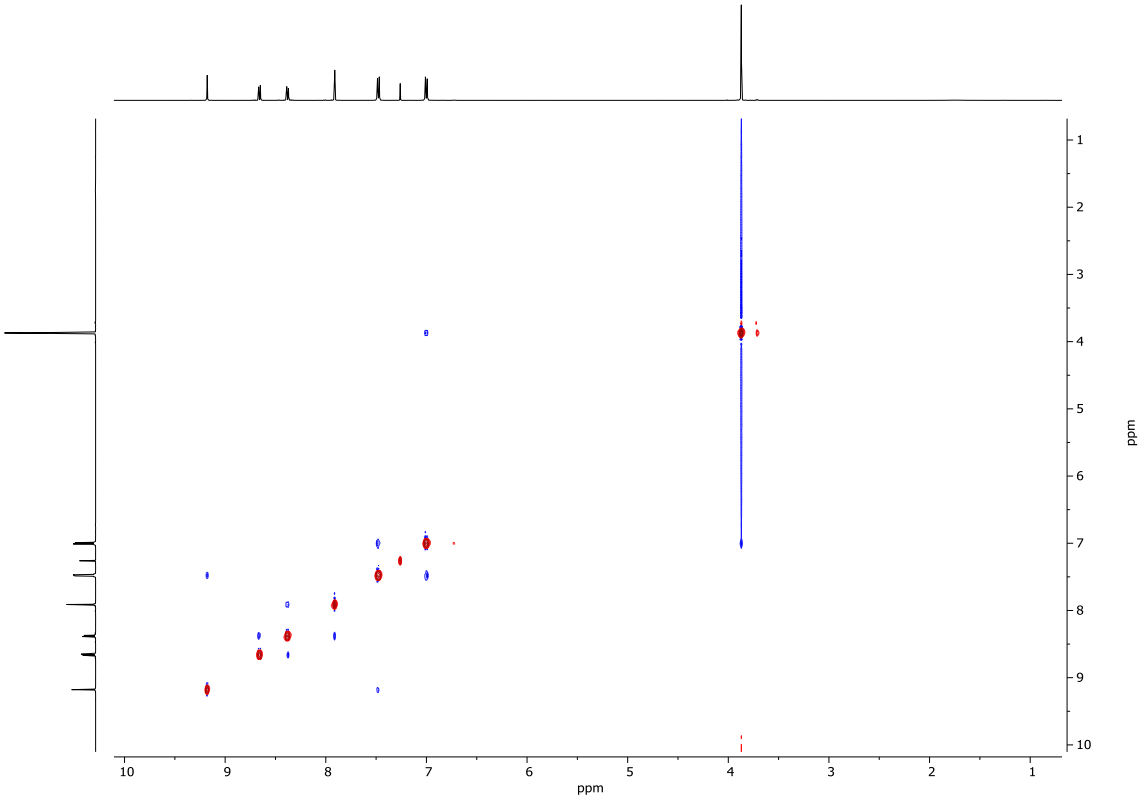


Figure B17. NOESY spectrum (500 MHz, $CDCl_3$) for compound **1b**.

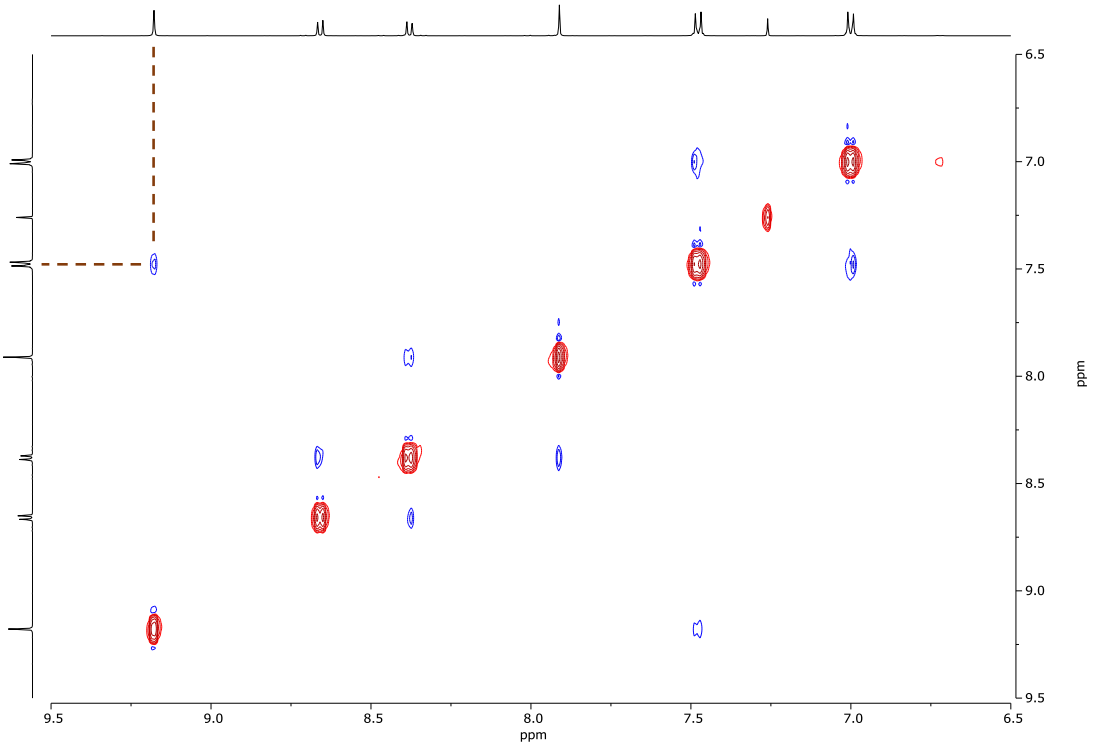


Figure B18. Enlargement of the aromatic region of the NOESY spectrum (500 MHz, $CDCl_3$) for compound **1b**, and correlation between the imine CH proton and the closest proton of the 4-methoxyphenyl ring.

Compound **2b**

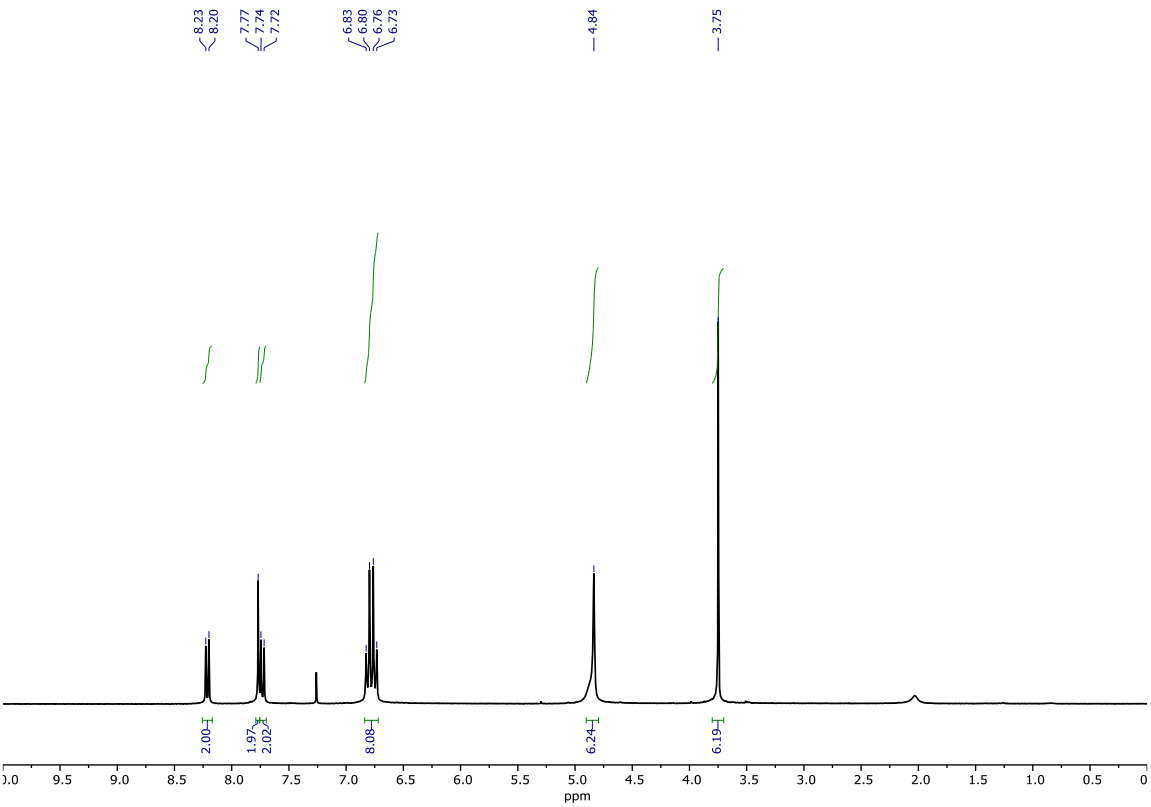


Figure B19. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **2b**.

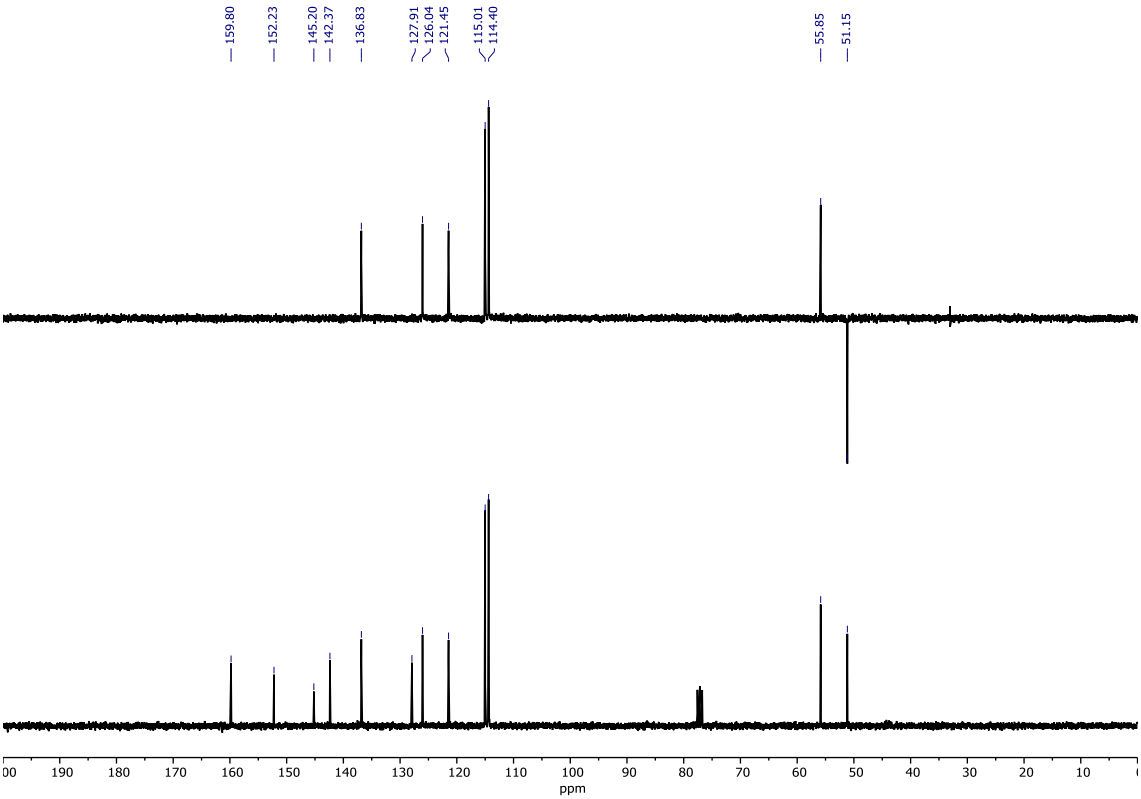
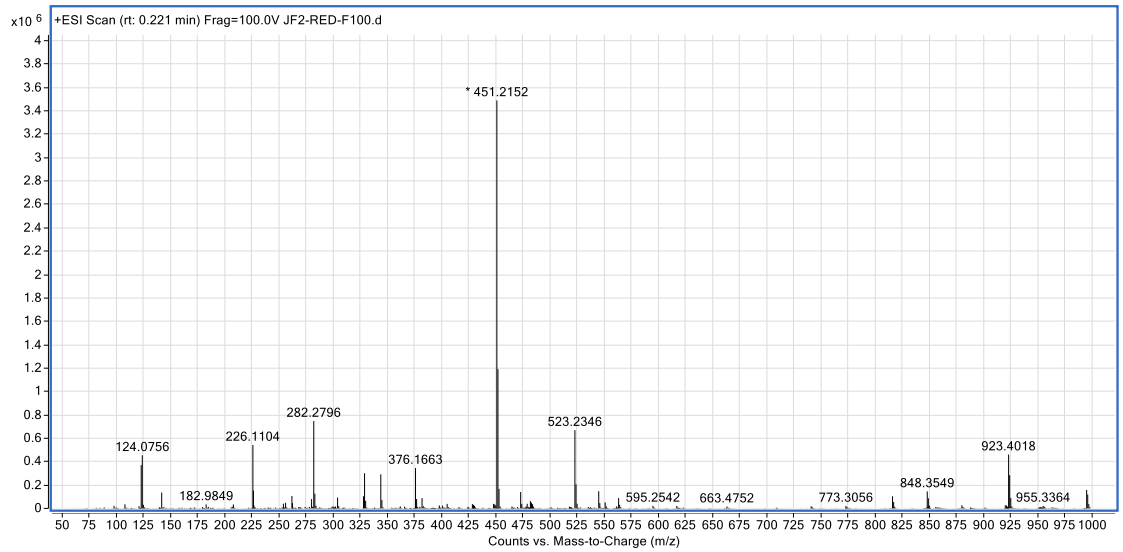
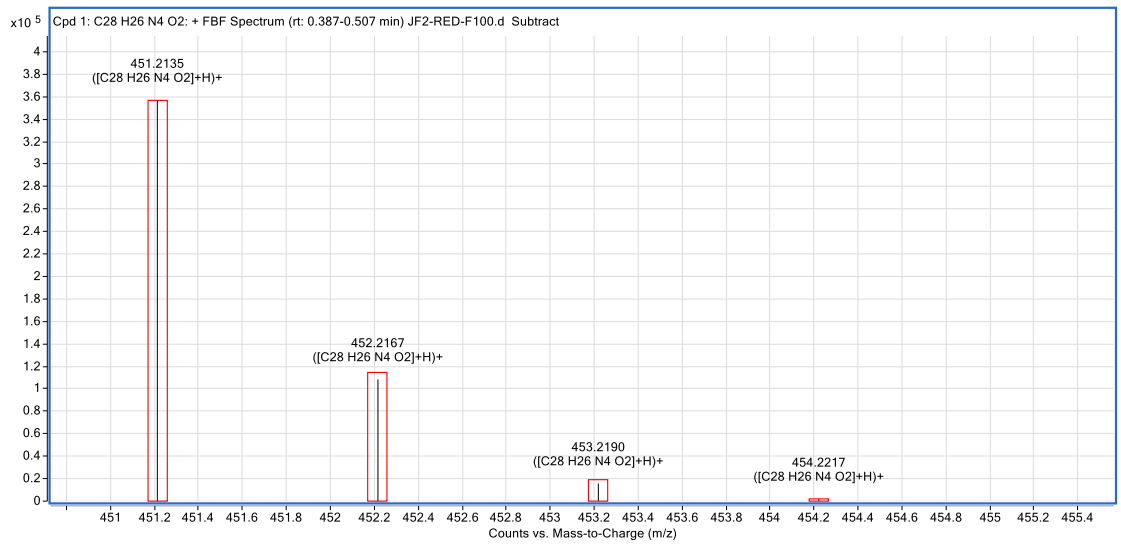


Figure B20. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **2b**.





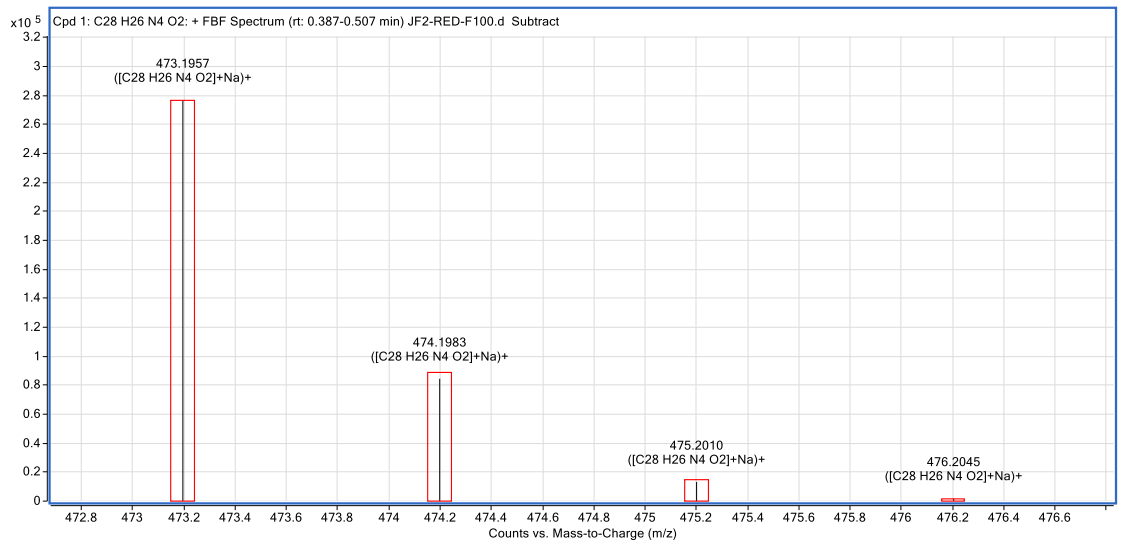


Figure B21. HR-MS (+ESI) spectra for compound **2b**.

Compound **1c**

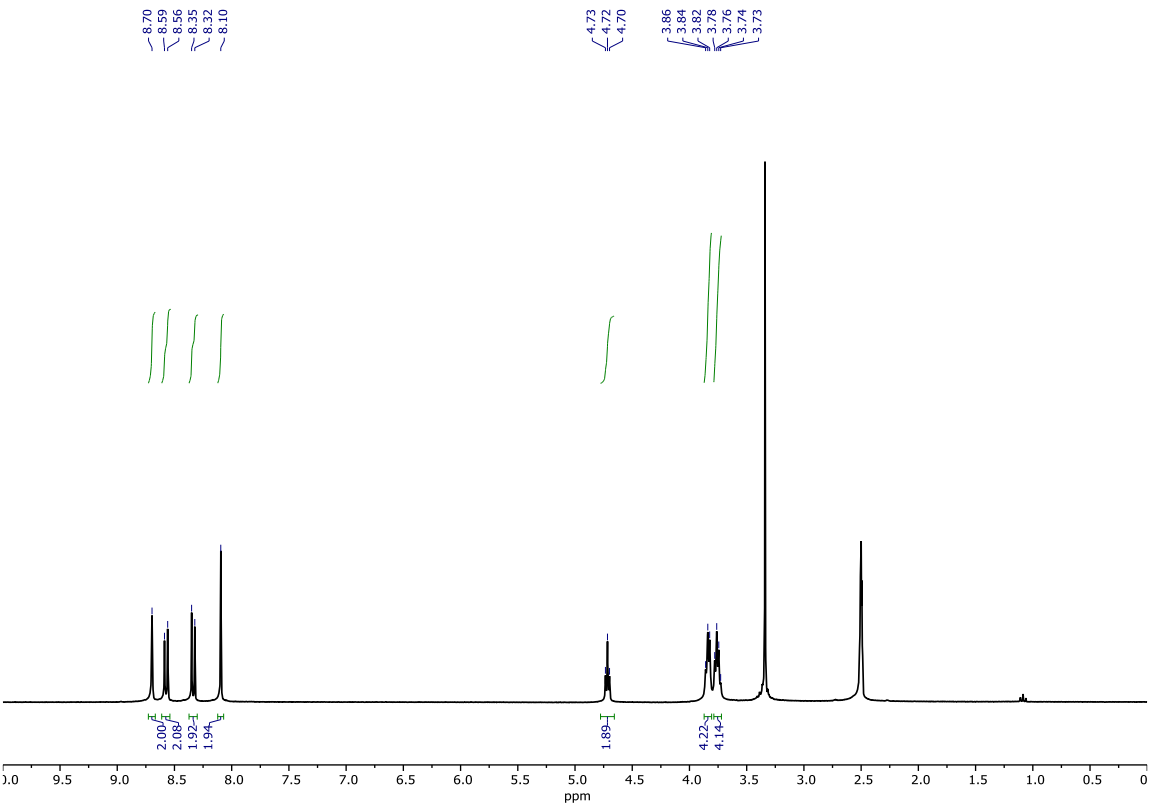


Figure B22. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **1c**.

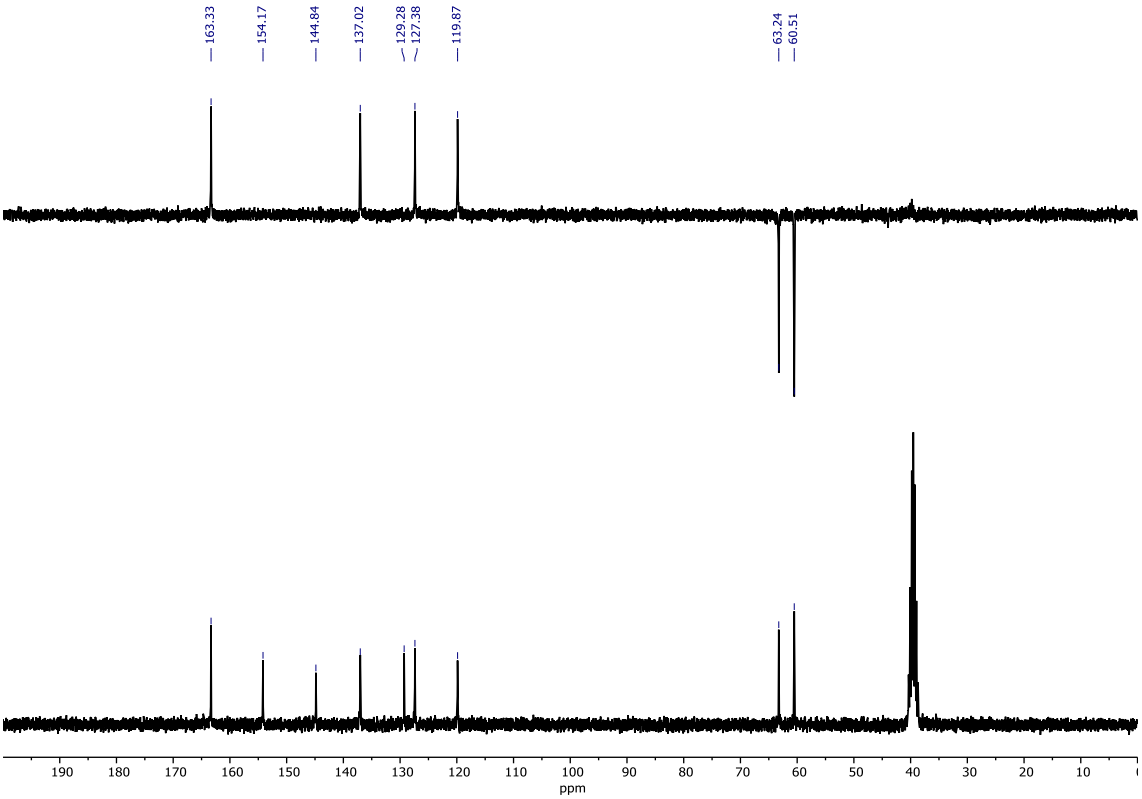


Figure B23. ^{13}C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **1c**.

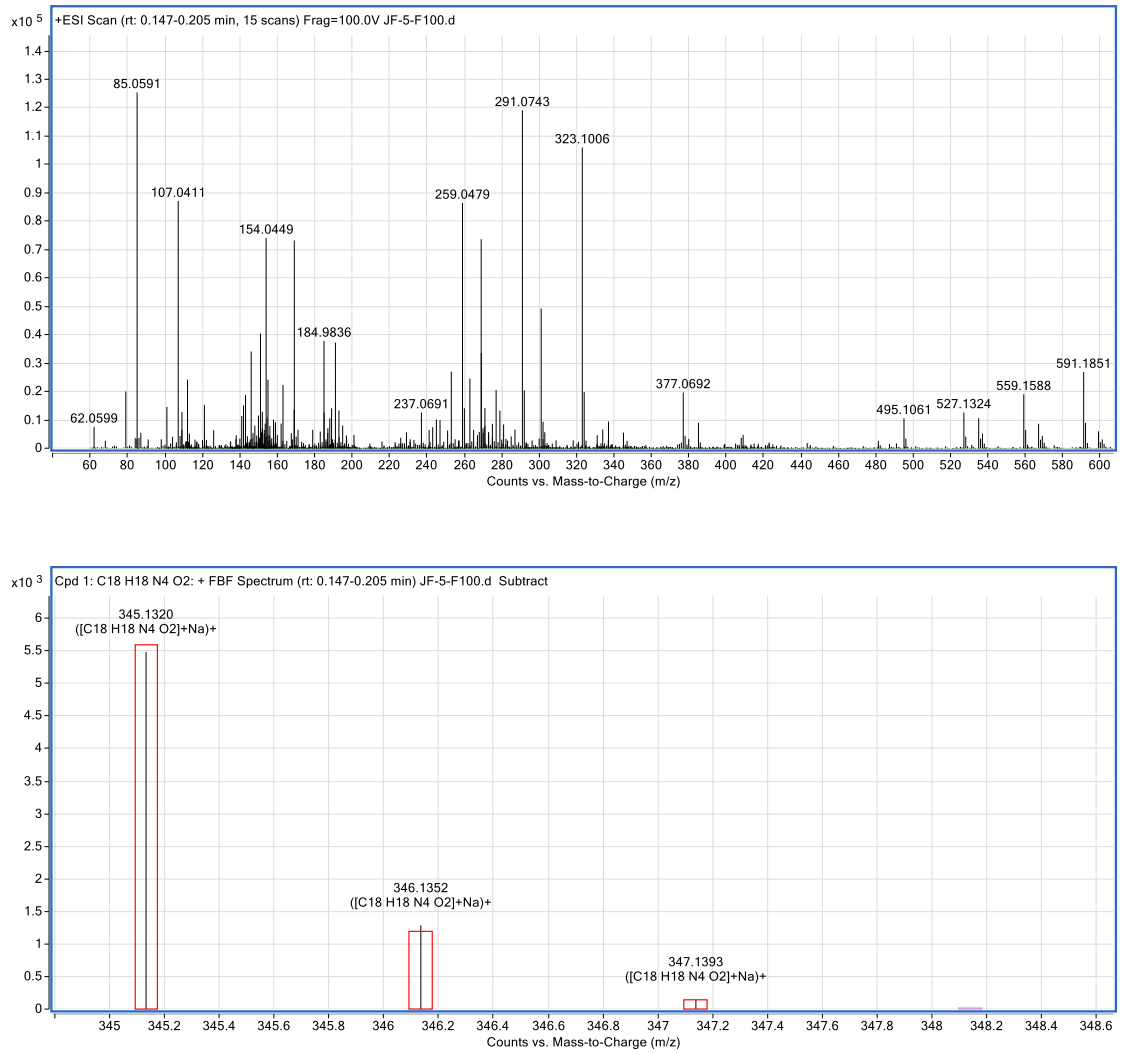


Figure B24. HR-MS (+ESI) spectra for compound **1c**.

Compound **2c**

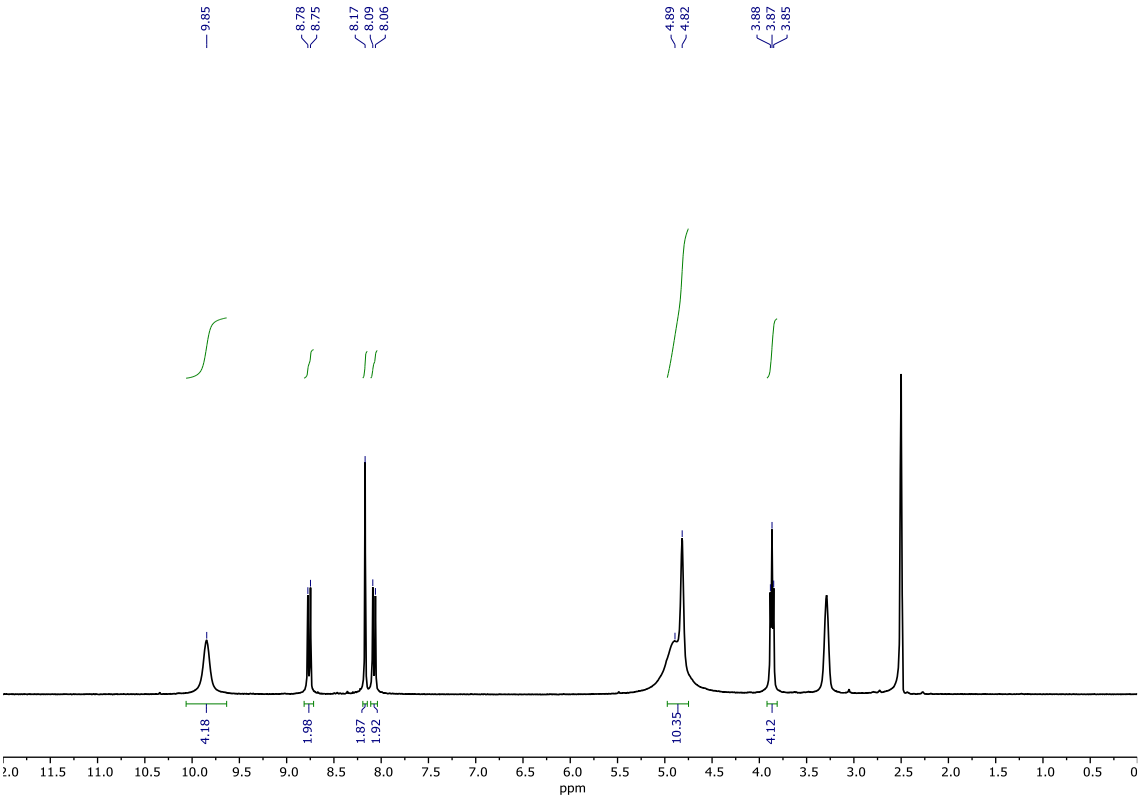


Figure B25. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **2c**.

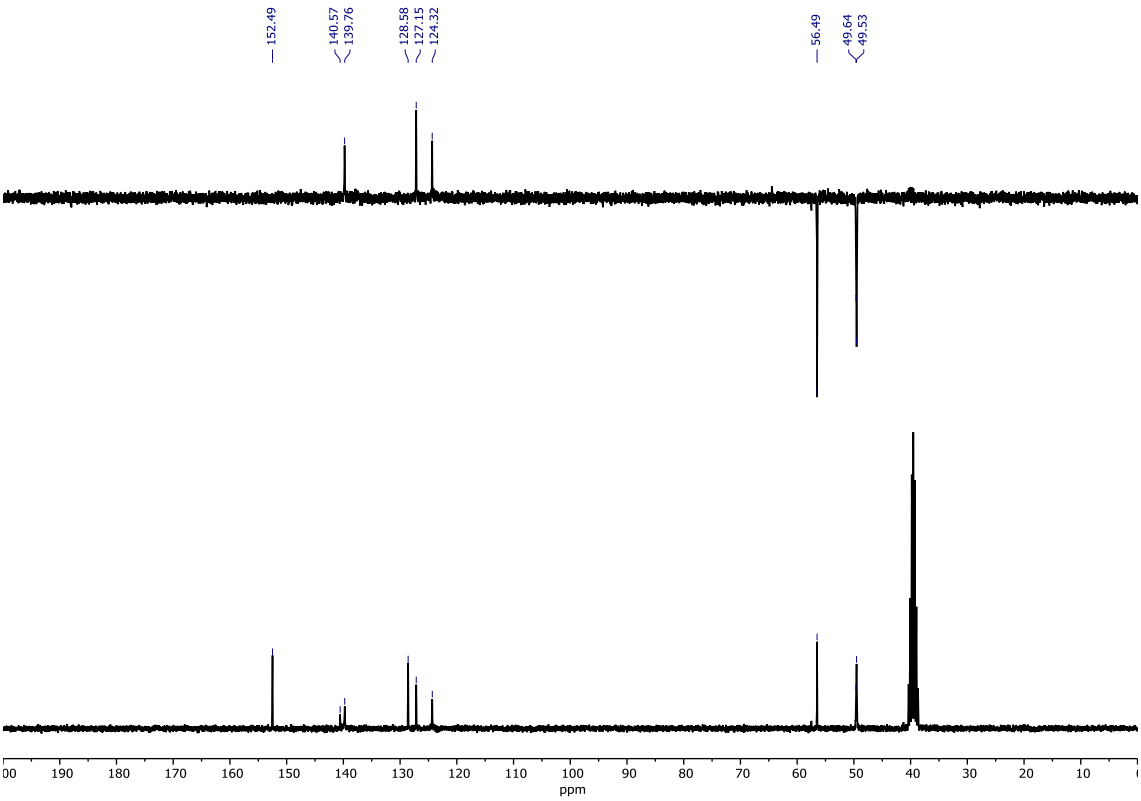
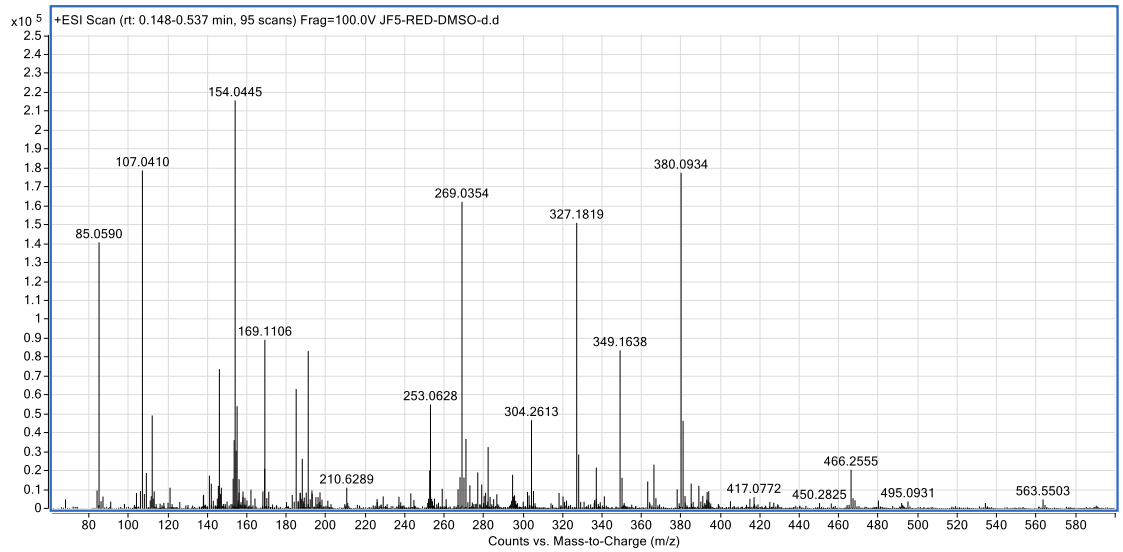


Figure B26. ^{13}C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **2c**.



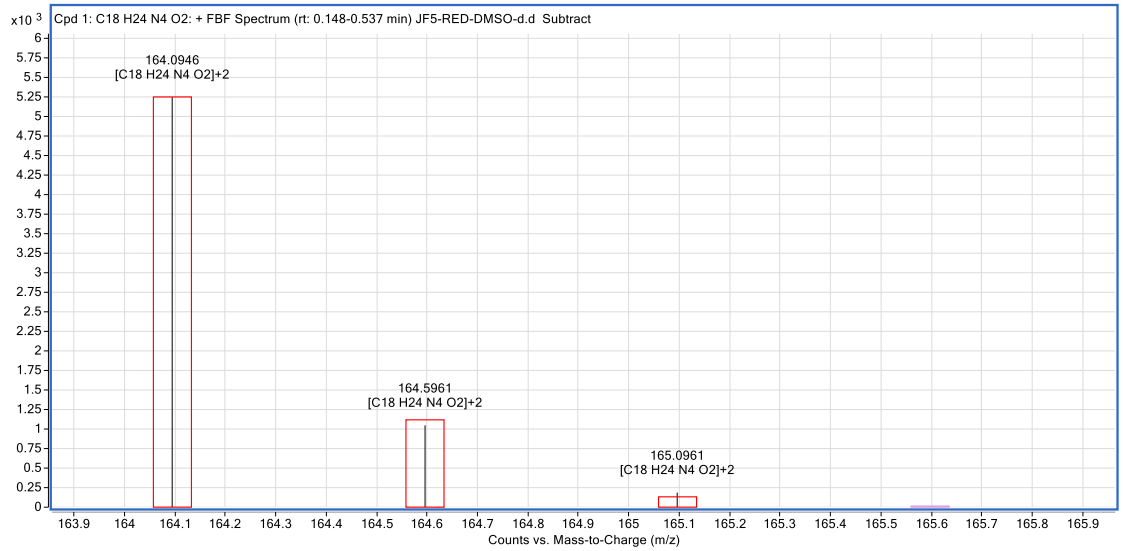


Figure B27. HR-MS (+ESI) spectra for compound **2c**.

Compound **1d**

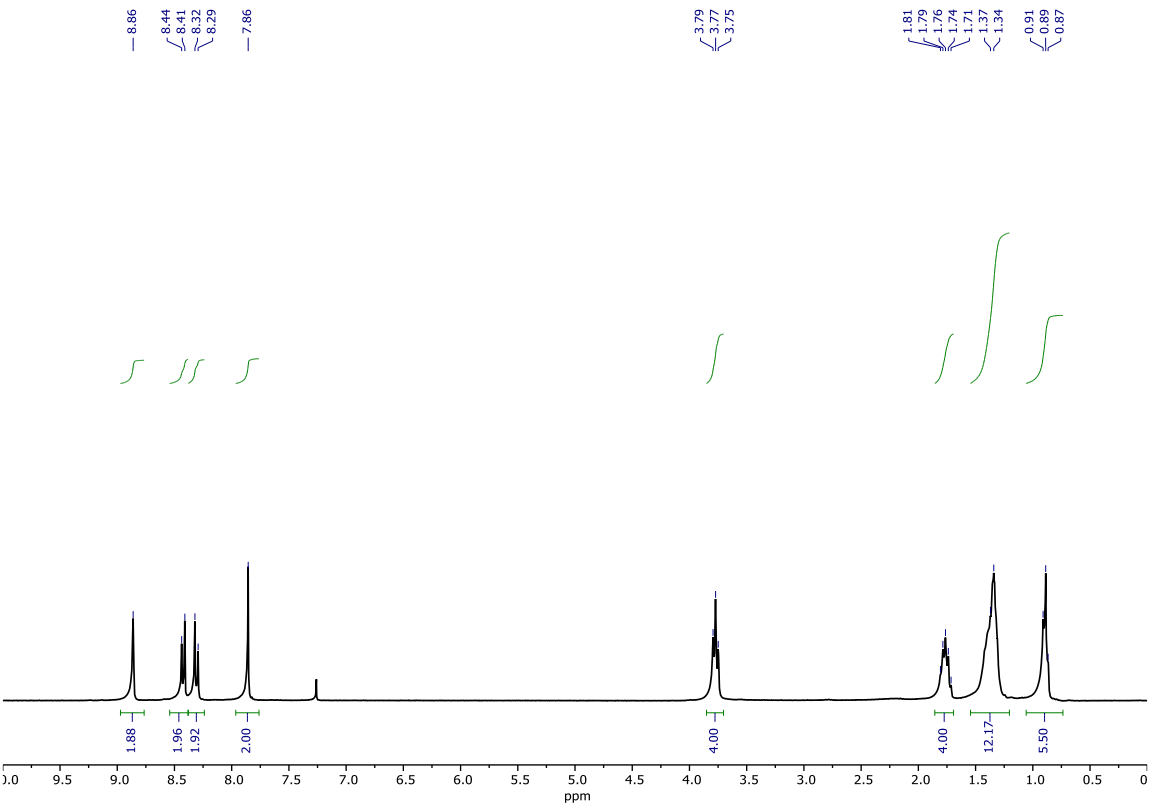


Figure B28. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **1d**.

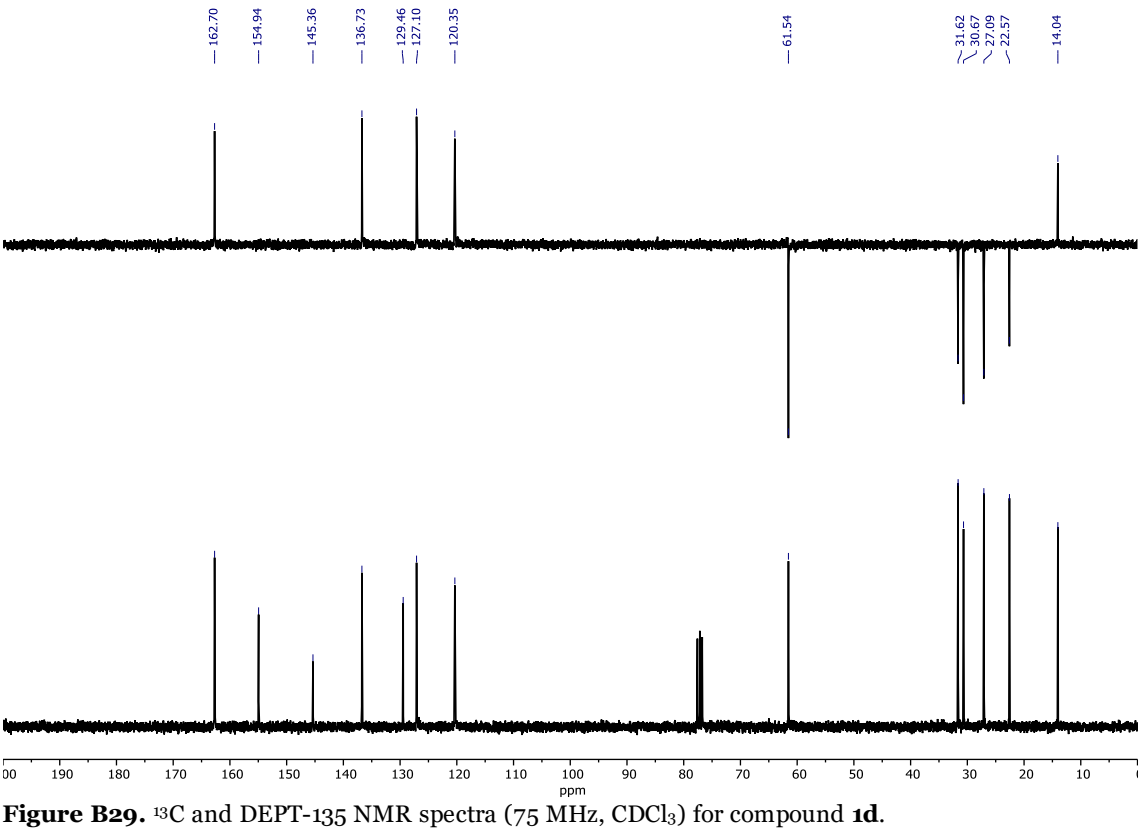
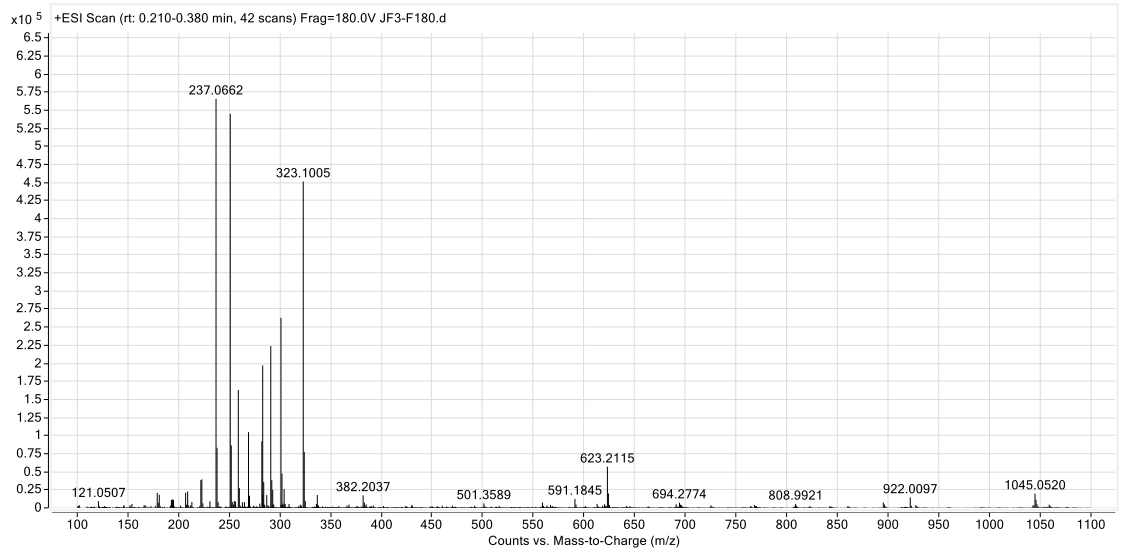


Figure B29. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **1d**.



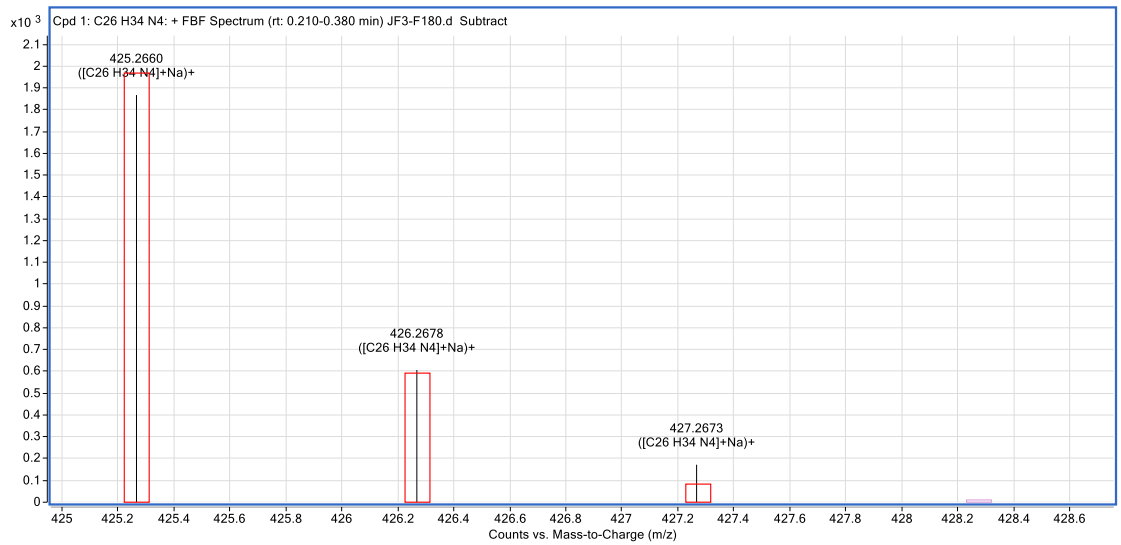


Figure B30. HR-MS (+ESI) spectra for compound **1d**.

Compound **2d**

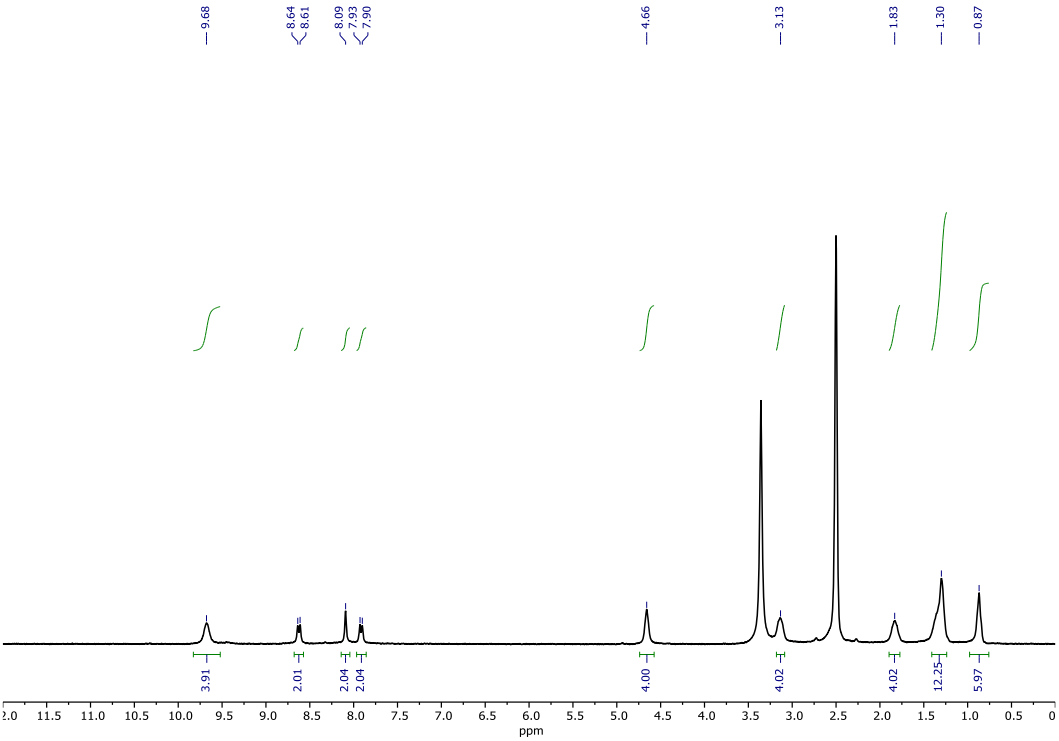


Figure B31. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **2d**.

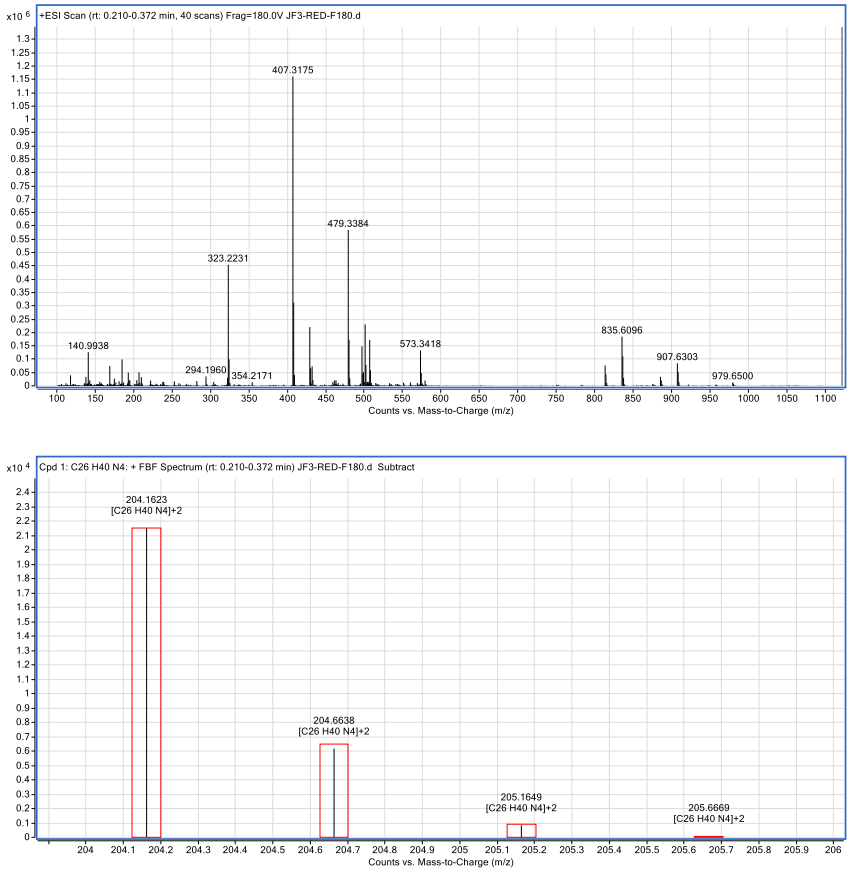


Figure B32. HR-MS (+ESI) spectra for compound **2d**.

Compound **1e**

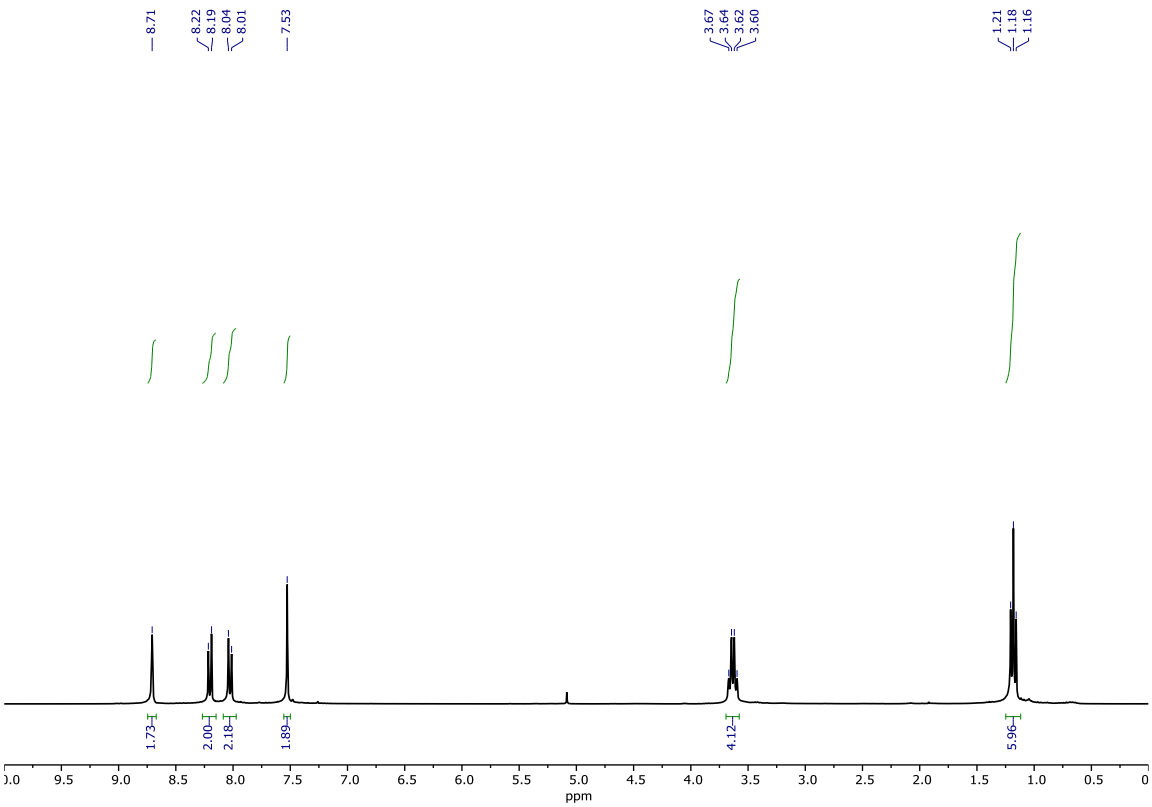


Figure B33. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **1e**.

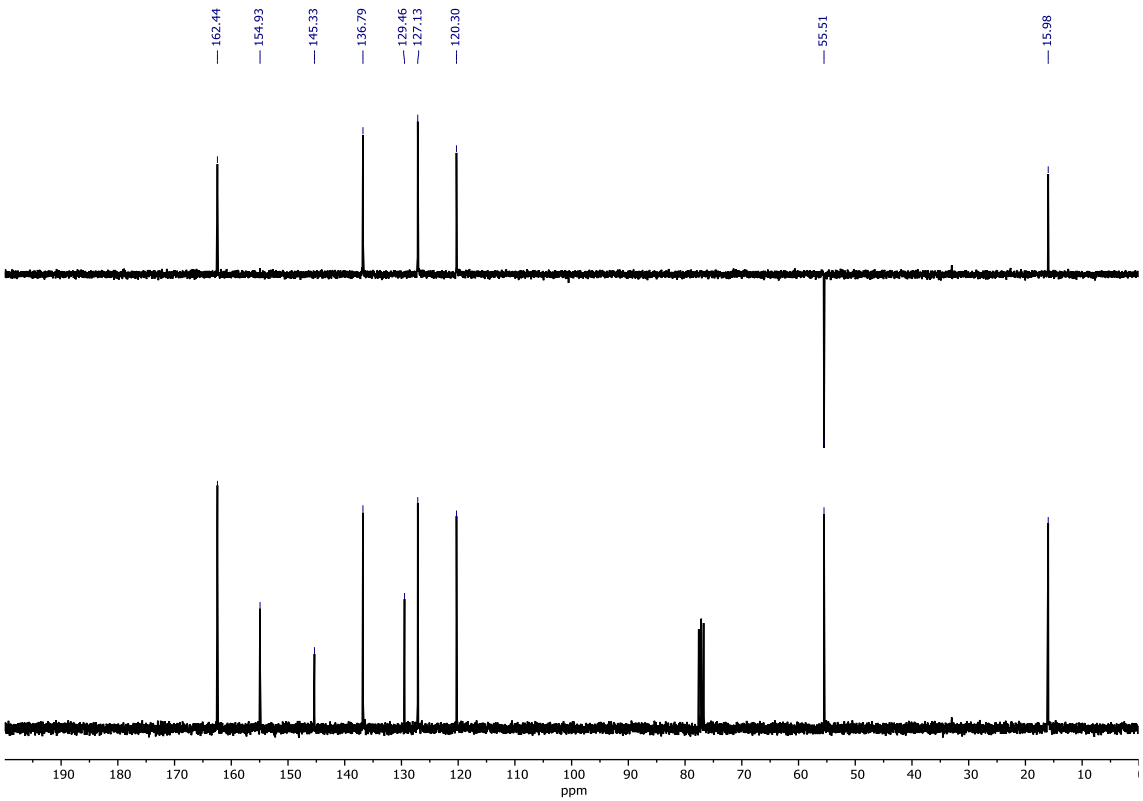


Figure B34. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **1e**.

Amides (including macrocycles)

Compound **4a**

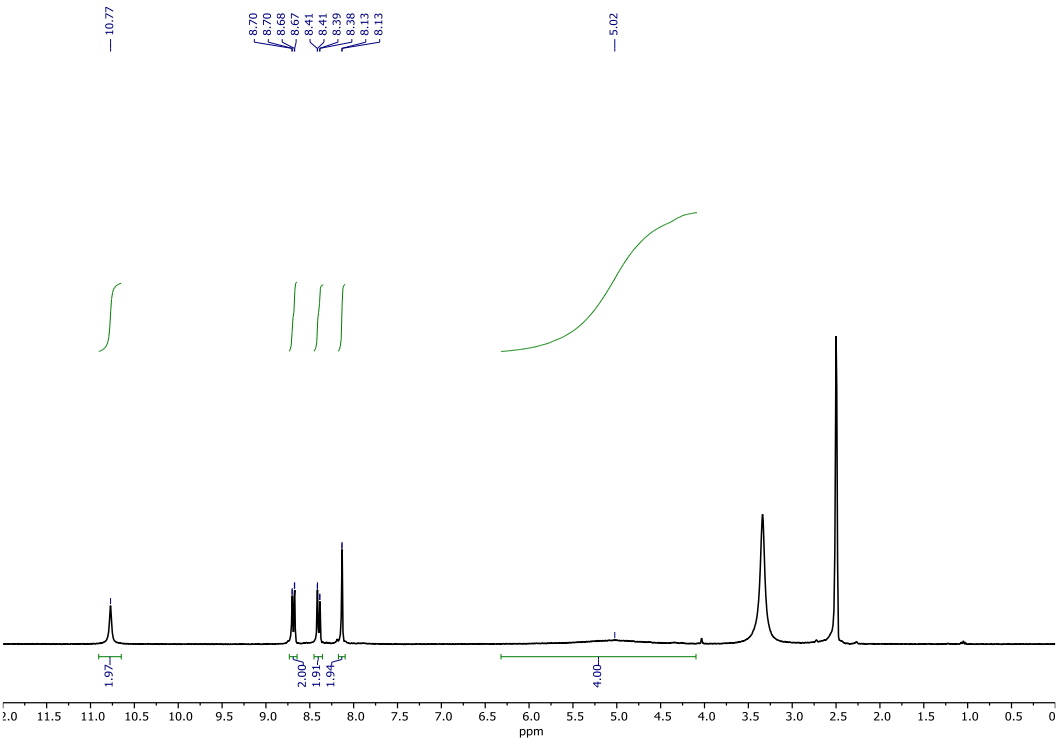


Figure B35. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **4a**.

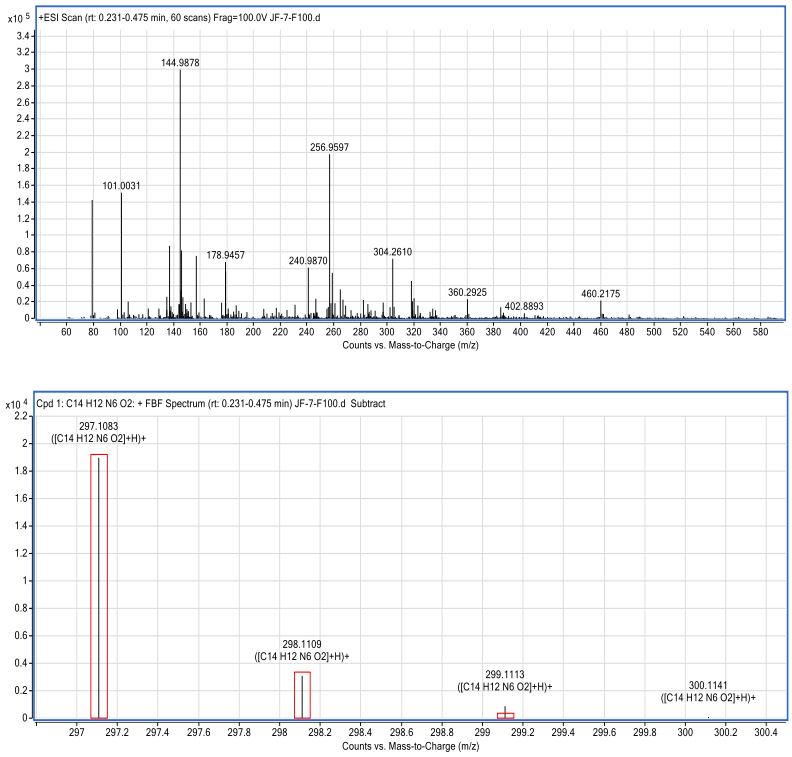


Figure B36. HR-MS (+ESI) spectra for compound **4a**.

Compound **5a**

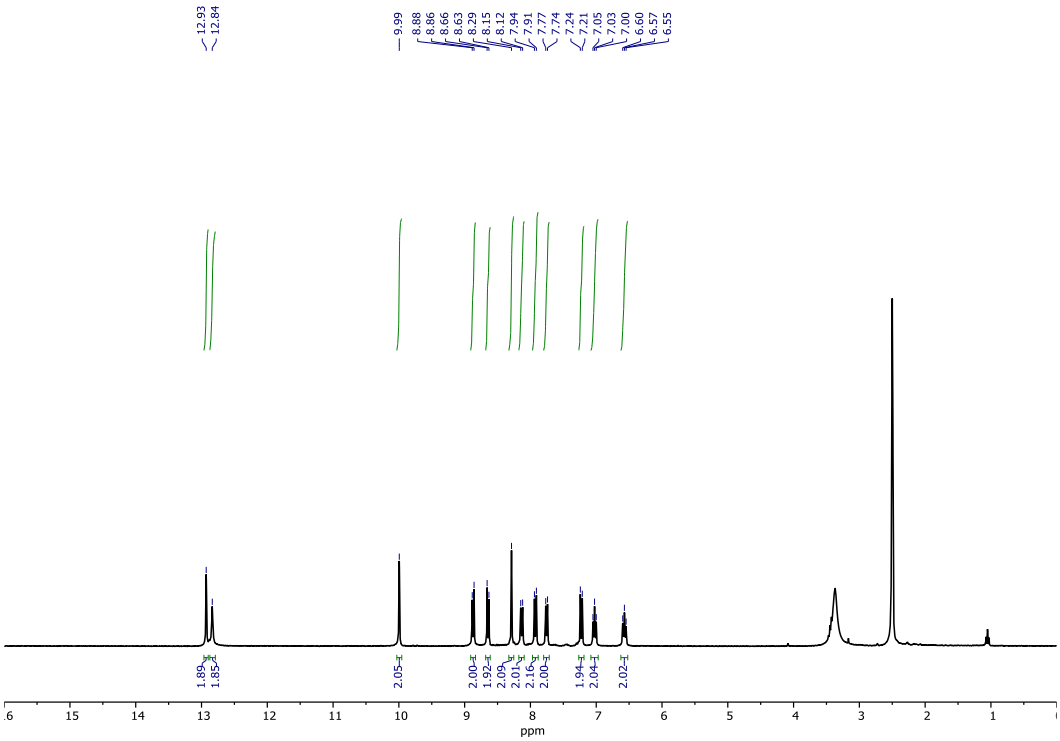


Figure B37. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **5a**.

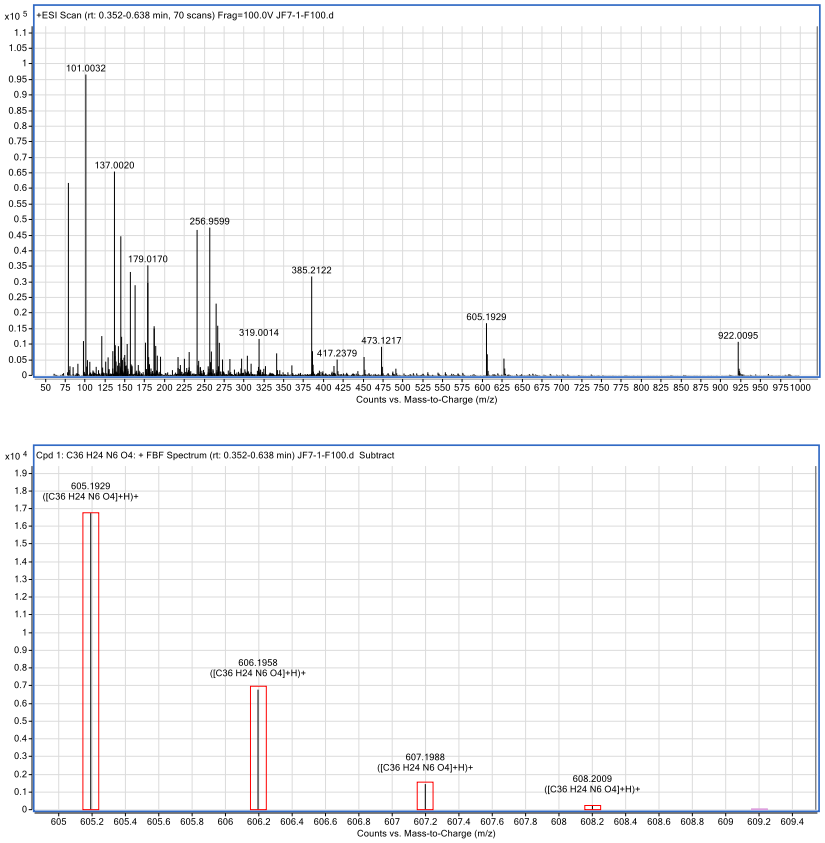


Figure B38. HR-MS (+ESI) spectra for compound **5a**.

Compound **4b**

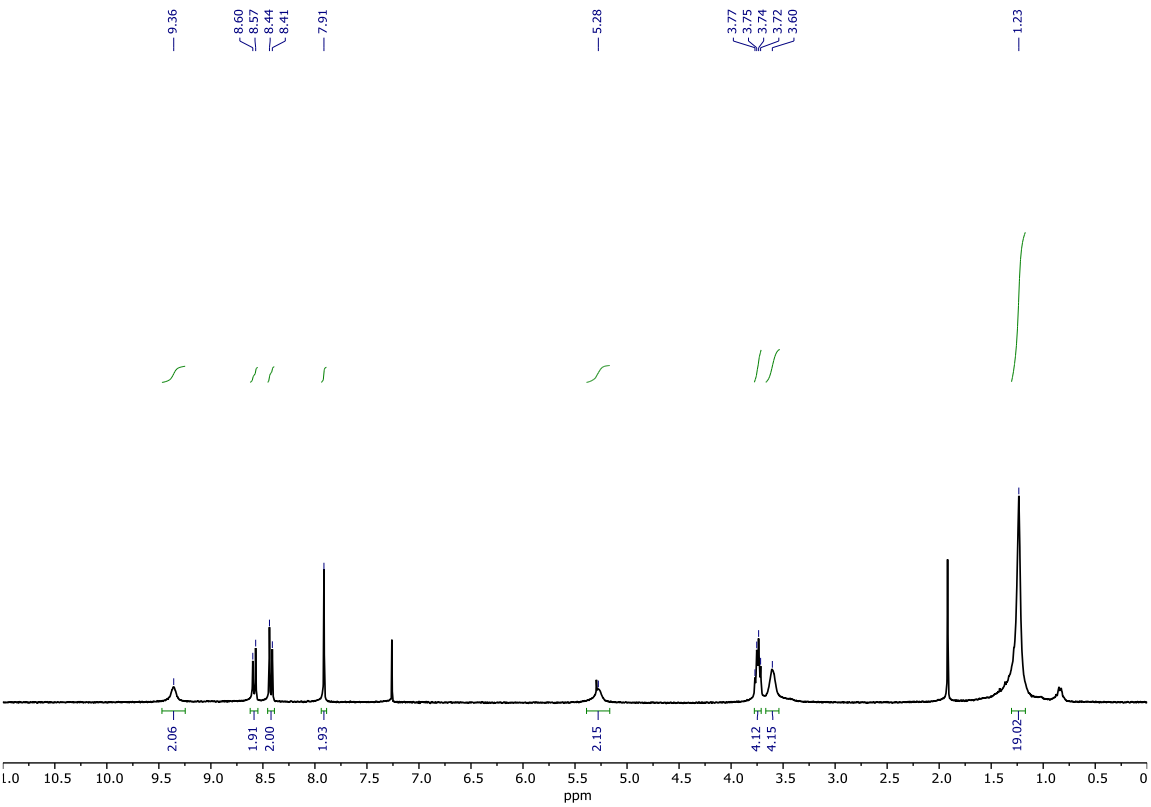


Figure B39. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **4b**.

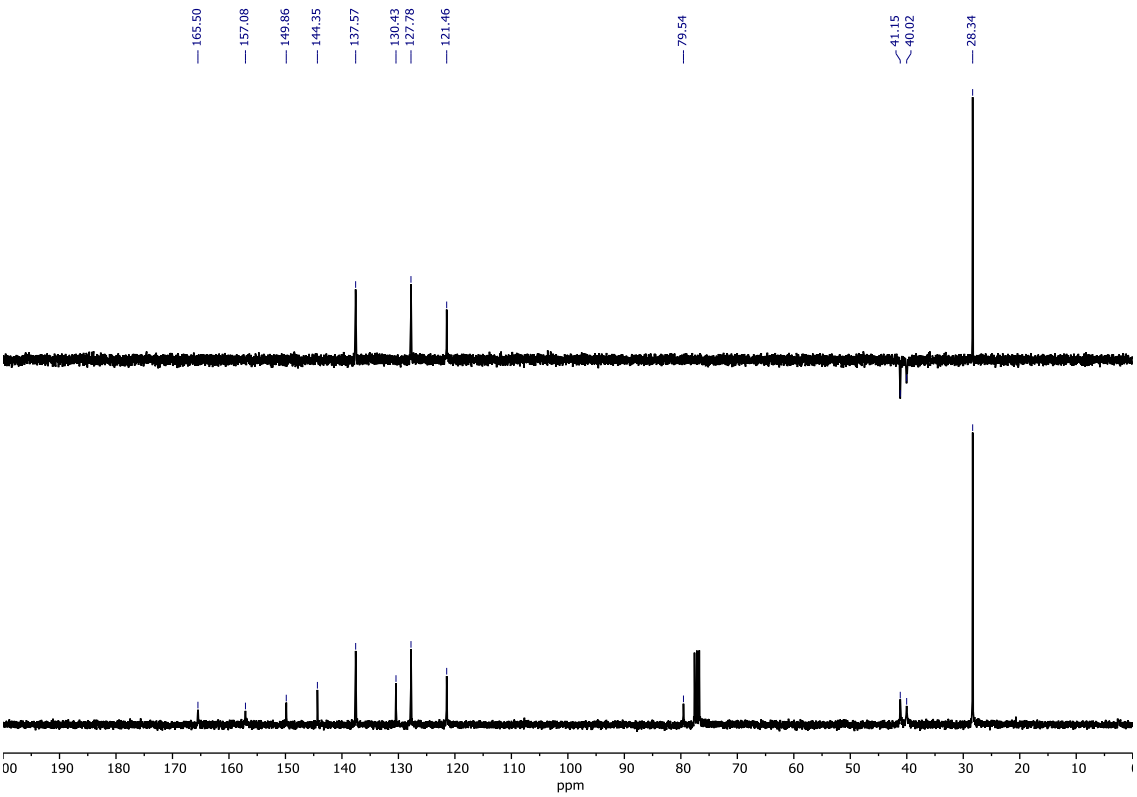


Figure B40. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **4b**.

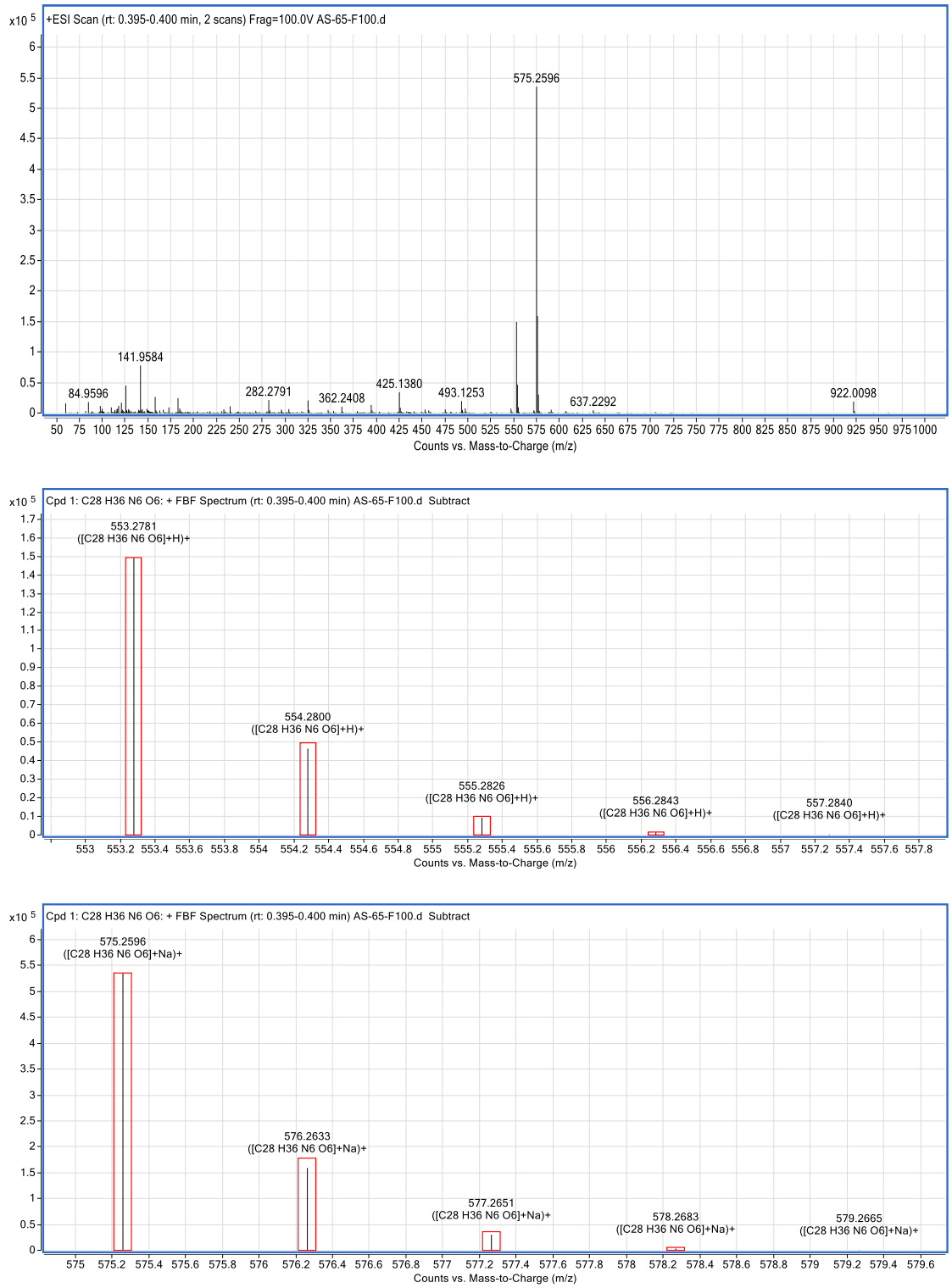


Figure B41. HR-MS (+ESI) spectra for compound **4b**.

Compound **5b**

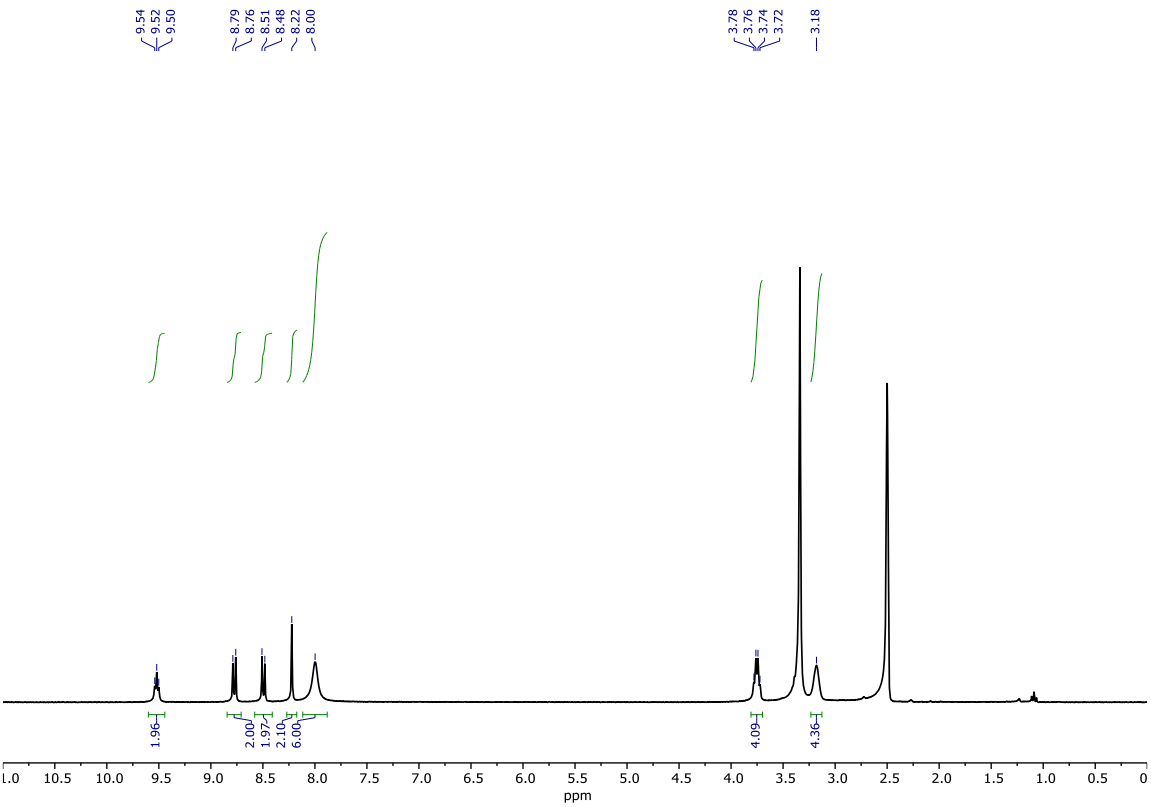


Figure B42. [1]H NMR spectrum (300 MHz, DMSO-d_6) for compound **5b**.

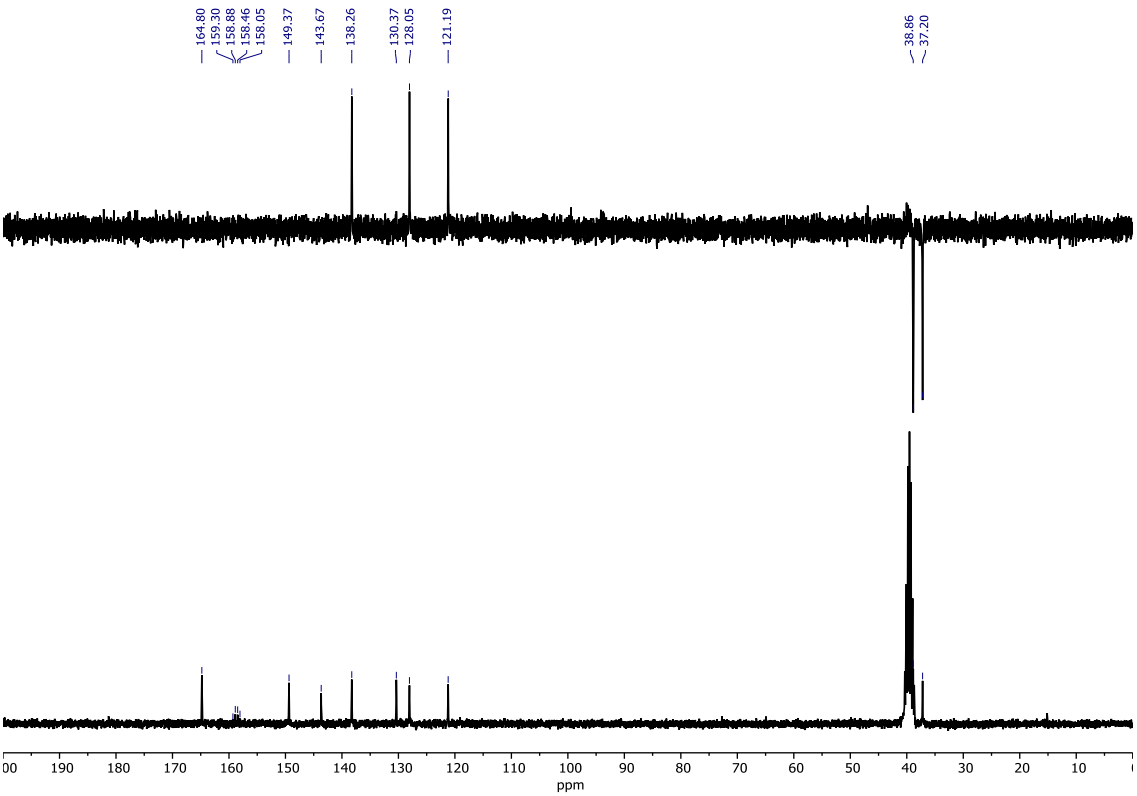


Figure B43. [13]C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **5b**.

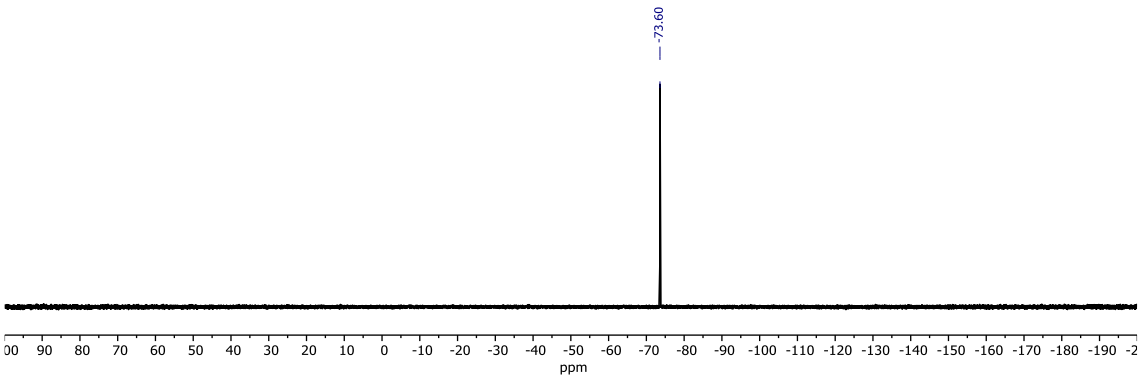


Figure B44. ^{19}F NMR spectrum (300 MHz, DMSO-d_6) for compound **5b**.

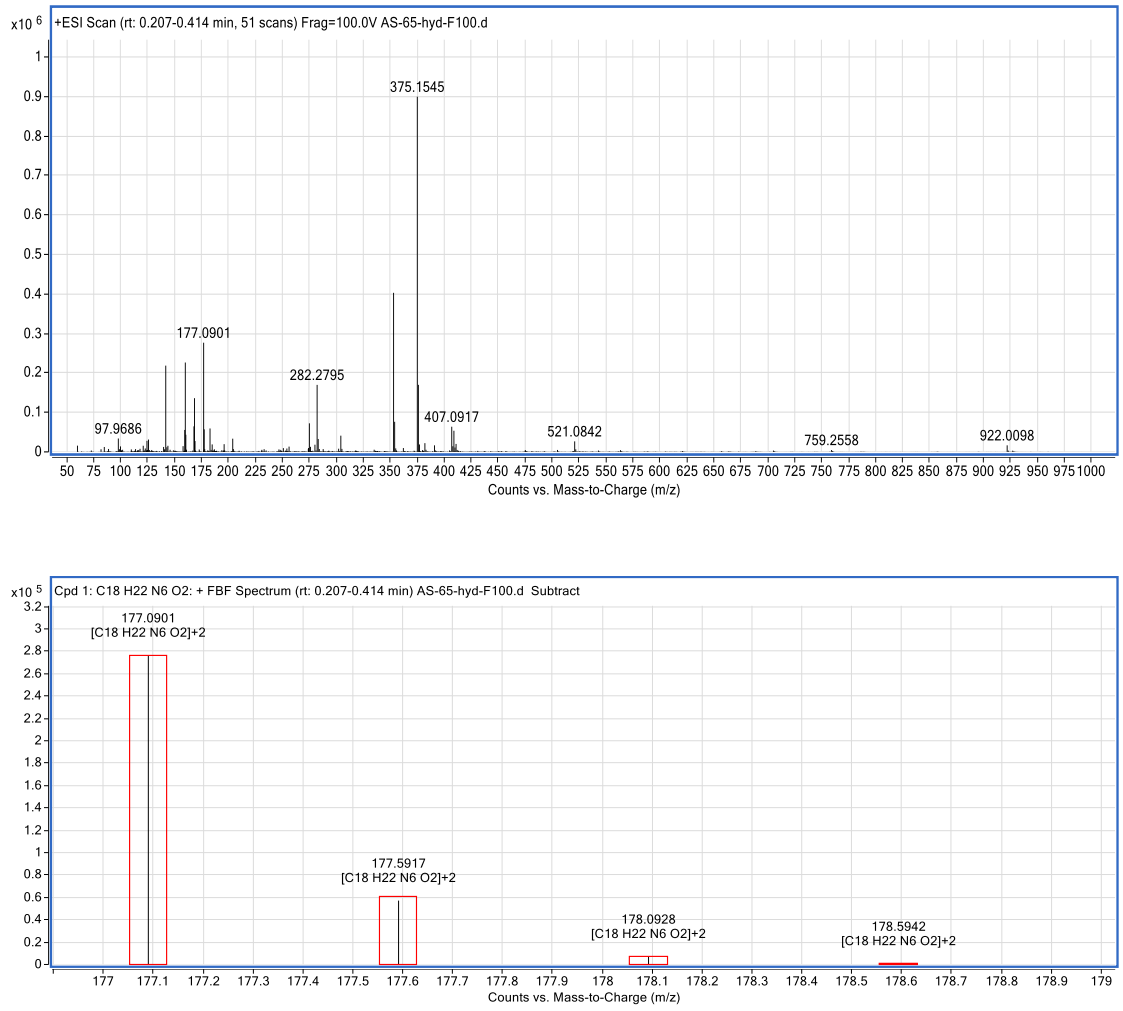


Figure B45. HR-MS (+ESI) spectra for compound **5b**.

Compound **4c**

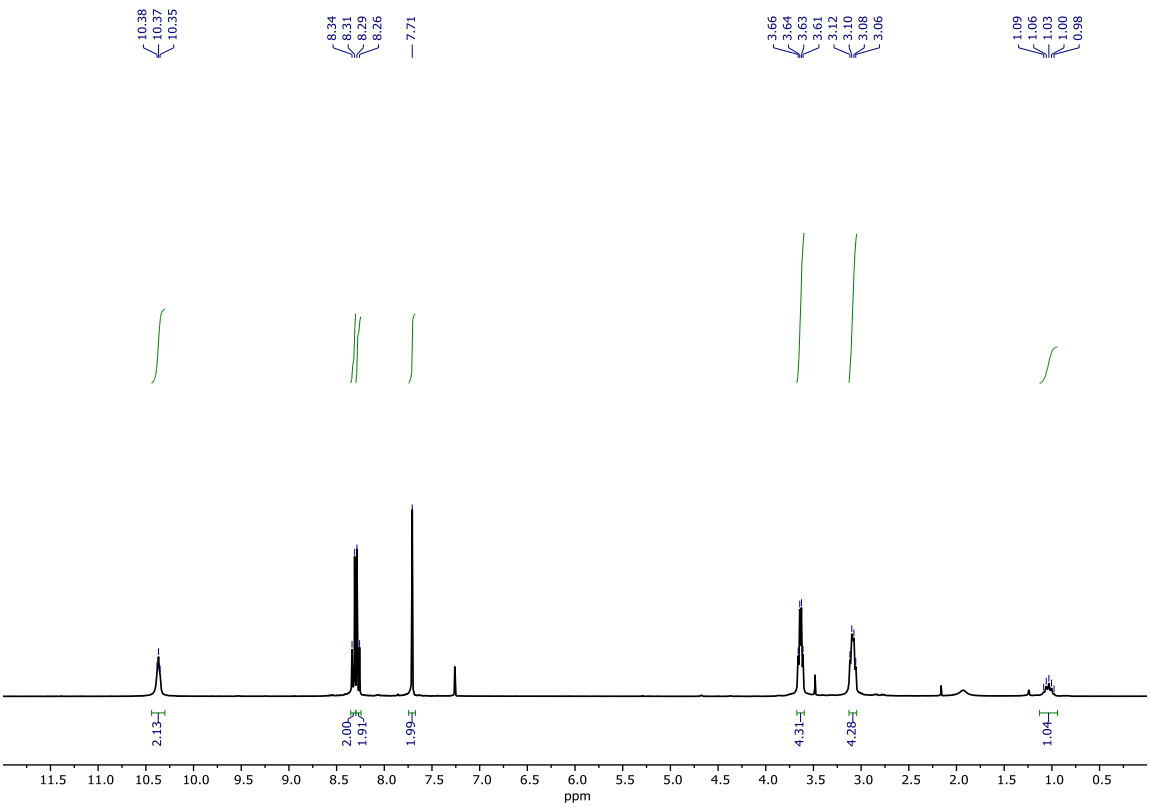


Figure B46. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **4c**.

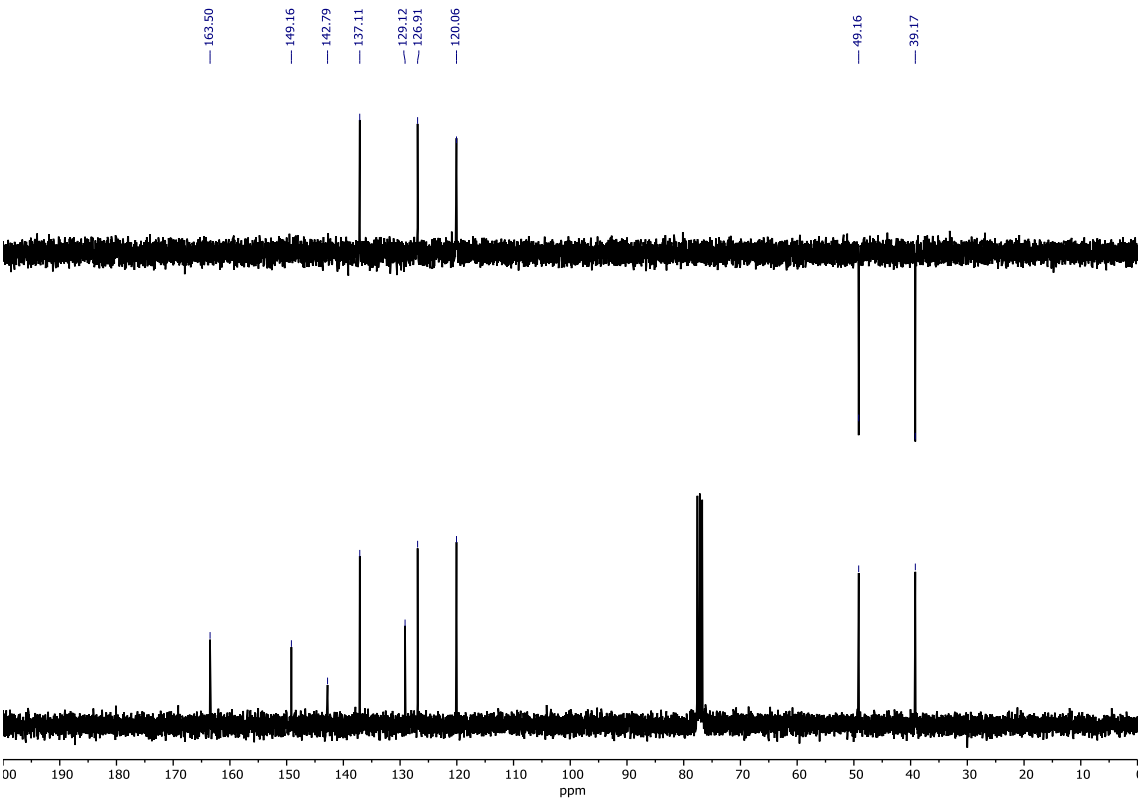
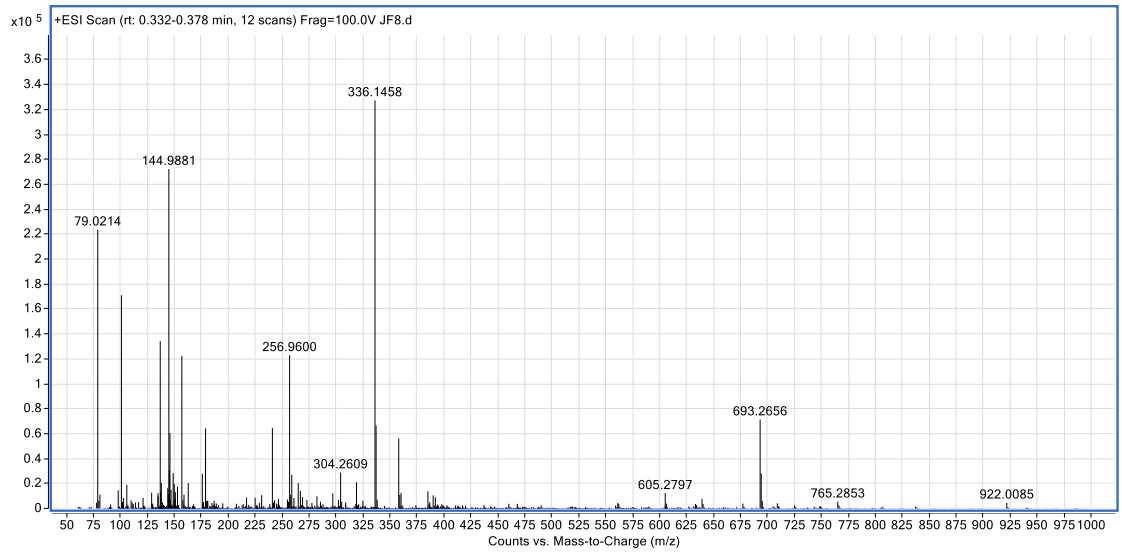
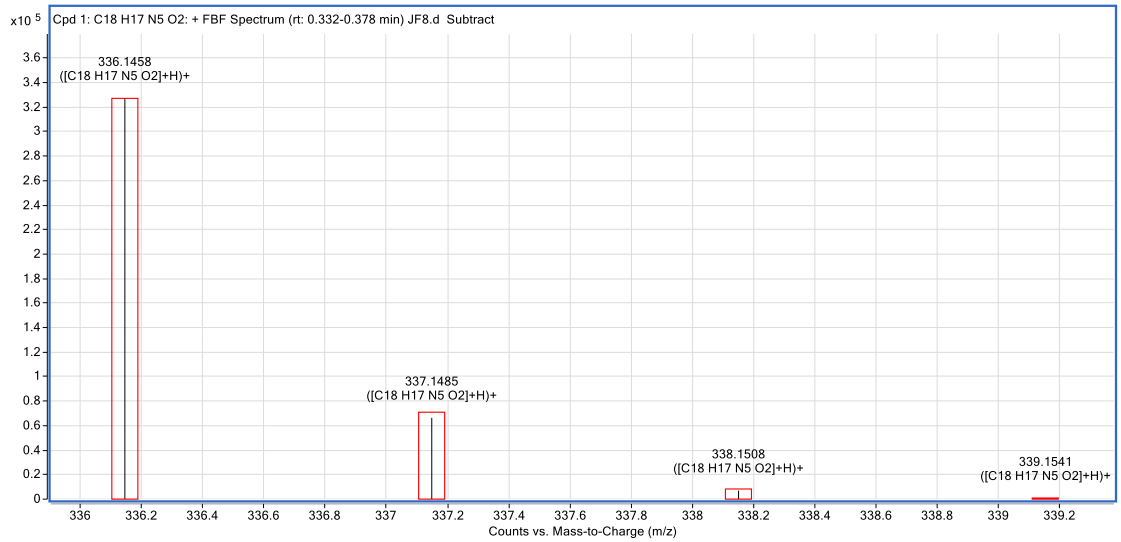


Figure B47. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **4c**.





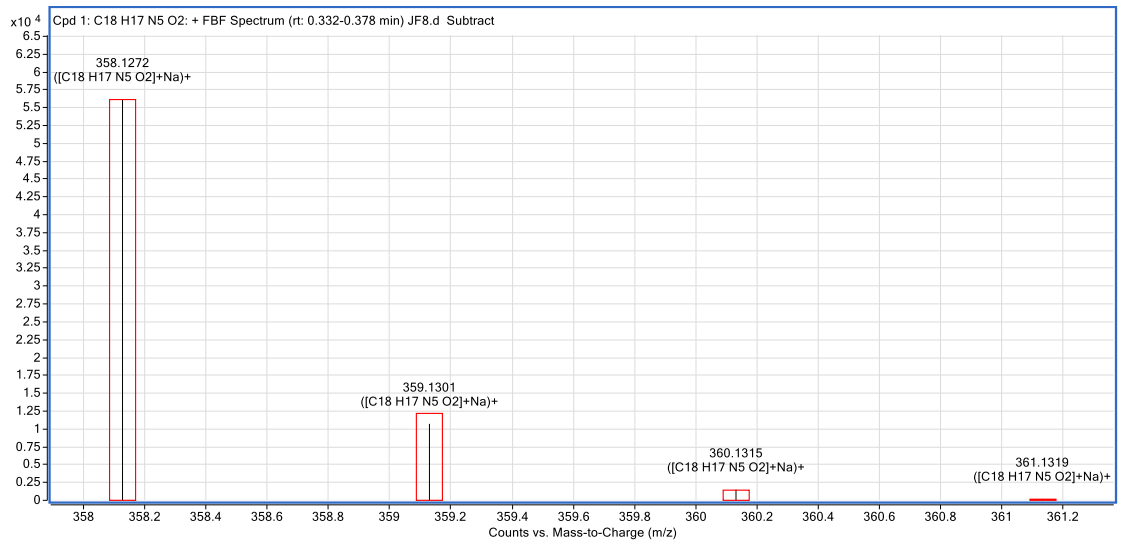


Figure B48. HR-MS (+ESI) spectra for compound **4c**.

Compound **4d**

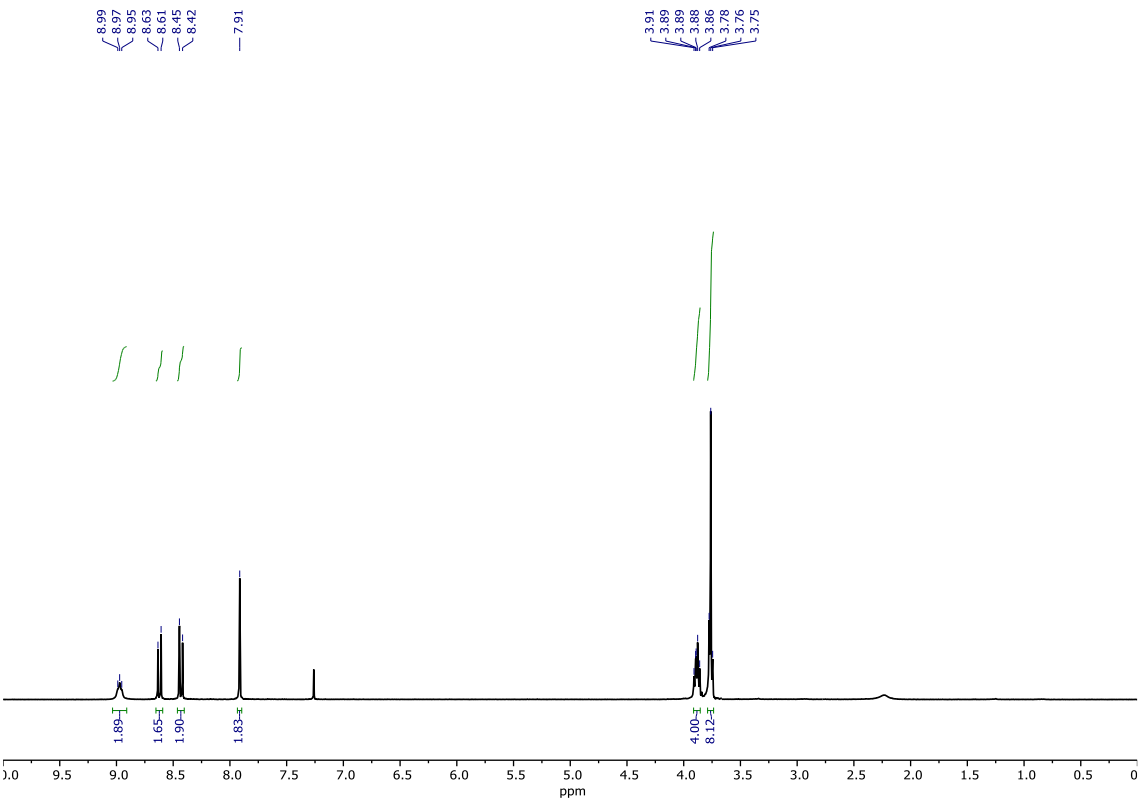


Figure B49. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **4d**.

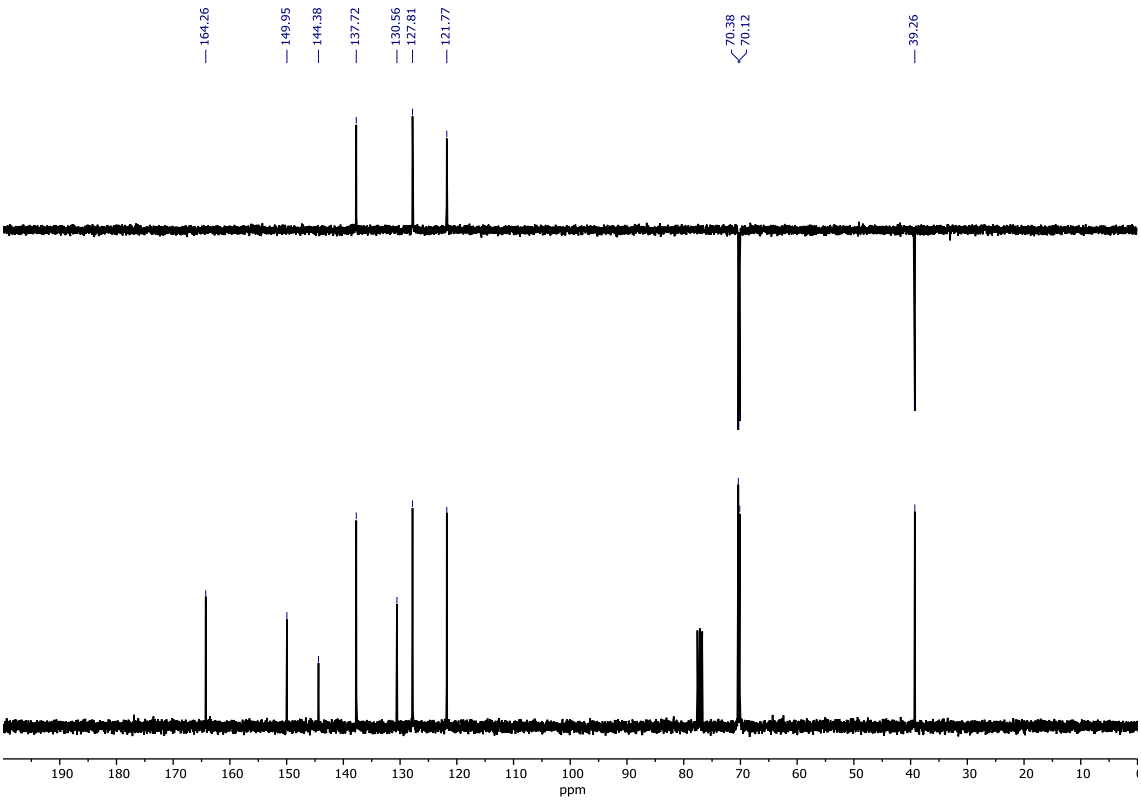
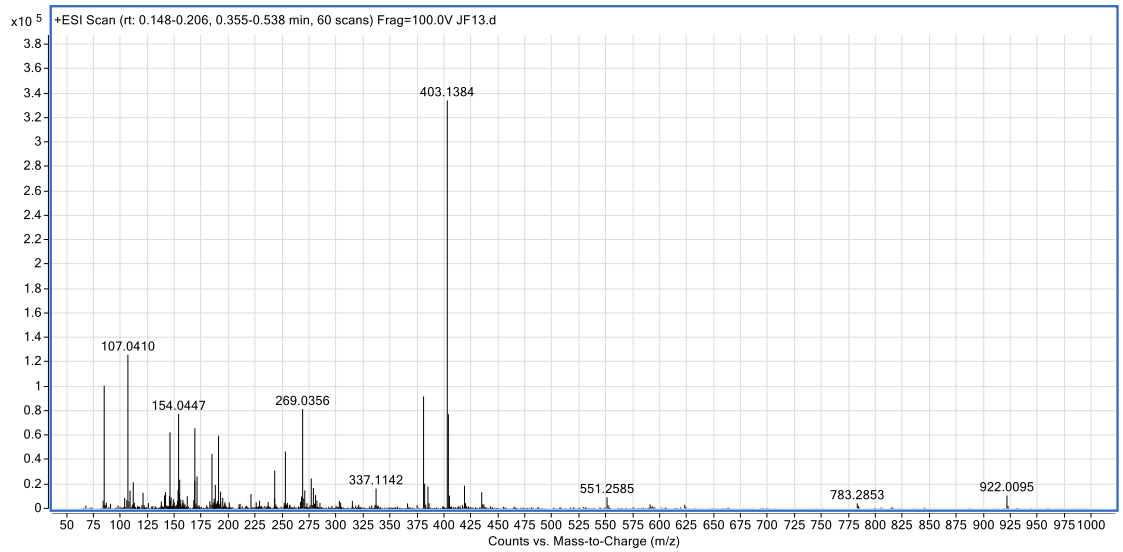
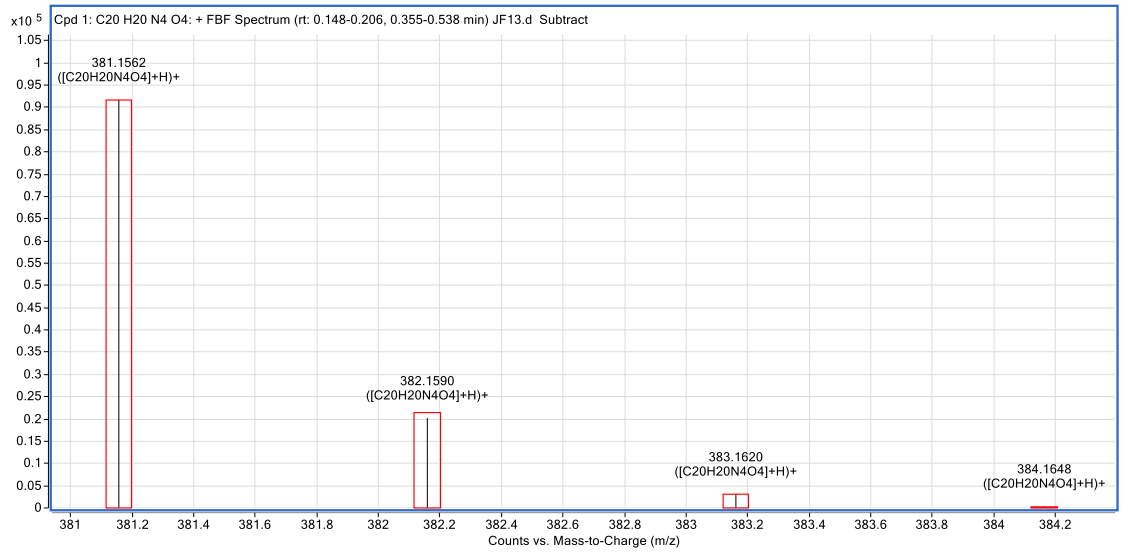


Figure B50. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **4d**.





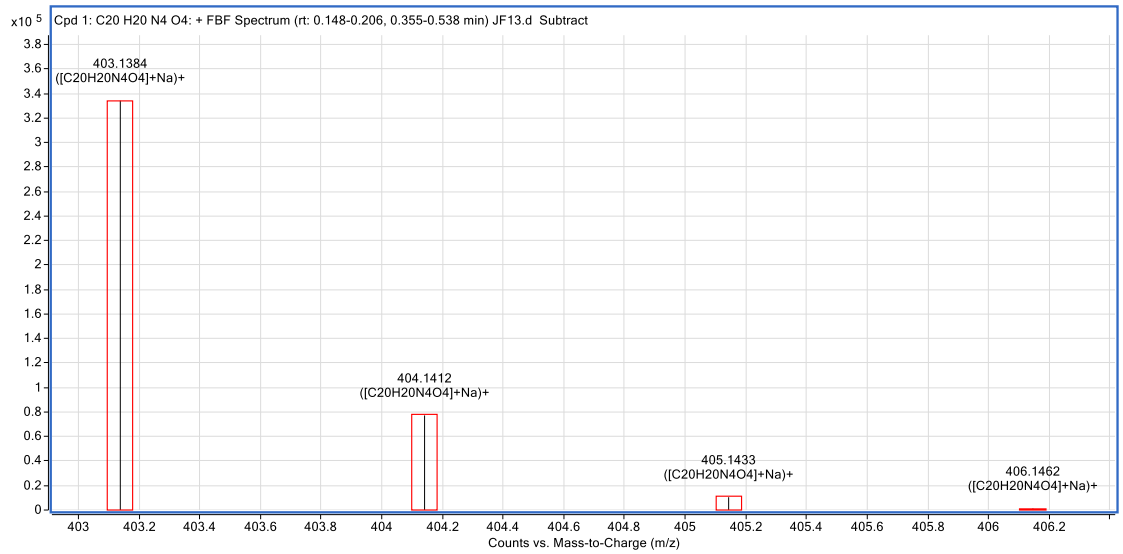


Figure B51. HR-MS (+ESI) spectra for compound **4d**.

In vitro studies

Oligonucleotide sequences

Table B1. List of all sequences tested.

Name	Sequence (5' => 3')	Topology
F21T	FAM-GGGTTAGGGTTAGGGTTAGGG-TAM	Hybrid
F21CTAT	FAM-GGGCTAGGGCTAGGGCTAGGG-TAM	Antiparallel
F25CebT	FAM-AGGGTGGGTGTAAGTGTGGGTGGGT-TAM	Parallel
FdxT	FAM-TATAGCTAT-PEG-TATAGCTATA-TAM	Duplex
22AG	AGGGTTAGGGTTAGGGTTAGGG	Hybrid
26Ceb	AAGGGTGGGTGTAAGTGTGGGTGGGT	Parallel
22CTA	AGGGCTAGGGCTAGGGCTAGGG	Antiparallel

FRET melting experiments

FRET melting curves

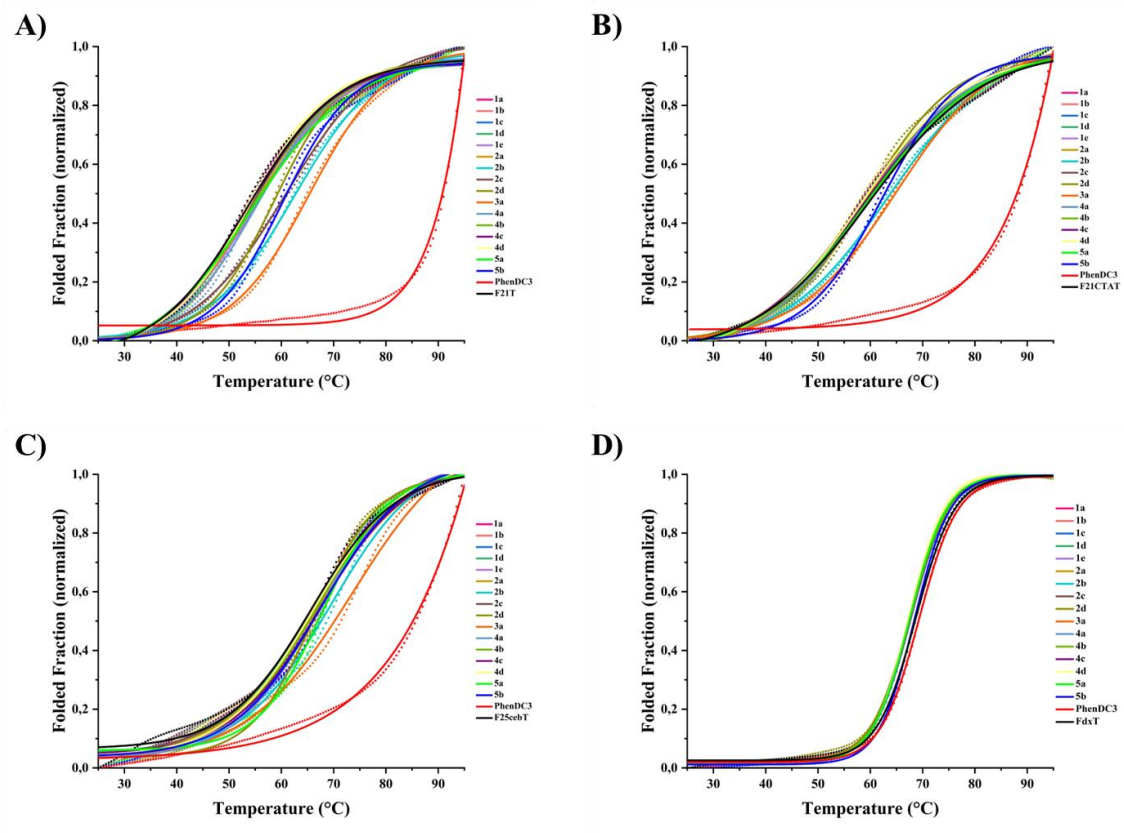


Figure B52. FRET melting curves of **A)** F21T, **B)** F21CTAT, **C)** F25cebT and D) FdxT in the absence and presence of molar equivalents of each ligand.

Circular dichroism experiments

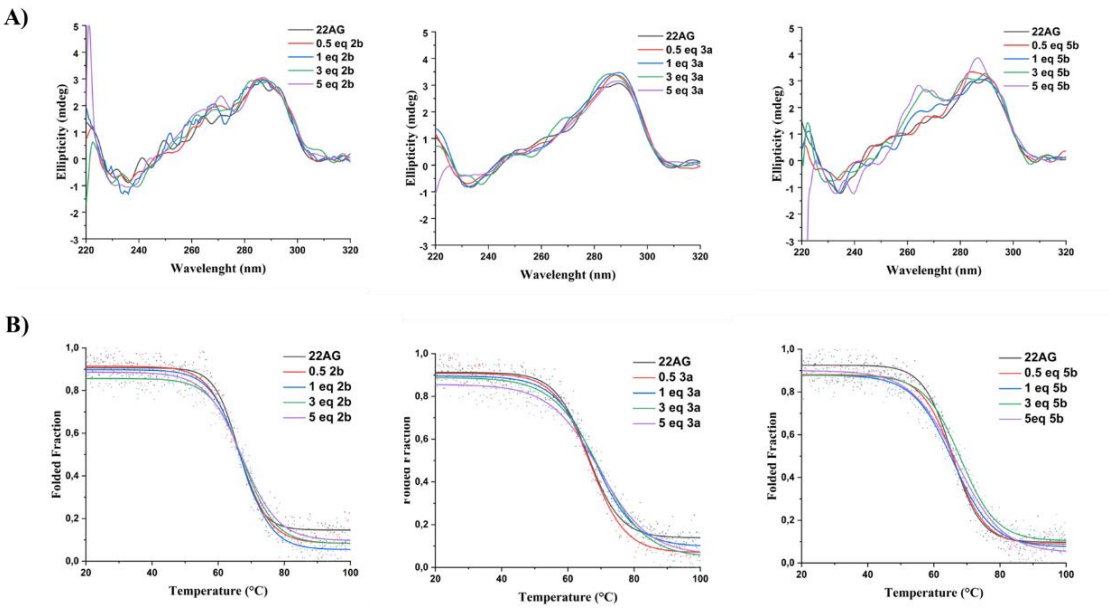


Figure B53. CD titration **(A)** and CD melting curves **(B)** for 22AG sequence in the presence of increasing molar equivalents of ligands **2b**, **3a** and **5b**.

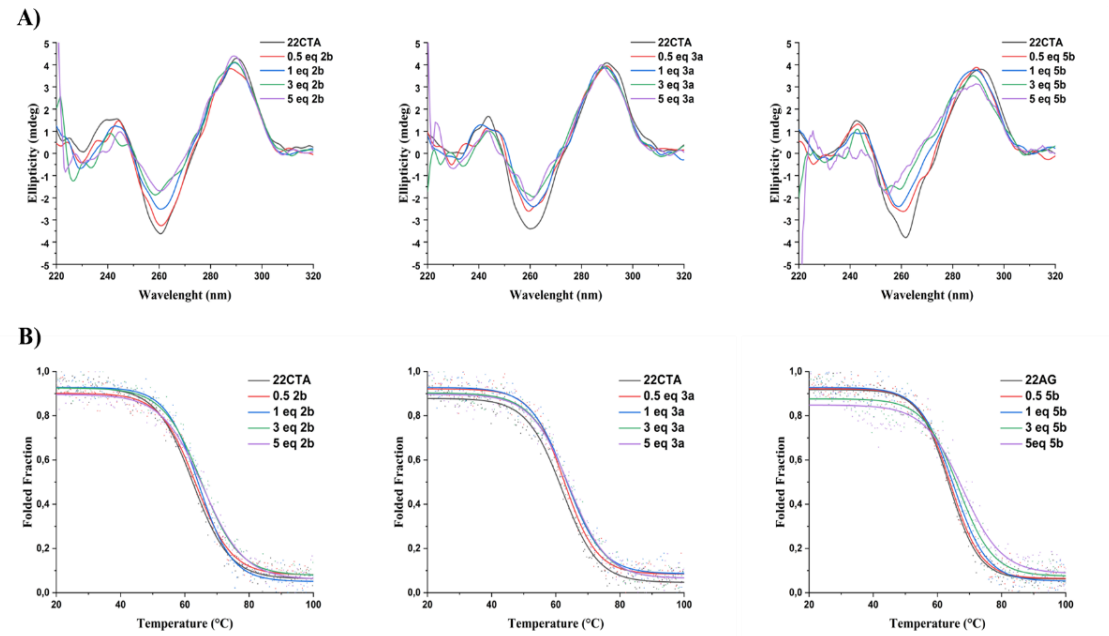


Figure B54. CD titration **(A)** and CD melting curves **(B)** for 22CTA sequence in the presence of increasing molar equivalents of ligands **2b**, **3a** and **5b**.

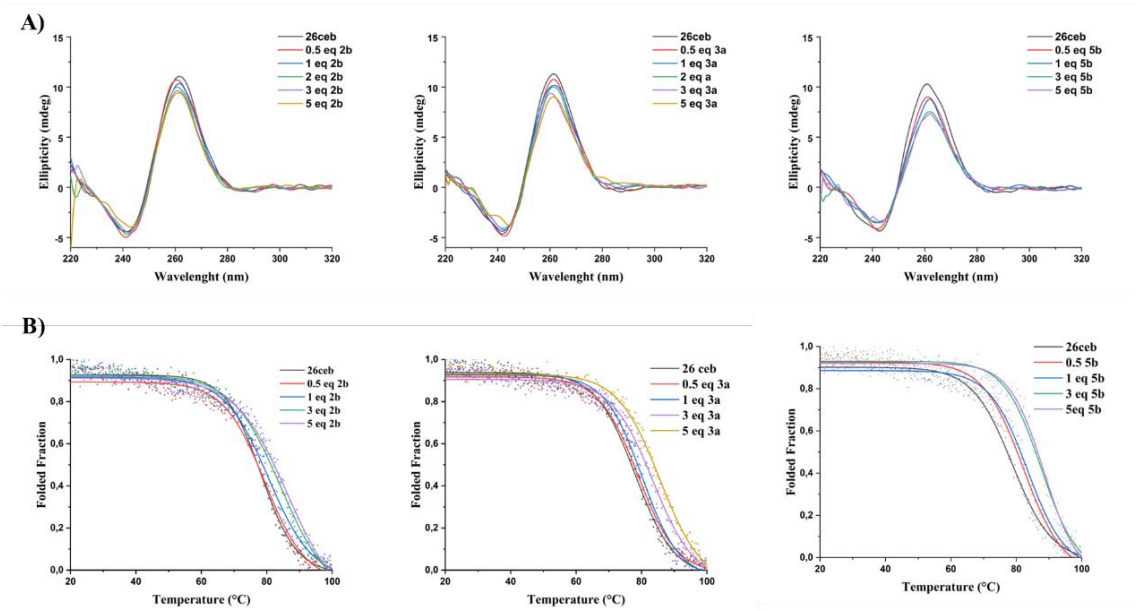


Figure B55. CD titration **(A)** and CD melting curves **(B)** for 26ceb sequence in the presence of increasing molar equivalents of ligands **2b**, **3a** and **5b**.

Cellular studies

Table B2. Percentage of cell viability for the compounds **1a-e**, **2a-d**, **3a**, **4a-d** and **5a-b**, measured by MTT assay.

	Percentage (%) of cell viability at 10 μM ligand concentration		
Ligand	**A549**	**H1299**	**NHDF**
1a	46.7 ± 9.3	53.1 ± 13.0	39.5 ± 14.4
1b	47.7 ± 6.1	48.3 ± 11.1	38.7 ± 6,9
1c	44.3 ± 7.0	39.4 ± 4.2	32.1 ± 10.9
1d	50.6 ± 3.7	57.2 ± 13.2	32.2 ± 11.2
1e	53.4 ± 12.5	41.7 ± 8.2	44.9 ± 14.0
2a	65.1 ± 2.1	34.4 ± 9.1	43.3 ± 13.3
2b	62.1 ± 9.8	47.6 ± 8.1	40.3 ± 7.3
2c	42.6 ± 3.9	43.4 ± 14.3	39.7 ± 9.1
2d	27.4 ± 5.6	42.9 ± 12.3	22.4 ± 12.2
3a	58.9 ± 7.6	59.1 ± 2.5	66.6 ± 8.8
4a	35.1 ± 7.9	60.5 ± 8.2	72.6 ± 19.7
4b	41.6 ± 6.6	68.2 ± 6.6	72.2 ± 4.0
4c	44.2 ± 7.6	73.9 ± 11.7	62.5 ± 7.8
4d	46.4 ± 4.4	84.8 ± 12.7	80.8 ± 31.8
5a	79 ± 13.8	45.3 ± 5.7	64.8 ± 3.4
5b	24.5 ± 2.2	70.7 ± 27.5	83.3 ± 10.7
PhenDC3	38.0 ± 1.9	72.8 ± 18.7	78.1 ± 10.6
Gemcitabine	19.4 ± 5.5	49.4 ± 10.9	38.7 ± 7.6

Results are presented as mean ± SD. Error margins correspond to SD of three replicates.

Appendix C

Supporting information for 7th chapter

^{1}H, ^{13}C, DEPT-135, NOESY and HR-MS spectra

Compound **1**

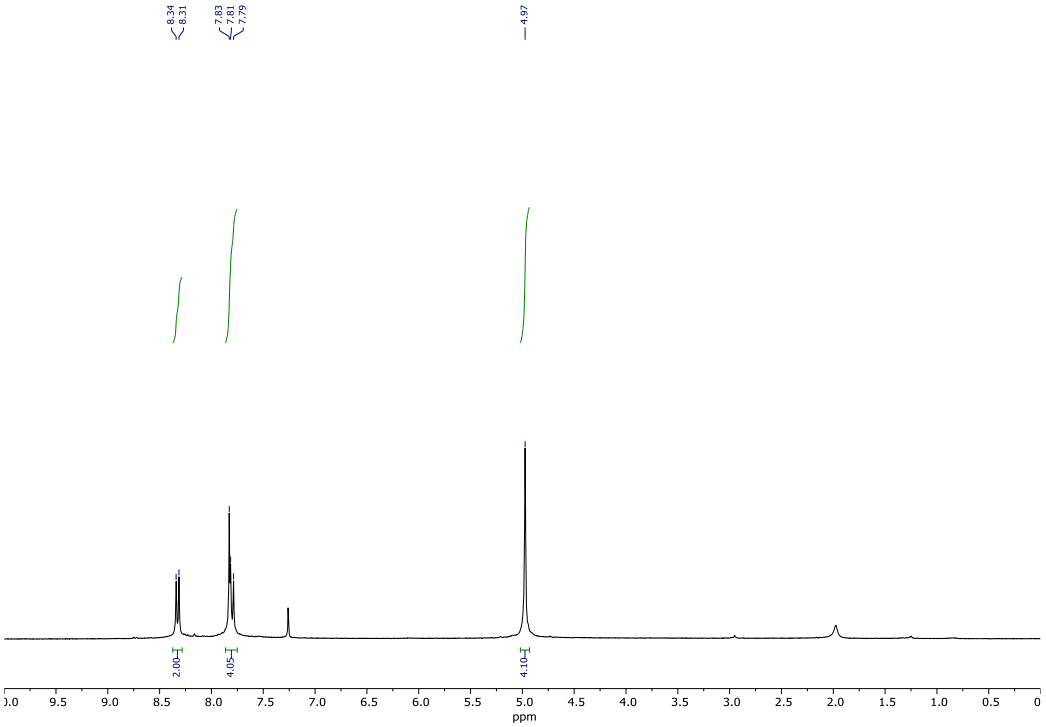


Figure C1. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **1**.

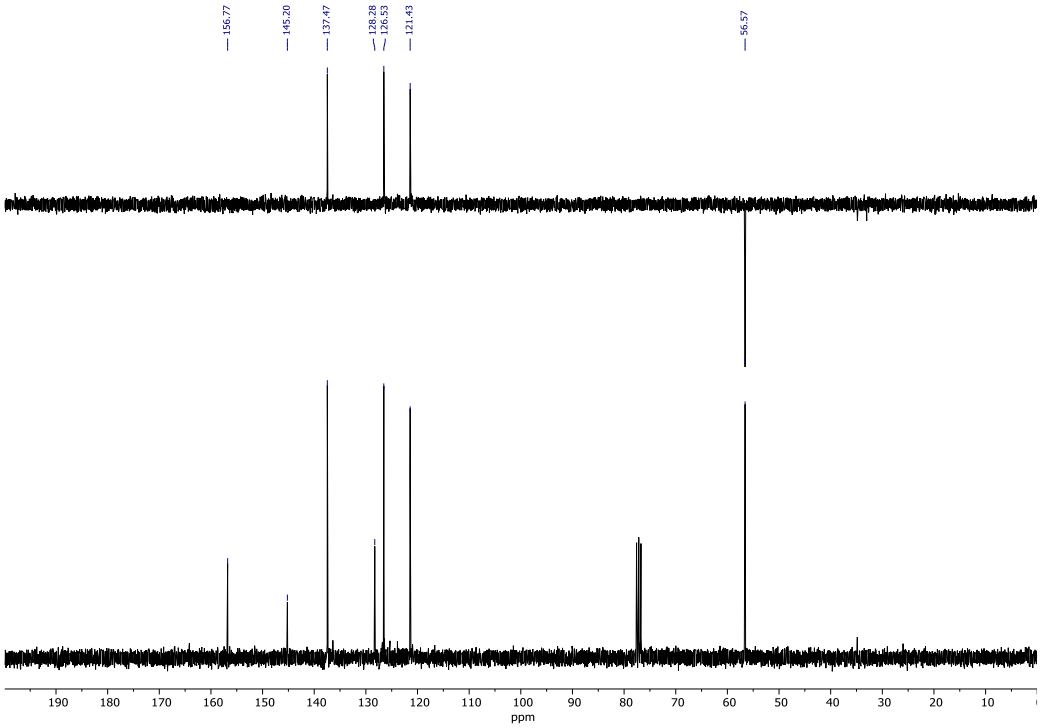


Figure C2. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **1**.

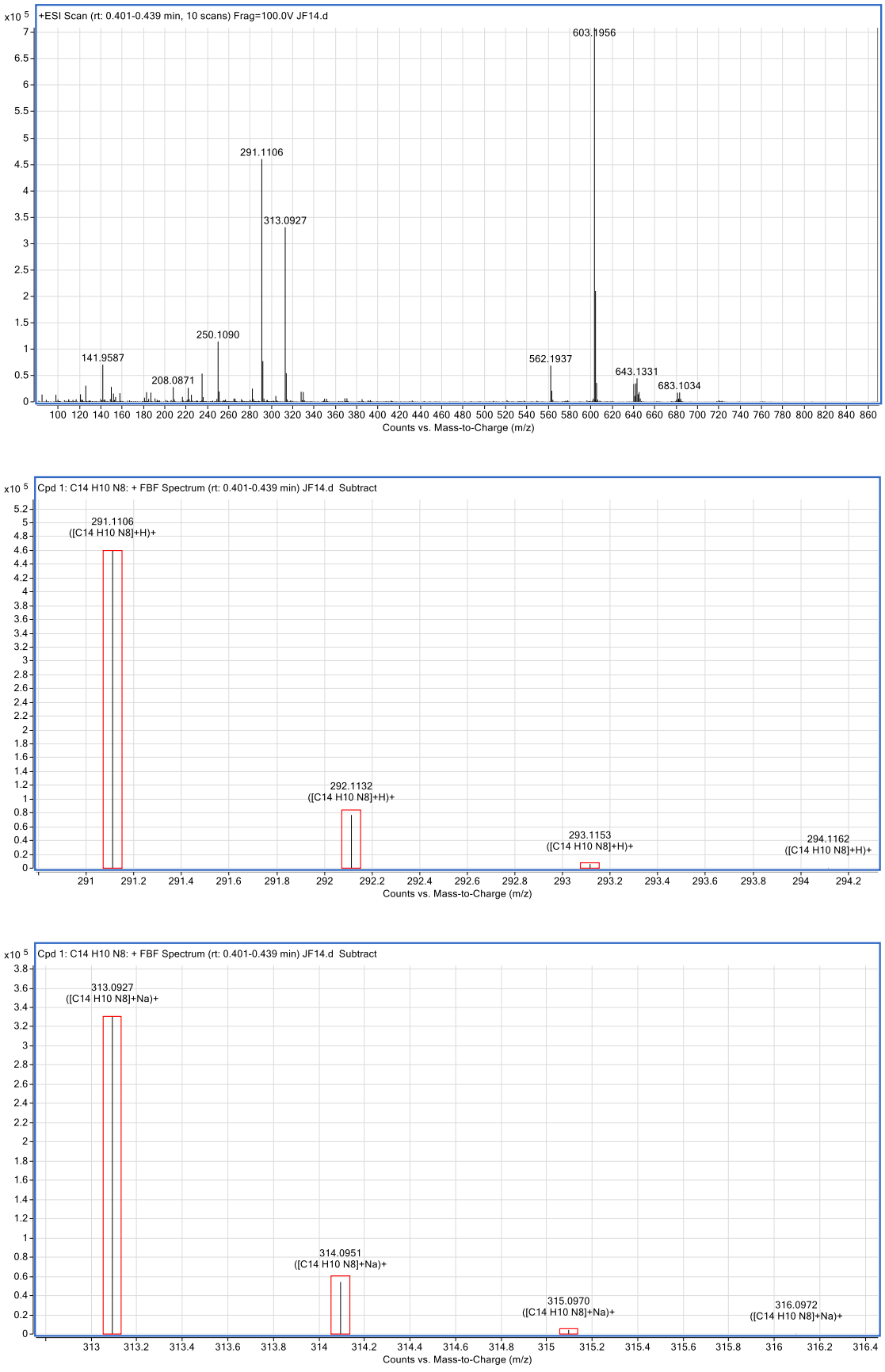


Figure C3. HR-MS (+ESI) spectra for compound **1**.

Compound **9**

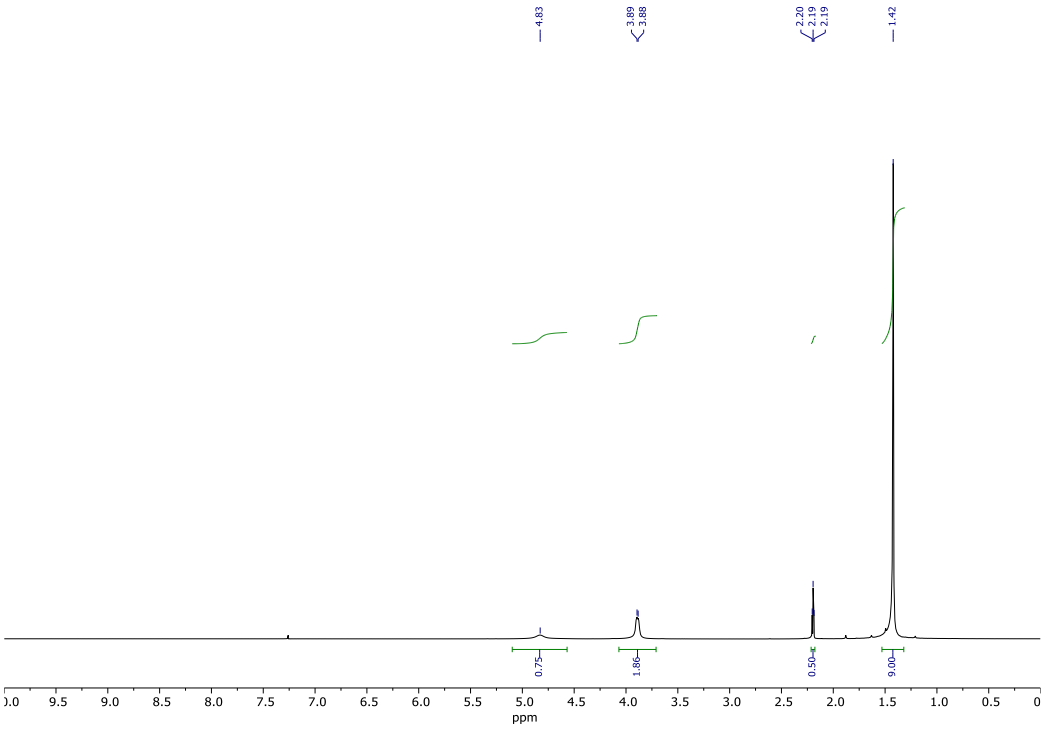


Figure C4. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **9**.

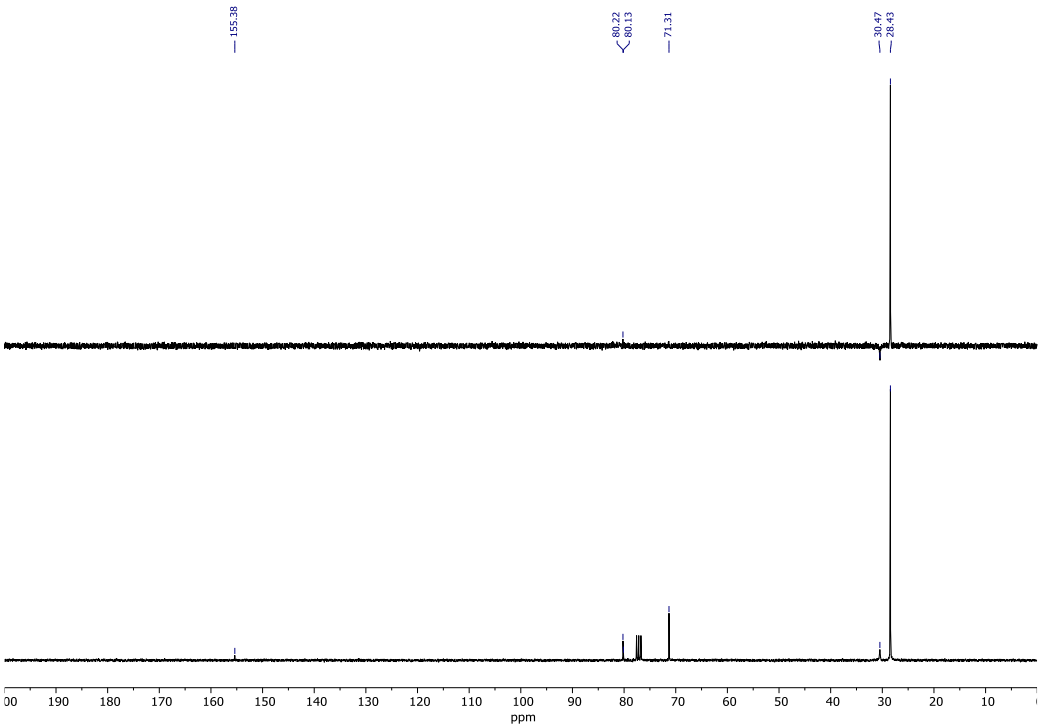


Figure C5. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **9**.

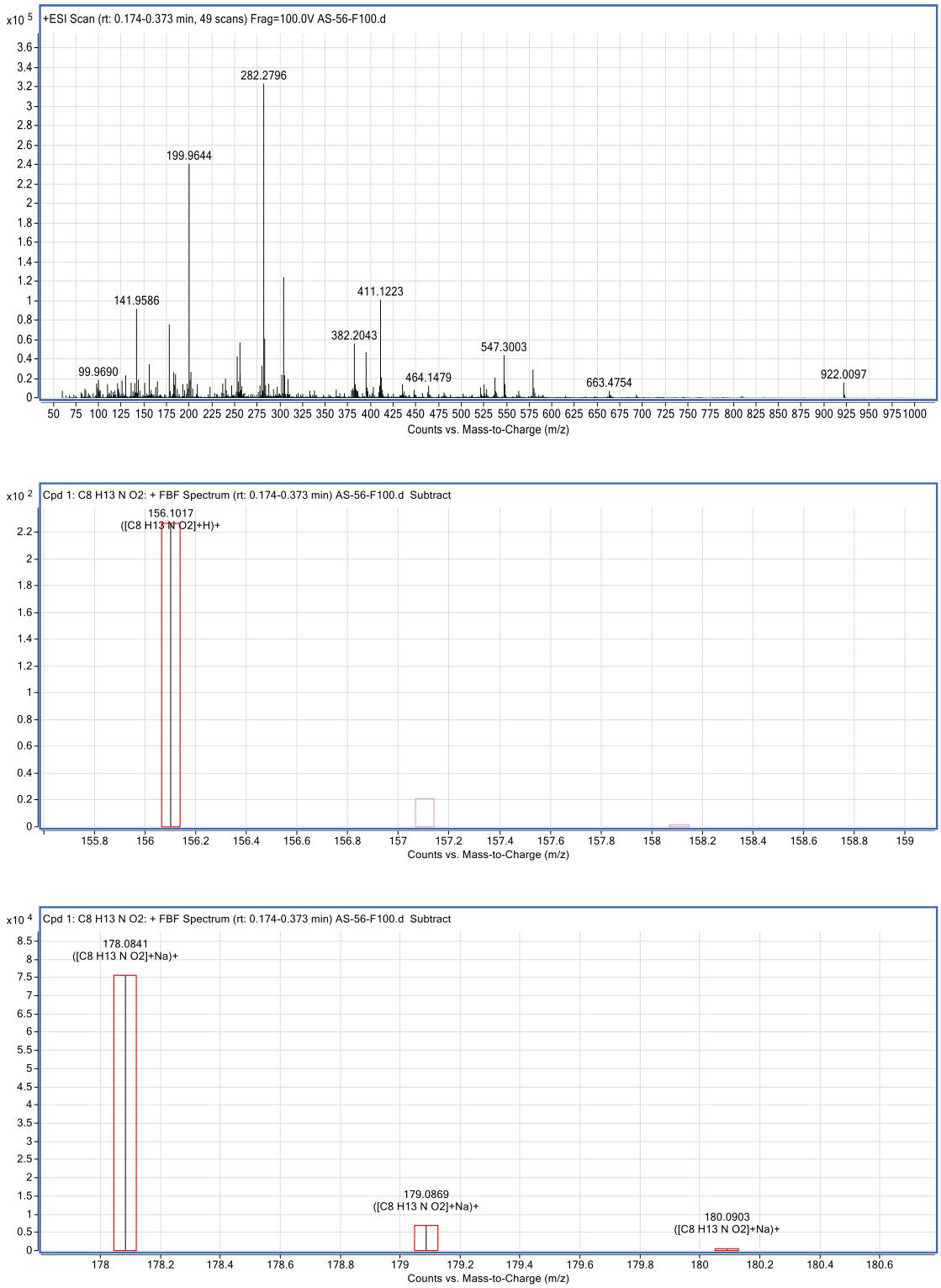


Figure C6. HR-MS (+ESI) spectra for compound **9**.

Compound **2a**

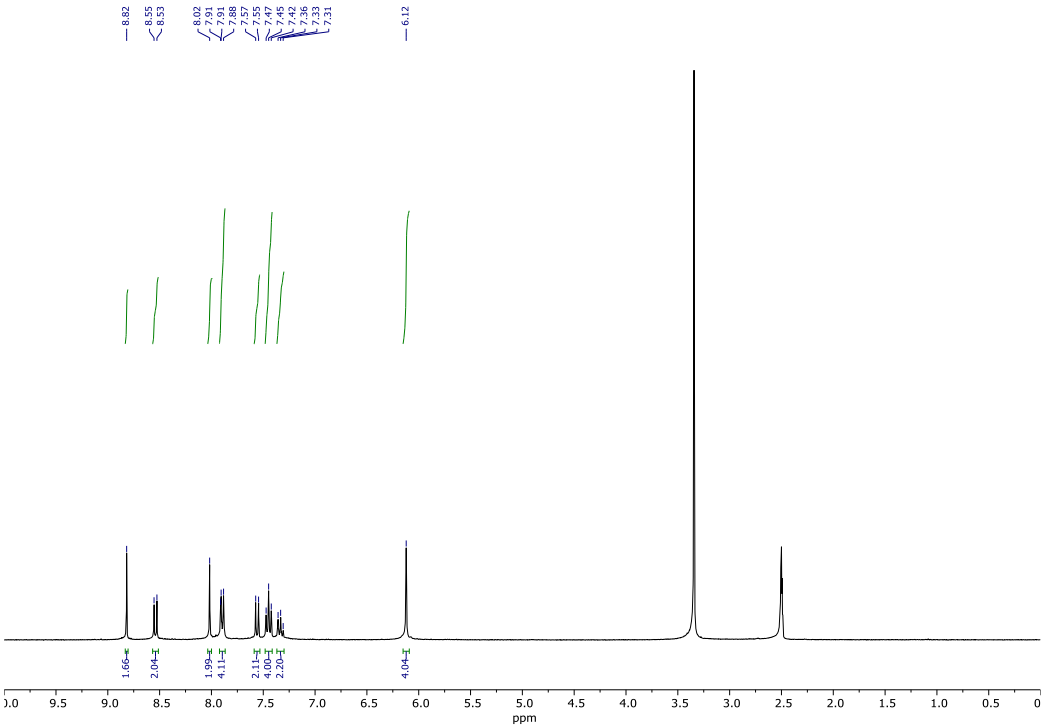


Figure C7. [1]H NMR spectrum (300 MHz, DMSO-d_6) for compound **2a**.

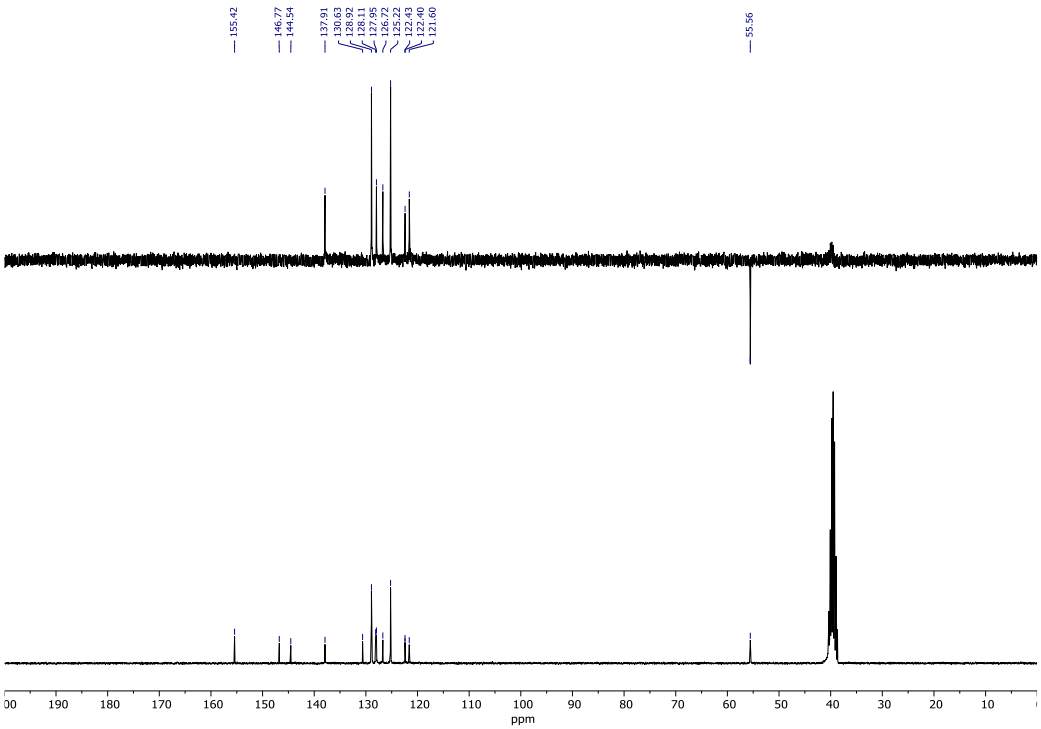


Figure C8. [13]C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **2a**.

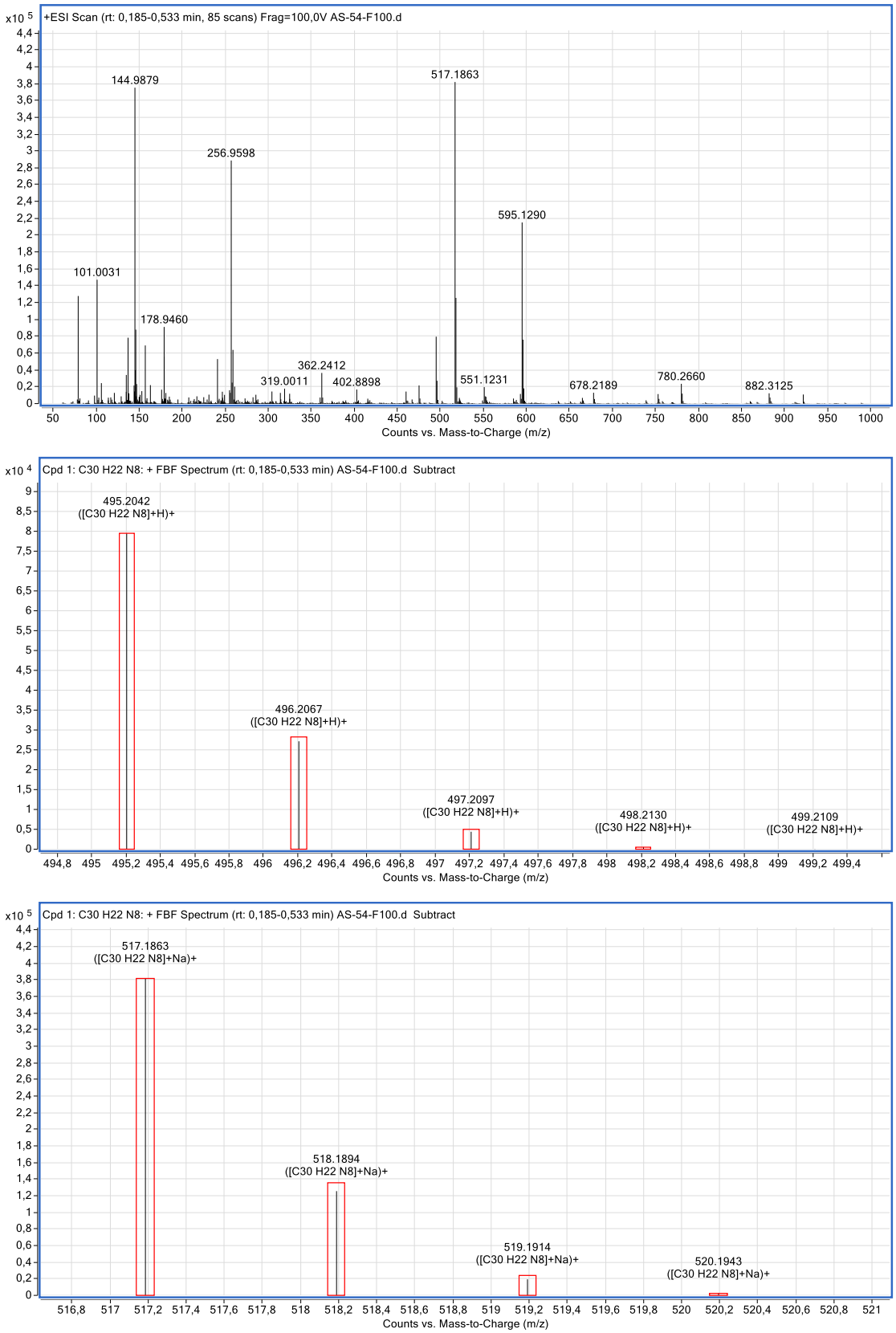


Figure C9. HR-MS (+ESI) spectra for compound **2a**.

Compound **3a**

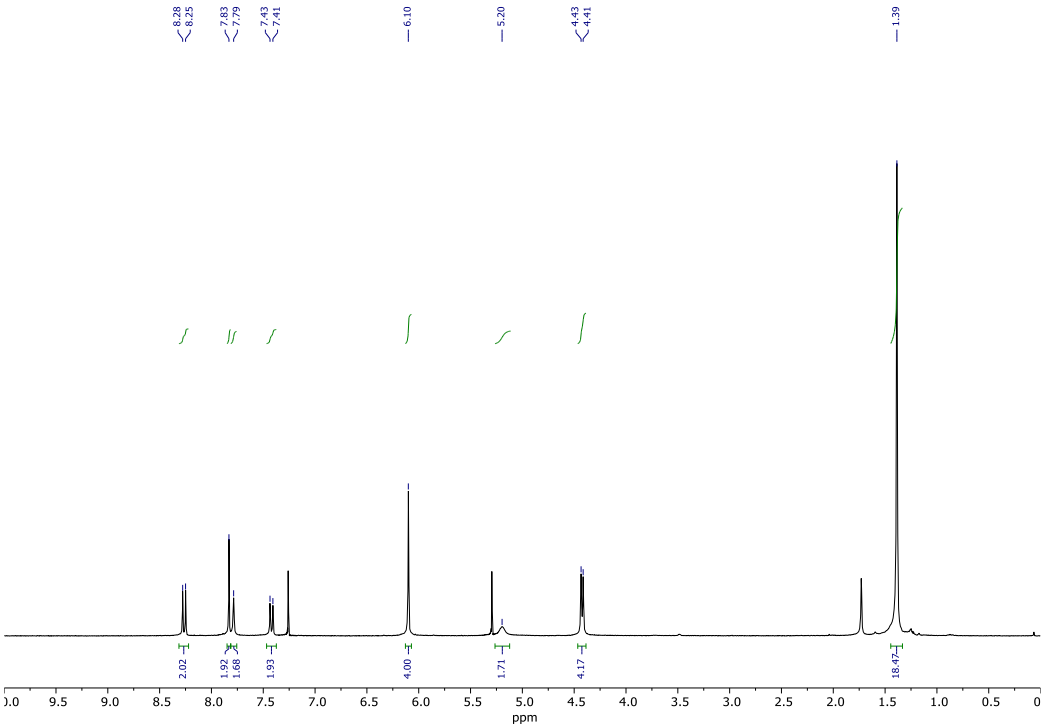


Figure C10. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **3a**.

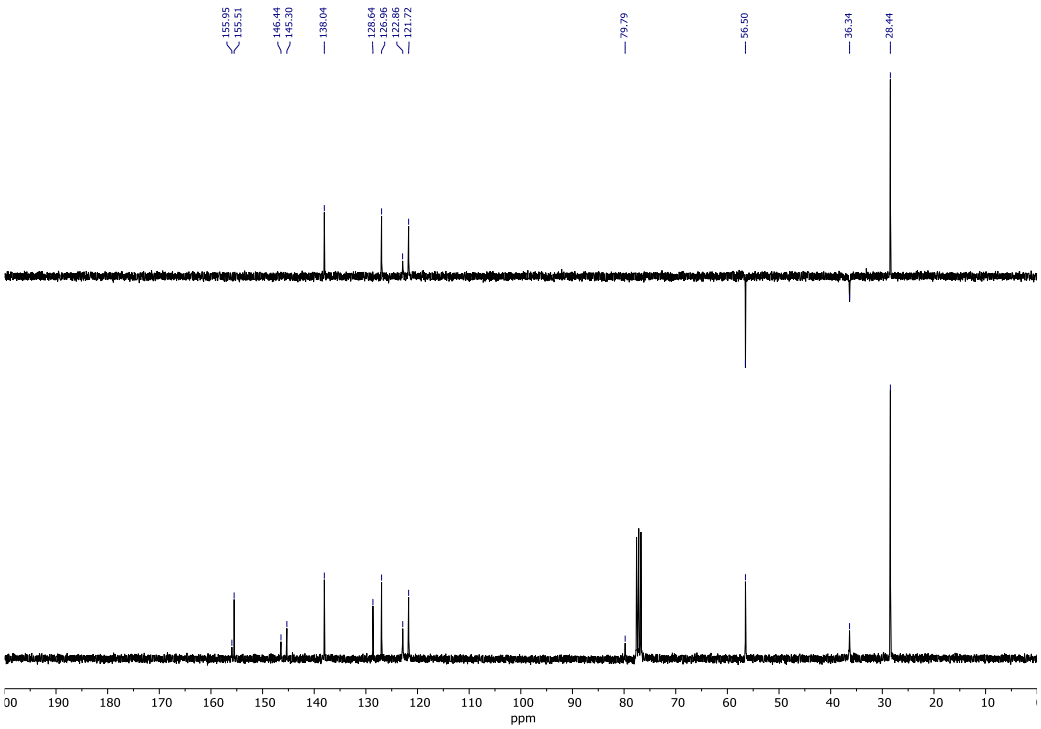


Figure C11. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **3a**.

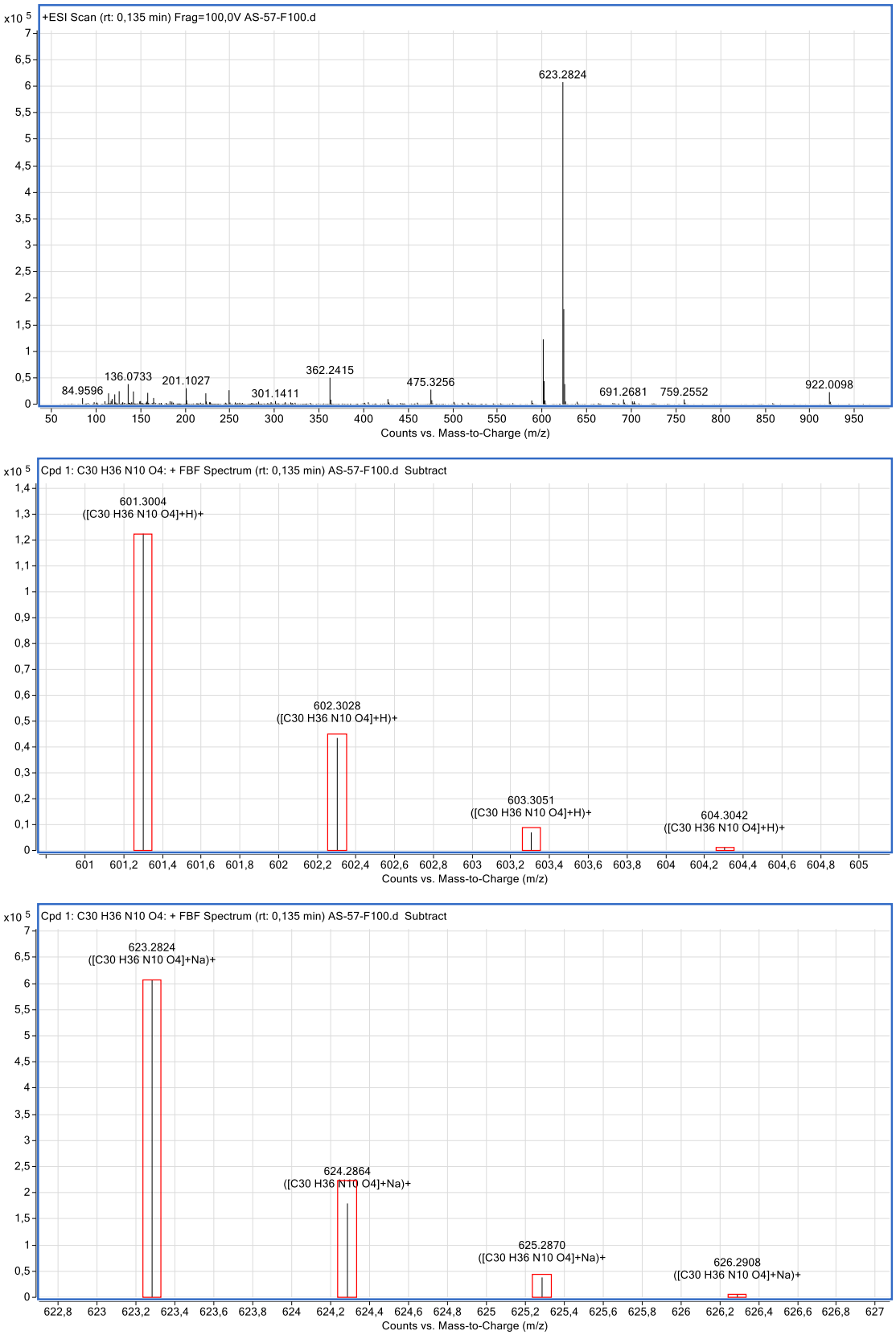


Figure C12. HR-MS (+ESI) spectra for compound **3a**.

Compound **3b**

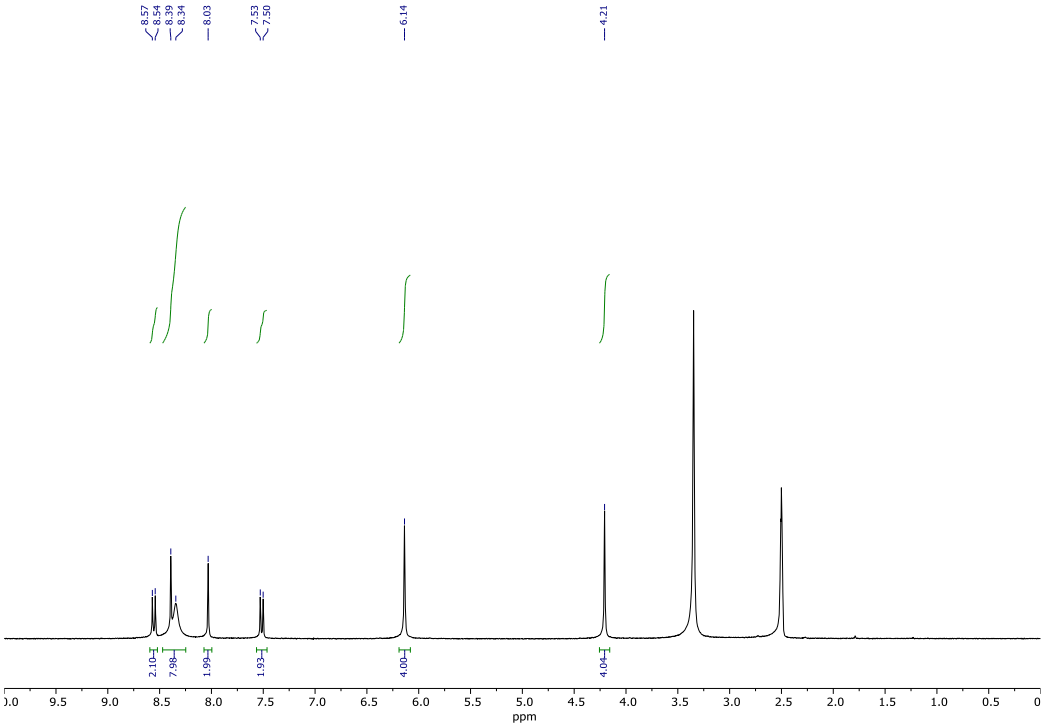


Figure C13. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **3b**.

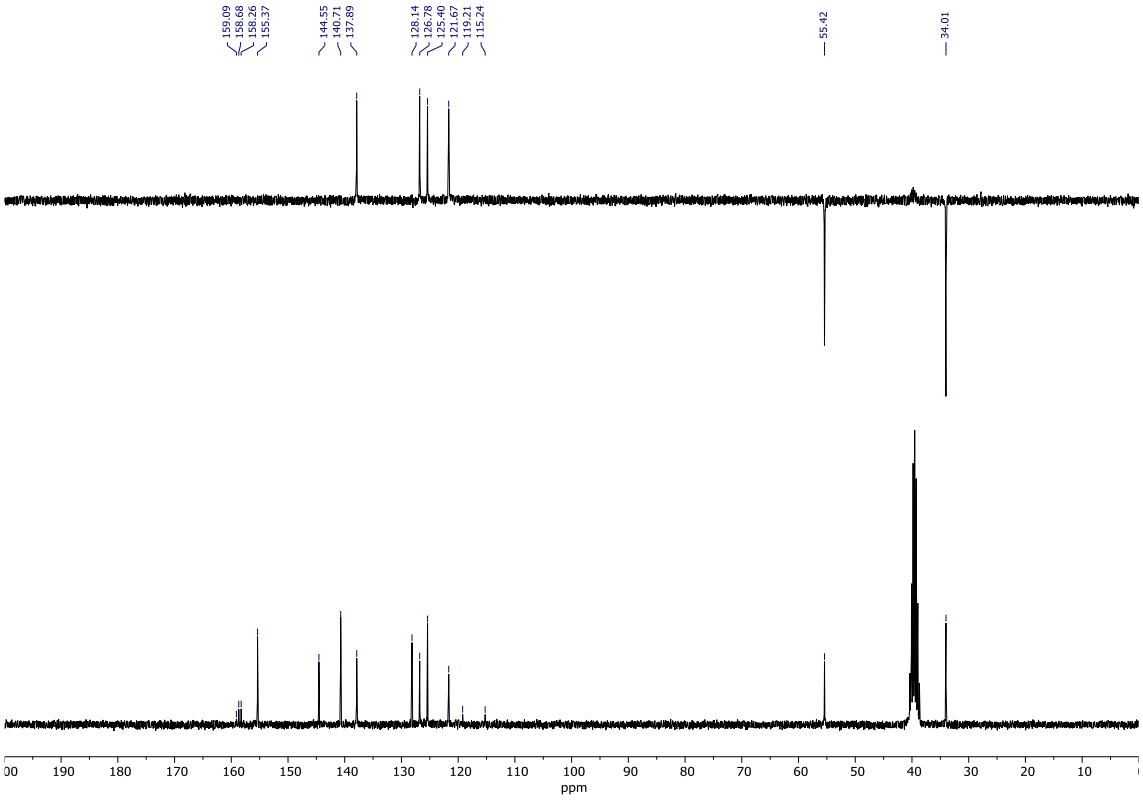


Figure C14. ^{13}C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **3b**.

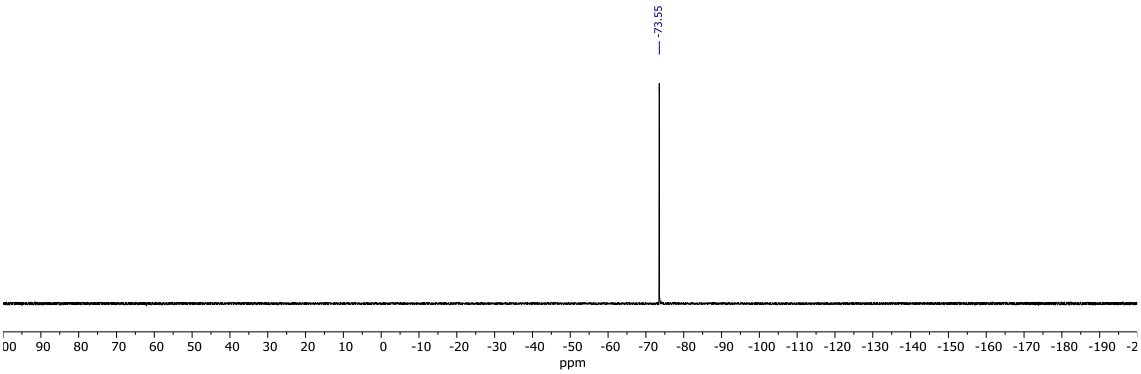


Figure C15. ^{19}F NMR spectrum (282 MHz, DMSO-d_6) for compound **3b**.

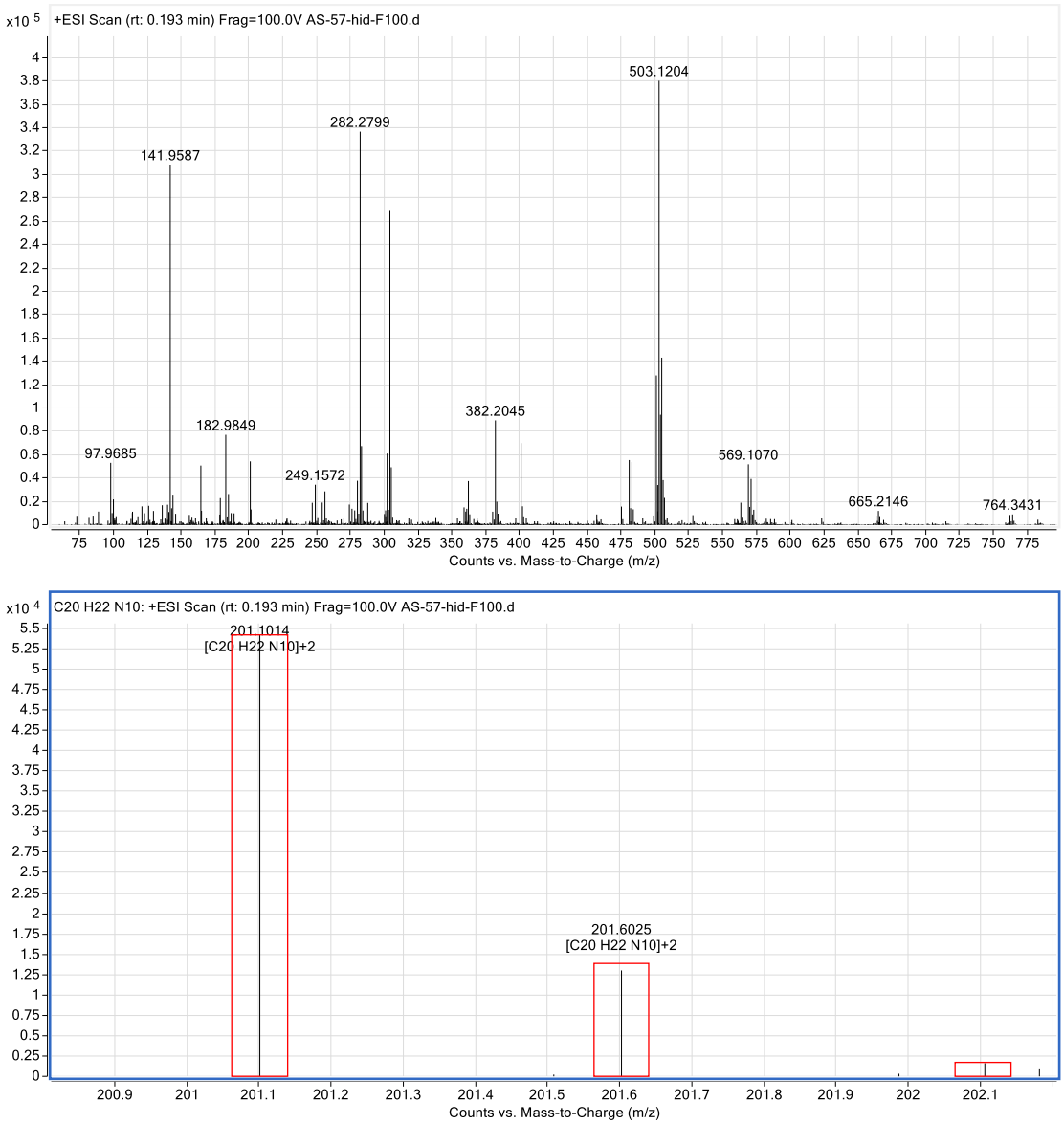


Figure C16. HR-MS (+ESI) spectra for compound **3b**.

Compound **4a**

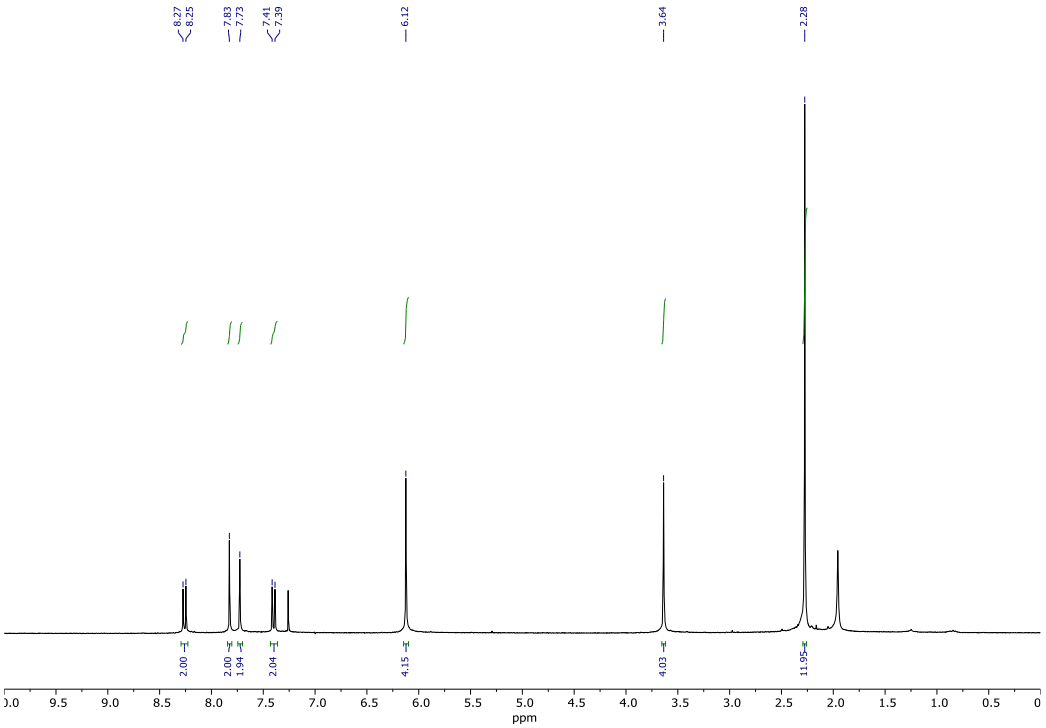


Figure C17. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **4a**.

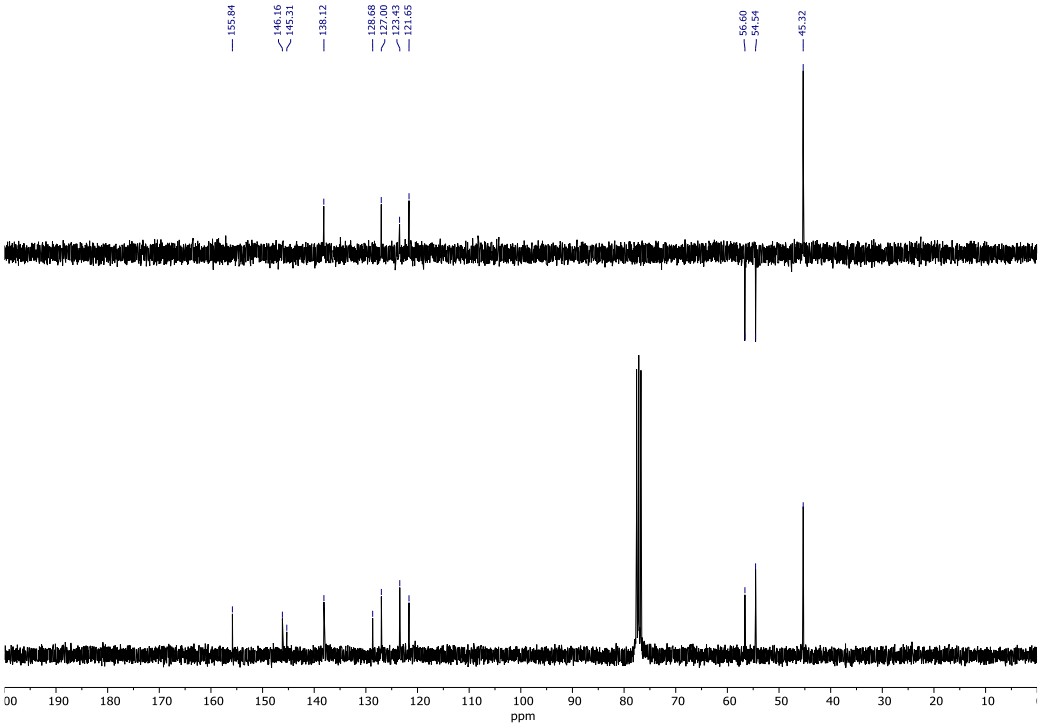


Figure C18. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **4a**.

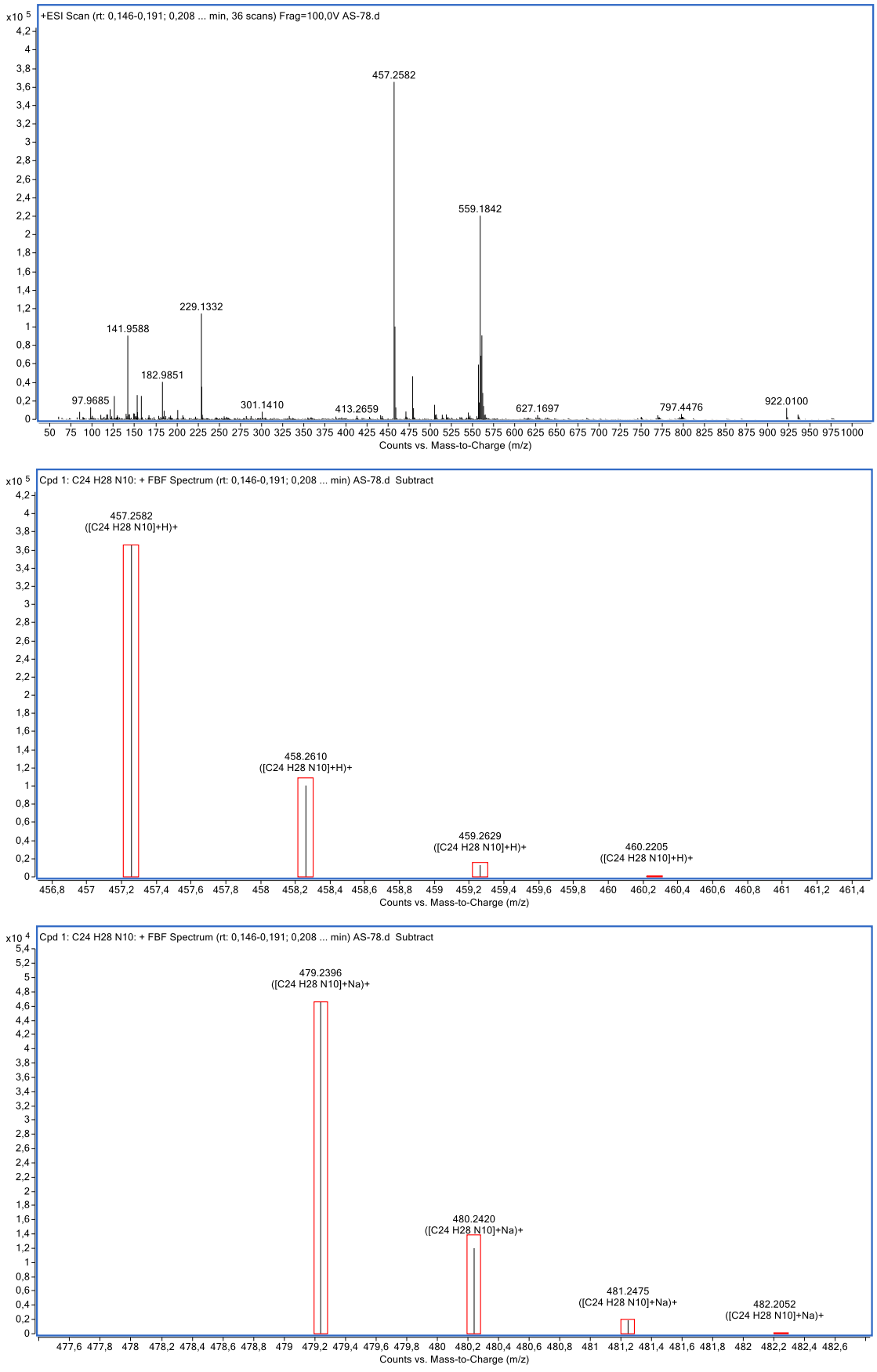


Figure C19. HR-MS (+ESI) spectra for compound **4a**.

Compound **4b**

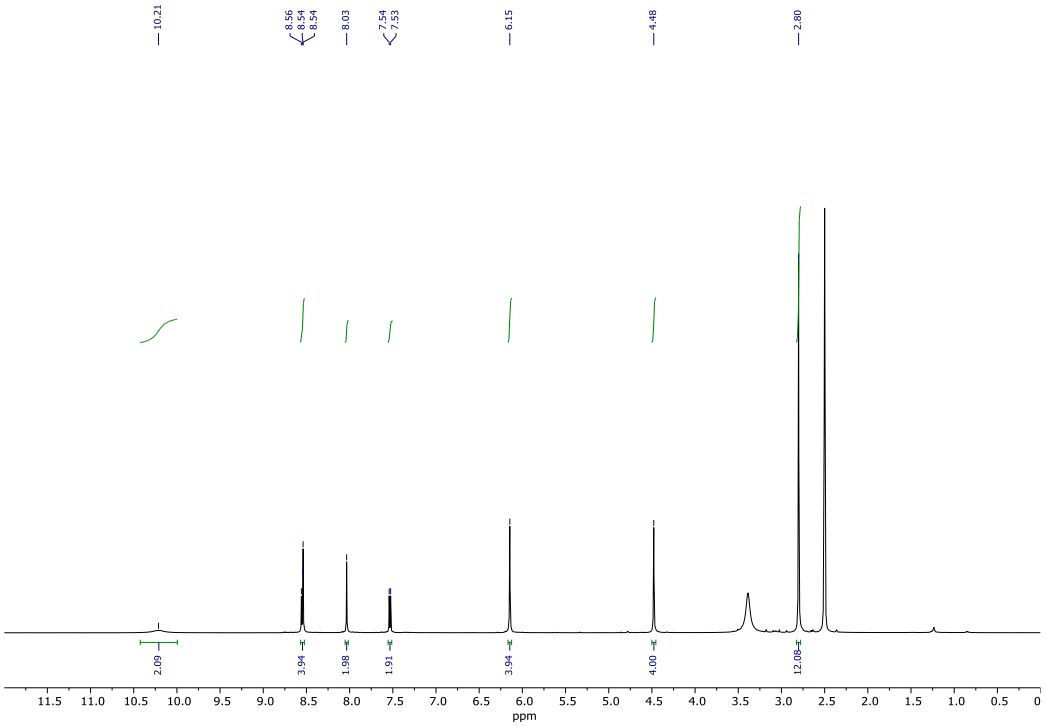


Figure C20. [1]H NMR spectrum (500 MHz, DMSO-*d*6) for compound **4b**.

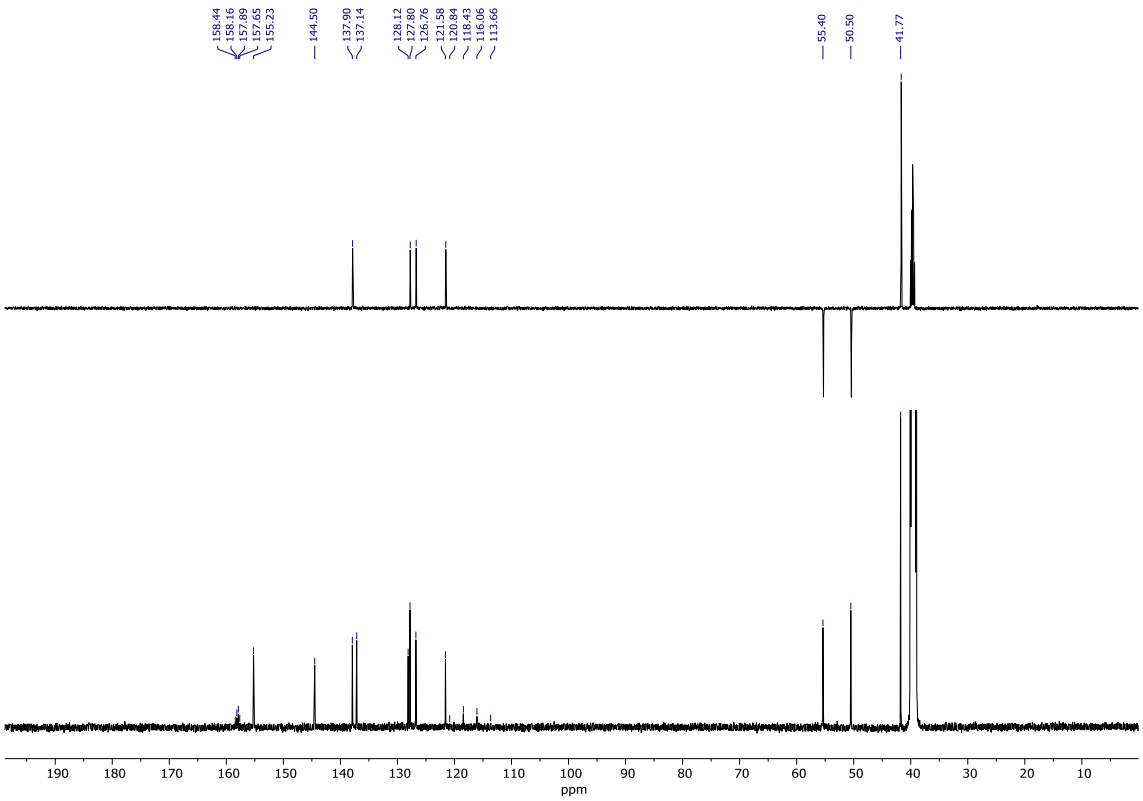


Figure C21. [13]C and DEPT-135 NMR spectra (126 MHz, DMSO-*d*6) for compound **4b**.

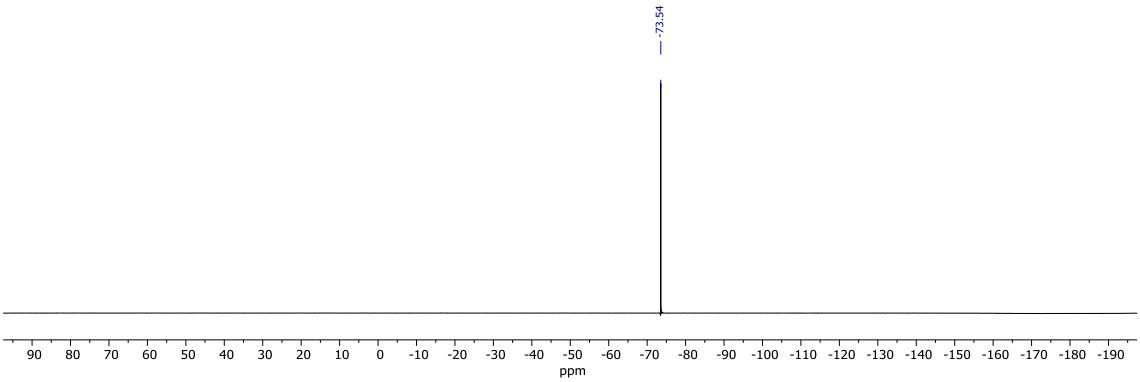


Figure C22. ^{19}F NMR spectrum (471 MHz, DMSO-d_6) for compound **4b**.

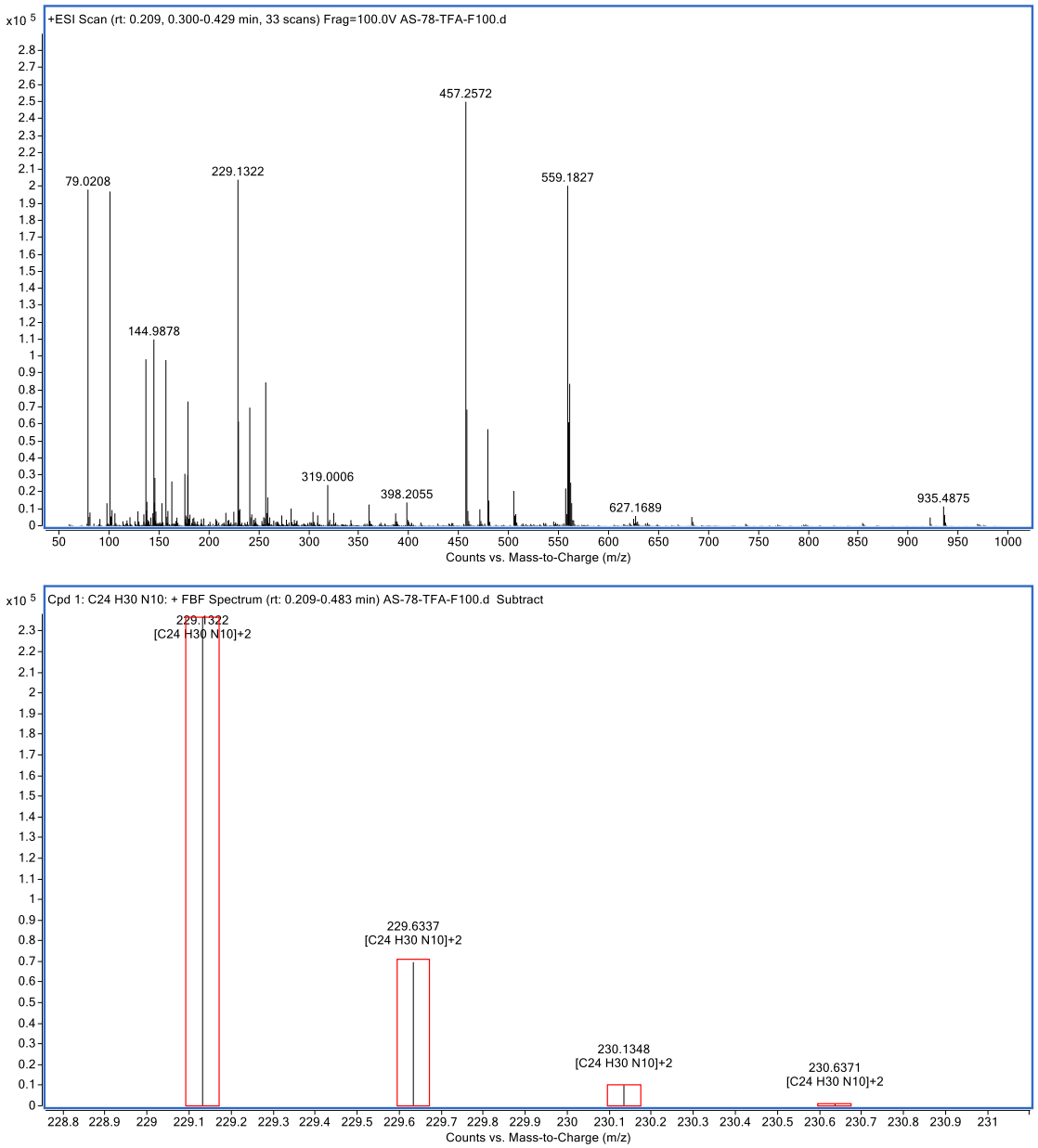


Figure C23. HR-MS (+ESI) spectra for compound **4b**.

Compound **5a**

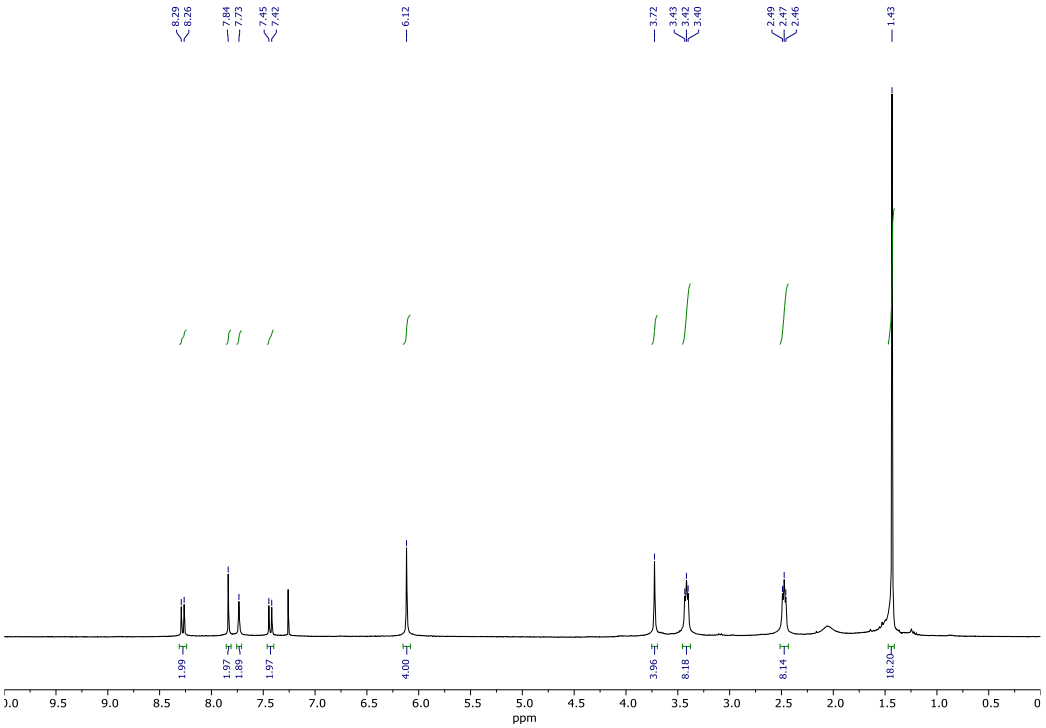


Figure C24. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **5a**.

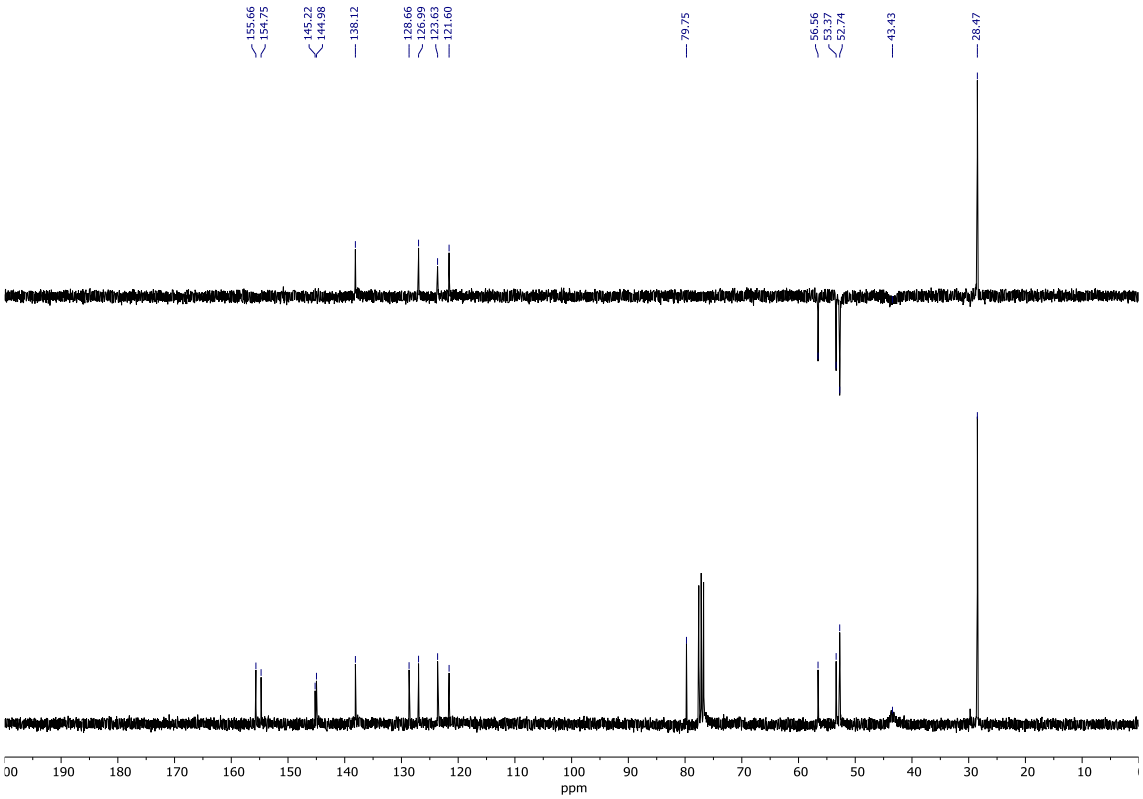


Figure C25. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **5a**.

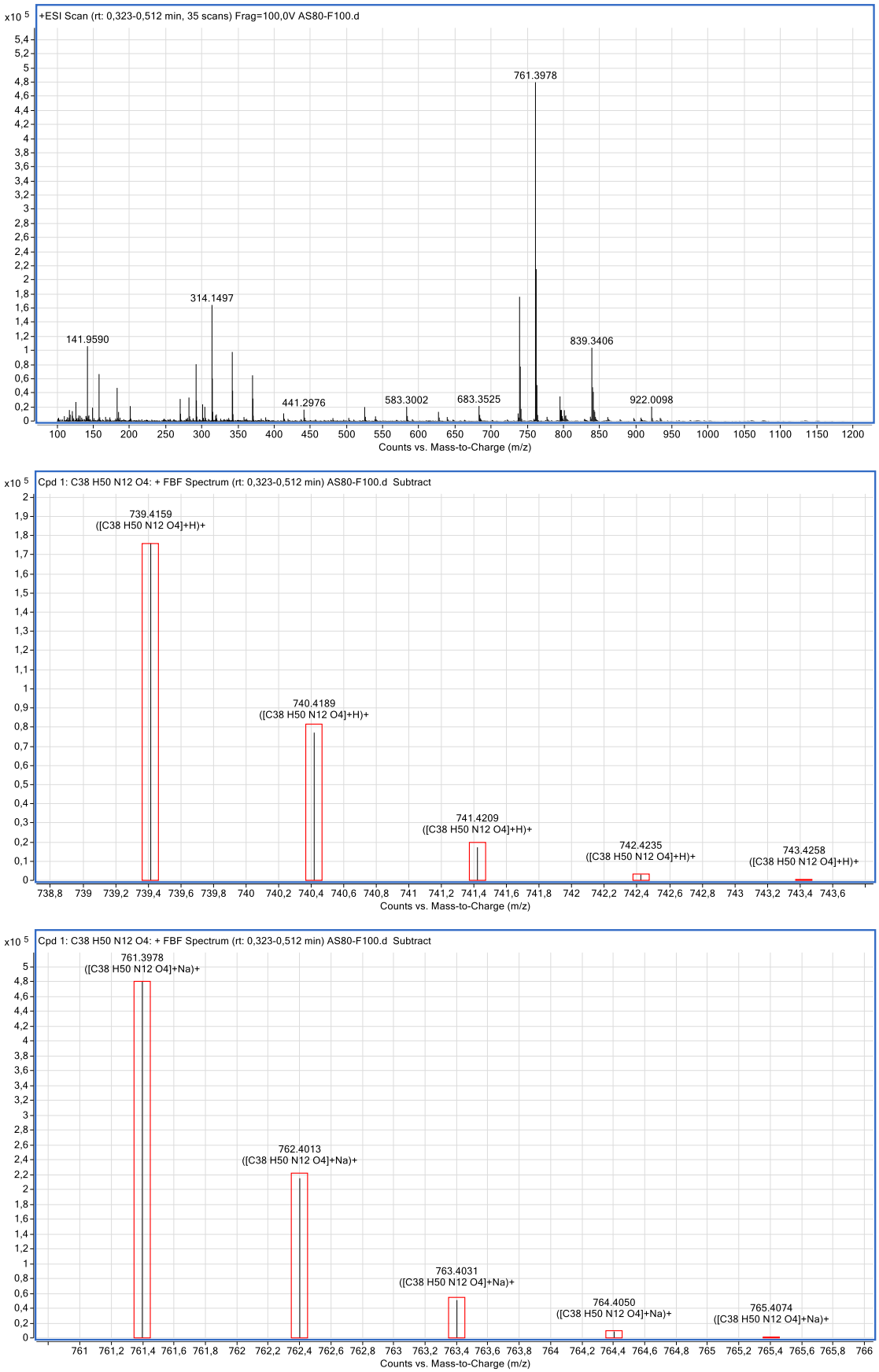


Figure C26. HR-MS (+ESI) spectra for compound **5a**.

Compound **5b**

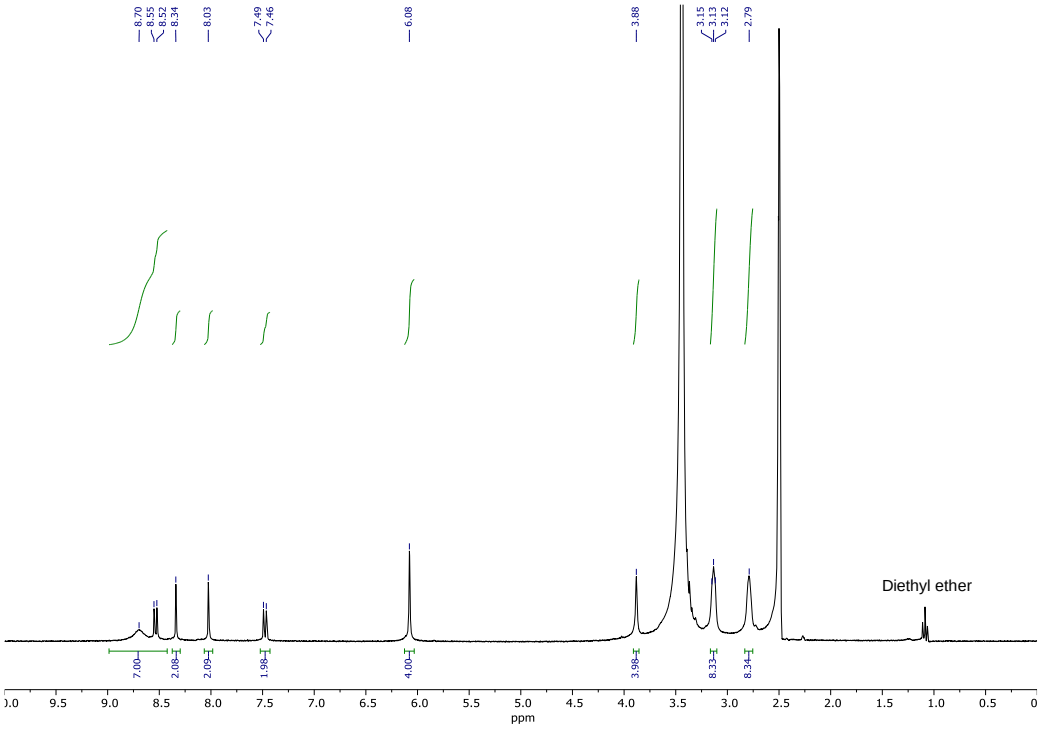


Figure C27. [1]H NMR spectrum (300 MHz, DMSO-d_6) for compound **5b**.

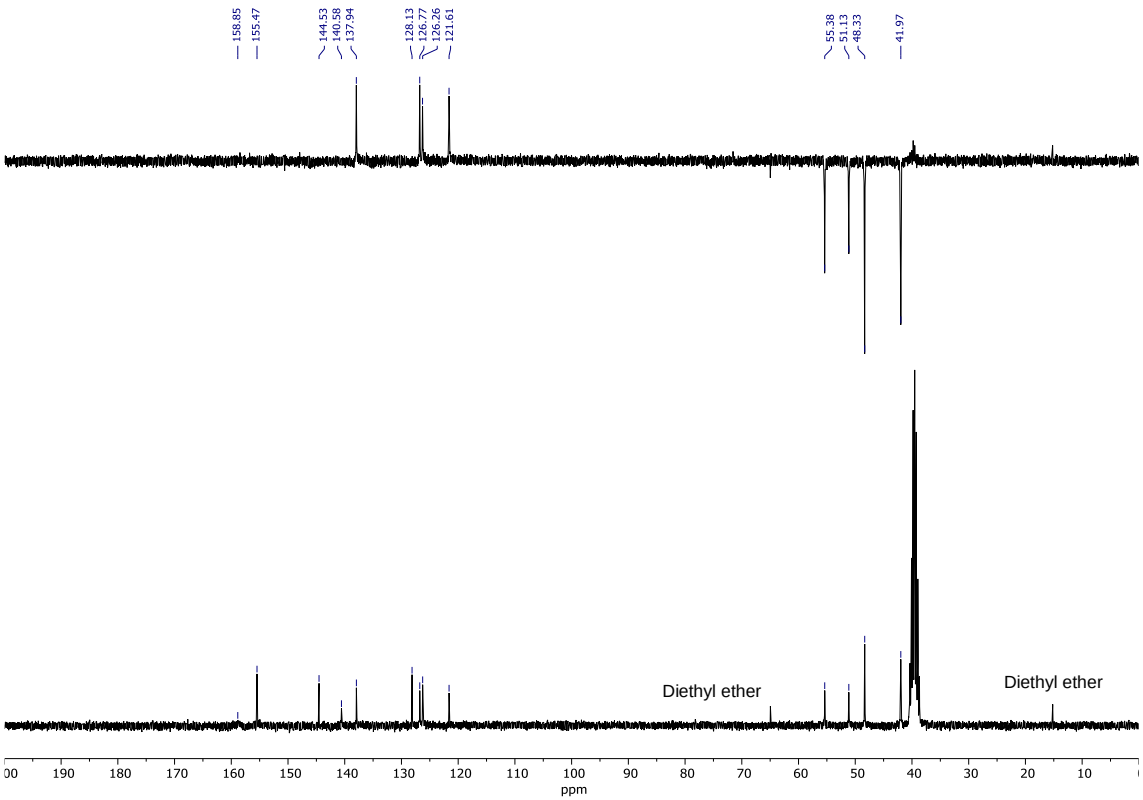


Figure C28. [13]C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **5b**.

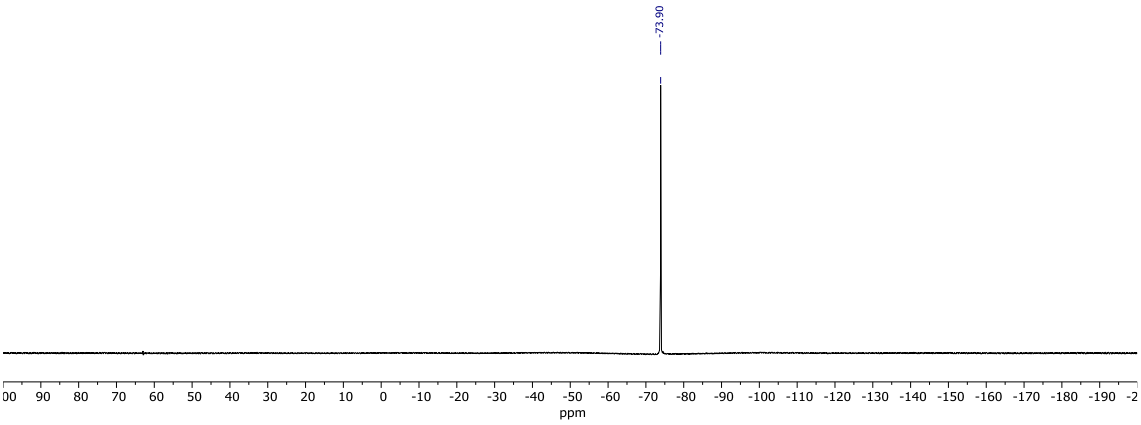


Figure C29. ^{19}F NMR spectrum (282 MHz, DMSO-d_6) for compound **5b**.

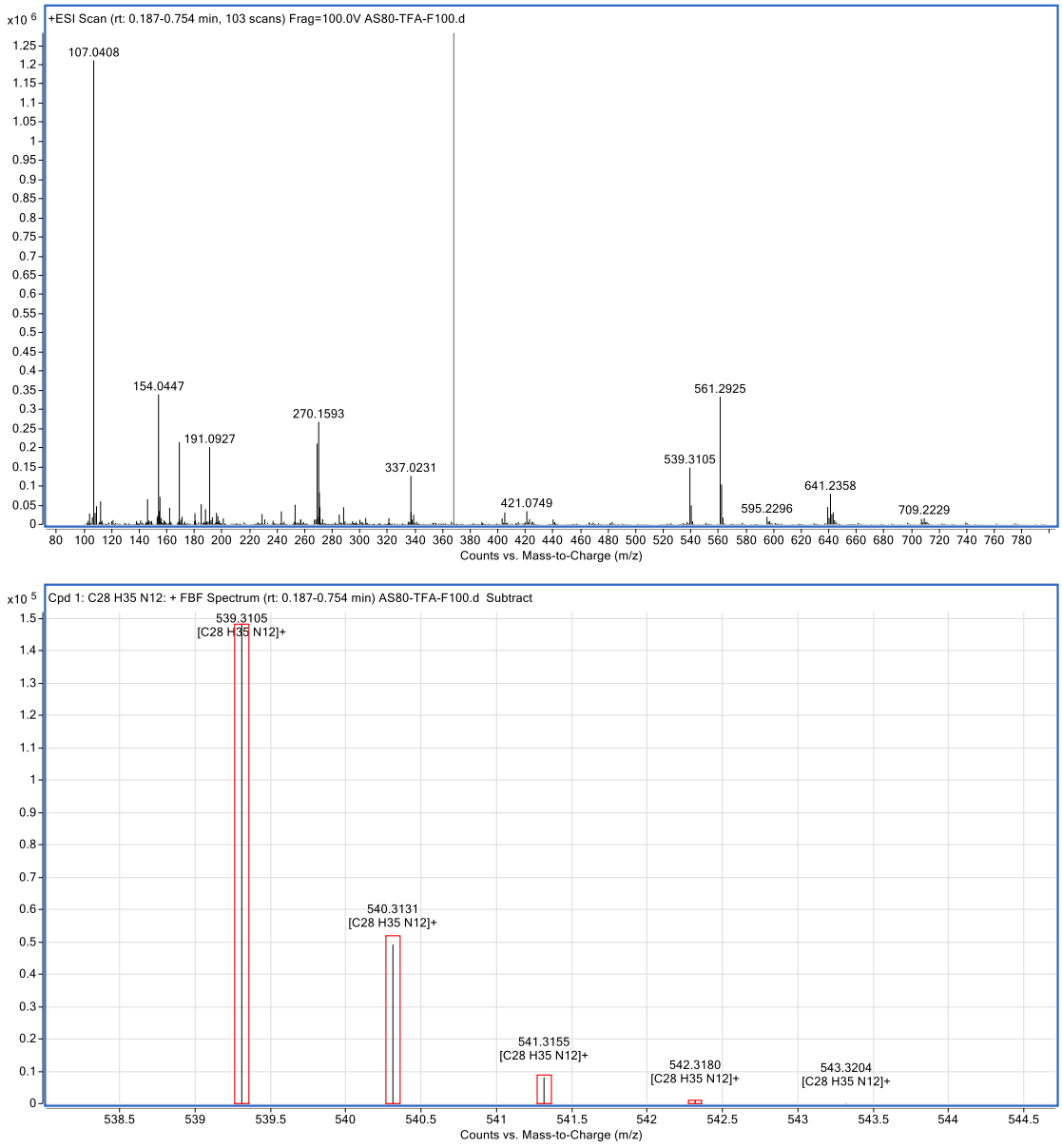


Figure C30. HR-MS (+ESI) spectra for compound **5b**.

Compound **6a**

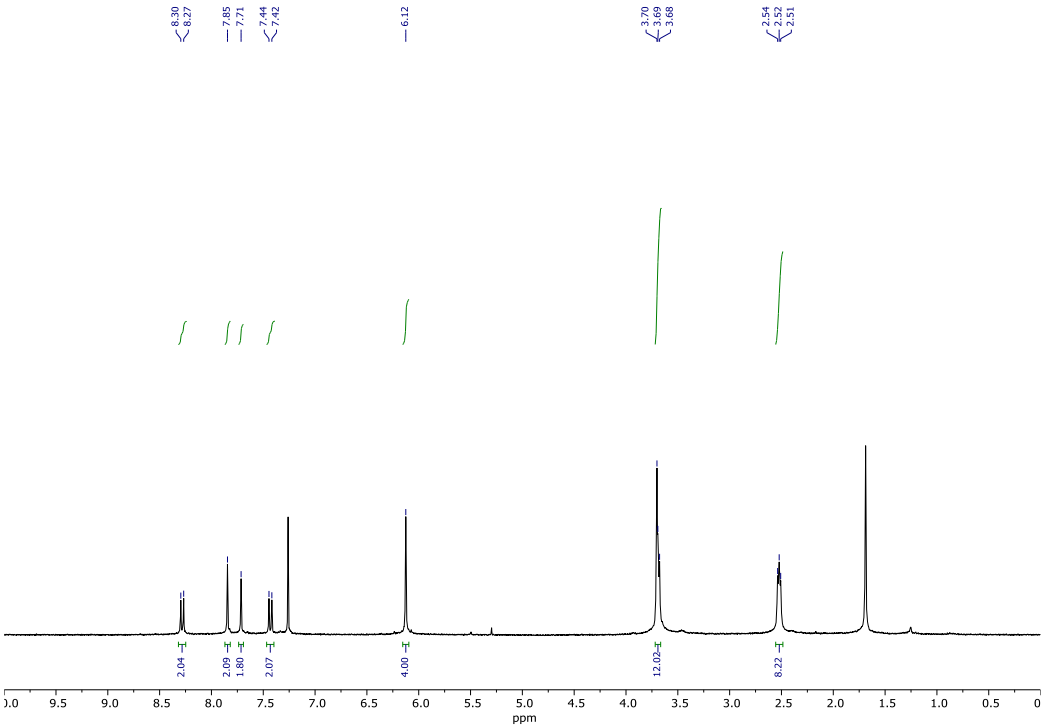


Figure C31. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **6a**.

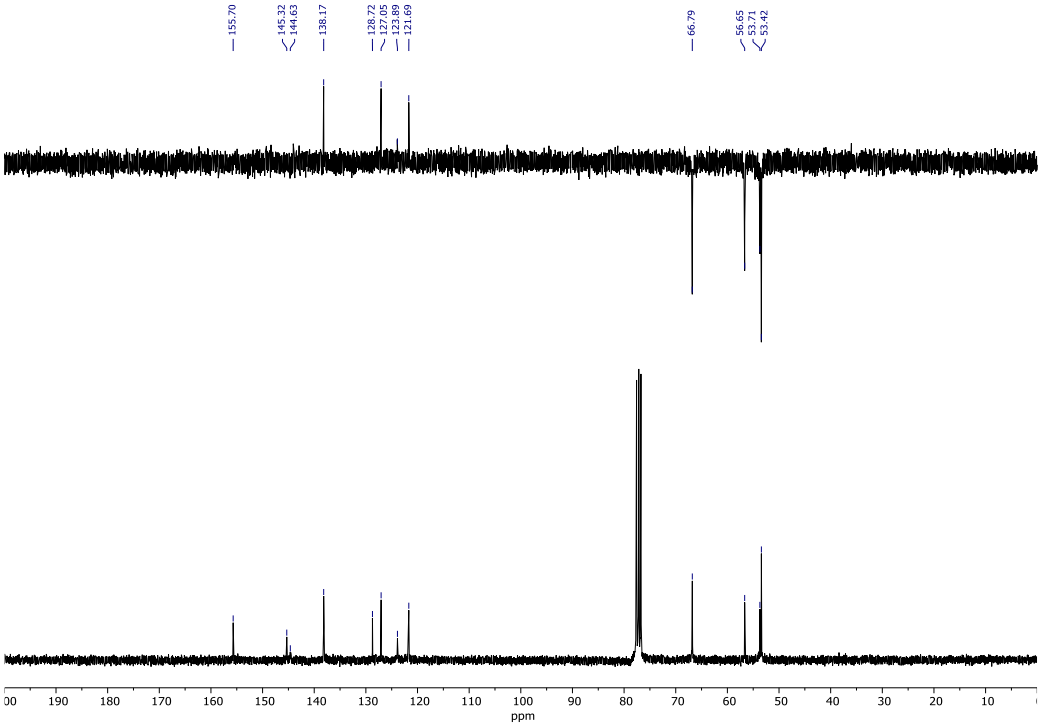


Figure C32. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **6a.**

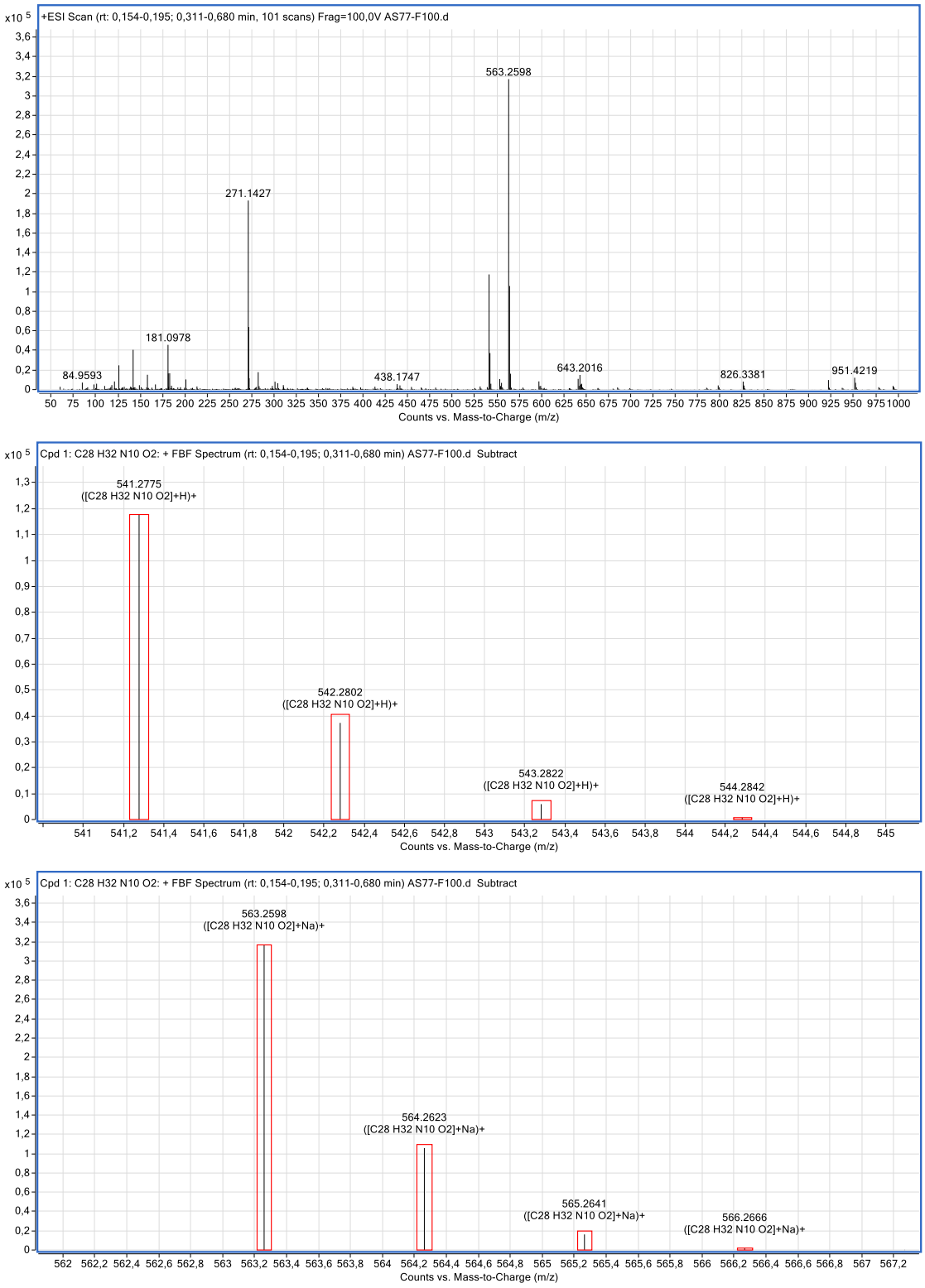


Figure C33. HR-MS (+ESI) spectra for compound **6a**.

Compound **6b**

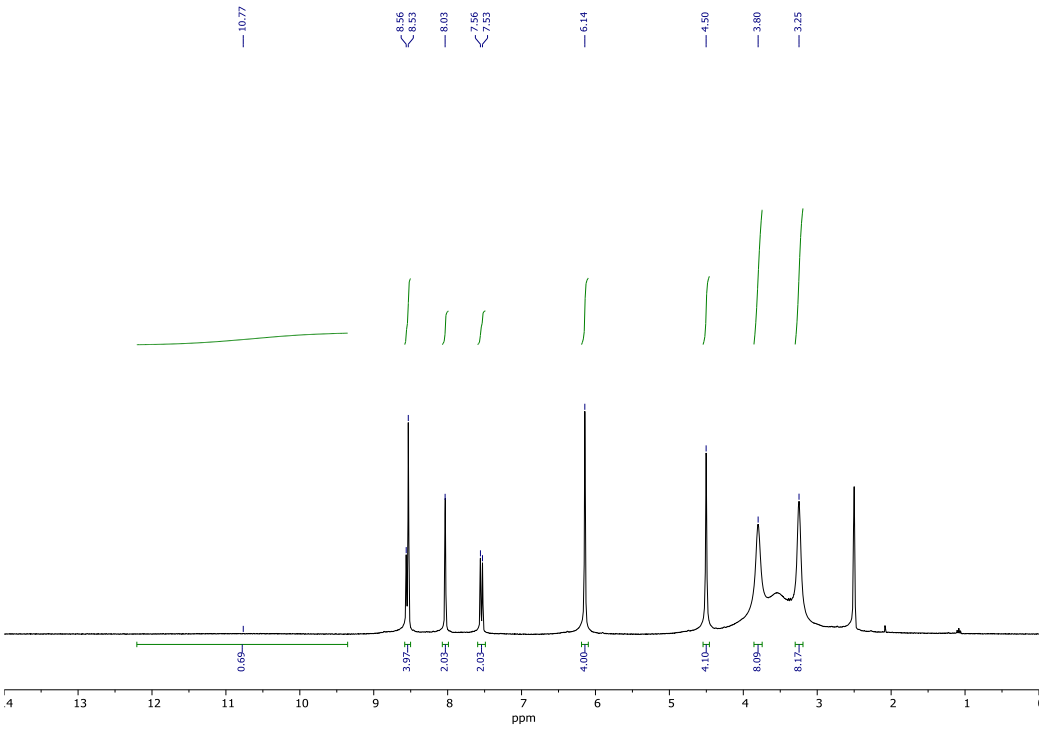


Figure C34. ^{1}H NMR spectrum (300 MHz, DMSO-d_6) for compound **6b**.

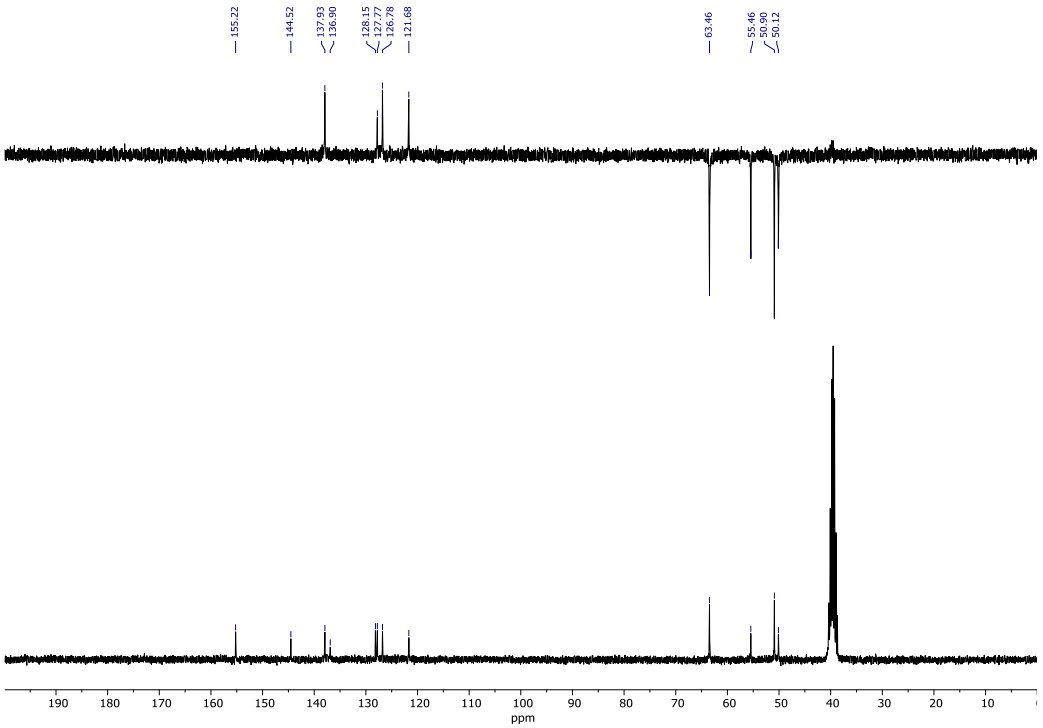


Figure C35. ^{13}C and DEPT-135 NMR spectra (75 MHz, DMSO-d_6) for compound **6b**.

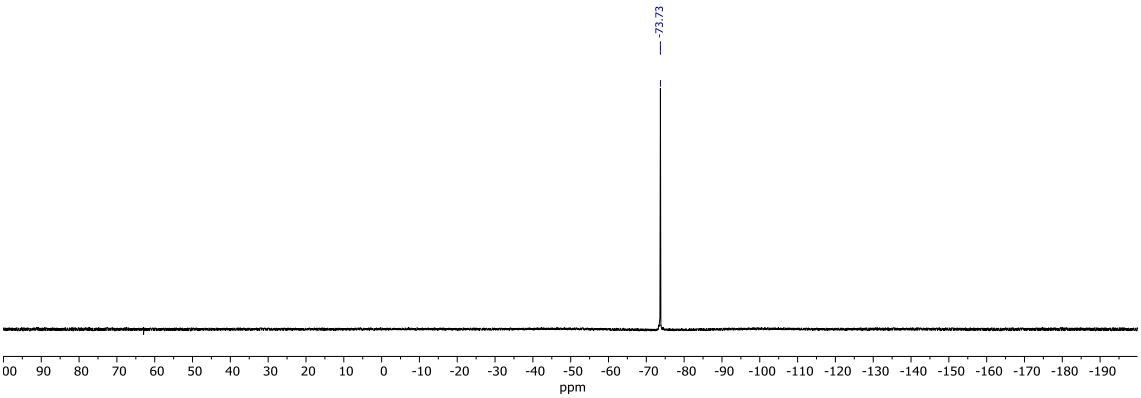


Figure C36. ^{19}F NMR spectrum (282 MHz, DMSO-d_6) for compound **6b**.

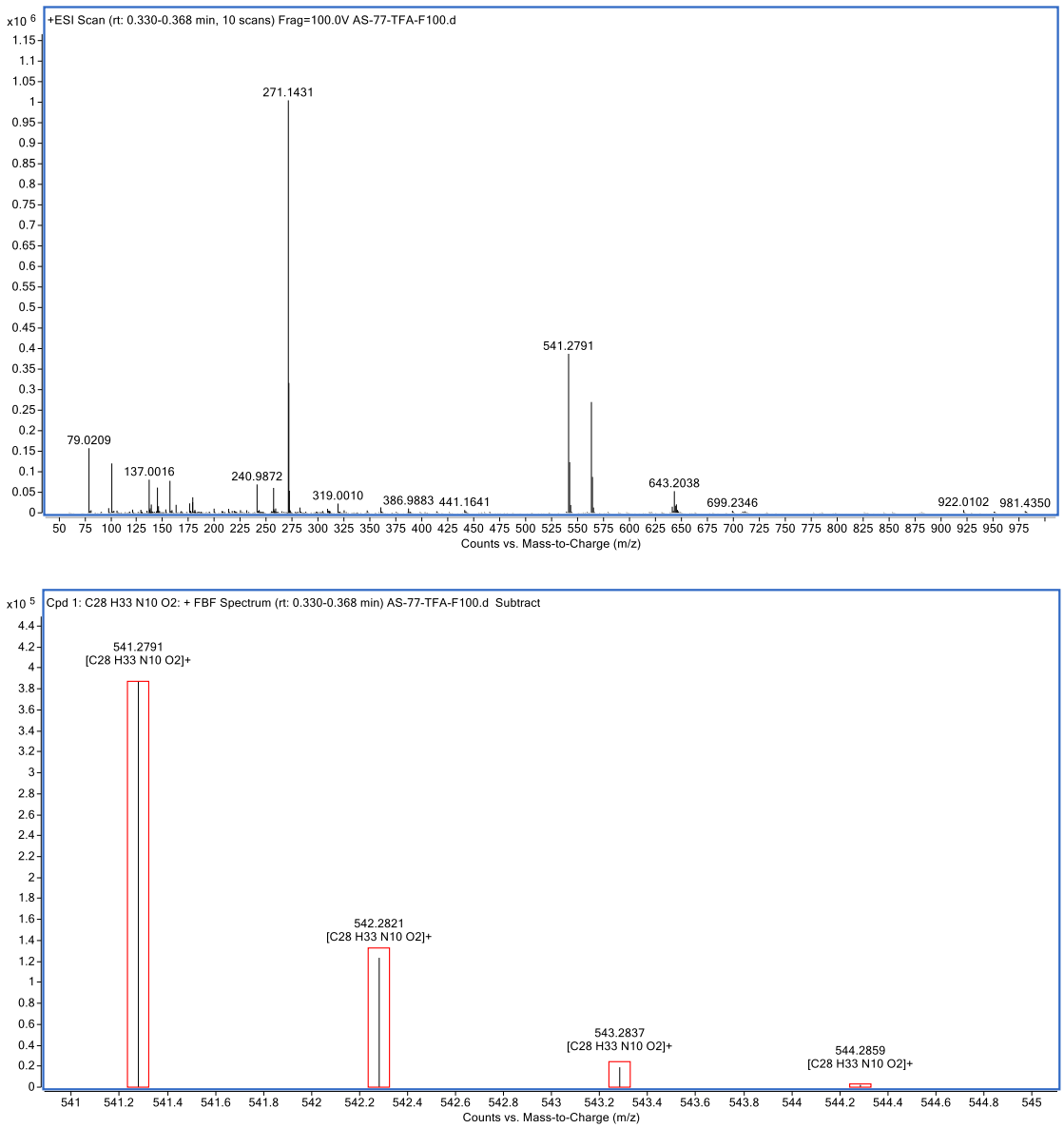


Figure C37. HR-MS (+ESI) spectra for compound **6b**.

Compound **7a**

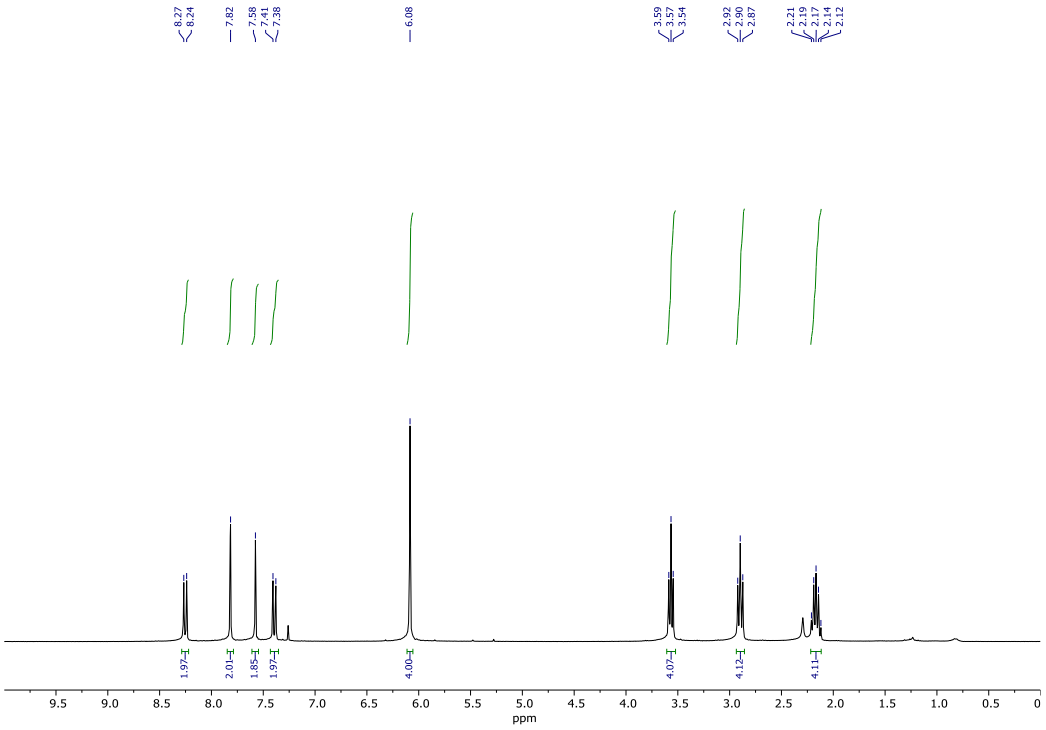


Figure C38. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **7a**.

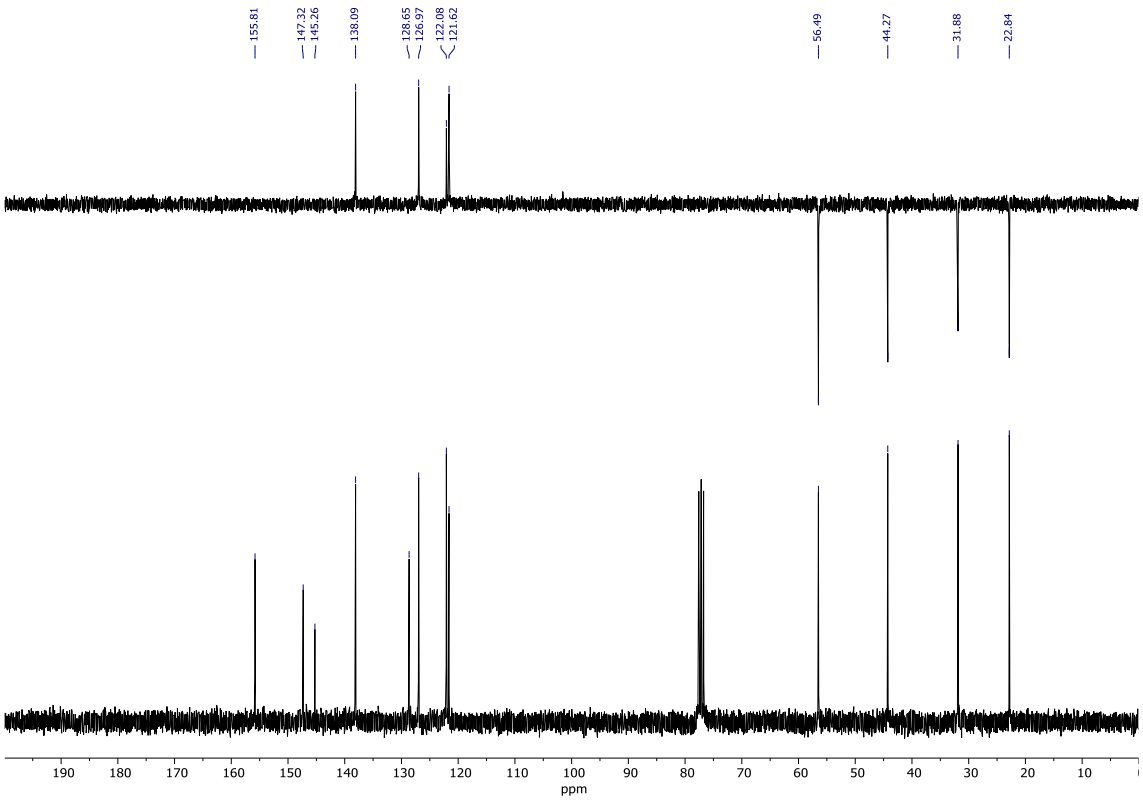


Figure C39. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **7a**.

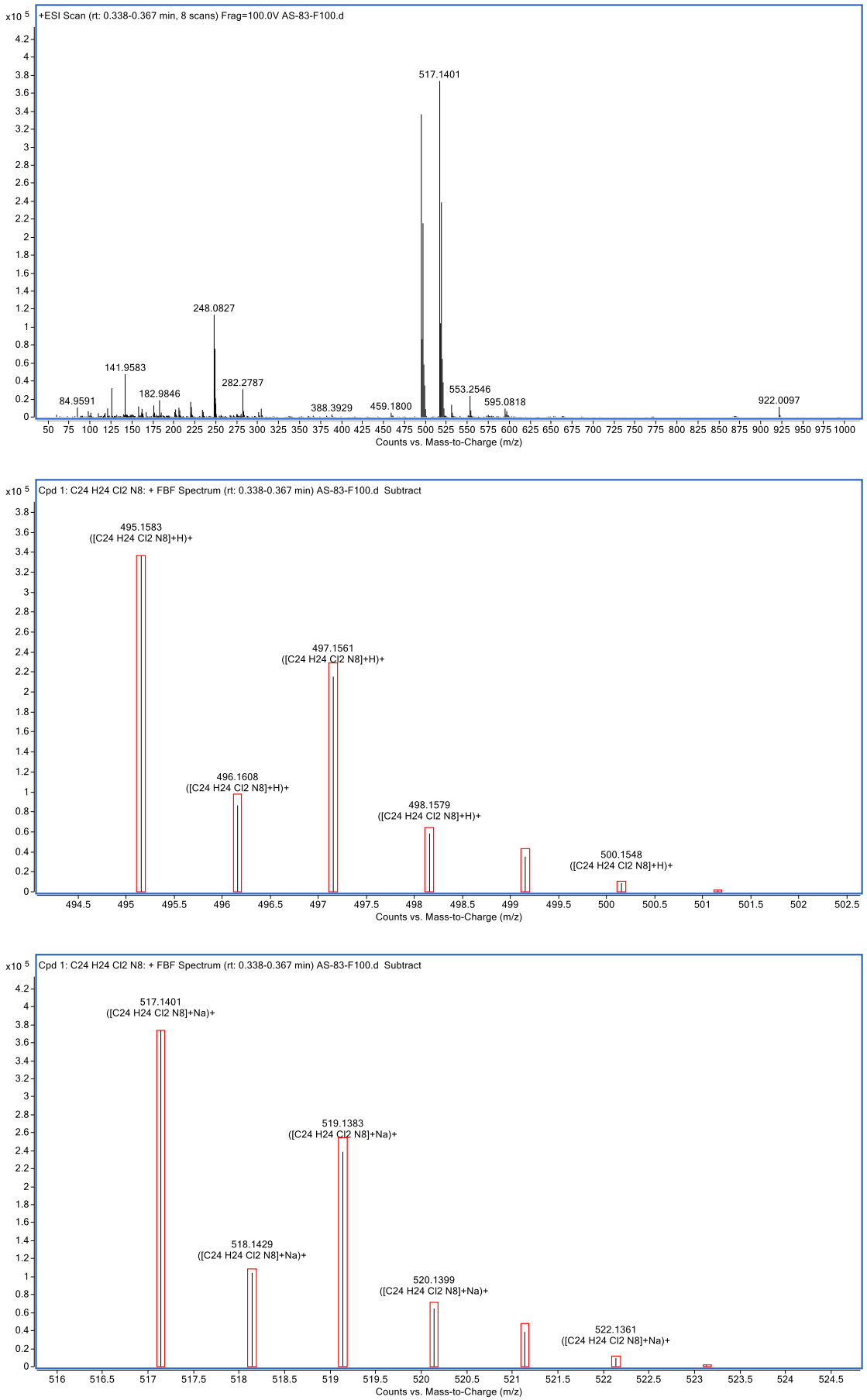


Figure C40. HR-MS (+ESI) spectra for compound **7a**.

Compound **8a**

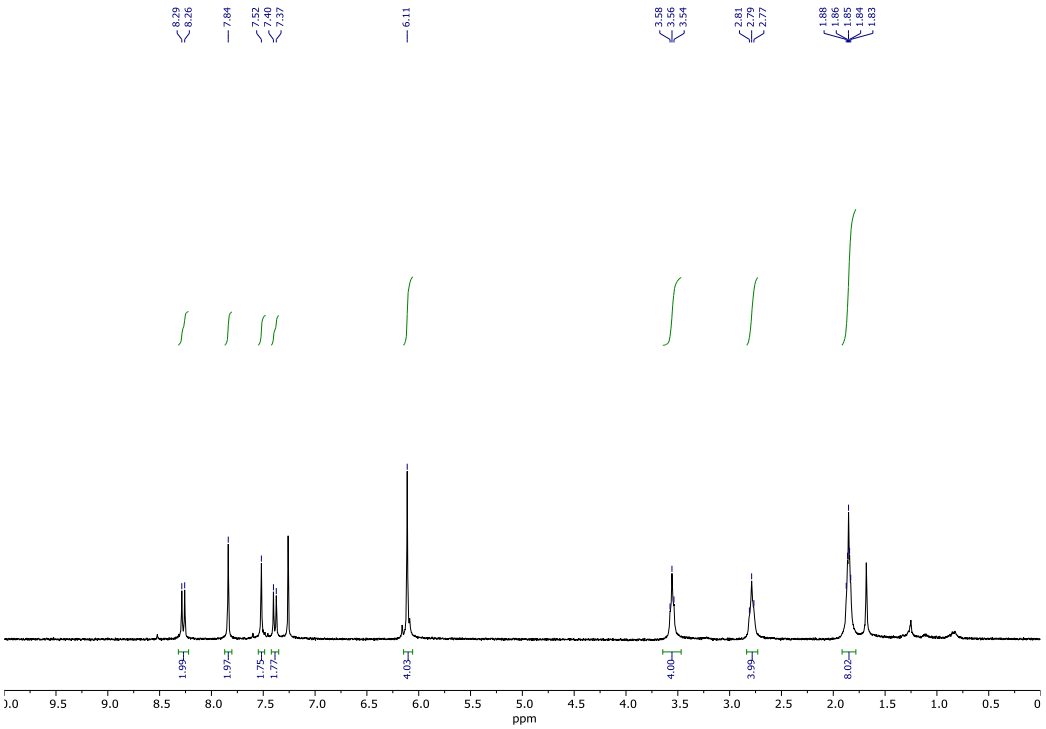


Figure C41. ^{1}H NMR spectrum (300 MHz, $CDCl_3$) for compound **8a**.

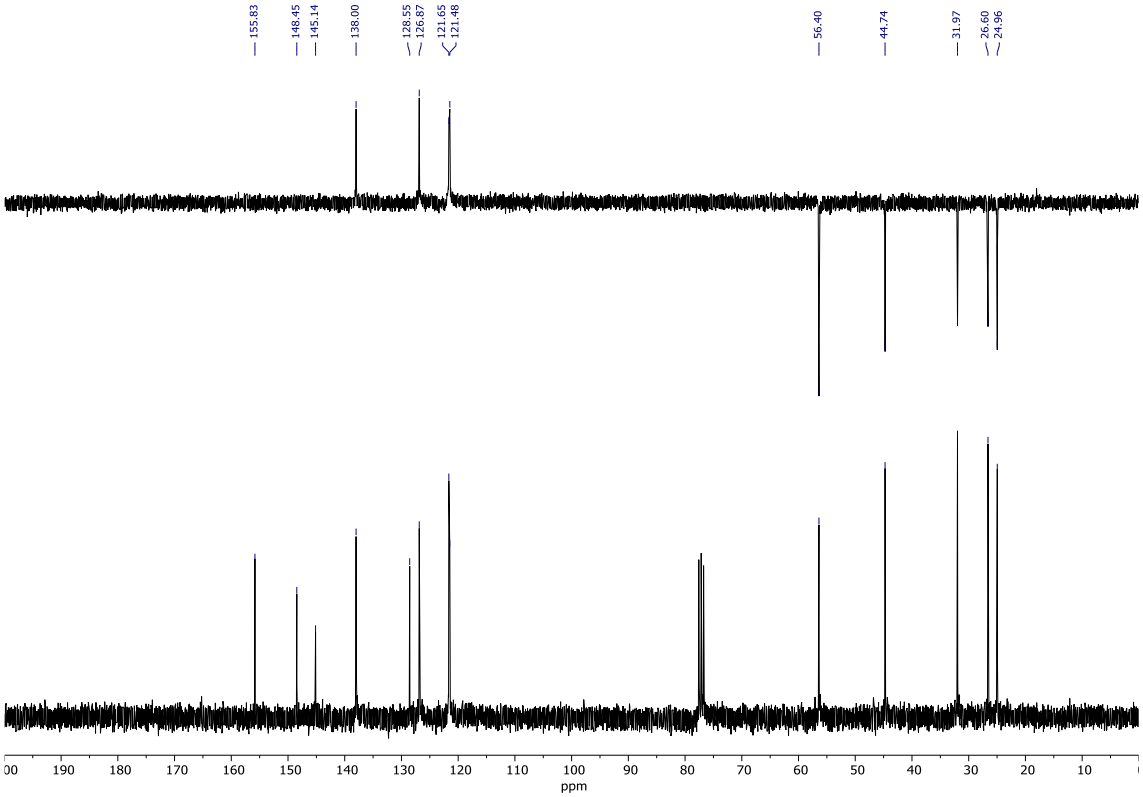


Figure C42. ^{13}C and DEPT-135 NMR spectra (75 MHz, $CDCl_3$) for compound **8a**.

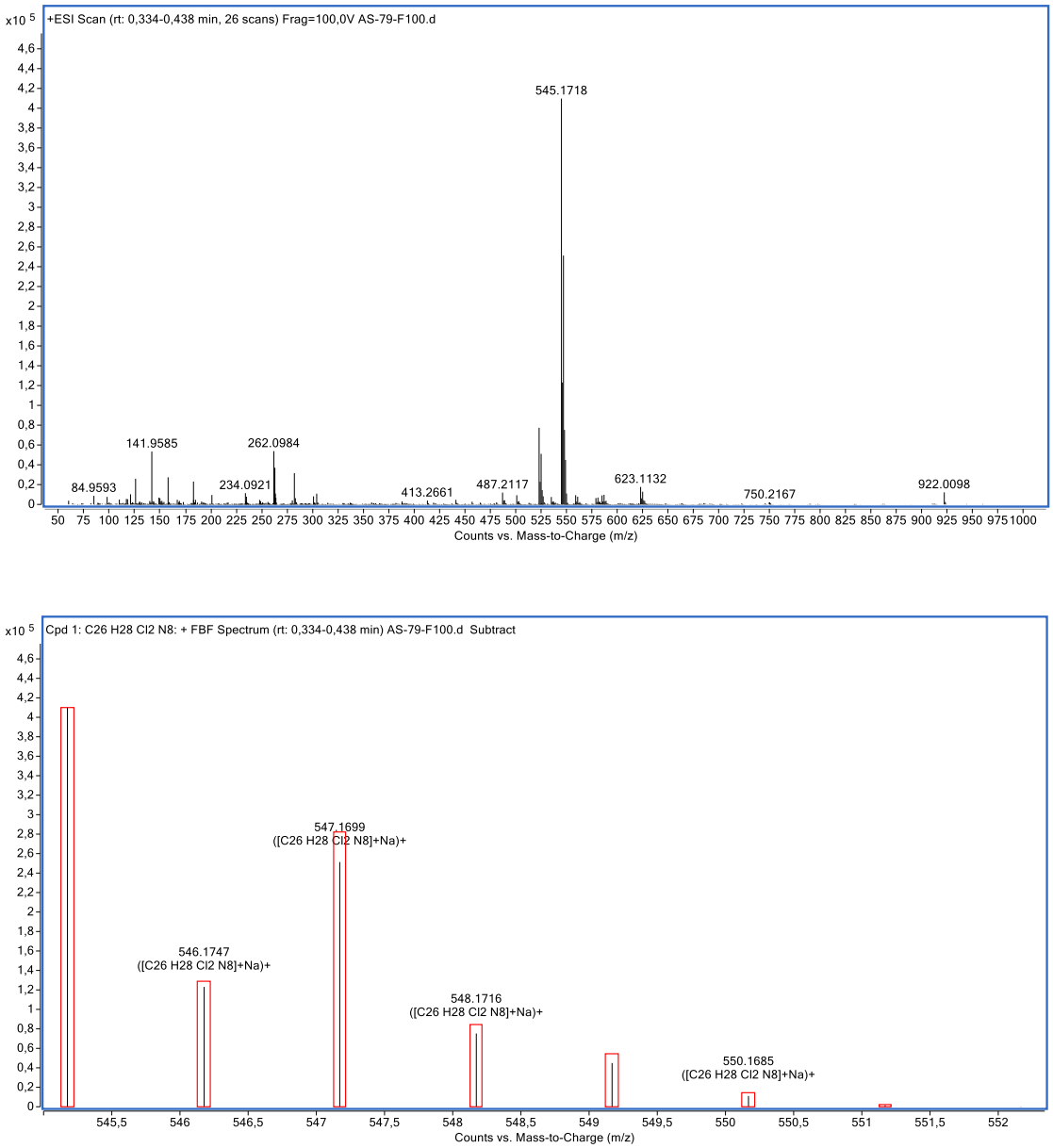


Figure C43. HR-MS (+ESI) spectra for compound **8a**.

In vitro studies

Oligonucleotide sequences

Table C1. List of all oligonucleotide sequences tested.

Name	Sequence (5' ➔ 3')	Topology
F21T	FAM-GGGTTAGGGTTAGGGTTAGGG-TAM	Hybrid
F21CTAT	FAM-GGGCTAGGGCTAGGGCTAGGG-TAM	Antiparallel
F25CebT	FAM-AGGGTGGGTGTAAGTGTGGGTGGGT-TAM	Parallel
FKRAS-21RT	FAM-AGGGCGGTGTGGGAAGAGGGA-TAM	Parallel
FKRAS-22RTT	FAM-AGGGCGGTGTGGGAATAGGGAA-TAM	Parallel
FKRAS-32RT	FAM- AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG-TAM	Parallel
FpreMIR150T	FAM-GGCCUGGGGGACAGGGACCUGGG-TAM	Parallel
FdxT	FAM-TATAGCTATA-hexaethyleneglycol-TATAGCTATA-TAM	Duplex
22AG	AGGGTTAGGGTTAGGGTTAGGG	Hybrid
26Ceb	AAGGGTGGGTGTAAGTGTGGGTGGGT	Parallel
22CTA	AGGGCTAGGGCTAGGGCTAGGG	Antiparallel
KRAS-21R	AGGGCGGTGTGGGAAGAGGGA	Parallel
KRAS-22RT	AGGGCGGTGTGGGAATAGGGAA	Parallel
KRAS-32R	AGGGCGGTGTGGGAAGAGGGAAGAGGGGGAGG	Parallel
FpreMIR150	GGCCUGGGGGACAGGGACCUGGG	Parallel

FRET-melting experiments

Table C2. Thermal stabilization values (°C) of the G4 sequences induced by 5 μM of the synthesized compounds (**2a-8a** and **3b-6b**), measured by FRET melting.

	ΔT_m (°C)[a]				
Ligand	**F21T**	**F21CTAT**	**F25cebT**	**FpreMIR150T**	**FdxT**
2a	1.0 ± 0.2	NS	0.9 ± 0.2	NS	NS
3a	0.9 ±0.2	NS	0.9 ± 0.1	NS	NS
3b	2.3 ± 0.5	NS	1.1 ± 0.2	NS	NS
4a	2.4 ± 0.3	NS	1.0 ± 0.2	NS	NS
4b	3.1 ± 0.4	NS	1.2 ± 0.2	NS	NS
5a	0.9 ± 0.1	NS	0.8 ± 0.1	NS	NS
5b	4.9 ± 0.4	NS	2.1 ± 0.5	NS	NS
6a	1.1 ± 0.1	NS	1.4 ± 0.7	NS	NS
6b	1.4 ± 0.3	NS	0.9 ± 0.1	NS	NS
7a	1.0 ± 0.3	NS	0.8 ± 0.1	NS	NS
8a	0.8 ± 0.4	NS	1.1 ± 0.5	NS	NS
PhenDC3	>30	>30	>30	16.0 ± 1.8	NS

Negative ΔT_m values are reported as NS (no stabilization).

Error margins correspond to SD of three independent experiments.

[a] ΔT_m values of F21T, F21CTAT, F25CebT, FpreMIR150T, and FdxT (0.2 μM) were performed in buffer containing 10 mM lithium cacodylate (pH 7.2), 10 mM KCl and 90 mM LiCl. T_m values for the G4 sequences are 52.1 ± 0.2 (F21T), 57.5 ± 0.1 (F21CTAT), 67.2 ± 0.1 (F25cebT), 69.8 ± 0.4 (FpreMIR150T), and 68.7 ± 0.3 (FdxT).

Table C3. Thermal stabilization values (°C) of the G4 sequences induced by 5 and 2 μM of the synthesized compounds (**2a-8a** and **3b-6b**), measured by FRET melting.

	ΔT_m (°C)[a]					
	5 μM ligand			**2 μM ligand**		
Ligand	**FKRAS-21RT**	**FKRAS-22RTT**	**FKRAS-32RT**	**FKRAS-21RT**	**FKRAS-22RTT**	**FKRAS-32RT**
2a	3.7 ± 0.6	3.3 ± 0.7	2.8 ± 0.6	2.0 ± 1.2	1.4 ± 0.3	0.7 ± 0.3
3a	2.3 ± 0.7	2.9 ± 0.2	3.0 ± 1.0	2.4 ± 1.0	1.0 ± 0.5	0.7 ± 0.5
3b	6.8 ± 0.1	6.3 ± 0.3	4.0 ± 1.0	3.2 ± 1.3	2.5 ± 0.5	1.4 ± 0.5
4a	7.0 ± 0.7	6.8 ± 0.4	4.7 ± 0.7	3.6 ± 1.1	2.9 ± 0.6	1.2 ± 0.7
4b	9.5 ± 0.4	8.5 ± 0.7	5.6 ± 0.6	4.6 ± 1.4	3.3 ± 0.5	1.5 ± 0.5
5a	3.3 ± 0.7	2.8 ± 0.5	2.8 ± 1.3	2.2 ± 0.3	2.0 ± 1.2	0.9 ± 0.4
5b	11.2 ± 1.1	10.6 ± 0.4	7.6 ± 1.0	5.1 ± 1.2	4.5 ± 0.5	1.9 ± 0.3
6a	4.1 ± 0.4	3.5 ± 1.1	2.7 ± 0.8	2.5 ± 0.5	2.1 ± 0.6	0.4 ± 0.3
6b	3.9 ± 0.7	3.8 ± 0.3	3.1 ± 0.5	3.4 ± 2.3	2.2 ± 0.2	0.9 ± 0.4
7a	4.1 ± 0.1	3.1 ± 0.9	2.7 ± 0.5	2.7± 1.2	1.9 ± 0.6	0.4 ± 0.4
8a	3.7 ± 0.1	3.2 ± 0.2	2.8 ± 1.1	2.6 ± 0.3	2.2 ± 0.3	0.9 ± 0.3
PhenDC3	>30	>30	>30	>30	>30	>30

Negative ΔT_m values are reported as NS (no stabilization).

Error margins correspond to SD of three independent experiments.

[a] ΔT_m values of FKras21T, FKras22RTT, FKras32T (0.2 μM) were performed in buffer containing 10 mM lithium cacodylate (pH 7.2), 10 mM KCl and 90 mM LiCl. T_m values for the G4 sequences are 43.2 ± 0.8 (FKras-21T), 43.3 ± 0.5 (FKras-22RT), 47.5 ± 0.6 (FKras-32T).

Circular dichroism studies

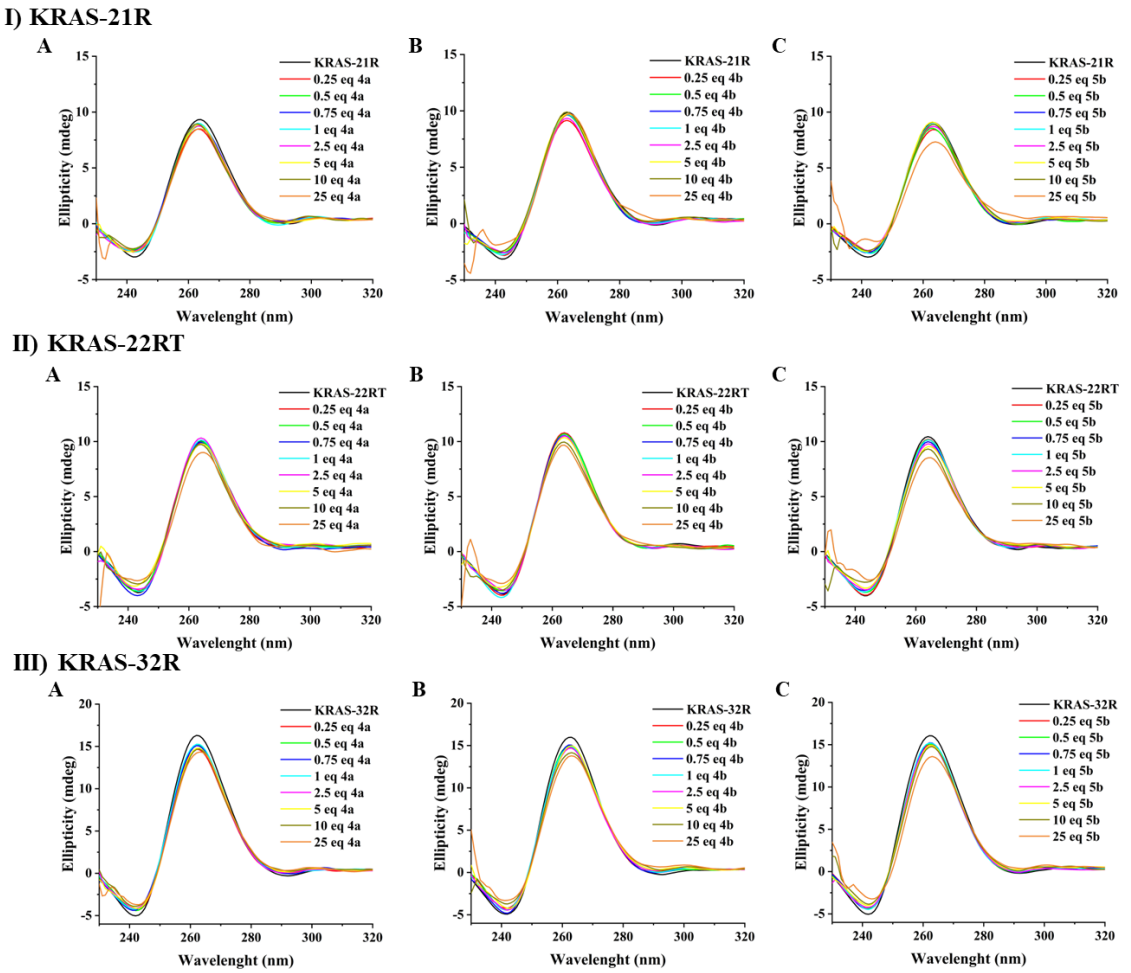


Figure C44. CD titration of I) KRAS-21R II) KRAS-22RT, and III) KRAS-32R G-quadruplex in the absence and presence of increasing molar equivalents of ligands **A) 4a**, **B) 4b**, and **C) 5b**. The experiments were performed in a buffer containing 10 mM lithium cacodylate (pH 7.2) and 100 mM KCl.

Table C4. Thermal stabilization values (°C) of the G4 sequences induced by 25 molar equivalents of the synthesized compounds **4a-b** and **5b**, measured by FRET-melting and CD-melting experiments.

	ΔT_m (°C)					
	FRET-melting			**CD-melting**		
Ligand	**FKRAS-21RT**	**FKRAS-22RTT**	**FKRAS-32RT**	**KRAS-21R**	**KRAS-22RT**	**KRAS-32R**
4a	7.0	6.8	4.7	9.9	11.0	4.3
4b	9.5	8.5	5.6	12.9	12.6	4.9
5b	11.2	10.6	7.6	18.6	18.4	9.5

Fluorimetric Titrations

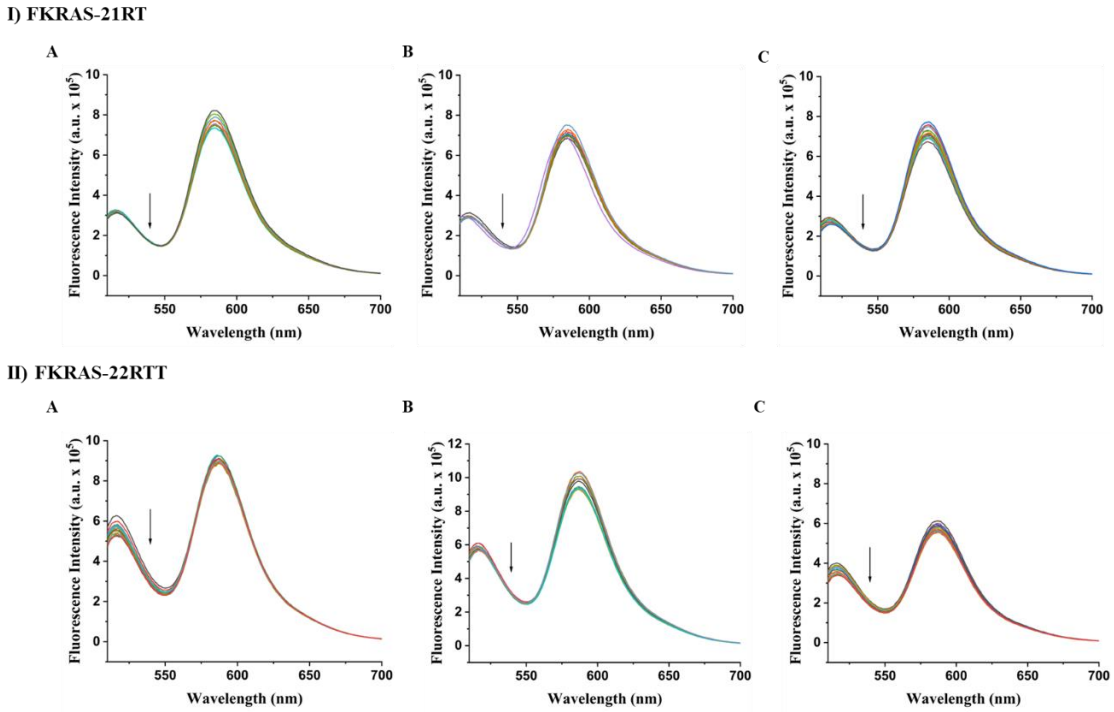


Figure C45. Fluorescence titration spectra of pre-folded I) FKRAS-21RT, and II) FKRAS-22RTT with increasing concentrations of ligands **A) 4a**, **B) 4b**, and **C) 5b**. The experiments were performed in a buffer containing 10 mM lithium cacodylate (pH 7.2) and 100 mM KCl.

Table C5. Apparent dissociation constants (K_D) and Hill coefficient (n) of pre-folded KRAS G4 sequences in the presence of ligands **4a-b** and **5b** obtained by fluorescence titrations.

Ligand	K_D (M)	n
	FKRAS-21RT	
4a	5.4×10^{-7}	0.3
4b	2.7×10^{-9}	0.1
5b	3.5×10^{-6}	0.1
	FKRAS-22RTT	
4a	7.5×10^{-7}	0.2
4b	4.1×10^{-8}	0.4
5b	2.1×10^{-6}	0.3

Cellular studies

Table C6. Percentage of cell viability for compounds **2a-8a** and **3b-6b**, measured by MTT assay after 48h incubation. The values were determined in three independent experiments.

	Percentage (%) of cell viability at 10 μM ligand concentration		
Ligand	**A549**	**H1299**	**MRC-5**
2a	66.0 ± 3.3	36.6 ± 6.3	70.3 ± 8.9
3a	85.0 ± 7.2	92.6 ± 8.9	105 ± 3.8
3b	93.5 ± 2.4	96.1 ± 9.5	101.2 ± 6.5
4a	94.1 ± 7.3	100.8 ± 7.8	104. 6 ± 10.0
4b	93.9 ± 9.7	90.3 ± 9.4	99.3 ± 8.8
5a	76.5 ± 15.5	86.0 ± 8.9	94.0 ± 11.4
5b	83.1 ± 8.7	84.3 ± 7.4	88.9 ± 7.1
6a	91.0 ± 1.7	103.5 ± 4.7	100.9 ± 2.2
6b	94.1 ± 3.8	98.2 ± 3.5	97.8 ± 7.9
7a	76.5 ± 4.8	55.8 ± 2.9	92.3 ± 10.2
8a	76.0 ± 9.2	86.1 ± 6.8	95.2 ± 10.3
PhenDC3	21.9 ± 4.5	55.0 ± 0.3	78.1 ± 0.5
Gemcitabine	17.8 ± 2.0	39.4 ± 4.6	58.2 ± 3.5

Results are presented as mean ± SD. Error margins correspond to SD of three replicates.