Elucidating the cation-dependent folding of DNA G-quadruplexes, and how to target specific topologies with synthetic ligands

Alexander König

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

SPÉCIALITÉ CHIMIE PHYSIQUE

Par Alexander KÖNIG

Elucidating the cation-dependent folding of DNA G-quadruplexes, and how to target specific topologies with synthetic ligands

Élucidant le repliement cation-dépendant des G-quadruplexes d'ADN et comment cibler des topologies spécifiques avec des ligands synthétiques

Sous la direction de : Valérie GABELICA
(co-encadrant : Eric LARGY)

Soutenue le 11 decembre 2023

Membres du jury :

M. FERRAND, Yann	Directeur de recherche, CBMN	président
Mme. VERGA, Daniela	Chargée de recherche, Institut Curie	rapporteur
M. ZENOBI, Renato	Professeur, ETH Zürich	rapporteur
M. HEDDI, Brahim	Chargé de recherche, ENS Paris-Saclay	examinateur
Mme. GABELICA, Valérie	Directeur de recherche, ARNA	directeur
M. MACKERETH, Cameron	Directeur de recherche, ARNA	invité
M. LARGY, Eric	Maître de conférences, ARNA	invité

ACKNOWLEDGEMENTS

This is a tribute to all the people that were part of this journey from my arrival in Bordeaux to my PhD defense. No matter how big or small their contributions were, their names and story matter to me, so I will do my best to acknowledge them all.

1) My thesis supervisors

Valérie and Eric, whose full names and titles are on the previous page. They are both ridiculously talented and I can only hope that some of their brilliance has rubbed off on me. Thank you for always being there when I reached out to you for scientific as well as emotional support. Over these last years I learned so many things from you and I just want to highlight some lessons that left a deep impression on me.

Valérie, thanks to you I will always remember to put the science first, and the aesthetics in second. You also made me pay attention to how I approach long-term projects. I need to learn to appreciate the journey from beginning to end and not just enjoy the relief at the end knowing everything is finished and dealt with.

Eric, thanks to you I learned how much time can be saved by doing a couple hours' worth of reading prior to jumping head-first into a new project. You also went to great efforts to sensibilize me on logical fallacies, especially *non sequitur*. Along with left-handed DNA helices, I now keep noticing them in my day-to-day life.

2) The Gabelica team

The work environment was non-competitive and amical, not only thanks to Valérie's leading style, but also because of the talented international researchers I got to work with in the past years.

Thank you, Fred and Corinne, our MS platform engineers, for keeping the instruments running smoothly and sharing some of your wisdom with me. Cheers to my fellow PhD students, Matthieu and Vincent. Vincent, thank you for synthesizing the molecules that ended up becoming my main research project. I hope your newfound love for analytical chemistry has come to fruition in Nancy. Matthieu, thank you for sticking around and creating the photo album that allows me to recollect my memories of the time spent in Bordeaux even as they fade over time. I'm looking forward to your defense!

Thank you to all the postdocs in our group. Big hugs and thanks to my big Indian sister, my Didi, Debasmita or, as we called her sometimes, Debbie. The energy and passion you brought into the team was like no other and things were never the same after you left. You deserve the world and I can't wait to see you fulfill your dream of becoming a professor at a prestigious Indian school. Another round of hugs to Dominika. The longer we worked together, the more I got to see your genuine kindness and goodwill. I wish for you to persevere through the difficult challenges that life has thrown at you. I hope our paths will cross again, especially since you are looking for opportunities in Germany, so I'm keeping my fingers crossed! Thank you, Nina, for looking after us all, all the best to you and your family. Thank you, Anirban, Sanae and Aram, for the company and the discussions.

Many thanks to all the interns and guest researchers who brought fresh wind into the team. Thank you, Oksana, for letting me discover my new-found taste for seafood. Thank you, Jessy, for the days we prepared samples together and bonded over our shared love for R&B music. You have a beautiful soul and I wish that Paris will be kind to you. Thank you, Cristina, for initiating group activities that brought us all closer together, especially the Holi night. Extra thanks for teaching me the basics of dyeing your hair and recommending me Italian classics that became part of my playlist. Thank you, Guilhem, it was such a pleasure to train you and work with you. You have the curiosity and open-mindedness that make up a great scientist. I hope by now you obtained your Master's degree and keep pushing forward! Thank you, Clarisse, for the immaculate vibes and fashion sense. Just like Matthieu, I'm looking forward to your defense! If I ever become filthy rich, I'll invite you to my luxury home in Dinard. Thank you, Miriam, Jonathan and Romane, for having been a part of this journey.

3) My scientific peers in Bordeaux

I was part of the ARNA laboratory, working in the IECB building. From IECB, I would like to thank the people whom I've seen regularly and shared friendly conversations with.

I want to thank Yann Ferrand and his team, including Vincent, my compatriots Robin and Janine as well as Gabrielle. I want to thank Cameron Mackereth and his team, in particular Amani. Many blessings to you and Tarek! Special thanks to Mikayel and his team, you were all friendly, helpful and a pleasure to talk with! Thank you, Carmelo, for being patient with my grievances over the SAFAS and teaching me about SPR/BLI along the way. Thank you, Pascale, for organizing my 'pôt the thèse' and being the ray of sunshine that you are. Your optimism and kind-heartedness helped bringing some light into my gloomier days. Thank you, Bikash, for the technical support on the instruments. I will miss your peaceful energy. Thank you, Sabrina and Parumita, for the lovely conversations we had. Thank you to Sébastian Campagne and his team. Many thanks to my fellow eccentric Florian, may your publications be seen, your raves be keen and your grass grow green. From the other teams I would like to thank Arunima, Sramona, Shashank, Bilal, Tarek and Mayara. Among the permanent staff I would like to thank Axelle and Estelle for the support on NMR-related questions and Stéphane for initiating the collaboration which led to the first crystal structure of a foldamer-G4-complex and for teaching me the art of crystallogenesis and crystal structure analysis.

I would also like to thank my fellow MS enthusiasts from the Tokarski lab. Thank you, Catherine, for being the best friend I made during this chapter in my life. As we move onwards and out of Bordeaux, I hope we get to stay long-term friends and see each other again someday. For now, I wish you the best on your 1-year tropical paradise retreat, you deserve it more than anyone! Thank you, Vašek, I loved your energy and wish you the best of luck during the final stretch of your PhD. Thank you, Francesca, and a happy future to you in Bordeaux. Congratulations Micheal on your recently defended PhD and I wish you and Yasmin a happy fresh start in Texas!

4) My scientific peers all over the world

Part of my PhD experience was building my network by connecting with fellow scientists all over the world. At the 'G4thering' I met Ines, who I hope to meet in Frankfurt soon. At IntSMS I was part of a group with 5 amazing fellow scientists: Alexia, Nina, Azar, Hui-Chung and Noora-Kaisa. I am so happy that the 6 of us ended up together as we keep chatting and updating each other in our little WhatsApp group chat even 2 years after the summer school! Among the faculty I have fond memories of Yury Tsybin and Ljiljana Paša-Tolić, whom I ended up seeing again at the MSBM summer school. There I also met Mengze again, which is by now the third conference we attended together! Thank you, Mengze, I am really happy we became good friends in the end (third time's the charm, am I right?). It was a pleasure to come visit you in Zürich after my PhD and I'm so honored to be at your wedding in 2 months!!! Thank you, Bo, for being an amazing roomie during the event. I hope to you see in Europe again someday. Many thanks to all the attendees who made MSBM one of my best experiences during my PhD, including Molly, Peng, Michał, Aleksandr, Nemanja, Gabriella, Merve, Olga, Chris and Ken.

5) My roommates

During my PhD I lived in a mansion shared among 12 people, mostly students and researchers, who came from all over the world. I enjoyed the experience of getting in touch with people of all origins and backgrounds, which allowed me to shape my international mindset (as well as expand my cooking skills). I want to thank Kosuke, Charlotte, Bauka, Carlos, Saji, Lisa, Verane, Martin, Pavlina, Benny, Alexandre, Maëlys, June, Masimba, Klára, Vitalina, Livie, Megan, Giorgia, Sudeep and Apurba for being good roommates and acquaintances.

Thank you, Lucile, for being there with me from beginning to end. Thank you, Wanmai, I admire your kindness and good will towards everybody and I am sad that the time we spent together was so short. Good luck with your thesis and stay awesome! Thank you, Clara, our energies matched so well and I enjoyed visiting you in Toulouse multiple times after you had moved out. I am happy you found your way and I will keep liking every drawing and cosplay you share on Instagram! Thank you, Raphaël, despite our rough start and me being impatient with you from time to time I grew to genuinely like your vibe. You've come so far from the young clueless teenager you were when you moved in and I am proud to see your progress! Thank you, Alice, you are an angel and although I was sad to see you go, I will keep rooting for you, wherever you are!

Many thanks to Lauri, who during his short stay got me started on calisthenics, motivated me to learn Finnish, taught me how to operate a grill, how to shine your shoes, as well as everything you need to know about viniculture, wine and alcohol in general. Lauri and I are still good friends and talk to each other almost daily. He came back to Bordeaux one year after he left. I'm looking forward to returning the favor and coming to see him in Finland in the near future.

6) Special thanks to those who supported me through challenging times

Life has its ups and downs. My PhD journey was not only a story of academic achievements, but also one of setbacks, disappointments and constant anxiety. I struggle with openly talking about these kinds of feelings, because I don't want to push my weight onto other people, taking them down with me. For that I want to voice my deepest, most genuine feelings of gratitude towards those who listened and consoled me when I needed it the most. Those people are my friends Claudi, Catherine and Dominika. When I felt like my job is controlling every aspect of my life and I am powerless to it, you three helped be find the courage to seek out a therapist. When I had an existential crisis over having no money, no job and no future home, you supported me by sending me job opportunities and helping me write my first application letters. When I needed someone to talk, whether it was grievances with the French administration, my over-controlling landlady or a failed experiment, you listened. From the bottom of my heart, thank you for being there, I don't know if I could have made it without you.

Many thanks to Anthon Granzhan and Chérine Bechara for being part of my thesis committee. Many thanks to my thesis jury, many of whom traveled to Bordeaux from as far as Zürich to attend my defense. Thank you, Daniela Verga and Renato Zenobi, for being my thesis examiners. Thank you, Brahim Heddi, Cameron Mackereth and Yann Ferrand for agreeing to be part of the jury. Thank you to the EDSC (my doctoral school) for funding my PhD position and thereby letting me have this experience.

THESIS CONTRIBUTIONS

Credit should be given when it's due. I would like to highlight everyone's contributions to my research

Valérie Gabelica, my thesis director, is an engaged team leader and thesis director. She raised me to her high standard of scientific conduct, rigor, integrity, writing and communication. She encouraged me to become an autonomous researcher and supported me when I had ideas for a new research project. I read somewhere that your PhD defines how you conduct science throughout the rest of your career, so I have to give credit to Valérie for leaving a long-lasting positive impact on my research and, on a grander scale, the scientific community.

Eric Largy, my co-supervisor, was equally guiding and encouraging me throughout my thesis. He has at several points contributed to my research with his outstanding data processing skills. I want to mention here that he performed the SVDs in chapter 2 and did a significant portion of the data processing for chapter 3, including: processing the ESI-MS ligand screening and ESI-MS titration data, calculating the K_D values, performing dynamic fitting, plotting all the mass spectra, plotting the heatmap (for the ligand screening). His database of G-quadruplex structures and reference data (that he created with Anirban) was an essential tool in helping me create adequate DNA sequence panels for my projects. He also makes great G-quadruplex schemes, which I used a couple times in my images (you can see them in Figures 17-19, for example).

Frédéric Rosu, our research engineer and lab manager, made sure all the MS instruments would be at peak performance. He would help out with any technical issues and share his expertise, both instrument- and research-sided.

Cameron Mackereth is the mastermind behind all my NMR experiments. I could present him the research project and the open questions to be addressed by NMR – he would find a way to make it work. Extra credit to him for helping me overcome my fear of NMR that was created from an extremely demanding syllabus at my *alma mater*.

Vincent Laffilé, under the supervision of Yann Ferrand, synthesized, purified and characterized the foldamers that play a key role in Chapter 3. Vincent could make any day feel like Christmas when he came by with a new batch of ligand.

Stéphane Thore contributed by resolving the crystal structure of the 222T/QQPQ complex. He taught me the foundations of crystallogenesis and x-ray crystallography.

TABLE OF CONTENTS

ABBREVIATIONS

BEM	Bead Ejection Model
CCS	Collisional Cross Section
CD	Circular Dichroism
CEM	Chain Ejection Model
COSY	Correlation Spectroscopy
CRM	Charged Residue Model
DNA	Deoxyribonucleic Acid
DTIMS	Drift Tube Ion Mobility Spectrometry
ESI	Electrospray Ionization
FID	Free Induction Decay
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Correlation
HOMO	Highest Occupied Molecular Orbital
HSQC	Heteronuclear Single Quantum Correlation
IDS	Isothermal Difference Spectra
IEM	Ion Evaporation Model
IMS	Ion Mobility Spectrometry
IP-RP-LC	Ion Pairing Reverse Phase Liquid Chromatography
LC	Liquid Chromatography
LUMO	Lowest Unoccupied Molecular Orbital
MALDI	Matrix-assisted Laser Desorption Ionization
MES	2-(N-morpholino)ethanesulfonic acid
MPD	2-Methyl-2,4-Pentanediol
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
PEG	Polyethylene Glycol
RNA	Ribonucleic Acid
RP	Reverse Phase
SVD	Singular Value Decomposition
TDS	Thermal Difference Spectra
TMAA	Trimethyl Ammonium Acetate
TO	Thiazole Orange
TOCSY	Total Correlation Spectroscopy
TWIMS	Travelling Wave Ion Mobility Spectrometry

THESIS ABSTRACT

Titre : *Élucidant le repliement cation-dépendant des G-quadruplexes d'ADN et comment cibler des topologies spécifiques avec des ligands synthétiques*

Résumé :

Cette thèse est consacrée au repliement des structures secondaires de l'ADN. L'objectif principal est de comprendre comment les cations, la séquence et les ligands de petites molécules influencent la topologie des G-quadruplexes. Les G-quadruplexes sont des structures secondaires uniques qui se forment dans les brins riches en guanine, avec quatre guanines s'associant en une tétrade (G-tetrad ; G4). En rationalisant nos résultats, nous avons essayé d'élucider certains aspects fondamentaux du repliement des G-quadruplex. La recherche sur les G-quadruplex reste très empirique, c'est pourquoi la compréhension des facteurs déterminants du repliement et de la stabilité est nécessaire pour la prédiction de la structure et la conception rationnelle de ligands. Nous utilisons notre expertise en chimie analytique, en particulier la spectrométrie de masse native, pour l'analyse biophysique de la structure de l'ADN. Nous avons étudié les effets des cations/ligands sur la topologie de l'ADN dans trois projets distincts.

Le Projet 1 analyse l'interaction entre les lanthanides et l'ADN G-quadruplex. Etant des cations trivalents, les lanthanides ont le plus grand potentiel pour réduire la densité de charge, ce qui est nécessaire pour le repliement des G-quadruplexes. Nous présentons la première analyse systématique porte sur (1) plusieurs ions lanthanides, (2) des brins d'ADN de composition de base et de richesse G différentes, et (3) une analyse structurale des complexes lanthanide-ADN. Nous déduisons des mesures de luminescence, de CD et de MS que Eu^{3+} et Tb^{3+} (mais ni La^{3+} ni Yb^{3+}) ciblent spécifiquement les guanines dans l'ADN riche en G, avec 1 Tb^{3+} pour 6 à 12 guanines. Nos données ni valident ni infirment la formation de G-quadruplex, mais nous proposons plusieurs structures hypothétiques.

Le Projet 2 analyse l'influence de la séquence sur la conformation des G-quadruplexes en milieu potassium. Ce projet a été motivé par le fait que les G-quadruplexes avec trois G-tétrades hétérostacking ont été décrits dans Na^+ mais jamais dans K^+. Nous avons modifié des séquences d'ADN qui forment des G-quadruplexes antiparallèles à deux tétrades, stabilisés par des triades de bases, en prolongeant les tracts G par des G supplémentaires. L'objectif était d'insérer une troisième G-tétrade, tout en maintenant la conformation antiparallèle (et l'hétérostacking). Les séquences modifiées présentent en deux conformères principaux : un G-quadruplex hybride à 3 tétrades et un G-quadruplex antiparallèle avec un nombre incertain de tétrades. La forme hybride est thermodynamiquement favorisée et obtenue par annealing, tandis que la forme antiparallèle est cinétiquement favorisée et obtenue par ajout de K^+ à basse température. Notre travail permet de comprendre comment le K^+ détermine le comportement biophysique de l'ADN G-quadruplex et permet de concevoir des séquences modèles optimisées pour les structures G-quadruplex atypiques.

Le Projet 3 est une étude exhaustive sur l'interaction non covalente entre de petites hélices oligo-aromatiques ("foldamères") et une large gamme de structures G-

quadruplex. Le criblage par spectrométrie de masse native révèle que nos foldamères ciblent sélectivement les G-quadruplexes parallèles et sont sensibles à l'accessibilité stérique des G-tétrades, les boucles ou les nucléotides flanquants empêchant la liaison des foldamères. Notre structure cristalline montre deux sous-unités de quinoléine interagissant avec un G-tétrade externe par empilement π. La RMN en solution confirme que le foldamère cible les extrémités 3' et 5' du G-quadruplex. La sélectivité conformationnelle des foldamères provient de leur forme hélicoïdale "volumineuse", qui impose des restrictions stériques à la liaison du G-quadruplex. Nos résultats présentent les foldamères comme des ligands sélectifs prometteurs du G-quadruplex avec un échafaudage unique qui peut être modifié de plusieurs manières pour améliorer encore l'affinité et la sélectivité.

Keywords: spectrométrie de masse, chimie supramoléculaire, G-quadruplex

---

Title: *Elucidating the cation-dependent folding of G-quadruplexes and how to target specific topologies with synthetic ligands*

Abstract:

This thesis is dedicated to the folding of secondary DNA structures. The central goal is to understand how cations, sequence and small molecule ligands influence G-quadruplex topology. G-quadruplexes are unique secondary structures that form in guanine-rich strands, with four guanines associating into a G-tetrad (G4). By rationalizing our findings, we strive to elucidate some fundamentals aspects of G-quadruplex folding. G-quadruplex research remains highly empirical, therefore understanding the determinants of folding and stability is strongly needed for structure prediction and rational ligand design. We utilize our expertise in analytical chemistry, in particular native mass spectrometry, for the biophysical analysis of DNA structure. We studied the cation/ligand effects on DNA topology in three separate projects.

Project 1 analyzes the interaction between lanthanides and G-quadruplex DNA. As trivalent cations, lanthanides have the greatest potential in reducing charge density, which is required for G-quadruplex folding. We present the first systematic analysis that features (1) several lanthanide ions, (2) DNA strands of different base composition and G-richness, and (3) a structural analysis of lanthanide-DNA complexes. We deduce from luminescence, CD and MS measurements that Eu^{3+} and Tb^{3+} (but not La^{3+} or Yb^{3+}) specifically target guanines in G-rich DNA, with 1 Tb^{3+} per 6 to 12 guanines. Our evidence cannot validate or disprove G-quadruplex formation, but lets us propose several hypothetical structures.

Project 2 analyzes the influence of the sequence on G-quadruplex conformation in K^+. This project was motivated by the fact that G-quadruplexes with three heterostacking G-tetrads form with Na^+ and were never observed in K^+. We modified DNA sequences that form 2-tetrad antiparallel G-quadruplexes, stabilized by base triads, by extending the G-tracts with additional Gs. The objective is the insertion of a third G-tetrad, while maintaining the antiparallel (and heterostacking) conformation. The modified sequences have two main conformers: a 3-tetrad hybrid G-quadruplex and an antiparallel G-quadruplex with an uncertain number of G-tetrads. The hybrid form is thermodynamically favored and obtained by annealing while the antiparallel form is kinetically favored and obtained by adding K^+ at low temperature. Our work gives insights for how K^+ determines the biophysical behavior of G-quadruplex DNA and enables the design of optimized model sequences for atypical G-quadruplex structures.

Project 3 is a comprehensive study on the non-covalent interaction between small oligo-aromatic helices (‘foldamers’) and a wide range of G-quadruplex structures. The native mass spectrometry screening reveals that our foldamers selectively target parallel G-quadruplexes and are sensitive to the steric accessibility of G-tetrads, with loops or flanking nucleotides preventing foldamer binding. Our crystal structure shows two quinoline subunits interacting with an external G-tetrad through π-stacking. Solution NMR confirms that the foldamer targets the 3’ and 5’ end of the G-quadruplex. The conformational selectivity of foldamers originates from their ‘bulky’ helical shape, which imposes steric restrictions on G-quadruplex binding. Our results introduce

foldamers as promising selective G-quadruplex ligands with a unique scaffold that can be modified in several ways to potentially further improve affinity and selectivity.

Keywords: mass spectrometry, supramolecular chemistry, G-quadruplex

---

Unité de recherche

ARNA - Acides nucléiques : régulations naturelles et artificielles. Université de Bordeaux, INSERM 1212, CNRS UMR 5320, 146 rue Léo Saignat, 33000 Bordeaux.

INTRODUCTION

Preamble

My PhD work encompasses three main projects, which are presented in this thesis manuscript. The manuscript does not feature my contribution to a review article published by the team (Mass Spectrometry of Nucleic Acid Noncovalent Complexes, Largy *et al.*, Chem. Rev. 2022, 122 (8), 7720-7839). Neither does it feature the talks and posters I had presented at the G4thering, IMSC and 'GQs made in France' conferences or the IntSMS and MSBM summer schools.

«The application of analytical chemistry to study the biophysics and topology of G-quadruplex DNA » is how I would formulate the quintessence of this thesis project. To give every reader an equal opportunity to understand and interpret my research, the manuscript starts with an introductory review, summarizing everything the reader needs to know on G-quadruplex biophysics and analytical chemistry before diving into the three main projects.

We start by distinguishing the G-quadruplex from other secondary DNA structures and introduce its main topologies. Then, we present all the characteristic features of a G-quadruplex: loops, flanking nucleotides, grooves, glycosidic bond angle, base stacking, H-bond rotation, handedness, molecularity and cations. The i-motif and its relation to the G-quadruplex is outlined, since it serves as an alternative target in the foldamer project (chapter 3). The final section presents G-quadruplex ligands in a simplified manner. They are only relevant to chapter 3, where they are reintroduced with greater detail.

The second half of the introduction is dedicated to the five main analytical methods that we used: native mass spectrometry (MS), circular dichroism (CD), UV melting, x-ray crystallography and NMR. We focus on the application to G-quadruplex analysis, following a practical approach of demonstrating what information can be extracted and how it is done. We include a theoretical introduction for native MS and CD, because these two methods will be discussed in great detail throughout every chapter. Native MS is a key method because our team has great expertise in native MS, challenging the limits of how much structural information can be obtained from MS alone. As for CD, I made it my personal mission to summarize the fundamentals of CD absorption in G-quadruplexes, because I could not find a proper review in the literature.

Methods that were used, but not featured in the introduction are 1) UV spectroscopy, because it was used to validate stock concentrations and was therefore not involved in biophysical analysis, and 2) luminescence spectroscopy, which is only relevant to the lanthanide project (chapter 1) and therefore introduced within chapter 1.

The main part of the thesis is separated into three different projects, a summary of whom is given in the thesis abstract. Chapter 1 is about lanthanides interacting with G-rich sequences, Chapter 2 is about trying to create a 3-tetrad antiparallel G-quadruplex in K^+ and Chapter 3 is a comprehensive study on foldamers as topology-selective G-quadruplex ligands.

The G-quadruplex

From oligonucleotides to G-quadruplexes

DNA and RNA are polymers of nucleotides. The monomeric unit is a nucleotide, with a DNA nucleotide depicted in Figure 1.

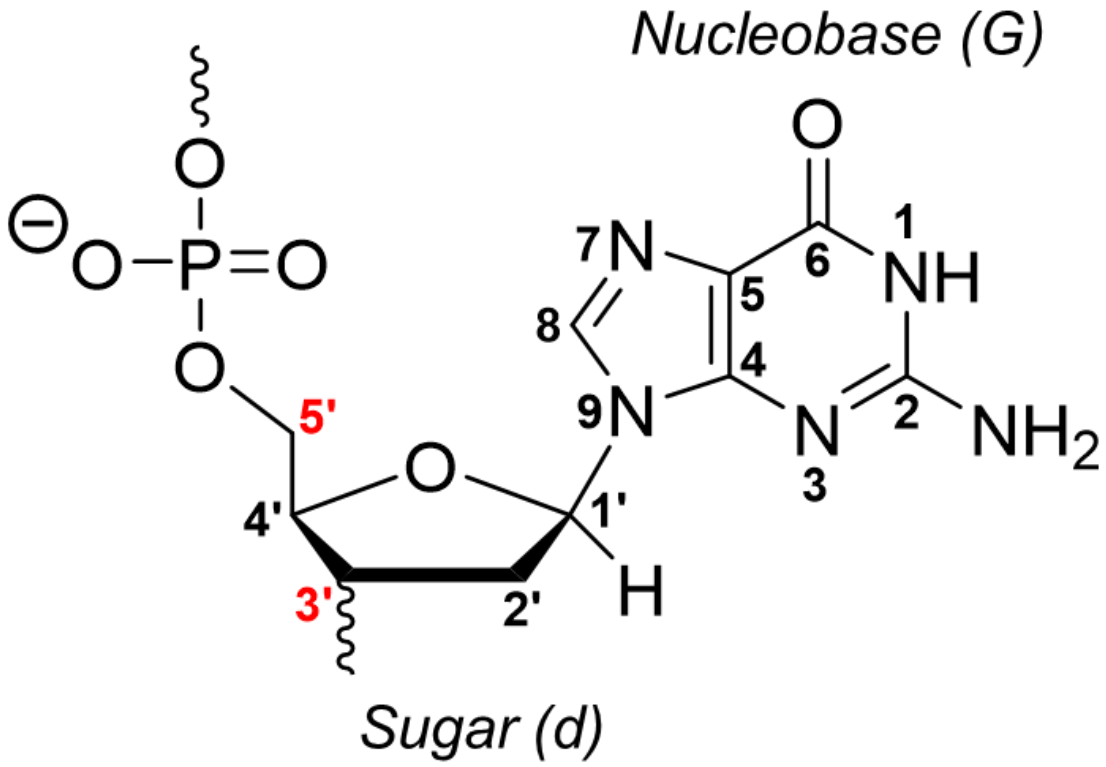


Figure 1 The guanine deoxynucleotide consists of a nucleobase (guanine or G) a sugar (deoxyribose or 'd') and a phosphate. Atom labels are shown, with 5' and 3' highlighted in red.

A nucleotide contains three structural elements: Nucleobase, sugar and phosphate. The phosphate serves as a connector between two sugar moieties. It also carries a negative charge, which is why oligonucleotides are polyanions. The sugar is deoxyribose (d) for DNA and ribose (r) for RNA (which has an -OH group in the 2' position). To differentiate sugar atoms from nucleobase atoms, the sugar atoms are labeled with an apostrophe. The glycosidic function is always 1' ("one prime"), the phosphate groups connect to the 3' and 5' atoms. By convention, oligonucleotide sequences are written from 5' to 3'. The sugar/phosphate chain is commonly referred to as the 'oligonucleotide backbone' or, more often, just 'backbone'. The backbone can be modified for analytical purposes (usually to enforce a desired conformation) or to increase environmental resistance (e.g. for oligo-therapeutics).[1,2] Nonetheless, backbone modifications are not within the scope of this thesis project.

The nucleobases are adenine, cytosine, guanine and thymine (for DNA) or uracil (for RNA). They carry the genetic information that is transcribed into the transcriptome and translated into the proteome. They are connected to the sugar through an N-C glycosidic bond and can associate with other nucleobases through non-covalent H-bonds. The most common base association *in vivo* is the G≡C and A=T base pairing, which is why it is referred to as "canonical base pairing" (Figure 2). Watson-Crick base pairing promotes the formation of the well-known DNA double helix.

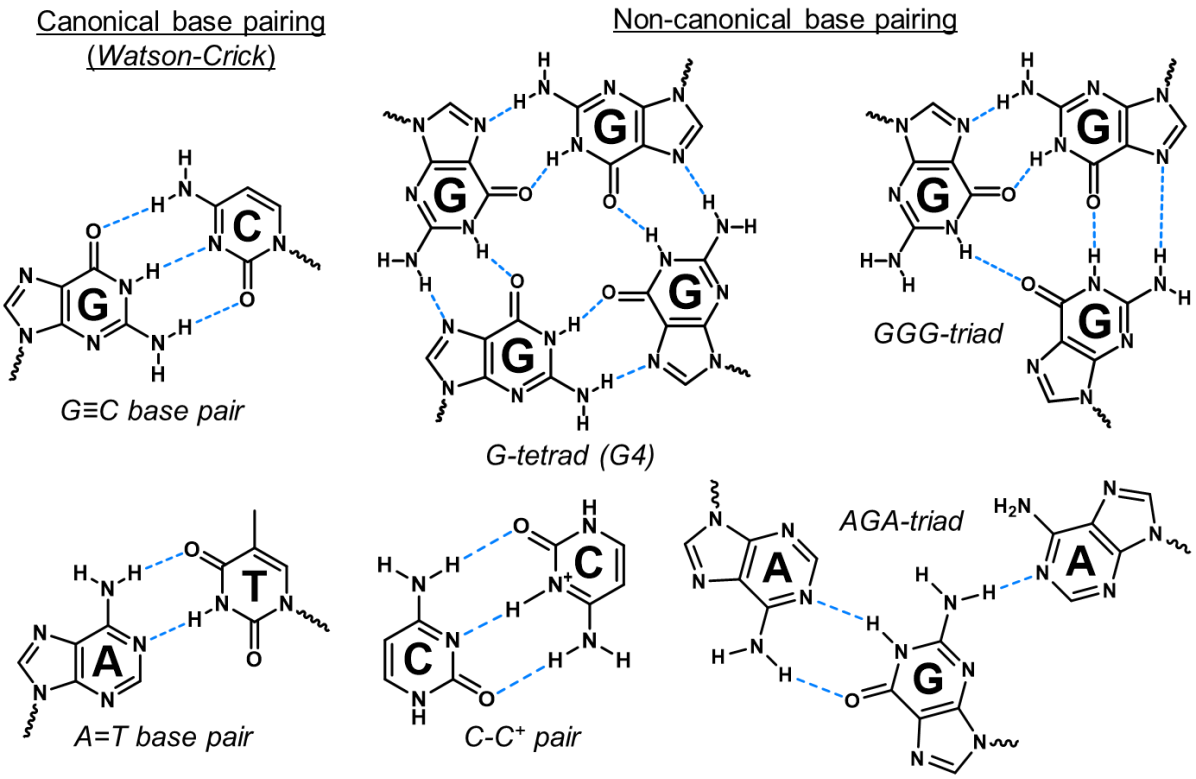


Figure 2 The canonical base pairing of Watson-Crick duplexes (left) and variations of non-canonical base pairing found in G-quadruplexes/i-motifs (right).

However, the duplex is not the only secondary oligonucleotide structure. In G-rich sequences, four guanines can associate with each other through Hoogsteen-type base pairing, leading to the formation of a G-tetrad. These G-tetrads can stack onto each other, creating a G-quadruplex (Figure 3). A cation sits between each pair of G-tetrads.

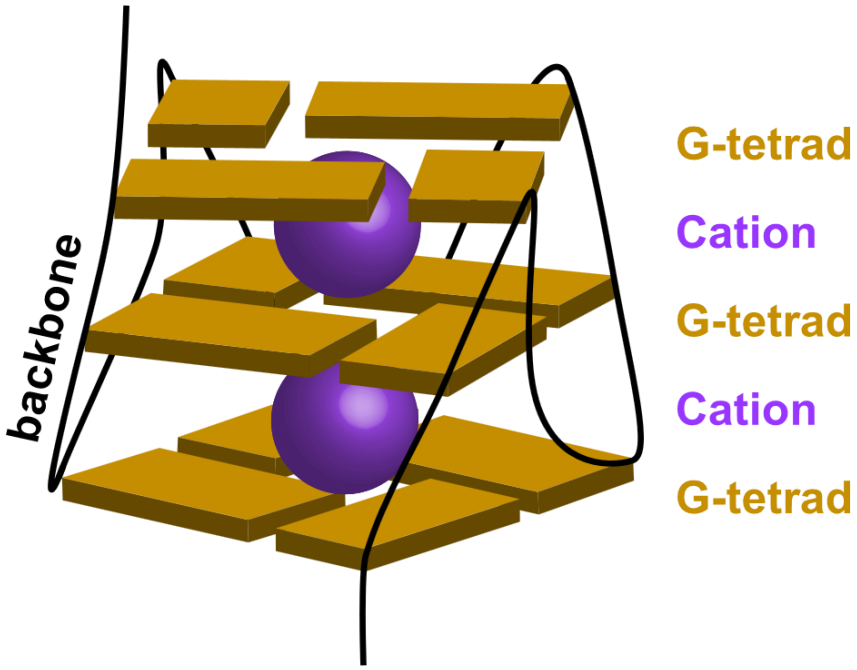


Figure 3 Simplified scheme of a G-quadruplex

Structural characteristics

Strand progression

G-quadruplexes are divided into three main topology classes: Parallel, antiparallel and hybrid. Each class is defined by their alignment of G-tracts. The G-tracts are the four strands that make up the quadruplex assembly. For example, the sequence 22AG (dA**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) forms a three-tetrad quadruplex in Na^+, whose G-tracts are four separate GGG repeats. In parallel topology, all G-tracts follow the same direction relative to one another. In antiparallel topology, the direction switches between each G-tract. Hybrid topology is a mix of the other two: three G-tracts face the same direction, one G-tract is opposed. A visual guide is provided in Figure 4.

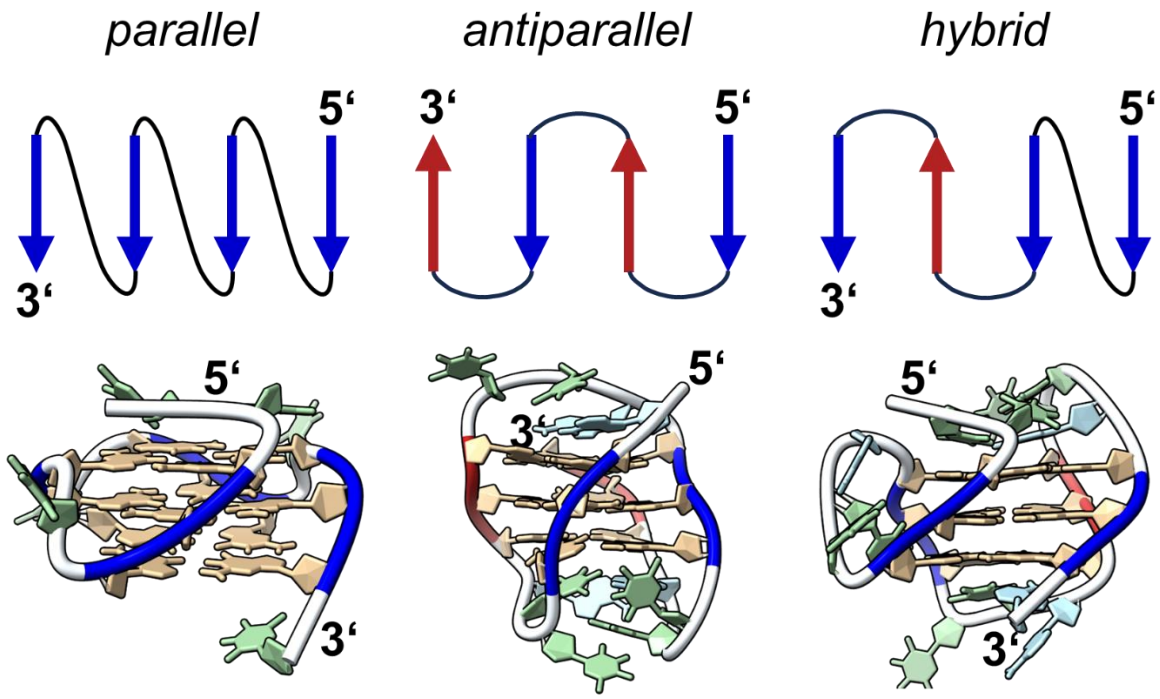


Figure 4 Strand progressions for the three main G-quadruplex topology classes with examples from resolved solution-phase structures shown below. PDB codes: 2LK7 (parallel),[3] 143D (antiparallel),[4] 2GKU (hybrid).[5] Guanines are tan, adenines are blue and thymines are green.

Strand progression is the standard method to assign G-quadruplex topology. Nonetheless, an important thing to keep in mind when discussing different topologies is that topology is not equivalent to secondary structure. Topology is just one of the many aspects of the secondary structure of a G-Quadruplex. Therefore, the topology alone is insufficient to describe the G-quadruplex structure and other structural features must be taken into account.

Loops

Loops are the connecting region between two G-tracts. Based on how they loop around the G-quadruplex core, they can be divided into four main types (Figure 5). A fifth type, the V-loop, will not be discussed since it is a very rare form of snapback occurring mostly in hybrid-2 G-quadruplexes.[6,7]

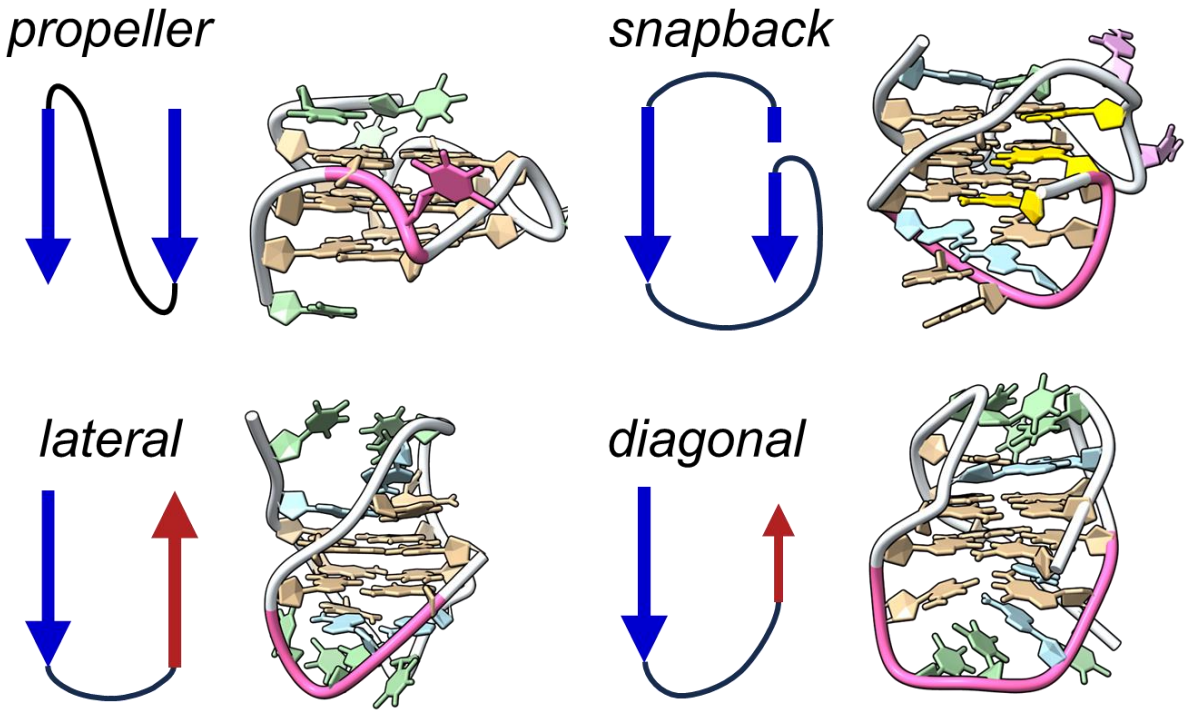


Figure 5 The four main types of G-quadruplex loops with real examples, where the loop is highlighted in pink. PDB codes: 2LK7 (propeller),[3] 2O3M (snapback),[8] 5LQG (lateral and diagonal).[9] Guanines are tan, adenines are blue and thymines are green. The G-tract split in two by the snapback loop is highlighted in gold.

Propeller loops are typically short, containing 1-2 nucleotides. Propeller loops connect two G-tracts that are aligned in the same direction, which is why most parallel G-quadruplexes have three propeller loops. When the G-tract of a parallel G-quadruplex is truncated, it might flip back on itself to replace the missing G with a G from elsewhere in the sequence. This leads to the formation of a snapback loop. They are most likely to form in G-rich sequences with short loops and inconsistent G-tract length.[10]

Longer loops (> 3 nucleotides) are typically lateral and diagonal loops. Since they change the direction of the G-tract, they are not found in parallel G-quadruplexes. Antiparallel G-quadruplexes have two defined sub-types based on the loop sequence: 1) Lateral-lateral-lateral ('chair-type') and 2) lateral-diagonal-lateral ('basket-type'). Hybrid G-quadruplexes are also divided into two main sub-types: 1) propeller-lateral-lateral ('hybrid-1') and 2) lateral-lateral-propeller ('hybrid-2'). The majority, but not all antiparallel/hybrid G-quadruplexes can be assigned to one of these sub-classes.

Loops can significantly contribute to the non-covalent G-quadruplex structure. Nucleotides in the loops can form AT base pairs or GGG/AGA triads that stack on top of the G-tetrads.[11,12] Long loops can form intramolecular Watson-Crick base pairs (a so-called 'hairpin'), creating what is called a duplex-quadruplex junction.[13] Nonetheless, the G-quadruplex stability decreases with loop length,[14] so the majority of known G-quadruplexes have loops at a length of 1-5 nucleotides.

Flanking nucleotides

While loops are located between G-tracts, the flanking nucleotides are what comes before and after the G-quadruplex forming strand. They may seem insignificant, but their role in G-quadruplex formation is detrimental in many cases. A 5' flanking thymine can template the first guanosine into a configuration that promotes parallel topology.[15] The introduction of nucleotides at the 5' end also prevents G-quadruplex dimerization

by covering the preferred stacking interface.[3] Flanking adenines can interact with nearby loops and contribute to the formation of base triads.[16]

In a genomic context, the flanking region spans beyond hundreds of nucleotides. Higher-order effects come into play, for example the supercoiling of DNA. One hypothesis is that the formation of G-quadruplexes relieves superhelical stress. To test this hypothesis, G-quadruplex DNA was injected into negatively supercoiled plasmids. Superhelical stress can stabilize a G-quadruplex in physiological condition,[17] but not induce G-quadruplex formation in an unfolded strand.[18]

Grooves

A G-quadruplex has four grooves, each running between a pair of G-tracts. Depending on their width, they are classified as narrow, medium or wide. As the width depends on the arrangement of the sugar backbone, it is defined as the distance between two phosphorus atoms of neighboring dG nucleotides of the same G-tetrad. A narrow groove measures around 7-10 Å, a medium groove 15-17 Å and a wide groove 20-23 Å (Figure 6).[19]

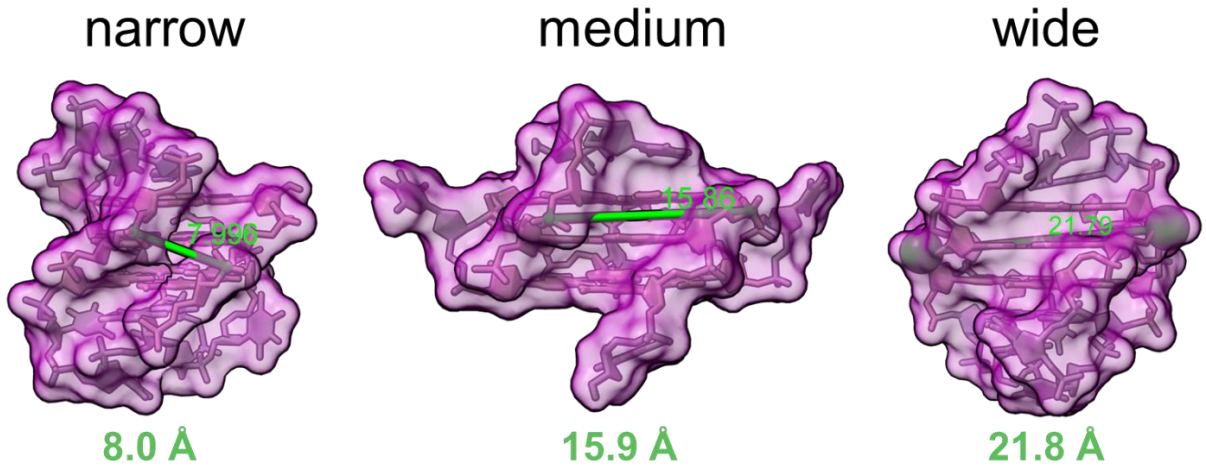


Figure 6 Examples of a narrow groove (PDB: 5YEY), a medium groove (PDB: 2LK7) and a wide groove (PDB:5YEY) with the distance measurement for the central G-tetrad showcased in lime.

Which grooves a G-quadruplex has is not up to chance. All topologies have their characteristic groove patterns. Parallel G-quadruplexes have four medium grooves. Antiparallel 'chair' type G-quadruplexes have two opposing wide and narrow grooves (there can never be two narrow or two wide grooves next to each other). The other topologies have one narrow, two medium and one wide groove(s). Each of them has their characteristic sequence of grooves from 5' to 3', which are: wide-medium-narrow-medium (antiparallel 'basket'), medium-wide-narrow-medium (hybrid-1) and wide-narrow-medium-medium (hybrid-2), respectively.

Glycosidic bond angle and base stacking

The glycosidic bond between the sugar moiety and the nucleobase can rotate, but tends to be locked around a certain angle due to intramolecular forces, such as G-tetrad formation, in the case of G-quadruplexes. When the glycosidic sugar proton (H1') and aromatic nucleobase proton (H8 for guanine) face in the same direction, the nucleotide is in *syn*-configuration. When they face in opposite direction, the nucleotide is in *anti*-configuration (Figure 7).

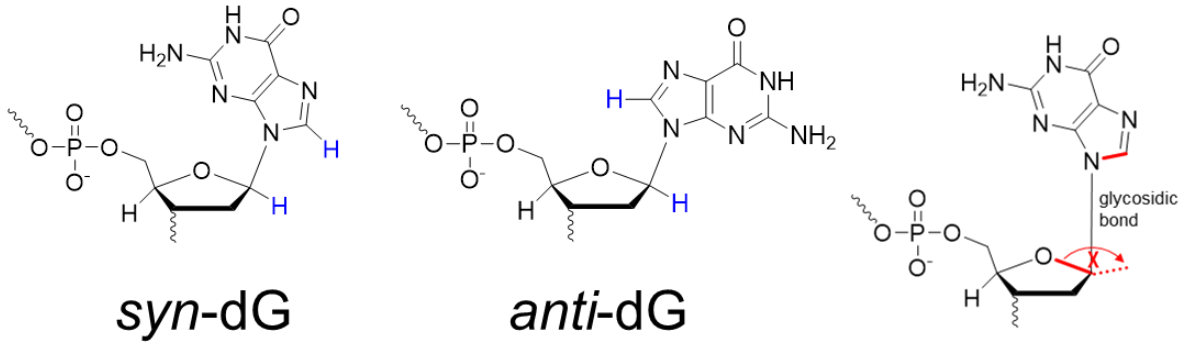


Figure 7 Desoxyguanosine in syn and anti-configuration. The H1' and H8 protons are highlighted in blue. The glycosidic bond angle χ is measured relative to the bonds shown in red.

Parallel G-quadruplexes only have *anti*-guanines. Antiparallel G-quadruplexes usually switch back and forth between *syn*- and *anti-guanines* as the strand progresses. In hybrid G-quadruplexes, three G-tracts have the same *syn*/*anti*/*anti* sequence, while the fourth G-tract has the inverse *anti*/*syn*/*syn* sequence (Figure 8).

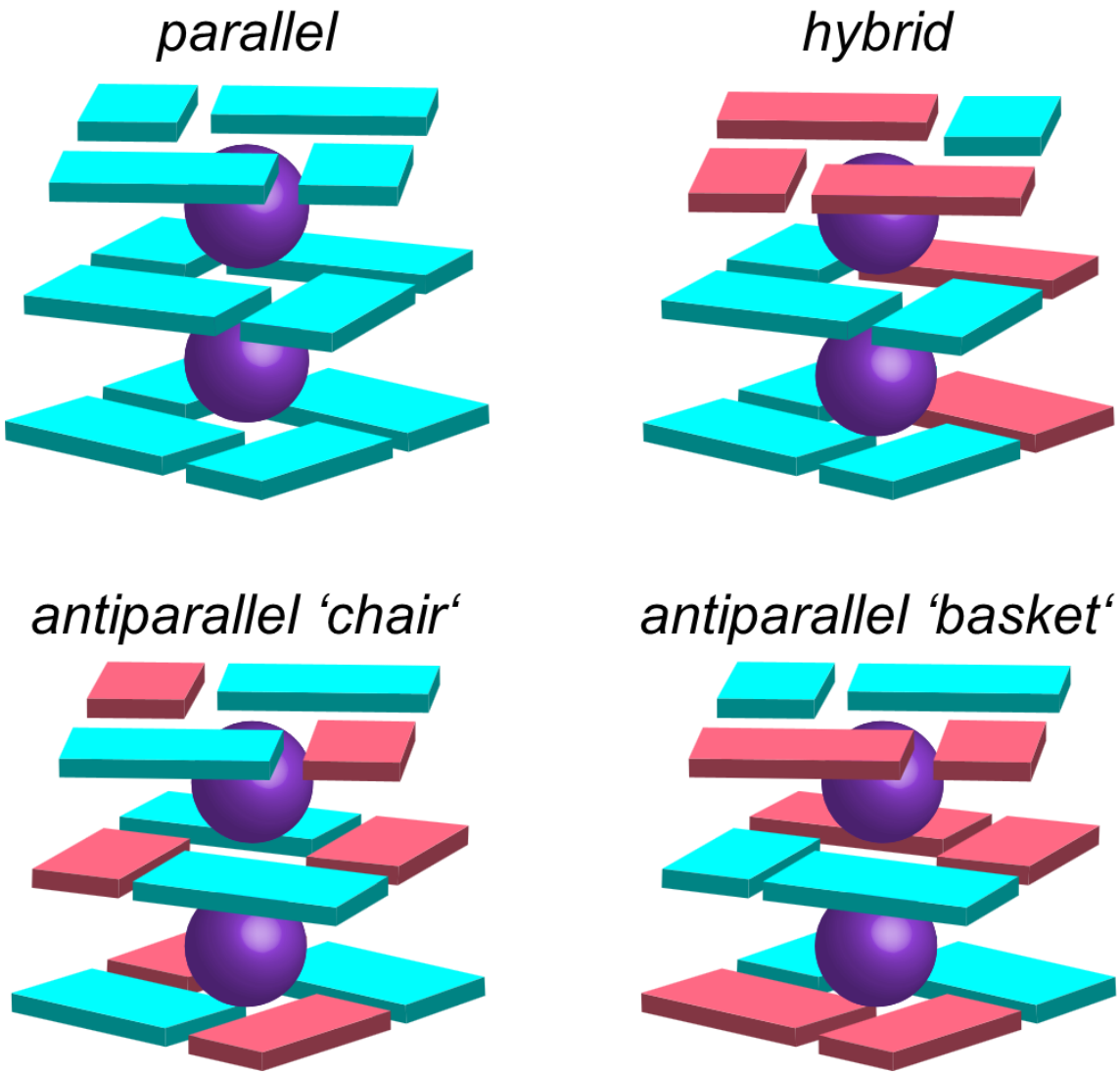


Figure 8 Standard base stacking patterns for G-quadruplexes based on their topology class, independent of strand orientation or loop patterns. syn-guanines in hot pink, anti-guanines in cyan, K^+ in purple.

The vast majority of G-quadruplexes follow a base stacking pattern like shown in Figure 8. The oligonucleotide strand has different ways to move along these scaffolds, which lets us divide G-quadruplexes into more sub-groups. Webba da Silva proposed a system classifying G-quadruplexes into categories I to VIII, according to base stacking and strand evolution.[20]

When two G-tetrads stacked on top of each other have the same *syn*/*anti* base pattern, it is called homo-stacking. When the *syn*/*anti* pattern is inversed, it is called hetero-stacking. Homo-stacking is characteristic for parallel G-quadruplexes, hetero-stacking is characteristic for antiparallel G-quadruplexes, while hybrid G-quadruplexes tend to have a mix of both.

Homo-stacking is typical for parallel G-quadruplexes, but why is it only *anti*-guanines? Why not have four *syn*-guanines instead, for example? *Ab initio* calculations showed that only *syn*→*anti* and *anti*→*anti* base stacks have favorable free energies, whereas *anti*→*syn* and *syn*→*syn* base stacks have unfavorable energy contributions (Figure 9).[21]

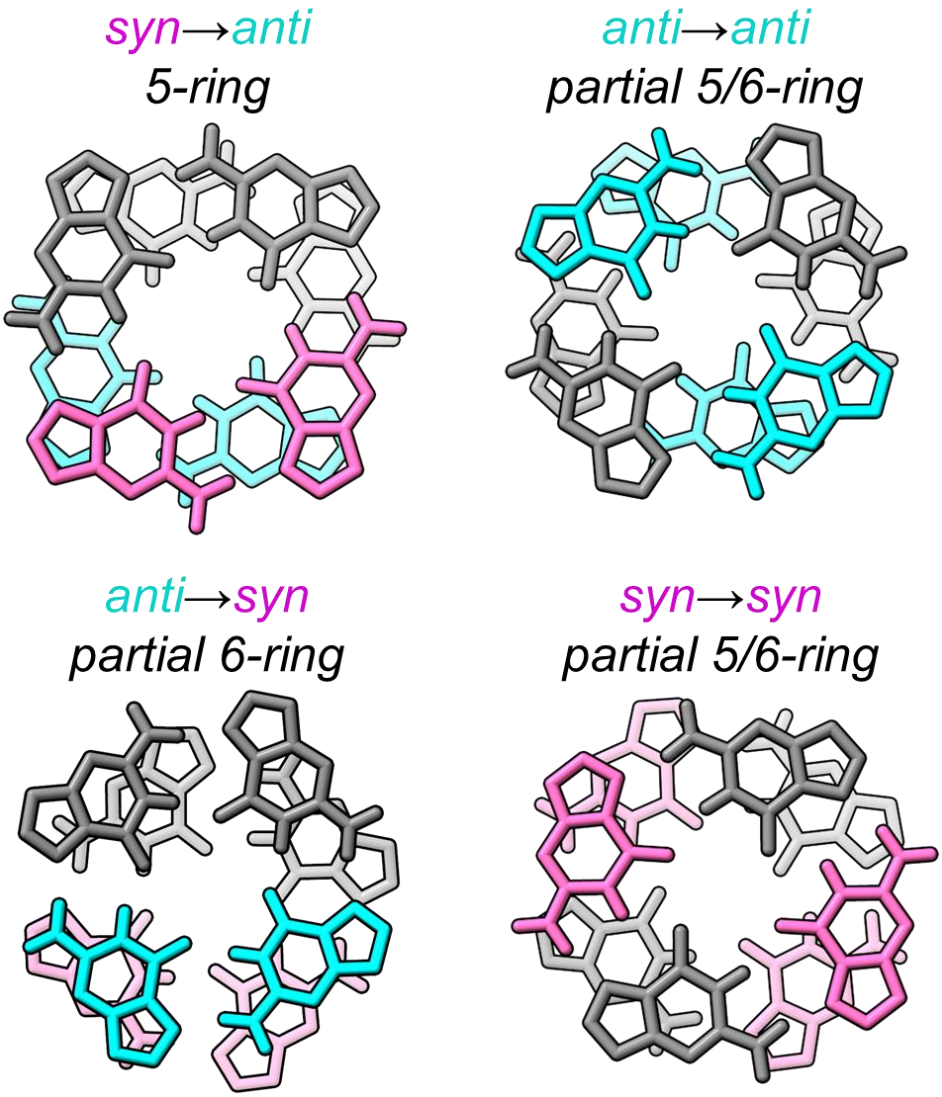


Figure 9 The four combinations of base stacking in G-quadruplexes (5'→3' direction). Strands pointing in the opposite direction are greyed out. PDB codes: 5LQG (syn→anti),[9] 143D (anti→syn),[4] 5YEY (anti→anti and syn→syn).[22] The base stacking pattern and ring overlap are two different terms found in the literature, but they refer to the same phenomenon.

Therefore, G-quadruplexes are thermodynamically driven to maximize the amount of *syn→anti* and *anti→anti* base stacks, while minimizing the other two, which is why we have the four main G-tetrad stacking patterns shown in Figure 8.

H-bond rotation

The H-bonds in the G-tetrad can rotate clockwise or anti-clockwise. We can express the H-bond rotation with the indices (+) and (-). Perspective is key here, so we assign rotation patterns as follows: We look down onto the G-tetrad stacks, with the 5' end in the back. We then assess the rotation pattern of the G-tetrad that is the furthest in the back and advance to the front from there (Figure 10).

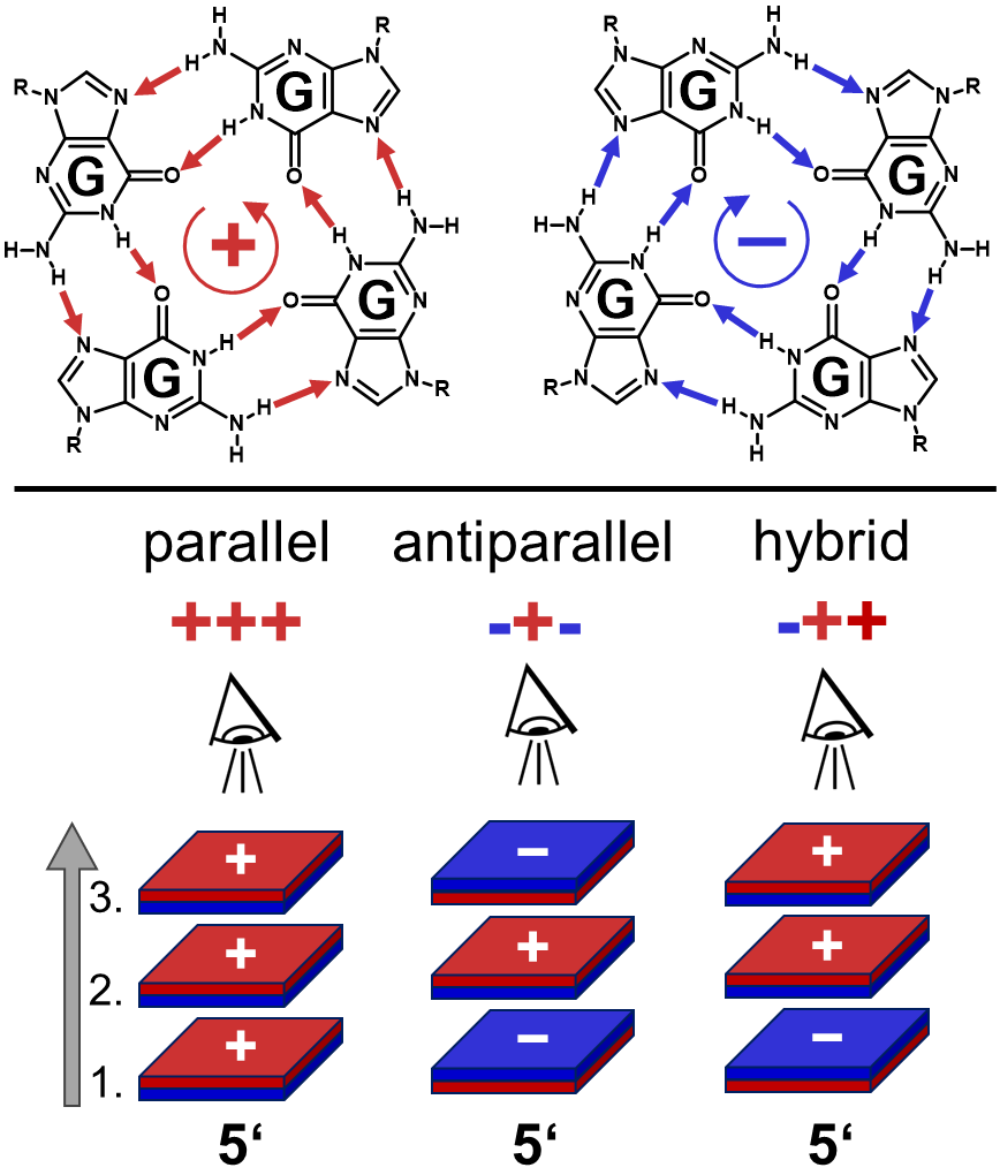


Figure 10 ***Top****: Positive and negative H-bond rotation pattern.* ***Bottom****: Assignment of rotation patterns for the three main topology classes.*

If one were to look from the other side instead, +++ becomes ---, -+- becomes +-+ and -++ becomes --+. Switches in polarity are equivalent to hetero-stacking and preservation of polarity is equivalent to homo-stacking. Although H-bond rotation is directly caused by the glycosidic bond angle succession it nevertheless offers a different perspective on G-tetrad stacking.

Handedness

By default, G-quadruplexes are right-handed. Right-handedness for G-quadruplexes means that the G-tracts and the G-tetrads are turning clockwise as the strand progresses, which is illustrated in Figure 8 and can also be seen in Figure 9.

Left-handed G-quadruplexes can be created by synthesizing oligonucleotides with L-sugars instead of D-sugars.[23,24] Natural DNA sequences with 'GG' and 'G' tracts interrupted by one 'T' can also fold into a left-handed G-quadruplex, such as the sequence Z-G4 (dT(GGT)$_4$TG(TGG)$_3$TGTT).[25] One notable difference between left- and right-handed G-quadruplex is that in right-handed G-quadruplexes, loops and flanking nucleotides are facing more outwards, whereas for left-handed G-quadruplexes, loops and flanking nucleotides are more bent inwards.[26]

Molecularity

Molecularity describes the number of oligonucleotide strands within a G-quadruplex assembly. G-quadruplexes with one strand are called intramolecular. Intramolecular G-quadruplexes can still stack onto each other, forming a dimer. G-quadruplexes prefer to dimerize at the 5' to 5' interface.[27] Parallel G-quadruplexes are the most likely to dimerize, since their G-tetrads are not covered by loops. G-quadruplexes can keep stacking onto each other, forming higher-order aggregates.[28] The human telomeric sequence contains a high density of G-quadruplex forming regions, so there is a possibility for the formation of multimeric G-quadruplexes in vivo.[29] G-quadruplex multimers, stabilized by chelating metal cations, are being utilized in nanotechnology as 'G-nanowires'.[30]

Multimolecular G-quadruplexes are formed by 2-4 oligonucleotide strands (Figure 11).

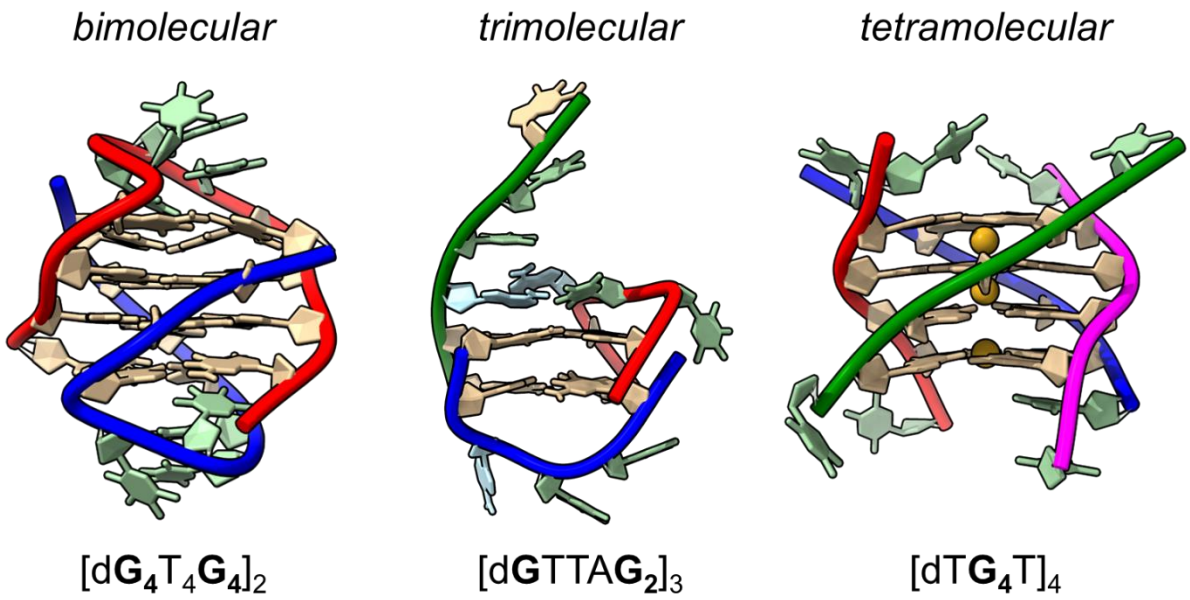


Figure 11 Left: Solution-phase structure of [dG$_4$T$_4$G$_4$]$_2$ in 50 mM NaCl, PDB: 156D.[31] Center: Solution-phase structure of [dGTTAG$_2$]$_3$ in 33 mM Na$_3$PO$_4$ (pH 6.8), PDB: 6M05.[32] Right: Crystal structure of [dTG$_4$T]$_4$ in 75 mM NaCl and 10 mM Tris-HCl (pH 7.4), PDB: 2O4F [33]. Guanines are tan, adenines are light blue and thymines are pale green.

Multimolecular G-quadruplexes are less dynamically constrained by their loops compared to intramolecular G-quadruplex and can thus form more dynamic assemblies. The *Oxytricha* telomeric repeat G4T4G4 is arguably the most researched

bimolecular G-quadruplex. It is an antiparallel four-tetrad G-quadruplex with two diagonal loops. Two diagonal loops are uncommon in intramolecular G-quadruplexes, but more common in the bimolecular assembly because 2 strands provide more conformational freedom.

TG4T itself is a stable and rigid tetramolecular G-quadruplex.[34] However, inserting more guanines enables the strands to 'slide' along the G-quadruplex core, creating a range of slipped-strand configurations.[35]

As a final example, the sequence $[dGCG_2AG_4AG_2]_2$ shows that multimolecular G-quadruplexes will not necessarily abide by the rules that apply for intramolecular G-quadruplexes (Figure 12). The core structure contains 2 G-tetrads as well as one AGGGGA hexad and a CGCG tetrad. The two strands mirror each other's direction in a very atypical pattern. It is tricky to even assign a topology although, if we go by strand orientation and base stacking pattern, hybrid topology would be the most sensible assignment.

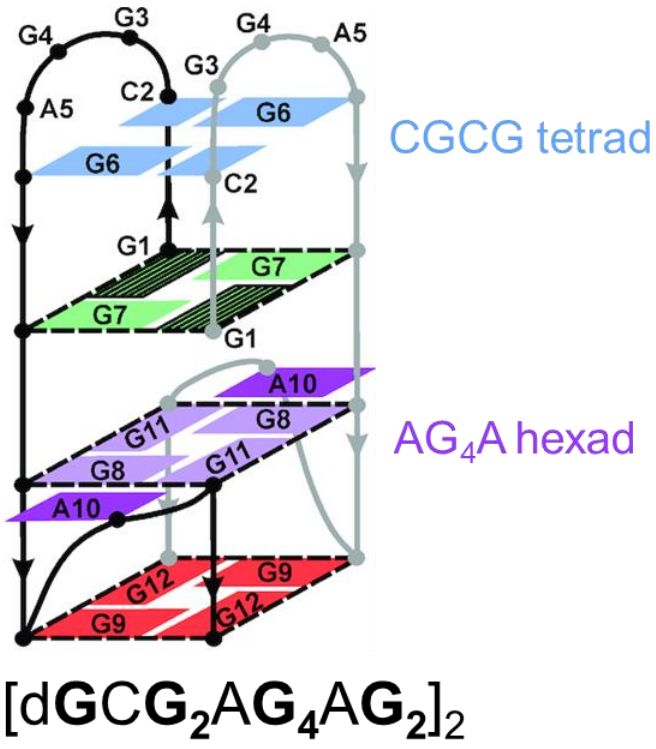


Figure 12 NMR-derived structure of $[dGCG_2AG_4AG_2]_2$ in 175 mM KCl and 10 mM K_3PO_4 (pH 6.8), PDB: 6SX6. Reproduced from Šket et al. NAR 2020, 48 (5), 2749-2761.[36] CC-BY-NC creative commons.

Cations

One cation between each pair of G-tetrads is essential for G-quadruplex formation. The energetic driving forces behind this are the 1) cation-π interaction between cation and G-tetrads 2) entropic release of solvent molecules from the cavity between the G-tetrads and 3) reduction of charge density by balancing the negative charges at the phosphate residues.[37,38] Only main-group cations with the right size can intercalate between two G-tetrads (Table 1). Transition metals initiate covalent interactions with nucleobases. Metal ions that target guanine, such as Ag^+, counteract the K^+-induced formation of G-quadruplexes by shifting the equilibrium towards other higher-order structures such as G·G mismatches.[39–41]

Table 1 Main-group cations and their effect on G-quadruplex formation.

Cation	Ionic radius [Å] (Shannon[42], coordination VIII)	Promotes G-quadruplex formation?
Li^+	0.92	No[43–46]
Na^+	1.18	Yes[43,44,46–48]
K^+	1.51	Yes[37,48,49]
Rb^+	1.61	Yes[43–46]
Cs^+	1.74	No[43–46]
Mg^{2+}	0.89	No[44,45,50–54]
Ca^{2+}	1.12	Not always[44–46,50–55]
Sr^{2+}	1.26	Yes[44,50,56,57]
Ba^{2+}	1.42	Yes[44,50]
Tl^+	1.59	Yes[58–60]
Pb^{2+}	1.29	Yes[61–63]
NH_4^+	1.54[64]	Yes[35,65–74]

We can derive from Table 1 that the 'sweet spot' for cation intercalation lies between 1.2-1.6 Å. The biologically relevant cations for G-quadruplex formation are Na^+ and K^+ in both intracellular (148 mM K^+, 8 mM Na^+) and extracellular (5 mM K^+, 144 mM Na^+) conditions.[75] Because G-quadruplex formation occurs inside the cell (for the most part), the default cation is K^+. K^+ ions form very stable G-quadruplexes, with K_D values in the µM to mM range,[76] so that K^+ concentrations as little as 1 mM can be enough to form a stable G-quadruplex.[77] Cation competition experiments shows that K^+ displaces most other cations from pre-folded G-quadruplexes e.g. sodium[46,47] or ammonium[78]. Only divalent cations of similar size such as Sr^{2+} and Pb^{2+} match with K^+ in terms of binding affinity.[44,56,79] Figure 13 shows the melting curves of the two same G-quadruplex sequences in 100 mM of various cations. K^+ clearly forms one of the most stable G-quadruplexes.

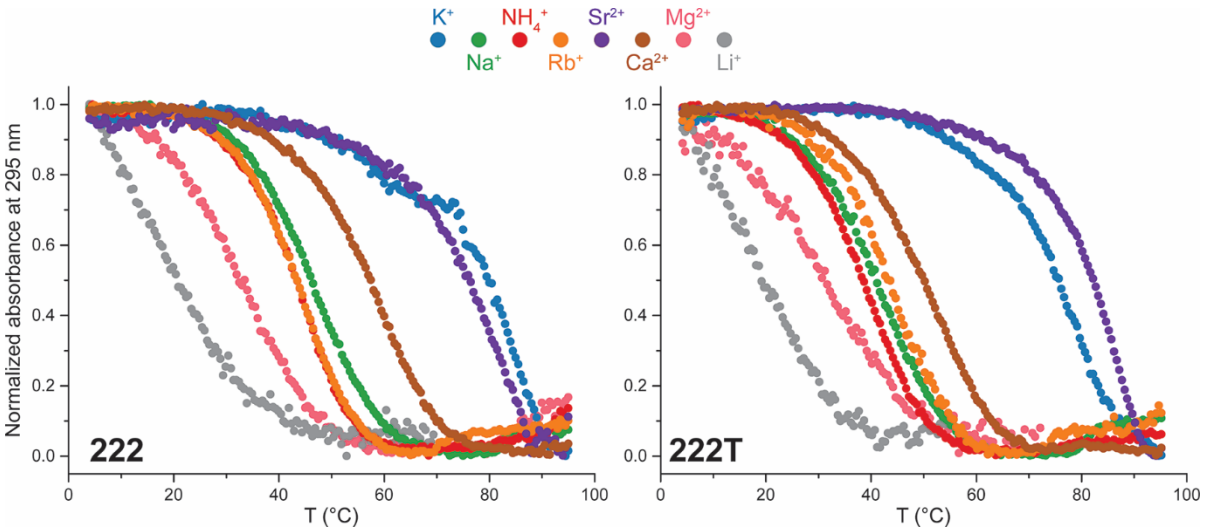


*Figure 13 The folded fraction of 222 (**G₃TTG₃TTG₃TTG₃**) and 222T (**TG₃TTG₃TTG₃TTG₃T**) as a function of temperature for different cations, determined by UV melting. Samples contain 10 µM DNA, 100 mM cation and 20 mM lithium cacodylate (pH 7.2). Reprinted with permission from Largy et al. JACS 2018, 138 (8), 2780-2792.[46] Copyright 2016 American Chemical Society.*

K^+ and Na^+ favor different topologies. 3-tetrad intramolecular G-quadruplexes in K^+ are mostly parallel or hybrid, while 2- and 4-tetrad G-quadruplexes are mostly

antiparallel.[77] Na^+ promotes antiparallel topology − sequences that assume parallel/hybrid topology in K^+ will often be antiparallel in Na^+.[57,80,81] While we still cannot fully explain why K^+ and Na^+ promote different G-quadruplex folds, we can point out differences between their binding mechanisms that can help us obtain a better understanding. First, Na^+ ions are much smaller and can fit inside the center of a G-tetrad (check out the TG4T structure in Figure 11 for reference). Meanwhile, K^+ ions are significantly larger and always located inbetween, but never inside G-tetrads.[37,48] Thermodynamic studies revealed that Na^+ induced G-quadruplex formation is enthalpy-driven, while K^+ induced G-quadruplex formation is entropy-driven.[82,83] That could be a contributing factor for why G-quadruplexes in K^+ are more stable at high temperature compared to Na^+, but gives no concrete information on topology preference.

NH_4^+ is not biologically relevant, but frequently used in mass spectrometry because it is volatile and can therefore be added without limitation. NH_4^+ is chemically similar to K^+ but will form less stable G-quadruplexes (Figure 13) that are more labile in the gas phase as they pass through the mass spectrometer due to neutral loss of NH_3.[65,66]

G-quadruplex vs. i-motif

When a G-quadruplex forms on a G-rich strand, an i-motif can form on the opposing C-rich strand. The core structure of the i-motif is not a stack of G-tetrads, but a stack of CC^+ base pairs, where one cytosine is protonated (Figure 14).

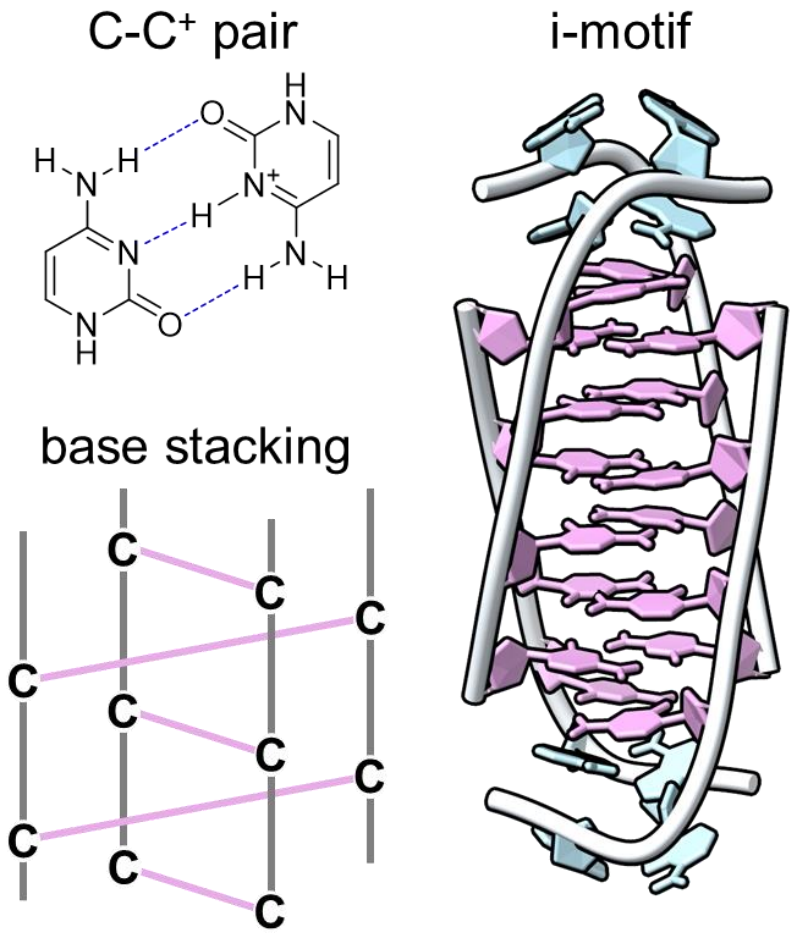


Figure 14 Overview showing a CC^+ base pair, the criss-cross stacking pattern of CC^+ pairs and the NMR structure of $(dAACCCC)_4$ at pH 4.5 (PDB: 1YBL).[84] Cytosines are pink, adenines are blue.

i-motifs require cytosine protonation and therefore need to be stabilized at slightly acidic pH (5-7) for *in vitro* analysis. The longer the C-tracts, the more stable the i-motif at neutral pH.[85] i-motifs are stable at neutral pH *in vivo,*[86] due to several factors that are not accounted for *in vitro* such as 1) Superhelical stress[17] 2) Molecular crowding[87] 3) Mg^{2+} as a potential co-factor in i-motif folding[87] 4) Stabilization through duplex-junction.[88]

G-quadruplex ligands

Due to their involvement in gene expression, G-quadruplexes are attractive drug targets. The first generation of G-quadruplex ligands was derived from duplex-intercalating drugs (Figure 15). As a consequence, those ligands were not able to selectively target G-quadruplex over duplex-DNA in a cellular environment.[89,90] Newer G-quadruplex ligands resolved this issue with aromatic systems designed to match the size of a G-tetrad, rather than a GC base pair. The difference in affinity between duplex and G-quadruplex is several orders of magnitude for PhenDC3,[91] leading to its application as a biomarker to detect G-quadruplexes in cells.[92] Cationic sidechains are often added to prevent charge repulsion from DNA polyanions and improve water solubility.[93]

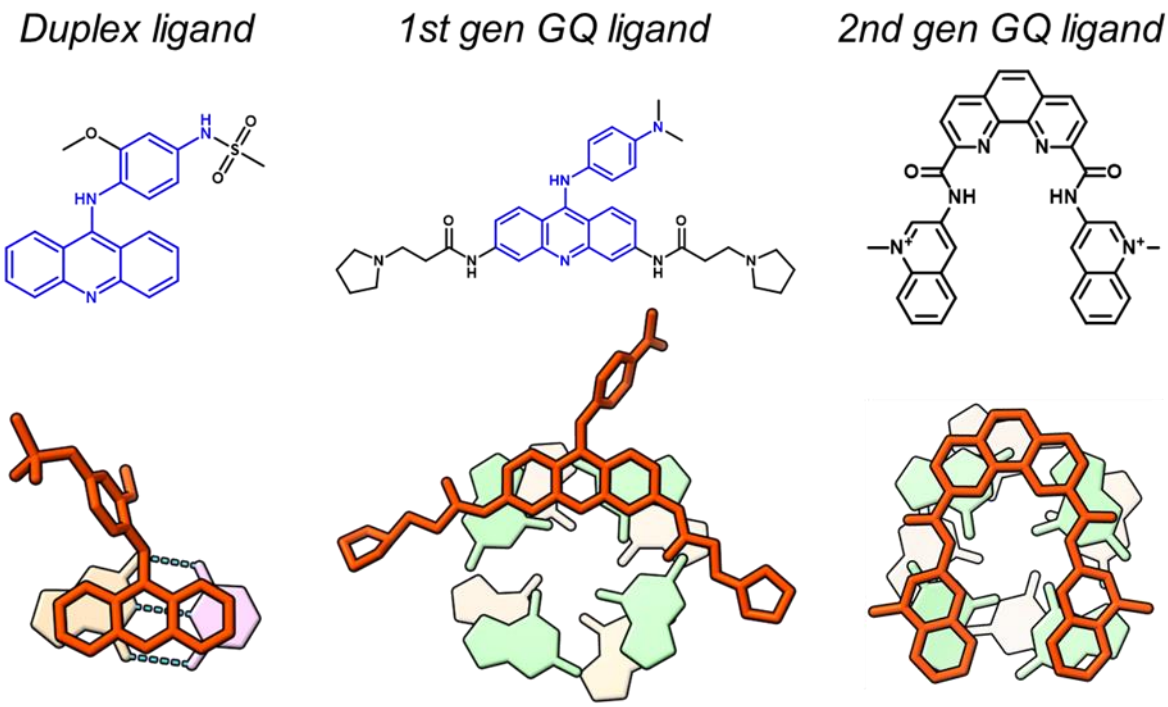


Figure 15 Evolution of G-quadruplex ligands. ***Left****: Duplex binding drug Amsacrine stacking onto a GC base pair (PDB: 4G0U).*[94] ***Center****: First-generation G-quadruplex ligand BRACO-19 stacked onto a G-quadruplex (PDB: 3CE5).*[95] ***Right****: Second-generation G-quadruplex ligand PhenDC3. The size of the aromatic system is optimized to stack onto a G-tetrad (PDB: 2MGN).*[96]

Newly introduced G-quadruplex ligands usually have two characteristic features: 1) A flat, heteroaromatic core structure that enables π-stacking onto the G-tetrads 2) Extended, flexible sidechains to target sequence-dependent structures, especially the grooves and the loop regions. Several classes of G-quadruplex ligands were discovered from this, including bisquinolinium-dicarboxyamides,[91,97–102] acridines,[103–106] porphyrins[107–110] and diimides.[111–114] Although out of scope for this thesis project, fluorescence is one of the main established methods to probe G-quadruplex folding and ligand targeting. Beyond generic fluorescence dyes such as thiazole orange or

thioflavin T, several ligands have been developed as G-quadruplex specific fluorescent probes.[115–117] Some notable G-quadruplex ligands are introduced in Figure 16.

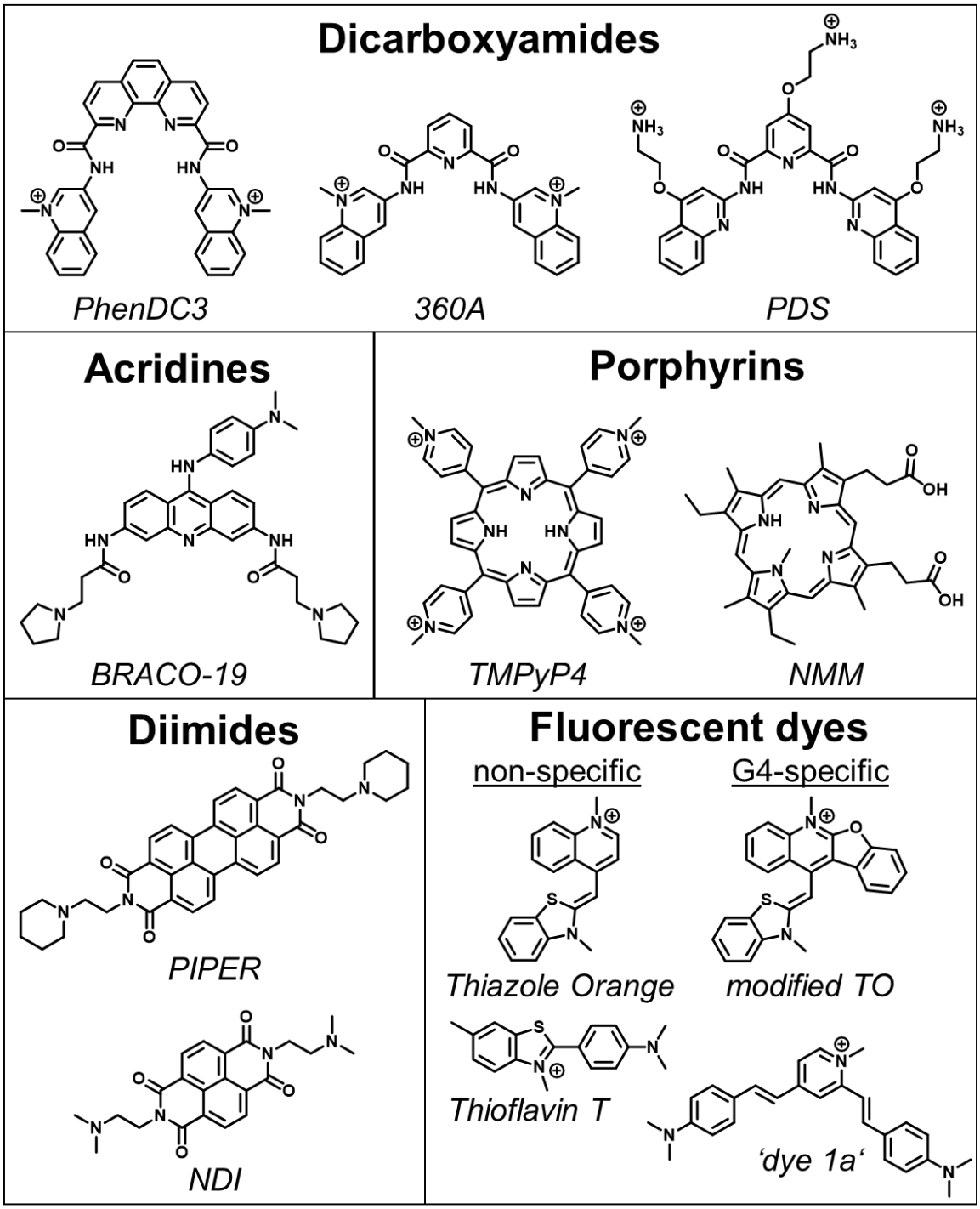


Figure 16 Introducing some of the most prominent G-quadruplex ligands. Characteristic features are 1) A flat, heteroaromatic core structure 2) Flexible sidechains at the perimeter 3) Positive charges.

Nonetheless, current ligands still lack the necessary selectivity among different G-quadruplexes for clinical use.[118,119] Ligands discriminating certain topologies[97,120,121] or topology-subclasses[122–124] have already been reported but rational design efforts have not yet yielded small molecules with significant selectivity for a specific G-quadruplex structure.[125] Chapter 3 introduces a new ligand class that deviates from the condensed aromatic structure paradigm, leading to improved topology selectivity.

Biophysical methods to study G-quadruplexes

Mass spectrometry (MS)

Native MS

Native mass spectrometry is all about preserving the analyte's non-covalently folded structure, i.e. its *native* structure, during analysis. In a mass spectrometer, the ion source is the highest risk factor for denaturing a biomolecule, because the analyte has to be brought into the gas phase, which requires the application of external force. Some ionization methods are harsh enough to break covalent bonds, others are soft enough to preserve non-covalent structures. The softest ionization method is electrospray ionization (ESI), for which a typical native MS workflow is depicted in Figure 17. Matrix-assisted laser desorption ionization (MALDI) is another ionization method that is compatible with native MS, however it is less common and out of scope for this thesis project.[126,127]

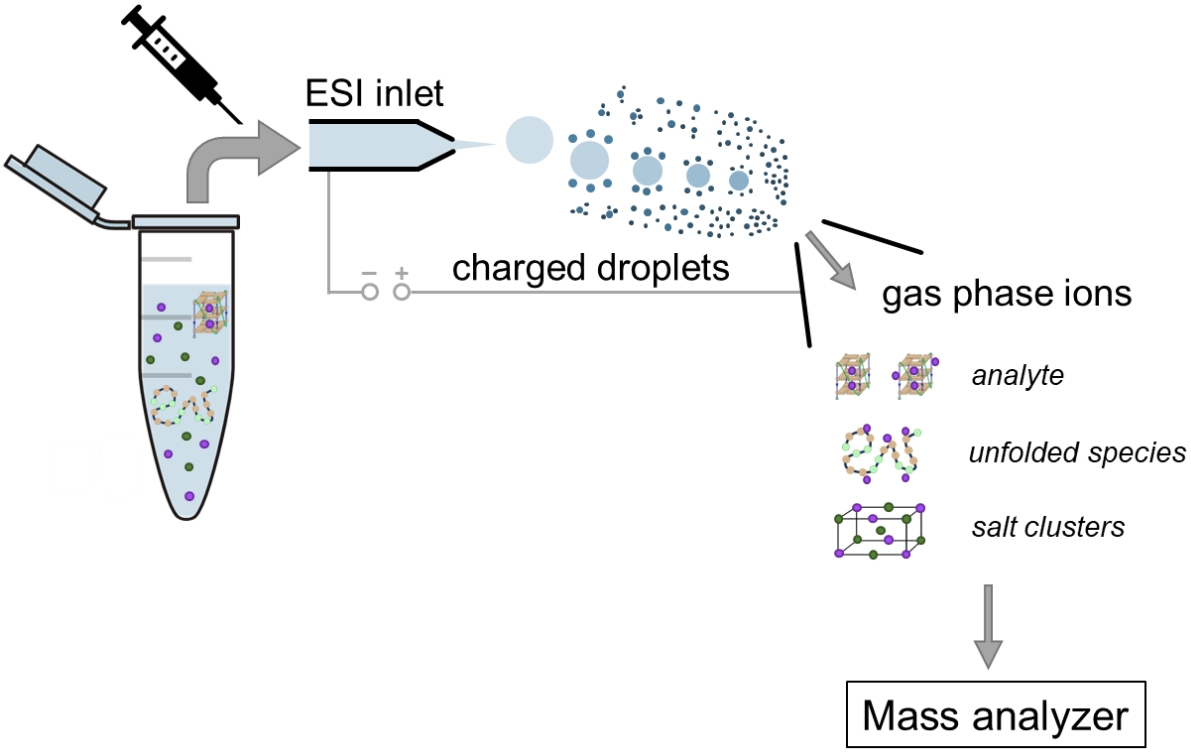


Figure 17 Native MS workflow for the study of G-quadruplexes using ESI-MS in negative mode. K^+ ions are purple, Cl^- ions are green.

To reduce the risk of denaturation, the samples are processed as little as possible and injected directly into the MS instrument. LC-MS coupling is usually incompatible with a native workflow due to the denaturing conditions inside the column (solvent, high temperature, additives). Double-stranded RNA has been preserved using reverse-phase LC with an ion pairing agent (IP-RP-LC), showing that native LC conditions can be achieved.[128,129] Nonetheless, we do not advise LC-MS coupling for native analysis. We ourselves do not work with *ex vivo* DNA, but with synthetic DNA sequences of interest, so a target separation/extraction is not required.

Our samples only contain what is needed to ensure native G-quadruplex folding. Note that in native conditions, the analyte exists in an ensemble of different conformational states, such as the folded and unfolded ensemble. Besides the analyte, everything in

the sample should be volatile. During droplet evaporation, non-volatile additives attach themselves electrostatically to the analyte (forming an adduct) or form salt clusters. High salt concentrations create a high ion current of something that is not the analyte, masking its detection. For optimal sensitivity, the major portion of the ion current reaching the detector should contain analyte ions. We therefore have to limit the quantity of non-volatile additives. Based on experience, we work with a limit of 1 mM salt concentration.[77] Ammonium acetate is volatile and can be thus added without restriction, making it a standard buffer for MS. Since NH_4^+ ions promote G-quadruplex formation as well, ammonium acetate is our go-to buffer for native MS of G-quadruplexes. When studying biologically relevant G-quadruplexes, we have to add at most 1 mM of K^+ along with trimethylammoniumacetate (TMAA) to obtain physiological ionic strength (ca. 100 mM). The $(CH_3)_3NH^+$ ion is too big to fit inside the G-quadruplex structure and therefore not interfering with G-quadruplex folding.

Ionic strength is not only important because of physiological concentration, but because of its effect on desolvation during the electrospray process. Despite the simple practical approach, the electrospray process is so complex in nature that it has yet to be fully understood. What we do know is that when the solution exits the capillary, the electric field pulls the liquid into a cone shape ('Taylor cone'). From the tip of the cone extends a filament that breaks up into the first set of droplets. An auxiliary gas flow helps break up the filament. The charged droplets are pulled towards the gas-phase inlet. Their polarity depends on the polarity of the electric field; for oligonucleotides we operate in 'negative mode', i.e. the droplets carry a net negative charge. On their way to the inlet, the droplets evaporate. A drying gas flow assists in droplet evaporation. The evaporation keeps going until the charge density hits a critical limit ('Rayleigh-limit'), where a series of small droplets ('offspring droplets') burst from the main droplet ('bulk'). The distribution of charges is asymmetrical and charges will accumulate in the offspring droplets, while the bulk droplets lose charges. As a result, any analyte inside the bulk droplets has its charges neutralized, making it undetectable. Response factors in ESI-MS heavily depend on how many analyte molecules end up in the offspring droplets and how many are left in the bulk solution. Sadly, we have little to no control over how much analyte migrates into the offspring droplets. The consecutive cycle of evaporation and charge explosion continues until the droplets have shrunk to a size that is close to the analyte molecule. At this point, the droplet only contains one analyte molecule which is about to transition from solution phase to gas phase. Several mechanisms can occur and the ionic strength of the solution is a deciding factor whether the analyte remains folded or becomes denatured (Figure 18).[130–133]

The Ion Evaporation Mechanism (IEM) only occurs for small, rigid ions and is therefore not relevant for biomolecules. It involves the ion migrating to the surface of the droplet, from where it breaches through the droplet surface and enters the gas phase. The energetic barrier for ion ejection is the surface tension of the solvent, while the driving force is charge repulsion. The Charged Residue Mechanism (CRM) is the desired mechanism for native MS. In this scenario, the analyte remains in the center of the droplet, while the solvent and all volatile compounds evaporate. The remaining charges and non-volatile ions are passed on to the analyte, which is now left in the gas phase. The gas-phase charge state has no correlation to the solution-phase charge state, but rather depends on the size of the final droplet, so it scales with molecule size.[130,131]

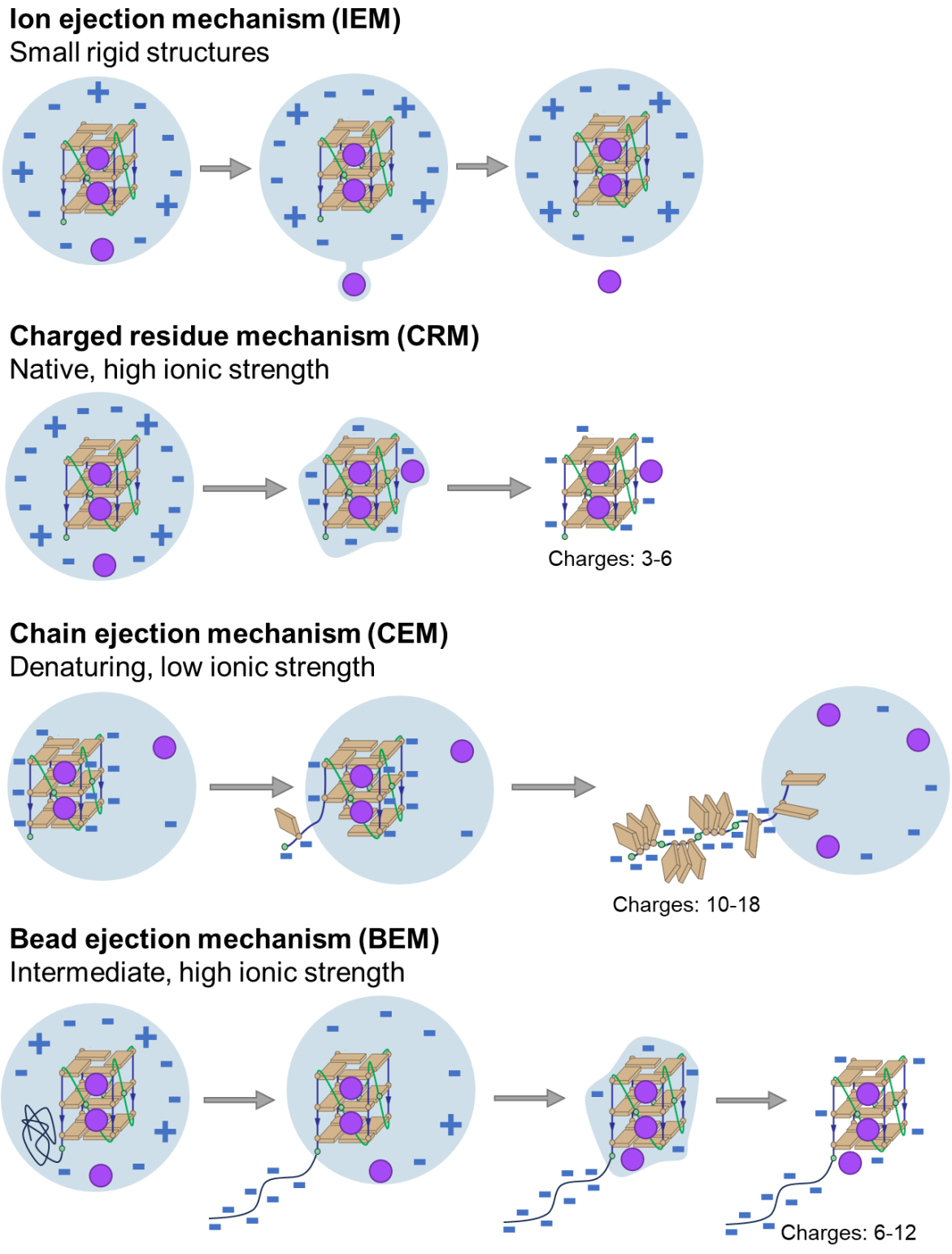


Figure 18 The four known electrospray mechanisms of ion transfer from solution into gas-phase. CRM is desirable for native analysis, CEM is desirable for top-down analysis. BEM is an intermediate mechanism for biomolecules that contain unstructured units. Charge states shown are in reference to a 24-mer oligonucleotide.

In low ionic strength, there are not enough charges spread across the droplet surface, so the analyte ion is drawn to the surface to add its own charges. The Chain Ejection Mechanism (CEM) is more or less a variation of the Ion Ejection Mechanism, but for big and flexible molecules. Both mechanisms are driven by charge repulsion, but while

the IEM is driven by intermolecular repulsion, the CEM is driven by intramolecular repulsion, caused by a high charge density within the analyte ion. That is because in low ionic strength, there are not enough counterions to neutralize excess charges. The CEM starts with one of the molecule termini (The 5' or 3' end for G-quadruplexes) being ejected from the droplet. When a terminus is ejected, it drags the nearby strand with it, which then drags out more of the strand etc. until the entire molecule has been ejected as an elongated strand. In the elongated form, the charge density is reduced while native folding is lost. Biomolecules with unstructured domains can undergo a mix of CRM and CEM, which is the Bead Ejection Mechanism (BEM). Structured domains follow the CRM, while unstructured domains undergo chain ejection. If all folded domains are preserved, bead ejection can still be considered a native mechanism.[133–135]

For native MS, we want to maximize the probability of the Charged Residue Mechanism. This is most effectively done by maintaining high ionic strength. With sufficient charges on the surface, the analyte ion will remain at the center of the droplet. Chain ejection is more likely to occur when the analyte ion migrates to the surface. This can be due to low ionic strength, but also due to hydrophobicity (not an issue for DNA, but a big problem with lipophilic proteins, for example). Chain ejection produces higher charge states, which have higher response factors. Therefore, denaturation by chain ejection can be desirable for top-down methods. Another method to obtain high charge states is the addition of supercharging agents. The exact mechanism behind supercharging remains unknown, but depending on the supercharging agent the process can be native or denaturing.[133,136]

Native MS of G-quadruplexes is a powerful tool to analyze species with a difference in mass. One of the key mass shifts is the number of cations bound to the G-quadruplex, from which we can track G-quadruplex folding (Figure 19). For each K^+ bound there is a mass shift of 38 Da (^1H detaches, ^{39}K attaches). K^+ ions bind in two ways: non-specific and specific. Non-specific K^+ ions attach to phosphate groups on the DNA backbone; their distribution follows a discrete statistical distribution (i.e. they look bell-shaped) that can be approximated with a Poisson function.[137] The sequence 24nonG4 exhibits a good example of an unspecific cation distribution. 0 K^+ is the most abundant adduct and the cation distribution pattern is bell-shaped. So even if 24nonG4 contains 50% guanines we can deduce from the mass spectrum that it does not form a G-quadruplex. Specific K^+ ions in G-Quadruplexes are the ones located between 2 G-tetrads. An adduct with n K^+ ions is considered specific when it is significantly more populated than its neighboring adducts n+1 and n-1. 24TTG forms a 3-tetrad G-quadruplex, which has 2 K^+ ions bound within the structure. The mass spectra clearly show 2 K^+ as the main adduct, with little population of 0/1 K^+, proving that the sequence indeed forms a G-quadruplex that is stable in the sample buffer.

Na^+ is present as a contamination and due to DNA strands being polyanions, cations like Na^+ will easily attach to them. In G-quadruplex DNA, specific K^+ adducts bind with a much higher affinity than unspecific Na^+ adducts, making Na^+ adduct peaks look small in comparison. In non-quadruplex DNA, Na^+ and K^+ compete as unspecific adducts, causing broad and difficult-to-read adduct distributions (bottom example in Figure 19). To mitigate the uncertainty of whether an adduct is specific or non-specific, we recommend checking the highest charge state (in native conditions). Note how in Figure 19 the 5- charge state carries significantly less cation adducts than the 4- charge state. Although there is no definite explanation for this, the higher charge states in native MS will always carry less unspecific cations. Quantification of specific cations is

possible but requires a control experiment with an unstructured sequence of similar base composition. For example, in Figure 19, 24nonG4 (bottom) would serve as a G-rich 24-mer control sequence to subtract unspecific adducts from 24TTG (top). Our lab developed this approach and quantified the K_D values of K^+ binding to 24TTG this way.[76]

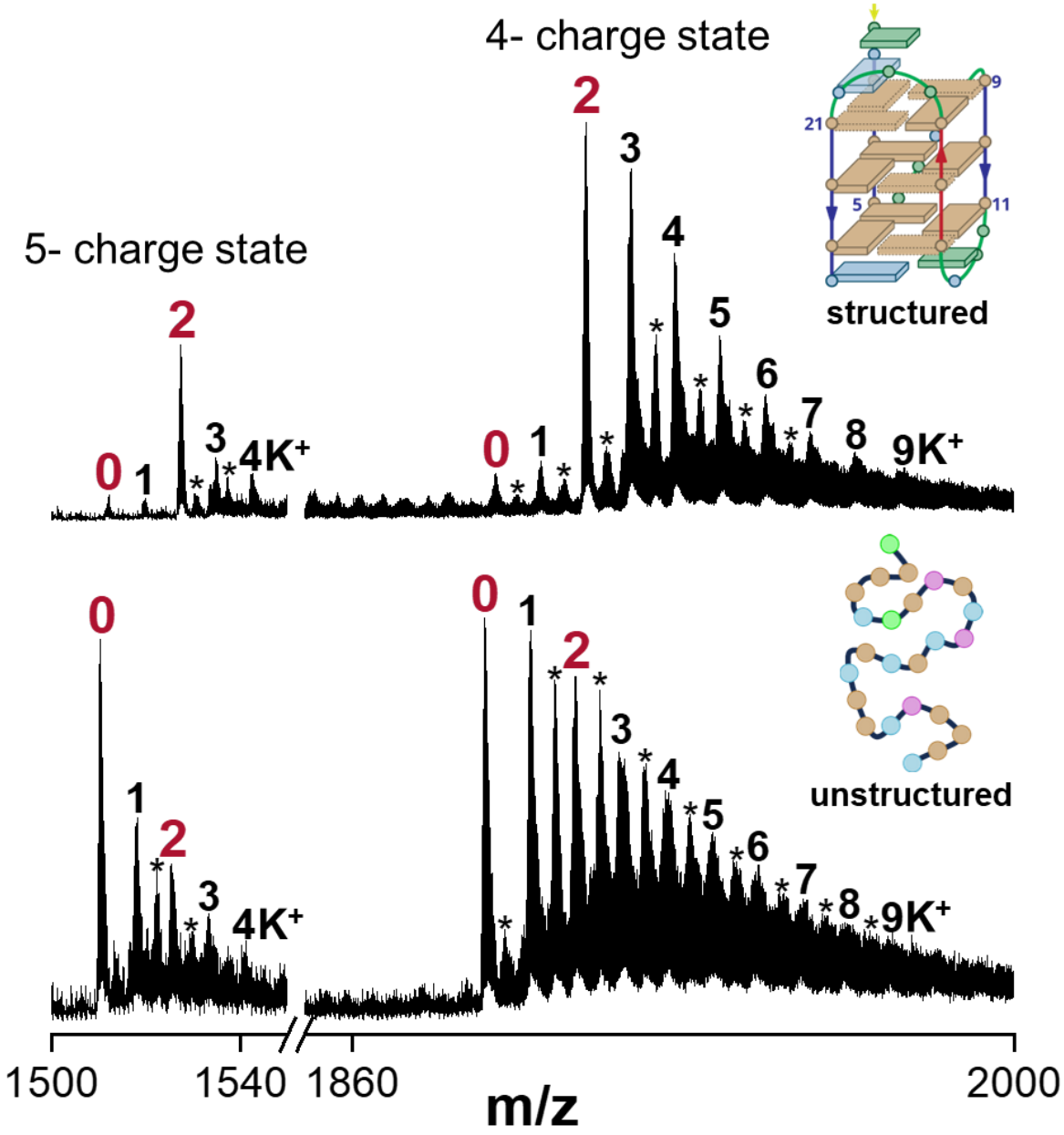


*Figure 19 Mass spectra of two 24mer G-rich DNA sequences. Top: 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A), Bottom: 24nonG4 (dT**G₃**AT**G**C**G**ACA**G**A**G**A**GG**AC**G₃**A). K^+ adducts are labeled with numbers. Na^+ adducts are labeled with (*).*

Ion mobility

Ion Mobility Spectrometry (IMS) is arguably the most powerful MS method for secondary structure analysis. An IMS setup has two key components: 1) an electric field accelerating the ions (higher charge state = faster acceleration) and 2) an inert gas medium. Analyte ions decelerate by transferring their kinetic energy to the surrounding inert gas molecules (mostly through direct collisions). The average ion velocity (v_D) is proportional to the acceleration caused by the electric field voltage (E)

and the deceleration caused by interactions between ion and inert gas, which is expressed by an ion-specific property called the ion mobility (*K*).

$$v_D = K \times E \quad (1)$$

The ion mobility (dimension: cm²/Vs) correlates to the shape of the molecule, in particular its compactness. The higher the ion mobility, the less the ion is slowed down by inert gas interactions. Since IMS separates the ions based on ion mobility, this unique method allows us to separate the conformational ensemble of G-quadruplexes and isolate the mass spectra of different conformation states. Our lab works with a Drift Tube IMS (DT-IMS) setup, which is a time-based separation method with a constant low-field voltage (Figure 20).

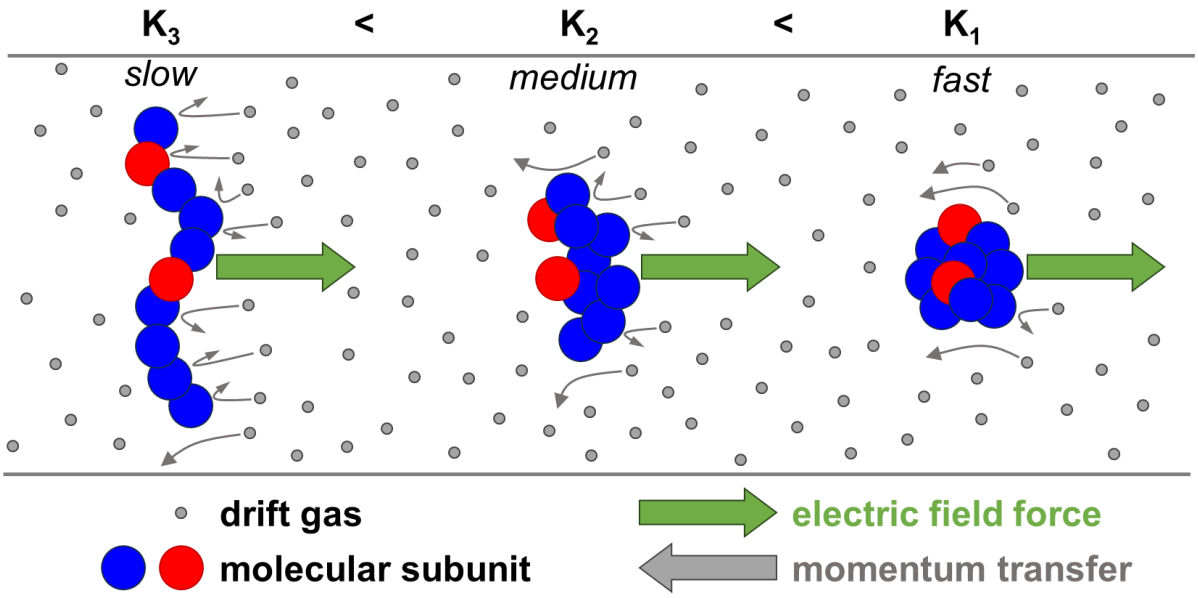


Figure 20 Illustration showing the basics of Drift Tube Ion Mobility Spectrometry (DT-IMS). Compact conformers of the same charge state will travel faster than extended conformers due to transferring less momentum to the drift gas. Ion-gas interactions are mostly collisions, but include non-covalent contacts as well.

The ions are separated based on their retention time inside the drift tube, which is linked to the ion mobility. The resolution of DT-IMS is proportional to the length of the drift tube. Nonetheless, increasing drift tube length is often impractical due to the dimensions of a normal research lab. One solution is to make the drift tube cyclic; instrumentation for cyclic IMS has very recently become commercially available.[138] Another way is increasing the retention time inside the drift tube through pulsed electric waves that push back the ions. This method is called Travelling Wave IMS (TWIMS).

Because drift time and ion mobility are dependent on the instrumental setup, the momentum transfer integral Ω is used as a reproducible reference parameter. It is more commonly referred to as the collision cross section (CCS) and has the dimension of a surface. The collision cross section can be seen as a hypothetical area of contact between analyte ion and drift gas. The majority of those contacts are collisions, but non-covalent contacts also contribute to the momentum transfer (as depicted in Figure 20). CCS values cannot be directly measured, but have to be calculated from the ion mobility.

Our DT-IMS instrument measures the arrival time (t_A) as the time between the release of ions from the trapping funnel until their detection at the TOF mass analyzer. The ion spends a portion of that time in the drift tube (drift time: t_D) and the other portion inside

the rest of the instrument (dead time: t_0). We separate those contributions by operating the drift tube at different drift voltages. Figure 21 shows the shift of arrival time distribution (ATD) at different drift voltages (ΔV) from our external calibrant, which is a solution of 40 mM dTG_4T in 150 mM ammonium acetate.

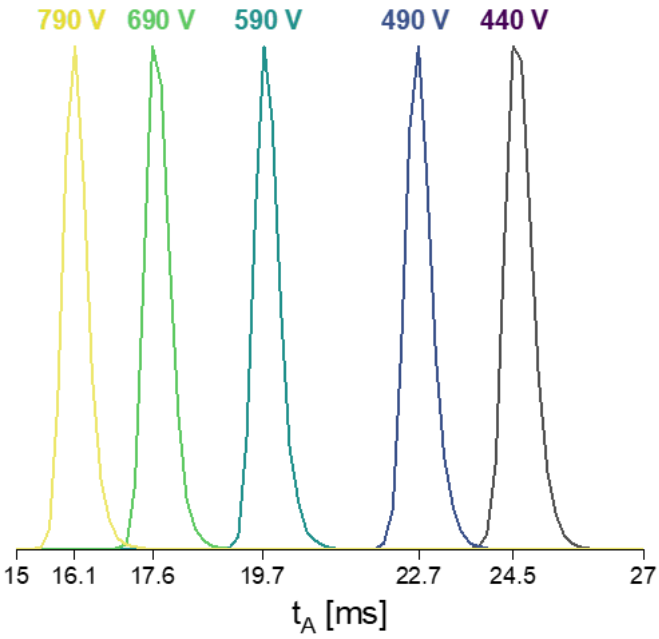


Figure 21 Arrival time distributions of the G-quadruplex species $[(dTG_4T)_4*3NH_4^+]^{5-}$ at different drift voltages.

The drift time t_D is dependent on the drift voltage ΔV, whereas the dead time is not.

$$t_D = \frac{L^2}{K \times \Delta V} \tag{2}$$

Where L is the length of the drift tube (78.1 cm) and K is the mobility of the ion species (In this case: $[(dTG_4T)_4*3NH_4^+]^{5-}$). Based on this, we can formulate the correlation between arrival time and drift voltage as a linear function.

$$t_A = t_0 + t_D = t_0 + \frac{L^2}{K} \times \frac{1}{\Delta V} \tag{3}$$

The plot is shown in Figure 22. We extract the dead time from the y-intercept and the ion mobility from the slope. We convert the ion mobility K to the collision cross section using the Mason-Schamp equation. Note that this equation follows the assumption of a static drift gas and low-field conditions.

$$CCS \cong \Omega = \frac{3}{16} \times \frac{z \times e}{N_0 \times K} \times \frac{p_0 \times T}{p \times T_0} \sqrt{\frac{2\pi}{\mu \times k_B \times T}} \tag{4}$$

The equation features several constants: e, k_B, N_0 (Loschmidt constant), p_0 (1 atm), T_0 (273.15 K). The temperature T, pressure p and the reduced mass μ ($\mu = (M_{ion}*M_{He})/(M_{ion}+M_{He})$) remain practically constant within the experimental conditions. The ion mobility K and charge state z will be different for every ion species.

This method produces a single CCS value which corresponds to the peak maximum of the arrival time distribution. For the calibrant TG_4T, our CCS values are 788.7 Å² for the 5- ion (Figure 22) and 740.4 Å² for the 4- ion (not shown). These values are in good agreement with previously published experimental results (787.5 Å² [5-], 735.7 Å² [4-]).[34]

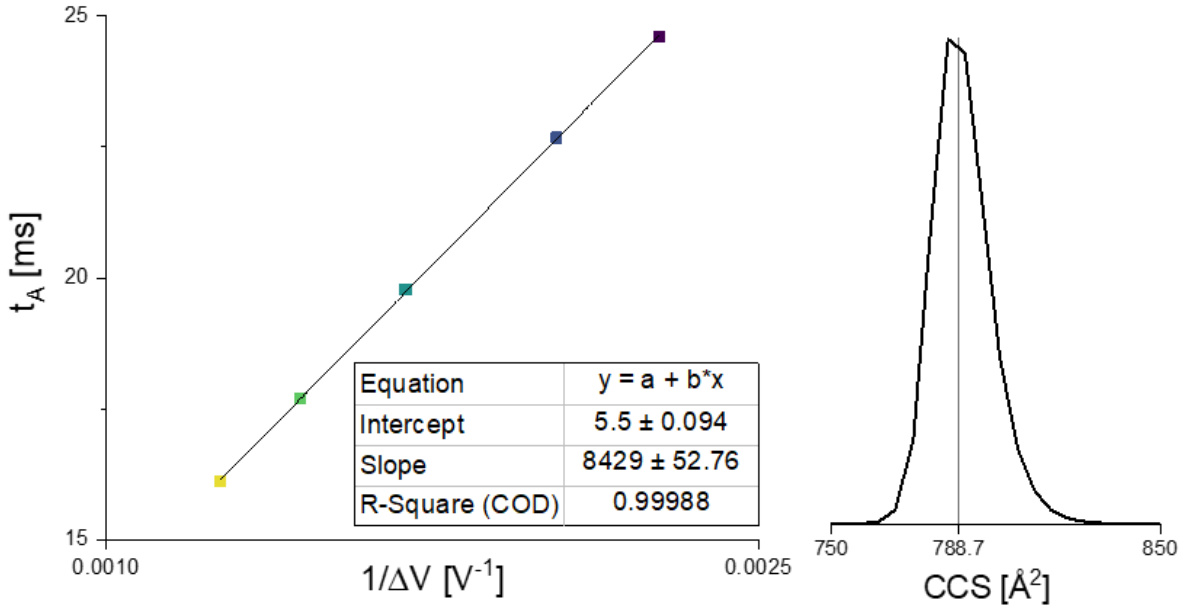


Figure 22 ***Left****: Obtaining drift time and reduced ion mobility by linearly fitting the arrival time at different drift voltages.* ***Right****: CCS distribution of [(dTG$_4$T)$_4$*3NH$_4^+$]$^{5-}$, calculated from the reduced ion mobility using the Mason-Schamp equation.*

We generate CCS distributions by converting the arrival time t_A to the respective CCS value using '*a*' as a conversion factor and $z/\sqrt{\mu}$ as a proportionality factor required for linear approximation. The conversion factor a is calculated from the CCS at peak maximum (CCS_{max}) that was determined at the previous step.[34]

$$CCS = a \times t_A \times \frac{z}{\sqrt{\mu}} \tag{5}$$

$$a = \frac{CCS_{max}}{t_{A,max}} \times \frac{\sqrt{\mu}}{z} \tag{6}$$

Conceptually, any of the five ATDs shown in Figure 21 can be converted using this method. To be as close as possible to the low-field limit, we consistently choose the ATD at the lowest drift voltage.

Changes in CCS values can be interpreted as conformational changes. Significant structural rearrangements, e.g. a DNA strand folding into a G-quadruplex, causes a significant CCS shift. In chapter 1, we utilize IMS to identify cation-induced rearrangements of G-rich DNA strand. In chapter 3, we analyze the effect of G-quadruplex ligands on the CCS distribution. Changes in the CCS distribution allowed us to identify ligand-induced G-quadruplex rearrangements. The separation of conformation states let us assess the conformational selectivity of our G-quadruplex ligands.

Although IMS cannot resolve unknown structures, experimental CCS values can be useful to validate or disprove calculated structures.[134,139] For the verification of *in silico* structures, we recommend measuring experimental CCS values in helium, since nitrogen (which is the standard drift gas for IMS) is polarizable, so that dipole-dipole interactions between analyte and N_2 will contribute to the experimental CCS value, making calculations more difficult and less accurate.[140]

Circular dichroism spectroscopy (CD)

Circular Dichroism (CD) is a spectroscopic method that uses circular polarized light. A chiral center will absorb either more left-handed or more right-handed circular polarized light. The difference in absorption (A) is measured and converted to a difference in extinction coefficient (Δε) using the Lambert-Beer law (c = concentration of chiral species, l = path length).

$$\Delta\varepsilon = \varepsilon_{left} - \varepsilon_{right} = \frac{A_{left} - A_{right}}{c \times l} \quad (7)$$

The asymmetric absorption of left-handed and right-handed circular polarized light creates elliptical polarized light. Historically, the difference in absorption was derived from the angle of ellipticity. Even though modern instruments directly measure the difference in absorption (ΔA), it is commonly converted into ellipticity (θ) for historical convenience (with the unit millidegrees or *mdeg*).[141]

$$\theta = 3298.2 \times \Delta A \quad (8)$$

Oligonucleotides absorb circular polarized light in a range of 220-300 nm. The excitation reaction is a HOMO-LUMO-transition between two nucleobases that interact through π-stacking (π→π* transition).[142] The chiral selectivity comes from the chiral arrangement of the nucleobases relative to one another. The main contributing factor is the base stacking pattern, which leads to different contact regions for π-stacking (shown in Figure 9).[143] The alignment of G-tetrads also affects the π-stacking interface: The nature of G-quadruplex stacking (hetero/homo) and the angle at which two G-tetrads are twisted relative to each other both have an effect on the CD spectrum (unpublished data). Figure 23 summarizes all structural aspects of G-quadruplexes that contribute to its signal in CD.

Although most intramolecular interactions originate from the G-quadruplex core, nucleobases in the loop and flanking regions can also form H-bond networks, such as base pairs, triads or even tetrads. These structures join the π-stacking network and contribute to the CD signal.

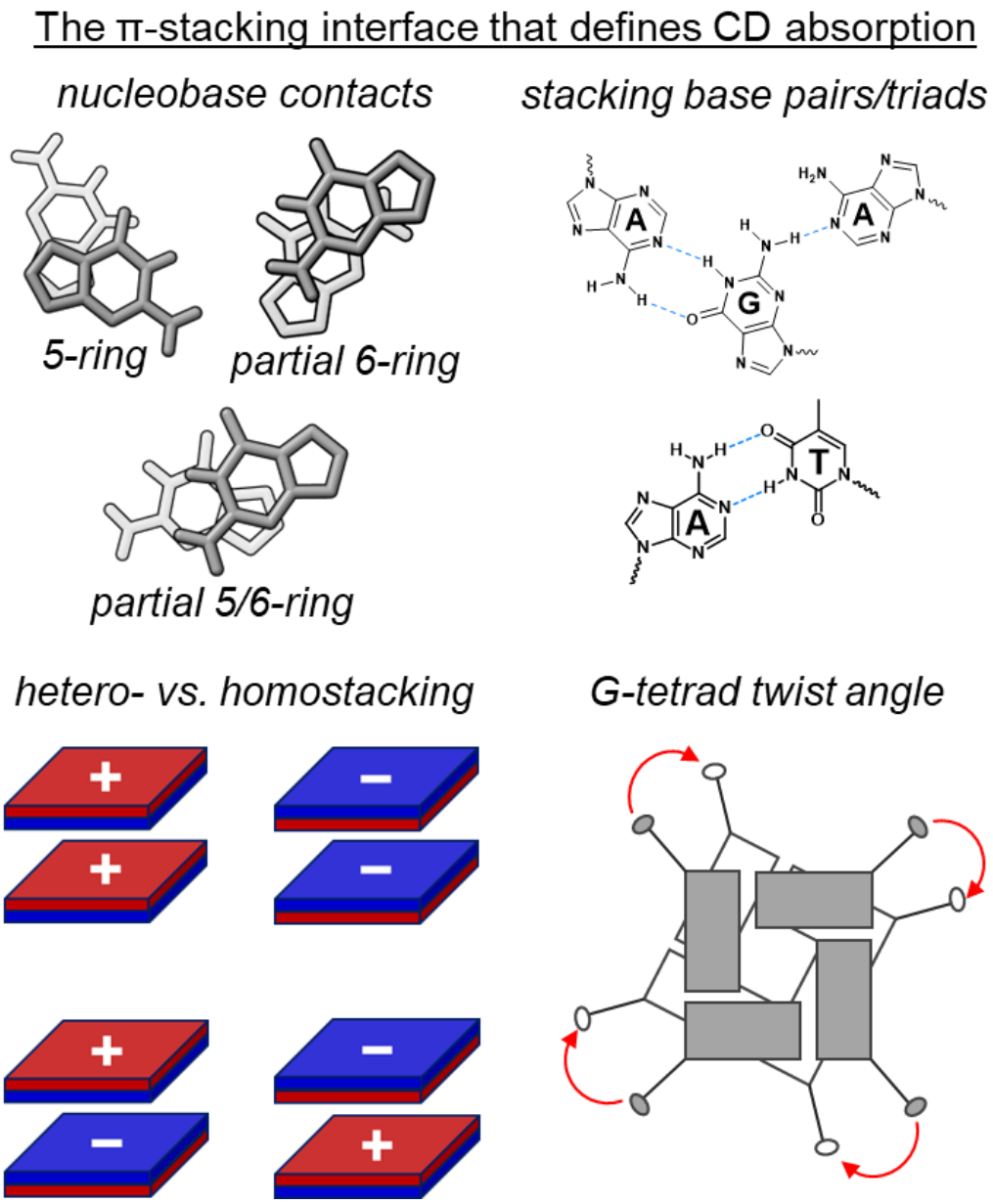


Figure 23 A schematic overview summarizing the structural origins of CD signal in G-quadruplexes.

Parallel, antiparallel and hybrid G-quadruplexes have characteristic base stacking patterns (note that the antiparallel 'chair' and 'basket' configurations both have *syn→anti* base stacks, even if their arrangement is different), leading to characteristic CD signatures for those three topology classes. Parallel G-quadruplexes have a strong positive band at 260 nm. Antiparallel G-quadruplexes have a positive band at 290 nm and a negative band at 260 nm. Hybrid G-quadruplexes have two positive bands at 260/290 nm (Figure 24).[144]

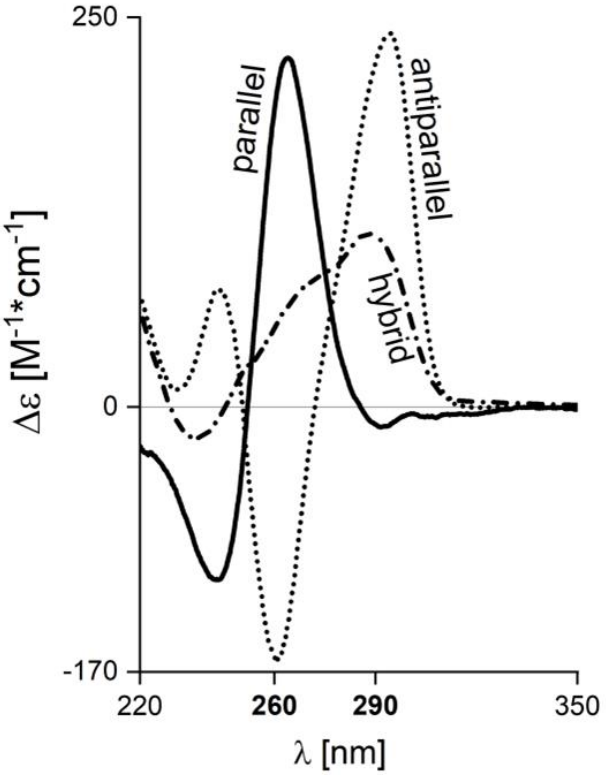


*Figure 24 Typical CD signatures for the three main topologies of G-quadruplexes. Sequences shown are dTT**G**T**GG**T**G**$_3$T**G**$_3$T**G**$_3$T (parallel, PDB: 2M4P), **G**$_4$TTA**G**$_4$TTA**G**$_4$TTA**G**$_4$T (antiparallel), TTA**G**$_3$TTA**G**$_3$TTA**G**$_3$TTA**G**$_3$TT (hybrid, PDB: 2JPZ). Samples contain 10 µM DNA and 100 mM KCl.*

CD is a simple yet powerful method for a qualitative secondary structure analysis of G-quadruplexes. CD is particularly useful in separating parallel from non-parallel G-quadruplexes. Antiparallel and hybrid G-quadruplexes, however, have similar base stacking (*syn/anti*), G-tetrad stacking (hetero/homo) and loop contributions (base pairs/triads). Furthermore, they are prone to polymorphism and can have multiple conformers hidden within a single signature. Therefore, topology assignments based on antiparallel/hybrid CD signatures should be done with caution and ideally cross-validation from other methods.

UV-melting

UV melting is a simple approach to probe the thermal stability of secondary DNA structures, including the G-quadrulex.[14,145,146] However, deeper thermodynamic analysis also yields ΔH and ΔS values of G-quadruplex folding/unfolding.[147–149] The data acquisition workflow is similar to a regular UV spectroscopy experiment, except for the setting of a temperature range and temperature ramp.[150] We will use the sequence 24TTG (dTT**G**$_3$TTA**G**$_3$TTA**G**$_3$TTA**G**$_3$A) as a model sequence throughout this section to showcase all the steps of data processing. The underlying physical phenomenon of UV melting is that G-quadruplex folding causes a change in UV absorption (Figure 25). The transition between folded and unfolded G-quadruplex is induced by a change in temperature. The G-quadruplex unfolds (denatures) at high temperature and refolds at low temperature.

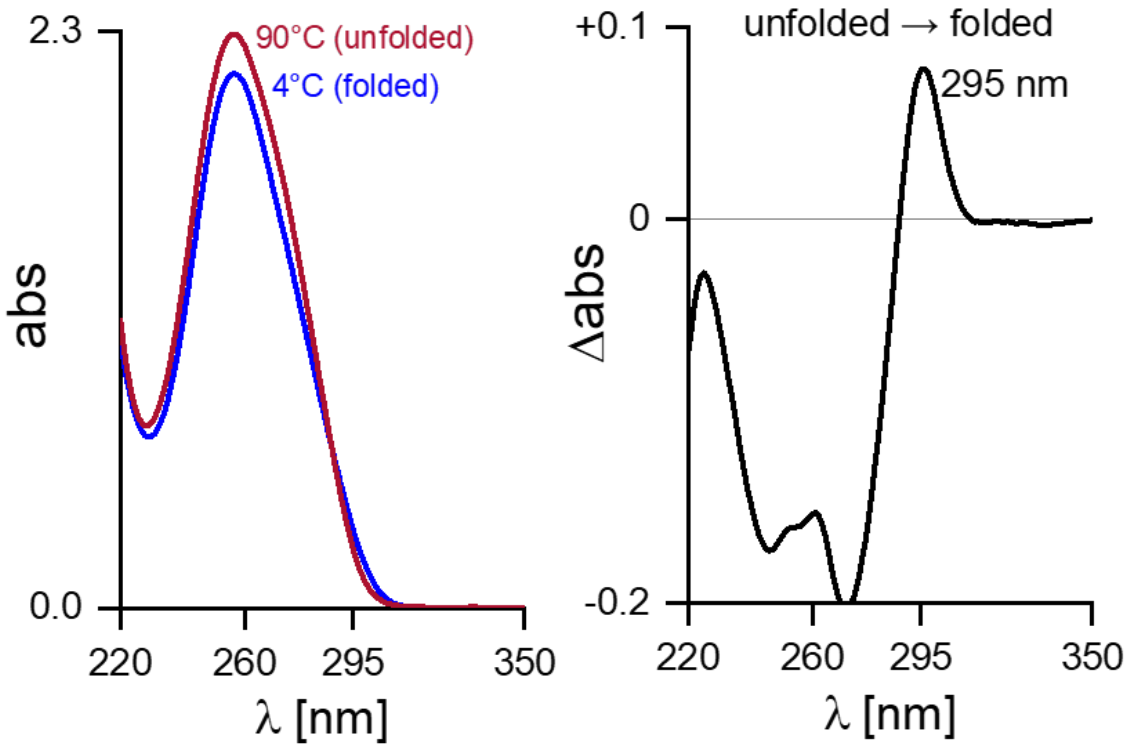


Figure 25 ***Left****: UV spectrum of 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A) at low and high temperature* ***Right****: Changes in UV absorption caused by temperature-induced G-quadruplex folding. Sample matrix contains 10 µM DNA, 0.5 mM KCl and 50 mM TMAA.*

We can use that change in UV absorption to track G-quadruplex (un)folding as a function of temperature. But first, we need to find a wavelength where the change in absorption is specific to G-quadruplex (un)folding (Figure 26).

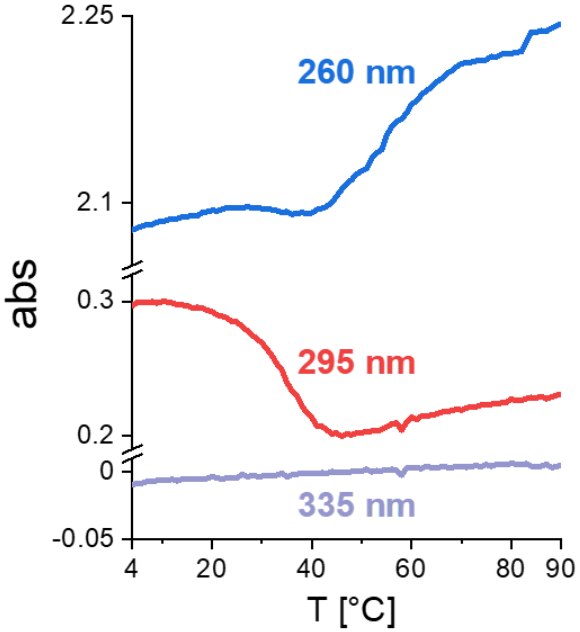


Figure 26 The temperature-dependent UV absorption of 24TTG at three different wavelengths. Sample matrix contains 10 µM DNA, 0.5 mM KCl and 50 mM TMAA.

The change in absorption is the highest around the absorption maximum at 260 nm. However, there appear to be multiple convoluted transitions that are correlated to

matrix effects, making it unfeasible for data processing. Meanwhile there is a singular, clean transition at 295 nm, indicating that the change in absorption solely originates from G-quadruplex (un)folding. 295 nm is indeed the standard wavelength to track G-quadruplex (un)folding, while a reference to remove instrumental artifacts is being measured at 335 nm.[150]

After subtracting the reference curve at 335 nm, the melting curve at 295 nm can be used to determine the melting temperature, at which 50% of the G-quadruplex is unfolded. A simple, but not fully accurate approach is to plot the first derivative of the melting curve, which results in a bell curve that has a maximum at the melting temperature.[150] This approach can be satisfying for a simple qualitative analysis of melting curves. For a more advanced analysis, we need to plot the fraction of folded DNA as a function of temperature. In order to calculate the folded fraction (θ), we use the low-temperature baseline $L_1(T)$ as a reference for θ = 1 (fully folded) and the high-temperature baseline $L_0(T)$ as a reference for θ = 0 (fully unfolded). We can fit the UV absorption A(T) on a scale of 0 to 1 with the following equation.

$$\theta(T) = \frac{L_0(T) - A(T)}{L_0(T) - L_1(T)} \tag{9}$$

Figure 27 shows the conversion of UV absorption to fraction folded for 24TTG.

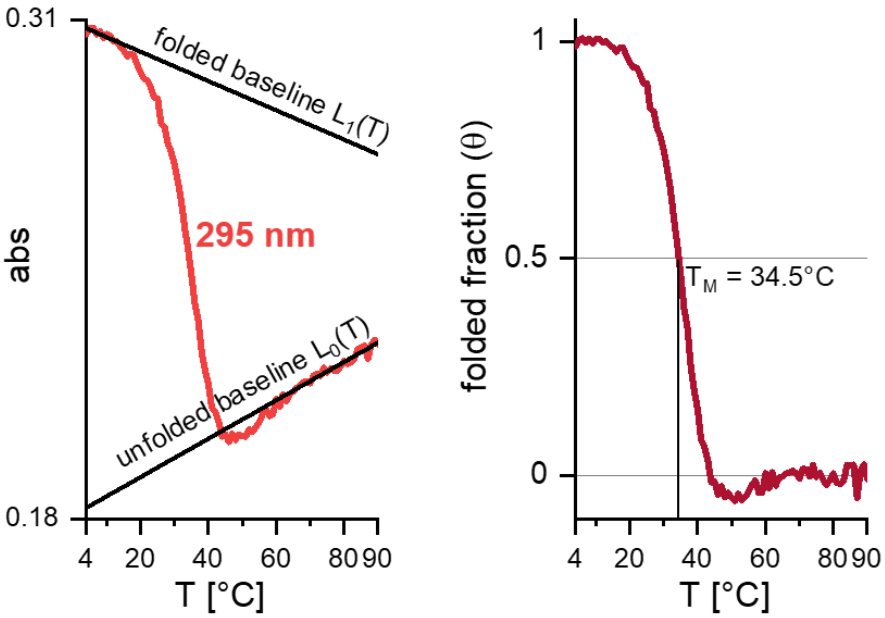


Figure 27 Left: Melting curve at 295 nm with linear functions to fit the low-temperature and high-temperature baseline. Right: Temperature-dependent folded fraction, with the melting temperature at θ = 0.5. Sample matrix contains 10 µM DNA, 0.5 mM KCl and 50 mM TMAA.

Note that the selection of datapoints to fit the temperature baselines is user-dependent. Therefore, melting temperatures are never perfectly repeatable, even with the same dataset. Our lab has developed a baseline-selection algorithm for a more consistent approach to melting data processing.[77] Also note that unexpected kinks can appear in the melting curve, which for example in Figure 27 leads to apparent negative θ values around 50°C. It is once again the user's choice whether to fit the baseline through the data points around 50°C or those from 60-90°C. As the user in question, I decided I would rather have negative θ values than a folded fraction around 10% at 90°C.

To check for slow kinetics (i.e. slower than the temperature gradient), the melting curve is often measured in both the heating and the cooling direction. It is advised to start with the cooling cycle, because that includes an annealing which removes the influence of sample storage time. When the (re)folding kinetics are slower than the temperature ramp, there will be a hysteresis between the two melting curves. 24TTG shows no remarkable hysteresis, but the low experimental precision of UV melting leads to small T_M differences between cooling and heating cycle regardless.[77] The melting temperature is commonly reported as the average between the two values from each curve.

For thermodynamic analysis the *θ(T)* curve is reprocessed as a van't-Hoff plot, with *1/T* (in 1/K) on the x-axis and *ln* K_A on the y-axis. Note that this only works when there is no significant hysteresis between heating and cooling curve (indicating slow kinetics).

$$ln\ K_A = -\frac{\Delta H}{R} \times \frac{1}{T} + \frac{\Delta S}{R} \tag{10}$$

We obtain the folding enthalpy *ΔH* from the slope and the folding entropy *ΔS* from the y-intercept. *R* is the universal gas constant. The association constant K_A is calculated from the folded fraction, assuming a simplified reaction model.

$$unfolded\ DNA \leftrightarrow folded\ DNA$$

$$K_A(T) = \frac{\theta(T)}{1-\theta(T)} \tag{11}$$

The van't-Hoff plot for 24TTG is shown in Figure 28. It should be noted that other authors suggest further restricting the analysis interval ($0.15 < \theta < 0.85$) in order to keep just the region where Ka values are most precise.[151]

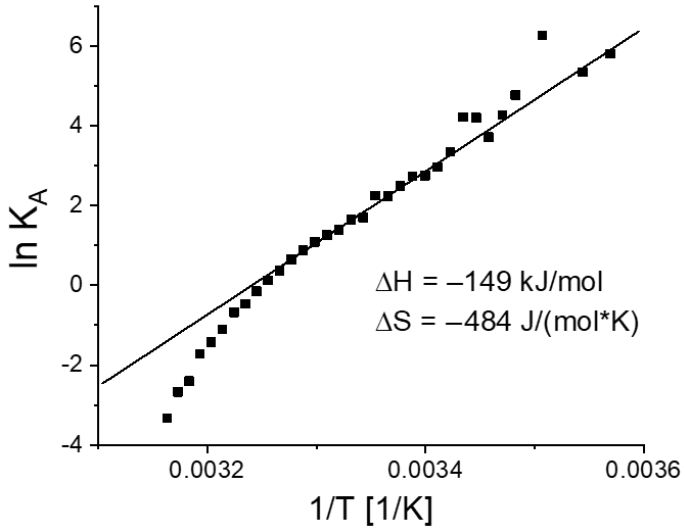


Figure 28 van't-Hoff plot of 24TTG to extract the folding enthalpy/entropy.

We obtain a folding enthalpy of -149 kJ/mol and a folding entropy of -484 J/(mol*K). Those values are comparable to the ones Mergny/Lacroix presented for 22AG (ΔH = -150 kJ/mol, ΔS = -485 J((mol*K)). 22AG (dA**G₃**TTA**G₃**TTA**G₃**TTA**G₃**) is a telomeric sequence that is similar to 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A).

X-ray crystallography

X-ray crystallography is one of the two methods used to fully resolve G-quadruplex structures (the other one is NMR). Two main steps are required to obtain a crystal structure: 1) crystallogenesis and 2) crystal structure determination.

The objective of crystallogenesis is to obtain single crystals of the analyte. For that the analyte is placed in a crystallization matrix, which has three main components.

1) Buffer to adjust the pH value, which is usually kept around 6-7 for G-quadruplexes. Standard buffers for oligonucleotides are cacodylate, Tris, HEPES or MES.[152,153] Phosphate buffers (typical for NMR) are avoided due to the risk of phosphates crystallizing.[154]

2) Precipitation agent. It is added to reduce water activity and thus facilitate the transition from solution to crystal phase. Most G-quadruplex crystals used PEG, MPD (often in combination with spermine) or ammonium sulfate as a precipitation agent.[152–154]

3) Additives – For G-quadruplexes: a mix of mono- and divalent cations. The monovalent cations are essential to G-quadruplex formation (K^+, Na^+, NH_4^+) and are often added at >100 mM concentration. The divalent cations neutralize negative charges at the phosphate backbone, reducing charge repulsion and promoting aggregation. Any alkaline earth metal (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}) works but most successful matrices contain Mg^{2+}. High concentrations of divalent cations destabilize the G-quadruplex, so they are usually kept around 10 mM.[152–154]

The analyte inside the crystallization matrix is dispensed as a small droplet that slowly evaporates over time until the analyte either precipitates (undesired outcome) or crystallizes (desired outcome). Trial-and-error remains the main strategy to obtain crystals and a wide range of crystallization matrices have to be tested before acquiring crystals suited for x-ray diffraction analysis. A detailed step-by-step tutorial on how to obtain and measure G-quadruplex crystals can be found in [155]. Figure 29 provides a simplified overview on the x-ray diffraction analysis workflow.

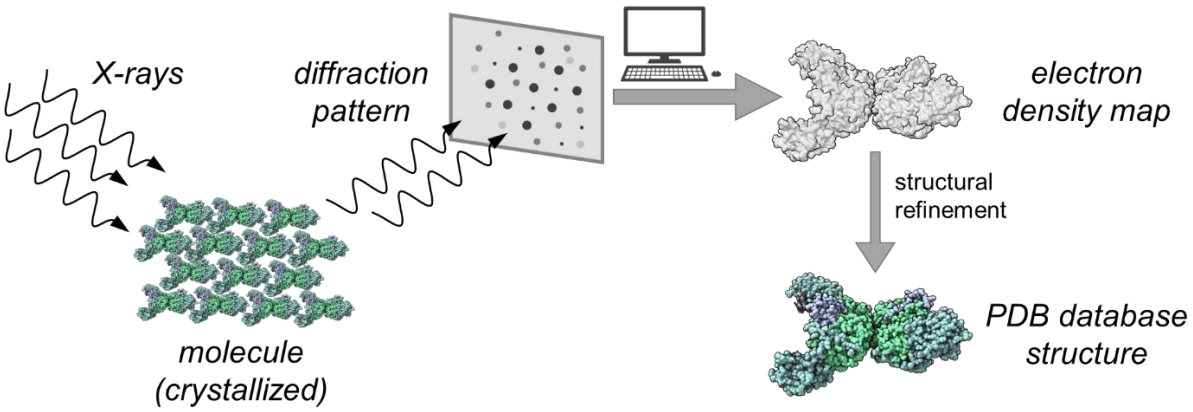


Figure 29 A simple flow chart of x-ray crystal structure analysis, depicting the x-ray diffraction experiment, the data processing to obtain the electron density map and the structural refinement based on the electron density map. The structure shown is a bacterial membrane signaling protein (PDB: 6YBU).

When a single crystal is hit by an x-ray beam, the x-rays will diffract at discrete angles: Those that meet the Bragg-condition. The resulting diffraction pattern is two-dimensional, but since the crystal is being rotated 360° in every direction, we obtain a three-dimensional projection. X-rays diffract off of electrons, and not atoms. The

diffraction pattern is therefore a projection of the crystal's electronic structure and not its atomic structure. The electronic structure is solved from the diffraction pattern which generates an electron density map. The atomic structure is then obtained by creating a structural model that fits in the electron density map and refining that structure until the best model to describe the electron density map is found.

G-quadruplexes are known to be polymorphic and flexible, so that analyzing their crystal structure can be somewhat tricky. Flexible loops or flanking nucleotides cannot be positioned based on the electron density map, because their electron density is spread thin across several conformation states. The positioning of such disordered units will entirely depend on the molecular modeling. As shown in the case of 22AG, disordering can even be a crystal structure artifact (Figure 30). The flanking nucleotide of the adjacent G-quadruplex clashes with the third TTA loop, causing internal disorder and a gap in the electron density map. Removing the flanking nucleotide in 21G restores the monomorphic loop arrangement, supported by a well-resolved electron density map.

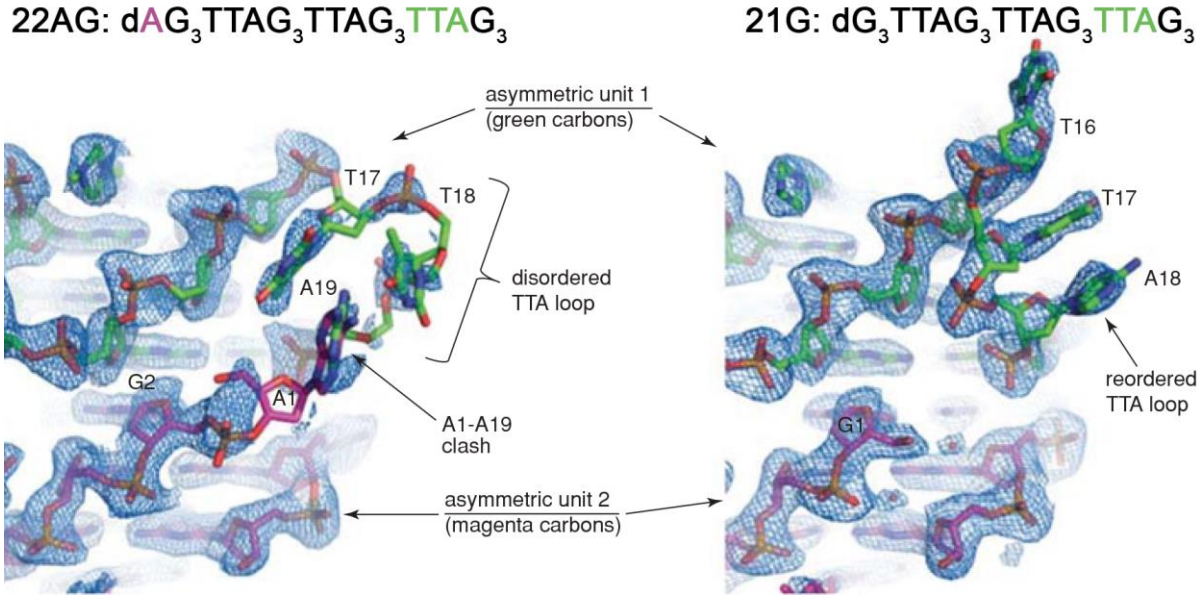


Figure 30 X-ray crystal structures of 22AG and 21G (sequences shown). One asymmetric unit shown in green, the other one in red. Clashes between the flanking A1 and the third TTA loop cause a disordering of the TTA loop and a gap in the electron density map for 22AG. Removing the flanking A1 in 21G removes the clash and leaves a well-resolved electron density map. Adapted with permission from Neidle et al. Curr. Protoc. Nucleic Acid Chem.2012, 50 (1), 1-22.[152] Copyright 2012 Wiley.

Telomeric sequences such as 22AG are parallel in crystal structures, but antiparallel/hybrid in solution-phase structures.[156] Molecular crowding appears to generally promote the parallel topology.[80] Thus, differences between crystal structures and solution-phase structures can occur for G-quadruplexes. That does not mean one structure is right and one is wrong, but that either structure is the favored configuration within its respective conditions. Which of those is a more accurate representation of the *in vivo* structure is an open debate, since there is not yet any definite answer.

The application of X-ray crystallography is mostly limited to structure analysis. One notable example beyond the scope of structure determination is racemic crystallography, where chiral ligands are crystallized with a mixture of left- and right-handed G-quadruplex. The crystal structure can be then used to probe enantiomeric preferences of G-quadruplex-ligand interactions and highlight different binding modes.[23]

NMR

NMR is the go-to method to resolve G-quadruplex structures in solution phase. The sample matrix typically contains phosphate buffer because phosphates do not have any protons. Depending on the desired cation, one would use sodium or potassium phosphate. If additional salt is needed, NaCl or KCl can be added. The standard solvent is 90/10 vol-% H_2O/D_2O. The D_2O is needed because NMR instruments use the ^2D signal to lock the magnetic field strength.[157]

Resolving solution-phase structures with NMR can be dissected into four key steps.[158,159]

1) Identifying H1 and H8 protons through labeling

The guanine nucleobase has three protons: The imino H1 proton, the amino H2 protons and the aromatic H8 proton. The H1/H2 protons are exchangeable. H1 protons are detectable at a characteristic range from 10-12 ppm, because their exchange rate is slower than the sampling rate. H2 protons are undetectable, because their exchange rate is faster than the sampling rate. Each G-quadruplex has at least 8 sets of H1/H8 protons, so to unambiguously identify which one is which, each guanine has to be labeled. Older methods would use chemical labels, such as 8-bromoguanine. However, the *-Br* residue forces the guanosine into *syn*-configuration and can thus alter the secondary structure, if not done carefully. The more common approach is isotope labeling (Figure 31).

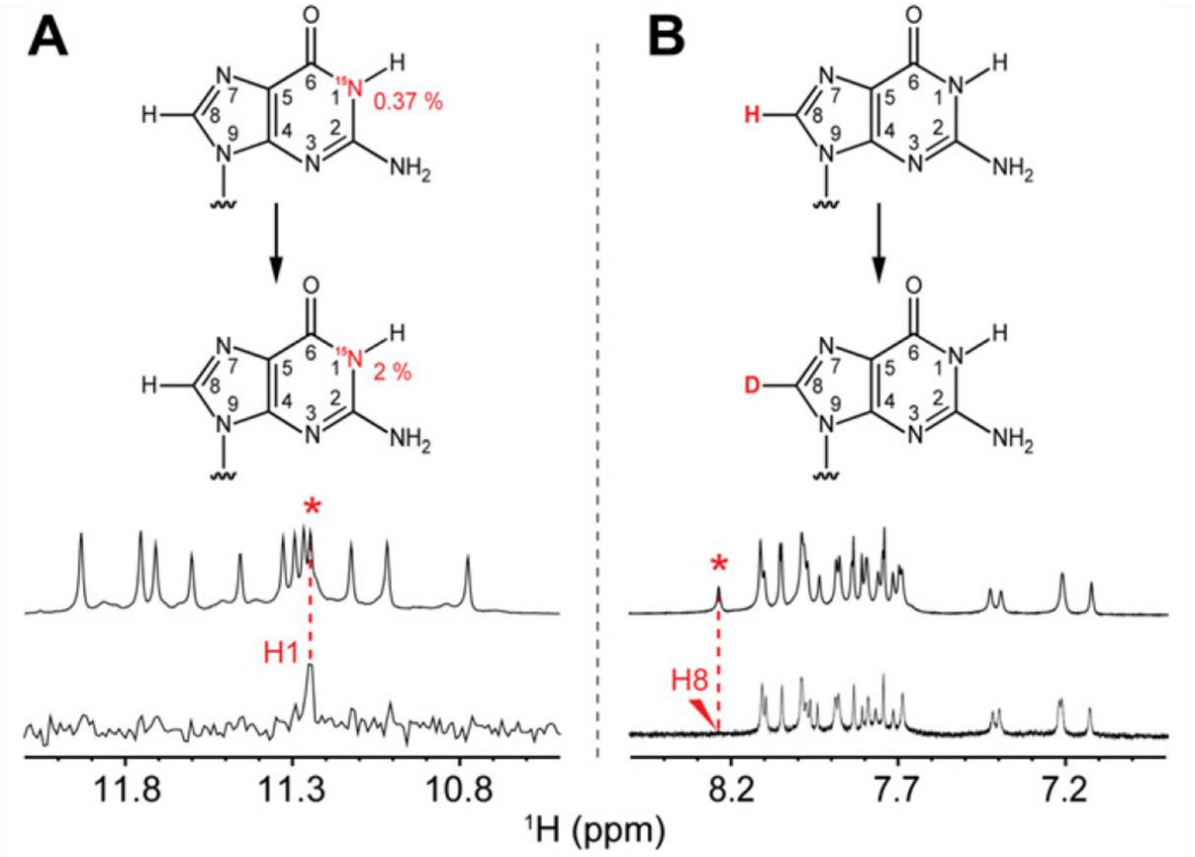


Figure 31 Isotope labeling of A) H1 protons through ^{15}N enrichment of the N1 atom and ^{15}N-^1H filtering with J-HMQC. B) H8 protons through D8 substitution. Reproduced with permission from Phan et al. Methods 2012, 57 (1), 11-24.[158] Copyright 2012 Elsevier.

To identify H1 protons, the N1 atom is $^{15}N/^{14}N$-enriched at the labeled nucleobase. Using J-HSQC, the spectra are filtered by ^{15}N-^{1}H-couplings, so that the H1 signal is several times higher on the isotope-labeled nucleobase (In Figure 31 it should be five times higher because it is enriched from 0.37% to 2%). The H8 protons are covalently bound, so when one of them is switched out with deuterium, the corresponding H8 signal will vanish. In order to unambiguously assign all H1/H8 protons, this labeling process has to be repeated for each guanine.

2) Map out the G-quadruplex core with H1-H8-H1' resonances

NOESY is a 2D-NMR method that is sensitive to the distance between two nuclei without the need for J-coupling. When two nuclei are close to each other (< 6 Å), an NOE correlation peak appears. When four guanines form a G-tetrad, they are close to one another. The NOE contacts between H1 and H8 protons let us identify which four guanines form a G-tetrad together and which guanines are next to each other (Figure 32).

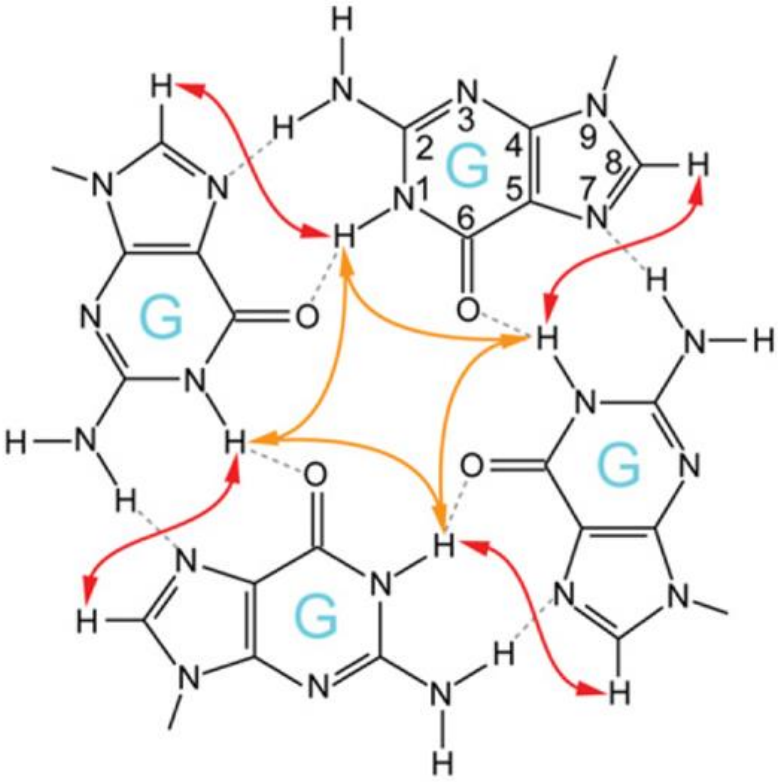


Figure 32 Characteristic NOE correlations within a G-tetrad. Reproduced with permission from Phan et al. Methods 2012, 57 (1), 11-24.[158] Copyright 2012 Elsevier.

Now we know which guanines form tetrads and how many tetrads there are, but we still lack information on the topology. To assess the topology, we can look at the glycosidic bond angle, which (for guanine) is defined between the H8 proton and the H1' proton on the sugar residue. Since each H8 proton is assigned, the corresponding C8 atoms can be assigned with ^{1}H-^{13}C-HSQC. The C8 atoms are ^{3}J-coupled to the H1' protons, so they can be linked by HMBC. That way each H1' proton is unambiguously assigned based on the H8 protons (that were previously assigned by labeling).

Even though the H1'-H8-NOE intensity is correlated to the glycosidic bond angle, the readout is not quantitative. Instead, the sequence-adjacent guanine is used to probe the *syn*/*anti* arrangement pattern. Each possible combination (*anti*→*anti*, *syn*→*syn*, *syn*→*anti*, *anti*→*syn*) has a characteristic NOE pattern connecting the two pairs of

H8/H1’ protons (Figure 33). The *syn/anti* stacking pattern provides strong evidence on G-quadruplex topology.

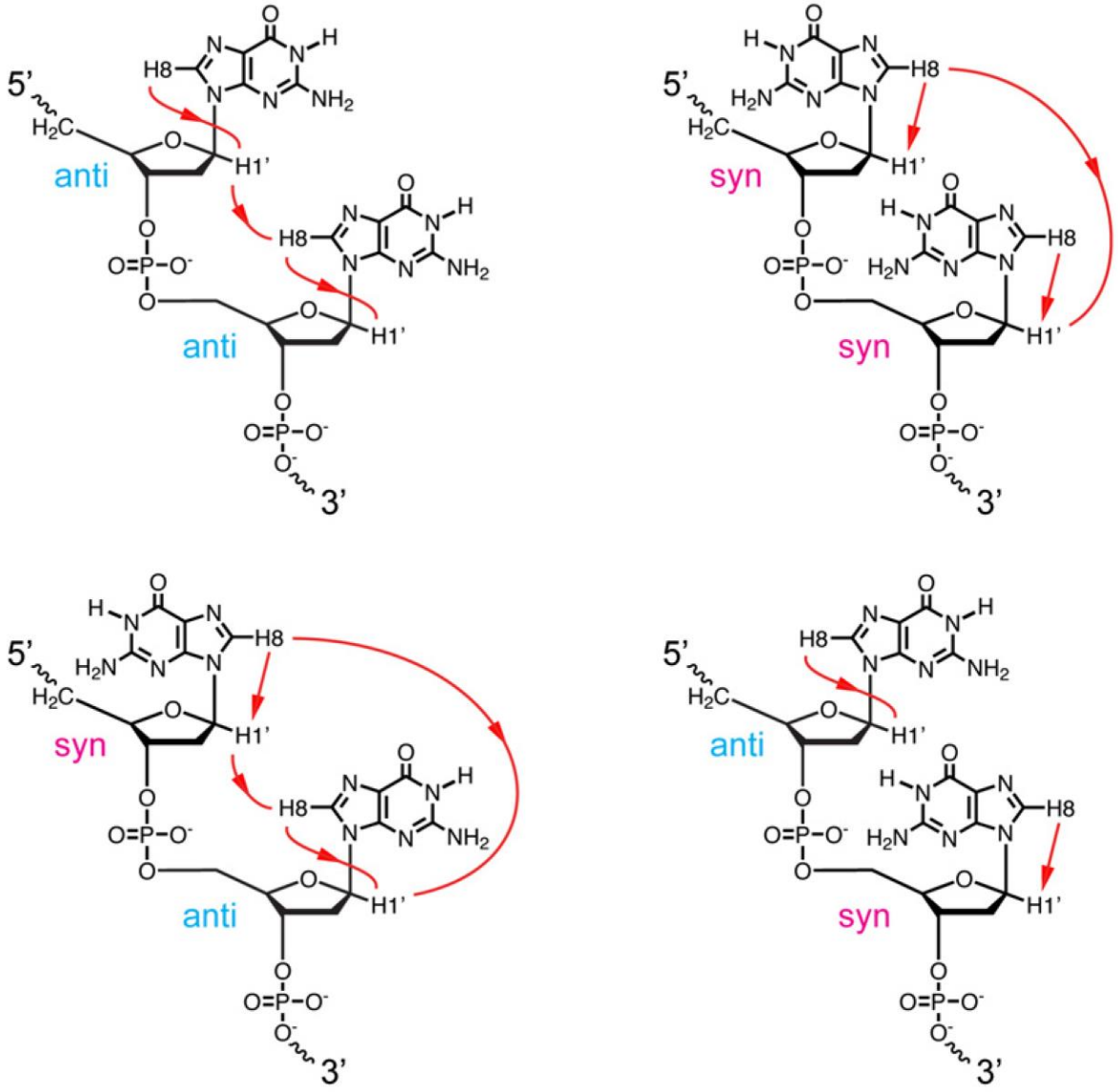


Figure 33 Characteristic NOE correlations between two sequence-adjacent guanines, depending on their syn/anti-pairing. Reproduced with permission from Phan et al. Methods 2012, 57 (1), 11-24.[158] Copyright 2012 Elsevier.

3) Identify every other proton and determine the NOE network

The H1, H8 and H1’ protons are already identified, leaving the remaining sugar protons (H2’, H2’’, H3’, H4’, H5’, H5’’). Since the sugar moiety is a singular spin system, the sugar protons on the same nucleobase are connected by TOCSY correlation peaks. With the H1’ already identified, the remaining sugar protons are affiliated through TOCSY and assigned by NOESY (proximity) and/or COSY (J-coupling).

With all protons assigned, the proximity network of protons is mapped out through NOESY contacts, like in the example shown in Figure 34.

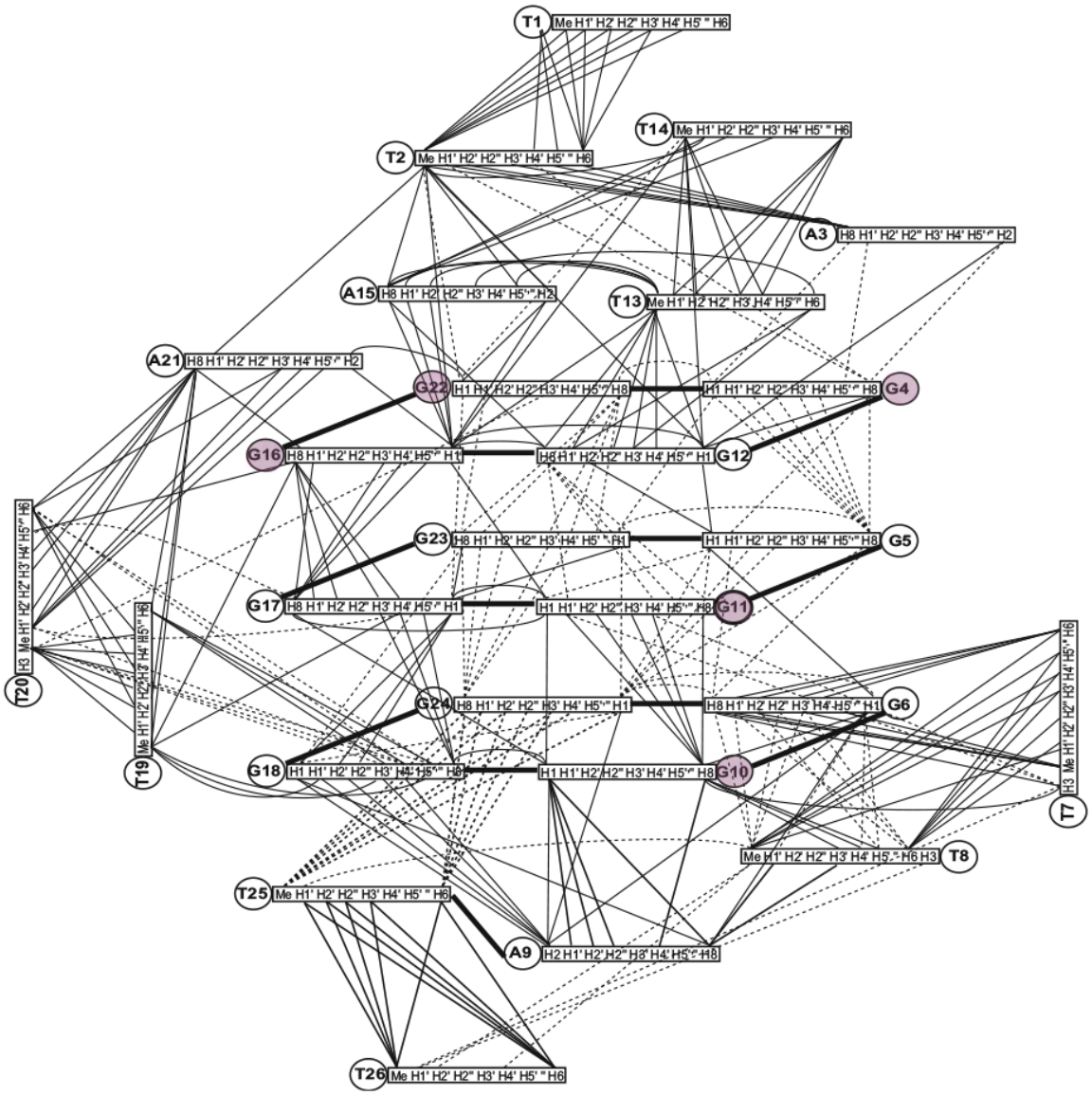


*Figure 34 NOE contacts mapping out the proximity network of 26TTA (dTTA**G₃**TTA**G₃**TTA**G₃**TTA**G₃**TT). syn-guanines are highlighted in pink. Adapted with permission from Yang et al. Methods in molecular biology 2019, vol. 2035.[159] Copyright 2019 Springer.*

4) Use the NOE constraints to model a structure

The NOE cross-peaks provide distance constraints that are used to model a resolved structure. Similar to X-ray crystallography, you start by building a model structure which you then refine until the resolved structure has the best possible correlation with the experimental data.

NMR can address questions on secondary structure at various levels of detail – one does not have to go for a fully resolved structure right off the bat. One can count the number of H1 protons to estimate the number of G-tetrads – or just look for H1 protons to see if a G-quadruplex forms in the first place. One can localize the binding site of G-quadruplex ligands and find that the binding site is different depending on the G-quadruplex topology.[160] The development of in-cell NMR opens up new prospects of assessing G-quadruplex folds *in vivo*.[161] In summary, the only limitations to structural analysis by NMR spectroscopy are resources and expertise.

References

(1) Evich, M.; Spring-Connell, A. M.; Germann, M. W. Impact of Modified Ribose Sugars on Nucleic Acid Conformation and Function. *Heterocyclic Communications* **2017**, *23* (3), 155–165. https://doi.org/10.1515/hc-2017-0056.

(2) Micklefield, J. Backbone Modification of Nucleic Acids: Synthesis, Structure and Therapeutic Applications. *Current Medicinal Chemistry* **2001**, *8* (10), 1157–1179.

(3) Do, N. Q.; Phan, A. T. Monomer–Dimer Equilibrium for the 5′–5′ Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study. *Chemistry – A European Journal* **2012**, *18* (46), 14752–14759. https://doi.org/10.1002/chem.201103295.

(4) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* **1993**, *1* (4), 263–282. https://doi.org/10.1016/0969-2126(93)90015-9.

(5) Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. Structure of the Human Telomere in K+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J Am Chem Soc* **2006**, *128* (30), 9963–9970. https://doi.org/10.1021/ja062791w.

(6) Haase, L.; Dickerhoff, J.; Weisz, K. Sugar Puckering Drives G-Quadruplex Refolding: Implications for V-Shaped Loops. *Chemistry – A European Journal* **2020**, *26* (2), 524–533. https://doi.org/10.1002/chem.201904044.

(7) Kuryavyi, V.; Patel, D. J. Solution Structure of a Unique G-Quadruplex Scaffold Adopted by a Guanosine-Rich Human Intronic Sequence. *Structure* **2010**, *18* (1), 73–82. https://doi.org/10.1016/j.str.2009.10.015.

(8) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter. *J. Am. Chem. Soc.* **2007**, *129* (14), 4386–4392. https://doi.org/10.1021/ja068739h.

(9) Galer, P.; Wang, B.; Šket, P.; Plavec, J. Reversible pH Switch of Two-Quartet G-Quadruplexes Formed by Human Telomere. *Angewandte Chemie International Edition* **2016**, *55* (6), 1993–1997. https://doi.org/10.1002/anie.201507569.

(10) Schnarr, L.; Jana, J.; Preckwinkel, P.; Weisz, K. Impact of a Snap-Back Loop on Stability and Ligand Binding to a Parallel G-Quadruplex. *J. Phys. Chem. B* **2020**, *124* (14), 2778–2787. https://doi.org/10.1021/acs.jpcb.0c00700.

(11) Lenarčič Živković, M.; Rozman, J.; Plavec, J. Adenine-Driven Structural Switch from a Two- to Three-Quartet DNA G-Quadruplex. *Angewandte Chemie International Edition* **2018**, *57* (47), 15395–15399. https://doi.org/10.1002/anie.201809328.

(12) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131* (12), 4301–4309. https://doi.org/10.1021/ja807503g.

(13) Ngoc Nguyen, T. Q.; Lim, K. W.; Phan, A. T. Duplex Formation in a G-Quadruplex Bulge. *Nucleic Acids Research* **2020**, *48* (18), 10567–10575. https://doi.org/10.1093/nar/gkaa738.

(14) Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability. *Nucleic Acids Research* **2010**, *38* (21), 7858–7868. https://doi.org/10.1093/nar/gkq639.

(15) Chen, J.; Cheng, M.; Salgado, G. F.; Stadlbauer, P.; Zhang, X.; Amrane, S.; Guédin, A.; He, F.; Šponer, J.; Ju, H.; Mergny, J.-L.; Zhou, J. The Beginning and the End: Flanking Nucleotides Induce a Parallel G-Quadruplex Topology. *Nucleic Acids Research* **2021**, No. gkab681. https://doi.org/10.1093/nar/gkab681.

(16) Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. Structure of the Intramolecular Human Telomeric G-Quadruplex in Potassium Solution: A Novel Adenine Triple Formation. *Nucleic Acids Research* **2007**, *35* (7), 2440–2450. https://doi.org/10.1093/nar/gkm009.
(17) Sun, D.; Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009**, *52* (9), 2863–2874. https://doi.org/10.1021/jm900055s.
(18) Sekibo, D. A. T.; Fox, K. R. The Effects of DNA Supercoiling on G-Quadruplex Formation. *Nucleic Acids Res* **2017**, *45* (21), 12069–12079. https://doi.org/10.1093/nar/gkx856.
(19) Li, K.; Yatsunyk, L.; Neidle, S. Water Spines and Networks in G-Quadruplex Structures. *Nucleic Acids Research* **2021**, *49* (1), 519–528. https://doi.org/10.1093/nar/gkaa1177.
(20) Dvorkin, S. A.; Karsisiotis, A. I.; Webba da Silva, M. Encoding Canonical DNA Quadruplex Structure. *Sci. Adv.* **2018**, *4* (8), eaat3007. https://doi.org/10.1126/sciadv.aat3007.
(21) Cang, X.; Šponer, J.; Cheatham, T. E., III. Explaining the Varied Glycosidic Conformational, G-Tract Length and Sequence Preferences for Anti-Parallel G-Quadruplexes. *Nucleic Acids Research* **2011**, *39* (10), 4499–4512. https://doi.org/10.1093/nar/gkr031.
(22) Liu, C.; Zhou, B.; Geng, Y.; Tam, D. Y.; Feng, R.; Miao, H.; Xu, N.; Shi, X.; You, Y.; Hong, Y.; Tang, B. Z.; Lo, P. K.; Kuryavyi, V.; Zhu, G. A Chair-Type G-Quadruplex Structure Formed by a Human Telomeric Variant DNA in K+ Solution. *Chem. Sci.* **2018**, *10* (1), 218–226. https://doi.org/10.1039/C8SC03813A.
(23) Mandal, P. K.; Collie, G. W.; Kauffmann, B.; Huc, I. Racemic DNA Crystallography. *Angewandte Chemie International Edition* **2014**, *53* (52), 14424–14427. https://doi.org/10.1002/anie.201409014.
(24) Daly, S.; Rosu, F.; Gabelica, V. Mass-Resolved Electronic Circular Dichroism Ion Spectroscopy. *Science*, 2020, *368*, 1465. https://doi.org/10.1126/science.abb1822.
(25) Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T. Structure of a Left-Handed DNA G-Quadruplex. *Proceedings of the National Academy of Sciences* **2015**, *112* (9), 2729–2733. https://doi.org/10.1073/pnas.1418718112.
(26) Das, P.; Phan, A. T. Tetrad-Binding Ligands Do Not Bind Specifically to Left-Handed G-Quadruplexes. *Chem. Commun.* **2022**, *58* (80), 11264–11267. https://doi.org/10.1039/D2CC03374G.
(27) Kogut, M.; Kleist, C.; Czub, J. Why Do G-Quadruplexes Dimerize through the 5'-Ends? Driving Forces for G4 DNA Dimerization Examined in Atomic Detail. *PLOS Computational Biology* **2019**, *15* (9), e1007383. https://doi.org/10.1371/journal.pcbi.1007383.
(28) Smargiasso, N.; Rosu, F.; Hsia, W.; Colson, P.; Baker, E. S.; Bowers, M. T.; De Pauw, E.; Gabelica, V. G-Quadruplex DNA Assemblies: Loop Length, Cation Identity, and Multimer Formation. *J. Am. Chem. Soc.* **2008**, *130* (31), 10208–10216. https://doi.org/10.1021/ja801535e.
(29) Frasson, I.; Pirota, V.; Richter, S. N.; Doria, F. Multimeric G-Quadruplexes: A Review on Their Biological Roles and Targeting. *International Journal of Biological Macromolecules* **2022**, *204*, 89–102. https://doi.org/10.1016/j.ijbiomac.2022.01.197.
(30) Karimata, H.; Miyoshi, D.; Fujimoto, T.; Koumoto, K.; Wang, Z.-M.; Sugimoto, N. Conformational Switch of a Functional Nanowire Based on the DNA G-Quadruplex.

Nucleic Acids Symposium Series **2007**, *51* (1), 251–252. https://doi.org/10.1093/nass/nrm126.
(31) Schultze, P.; Smith, F. W.; Feigon, J. Refined Solution Structure of the Dimeric Quadruplex Formed from the Oxytricha Telomeric Oligonucleotide d(GGGGTTTTGGGG). *Structure* **1994**, *2* (3), 221–233. https://doi.org/10.1016/S0969-2126(00)00023-X.
(32) Jing, H.; Fu, W.; Hu, W.; Xu, S.; Xu, X.; He, M.; Liu, Y.; Zhang, N. NMR Structural Study on the Self-Trimerization of d(GTTAGG) into a Dynamic Trimolecular G-Quadruplex Assembly Preferentially in Na+ Solution with a Moderate K+ Tolerance. *Nucleic Acids Research* **2021**, No. gkab028. https://doi.org/10.1093/nar/gkab028.
(33) Creze, C.; Rinaldi, B.; Haser, R.; Bouvet, P.; Gouet, P. Structure of a d(TGGGGT) Quadruplex Crystallized in the Presence of Li+ Ions. *Acta Cryst D* **2007**, *63* (6), 682–688. https://doi.org/10.1107/S0907444907013315.
(34) Marchand, A.; Livet, S.; Rosu, F.; Gabelica, V. Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* **2017**, *89* (23), 12674–12681. https://doi.org/10.1021/acs.analchem.7b01736.
(35) Rosu, F.; Gabelica, V.; Poncelet, H.; De Pauw, E. Tetramolecular G-Quadruplex Formation Pathways Studied by Electrospray Mass Spectrometry. *Nucleic Acids Research* **2010**, *38* (15), 5217–5225. https://doi.org/10.1093/nar/gkq208.
(36) Pavc, D.; Wang, B.; Spindler, L.; Drevenšek-Olenik, I.; Plavec, J.; Šket, P. GC Ends Control Topology of DNA G-Quadruplexes and Their Cation-Dependent Assembly. *Nucleic Acids Research* **2020**, *48* (5), 2749–2761. https://doi.org/10.1093/nar/gkaa058.
(37) Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *The Alkali Metal Ions: Their Role for Life*, 2016, 203–258. https://doi.org/10.1007/978-3-319-21756-7_7.
(38) Hadži, S.; Lah, J. Origin of Heat Capacity Increment in DNA Folding: The Hydration Effect. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2021**, *1865* (1), 129774. https://doi.org/10.1016/j.bbagen.2020.129774.
(39) Swasey, S. M.; Leal, L. E.; Lopez-Acevedo, O.; Pavlovich, J.; Gwinn, E. G. Silver (I) as DNA Glue: Ag + -Mediated Guanine Pairing Revealed by Removing Watson-Crick Constraints. *Scientific Reports* **2015**, *5* (1), 10163. https://doi.org/10.1038/srep10163.
(40) Swasey, S. M.; Rosu, F.; Copp, S. M.; Gabelica, V.; Gwinn, E. G. Parallel Guanine Duplex and Cytosine Duplex DNA with Uninterrupted Spines of AgI -Mediated Base Pairs. *J. Phys. Chem. Lett.* **2018**, *9* (22), 6605–6610. https://doi.org/10.1021/acs.jpclett.8b02851.
(41) Blume, S. W.; Guarcello, V.; Zacharias, W.; Miller, D. M. Divalent Transition Metal Cations Counteract Potassium-Induced Quadruplex Assembly of Oligo(dG) Sequences. *Nucleic Acids Res* **1997**, *25* (3), 617–625.
(42) Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallographica Section A* **1976**, *32* (5), 751–767. https://doi.org/10.1107/S0567739476001551.
(43) Hong, E. S.; Yoon, H.-J.; Kim, B.; Yim, Y.-H.; So, H.-Y.; Shin, S. K. Mass Spectrometric Studies of Alkali Metal Ion Binding on Thrombin-Binding Aptamer DNA. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (7), 1245–1255. https://doi.org/10.1016/j.jasms.2010.03.035.

(44) Vairamani, M.; Gross, M. L. G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **2003**, *125* (1), 42–43. https://doi.org/10.1021/ja0284299.
(45) Piekarczyk, A.; Bald, I.; Flosadóttir, H. D.; Ómarsson, B.; Lafosse, A.; Ingólfsson, O. Influence of Metal Ion Complexation on the Metastable Fragmentation of DNA Hexamers. *Eur. Phys. J. D* **2014**, *68* (6), 146. https://doi.org/10.1140/epjd/e2014-40838-7.
(46) Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *Journal of the American Chemical Society*, 2016, *138*, 2780–2792. https://doi.org/10.1021/jacs.5b13130.
(47) Ma, G.; Yu, Z.; Zhou, W.; Li, Y.; Fan, L.; Li, X. Investigation of Na+ and K+ Competitively Binding with a G-Quadruplex and Discovery of a Stable K+–Na+-Quadruplex. *J. Phys. Chem. B* **2019**, *123* (26), 5405–5411. https://doi.org/10.1021/acs.jpcb.9b02823.
(48) Auffinger, P.; D'Ascenzo, L.; Ennifar, E. Sodium and Potassium Interactions with Nucleic Acids. In *The Alkali Metal Ions: Their Role for Life*; Sigel, A., Sigel, H., Sigel, R. K. O., Eds.; Metal Ions in Life Sciences; Springer International Publishing: Cham, 2016; Vol. 16, pp 167–201. https://doi.org/10.1007/978-3-319-21756-7_6.
(49) Auffinger, P.; Grover, N.; Westhof, E. 1:Metal Ion Binding to RNA. In *Structural and Catalytic Roles of Metal Ions in RNA*; 2011; pp 1–36. https://doi.org/10.1039/9781849732512-00001.
(50) Seo, J.; Hong, E. S.; Yoon, H.-J.; Shin, S. K. Specific and Nonspecific Bindings of Alkaline-Earth Metal Ions to Guanine-Quadruplex Thrombin-Binding Aptamer DNA. *International Journal of Mass Spectrometry* **2012**, *330–332*, 262–270. https://doi.org/10.1016/j.ijms.2012.09.002.
(51) Kenderdine, T.; Xia, Z.; Williams, E. R.; Fabris, D. Submicrometer Nanospray Emitters Provide New Insights into the Mechanism of Cation Adduction to Anionic Oligonucleotides. *Anal. Chem.* **2018**, *90* (22), 13541–13548. https://doi.org/10.1021/acs.analchem.8b03632.
(52) Wu, Q.; Cheng, X.; Hofstadler, S. A.; Smith, R. D. Specific Metal-Oligonucleotide Binding Studied by High Resolution Tandem Mass Spectrometry. *Journal of Mass Spectrometry* **1996**, *31* (6), 669–675. https://doi.org/10.1002/(sici)1096-9888(199606)31:6<669::aid-jms340>3.0.co;2-5.
(53) Wang, Y.; Taylor, J.-S.; Gross, M. L. Fragmentation of Electrospray-Produced Oligodeoxynucleotide Ions Adducted to Metal Ions. *Journal of the American Society for Mass Spectrometry* **2001**, *12* (5), 550–556. https://doi.org/10.1016/S1044-0305(01)00231-8.
(54) Keller, K. M.; Brodbelt, J. S. Charge State-Dependent Fragmentation of Oligonucleotide/Metal Complexes. *J Am Soc Mass Spectrom* **2005**, *16* (1), 28–37. https://doi.org/10.1016/j.jasms.2004.09.016.
(55) Miyoshi, D.; Nakao, A.; Sugimoto, N. Structural Transition from Antiparallel to Parallel G-Quadruplex of d(G4T4G4) Induced by Ca2+. *Nucleic Acids Res* **2003**, *31* (4), 1156–1163.
(56) Wilcox, J. M.; Rempel, D. L.; Gross, M. L. Method of Measuring Oligonucleotide−Metal Affinities: Interactions of the Thrombin Binding Aptamer with K+ and Sr2+. *Anal. Chem.* **2008**, *80* (7), 2365–2371. https://doi.org/10.1021/ac701903w.
(57) Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double

Switches. *Journal of the American Chemical Society*, 2016, *138*, 2780–2792. https://doi.org/10.1021/jacs.5b13130.
(58) Gill, M. L.; Strobel, S. A.; Loria, J. P. 205Tl NMR Methods for the Characterization of Monovalent Cation Binding to Nucleic Acids. *J. Am. Chem. Soc.* **2005**, *127* (47), 16723–16732. https://doi.org/10.1021/ja055358f.
(59) Gill, M. L.; Strobel, S. A.; Loria, J. P. Crystallization and Characterization of the Thallium Form of the Oxytricha Nova G-Quadruplex. *Nucleic Acids Res* **2006**, *34* (16), 4506–4514. https://doi.org/10.1093/nar/gkl616.
(60) Hoang, M.; Huang, P.-J. J.; Liu, J. G-Quadruplex DNA for Fluorescent and Colorimetric Detection of Thallium(I). *ACS Sens.* **2016**, *1* (2), 137–143. https://doi.org/10.1021/acssensors.5b00147.
(61) Pan, J.; Zhang, S. Interaction between Cationic Zinc Porphyrin and Lead Ion Induced Telomeric Guanine Quadruplexes: Evidence for End-Stacking. *J Biol Inorg Chem* **2009**, *14* (3), 401–407. https://doi.org/10.1007/s00775-008-0457-5.
(62) Yu, Z.; Zhou, W.; Ma, G.; Li, Y.; Fan, L.; Li, X.; Lu, Y. Insights into the Competition between K^+ and Pb^{2+} Binding to a G-Quadruplex and Discovery of a Novel K^+–Pb^{2+}–Quadruplex Intermediate. *Journal of Physical Chemistry B* **2018**, *122* (40), 9382–9388. https://doi.org/10.1021/acs.jpcb.8b08161.
(63) Yu, Z.; Zhou, W.; Han, J.; Li, Y.; Fan, L.; Li, X. Na+-Induced Conformational Change of Pb2+-Stabilized G-Quadruplex and Its Influence on Pb2+ Detection. *Anal. Chem.* **2016**, *88* (19), 9375–9380. https://doi.org/10.1021/acs.analchem.6b02466.
(64) Sidey, V. On the Effective Ionic Radii for Ammonium. *Acta Cryst B* **2016**, *72* (4), 626–633. https://doi.org/10.1107/S2052520616008064.
(65) Balthasart, F.; Plavec, J.; Gabelica, V. Ammonium Ion Binding to DNA G-Quadruplexes: Do Electrospray Mass Spectra Faithfully Reflect the Solution-Phase Species? *J. Am. Soc. Mass Spectrom.* **2013**, *24* (1), 1–8. https://doi.org/10.1007/s13361-012-0499-3.
(66) Ferreira, R.; Marchand, A.; Gabelica, V. Mass Spectrometry and Ion Mobility Spectrometry of G-Quadruplexes. A Study of Solvent Effects on Dimer Formation and Structural Transitions in the Telomeric DNA Sequence d(TAGGGTTAGGGT). *Methods* **2012**, *57* (1), 56–63. https://doi.org/10.1016/j.ymeth.2012.03.021.
(67) Marchand, A.; Ferreira, R.; Tateishi-Karimata, H.; Miyoshi, D.; Sugimoto, N.; Gabelica, V. Sequence and Solvent Effects on Telomeric DNA Bimolecular G-Quadruplex Folding Kinetics. *J. Phys. Chem. B* **2013**, *117* (41), 12391–12401. https://doi.org/10.1021/jp406857s.
(68) Mazzitelli, C. L.; Wang, J.; Smith, S. I.; Brodbelt, J. S. Gas-Phase Stability of G-Quadruplex DNA Determined by Electrospray Ionization Tandem Mass Spectrometry and Molecular Dynamics Simulations. *Journal of The American Society for Mass Spectrometry* **2007**, *18* (10), 1760–1773. https://doi.org/10.1016/j.jasms.2007.07.008.
(69) Collie, G. W.; Parkinson, G. N.; Neidle, S.; Rosu, F.; De Pauw, E.; Gabelica, V. Electrospray Mass Spectrometry of Telomeric RNA (TERRA) Reveals the Formation of Stable Multimeric G-Quadruplex Structures. *J. Am. Chem. Soc.* **2010**, *132* (27), 9328–9334. https://doi.org/10.1021/ja100345z.
(70) Guo, X.; Liu, S.; Yu, Z. Bimolecular Quadruplexes and Their Transitions to Higher-Order Molecular Structures Detected by ESI-FTICR-MS. *J. Am. Soc. Mass Spectrom.* **2007**, *18* (8), 1467–1476. https://doi.org/10.1016/j.jasms.2007.05.003.
(71) Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J.-L. Guanines Are a Quartet's Best Friend: Impact of Base Substitutions on the Kinetics and Stability of Tetramolecular Quadruplexes. *Nucleic Acids Research* **2007**, *35* (9), 3064–3075. https://doi.org/10.1093/nar/gkm111.

(72) Gabelica, V.; Rosu, F.; Witt, M.; Baykut, G.; Pauw, E. D. Fast Gas-Phase Hydrogen/Deuterium Exchange Observed for a DNA G-Quadruplex. *Rapid Communications in Mass Spectrometry* **2005**, *19* (2), 201–208. https://doi.org/10.1002/rcm.1772.
(73) Joly, L.; Rosu, F.; Gabelica, V. D(TGnT) DNA Sequences Do Not Necessarily Form Tetramolecular G-Quadruplexes. *Chem. Commun.* **2012**, *48* (67), 8386–8388. https://doi.org/10.1039/c2cc33316c.
(74) Borbone, N.; Amato, J.; Oliviero, G.; D'Atri, V.; Gabelica, V.; De Pauw, E.; Piccialli, G.; Mayol, L. D(CGGTGGT) Forms an Octameric Parallel G-Quadruplex via Stacking of Unusual G(:C):G(:C):G(:C):G(:C) Octads. *Nucleic Acids Research* **2011**, *39* (17), 7848–7857. https://doi.org/10.1093/nar/gkr489.
(75) Freund, J.; Kalbitzer, H. Physiological Buffers for NMR Spectroscopy. *J Biomol NMR* **1995**, *5* (3), 321–322. https://doi.org/10.1007/bf00211760.
(76) Marchand, A.; Gabelica, V. Folding and Misfolding Pathways of G-Quadruplex DNA. *Nucleic Acids Research*, 2016, *44*, 10999–11012. https://doi.org/10.1093/nar/gkw970 %J Nucleic Acids Research.
(77) Ghosh, A.; Largy, E.; Gabelica, V. DNA G-Quadruplexes for Native Mass Spectrometry in Potassium: A Database of Validated Structures in Electrospray-Compatible Conditions. *Nucleic Acids Research* **2021**, *49* (4), 2333–2345. https://doi.org/10.1093/nar/gkab039.
(78) Marchand, A.; Gabelica, V. Native Electrospray Mass Spectrometry of DNA G-Quadruplexes in Potassium Solution. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (7), 1146–1154. https://doi.org/10.1007/s13361-014-0890-3.
(79) Davis, J. T. G-Quartets 40 Years Later: From 5′-GMP to Molecular Biology and Supramolecular Chemistry. *Angewandte Chemie International Edition* **2004**, *43* (6), 668–698. https://doi.org/10.1002/anie.200300589.
(80) Zhang, D.-H.; Fujimoto, T.; Saxena, S.; Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **2010**, *49* (21), 4554–4563. https://doi.org/10.1021/bi1002822.
(81) Fujii, T.; Podbevšek, P.; Plavec, J.; Sugimoto, N. Effects of Metal Ions and Cosolutes on G-Quadruplex Topology. *Journal of Inorganic Biochemistry* **2017**, *166*, 190–198. https://doi.org/10.1016/j.jinorgbio.2016.09.001.
(82) Nicholson, D. A.; Nesbitt, D. J. Kinetic and Thermodynamic Control of G-Quadruplex Polymorphism by Na+ and K+ Cations. *J. Phys. Chem. B* **2023**. https://doi.org/10.1021/acs.jpcb.3c01001.
(83) Laouer, K.; Schmid, M.; Wien, F.; Changenet, P.; Hache, F. Folding Dynamics of DNA G-Quadruplexes Probed by Millisecond Temperature Jump Circular Dichroism. *J. Phys. Chem. B* **2021**, *125* (29), 8088–8098. https://doi.org/10.1021/acs.jpcb.1c01993.
(84) Esmaili, N.; Leroy, J. L. I-Motif Solution Structure and Dynamics of the d(AACCCC) and d(CCCCAA) Tetrahymena Telomeric Repeats. *Nucleic Acids Research* **2005**, *33* (1), 213–224. https://doi.org/10.1093/nar/gki160.
(85) Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Research* **2018**, *46* (16), 8038–8056. https://doi.org/10.1093/nar/gky735.
(86) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nature Chem* **2018**, *10* (6), 631–637. https://doi.org/10.1038/s41557-018-0046-3.

(87) Saxena, S.; Joshi, S.; Shankaraswamy, J.; Tyagi, S.; Kukreti, S. Magnesium and Molecular Crowding of the Cosolutes Stabilize the I-Motif Structure at Physiological pH. *Biopolymers* **2017**, *107* (7), e23018. https://doi.org/10.1002/bip.23018.
(88) Serrano-Chacón, I.; Mir, B.; Escaja, N.; González, C. Structure of I-Motif/Duplex Junctions at Neutral pH. *J. Am. Chem. Soc.* **2021**, *143* (33), 12919–12923. https://doi.org/10.1021/jacs.1c04679.
(89) Machireddy, B.; Sullivan, H.-J.; Wu, C. Binding of BRACO19 to a Telomeric G-Quadruplex DNA Probed by All-Atom Molecular Dynamics Simulations with Explicit Solvent. *Molecules* **2019**, *24* (6), 1010. https://doi.org/10.3390/molecules24061010.
(90) Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors. *Proceedings of the National Academy of Sciences* **2001**, *98* (9), 4844–4849. https://doi.org/10.1073/pnas.081560598.
(91) Monchaud, D.; Allain, C.; Bertrand, H.; Smargiasso, N.; Rosu, F.; Gabelica, V.; De Cian, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. Ligands Playing Musical Chairs with G-Quadruplex DNA: A Rapid and Simple Displacement Assay for Identifying Selective G-Quadruplex Binders. *Biochimie* **2008**, *90* (8), 1207–1223. https://doi.org/10.1016/j.biochi.2008.02.019.
(92) Verga, D.; Hamon, F.; Nicoleau, C.; Guetta, C.; Wu, T.; Guerquin-Kern; Marco, S.; Teulade-Fichou, M.-P. Chemical Imaging by NanoSIMS Provides High-Resolution Localization of the G-Quadruplex Interactive Drug (Br)-PhenDC3 on Human Chromosomes. *J. Mol. Biol. & Mol. Imaging* **2017**, *4* (1), 1029–1035.
(93) Micheli, E.; Lombardo, C. M.; D'Ambrosio, D.; Franceschin, M.; Neidle, S.; Savino, M. Selective G-Quadruplex Ligands: The Significant Role of Side Chain Charge Density in a Series of Perylene Derivatives. *Bioorganic & Medicinal Chemistry Letters* **2009**, *19* (14), 3903–3908. https://doi.org/10.1016/j.bmcl.2009.03.106.
(94) Wu, C.-C.; Li, Y.-C.; Wang, Y.-R.; Li, T.-K.; Chan, N.-L. On the Structural Basis and Design Guidelines for Type II Topoisomerase-Targeting Anticancer Drugs. *Nucleic Acids Research* **2013**, *41* (22), 10630–10640. https://doi.org/10.1093/nar/gkt828.
(95) Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Structural Basis of DNA Quadruplex Recognition by an Acridine Drug. *J. Am. Chem. Soc.* **2008**, *130* (21), 6722–6724. https://doi.org/10.1021/ja8016973.
(96) Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. Solution Structure of a G-Quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angewandte Chemie International Edition* **2014**, *53* (4), 999–1002. https://doi.org/10.1002/anie.201308063.
(97) Reznichenko, O.; Cucchiarini, A.; Gabelica, V.; Granzhan, A. Quadruplex DNA-Guided Ligand Selection from Dynamic Combinatorial Libraries of Acylhydrazones. *Org. Biomol. Chem.* **2021**, *19* (2), 379–386. https://doi.org/10.1039/D0OB01908A.
(98) Lecours, M. J.; Marchand, A.; Anwar, A.; Guetta, C.; Hopkins, W. S.; Gabelica, V. What Stoichiometries Determined by Mass Spectrometry Reveal about the Ligand Binding Mode to G-Quadruplex Nucleic Acids. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2017**, *1861* (5, Part B), 1353–1361. https://doi.org/10.1016/j.bbagen.2017.01.010.
(99) Müller, S.; Sanders, D. A.; Di Antonio, M.; Matsis, S.; Riou, J.-F.; Rodriguez, R.; Balasubramanian, S. Pyridostatin Analogues Promote Telomere Dysfunction and Long-Term Growth Inhibition in Human Cancer Cells †Electronic Supplementary Information (ESI) Available: Experimental Procedures and Characterization of Building

Blocks and Pyridostatin Analogues and FRET-Melting Profiles. See DOI: 10.1039/C2ob25830g Click Here for Additional Data File. *Org Biomol Chem* **2012**, *10* (32), 6537–6546. https://doi.org/10.1039/c2ob25830g.
(100) Escarcega, R. D.; Patil, A. A.; Moruno-Manchon, J. F.; Urayama, A.; Marrelli, S. P.; Kim, N.; Monchaud, D.; McCullough, L. D.; Tsvetkov, A. S. Pirh2-Dependent DNA Damage in Neurons Induced by the G-Quadruplex Ligand Pyridostatin. *Journal of Biological Chemistry* **2023**, *299* (10). https://doi.org/10.1016/j.jbc.2023.105157.
(101) Marchand, A.; Granzhan, A.; Iida, K.; Tsushima, Y.; Ma, Y.; Nagasawa, K.; Teulade-Fichou, M.-P.; Gabelica, V. Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* **2015**, *137* (2), 750–756. https://doi.org/10.1021/ja5099403.
(102) Verga, D.; Hamon, F.; Poyer, F.; Bombard, S.; Teulade-Fichou, M.-P. Photo-Cross-Linking Probes for Trapping G-Quadruplex DNA. *Angewandte Chemie International Edition* **2014**, *53* (4), 994–998. https://doi.org/10.1002/anie.201307413.
(103) Ungvarsky, J.; Plsikova, J.; Janovec, L.; Koval, J.; Mikes, J.; Mikesová, L.; Harvanova, D.; Fedorocko, P.; Kristian, P.; Kasparkova, J.; Brabec, V.; Vojtickova, M.; Sabolova, D.; Stramova, Z.; Rosocha, J.; Imrich, J.; Kozurkova, M. Novel Trisubstituted Acridines as Human Telomeric Quadruplex Binding Ligands. *Bioorganic Chemistry* **2014**, *57*, 13–29. https://doi.org/10.1016/j.bioorg.2014.07.010.
(104) Perrone, R.; Butovskaya, E.; Daelemans, D.; Palù, G.; Pannecouque, C.; Richter, S. N. Anti-HIV-1 Activity of the G-Quadruplex Ligand BRACO-19. *Journal of Antimicrobial Chemotherapy* **2014**, *69* (12), 3248–3258. https://doi.org/10.1093/jac/dku280.
(105) Debray, J.; Zeghida, W.; Jourdan, M.; Monchaud, D.; Dheu-Andries, M.-L.; Dumy, P.; Teulade-Fichou, M.-P.; Demeunynck, M. Synthesis and Evaluation of Fused Bispyrimidinoacridines as Novel Pentacyclic Analogues of Quadruplex-Binder BRACO-19. *Org. Biomol. Chem.* **2009**, *7* (24), 5219–5228. https://doi.org/10.1039/B912716J.
(106) Gunaratnam, M.; Greciano, O.; Martins, C.; Reszka, A. P.; Schultes, C. M.; Morjani, H.; Riou, J.-F.; Neidle, S. Mechanism of Acridine-Based Telomerase Inhibition and Telomere Shortening. *Biochemical Pharmacology* **2007**, *74* (5), 679–689. https://doi.org/10.1016/j.bcp.2007.06.011.
(107) Monchaud, D.; Granzhan, A.; Saettel, N.; Guédin, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. “One Ring to Bind Them All”—Part I: The Efficiency of the Macrocyclic Scaffold for G-Quadruplex DNA Recognition. *J Nucleic Acids* **2010**, *2010*, 525862. https://doi.org/10.4061/2010/525862.
(108) Nicoludis, J. M.; Barrett, S. P.; Mergny, J.-L.; Yatsunyk, L. A. Interaction of Human Telomeric DNA with N-Methyl Mesoporphyrin IX. *Nucleic Acids Research* **2012**, *40* (12), 5432. https://doi.org/10.1093/nar/gks152.
(109) Parkinson, G. N.; Ghosh, R.; Neidle, S. Structural Basis for Binding of Porphyrin to Human Telomeres,. *Biochemistry* **2007**, *46* (9), 2390–2397. https://doi.org/10.1021/bi062244n.
(110) Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human MYC Promoter. *Nat Chem Biol* **2005**, *1* (3), 167–173. https://doi.org/10.1038/nchembio723.
(111) Pillich, R. T.; Scarsella, G.; Galati, G.; Izzo, L.; Iacoangeli, A.; Castelli, M.; Risuleo, G. The Diimide Drug PIPER Has a Cytotoxic Dose-Dependent Effect In Vitro and Inhibits Telomere Elongation in HELA Cells. *Anticancer Research* **2005**, *25* (5), 3341–3346.

(112) Parkinson, G. N.; Cuenca, F.; Neidle, S. Topology Conservation and Loop Flexibility in Quadruplex–Drug Recognition: Crystal Structures of Inter- and Intramolecular Telomeric DNA Quadruplex–Drug Complexes. *Journal of Molecular Biology* **2008**, *381* (5), 1145–1156. https://doi.org/10.1016/j.jmb.2008.06.022.
(113) Doria, F.; Nadai, M.; Sattin, G.; Pasotti, L.; Richter, S. N.; Freccero, M. Water Soluble Extended Naphthalene Diimides as pH Fluorescent Sensors and G-Quadruplex Ligands. *Org. Biomol. Chem.* **2012**, *10* (19), 3830–3840. https://doi.org/10.1039/C2OB07006E.
(114) Nadai, M.; Doria, F.; Di Antonio, M.; Sattin, G.; Germani, L.; Percivalle, C.; Palumbo, M.; Richter, S. N.; Freccero, M. Naphthalene Diimide Scaffolds with Dual Reversible and Covalent Interaction Properties towards G-Quadruplex. *Biochimie* **2011**, *93* (8), 1328–1340. https://doi.org/10.1016/j.biochi.2011.06.015.
(115) Largy, E.; Granzhan, A.; Hamon, F.; Verga, D.; Teulade-Fichou, M.-P. Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes. In *Quadruplex Nucleic Acids*; Chaires, J. B., Graves, D., Eds.; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 2013; pp 111–177. https://doi.org/10.1007/128_2012_346.
(116) Xie, X.; Choi, B.; Largy, E.; Guillot, R.; Granzhan, A.; Teulade-Fichou, M.-P. Asymmetric Distyrylpyridinium Dyes as Red-Emitting Fluorescent Probes for Quadruplex DNA. *Chemistry – A European Journal* **2013**, *19* (4), 1214–1226. https://doi.org/10.1002/chem.201203710.
(117) Arthanari, H.; Basu, S.; Kawano, T. L.; Bolton, P. H. Fluorescent Dyes Specific for Quadruplex DNA. *Nucleic Acids Res* **1998**, *26* (16), 3724–3728.
(118) Ma, Y.; Iida, K.; Nagasawa, K. Topologies of G-Quadruplex: Biological Functions and Regulation by Ligands. *Biochemical and Biophysical Research Communications* **2020**, *531* (1), 3–17. https://doi.org/10.1016/j.bbrc.2019.12.103.
(119) Müller, D.; Saha, P.; Panda, D.; Dash, J.; Schwalbe, H. Insights from Binding on Quadruplex Selective Carbazole Ligands. *Chemistry – A European Journal* **2021**, *27* (50), 12726–12736. https://doi.org/10.1002/chem.202101866.
(120) Dickerhoff, J.; Dai, J.; Yang, D. Structural Recognition of the MYC Promoter G-Quadruplex by a Quinoline Derivative: Insights into Molecular Targeting of Parallel G-Quadruplexes. *Nucleic Acids Research* **2021**, *49* (10), 5905–5915. https://doi.org/10.1093/nar/gkab330.
(121) Ritson, D. J.; Moses, J. E. A Fragment Based Click Chemistry Approach towards Hybrid G-Quadruplex Ligands: Design, Synthesis and Biophysical Evaluation. *Tetrahedron* **2012**, *68* (1), 197–203. https://doi.org/10.1016/j.tet.2011.10.066.
(122) Hamon, F.; Largy, E.; Guédin-Beaurepaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J.-L.; Riou, J.-F.; Nguyen, C.-H.; Teulade-Fichou, M.-P. An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies. *Angewandte Chemie International Edition* **2011**, *50* (37), 8745–8749. https://doi.org/10.1002/anie.201103422.
(123) Lin, C.; Wu, G.; Wang, K.; Onel, B.; Sakai, S.; Shao, Y.; Yang, D. Molecular Recognition of the Hybrid-2 Human Telomeric G-Quadruplex by Epiberberine: Insights into Conversion of Telomeric G-Quadruplex Structures. *Angewandte Chemie International Edition* **2018**, *57* (34), 10888–10893. https://doi.org/10.1002/anie.201804667.
(124) McQuaid, K. T.; Takahashi, S.; Baumgaertner, L.; Cardin, D. J.; Paterson, N. G.; Hall, J. P.; Sugimoto, N.; Cardin, C. J. Ruthenium Polypyridyl Complex Bound to a Unimolecular Chair-Form G-Quadruplex. *J. Am. Chem. Soc.* **2022**, *144* (13), 5956–5964. https://doi.org/10.1021/jacs.2c00178.

(125) Largy, E.; Teulade-Fichou, M.-P. Screening for Quadruplex Binding Ligands: A Game of Chance? **2012**. https://doi.org/10.1039/9781849736954-00248.
(126) Beaufour, M.; Ginguené, D.; Le Meur, R.; Castaing, B.; Cadene, M. Liquid Native MALDI Mass Spectrometry for the Detection of Protein-Protein Complexes. *J Am Soc Mass Spectrom* **2018**, *29* (10), 1981–1994. https://doi.org/10.1007/s13361-018-2015-x.
(127) Sturiale, L.; Palmigiano, A.; Silipo, A.; Knirel, Y. A.; Anisimov, A. P.; Lanzetta, R.; Parrilli, M.; Molinaro, A.; Garozzo, D. Reflectron MALDI TOF and MALDI TOF/TOF Mass Spectrometry Reveal Novel Structural Details of Native Lipooligosaccharides. *Journal of Mass Spectrometry* **2011**, *46* (11), 1135–1142. https://doi.org/10.1002/jms.2000.
(128) Beverly, M.; Hartsough, K.; Machemer, L. Liquid Chromatography/Electrospray Mass Spectrometric Analysis of Metabolites from an Inhibitory RNA Duplex. *Rapid Communications in Mass Spectrometry* **2005**, *19* (12), 1675–1682. https://doi.org/10.1002/rcm.1972.
(129) Noll, B.; Seiffert, S.; Vornlocher, H.-P.; Roehl, I. Characterization of Small Interfering RNA by Non-Denaturing Ion-Pair Reversed-Phase Liquid Chromatography. *Journal of Chromatography A* **2011**, *1218* (33), 5609–5617. https://doi.org/10.1016/j.chroma.2011.06.057.
(130) Schalley, C. A. *Analytical Methods in Supramolecular Chemistry*, 2., compl. rev. and enl. ed.; Wiley-VCH: Weinheim, 2012.
(131) Schmidt, A. Untersuchungen des Elektrospray-Ionisierungsprozesses zum massenspektrometrischen Nachweis nicht-kovalenter Komplexe. Dissertation, Johann Wolfgang Goethe Universität, Frankfurt am Main, 2000.
(132) Kebarle, P. A Brief Overview of the Present Status of the Mechanisms Involved in Electrospray Mass Spectrometry. *Journal of Mass Spectrometry* **2000**, *35* (7), 804–817. https://doi.org/10.1002/1096-9888(200007)35:7<804::aid-jms22>3.0.co;2-q.
(133) Largy, E.; König, A.; Ghosh, A.; Ghosh, D.; Benabou, S.; Rosu, F.; Gabelica, V. Mass Spectrometry of Nucleic Acid Noncovalent Complexes. *Chem. Rev.* **2021**. https://doi.org/10.1021/acs.chemrev.1c00386.
(134) Khristenko, N.; Rosu, F.; Largy, E.; Haustant, J.; Mesmin, C.; Gabelica, V. Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism. *J. Am. Chem. Soc.* **2023**, *145* (1), 498–506. https://doi.org/10.1021/jacs.2c10762.
(135) Bončina, M.; Vesnaver, G.; Chaires, J. B.; Lah, J. Unraveling the Thermodynamics of the Folding and Interconversion of Human Telomere G-Quadruplexes. *Angewandte Chemie International Edition* **2016**, *55* (35), 10340–10344. https://doi.org/10.1002/anie.201605350.
(136) Ghosh, D.; Rosu, F.; Gabelica, V. Negative Electrospray Supercharging Mechanisms of Nucleic Acid Structures. *Anal. Chem.* **2022**. https://doi.org/10.1021/acs.analchem.2c03187.
(137) Wang, W.; Kitova, E. N.; Klassen, J. S. Nonspecific Protein−Carbohydrate Complexes Produced by Nanoelectrospray Ionization. Factors Influencing Their Formation and Stability. *Anal. Chem.* **2005**, *77* (10), 3060–3071. https://doi.org/10.1021/ac048433y.
(138) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564–8573. https://doi.org/10.1021/acs.analchem.9b01838.
(139) Baker, E. S.; Manard, M. J.; Gidden, J.; Bowers, M. T. Structural Analysis of Metal Interactions with the Dinucleotide Duplex, dCG·dCG, Using Ion Mobility Mass

Spectrometry. *J. Phys. Chem. B* **2005**, *109* (11), 4808–4810. https://doi.org/10.1021/jp0501190.
(140) Warnke, S.; Seo, J.; Boschmans, J.; Sobott, F.; Scrivens, J. H.; Bleiholder, C.; Bowers, M. T.; Gewinner, S.; Schöllkopf, W.; Pagel, K.; von Helden, G. Protomers of Benzocaine: Solvent and Permittivity Dependence. *J. Am. Chem. Soc.* **2015**, *137* (12), 4236–4242. https://doi.org/10.1021/jacs.5b01338.
(141) del Villar-Guerra, R.; Gray, R. D.; Chaires, J. B. Characterization of Quadruplex DNA Structure by Circular Dichroism. *Current Protocols in Nucleic Acid Chemistry* **2017**, *68* (1), 17.8.1-17.8.16. https://doi.org/10.1002/cpnc.23.
(142) Improta, R. Quantum Mechanical Calculations Unveil the Structure and Properties of the Absorbing and Emitting Excited Electronic States of Guanine Quadruplex. *Chemistry – A European Journal* **2014**, *20* (26), 8106–8115. https://doi.org/10.1002/chem.201400065.
(143) Lech, C. J.; Heddi, B.; Phan, A. T. Guanine Base Stacking in G-Quadruplex Nucleic Acids. *Nucleic Acids Research* **2013**, *41* (3), 2034–2046. https://doi.org/10.1093/nar/gks1110.
(144) del Villar-Guerra, R.; Trent, J. O.; Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angewandte Chemie International Edition* **2018**, *57* (24), 7171–7175. https://doi.org/10.1002/anie.201709184.
(145) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135* (13), 5017–5028. https://doi.org/10.1021/ja310251r.
(146) Agarwal, T.; Kumar, S.; Maiti, S. Unlocking G-Quadruplex: Effect of Unlocked Nucleic Acid on G-Quadruplex Stability. *Biochimie* **2011**, *93* (10), 1694–1700. https://doi.org/10.1016/j.biochi.2011.05.036.
(147) Mergny, L., Jean-Louis; Lacroix. Analysis of Thermal Melting Curves. *Oligonucleotides*, 2003, 515–537.
(148) Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Characterization of Structure and Stability of Long Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* **2006**, *128* (48), 15461–15468. https://doi.org/10.1021/ja064536h.
(149) Olsen, C. M.; Lee, H.-T.; Marky, L. A. Unfolding Thermodynamics of Intramolecular G-Quadruplexes: Base Sequence Contributions of the Loops. *J. Phys. Chem. B* **2009**, *113* (9), 2587–2595. https://doi.org/10.1021/jp806853n.
(150) Mergny, J. L.; Lacroix, L. UV Melting of G-Quadruplexes. *Curr Protoc Nucleic Acid Chem*, 2009, *Chapter 17*, Unit 17.1. https://doi.org/10.1002/0471142700.nc1701s37.
(151) Puglisi, J. D.; Tinoco, I. [22] Absorbance Melting Curves of RNA. In *Methods in Enzymology*; RNA Processing Part A: General Methods; Academic Press, 1989; Vol. 180, pp 304–325. https://doi.org/10.1016/0076-6879(89)80108-9.
(152) Campbell, N.; Collie, G. W.; Neidle, S. Crystallography of DNA and RNA G-Quadruplex Nucleic Acids and Their Ligand Complexes. *Current Protocols in Nucleic Acid Chemistry* **2012**, *50* (1), 17.6.1-17.6.22. https://doi.org/10.1002/0471142700.nc1706s50.
(153) Timsit, Y.; Moras, D. [20] Crystallization of DNA. In *Methods in Enzymology*; DNA Structures Part A: Synthesis and Physical Analysis of DNA; Academic Press, 1992; Vol. 211, pp 409–429. https://doi.org/10.1016/0076-6879(92)11022-B.
(154) Campbell, N. H.; Parkinson, G. N. Crystallographic Studies of Quadruplex Nucleic Acids. *Methods* **2007**, *43* (4), 252–263. https://doi.org/10.1016/j.ymeth.2007.08.005.

(155) Parkinson, G. N.; Collie, G. W. X-Ray Crystallographic Studies of G-Quadruplex Structures. In *G-Quadruplex Nucleic Acids: Methods and Protocols*; Yang, D., Lin, C., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2019; pp 131–155. https://doi.org/10.1007/978-1-4939-9666-7_8.
(156) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. Not so Crystal Clear: The Structure of the Human Telomere G-Quadruplex in Solution Differs from That Present in a Crystal. *Nucleic Acids Research* **2005**, *33* (14), 4649–4659. https://doi.org/10.1093/nar/gki782.
(157) Haslauer, K. E.; Hemmler, D.; Schmitt-Kopplin, P.; Heinzmann, S. S. Guidelines for the Use of Deuterium Oxide (D_2O) in ^1H NMR Metabolomics. *Anal. Chem.* **2019**, *91* (17), 11063–11069. https://doi.org/10.1021/acs.analchem.9b01580.
(158) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods*, 2012, *57*, 11–24. https://doi.org/10.1016/j.ymeth.2012.05.003.
(159) Lin, C.; Dickerhoff, J.; Yang, D. NMR Studies of G-Quadruplex Structures and G-Quadruplex-Interactive Compounds. In *G-Quadruplex Nucleic Acids: Methods and Protocols*; Yang, D., Lin, C., Eds.; Methods in Molecular Biology; Springer: New York, NY, 2019; pp 157–176. https://doi.org/10.1007/978-1-4939-9666-7_9.
(160) Ghosh, A.; Trajkovski, M.; Teulade-Fichou, M.-P.; Gabelica, V.; Plavec, J. Phen-DC3 Induces Refolding of Human Telomeric DNA into a Chair-Type Antiparallel G-Quadruplex through Ligand Intercalation. *Angewandte Chemie International Edition* **2022**, *61* (40), e202207384. https://doi.org/10.1002/anie.202207384.
(161) Hänsel, R.; Foldynová-Trantírková, S.; Dötsch, V.; Trantírek, L. Investigation of Quadruplex Structure Under Physiological Conditions Using In-Cell NMR. In *Quadruplex Nucleic Acids*; Chaires, J. B., Graves, D., Eds.; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 2013; pp 47–65. https://doi.org/10.1007/128_2012_33

CHAPTER 1: INVESTIGATING THE SECONDARY STRUCTURES FORMED BY TRIVALENT LANTHANIDE IONS AND G-QUADRUPLEX DNA SEQUENCES.

Motivation

Cations stabilize the G-quadruplex structure in several ways: 1) coordination bonds with O6 carbonyl atoms. The interaction is comparable to alkali metal ions being coordinated by crown ethers. 2) reduction of charge density within the compact G-quadruplex core.[1] The main factor determining whether a cation is deemed 'fit' or 'unfit' for G-quadruplex stabilization is the ionic radius. When the ion is too small, it cannot coordinate enough guanine bases and its solvation forces are stronger. When the ion is too big, it does not fit in the space between two G-tetrads. Another important property of any ion is its charge, which often remains unacknowledged in the context of G-quadruplex cations, because high ionic strength in solution is sufficient to regulate DNA charge density (i.e., no additional positive charges needed). And yet, the divalent cations Sr^{2+} and Pb^{2+} form more stable G-quadruplexes than the monovalent cations K^{+} or Na^{+}.[2–5] We are curious whether this trend continues and trivalent cations stabilize G-quadruplexes even more compared to Sr^{2+} or Pb^{2+}. Ionic radius is important here, because most trivalent cations (e.g. Fe^{3+}) are too small to be 'fit' for G-quadruplex stabilization.[6–8] The most promising candidates in terms of ionic radius are the lanthanides (Ln^{3+}). For reference, in the VIII coordination state, Na^{+} has an ionic radius of 1.18 Å (Shannon[9]), while the lanthanides range from 1.16 Å (La^{3+}) to 0.98 Å (Yb^{3+}). Even more promising is the fact that Ln^{3+} ions form clusters containing 8 or 12 guanines, which is consistent with a 2- or 3-tetrad G-quadruplex.[10] The literature provides some hints for specific Ln^{3+}-guanine interactions but with little structural investigation. We want to use our expertise in structural biophysics to investigate the secondary structures formed by G-rich DNA sequences in the presence of Ln^{3+} cations. The two main objectives are 1) provide evidence that proves or disproves the formation of G-quadruplex in the presence of trivalent cations 2) assess the stability of those secondary structures compared to G-quadruplexes folded in a monovalent (K^{+}) or a divalent (Sr^{2+}) cation.

State of the art

Lanthanide cations and single nucleotides form clusters whose stoichiometries are multiples of four, which is supporting evidence for the formation of G-tetrads (Figure 35).[10]

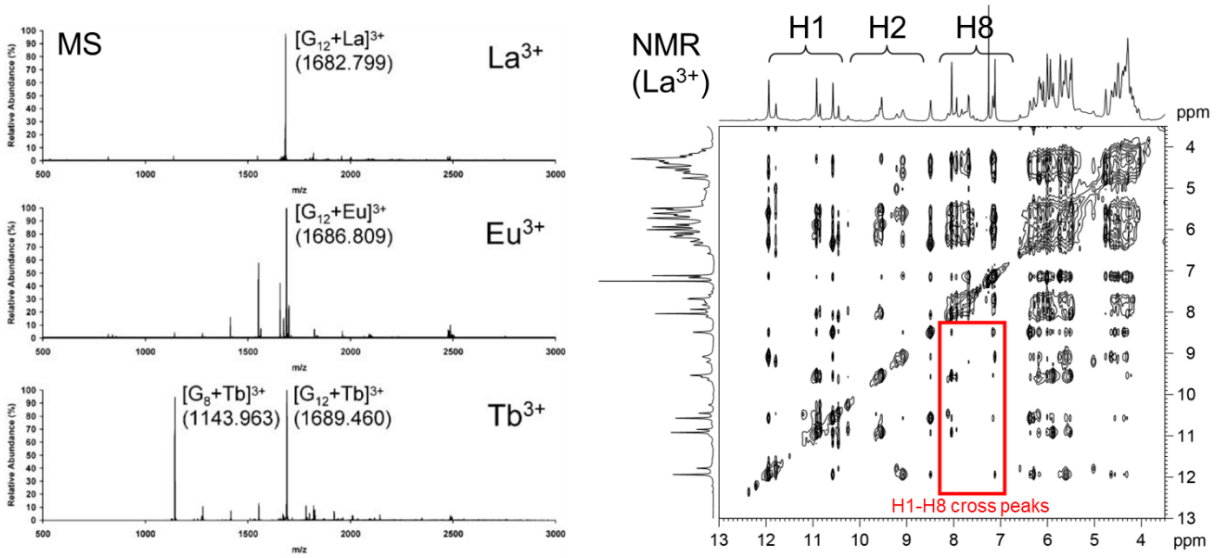


Figure 35 Clusters of 2',3',5'-triacetylguanosine and lanthanide cations were prepared "using a solid-liquid extraction method in $CHCl_3$" [sic] and obtained as a solid after $CHCl_3$ evaporation. ***Left****: MS spectra of clusters containing different templating cations. Sample was dissolved in nitromethane (concentration not provided) and measured on a Qq-TOF instrument in positive mode.* ***Right****: ^1H-^1H-NOE spectrum of the lanthanum-triacetylguanosine cluster in $CDCl_3$ at 268.2 K on a Bruker Advance 600 MHz instrument. Adapted with permission from Wu et al. Chem. Comm. 2007, 41, 4286-4288.[10] Copyright 2007 RSC Publishing.*

The NMR was recorded in aprotic solvent ($CDCl_3$), so with the absence of solvent exchange, the detection of H1 protons is no proof of G-tetrad formation. The main piece of evidence is the proximity of H1 and H8 protons, proven by the correlation peaks that are highlighted in the NOESY spectrum. According to the mass spectra, one lanthanide cation can coordinate 8 or 12 guanine nucleotides. The coordination number 8 is typical for G-quadruplex formation, while 12 is too high, unless there were one ion for every three tetrads, which has thus far never been described.

There is experimental evidence of Eu^{3+} and Tb^{3+} specifically targeting guanine over 1) other nucleobases and 2) phosphates at the nucleotide backbone (Figure 36).[11] The addition of double-stranded DNA causes a slight decline in Tb^{3+} emission, while it increases by one order of magnitude with the single-stranded variant. To understand this observation, we have to discuss how the luminescence of lanthanides is tightly linked to their hydration state. The excited state of lanthanide ions undergoes relaxation through the O-H vibrational states of coordinated water molecules.[11] Thus, the lanthanide emission is enhanced when water molecules are removed from the coordination sphere. The number of water molecules in the coordination sphere can also be quantified by measuring the difference of luminescence lifetime in H_2O vs. D_2O. This method utilizes the fact that lanthanide ions undergo relaxation through O-H vibrational states, but not O-D vibrational states.[12]

Tb^{3+} keeps most of its hydration sphere when targeting the backbone, but loses most of it when targeting the nucleobases.[13] This hypothesis was validated by quantifying the number of H_2O molecules coordinated by Tb^{3+} using the introduced luminescence lifetime method.[12] At pH 7 and 25°C, one Tb^{3+} ion coordinates 9 water molecules on average. Upon adding double-stranded DNA (ca. 2000 base pairs, extracted from

salmon), one water molecule was removed. Since the Tb^{3+} in solution can only access the DNA backbone, the Tb^{3+} appears to lose one H_2O molecule from backbone coordination. To make the nucleobases solvent-accessible, the duplex was denatured at 90°C and cooled down rapidly. Now the Tb^{3+} ions lost 6 water molecules, when the single-stranded species was added.[13] In conclusion, luminescence experiments can differentiate backbone from nucleobase binding due to different degrees of Tb^{3+} dehydration.[13,14]

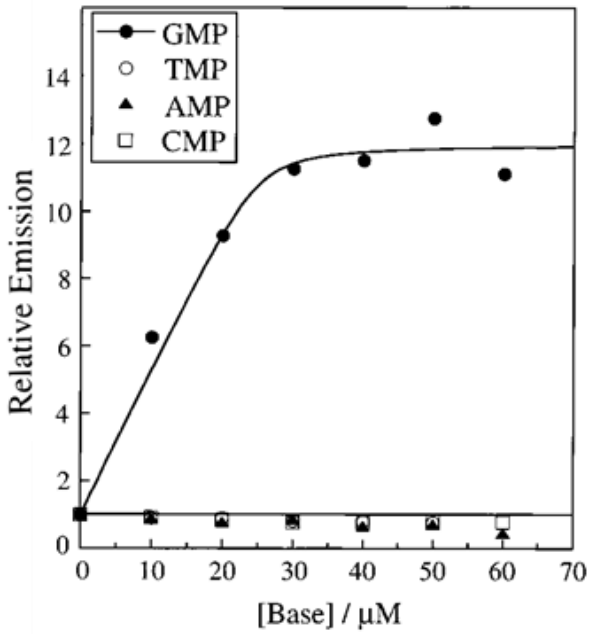


Figure 36 Relative emission of 25 µM Tb^{3+} as a function of deoxymonophosphate concentration. Samples contain 50 mM NaCl and 2 mM Tris buffer (pH 7.5). Excitation: 260 nm; Emission: 545 nm. Reprinted with permission from Fu and Turro JACS 1999, 121 (1), 1-7.[11] Copyright 1999 JACS.

Back to Figure 36, which compares the Tb^{3+} emission with different mononucleotides. For thymine, adenine and cytosine, the lack of change in emission indicates that Tb^{3+}, if anything, targets the phosphate group over the nucleobase. Most importantly, GMP is the only nucleotide where Tb^{3+} targets the nucleobase, as indicated by the shift in luminescence when GMP is added. Tb^{3+} and Eu^{3+} have both been characterized as guanine-specific probes to detect G-rich DNA.[15–17] Modifications of the N1, N2 or N7 atom resulted in partial losses of Tb^{3+} fluorescence, indicating that the guanine N-atoms are important π-donors for the Ln^{3+}-guanine interaction.[18]

The guanine specificity and the clusters with 8/12 nucleobases are solid grounds to suspect the formation of G-quadruplexes. In slight excess of Tb^{3+}, the telomeric sequence 21G displays a CD signature resembling an antiparallel G-quadruplex (Figure 37, left). DNA aggregation occurs in 10-fold or higher excess of Tb^{3+} and has been characterized by 1) high molecular weight bands appearing in PAGE[19] 2) loss of CD signature[20] and 3) visible Tb-induced DNA precipitation[21]. The most plausible explanation is that Tb^{3+} first targets the guanines, but once the guanines are saturated, they attach to the phosphates at the backbone. Since lanthanides carry three positive charges, they easily neutralize the negative charges on the DNA strand, reducing solubility and charge repulsion between DNA strands, thus promoting aggregation. This effect is strong enough for 50 µM Tb^{3+} to disrupt 2 µM of G-quadruplex in the presence of 50 mM Na^+ or 2 mM K^+. Note that there are 2-3 orders of magnitude between the Tb^{3+} and the Na^+/K^+ concentration.[20]

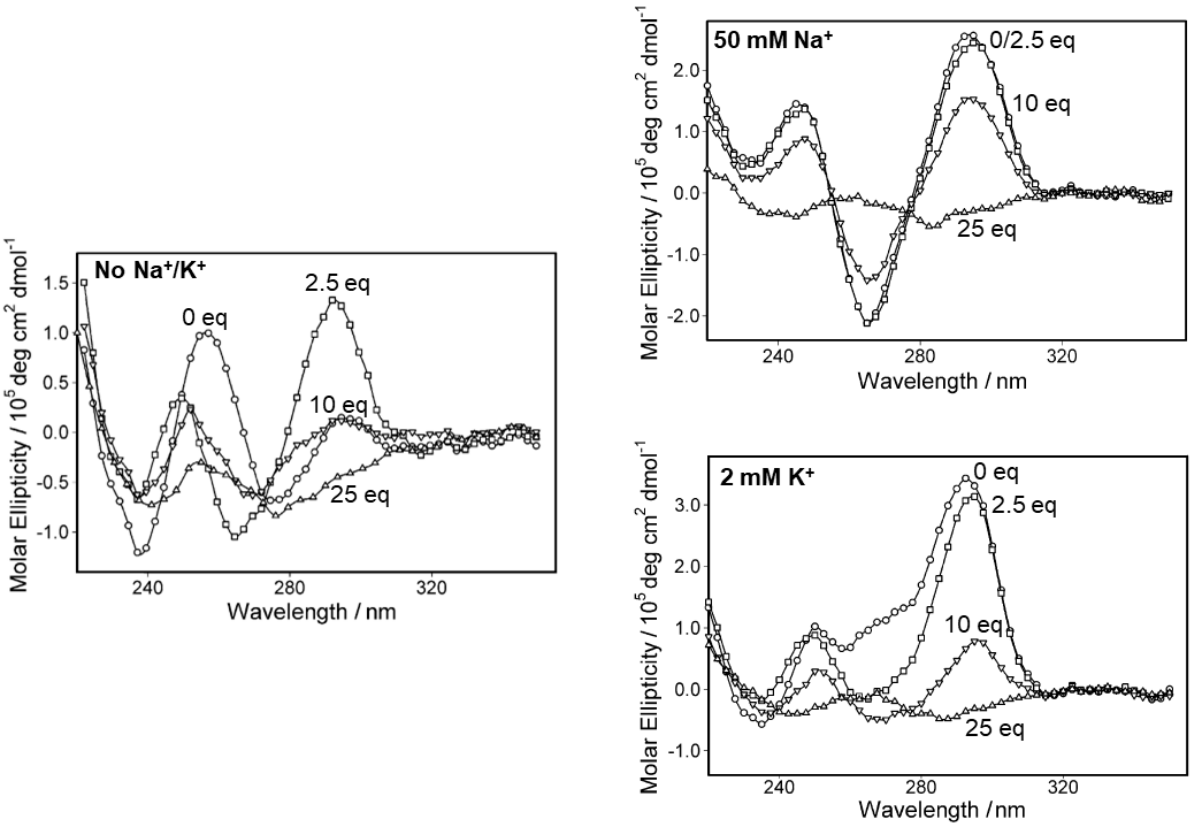


*Figure 37 CD spectra of 2 µM 21G (d**G₃**TTA**G₃**TTA**G₃**TTA**G₃**) with four different equivalents of Tb^{3+}. Unfolded strand on the left panel, pre-folded G-quadruplex in Na^+/K^+ on the right panel. T = 25°C Adapted with permission from Juskowiak et al. J. Inorg. Biochem. 2007, 101 (4), 678-685.[20] Copyright 2007 Elsevier.*

This experiment was also done in reverse, trying to displace a pre-folded structure in Tb^{3+} (Figure 38). In low Tb^{3+} excess, presumably before Tb^{3+} induces DNA aggregation, Na^+ and K^+ can displace Tb^{3+} from the DNA strand. In high Tb^{3+} excess, the aggregation becomes irreversible and the effect of Na^+/K^+ on Tb^{3+} luminescence is comparable to Li^+ and thus unlikely to be related to G-quadruplex formation. Only Mg^{2+} was a significant competitor in both conditions, probably because Mg^{2+} and Tb^{3+} are similar in terms of charge-to-size ratio.

In summary, Tb^{3+} and K^+/Na^+ compete for guanine interaction, but the Tb-induced DNA aggregation is an irreversible process that should be avoided by keeping the Tb:DNA ratio below 10:1. Despite several demonstrations of guanine specificity there is little knowledge on the secondary structure formed by lanthanides and G-rich sequences and there is no explicit proof of G-quadruplex formation.

Mass spectrometry will be effective in identifying specifically bound Ln^{3+} ions and determining their stoichiometry. We will be more methodological and comprehensive in our selection of lanthanides and DNA sequences. Instead of just picking 1 or 2 random lanthanides, we will better cover the range of the f-block. Our sequences will feature G-quadruplex sequences of different topology and molecularity, as well as controls of varying G-richness to better assess the Ln^{3+} binding mode.

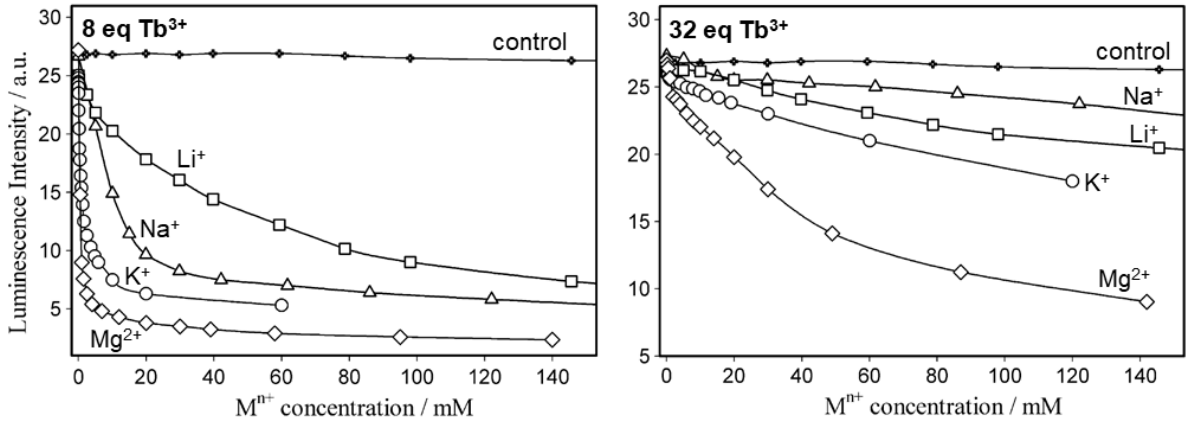


*Figure 38 Luminescence titration of different cations onto 1 µM 21G (d**G₃**TTA**G₃**TTA**G₃**TTA**G₃**) equilibrated in left: 8 µM Tb^{3+} Right: 32 µM Tb^{3+}. T = 25°C. Adapted with permission from Juskowiak et al. J. Inorg. Biochem. 2007, 101 (4), 678-685.[20] Copyright 2007 Elsevier.*

Materials and Methods

Selection of DNA sequences and lanthanides

The DNA sequences used in this project are summarized in Table 2. Since phosphates and nucleobases are competing targets, we aim to keep the number of phosphates around 20, while varying the G-richness. For that reason, TG4T and G4T4G4 received a 5' phosphate modification (while natural oligonucleotides normally have a phosphate group in the 5' position, our synthetic ones just have an -OH group in 5').

Table 2 Overview of DNA sequences used in this study

Name	Sequence	Bases	Structure in K^+ [1]	G-content
24TTG	dTT**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**A	24	hybrid-1 GQ	50%
2LBY	dTA**G$_3$**A**G$_3$**TA**G$_3$**A**G$_3$**T	19	parallel GQ	63%
2LK7	dTT**G$_3$**T**G$_3$**T**G$_3$**T**G$_3$**T	18	parallel GQ	67%
G4T4G4	(d**G$_4$**T$_4$**G$_4$**)$_2$	24 (2×12)	antiparallel GQ	67%
TG4T	(dT**G$_4$**T)$_4$	24 (4×6)	parallel GQ	67%
DK66	(dC**G**C**G**AATTC**G**C**G**)$_2$	24 (2×12)	ds-DNA/hairpin	33%
24nonG4	dT**G$_3$**AT**G**C**G**ACA**G**A **G**A**GG**AC**G$_3$**A	24	ss-DNA	50%
ss24	dT**G**CCAT**G**CTACT**G** A**G**AT**G**AC**G**CTA	24	ss-DNA	25%
T24	dT$_{24}$	24	ss-DNA	0%
A24	dA$_{24}$	24	ss-DNA	0%

The DNA sequences were ordered from Eurogentec and dissolved in UPLC grade water (Biosolve). The stock solutions were annealed at 85°C and then washed 4 times in 500 mM ammonium acetate solution, then 6 times with water using Amicon centrifugal filter units. The DNA concentration of the desalted stock solutions was determined on a Uvikon XS spectrophotometer, with absorption coefficients being calculated based on nearest-neighbor method.

Lanthanide salts were ordered from Sigma Aldrich ($LaCl_3 \cdot 7H_2O$, $EuCl_3 \cdot 6H_2O$, $TbCl_3 \cdot 6H_2O$, $YbCl_3 \cdot 6H_2O$) and weighed in to create 1 M stock solutions in water.

Instrumental analysis and data treatment

Circular dichroism (CD) is a spectroscopic method sensitive to chirality. CD spectra allow us to track changes in the secondary structure of the analyte. CD samples contain 5 or 10 µM DNA, 0-100 mM metal chloride and 50 mM trimethylammonium acetate (TMAA, pH 6.8) electrolyte to provide ionic strength. The trimethylammonium ion is too large to fit inside a G-quadruplex and will therefore not compete with K^+ for G-quadruplex formation. The CD samples were placed in optical cuvettes (Hellma, l = 10 mm) and measured on a Jasco J-815 spectrophotometer at a range from 220 to 350 nm at 50 nm/min scanning speed, 0.2 nm data pitch, 2 nm bandwidth, 2 s data integration time, 22°C in the sample holder, 3 acquisitions. Samples for blank correction were prepared without DNA and measured in the same manner. For titration experiments, cation stock solutions were added stepwise to the sample and left to equilibrate for ca. 10 minutes.

[1] GQ = G-quadruplex, ds = double-stranded. ss = single-stranded

CD melting experiments were used to separate signal contributions of different conformers. The melting was done in a temperature range from 4-90°C, starting at 90°C with the cooling ramp. The temperature gradient was 1°C/min, but was halted every 1°C for spectra acquisition from 220-350 nm with a scan rate of 200 nm/min (1 nm data pitch, 2 nm bandwidth, 2 s data integration time, 1 acquisition). The global temperature ramp is the average of the temperature gradient (1 °C/min) and the spectra acquisition time, during which the temperature gradient is halted (0.65 min per spectrum) and lies between 0.20-0.25 °C/min.

MS can determine the exact number of Tb^{3+} ions bound to the DNA strand, since the adducts are separated by mass.[22] We carried out an MS titration experiment to identify specific Tb^{3+} adducts. Tb^{3+} was chosen over Eu^{3+} because it is monoisotopic and thus generates less isotopic peaks, ensuring a better S/N ratio and making the mass spectra easier to read overall. 10 µM of DNA were doped with 0/5/10/20/50 µM $TbCl_3$ and 50 mM TMAA in H_2O and directly injected the sample into an Exactive Orbitrap Mass Analyzer (ESI(-)-FTMS). For ion mobility studies, we injected the sample into an Agilent 6560 IMS-QTOF. Ion mobility data processing was done as presented in the thesis introduction.

Under the supervision of Bikash Swain and with permission from Mikayel Aznauryan, luminescence experiments were performed on a FS5 Spectrofluorometer (Edinburgh Instruments) equipped with an SC25 temperature-controlled sample holder. The excitation wavelength was 290 nm (taken from Juskowiak *et al.*) and emission spectra were measured in a range of 400-650 nm at 20°C with 1.5 nm slits. Samples contained 10 µM DNA (20 µM for DK66), 50 µM $TbCl_3$ and 50 mM TMAA. A blank was measured in absence of DNA.

To investigate the formation of G-tetrads, Cameron Mackereth performed an NMR titration experiment where Tb^{3+} is added to unfolded 24TTG and 24nonG4 strand. NMR samples contained 200 µM DNA and 50 mM TMAA (pH 6.8) in 90/10 H_2O/D_2O and were measured on a Bruker Avance NEO (700 MHz) at 298 K. The experiment was prematurely stopped after increasing the Tb^{3+} concentration from 0.5 mM to 1 mM due to visible DNA precipitation and the inability to complete the shimming process.

Results and Discussion

Size matters: Not all lanthanides bind equally

Similar to the experiment done by Juskowiak *et al.* (Figure 37), we performed CD titrations where we added lanthanide cation to a telomeric G-quadruplex sequence (Figure 39). For the sake of simplicity, only four titration points are shown. The full dataset is deposited in the supporting information (Figure S1).

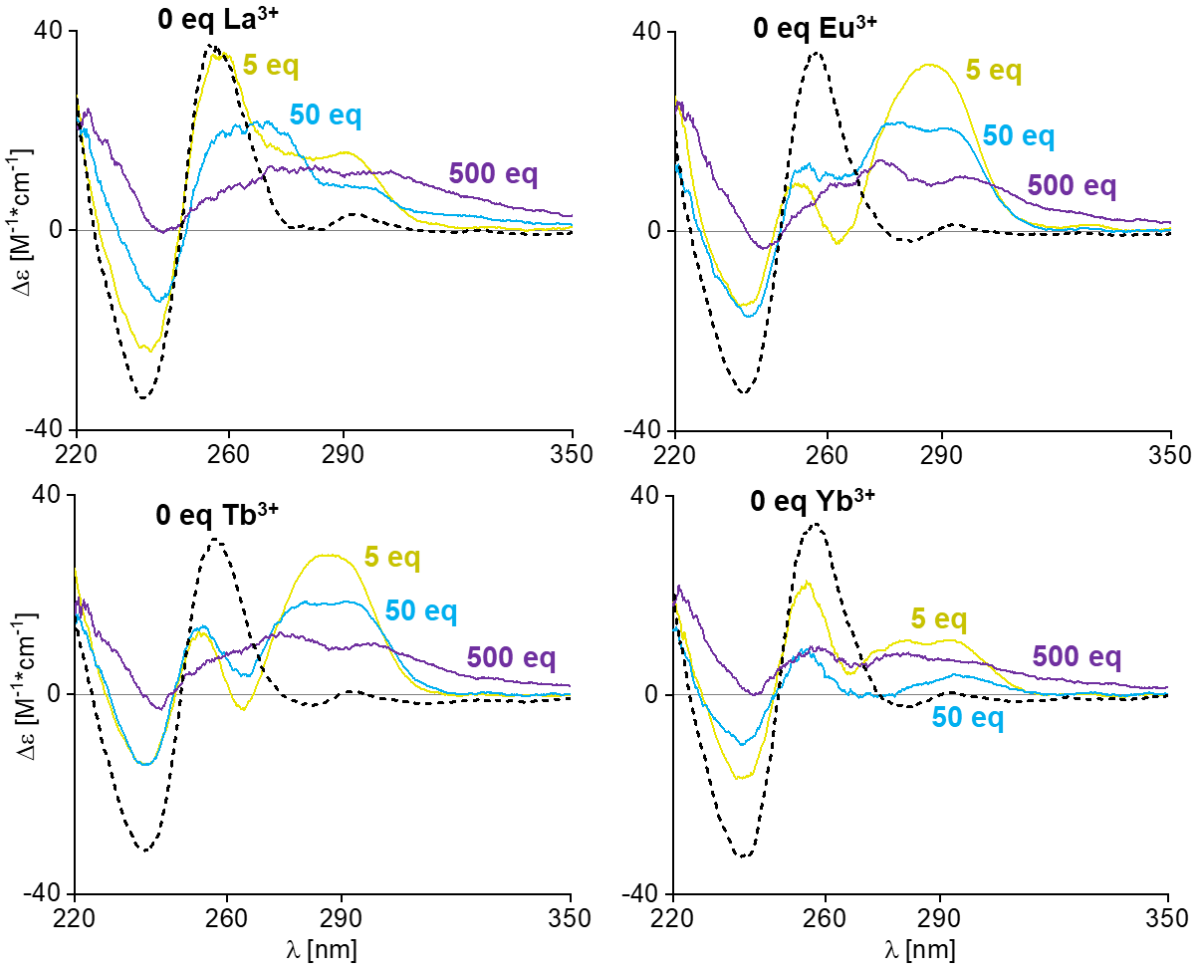


*Figure 39 CD titrations of 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A) with increasing cation equivalents (1 eq = 10 µM). Samples contain 10 µM DNA, 50 mM TMAA (pH 6.8), 0-10 mM metal chloride.*

Instead of just Tb^{3+}, we picked four lanthanides from different positions in the periodic table: lanthanum (left), europium (middle), terbium (middle) and ytterbium (right). The ionic radii of lanthanide cations significantly decrease from left to right. This phenomenon is known as lanthanide contraction.[23] Figure 39 shows that size matters, as the four lanthanides are not behaving the same way. The results for Tb^{3+} and Eu^{3+} are consistent with those that Juskowiak *et al.* have demonstrated previously: Upon adding 5-10 equivalents of Tb^{3+}/Eu^{3+}, the CD spectrum shifts to something resembling the signature of an antiparallel G-quadruplex. At higher concentrations, the CD signature decreases, indicating a disruption of the secondary structure, probably due to DNA aggregation/precipitation. Trivalent cations (Al^{3+} and Fe^{3+}) are used as coagulation agents in water treatment, causing the aggregation of colloidal particles by neutralizing their surface charges.[24] It is reasonable to assume that lanthanide ions would have the same effect on solubilized DNA.

La^{3+} and Yb^{3+} are different. Instead of a characteristic CD signature, there is only a continuous decrease of CD signal. There could still be a secondary structure, but if the cation affinity is too low, the required cation concentration falls into a range where DNA aggregation takes precedence. The La^{3+} dataset shows hints of a characteristic band at 260-270 nm. The smallest cation among the four, Yb^{3+}, was the most effective in reducing CD signal. These are new, unpublished results highlighting that whatever secondary structure forms between lanthanides and telomeric G-quadruplex DNA is specific to Eu^{3+} and Tb^{3+}, but not La^{3+} or Yb^{3+}. Juskowiak *et al.* argue that judging by the CD signature we are dealing with an antiparallel G-quadruplex. We formulate the counter-argument that if an antiparallel G-quadruplex folds in Eu^{3+}/Tb^{3+}, it should also fold in La^{3+}. G-quadruplex stabilizing cations fall in a size range somewhere from 1.10 to 1.60 Å. Lanthanides are on the lower end of that scale. Yb^{3+} (0.98 Å, $VIII^{9}$) is probably too small to coordinate multiple guanines. La^{3+} is the largest lanthanide cation (1.16 Å, $VIII^{9}$) and approximately the size of Na^{+} (1.18 Å, $VIII^{9}$). Therefore, the absence of the 'antiparallel' CD signature in our La^{3+} dataset contradicts the idea that the structure in Eu^{3+}/Tb^{3+} is an antiparallel G-quadruplex.

We replicated the experiment with the parallel G-quadruplex sequences 2LBY (Figure S2) and 2LK7 (Figure S3). All titration datasets start with a defined parallel CD signature, with Δε values of 130-150 at 260 nm. Adding 1-10 equivalents of La^{3+}/Eu^{3+}/Tb^{3+} (but not Yb^{3+}) slightly raises the band at 260 nm, until disruption commences at higher equivalents. We suspect that the 'unfolded' strand and the Ln^{3+}-DNA complex both have a characteristic band at 260 nm, which makes the CD spectra convoluted. Some mass spectra of 2LBY/2LK7 in La^{3+}/Eu^{3+}/Tb^{3+} would be really helpful to address these suspicions, but sadly that did not happen. However, we have mass spectra that proof 2LK7 is partially folded into a G-quadruplex with 2 K^{+} (Figure S4), while 2LBY is not (Figure S5). We assume that the desalting procedure was not fully effective on 2LK7, as K^{+} ions bind to parallel G-quadruplex DNA with K_D values in the low µM range,[25] which makes it very difficult to remove K^{+} contaminations. 2LK7 is one of the most stable G-quadruplexes and often called T95-2T because it has a T_M above 95°C in 100 mM K^{+}. We intentionally picked 2LBY, because it is the least stable parallel G-quadruplex in our panel (T_M = 43°C in 1 mM K^{+})[26] and would therefore be the easiest one to desalt properly.

We performed titrations in Li^{+} as a control experiment, showing no change in CD signature upon adding Li^{+} (Figures S1-S3). This validates that the effects on CD signature are correlated to the physicochemical nature of the lanthanides and not just correlated to ionic strength.

Tb^{3+} targets single-stranded G-rich sequences

We discarded La^{3+} and Yb^{3+} to focus on the secondary structure stabilized in Eu^{3+} and Tb^{3+}. Since the titration datasets for Eu^{3+}/Tb^{3+} look very similar, we stuck with Tb^{3+}, because we already picked it for mass spectrometry analysis.

As described in the introduction, an increase in Tb^{3+} luminescence signifies nucleobase interaction, while an insignificant shift indicates backbone binding or no binding at all. Figure 40 displays the Tb^{3+} luminescence emission spectra of 6 different 24-mer sequences (DK-66 is counted as a 2x12-mer) that vary in G-richness.

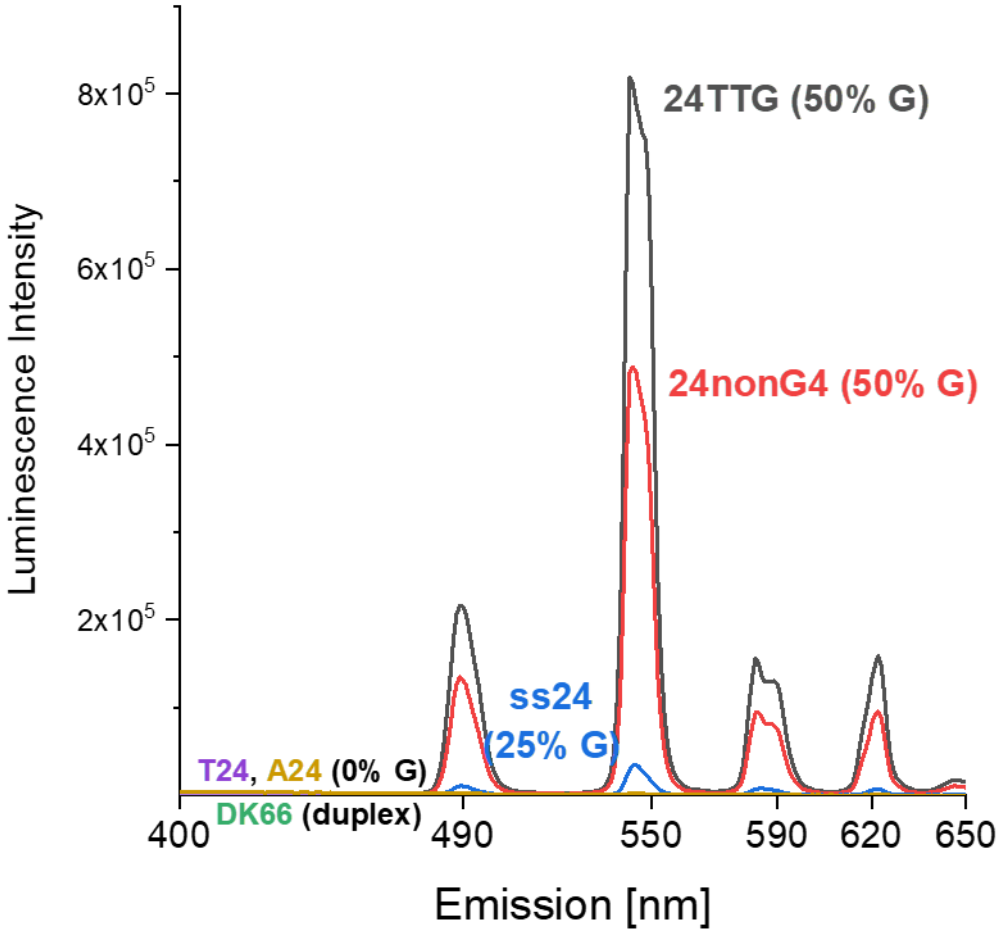


Figure 40 Emission spectra for the Tb^{3+} luminescence with different DNA sequences. Parenthesis show the G-richness of each sequence (DK66 has 33% but is also the only double-stranded species). Samples contain 10 µM DNA, 50 µM $TbCl_3$ and 50 mM TMAA (pH 6.8). Excitation wavelength is 290 nm. T = 20°C.

Seeing no significant change in emission for DK66 was expected, as it was already established that lanthanides do not light up on double-stranded DNA because the nucleobases are solvent-inaccessible.[11,13,14,27] It was also demonstrated on single nucleotides that Tb^{3+} specifically targets guanine (Figure 36). Figure 40 validates the guanine-specificity not just for single nucleotides, but for short single-stranded DNA. First of all, we can disprove any specificity towards thymine or adenine, because the emission barely changes upon adding T24 or A24 (see Figure S6 for emission spectra with logarithmic scaling). Second, the Tb^{3+} luminescence scales with guanine content when comparing ss24 (25% guanine) to 24TTG or 24nonG4 (50% guanine). And since 24TTG does not contain any cytosine, the luminescence effect is clearly caused by the abundance of guanine. In 24nonG4, the Gs are randomly distributed across the sequence (to prevent G-quadruplex formation in K^+), while they are organized into four

GGG tracts in 24TTG. The 2-fold luminescence increase from 24nonG4 to 24TTG suggests that the interaction with Tb^{3+} involves some sort of intramolecular organization of the DNA strand, otherwise the positioning of the Gs in the sequence would not matter.

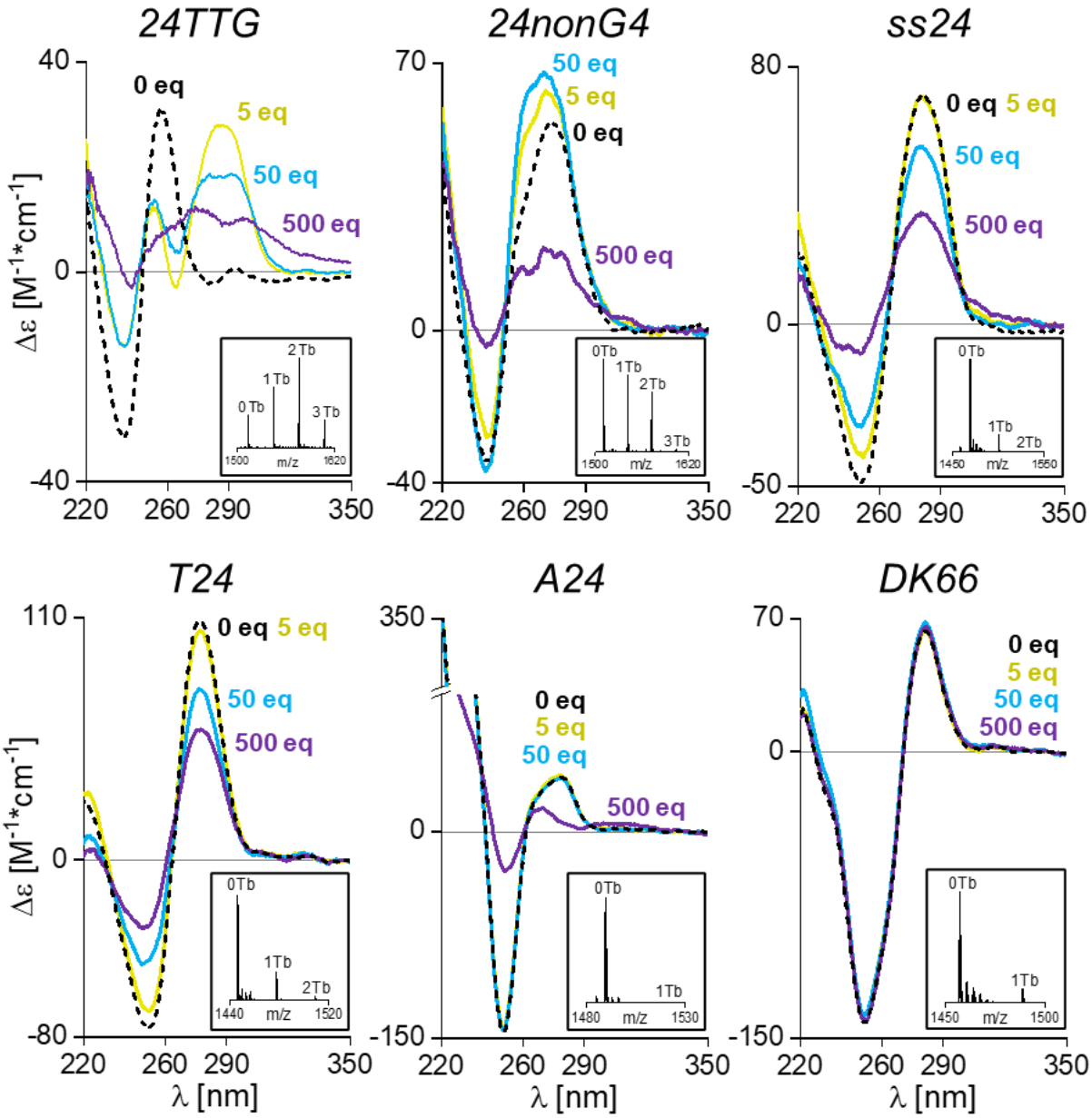


Figure 41 CD titrations of increasing Tb^{3+} equivalents (eq) onto six different 24-mer sequences. The small panel showcases the cation distribution at 5 equivalents, obtained from mass spectrometry (showing the 5- charge state). Samples contain 10 µM DNA (20 µM DK66), 0-10 mM $TbCl_3$ and 50 mM TMAA (pH 6.8) in H_2O.

We further analyze the undefined secondary structure through circular dichroism (CD) and mass spectrometry (MS). We performed CD titrations similar to the ones done in Figure 39 to assess Tb-induced changes in secondary structure for all sequences of interest. Up to 5 equivalents, we recorded mass spectra in order to quantify how many Tb^{3+} ions are involved in this secondary structure and to make qualitative assessments on Tb^{3+} affinity to the DNA sequences, because we cannot do that from CD spectra or a single luminescence spectrum. Figure 41 shows a side-by-side comparison of CD and mass spectra in the same Tb^{3+} concentration. Full CD and MS titration datasets are deposited in the supporting information (Figure S7 to S18).

24TTG stands out from the group with a defined change in CD signature that is more than just disruption. The mass spectrum shows 0-3 Tb^{3+} ions bound. Based on their relative abundance, the 1 Tb^{3+} and 2 Tb^{3+} adducts appear to be specific. In the context of G-quadruplexes, that would be consistent with a 2-tetrad and 3-tetrad G-quadruplex, but neither this nor the antiparallel CD signature is sufficient evidence for G-quadruplex formation. Under normal circumstances the K_{D1} and K_{D2} values of Tb^{3+} binding to 24TTG could be determined from the MS titrations and compared to K^+. However, Tb^{3+} also interacts with the acetate in the TMAA buffer, forming an organometallic bond (Figure 42).[28–30] Reducing the TMAA concentration from 100 mM to 50 mM mitigated the issue, but TMAA has to be kept in to regulate ionic strength, as it is essential for DNA folding. Substituting TMAA with ammonium bromide led to the formation of bromide clusters that reduced the S/N ratio for DNA (data not shown). The most promising electrolyte would be trimethylammonium carbonate, but since determining K_D values is not a priority in this project, it has not been purchased for the attempt.

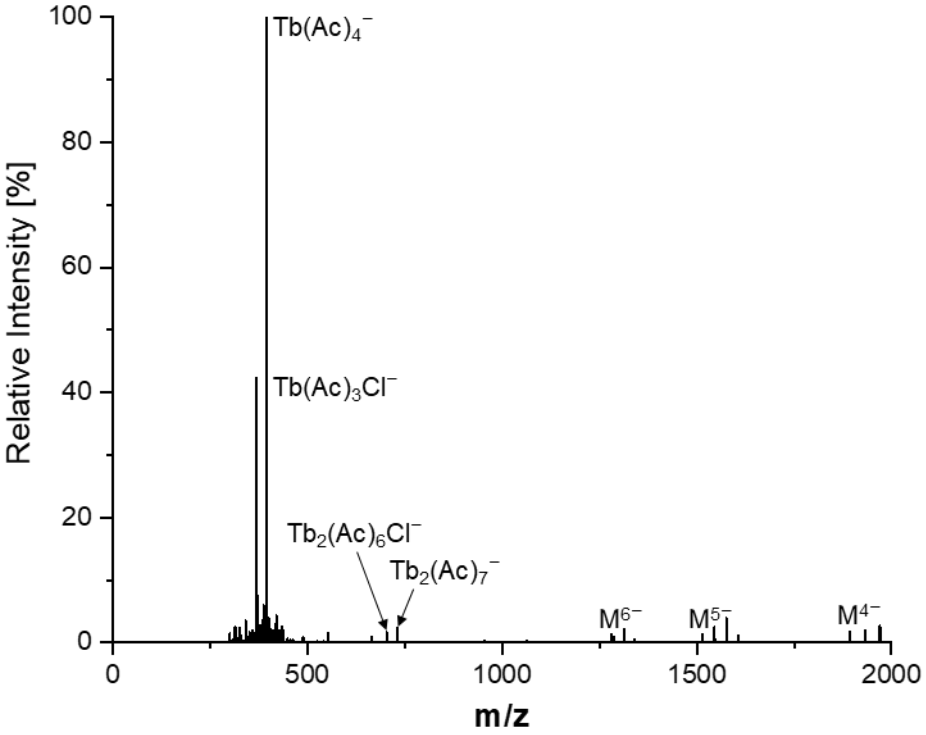


*Figure 42 Full MS spectrum of 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A) and 5 eq Tb^{3+}, revealing organometallic acetate (Ac) complexes. Sample contains 10 µM DNA (labeled 'M'), 50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.*

24nonG4 makes it clear that the arrangement of Gs within the sequence does affect Tb^{3+} binding significantly, thus the Tb^{3+} ions in the secondary structure must coordinate with more than one guanine. Otherwise, the arrangement of Gs would not matter when 24TTG and 24nonG4 have the same length and G-content. Like 24TTG, 24nonG4 has 1 and 2 specifically bound Tb^{3+} ions (at 5 eq), but also a much higher fraction of unbound strand (0 Tb). Therefore, the association constant of Tb^{3+} binding 24nonG4 is much lower than for binding 24TTG. Note how the CD signal is still increasing for 24nonG4 from 5 to 50 eq, while it already declines for 24TTG. With lower guanine affinity we need higher Tb^{3+} concentrations at which secondary structure formation competes with aggregation. For that reason, changes in CD signature appear to be less significant for 24nonG4 compared to 24TTG. For ss24, the association constant is so low that there is no remarkable change in the CD spectrum until aggregation commences.

For the control sequences, adding Tb^{3+} just leads to a loss of CD signal intensity. Only DK66 remains indifferent to the presence of Tb^{3+}, confirming yet again that Tb^{3+} cannot rearrange double-stranded DNA.[11,13,27] At 5 equivalents Tb^{3+}, there are sparsely populated adducts carrying 1, sometimes 2 Tb^{3+}. Judging by the adduct distribution and the lack of Tb^{3+} luminescence (Figure 40), these are unspecific adducts where Tb^{3+} attaches to the negatively charged backbone.

G-quadruplex formation in Tb^{3+} is unlikely, but not disproven

We want to clarify whether Tb^{3+} promotes a G-quadruplex structure or not. NMR is the fastest approach to validating G-quadruplex formation through the characteristic shift of H1-protons.[31] In a titration experiment, the ^{1}H-NMR signal for 24TTG and 24nonG4 was disappearing with increasing Tb^{3+} concentration (data not shown). This phenomenon is known as paramagnetic relaxation enhancement, where paramagnetic ions such as Tb^{3+} enhance the relaxation of protons in a range of 5-20 Å, causing signal suppression.[32,33] So while this experiment confirms that Tb^{3+} is interacting with the DNA strand, any sort of structural information is lost.

For G4T4G4 and TG4T, molecularity serves as a way to confirm G-quadruplex formation. If Tb^{3+} promotes G-quadruplex formation with e.g. G4T4G4, there will be 2 strands of G4T4G4 associated with Tb^{3+}, instead of just one (Figure 43).

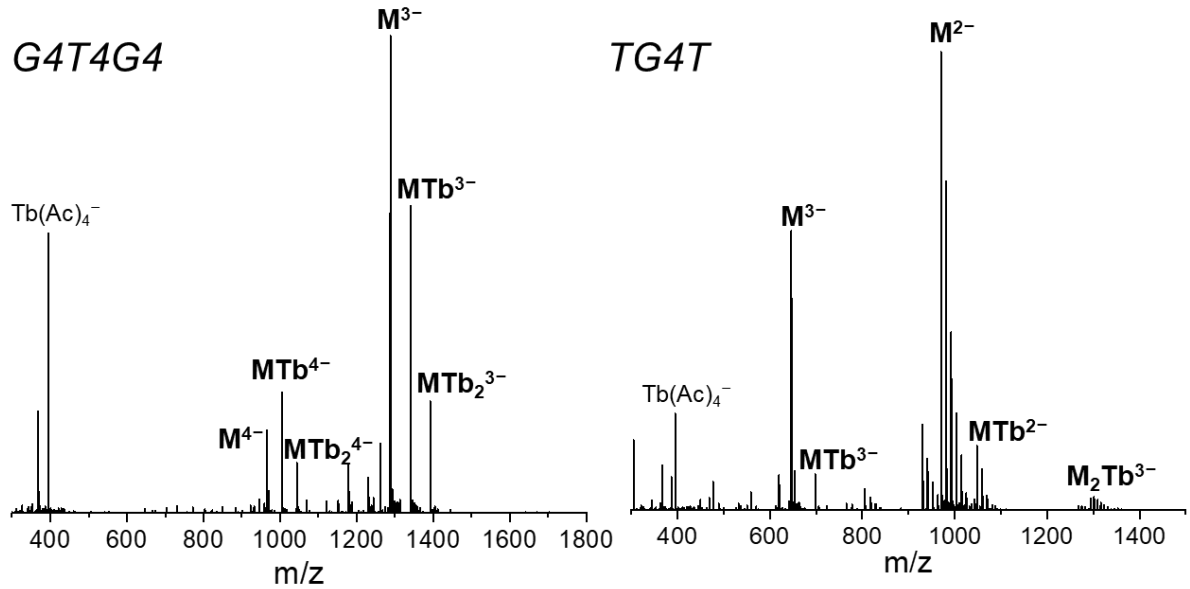


*Figure 43 Mass spectra of d(PO₄)**G4**T₄**G₄** and d(PO₄)T**G₄**T (M = DNA) after equilibrating in Tb^{3+} for 10 weeks. Samples contain 20 µM G4T4G4/40 µM TG4T, 50 µM $TbCl_3$ and 50 mM TMAA (pH 6.8) in H_2O. Note that TG4T is too small for the desalting procedure, hence the abundance of Na^+ adducts.*

Tb^{3+} interacts with the single strand. No multimolecular assemblies were identified, which is evidence against Tb^{3+} promoting G-quadruplex formation. The quantity of Tb^{3+} adducts is suspiciously low for TG4T. Combined with the presence of M_2Tb species it indicates that 4 guanines are not enough and Tb^{3+} prefers to coordinate at least 6 guanines. The Tb^{3+} coordination state with single nucleotide guanines was 8 and 12 (Figure 35), so we hypothesize that Tb^{3+} coordinates at least 6 but no more than 12 guanines.

We added Tb^{3+} to a G-quadruplex pre-folded in K^+ to assess whether Tb^{3+} can integrate into a K^+-mediated G-quadruplex (Figure 44). Cations like Na^+ or Pb^{2+} can integrate by displacing K^+, forming mixed specific adducts in a 3-tetrad G-quadruplex.[3,34] Tb^{3+} and K^+ do not mix, the low quantity of the MKTb highlights that it is not a mixed K-Tb-quadruplex, but an unspecific K^+ adduct of the MTb species. It was previously established that Tb^{3+} converts a K^+-mediated G-quadruplex into a different unknown structure.[20,35] Our results reconfirm this, while adding K^+ does not specifically

bind to the Tb^{3+}-mediated structure, which speaks against G-quadruplex formation in Tb^{3+}.

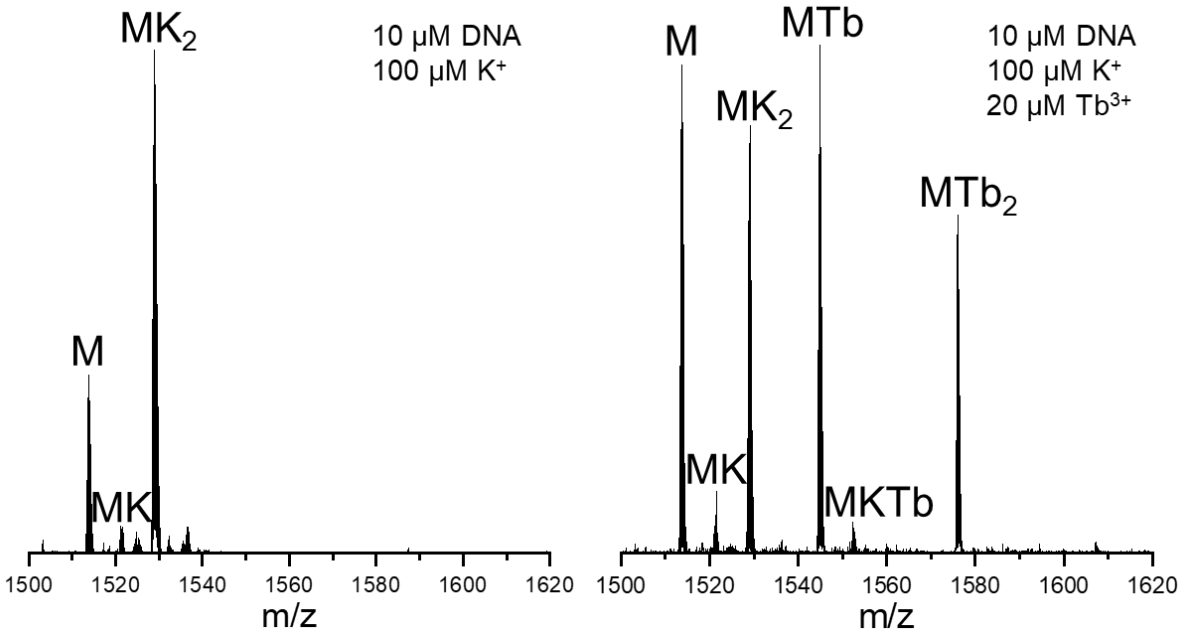


*Figure 44 Mass spectrum of 10 µM 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A) in 100 µM K^+ before and after adding 20 µM Tb^{3+}. Sample matrix contains 50 mM TMAA (pH 6.8) in H_2O.*

Single-stranded DNA folding into a G-quadruplex comes with a change in conformation, that can be tracked with ion mobility. In Figure 45, we compare the change in CCS distribution for K^+ and Tb^{3+} with three sequences of interest.

ss24 and 24nonG4 do not form G-quadruplexes in K^+, therefore the CCS distributions for 0K and 2K are superimposed (both adduct states correspond to unfolded strand). In 24TTG, the 2K adduct corresponds to a G-quadruplex and thus there is a significant CCS shift compared to the unfolded strand (0K). The 2K and the 2Tb species have matching CCS values at peak maximum (ca. 764 $Å^2$), which speaks in favor of Tb^{3+} promoting a G-quadruplex arrangement (although it does not prove it). There is also a slight increase in CCS value for the Tb^{3+} adducts of 24nonG4 and ss24, but the broadness of the CCS peaks makes it difficult to determine whether these shifts correlate to a significant conformational change.

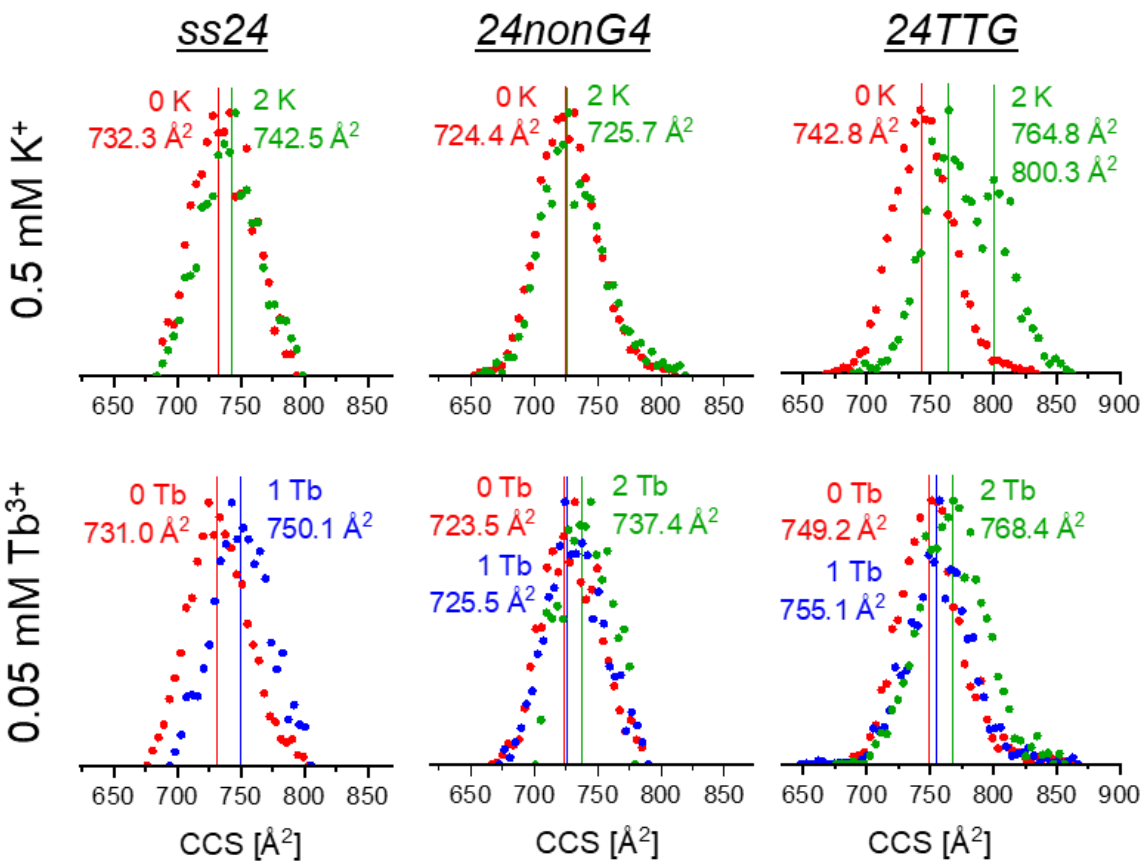


Figure 45 Normalized CCS distributions (4- charge state) of unfolded DNA (0 cation) and putatively specific adduct states (1/2Tb, 2K) of the three DNA sequences that form Tb-G complex. CCS shifts can signify a conformational change. ***Top row****: K^+-induced CCS changes.* ***Bottom row****: Tb^{3+}-induced CCS changes. Samples contain 10 µM DNA, 50 mM TMAA (pH 6.8) in H_2O and 500 µM KCl (top row) or 50 µM $TbCl_3$ (bottom row).*

Finally, Tb^{3+}, Sr^{2+} and K^+ are probed for their melting behavior with 24TTG to have a direct comparison between mono-, di- and trivalent cation. Figure 46 displays the CD melting spectra and the melting curves taken at λ_{max}. All cations are at the same concentration (50 µM). Visually, Tb^{3+} and Sr^{2+} have fairly similar melting curves. But the T_M for Tb^{3+} is closer to K^+ (T_M = 22°C), while Sr^{2+} (T_M = 35°C) has the highest T_M value. G-quadruplexes are often more thermodynamically stable in Sr^{2+} than in K^+ and 24TTG is no exception.[4] The T_M for Tb^{3+} cannot be quantified, because the structure is not fully folded at 4°C. The transition is very broad, compared to K^+/Sr^{2+}, meaning that the affinity constant is less temperature dependent.[36] A lower temperature dependency could mean that the Tb^{3+}-guanine structure is less entropy-driven than a standard G-quadruplex.

All in all, we presented several pieces of evidence for and against G-quadruplex formation in the presence of Tb^{3+}. The evidence is thus insufficient to validate or disprove the existence of a G-quadruplex in Tb^{3+}. The absence of a tetramer in TG4T makes me personally believe we are not dealing with a G-quadruplex, but something completely different.

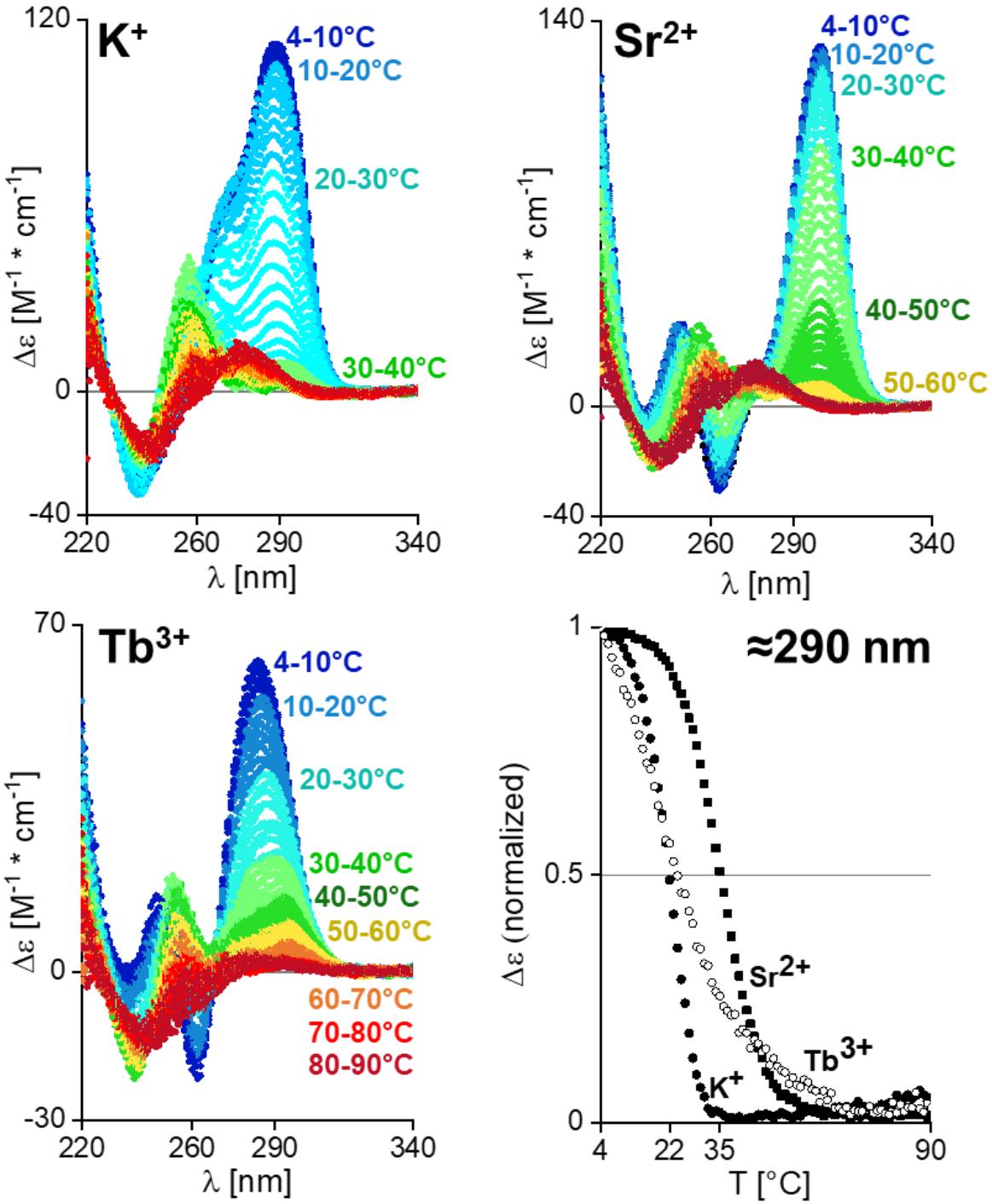


*Figure 46 CD melting transition from 4°C to 90°C of 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A) in 50 µM of mono-, di- and trivalent cation. Samples contain 10 µM DNA, 50 µM KCl/SrAc₂/TbCl₃ and 50 mM TMAA (pH 6.8) in H₂O. Melting curves were taken at the absorption maxima (288 nm for K^+, 300 nm for Sr^{2+} and 283 nm for Tb^{3+}).*

Conclusion

We investigated the interaction between G-rich DNA sequences and different lanthanides to assess whether or not they promote G-quadruplex formation. Eu^{3+} and Tb^{3+} induced the folding of G-rich DNA into a so far unknown secondary structure, while La^{3+} and Yb^{3+} did not. In-depth analysis with Tb^{3+} confirmed that Tb^{3+} specifically targets the guanine nucleobase and coordinates multiple guanines. A DNA sequence with 12 G bases coordinates 1 or 2 Tb^{3+} ions. G-quadruplex formation is unlikely, but cannot be disproven. Based on the number of specific Tb^{3+} ions and the confirmed multi-coordination we propose some possible structures in Figure 47.

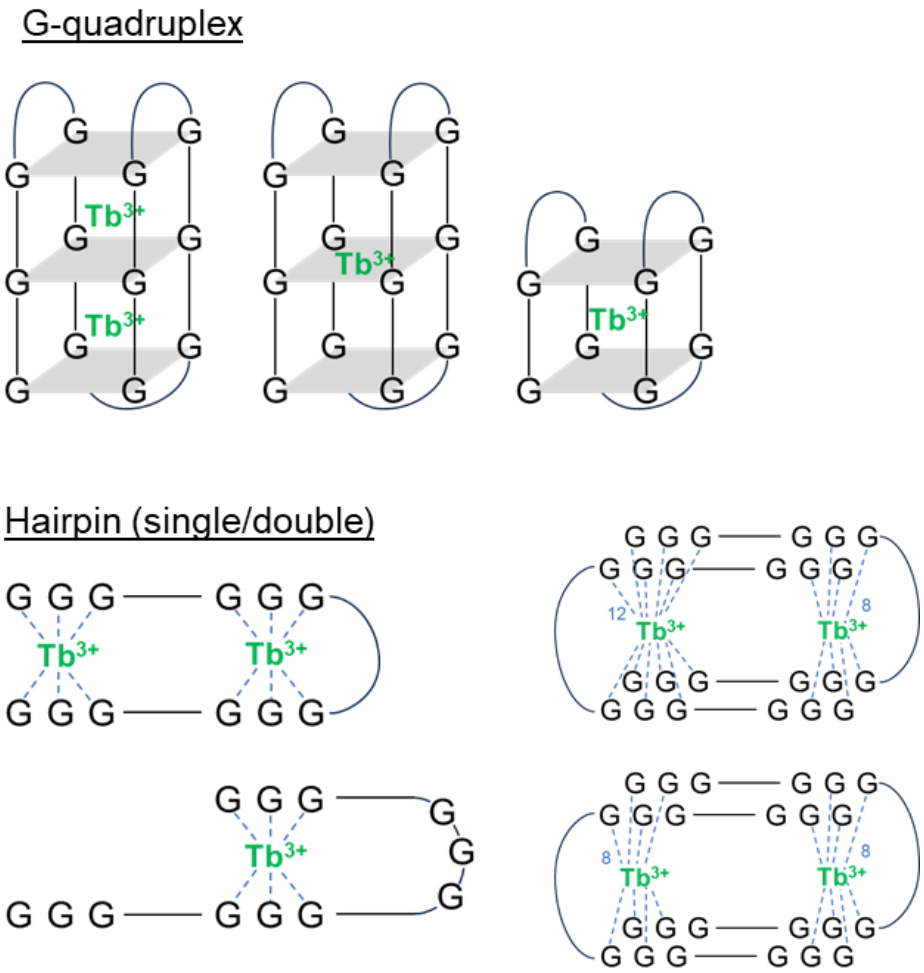


*Figure 47 Propositions for the structure formed by 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**A) and Tb^{3+}, taking into account that there are 1-2 specifically bound Tb^{3+} ions.*

We have exhausted many methods trying to determine the structure of the 24TTG-Tb^{3+} complex. Nonetheless, there is always room left for improvement, such as: 1) obtain a crystal structure. Tb^{3+} evidently facilitates DNA aggregation, so it is a good candidate for crystallization. 2) investigate specific DNA interactions with La^{3+}, which is diamagnetic so NMR structural analysis is possible (like in Figure 35). 3) substitute TMAA with TMAC (trimethylammonium carbonate) to prevent Tb^{3+} from binding to the buffer matrix. 4) measure mass spectra at a higher m/z range to look for double hairpin formation, as proposed in Figure 47.

References

(1) Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *The Alkali Metal Ions: Their Role for Life*, 2016, 203–258. https://doi.org/10.1007/978-3-319-21756-7_7.

(2) Yu, Z.; Zhou, W.; Han, J.; Li, Y.; Fan, L.; Li, X. Na+-Induced Conformational Change of Pb2+-Stabilized G-Quadruplex and Its Influence on Pb2+ Detection. *Anal. Chem.* **2016**, *88* (19), 9375–9380. https://doi.org/10.1021/acs.analchem.6b02466.

(3) Yu, Z.; Zhou, W.; Ma, G.; Li, Y.; Fan, L.; Li, X.; Lu, Y. Insights into the Competition between K^+ and Pb^{2+} Binding to a G-Quadruplex and Discovery of a Novel K^+–Pb^{2+}–Quadruplex Intermediate. *Journal of Physical Chemistry B* **2018**, *122* (40), 9382–9388. https://doi.org/10.1021/acs.jpcb.8b08161.

(4) Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *Journal of the American Chemical Society*, 2016, *138*, 2780–2792. https://doi.org/10.1021/jacs.5b13130.

(5) Wilcox, J. M.; Rempel, D. L.; Gross, M. L. Method of Measuring Oligonucleotide−Metal Affinities: Interactions of the Thrombin Binding Aptamer with K+ and Sr2+. *Anal. Chem.* **2008**, *80* (7), 2365–2371. https://doi.org/10.1021/ac701903w.

(6) Vairamani, M.; Gross, M. L. G-Quadruplex Formation of Thrombin-Binding Aptamer Detected by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **2003**, *125* (1), 42–43. https://doi.org/10.1021/ja0284299.

(7) Monn, S. T. M.; Schürch, S. Investigation of Metal-Oligonucleotide Complexes by Nanoelectrospray Tandem Mass Spectrometry in the Positive Mode. *J Am Soc Mass Spectrom* **2005**, *16* (3), 370–378. https://doi.org/10.1016/j.jasms.2004.11.019.

(8) Hettich, R. L. Investigating the Effect of Transition Metal Ion Oxidation State on Oligodeoxyribonucleotide Binding by Matrix-Assisted Laser Desorption/Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *International Journal of Mass Spectrometry* **2001**, *204* (1), 55–75. https://doi.org/10.1016/S1387-3806(00)00338-9.

(9) Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallographica Section A* **1976**, *32* (5), 751–767. https://doi.org/10.1107/s0567739476001551.

(10) Kwan, I. C. M.; She, Y.-M.; Wu, G. Trivalent Lanthanide Metal Ions Promote Formation of Stacking G-Quartets. *Chem. Commun.*, 2007, 4286–4288. https://doi.org/10.1039/b710299b.

(11) Fu, P. K.-L.; Turro, C. Energy Transfer from Nucleic Acids to Tb(III): Selective Emission Enhancement by Single DNA Mismatches. *J. Am. Chem. Soc.* **1999**, *121* (1), 1–7. https://doi.org/10.1021/ja9826082.

(12) Horrocks, W. DeW. Jr.; Sudnick, D. R. Lanthanide Ion Probes of Structure in Biology. Laser-Induced Luminescence Decay Constants Provide a Direct Measure of the Number of Metal-Coordinated Water Molecules. *J. Am. Chem. Soc.* **1979**, *101* (2), 334–340. https://doi.org/10.1021/ja00496a010.

(13) Costa, D.; Burrows, H. D.; da Graça Miguel, M. Changes in Hydration of Lanthanide Ions on Binding to DNA in Aqueous Solution. *Langmuir* **2005**, *21* (23), 10492–10496. https://doi.org/10.1021/la051493u.

(14) Klakamp, S. L.; Horrocks, W. DeW. Lanthanide Ion Luminescence as a Probe of DNA Structure. 1. Guanine-Containing Oligomers and Nucleotides. *Journal of Inorganic Biochemistry* **1992**, *46* (3), 175–192. https://doi.org/10.1016/0162-0134(92)80028-T.

(15) Zhang, M.; Le, H.-N.; Jiang, X.-Q.; Yin, B.-C.; Ye, B.-C. Time-Resolved Probes Based on Guanine/Thymine-Rich DNA-Sensitized Luminescence of Terbium(III). *Anal. Chem.* **2013**, *85* (23), 11665–11674. https://doi.org/10.1021/ac4034054.

(16) Zhang, J.; Chen, J.; Chen, R.; Chen, G.; Fu, F. An Electrochemical Biosensor for Ultratrace Terbium Based on Tb3+ Promoted Conformational Change of Human Telomeric G-Quadruplex. *Biosensors and Bioelectronics* **2009**, *25* (2), 378–382. https://doi.org/10.1016/j.bios.2009.07.029.

(17) Worlinsky, J. L.; Basu, S. Detection of Quadruplex DNA by Luminescence Enhancement of Lanthanide Ions and Energy Transfer from Lanthanide Chelates. *J. Phys. Chem. B* **2009**, *113* (4), 865–868. https://doi.org/10.1021/jp8100464.
(18) Ringer, D. P.; Burchett, S.; Kizer, D. E. Use of Terbium(III) Fluorescence Enhancement to Selectively Monitor DNA and RNA Guanine Residues and Their Alteration by Chemical Modification. *Biochemistry* **1978**, *17* (22), 4818–4825. https://doi.org/10.1021/bi00615a032.
(19) Nagesh, N.; Bhargava, P.; Chatterji, D. Terbium(III)-Induced Fluorescence of Four-Stranded G4-DNA. *Biopolymers* **1992**, *32* (10), 1421–1424. https://doi.org/10.1002/bip.360321015.
(20) Galezowska, E.; Gluszynska, A.; Juskowiak, B. Luminescence Study of G-Quadruplex Formation in the Presence of Tb3+ Ion. *Journal of Inorganic Biochemistry*, 2007, *101*, 678–685. https://doi.org/10.1016/j.jinorgbio.2006.12.013.
(21) Yonuschot, G.; Mushrush, G. W. Terbium as a Fluorescent Probe for DNA and Chromatin. *Biochemistry* **1975**, *14* (8), 1677–1681. https://doi.org/10.1021/bi00679a020.
(22) Largy, E.; König, A.; Ghosh, A.; Ghosh, D.; Benabou, S.; Rosu, F.; Gabelica, V. Mass Spectrometry of Nucleic Acid Noncovalent Complexes. *Chem. Rev.* **2021**. https://doi.org/10.1021/acs.chemrev.1c00386.
(23) Seitz, M.; Oliver, A. G.; Raymond, K. N. The Lanthanide Contraction Revisited. *J. Am. Chem. Soc.* **2007**, *129* (36), 11153–11160. https://doi.org/10.1021/ja072750f.
(24) Khairul Zaman, N.; Rohani, R.; Izni Yusoff, I.; Kamsol, M. A.; Basiron, S. A.; Abd. Rashid, A. I. Eco-Friendly Coagulant versus Industrially Used Coagulants: Identification of Their Coagulation Performance, Mechanism and Optimization in Water Treatment Process. *Int J Environ Res Public Health* **2021**, *18* (17), 9164. https://doi.org/10.3390/ijerph18179164.
(25) Marchand, A.; Gabelica, V. Folding and Misfolding Pathways of G-Quadruplex DNA. *Nucleic Acids Research*, 2016, *44*, 10999–11012. https://doi.org/10.1093/nar/gkw970 %J Nucleic Acids Research.
(26) Ghosh, A.; Largy, E.; Gabelica, V. DNA G-Quadruplexes for Native Mass Spectrometry in Potassium: A Database of Validated Structures in Electrospray-Compatible Conditions. *Nucleic Acids Research* **2021**, *49* (4), 2333–2345. https://doi.org/10.1093/nar/gkab039.
(27) Ringer, D. P.; Etheredge, J. L.; Kizer, D. E. The Influence of DNA Sequence on Terbium (III) Fluorescence Enhancement by DNA. *Journal of Inorganic Biochemistry* **1985**, *24* (2), 137–145. https://doi.org/10.1016/0162-0134(85)80005-2.
(28) Shintoyo, S.; Murakami, K.; Fujinami, T.; Matsumoto, N.; Mochida, N.; Ishida, T.; Sunatsuki, Y.; Watanabe, M.; Tsuchimoto, M.; Mrozinski, J.; Coletti, C.; Re, N. Crystal Field Splitting of the Ground State of Terbium(III) and Dysprosium(III) Complexes with a Triimidazolyl Tripod Ligand and an Acetate Determined by Magnetic Analysis and Luminescence. *Inorg. Chem.* **2014**, *53* (19), 10359–10369. https://doi.org/10.1021/ic501453h.
(29) Vojtíšek, P.; Cígler, P.; Kotek, J.; Rudovský, J.; Hermann, P.; Lukeš, I. Crystal Structures of Lanthanide(III) Complexes with Cyclen Derivative Bearing Three Acetate and One Methylphosphonate Pendants. *Inorg. Chem.* **2005**, *44* (16), 5591–5599. https://doi.org/10.1021/ic048190s.
(30) Gueye, Nd. M.; Moussa, D.; Thiam, E. I.; Barry, A. H.; Gaye, M.; Retailleau, P. Crystal Structure of Bis-(Acetato-K2 O,O′)Di-aqua-[1-(Pyridin-2-Yl-methyl-idene-κN)-2-(Pyridin-2-Yl-κN)Hydrazine-κN 1]Terbium(III) Nitrate Monohydrate. *Acta Crystallogr E Crystallogr Commun* **2017**, *73* (Pt 8), 1121–1124. https://doi.org/10.1107/S2056989017009653.
(31) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods*, 2012, *57*, 11–24. https://doi.org/10.1016/j.ymeth.2012.05.003.
(32) Clore, G. M.; Iwahara, J. Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* **2009**, *109* (9), 4108–4139. https://doi.org/10.1021/cr900033p.
(33) Venu, A. C.; Nasser Din, R.; Rudszuck, T.; Picchetti, P.; Chakraborty, P.; Powell, A. K.; Krämer, S.; Guthausen, G.; Ibrahim, M. NMR Relaxivities of Paramagnetic Lanthanide-Containing Polyoxometalates. *Molecules* **2021**, *26* (24), 7481. https://doi.org/10.3390/molecules26247481.

(34) Ma, G.; Yu, Z.; Zhou, W.; Li, Y.; Fan, L.; Li, X. Investigation of Na+ and K+ Competitively Binding with a G-Quadruplex and Discovery of a Stable K+–Na+-Quadruplex. *J. Phys. Chem. B* **2019**, *123* (26), 5405–5411. https://doi.org/10.1021/acs.jpcb.9b02823.
(35) Chen, Q.; Zuo, J.; Chen, J.; Tong, P.; Mo, X.; Zhang, L.; Li, J. A Label-Free Fluorescent Biosensor for Ultratrace Detection of Terbium (ш) Based on Structural Conversion of G-Quadruplex DNA Mediated by ThT and Terbium (ш). *Biosensors and Bioelectronics* **2015**, *72*, 326–331. https://doi.org/10.1016/j.bios.2015.04.039.
(36) Mergny, L., Jean-Louis; Lacroix. Analysis of Thermal Melting Curves. *Oligonucleotides*, 2003, 515–537.

CHAPTER 2: CHALLENGING THE TOPOLOGICAL BIAS OF *N*-TETRAD G-QUADRUPLEXES BY SELECTIVE EXTENSION OF G-TRACTS.

Motivation

Intramolecular 3-tetrad G-quadruplexes are a unique case where full heterostacking (i.e., alternating *syn*/*anti* stacks) has never been documented in K^+. The most likely explanation is that *syn*→*anti* stacking is energetically favored, while *anti*→*syn* stacking is not.[1] Therefore, the heterostacking pattern is favorable in 2-tetrad (4x *syn*→*anti*) or 4-tetrad (4x *syn*→*anti*→*syn*→*anti*) quadruplexes. But 3-tetrad quadruplexes (2x *syn*→*anti*→*syn* and 2x *anti*→*syn*→*anti*) have an equal ratio of favorable and unfavorable base stacks. As a result, antiparallel 3-tetrad G-quadruplexes in K^+ are very rare. One of those exceptions is the G-quadruplex 5YEY (Figure 48),[2] which adopts a hybrid base stacking.

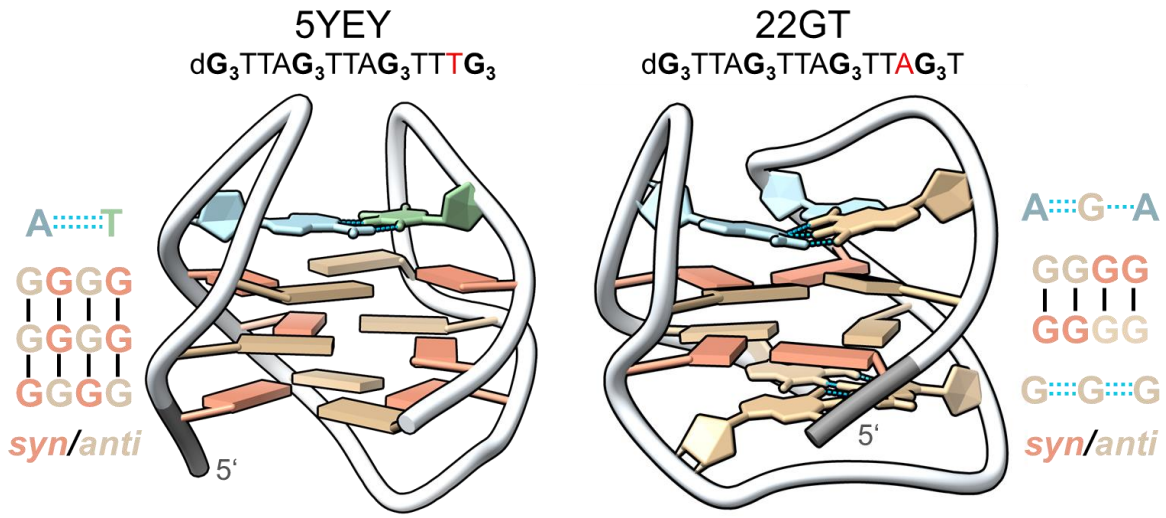


Figure 48 ***Left****: Solution NMR structure of antiparallel 3-tetrad G-quadruplex 5YEY (d**G₃**TTA**G₃**TTA**G₃**TTT**G₃**) (PDB: 5YEY).[2]* ***Right****: Solution NMR structure of antiparallel 2-tetrad G-quadruplex 22GT (d**G₃**TTA**G₃**TTA**G₃**TTA**G₃**T) (PDB: 2KF8).[3]*

5YEY is a human telomeric repeat 21-mer sequence with an A→T mutation in the third loop. This mutation enables the formation of an AT base pair, creating a contact between the first and the third loop, which can only be made in an antiparallel loop arrangement (lateral-diagonal-lateral or lateral-lateral-lateral). It appears that the energetic contribution of the AT base pair formation is high enough to stabilize the unusual antiparallel topology. However, 5YEY adopts a hybrid base stacking pattern, so the unfavorable energetics of heterostacking in a 3-tetrad K^+ G-quadruplex do not apply here. In fact, 5YEY possesses characteristic attributes of both antiparallel and hybrid topology, which are summarized in Table 3.

Table 3 Separation of antiparallel and hybrid characteristics for 5YEY.

Antiparallel attributes	**Hybrid attributes**
Loop progression (*lateral-lateral-lateral*)	Hybrid G-tetrad stacking (-++)
G-tetrad composition (2 *syn/2 anti*)	Number of G-tetrads (3)
Groove width (*narrow-wide-narrow-wide*)	Glycosidic bond angle progression (*saa-ssa-saa-ssa, s = syn, a = anti*)

The unmutated version of 5YEY (d**G$_3$**TTA**G$_3$**TTA**G$_3$**TT<u>T</u>**G$_3$**), which is 21G (d**G$_3$**TTA**G$_3$**TTA**G$_3$**TT<u>A</u>**G$_3$**), has no documented solution NMR structure in K^+ (in absence of ligand). As a substitute, Figure 48 shows the solution NMR structure of 22GT[3] (d**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**T) which only has an additional flanking T base. Supporting evidence for the claim that 21G and 22GT produce similar G-quadruplex structures is that the two sequences produce similar CD spectra, NMR spectra and K^+ distributions in MS.[4] 22GT is a 2-tetrad antiparallel G-quadruplex capped by a GGG triad on the 5' end and an AGA triad on the 3' end. Disrupting the AGA triad induces a structural conversion into a 3-tetrad hybrid G-quadruplex.[5]

The fact that a 2-tetrad antiparallel G-quadruplex is the preferred topology over a 3-tetrad hybrid G-quadruplex raises our interest. The main goal of this project wass to design a sequence that forms a 3-tetrad heterostacking G-quadruplex and validate its existence. The best way to understand a rule is by studying the exceptions – we hope that finding such an exception will lead towards a fundamental understanding as to why 3-tetrad heterostacking occurs in Na^+, but not in K^+. Such findings could elucidate how K^+ favors certain topologies.

Trying to model a sequence after 5YEY will likely lead to a hybrid stacking G-quadruplex. The 22GT structure looks more promising. We wanted to insert a third G-tetrad into the 22GT structure, which hopefully yields a 3-tetrad heterostacking G-quadruplex that is energetically stabilized by the AGA/GGG triads. For more sequence variation, we also added the RAN4 sequence (d**G$_3$**TA**G$_3$**AGC**G$_3$**AGA**G$_3$**), which like 22GT folds into a 2-tetrad antiparallel G-quadruplex capped by AGA/GGG triads in K^+ (PDB:6GZN).[5]

State of the art

Our central goal is to create a 3-tetrad heterostacking G-quadruplex in K^+ by extending the GGG tracts in 22GT/RAN4 into GGGG tracts. In this section we review what topologies are adopted by G-quadruplex sequences that contain at least one GGGG tract. Some of those examples feature 3-tetrad antiparallel G-quadruplexes and we will elaborate why we did not choose them as model systems for this study.

Short GGGG-containing DNA strands with less than 15 nucleotides will form multi-stranded G-quadruplexes which have fewer loops that impose dynamic constraints and can thus form assemblies that would not be possible for intramolecular G-quadruplexes (Figure 49). The $[dTG_4T]_4$ tetramer is a 4-tetrad parallel G-quadruplex, even in Na^+. The *Oxytricha* telomeric quadruplex $[dG_4T_4G_4]_2$ has two diagonal loops, which is very uncommon for intramolecular G-quadruplexes. The Gs in the $[dGCG_2AG_4AG_2]_2$ dimer are unevenly distributed, creating a peculiar 4-tetrad G-quadruplex that includes atypical secondary structures such a GCGC-tetrad (blue, Figure 49) and an AG_4A-hexad (purple, Figure 49).[6–8]

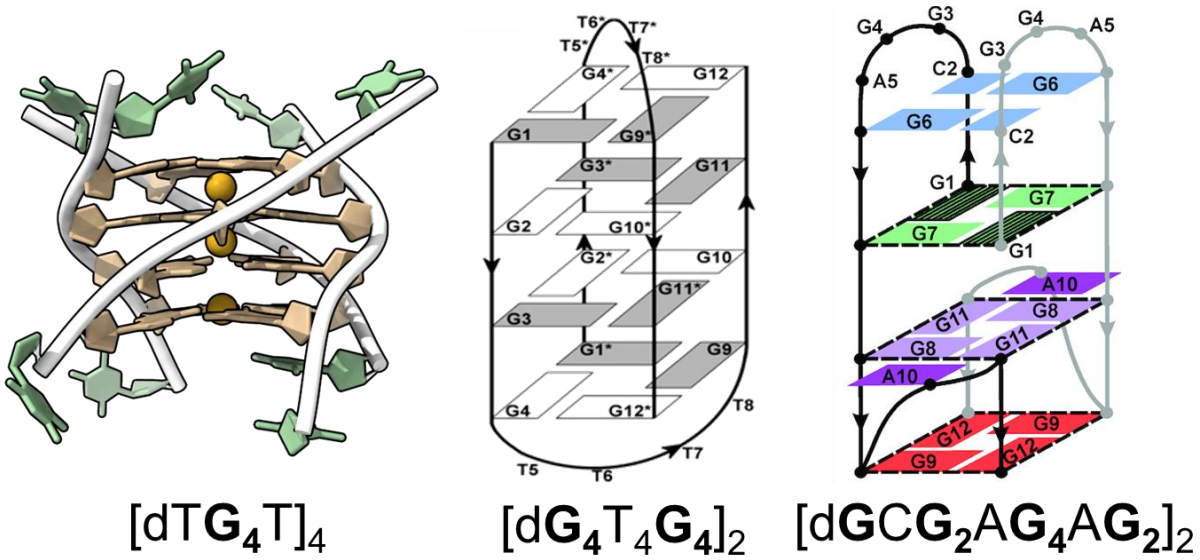


Figure 49 ***Left****: Crystal structure of $[dTG_4T]_4$ in 75 mM NaCl and 10 mM Tris-HCl (pH 7.4). PDB: 2O4F.[9]* ***Center****: NMR-resolved structure of $[dG_4T_4G_4]_2$ in 50 mM NaCl. Adapted with permission from Plavec et al. JACS 2003, 125 (26), 7866-7871.[10] Copyright 2003 ACS.* ***Right****: NMR-derived structure of $[dGCG_2AG_4AG_2]_2$ in 175 mM KCl and 10 mM K_3PO_4 (pH 6.8). Reproduced from Šket et al. NAR 2020, 48 (5), 2749-2761.[6] CC-BY-NC creative commons.*

Removing one guanine from $[dG_4T_4G_4]_2$ limits the highest possible number of tetrads to three. Different 3-tetrad structures were obtained depending on which Gs are being removed (Figure 50).[10] Both heterostacking ($[dG_4T_4G_3]_2$) and hybrid stacking ($[dG_3T_4G_3]_2$) were observed in Na^+. $[dG_3T_4G_4]_2$ adopts hybrid stacking in K^+ (like 5YEY), while displaying other bizarre features, which are 1) inconsistent G-tetrad composition (bottom tetrad is 2+2, middle and upper tetrad are 3+1 *syn/anti*) and 2) a rare v-loop from G19 to G20. This goes to show how removing 1 or 2 G-bases from a G-quadruplex forming sequence can induce significant changes in the final structure.

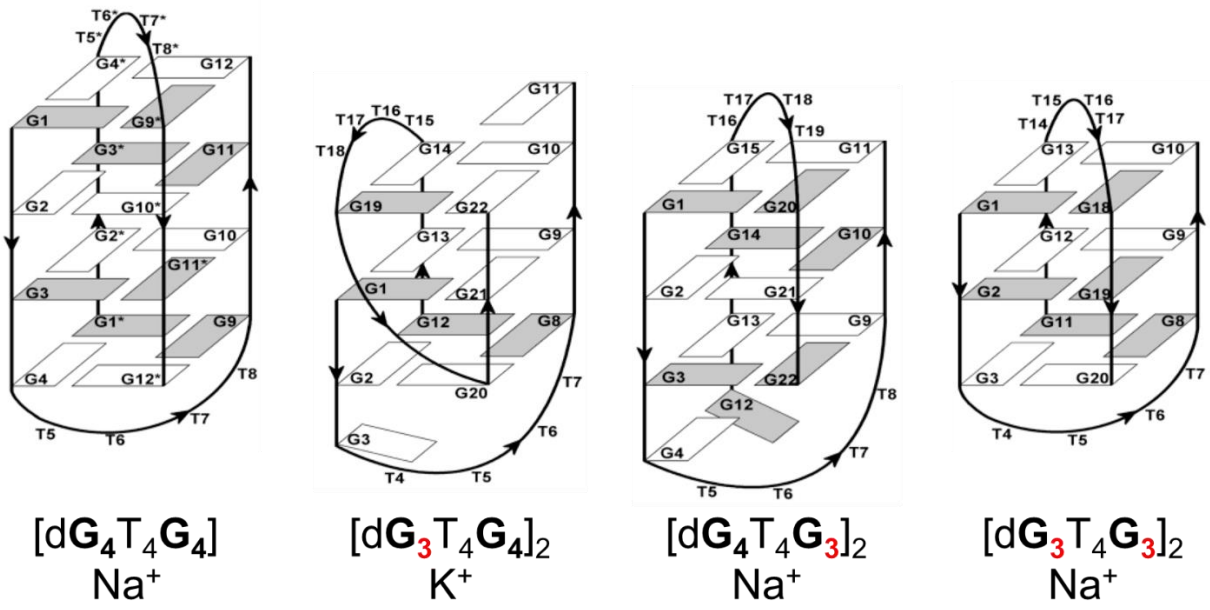


*Figure 50 NMR-resolved solution structure of **1)** $[dG_4T_4G_4]_2$ in 50 mM NaCl at pH 6.0 **2)** $[dG_3T_4G_4]_2$ in 10 mM KCl at pH 5.5 **3)** $[dG_4T_4G_3]_2$ in 50 mM NaCl at pH 7.2 **4)** $[dG_3T_4G_3]_2$ in 65 mM NaCl at pH 6.7. syn-guanines are highlighted in grey. Adapted with permission from Plavec et al. JACS 2003, 125 (26), 7866-7871.[10] Copyright 2003 ACS.*

$[dG_4T_4G_4]_2$ maintains the antiparallel topology in its intramolecular form ($dG_4T_4G_4T_4G_4T_4G_4$), which is commonly referred to as 'oxy28' (Figure 51).[11] Exchanging Na^+ for K^+ has little effect on the CD signature of oxy28, indicating that it is a monomorphic structure (just like $[dTG_4T]_4$).[12] We could have designed sequence candidates for 3-tetrad heterostacking by conjoining two $G_4T_4G_4$ derivatives into one intramolecular stand, for example $dG_4T_4G_3T_4G_4T_4G_3$, and then vary at which position we remove Gs. This is precisely what we ended up doing with 22GT/RAN4 anyway. Using $G_4T_4G_4$ would have left us with two disadvantages compared to 22GT/RAN4: 1) no loop interactions (base pairs/triads) to stabilize unusual topologies and 2) insufficient NMR solution structures in K^+ as a starting point.

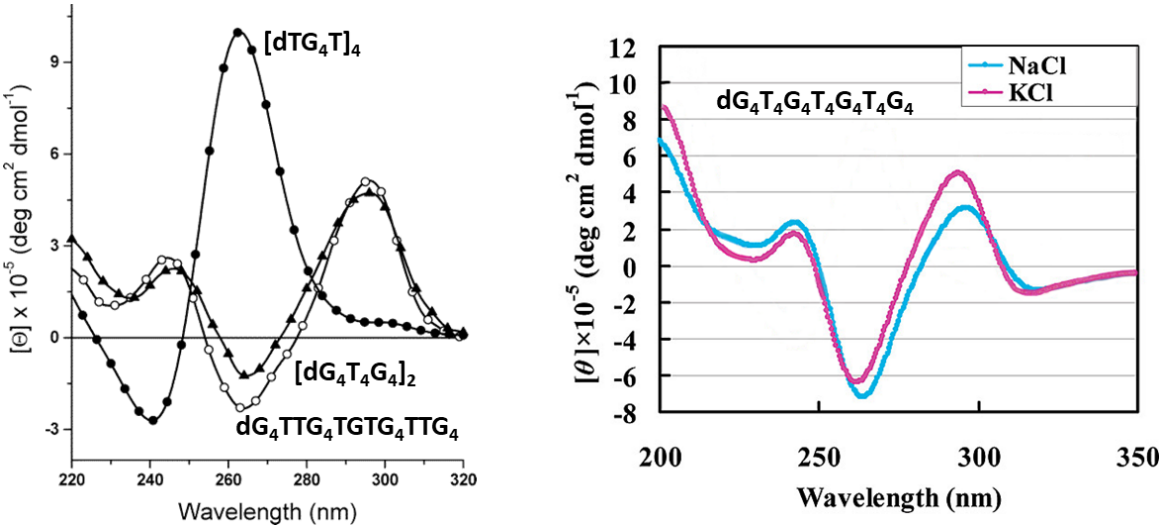
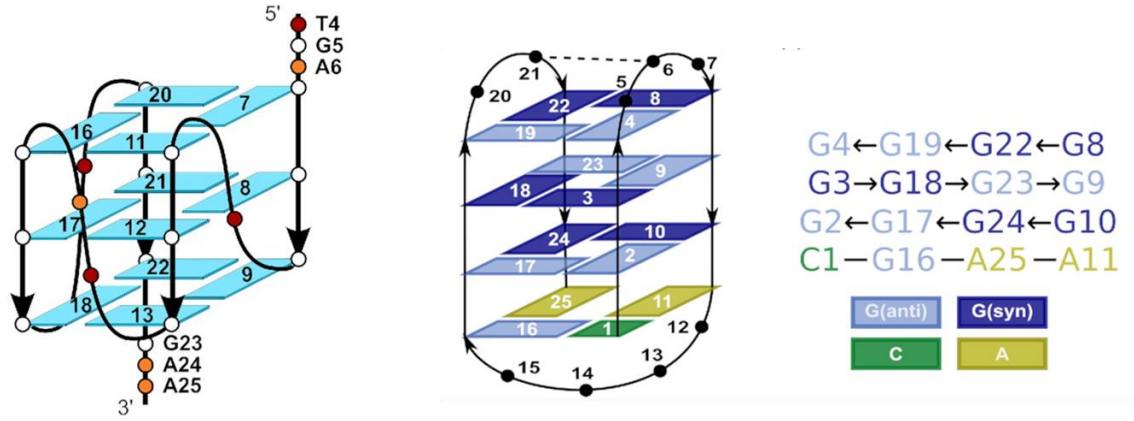


*Figure 51 **Left**: CD spectra obtained in 5 µM DNA, 200 mM NaCl, 10 mM Na_3PO_4 (pH 7.0) and 0.1 mM EDTA. Adapted with permission from Giancola et al. Biochemistry 2004, 43 (16), 4877-4884.[11] Copyright 2004 ACS. **Right**: CD spectra obtained in 6 µM DNA, 100 mM NaCl/KCl and 30 mM Tris-HCl (pH 7.0). Adapted with permission from Sugimoto et al. Biochemistry 2010, 49 (21), 4554-4563.[12] Copyright 2010 ACS.*

The GGGG motif is also present in the promoter regions of the human myc[13] and the bcl2[14] oncogene (Figure 52). However, since they only have one extended G-tract, they cannot form 4-tetrad G-quadruplexes.[13,14]



dT**G**A**G**$_3$T**G**$_3$TA**G**$_3$T**G**$_4$AA dC**G**$_3$CGC**G**$_3$AG$_2$AA**G**$_5$C**G**$_3$A

Figure 52 ***Left****: NMR-resolved structure of myc promoter G-quadruplex in 10 mM K_3PO_4 (pH 7.0, PDB: 7KBV). Reproduced with permission from Yang et al. NAR 2021, 49 (10), 5905-5915.[13] CC-BY-NC creative commons.* ***Right****: NMR-resolved structure of bcl2 promoter G-quadruplex in 70 mM KCl and 20 mM K_3PO_4 (pH 7, PDB: 6ZX7). Reproduced with permission from Plavec et al. NAR 2021, 49(4), 2346-2356.[14] CC-BY-NC creative commons.*

The myc-promoter sequence (Figure 52, left) forms a parallel G-quadruplex with three propeller loops which typically occur when the loop length is restricted to 1-2 nucleotides. The fourth G-base becomes a flanking nucleotide, having no effect on the G-quadruplex core structure (although G23 does form a base pair with A25, so overall it contributes to the stability of this structure).[13,15]

The bcl2 promoter sequence (right panel) has a long central loop with 6 nucleotides, enabling the formation of a diagonal loop for a basket-type antiparallel G-quadruplex. The 6ZX7 structure is so far the only documented instance of a heterostacking 3-tetrad G-quadruplex in K^+. However, the presence of a fourth AGCA tetrad raises the open question whether the 6ZX7 structure should be considered a 3-tetrad or a 4-tetrad G-quadruplex. The AGCA tetrad comes with unique base stacks (From 5' to 3': *syn*-C→*anti*-G, *syn*-G→*anti*-A, *anti*-G→*anti*-G, *syn*-G→*anti*-A), whose energetic contributions are unknown in 3 out of 4 cases. We did not experiment on the 6ZX7 structure, because we were unsure if it counts as a 3-tetrad G-quadruplex. Nevertheless, it is an exceptional structure which could be of great use to analyze the energetics of A/G and C/G base stacking.

Materials and Methods

Sequence selection and preparation

Table 4 summarizes the DNA sequences that were analyzed in this study. The G-tracts of the wild-type sequences contain 12 guanines, the 4444 variants contain 16 guanines and the others contain 14/15 guanines. 22GT_4443 and 22GT_4434 had one guanine removed to ensure they cannot form a 4-tetrad G-quadruplex. RAN4_3443 and RAN_4334 had two guanines removed to reduce the risk of a 4-tetrad G-quadruplex forming with the guanines in the loop regions of the RAN4 sequence.

Table 4 Overview of G-Quadruplex sequences used in this study

22GT	**GGG**	TTA	**GGG**	TTA	**GGG**	TTA	**GGG**	T
22GT_4444	**GGGG**	TTA	**GGGG**	TTA	**GGGG**	TTA	**GGGG**	T
22GT_4443	**GGGG**	TTA	**GGGG**	TTA	**GGGG**	TTA	**GGG**	T
22GT_4434	**GGGG**	TTA	**GGGG**	TTA	**GGG**	TTA	**GGGG**	T

RAN4	**GGG**	TA	**GGG**	AGC	**GGG**	AGA	**GGG**
RAN4_4444	**GGGG**	TA	**GGGG**	AGC	**GGGG**	AGA	**GGGG**
RAN4_3443	**GGG**	TA	**GGGG**	AGC	**GGGG**	AGA	**GGG**
RAN4_4434	**GGGG**	TA	**GGG**	AGC	**GGG**	AGA	**GGGG**

22GT is fully folded at room temperature in MS sample conditions, while RAN4 is a mix of folded and unfolded species (Table 5).

Table 5 Comparing the antiparallel basket-type model sequences used for this study. Melting data taken from [4].

	PDB	*Topology (K^+)*	*loops*	*grooves*	*length*	*M*	*T_M (1 mM K^+)*
RAN4	6GZN	2-tetrad antiparallel	l-d-l	wmnm	20 nt	6393 Da	< 25°C
22GT	2KF8				22 nt	6958 Da	41°C

The DNA sequences (See Table 2) were ordered from Eurogentec and dissolved in UPLC grade water (Biosolve). The stock solutions were annealed at 85°C and then desalted using Amicon centrifugal filter units. Cation contaminations were purged by washing four times with a 200 mM ammonium acetate solution and then six more times with water. The DNA concentration of the desalted stock solutions was determined on a Uvikon XS spectrophotometer, with absorption coefficients being calculated based on nearest-neighbor method.

Instrumental analysis and data processing

Circular dichroism (CD) samples contain 5 or 10 µM DNA, between 0 and 100 mM KCl and, in < 50 mM K^+ concentration, 100 mM trimethylammonium acetate (TMAA, pH 6.8) electrolyte to provide ionic strength. The CD samples were placed in optical cuvettes (Hellma, d = 10 mm) and measured on a Jasco J-815 spectrophotometer at a range from 220 to 350 nm at 50 nm/min scanning speed, 0.2 nm data pitch, 2 nm bandwidth, 2 s data integration time, 22°C temperature in the sample holder, 3

acquisitions. Samples for blank correction were prepared without DNA and measured in the same manner.

CD melting experiments were used to separate signal contributions of different conformers. Meltings were done in a temperature range from 4-90°C, starting at 90°C with the cooling ramp. The temperature gradient was 1°C/min, but was halted every 1°C for spectra acquisition from 220-350 nm with a scan rate of 200 nm/min (1 nm data pitch, 2 nm bandwidth, 2 s data integration time, 1 acquisition). The global temperature ramp is the average of the temperature gradient (1°C/min) and the spectra acquisition time, during which the temperature gradient is halted (0.65 min per spectrum) and lies between 0.20-0.25 °C/min.

The separation of CD signatures from CD melting data was done with singular value decomposition (SVD) by Eric Largy.[16]

$$D = U \times S \times V^T \tag{12}$$

The data matrix *D* contains the y-values of our CD spectra (in our case: Δε) at each wavelength and temperature point. The wavelength is scaled along the rows (*m*) and the temperature is scaled along the columns (*n*) i.e., each column of *D* contains a CD spectrum at a certain temperature. SVD decomposes the data matrix into a product of three matrices. The U-matrix is the most important one. It is an *m*×*m* matrix that contains the so-called basis spectra, which are (normalized component) CD signatures. The experimental melting data can be modeled as a linear combination of the basis spectra. The S-matrix contains the singular values, which can be interpreted as the weights for each basis spectrum. The singular values are used to assess how many components (basis spectra) are significantly contributing to the temperature-induced variance of the CD spectrum. During data processing, we phase-corrected the singular vectors to ensure all basis spectra would have positive bands at 290 nm. The V^T-matrix is an *n*×*n* matrix and more or less the counterpart to the U-matrix. If the U-matrix contains the basis spectra, the V^T-matrix contains the basis melting curves, which are commonly referred to as amplitude vectors.

For MS analysis, we doped 10 µM of DNA with 50 or 500 µM KCl and 100 mM TMAA in H_2O and directly injected the sample into an Exactive Orbitrap Mass Analyzer (ESI(-)-FTMS).

Since MS is limited to K^+ concentrations below 1 mM, Cameron Mackereth recorded the H1 proton region in ^{1}H-NMR to assess the number of G-quadruplex species and, if possible, the number of G-tetrads in concentrated KCl solution (> 1 mM K^+). NMR samples contain 100 µM DNA and 10 mM potassium phosphate (pH 7.0) in 90/10 H_2O/D_2O and were measured on a Bruker Avance NEO (700 MHz) at 278 K.

Results and Discussion

A three-tetrad quadruplex forms only when the four-tetrad quadruplex is disrupted

Figure 53 shows the CD and MS spectra of 22GT and its derivatives in 0.5 mM K^+. Each sequence is assigned a four-digit number indicating the number of guanines for each G-tract. The results for RAN4 and its derivatives are qualitatively similar (Figure S21), but more convoluted because the RAN4 G-quadruplex is unstable in low K^+ concentration. All DNA sequences (Table 2) had their CD spectra taken at 100 mM, 0.5 mM and 0.05 mM K^+ (Figure S22). Mass spectra in 0.05 mM K^+ are deposited in Figure S23 and Figure S24.

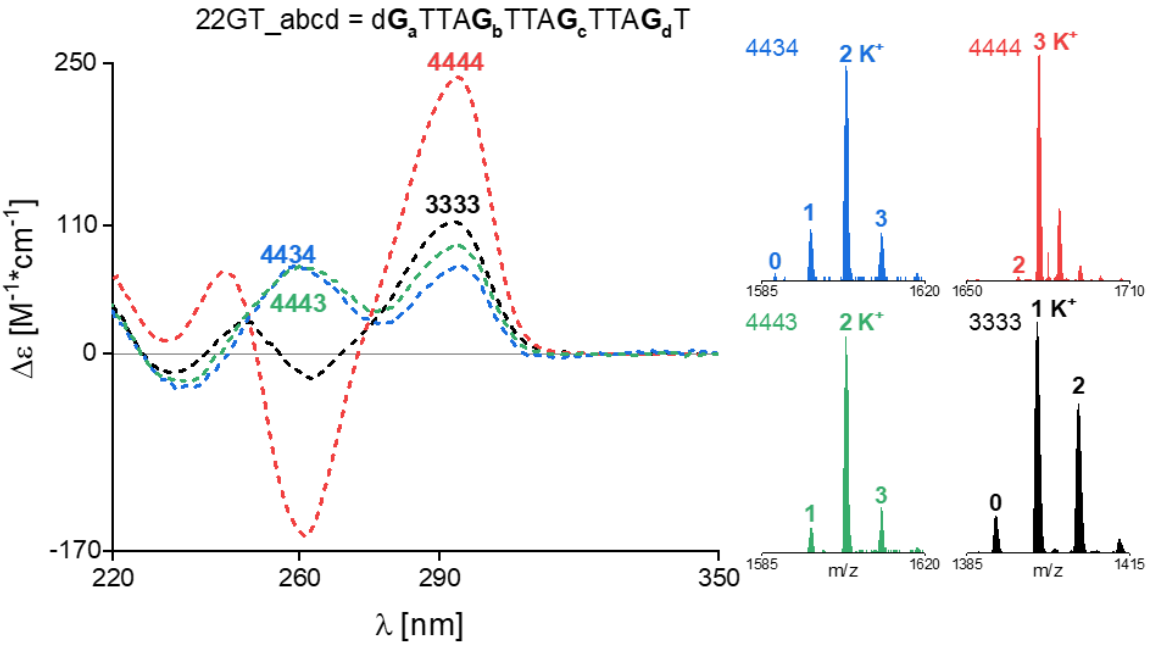


Figure 53 ***Left panel****: CD spectra of the 22GT sequence motif with varying G-tract lengths* ***Right panel****: Mass spectra corresponding to the CD signatures, showing the 5- charge state. Samples contain 10 µM DNA, 0.5 mM KCl and 100 mM TMAA (pH 6.8).*

The wild-type sequence 22GT ('3333') has some unfolded strand (0 K^+) and 1/2 specific K^+ adducts. The CD signature and specific 1 K^+ adduct are consistent with the 2-tetrad antiparallel G-quadruplex structure resolved by NMR in 96 mM KCl (PDB: 2KF8).[3] The second specific K^+ ion is also part of the 2-tetrad G-quadruplex: it is located between the GGG triad and the adjacent G-tetrad, where it is protected from solvent exchange.[17]

22GT_4444 has a characteristic antiparallel CD signature with a maximum close to 250 $M^{-1}cm^{-1}$. That intensity and the 3 specific K^+ adducts in the mass spectrum are sufficient evidence to confirm a 4-tetrad antiparallel G-quadruplex. Thus, injecting 4 guanines into the 22GT sequence converted the structure from a 2-tetrad to a 4-tetrad G-quadruplex, skipping over the 3-tetrad conformation. To force the formation of a 3-tetrad G-quadruplex, the number of guanines in the G-tracts needs to be more than 12, but less than 16.

The G-tracts of 22GT_4434 and 22GT_4443 contain 15 guanines, preventing the formation of 4-tetrad G-quadruplex. The main adduct is 2 K^+, which can be either a 3-tetrad G-quadruplex or a 2-tetrad G-quadruplex with a GGG triad. The CD spectra are

insufficient to assign a topology, but let us deduce the following: not parallel, not entirely antiparallel – it can be any mix of antiparallel/hybrid species. Further investigation is needed to comment on the number of species and their respective topologies.

Separating the two main conformers of the three-tetrad G-quadruplexes

We recorded NMR spectra of 22GT_4434, 22GT_4443, RAN4_3443 and RAN4_4334 in excess of K^+, showing the H1 proton region in Figure 54. None of these sequences are fully monomorphic, having at least 20 distinct H1 peaks. RAN4_3443 has 12 high-intensity H1 peaks, so the main species is likely a 3-tetrad G-quadruplex. If we limit the number of species (significantly) contributing to the CD signature of RAN4_3443 (Figure S22) to only one, we can deduce that it is a hybrid G-quadruplex. 22GT_4434 has several high-intensity peaks, indicating there is one major species and at least one other minor species. The peaks are not resolved enough to determine whether the major species has 8 or 12 H1 protons. 22GT_4443 and RAN4_4334 show a high degree of polymorphism.

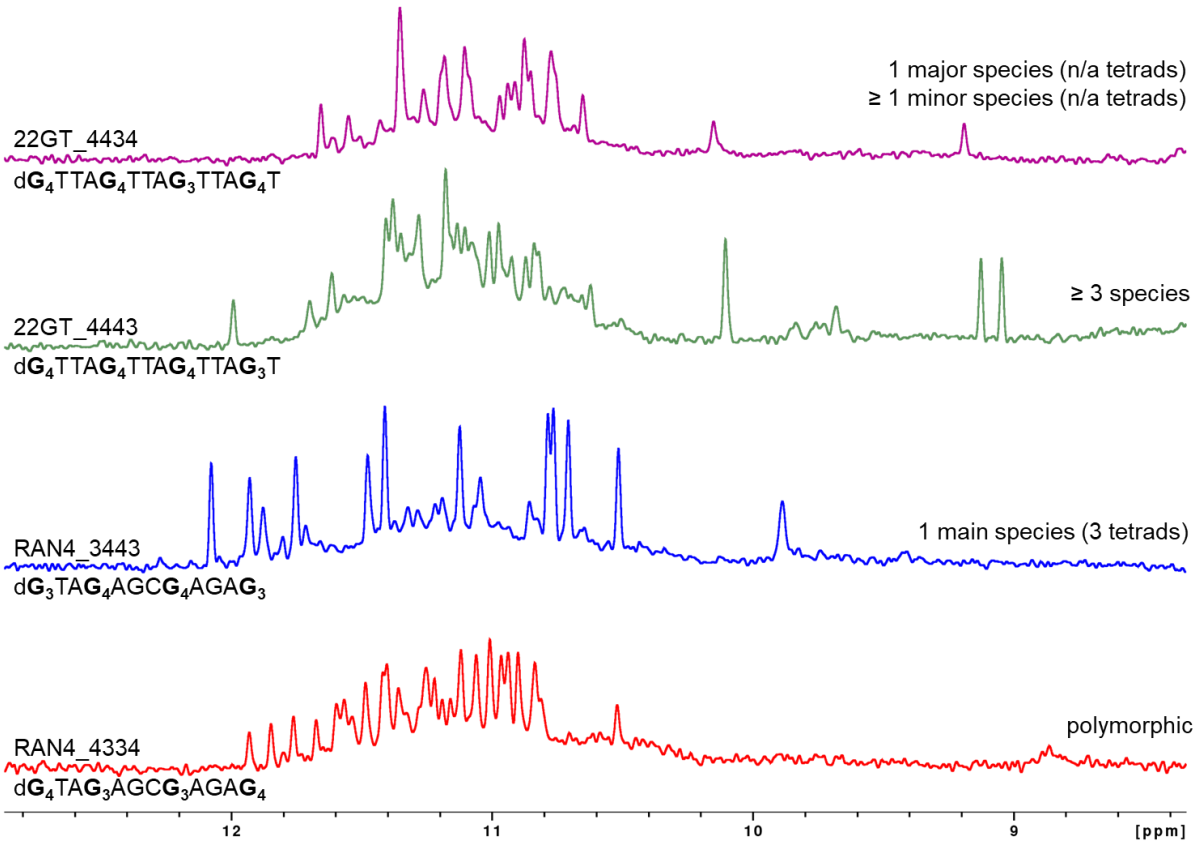


Figure 54 ^{1}H-NMR spectra of partially extended G-quadruplex sequences, showing the H1 proton region. Samples contain 100 µM DNA and 10 mM potassium phosphate (pH 7.0, 16 mM K^+) in 90/10 H_2O/D_2O. T = 278 K.

Figure 55 shows the temperature-dependent CD spectra and melting curves at 260/290 nm for 22GT_4443 in 0.5 and 100 mM K^+. Full datasets for all four sequences are deposited in the supporting information (Figure S25 to Figure S29).

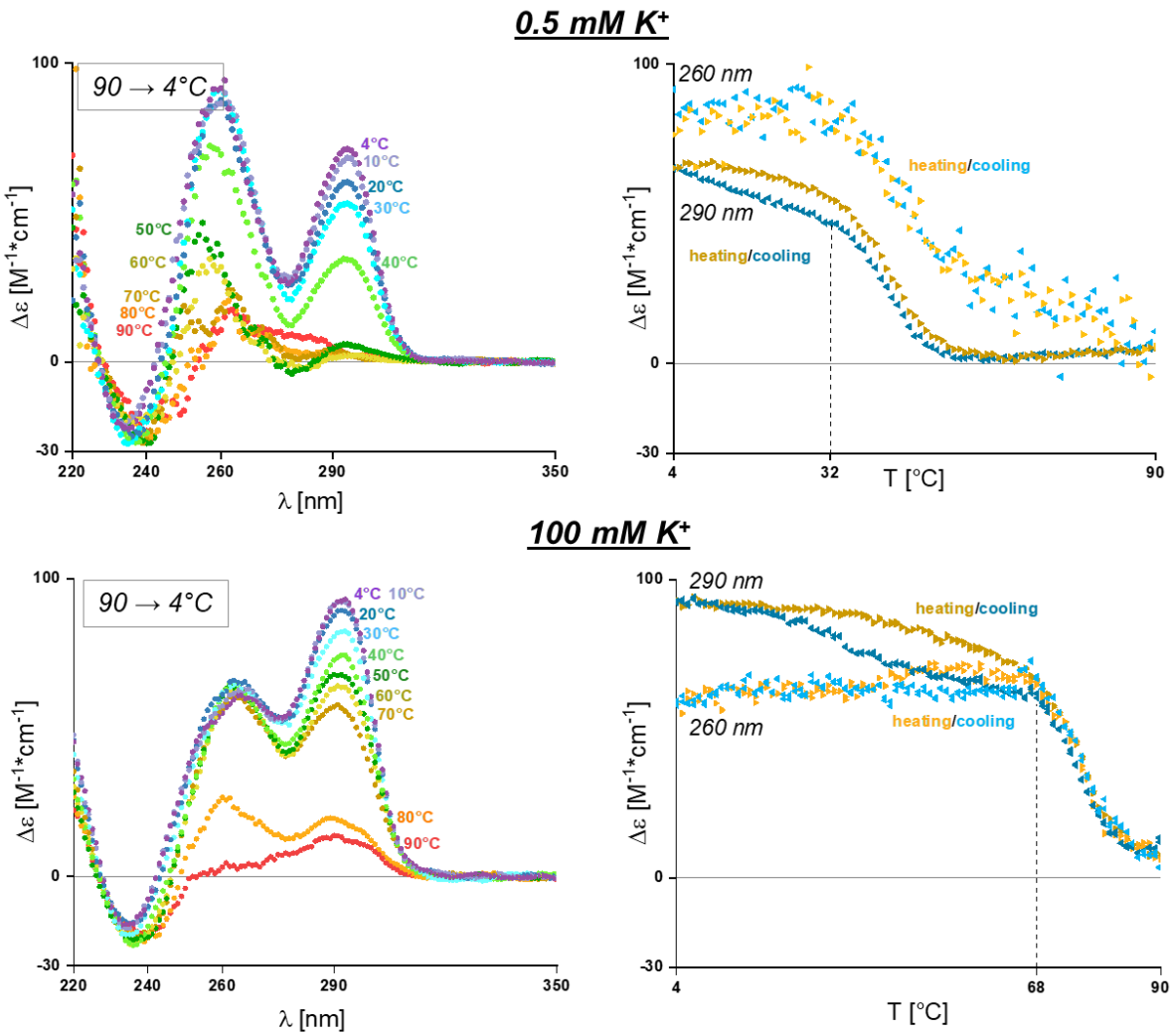


*Figure 55 CD melting data for 22GT_4443 (d**G₄**TTA**G₄**TTA**G₄**TTA**G₃**T) showing temperature-dependent CD spectra for the cooling ramp on the left and the transition curves around the two peak maxima at 260/290 nm on the right. The experiment was done in 100 (top) and 0.5 mM K⁺ (bottom). Sample conditions are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8).*

In both K^+ concentrations, we appear to have two species that transition in separate temperature ranges. First transition: the signal decreases at 290 nm and does not change at 260 nm. Ergo, this species has a positive band at 290 nm and a neutral band at 260 nm. It is probably an antiparallel G-quadruplexes which, like the 22GT wild-type sequence, has a weak negative band at 260 nm. The hysteresis is an indicator of slow folding or unfolding kinetics. Second transition: signal decrease at 260 and 290 nm. Ergo, this species has two positive bands at 260 and 290 nm, it is probably a hybrid G-quadruplex. The melting transitions are shifted by ca. 30°C between the two K^+ concentrations, but look qualitatively similar. The samples in 100 mM and 0.5 mM K^+ thus likely contain the same conformers, but the equilibria are not the same because of the different K^+ concentrations.

For a better separation of the two signatures, we performed a singular value decomposition (SVD) on the CD melting data. Figure 56 shows the first two basis spectra ('components') that were extracted from the U-matrix. From the singular values (Figure 57, Figure S30) we assessed that the first two components are sufficient to rationalize the temperature-dependent changes of CD spectra.

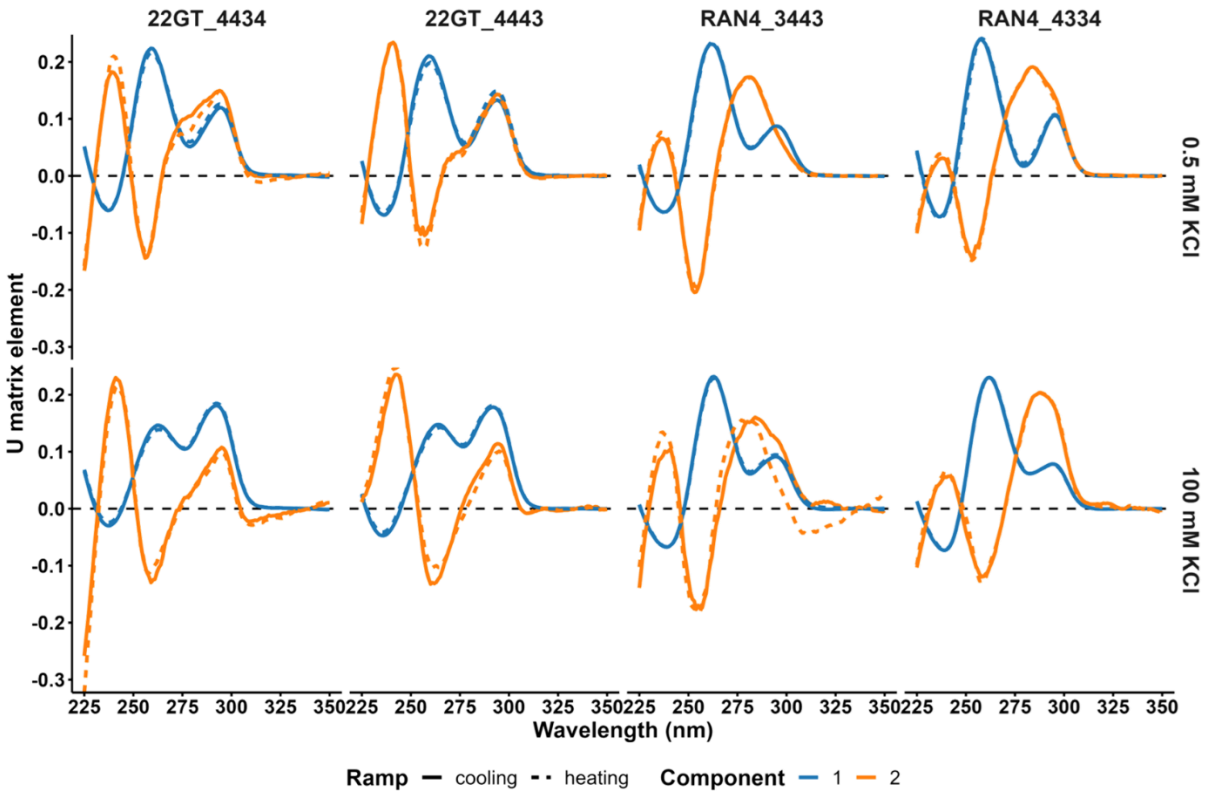


Figure 56 Basis spectra of the first two components from the U-matrix.

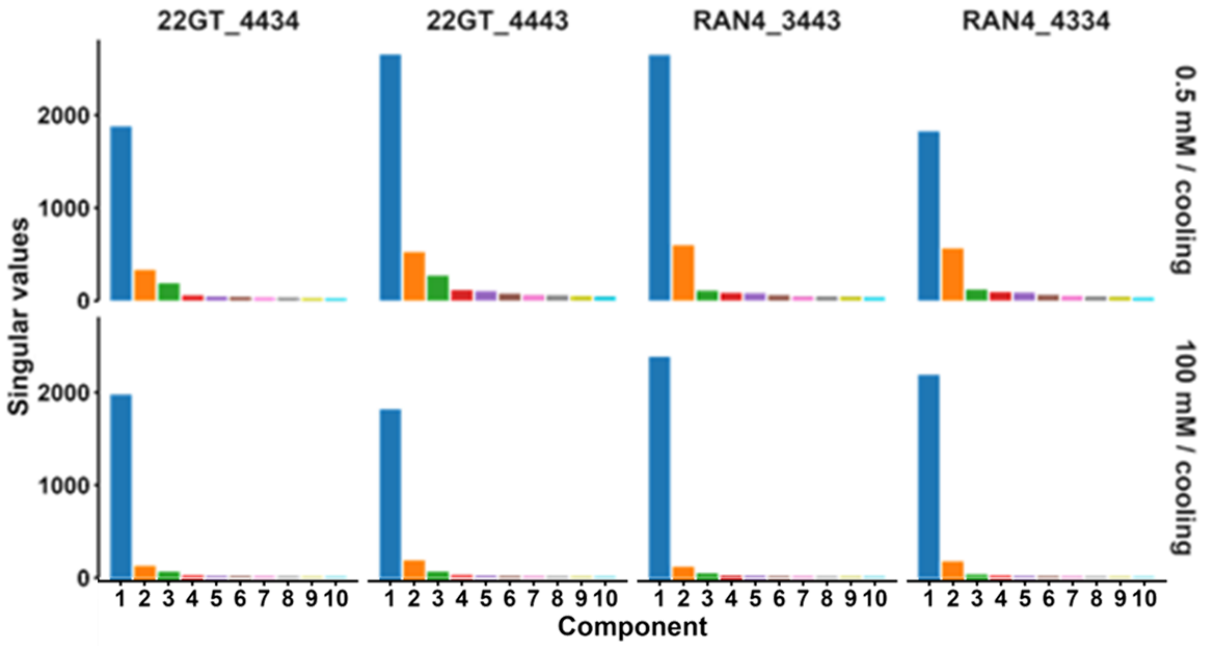


Figure 57 Singular values of the first 10 components for each CD melting dataset, taken from the S-matrix.

The CD signatures in Figure 56 are qualitative and cannot be translated into Δε values. Component 1 has a hybrid signature and the highest singular value, signifying that it has the highest mean signal contribution across the entire temperature range. We advise against guessing whether component 1 will be the most abundant species at room temperature, because the conformational ensemble changes with temperature and both the U- and S-matrix are temperature-independent by design. We will therefore not use the mass spectra measured at 25°C to assess how many specific K^+

ions component 1 has. What we should have done is perform an MS melting experiment in parallel and obtain basis component mass spectra. Based on the singular values, we could then assign the component mass spectra to their respective component CD spectra.

The NMR clearly showed a 3-tetrad structure for RAN4_3443, so component 1 has to be a 3-tetrad hybrid G-quadruplex. With all instances of component 1 being hybrid signatures and RAN4_3443 being a confirmed 3-tetrad G-quadruplex, we are going to make an educated guess that all instances of component 1 are in fact 3-tetrad hybrid G-quadruplexes. Another hypothesis on RAN4_3443 (d**G$_3$**TA**G$_3$**GAGC**G$_4$**AGA**G$_3$**) is that because its CD signature and sequence motif are similar to 2LOD (d**G$_3$**AT**G$_3$**ACACA**G$_4$**AC**G$_3$**), these two sequences might have a similar structure, which is a hybrid G-quadruplex with a propeller loop, followed by a diagonal loop and a lateral loop.[18]

Component 2 has an antiparallel CD signature, which does not necessarily mean heterostacking, because 5YEY also has an antiparallel CD signature but is hybrid stacking.[4] Nonetheless, component 2 seems like a promising candidate for a 3-tetrad heterostacking G-quadruplex. Although we cannot determine the quantity of component 2 at 25°C, the mass spectra in 0.5 mM K^+ (Figure 53, Figure S21) limit our options to 1 or 2 specific K^+ ions. It therefore has to be a 2-tetrad or a 3-tetrad antiparallel G-quadruplex.

The antiparallel conformer of 22GT_4443 is metastable in 100 mM K^+

We prepared and measured our G-quadruplexes at room temperature. In some instances, the CD signature underwent significant changes after the melting experiment (Figure 58)

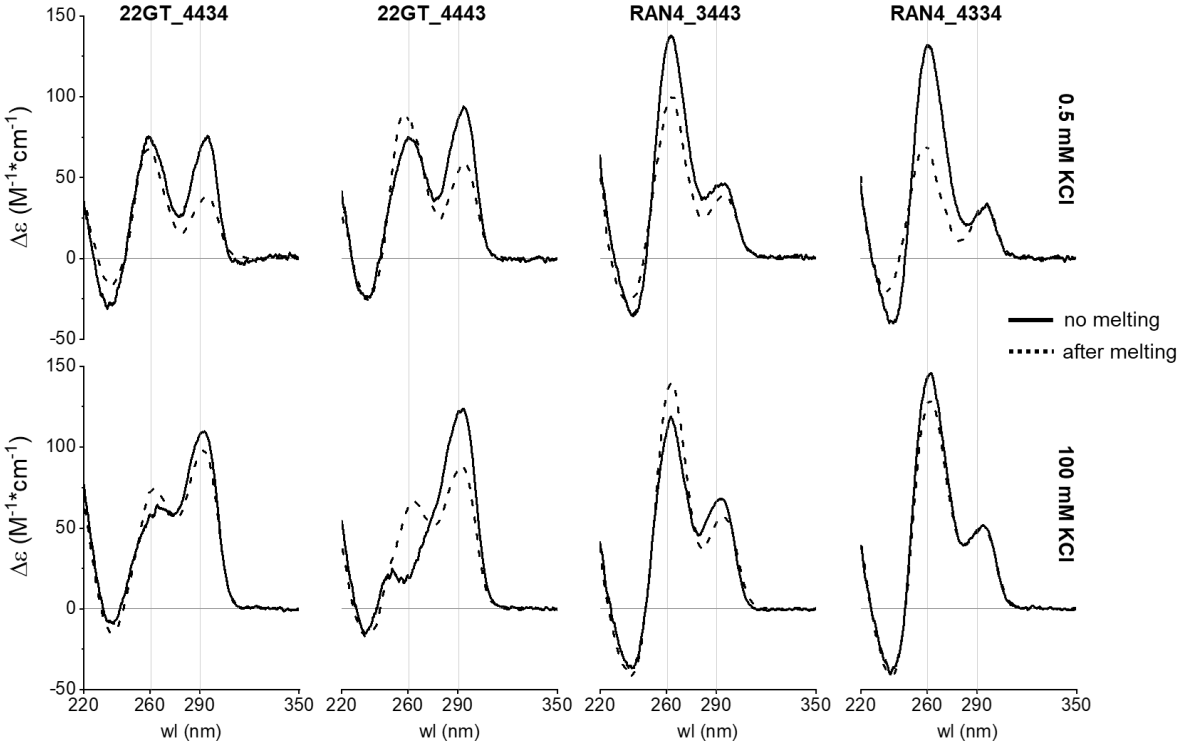


Figure 58 Comparing CD signatures of the G-quadruplexes folding in the same sample matrix at room temperature (full line) and after cooling from 90 to 22°C at a rate of ca. 0.2 °C/min (dashed line). Sample conditions are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8). T = 22°C.

The RAN4 derivatives qualitatively retain the same CD signature after CD melting, even if quantitative fluctuation is seen. 22GT_4443 undergoes the most significant changes: the CD band at 260 nm has increased, while the CD band at 290 nm has decreased. 22GT_4434 undergoes similar changes, but they are less remarkable. The changes imply that annealing and slow cooling causes an antiparallel to hybrid conversion. The hysteresis during the CD melting at 290 nm (Figure 55) is a sign of slow kinetics affecting the antiparallel conformer. Were the hybrid form affected, the hysteresis would also manifest at 260 nm, but it does not.

Antiparallel G-quadruplexes (especially those with 2 G-tetrads) are characterized by fast folding kinetics.[17,19] Therefore, the cause of the hysteresis must be slow unfolding kinetics of the antiparallel conformer. The annealing and slow cooling promotes the formation of the hybrid conformer, because it is thermodynamically favored. When K^+ is added at room temperature, the antiparallel conformer folds fast and does not unfold to allow the formation of the energetically more favorable hybrid conformer. This phenomenon is known as kinetic trapping.

To validate our claim that the antiparallel conformer is a kinetic trap, we performed a kinetics experiment where we added 100 mM K^+ to 22GT_4443 at 4°C (Figure 59).

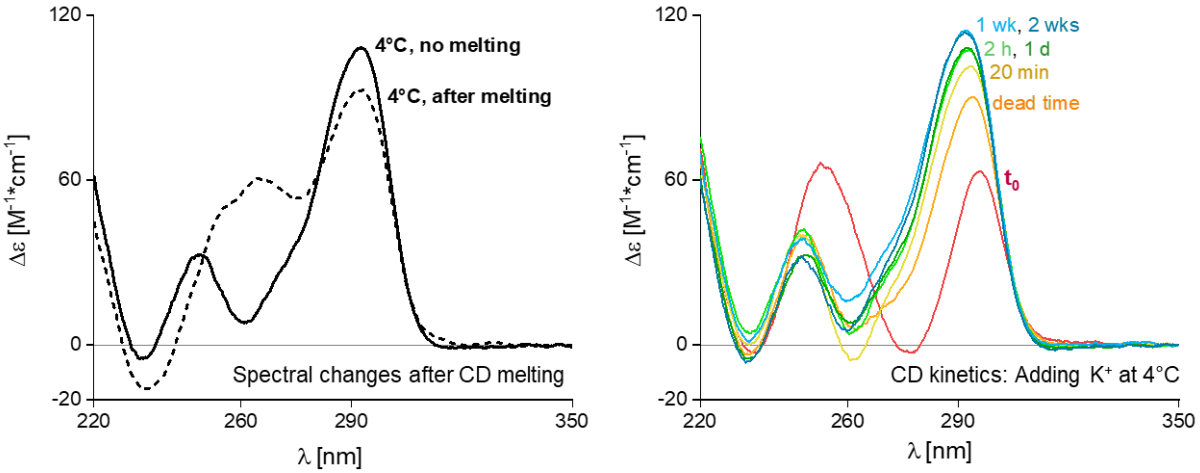


Figure 59 ***Left****: Comparing the CD signatures of 22GT_4443 (d****G₄****TTA****G₄****TTA****G₄****TTA****G₃****T) without and after a CD melting experiment.* ***Right****: CD spectra of 22GT_4443 measured before (t0) and after adding 100 mM KCl at certain time points (dead time is 2-3 min). Sample conditions are 10 µM DNA, 100 mM KCl, 10 mM TMAA (pH 6.8). T = 4°C.*

When formed at 4°C, 22GT_4443 has an authentic antiparallel CD signature. Over the course of 2 weeks, there were no changes that would indicate an antiparallel to hybrid conversion. So the antiparallel conformer is indeed a kinetic trap. We summarize our findings in Figure 60 by proposing a folding pathway for the 22GT_4443 G-quadruplex.

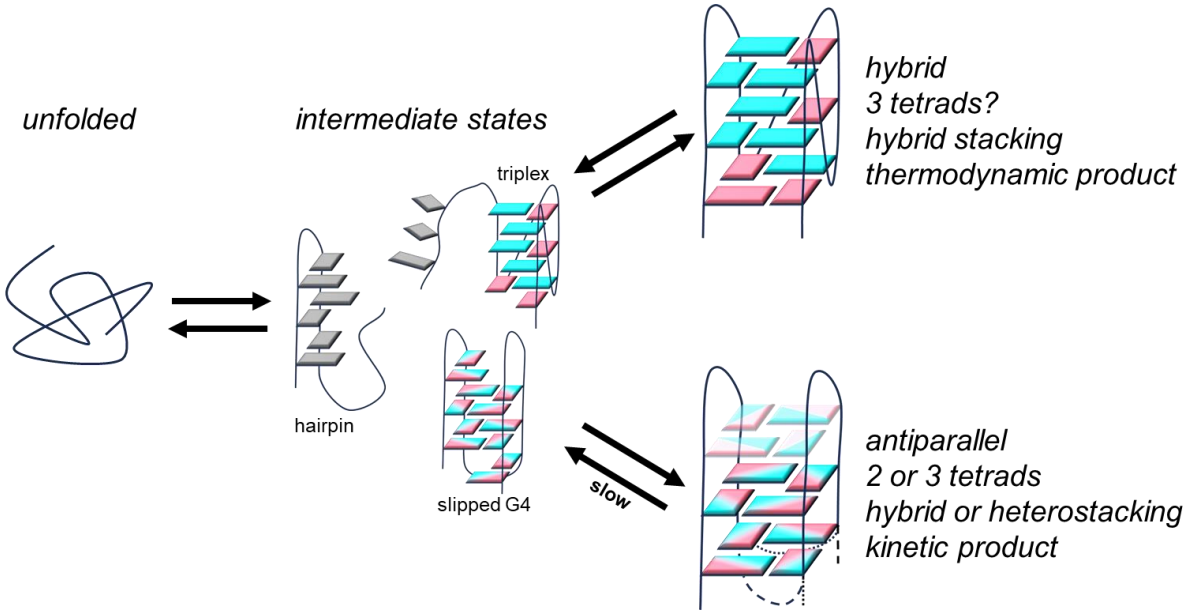


Figure 60 The proposed folding pathway for 22GT_4443.

Two separate folding pathways lead to either the hybrid or the antiparallel conformer, probably through different intermediate states. Theoretical studies suggest a triplex as the intermediate for hybrid/parallel G-quadruplexes, while antiparallel G-quadruplexes arrange through double hairpins or slipped-strand quadruplexes.[20,21] The different intermediate states are the most likely cause as to why antiparallel G-quadruplexes fold quicker than hybrid/parallel G-quadruplexes. The antiparallel conformer is a kinetic trap and obtained by G-quadruplex formation at low temperature. It has one or two specific K^+ ions and thus 2 or 3 G-tetrads. The hybrid conformer is energetically favored and obtained by annealing and slow cooling. We speculate that it has 3 G-tetrads.

Conclusion

Our model sequences, 22GT and RAN4, have G-tracts containing 12 guanines. They form 2-tetrad antiparallel G-quadruplexes stabilized by a GGG and an AGA triad. We injected four guanines into the G-tracts, which caused a conversion from a 2-tetrad to a 4-tetrad antiparallel G-quadruplex. In order to obtain 3-tetrad G-quadruplexes we had to disrupt the 4-tetrad conformation by leaving 14-15 guanines in the G-tracts.

We identified several G-quadruplex species with two specific K^+ ions, which can either mean three G-tetrads or two G-tetrads and a GGG triad. We separated the CD signatures of coexisting species by singular value decomposition of CD melting data. We identified two main conformers, which we hypothesize are a 3-tetrad hybrid G-quadruplex and an antiparallel G-quadruplex with an uncertain number of G-tetrads. The hybrid form is thermodynamically favored and obtained by annealing while the antiparallel form is kinetically favored and obtained by adding K^+ at low temperature.

We were unable to prove the existence of a 3-tetrad heterostacking G-quadruplex, but learned that the cation determines which G-quadruplex conformations are possible and which ones are not. We also learned that CD-melting is not ideal to analyze the antiparallel species because different equilibria are formed at different temperatures. Isothermal K^+ titrations and/or kinetics are more appropriate and could have been directly compared to MS experiments. The only thing to keep in mind is that one species is a kinetic trap, so the titrations should be repeated with and without annealing. Kinetic experiments should involve a temperature-jump setup, to better deal with the fast kinetics of the antiparallel conformer.

Trying to model a structure where AGA/GGG triads are supposed to stabilize heterostacking G-tetrads made us realize how little we know on the influence of base pairs/triads/tetrads on the arrangement of the adjacent G-tetrad. We believe that these loop interactions are the key to creating exceptional G-quadruplex conformations, so here are three project ideas:

1) a systematic review of solution NMR structures that feature base pairs or triads. A few simple questions to address: are adenines (in base pairs/triads) always in *anti*-configuration? Do adenines preferentially stack onto *syn*- or *anti*-guanines? Do triads have a defined H-bond rotation? If so, are they hetero- or homostacking onto the adjacent G-tetrad?

2) *in silico* experiments using 6ZX7 as a structural model to quantify the energy contributions of A/G and C/G base stacking. Flipping the bases between *syn*/*anti* configuration might elucidate which base stacks (*syn*/*anti*, *anti*/*syn*, *anti*/*anti*, *syn*/*syn*) are favorable or unfavorable.

3) modifying the 5YEY sequence to enable the formation of an AGA triad at the 5' end (d<u>A</u>**G₃**TTA**G₃**TTA**G₃**TTT**G₃**<u>G</u>). AGA triads require antiparallel loop arrangements, hence they promote a 2-tetrad antiparallel topology over a 3-tetrad hybrid topology.[5] 5YEY already has l-l-l loops, so an AGA triad should not convert the 5YEY structure into a 2-tetrad G-quadruplex. It would be interesting to check the modified 5YEY sequence by CD/MS to compare.

References

(1) Cang, X.; Šponer, J.; Cheatham, T. E., III. Explaining the Varied Glycosidic Conformational, G-Tract Length and Sequence Preferences for Anti-Parallel G-Quadruplexes. *Nucleic Acids Research* **2011**, *39* (10), 4499–4512. https://doi.org/10.1093/nar/gkr031.

(2) Liu, C.; Zhou, B.; Geng, Y.; Tam, D. Y.; Feng, R.; Miao, H.; Xu, N.; Shi, X.; You, Y.; Hong, Y.; Tang, B. Z.; Lo, P. K.; Kuryavyi, V.; Zhu, G. A Chair-Type G-Quadruplex Structure Formed by a Human Telomeric Variant DNA in K+ Solution. *Chem. Sci.* **2018**, *10* (1), 218–226. https://doi.org/10.1039/C8SC03813A.

(3) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131* (12), 4301–4309. https://doi.org/10.1021/ja807503g.

(4) Ghosh, A.; Largy, E.; Gabelica, V. DNA G-Quadruplexes for Native Mass Spectrometry in Potassium: A Database of Validated Structures in Electrospray-Compatible Conditions. *Nucleic Acids Research* **2021**, *49* (4), 2333–2345. https://doi.org/10.1093/nar/gkab039.

(5) Lenarčič Živković, M.; Rozman, J.; Plavec, J. Adenine-Driven Structural Switch from a Two- to Three-Quartet DNA G-Quadruplex. *Angewandte Chemie International Edition* **2018**, *57* (47), 15395–15399. https://doi.org/10.1002/anie.201809328.

(6) Pavc, D.; Wang, B.; Spindler, L.; Drevenšek-Olenik, I.; Plavec, J.; Šket, P. GC Ends Control Topology of DNA G-Quadruplexes and Their Cation-Dependent Assembly. *Nucleic Acids Research* **2020**, *48* (5), 2749–2761. https://doi.org/10.1093/nar/gkaa058.

(7) Kettani, A.; Gorin, A.; Majumdar, A.; Hermann, T.; Skripkin, E.; Zhao, H.; Jones, R.; Patel, D. J. A Dimeric DNA Interface Stabilized by Stacked A · (G · G · G · G) · A Hexads and Coordinated Monovalent cations11Edited by P. E. Wright. *Journal of Molecular Biology* **2000**, *297* (3), 627–644. https://doi.org/10.1006/jmbi.2000.3524.

(8) Jing, H.; Fu, W.; Hu, W.; Xu, S.; Xu, X.; He, M.; Liu, Y.; Zhang, N. NMR Structural Study on the Self-Trimerization of d(GTTAGG) into a Dynamic Trimolecular G-Quadruplex Assembly Preferentially in Na+ Solution with a Moderate K+ Tolerance. *Nucleic Acids Research* **2021**, No. gkab028. https://doi.org/10.1093/nar/gkab028.

(9) Creze, C.; Rinaldi, B.; Haser, R.; Bouvet, P.; Gouet, P. Structure of a d(TGGGGT) Quadruplex Crystallized in the Presence of Li+ Ions. *Acta Cryst D* **2007**, *63* (6), 682–688. https://doi.org/10.1107/S0907444907013315.

(10) Črnugelj, M.; Šket, P.; Plavec, J. Small Change in a G-Rich Sequence, a Dramatic Change in Topology: New Dimeric G-Quadruplex Folding Motif with Unique Loop Orientations. *J. Am. Chem. Soc.* **2003**, *125* (26), 7866–7871. https://doi.org/10.1021/ja0348694.

(11) Petraccone, L.; Erra, E.; Esposito, V.; Randazzo, A.; Mayol, L.; Nasti, L.; Barone, G.; Giancola, C. Stability and Structure of Telomeric DNA Sequences Forming Quadruplexes Containing Four G-Tetrads with Different Topological Arrangements. *Biochemistry* **2004**, *43* (16), 4877–4884. https://doi.org/10.1021/bi0300985.

(12) Zhang, D.-H.; Fujimoto, T.; Saxena, S.; Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **2010**, *49* (21), 4554–4563. https://doi.org/10.1021/bi1002822.

(13) Dickerhoff, J.; Dai, J.; Yang, D. Structural Recognition of the MYC Promoter G-Quadruplex by a Quinoline Derivative: Insights into Molecular Targeting of Parallel G-Quadruplexes. *Nucleic Acids Research* **2021**, *49* (10), 5905–5915. https://doi.org/10.1093/nar/gkab330.

(14) Bielskutė, S.; Plavec, J.; Podbevšek, P. Oxidative Lesions Modulate G-Quadruplex Stability and Structure in the Human BCL2 Promoter. *Nucleic Acids Research* **2021**, *49* (4), 2346–2356. https://doi.org/10.1093/nar/gkab057.

(15) Soenarjo, A. L.; Lan, Z.; Sazanovich, I. V.; Chan, Y. S.; Ringholm, M.; Jha, A.; Klug, D. R. The Transition from Unfolded to Folded G-Quadruplex DNA Analyzed and Interpreted by Two-Dimensional Infrared Spectroscopy. *J. Am. Chem. Soc.* **2023**. https://doi.org/10.1021/jacs.3c04044.

(16) Gray, R. D.; Chaires, J. B. Analysis of Multidimensional G-Quadruplex Melting Curves. *Curr Protoc Nucleic Acid Chem* **2011**, *CHAPTER*, Unit17.4. https://doi.org/10.1002/0471142700.nc1704s45.
(17) Marchand, A.; Gabelica, V. Folding and Misfolding Pathways of G-Quadruplex DNA. *Nucleic Acids Research*, 2016, *44*, 10999–11012. https://doi.org/10.1093/nar/gkw970 %J Nucleic Acids Research.
(18) Marušič, M.; Šket, P.; Bauer, L.; Viglasky, V.; Plavec, J. Solution-State Structure of an Intramolecular G-Quadruplex with Propeller, Diagonal and Edgewise Loops. *Nucleic Acids Research* **2012**, *40* (14), 6946–6956. https://doi.org/10.1093/nar/gks329.
(19) Gray, R. D.; Trent, J. O.; Chaires, J. B. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *Journal of Molecular Biology* **2014**, *426* (8), 1629–1650. https://doi.org/10.1016/j.jmb.2014.01.009.
(20) Stadlbauer, P.; Kührová, P.; Vicherek, L.; Banáš, P.; Otyepka, M.; Trantírek, L.; Šponer, J. Parallel G-Triplexes and G-Hairpins as Potential Transitory Ensembles in the Folding of Parallel-Stranded DNA G-Quadruplexes. *Nucleic Acids Research* **2019**, *47* (14), 7276–7293. https://doi.org/10.1093/nar/gkz610.
(21) Grün, J. T.; Schwalbe, H. Folding Dynamics of Polymorphic G-Quadruplex Structures. *Biopolymers n/a* (n/a), e23477. https://doi.org/10.1002/bip.23477.

CHAPTER 3: HELICAL FOLDAMERS AS SELECTIVE G-QUADRUPLEX LIGANDS.

Abstract

Our lab discovered (unintentionally) that small quinoline oligoamides, which fold into helices, bind to G-quadruplexes (GQs). Their helical structure greatly deviates from that of a typical G-quadruplex ligand, which is usually a flat aromatic heterocycle. We therefore decided to perform an MS ligand screening to address the following questions: 1) are foldamers GQ-specific? 2) are foldamers selective towards certain GQ topologies? 3) can foldamers compete with concurrent GQ ligands in terms of binding affinity? 4) do modifications of the oligomer sequence significantly influence GQ targeting? Our initial investigation concluded that foldamers are promising topology-selective G-quadruplex ligands. We then carried out structural investigations by CD and IMS to unravel what causes the topology selectivity of foldamers. We crystallized a foldamer-GQ-complex and resolved its crystal structure to validate the binding mode. NMR helped us sort out discrepancies between crystal structure and solution-phase structure. Overall, this project is a comprehensive analysis on the topology-selective interaction between foldamers and G-quadruplexes, establishing them as a novel ligand class.

State of the art

The free energy landscape of GQ folding results in many GQs co-existing in different folded states.[10–15] The structure of a singular GQ forming DNA sequence can depend on 1) The concentration of DNA strand 2) The type and concentration of cation[10,16–18] 3) The ionic strength and pH of the buffer[19,20] and 4) the nature of the solvent.[21,22] Over the decades, the solution-phase structures of several GQs have been resolved, which gives us a foundation of sequences that we can use to produce defined GQ structures for selectivity screening. We will use a previously assembled a database of resolved GQ structures featuring CD, MS, UV-melting and NMR data in both NMR and mass spectrometry conditions.[23]

Due to their involvement in gene expression, G-quadruplexes are attractive drug targets. G-quadruplex ligands are generally flat, condensed heteroaromatic systems that allow efficient π-stacking on G-tetrads while inhibiting dsDNA intercalation.[24] Cationic sidechains are often added to avoid charge repulsion and improve water solubility.[25] However, current ligands still lack the necessary selectivity among different G-quadruplexes for clinical use.[26,27] Ligands discriminating certain topologies[28–30] or topology-subclasses[31–33] have already been reported but rational design efforts have not yet yielded small molecules with significant selectivity for a specific G-quadruplex structure.[34] Molecular scaffolds that deviate from the condensed aromatic structure paradigm may be required to produce ligands with improved topology selectivity.[31]

Inspired by biopolymers, foldamers are synthetically constructed oligomers that self-organize into a defined folded structure.[35] We studied quinoline-based oligoamide foldamers that fold into a helix through varying interactions including electrostatic repulsions, intramolecular hydrogen bonds, conjugation and extensive aromatic stacking.[36–38] In absence of any chirality inducers, helical foldamers exist as a racemic mixture of left- and right-handed helix.[39] Foldamers possess two characteristic traits of GQ ligands: 1) A heteroaromatic core structure, which is derived from quinoline and 2)

Extended, flexible sidechains. But unlike most GQ ligands, foldamers are not flat. Our foldamers have a positively charged sidechain (derived from ornithine) to enhance water solubility (Figure 61).

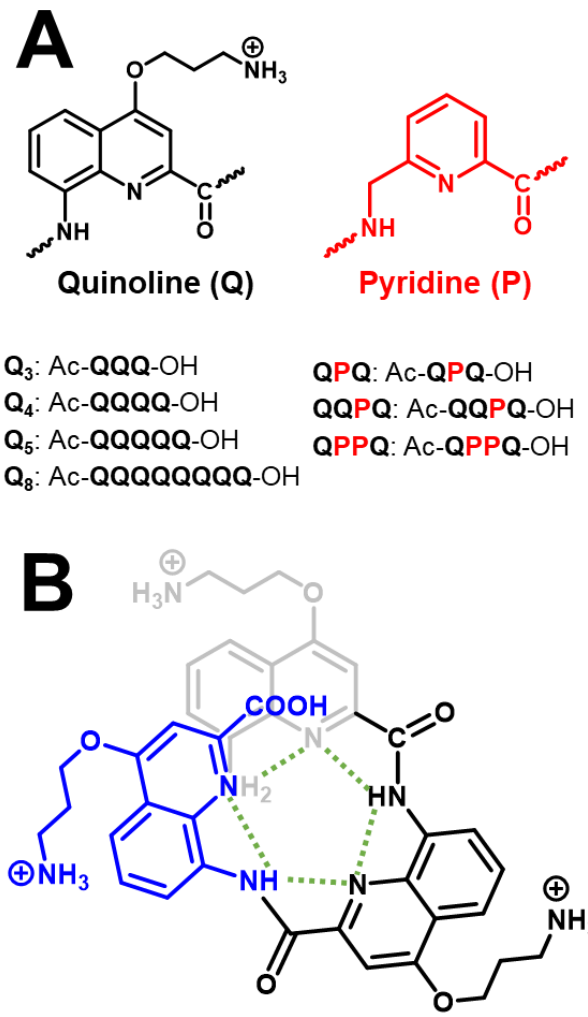


Figure 61 ***A)*** *Oligoquinoline foldamers used in this study, which contain quinoline (Q) and pyridine (P) subunits.* ***B)*** *Intramolecular H-bonds (green) sterically prohibit a planar structure, forcing the molecule into a (racemic) helical assembly.*

The interaction between quinoline oligoamides and GQs has been studied before. These foldamers stabilize GQs, but have no effect on DNA duplexes – based on FRET melting studies.[36,37,40] The foldamers display a remarkable selectivity among the probed GQ sequences. Unfortunately, the different structures/topologies of those sequences were not taken into account, so we cannot profoundly analyze the GQ selectivity. An X-ray crystal structure of a foldamer-GQ system showed co-crystallization with antiparallel GQ, but no specific interaction between ligand and target.[38] Negatively charged foldamers do not interact with GQs, likely due to charge repulsion.[37,41] Hence, aromatic foldamers designed as GQ ligands should carry positive charges, which are implemented through the side chain. The effect of foldamer helicity was probed by inducing the handedness quantitatively into a left *(M)* or right *(P)* configuration by covalently attaching either *(R)* or *(S)*-camphanic acid, respectively. However, introducing the camphanyl residue disrupts the foldamer-GQ interaction, indicating that the binding region is located at the *N*-terminus of the foldamer.[41] Helicenes derived from diazaoxatriangulenium (DAOTA) were separated using

enantioselective precipitation. The *(M)* and *(P)* helices have similar binding affinities to G-quadruplex, with K_D value differences less than factor 1.5, although molecular modeling implies that the flanking nucleotides interact differently with each form.[42]

Materials and Methods

Foldamer synthesis

The foldamers were synthesized and characterized by Vincent Lafflilé under the supervision of Yann Ferrand. Seven foldamer sequences (see Figure S34 and Figure S35) were prepared according to the synthetic schemes in the supporting information (Figure S36 to Figure S39), while also providing NMR spectra, HPLC traces and accurate masses for each ligand (Figure S40 to Figure S53).

The aromatic oligoamides were synthesized using a microwave-assisted solid phase synthesis approach. The matrix is a low loading Cl-MPA ProTide® resin. The first monomer unit was attached to the resin using CsI/DIEA. Subsequently, each monomer was coupled to the oligomer chain iteratively through periodic cycles of deprotection and then coupling the next monomer unit using PPh_3/Trichloroacetonitrile. Finally, trifluoroacetic acid was used to simultaneously cleave the oligomer from the resin and deprotect the Boc groups of the side chains to yield ammonium salts. After filtration of the remaining solids, TFA was evaporated leaving a residue, which was suspended in Et_2O and centrifuged. The residual solid was dissolved in water and freeze-dried.

All sequences were purified using semi-preparative HPLC on a reverse phase C18 column (mobile phase: H_2O/ACN/TFA). The purified foldamers were freeze-dried twice more to remove TFA traces. The final product was obtained as a yellow solid with a cotton-like texture, consisting of the foldamer with one trifluoroacetate counterion per ammonium groups.

DNA & sample preparation

All oligonucleotide sequences were purchased from Eurogentec (Belgium) and dissolved in UPLC/MS grade pure water (Biosolve Chimie, France) to 1 mM DNA. Stock solutions were annealed for 3 minutes at 85°C, then any residual Na^+ ions were exchanged with 500 mM NH_4OAc, which we then flushed out with water using cellulose-matrix 3K centrifugal filter units (MerckMillipore, Ireland). We validated the desalting method by analyzing a DNA stock solution before and after desalting (Figure S54). The concentration of each DNA stock solution was determined with a UV spectrophotometer (Uvikon XL Secomam), utilizing extinction coefficients at 260 nm that were calculated from the nucleotide sequence via nearest-neighbor method.[43] An overview of the sequences studied is provided in Table 6.
Foldamer stock solutions were prepared by weighing in the purified product on a microbalance (Sartorius ME5) and dissolving them in UPLC/MS grade pure water.

To form GQ, DNA solutions were doped with either potassium chloride from Sigma Aldrich and Trimethylammonium acetate solution (TMAA) from ChemCruz or ammonium acetate solution (NH_4OAc) from Sigma Aldrich (Switzerland).

Table 6 Overview of the DNA sequences used for the foldamer ligand screening experiments.

Name	PDB	Sequence	Topology[2]	Loops[3]	θ (20°C)[4]
1XAV	1XAV	dT**G**A**G**$_3$T**G**$_3$TA**G**$_3$T**G**$_3$TAA	Parallel GQ	ppp	100%
222T		dT**G**$_3$TT**G**$_3$TT**G**$_3$TT**G**$_3$T	Polymorphic GQ	ppp	98%

[2] Most abundant topology in MS sample conditions
[3] p = propeller, l = lateral, d = diagonal, s = snapback, v = v-loop
[4] Fraction of folded DNA at 20°C in 1 mM KCl, based on UV melting data

222T_mA		d**TG_3TTG_3AAG_3TTG_3**T	Polymorphic GQ	ppp	92%
222T_mC		d**TG_3TTG_3CCG_3TTG_3**T	Polymorphic GQ	ppp	96%
T30177TT	2M4P	d**TTGTGGTG$_3$TG$_3$TG$_3$**T	Parallel GQ	ppp	100%
26CEB	2LPW	d**AAG$_3$TG$_3$TGTAAGTGTG$_3$TG$_3$**T	Parallel GQ	ppp	100%
26CEB-mT		d**AAG$_3$TG$_3$TTTTTTTGTG$_3$TG$_3$**T	Parallel GQ	ppp	100%
2KYP	2KYP	d**CG$_3$CG$_3$CGCTAG$_3$AG$_3$**T	Parallel GQ	ppp	93%
2O3M	2O3M	d**AG$_3$AG$_3$CGCTG$_3$AGGAG$_3$**	Parallel GQ	ppls	84%
TG4T	2O4F	(dT**G$_4$**T)$_4$	Parallel GQ	none	N/A
21G		d**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**	Polymorphic GQ		100%
5YEY	5YEY	d**G$_3$**TTA**G$_3$**TTA**G$_3$**TTT**G$_3$**	Antiparallel GQ	lll	100%
22GT	2KF8	d**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**T	Antiparallel GQ	ldl	98%
22GT_18T		d**G$_3$**TTA**G$_3$**TTA**G$_3$**TTT**G$_3$**T	Antiparallel GQ		100%
22CTA	2KM3	dA**G$_3$**CTA**G$_3$**CTA**G$_3$**CTA**G$_3$**	Antiparallel GQ	lll	88%
TBA	148D	d**GG**TT**GG**T**GT**GG**TT**GG**	Antiparallel GQ	lll	90%
G4T4G4	4R47	(d**G$_4$**T$_4$**G$_4$**)$_2$	Antiparallel GQ	d/d	N/A
26TTA	2JPZ	dTTA**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**TT	Hybrid GQ	llp	66%
Bcl2	2F8U	d**G$_3$**C**G**C**G$_3$**A**GG**AATT**G$_3$**C**G$_3$**	Hybrid GQ	llp	97%
24TTG	2GKU	dTT**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**A	Hybrid GQ	pll	99%
24TTG_20T		dTT**G$_3$**TTA**G$_3$**TTA**G$_3$**TTT**G$_3$**A	Hybrid GQ		98%
23TAG	2JSM	dTA**G$_3$**TTA**G$_3$**TTA**G$_3$**TTA**G$_3$**	Hybrid GQ	pll	94%
2KPR	2KPR	d**G$_3$**T**G$_4$**AA**G$_4$**T**G$_3$**T	Hybrid GQ	llvp	100%
21CCC		d**C$_3$**TAA**C$_3$**TAA**C$_3$**TAA**C$_3$**	i-motif (pH 5.5) single strand (pH 7)		100%
ds26		(dCAATCGGATCGAA TTCGATCCGATTG)$_2$	Duplex/Hairpin		100%
DK-33		(dCGTAAATTTACG)$_2$	Duplex		100%
DK-66	1FQ2	(dCGCGAATTCGCG)$_2$	Duplex		100%
DK-100		(dCGCGGGCCCGCG)$_2$	Duplex		100%
ss24		dTGCCATGCTAC TGAGATGACGCTA	Single strand		
24nonG4		dTGGGATGCGACA GAGAGGACGGGA	Single strand		
T24		dT$_{24}$	Single strand		
A24		dA$_{24}$	Single strand		
T6		dT$_6$	Single strand		

Circular Dichroism spectroscopy

CD measurements were performed on a JASCO J-815 spectrophotometer equipped with a Lauda RE 305 temperature control system with the following parameters: 220-500 nm scan range, 50 nm/min scanning speed, 0.2 nm data pitch, 2 nm bandwidth, 2 s data integration time, 22°C temperature in the sample holder, 3 accumulations. The samples, containing 10 µM of DNA, were placed in Suprasil quartz cuvettes (Hellma) with 2 or 10 mm pathlength.

Raw data was blank subtracted then converted to molar ellipticity with Equation 12, where θ is the ellipticity (mdeg), c the molar concentration of oligonucleotide (mol/l), and l the pathlength (cm).

$$\Delta\varepsilon\ [M^{-1} \times cm^{-1}] = \frac{\theta}{3298.2 \times c \times l} \quad (12)$$

We provide CD spectra in MS conditions for all intramolecular GQ sequences listed in Table 6. They have either been published previously,[23] or are featured in the supporting information (Figure S55 to Figure S62).

Thermal denaturation

Thermal denaturation was monitored by UV absorption spectroscopy on a SAFAS UVmc2 double-beam spectrophotometer equipped with a high-performance Peltier temperature control unit. The melting ramps were: 20→90°C (10 °C/min), 90→4°C (0.2 °C/min), 4→90°C (0.2 °C/min); 90 s data reading interval, 0.5 s averaging time. The samples, containing 10 µM of DNA, were placed in 10 mm pathlength Suprasil quartz cuvettes (Hellma). The UV absorption was taken at 260, 295 and 335 nm. The 260 nm wavelength corresponds to the DNA duplex, the 295 nm one to the GQ and the absorption at 335 nm serves as an internal reference to correct for instrumental artifacts. In addition, the buffer alone (100 mM TMAA and 0.5/1 mM KCl) was measured for blank correction. Raw data, at 260 nm for dsDNA and 295 nm for G4s, was blank subtracted, and corrected with the 335 nm data, then converted to folded fraction θ (Equation 9).[44,45]

We provide UV melting curves for all intramolecular GQ sequences in MS conditions in Table 6. They have either been published previously,[23] or are featured in the supporting information (Figure S55 to Figure S60).

Native mass spectrometry: Instrumental setup

MS experiments were carried out on an Agilent 6560 IMS-QTOF. The instrumental settings were optimized to preserve non-covalent DNA structures, albeit making a compromise between signal intensity and softness of the ion transfer (Table S1).[46] Samples were infused at 3 µl/min for 11 minutes, working ligand-by-ligand to avoid cross-contaminations. The instrument is equipped with a drift tube for ion mobility spectrometry (DT-IMS). The ions pass through a tube, filled with helium, and we measure the time between ions being released into the drift tube until arriving at the detector. This time is the arrival time t_A, which we convert into the collisional cross section (CCS) through the Mason-Schamp equation (Equation 4).

We validated our instrumental settings before each session using $(dTG_4T)_4$*$3NH_4^+$ as an external calibrant. We accepted the instrumental settings when the experimental CCS values were within 2% of error to our previously published CCS values (787.5 Å² [5-] ion, 735.7 Å² [4-] ion).[47]

Ligand screening via native mass spectrometry

We designed a DNA panel for G-quadruplex ligand screening reflecting the diversity of G-quadruplex topologies (Table 6). We selected 23 G4-forming sequences featuring different strand stoichiometries (1, 2, 4), numbers of tetrads (2 to 4), topologies (parallel, antiparallel, hybrid) and origins (*e.g.*, oncogene promoter, telomeres, aptamers). The panel features control sequences of alternative secondary DNA structures (duplex, i-motif, single strand).

We screened the DNA panel against 4 different foldamer sequences: QPQ, QQPQ, QPPQ and QQQQ. Screening samples contained 10 µM DNA, 20 µM ligand, 4 µM dT_6, 0.5 mM KCl and 100 mM TMAA (pH 6.8) in water. 150 mM ammonium acetate (pH 6.8) was used for the sequences T_6, T_{24}, ss24, DK-66 and 21CCC. The multi-stranded GQs TG_4T and $G_4T_4G_4$ were screened in both TMAA/KCl and ammonium acetate.

Eric Largy processed the mass spectra by determining the concentration of any DNA species (M*) using Equation 13, where $[M]_0$ is the total concentration of DNA in the sample and *I* is the intensity of the different species, obtained by assigning a m/z window to every species and integrating all signal inside that window. (monomeric DNA: M, ligand: L).[48] The concentration of unbound ligand is determined by equation 14, where $[L]_0$ is the total concentration of ligand in the sample. The dissociation constants K_{D1} and K_{D2} values for the 1:1 and 2:1 (*L*:*M*) complexes were then determined from Equations 15/16. The fraction of bound DNA (Equation 17) is a more concise parameter to present ligand screening results.

$$[M^*] = [M]_0 \times \frac{I(M^*)}{\sum I(M)+I(ML)+I(ML_2)} \quad (13)$$

$$[L] = [L]_0 - [L]_{bound} = [L]_0 - [ML] - 2 \times [ML_2] \quad (14)$$

$$ML \leftrightarrow M + L;\ K_{D1} = \frac{[M]\times[L]}{[ML]} \quad (15)$$

$$ML_2 \leftrightarrow ML + L; K_{D2} = \frac{[ML]\times[L]}{[ML_2]} \quad (16)$$

$$fraction\ of\ bound\ DNA = \frac{[ML]+[ML_2]}{[M]+[ML]+[ML_2]} \quad (17)$$

The difference in K_{D1} and K_{D2} highlights ligand cooperativity. We speak of positive cooperativity when $K_{D2} < 4*K_{D1}$ and negative cooperativity when $K_{D2} > 4*K_{D1}$.[49]

ESI-MS titrations

The six sequences 1XAV/222T/21G/5YEY/T24/ss24 were titrated against 7 foldamers: QPQ, QPPQ, QQPQ, QQQ, QQQQ, QQQQQ, QQQQQQQQ. The latter will from now on be called Q_3, Q_4, Q_5 and Q_8. The MS titration experiments follow two main objectives: 1) Obtain more reliable K_D values compared to using a single spectrum and 2) Study the effect of foldamer length on GQ affinity to aid rational ligand design. We kept the DNA concentration at 10 µM and measured at 7 different ligand concentrations: 0/5/10/15/20/30/40 µM.

Eric Largy processed the MS titration data in three main steps: 1. Noise correction. To quantify the noise, we picked a 'signal-free' area in each mass spectrum and calculated the standard deviation of intensities. He then subtracted three times the standard deviation from the integrated DNA signals to remove noise contributions. The ligand screening data was noise corrected as well. 2. Response analysis. Response factors depend on the gas-phase transmission yield during electrospray ionization which can be different for the M/ML/ML_2 species. Changes in response were quantified by adding dT_6 as an internal calibrant.[50] The ratio of DNA to dT_6 intensity was monitored as a function of ligand concentration and from changes in that ratio Eric Largy derived the effect of ligand concentration on DNA response (Figure S63). Note that the ligands not

interact with dT_6 3. Dynamic fitting. Eric Largy used the software DynaFit which derives a mathematical model from the chemical equilibria 15/16 and matches it to the experimental data with minimal residual error (Figure S64). His model allows the response factors of the complexes (ML, ML_2) to vary between 0.8 to 1.1. From the fitted data he extracted K_{D1} and K_{D2} values.

X-ray crystallography and structure determination

Crystallogenesis was performed under supervision of Stéphane Thore. We used a sitting-drop well plate (SWISSCI MRC 2 Lens Crystallisation Plate) and obtained droplets by mixing 200 nl of a solution containing 400 µM DNA (222T), 900 µM ligand (QQPQ) and 20 mM potassium cacodylate (pH 7.0) with an equivalent volume of crystallization screening solutions from various commercial kits. The original condition was the No. 41 of the Natrix crystallizazion screen (Hampton Research) and was composed of 100 mM KCl, 15 mM $MgCl_2$, 50 mM Tris (pH 7.5) and 10% PEG-550. We obtained clear, cuboid crystals that were 10-20 µm in size. The crystals were transferred into a cryoprotective solution containing 100 mM KCl, 15 mM $MgCl_2$, 50 mM Tris (pH 7.5), 20% PEG-550 and 20% ethylene glycol. Finally, the crystals were flash frozen at -196°C and analyzed at the SOLEIL synchrotron (Saint-Aubin, France).

Several data sets were collected at the French synchrotron Soleil on the beamline Proxima I to a maximum resolution of 2.20 Å. The raw images were processed with XDS.[51] Stéphane Thore solved the crystal structure using the Phaser-Molecular Replacement procedure from the software package Phenix.[52] The structure was solved using the GQ core from the previously solved structure of the 222T sequence (PDB: 6P45).[53] He tested various models with and without the connecting nucleotide to help remove model bias. He improved the GQ model by several rounds of manual rebuilding and added the QQPQ ligand once the electron density from the GQ was filled. QQPQ ligand restrains were generated using the webserver from GlobalPhasing[54] and added to the refinement procedure in Phenix. The final model included two potassium ions located between the G-tetrads of the same GQ and one potassium ion located between the G-tetrads of two dimerizing GQs coordinated by the guanine bases, several magnesium ions and 18 nucleotides of the GQ sequence (missing only the first and the last thymine nucleotide). The summary of data collection and refinement statistics is presented in Table S5. Figures were prepared using ChimeraX.[55]

NMR spectroscopy

NMR samples contain 100 µM DNA, 300 µM QQPQ and 10 mM potassium phosphate buffer (pH 7, contains 16 mM K^+) in 90%/10% H_2O/D_2O. We used the similar sequence T95-2T ($dTTG_3TG_3TG_3TG_3T$, PDB: 2LK7) because it is monomorphic and produces clean NMR spectra.[56] We labeled the flanking thymines through T/U switches, adding the sequences 2LK7_1U ($d\mathbf{\underline{U}}TG_3TG_3TG_3TG_3T$), 2LK7_2U ($dT\mathbf{\underline{U}}G_3TG_3TG_3TG_3T$) and 2LK7_18U ($dTTG_3TG_3TG_3TG_3\mathbf{\underline{U}}$).

Cameron Mackereth measured every DNA sequence with and without QQPQ. He measured QQPQ by itself in d6-DMSO instead of H_2O/D_2O and KP. 1D-NMR spectra were recorded on a Bruker Avance NEO 700 MHz at 278 or 298 K. ^{1}H-NOESY and TOCSY measurements were recorded on a Bruker Avance-III 800 MHz at 278 K.

Results and Discussion

Ligand Screening

Introduction

Mass spectrometry can directly validate the presence of a species of interest, e.g. a DNA sequence, based on its mass. If different species have different masses, they are separated in a mass spectrum. In the context of foldamer-GQ systems, we can determine both the number of ligands and number of cations bound to the DNA sequence.

For example, the m/z shift for one ligand in Figure 62 is equivalent to the mass of the ligand (1031 Da) divided by the number of charges (4), which is ca. 258 m/z. We see a 1:1 and 2:1 complex, so we can assume there are two accessible foldamer binding sites on the GQ. In addition, the mass difference matches the mass of the ligand, validating that it is the ligand binding and not an impurity or a by-product.

Next, each K^+ bound there causes a mass shift of 38 Da (^1H detaches, ^{39}K attaches). K^+ ions bind in two ways: non-specific and specific. Non-specific K^+ ions interact electrostatically with negative phosphate groups on the DNA backbone; their distribution follows a discrete probability, which can be approximated as a Poisson distribution.[57] Specific K^+ ions in GQs are the ones located between 2 G-tetrads. An adduct with *n* K^+ ions is considered specific when it is significantly more populated than its neighboring adducts *n+1* and *n-1*.

In Figure 62, the main adduct of the unbound DNA is 1 K^+ which corresponds to a 2-tetrad antiparallel GQ.[10] But the 1 K^+ peak is not purely GQ. When the neighboring 0 K^+ peak is that intense, there will be a significant portion of unfolded strand with 1 unspecific K^+ which contributes to that signal. The main adduct of the complex is 2 K^+, which corresponds to a 3-tetrad parallel GQ. The absence of 0/1 K^+ adducts indicates ligand selectivity towards 3-tetrad parallel GQs over 2-tetrad antiparallel GQs and unfolded DNA.

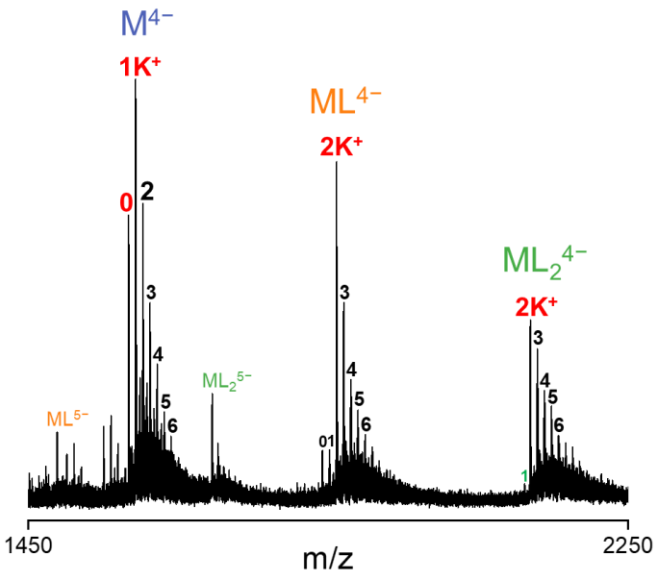


*Figure 62 Mass spectrum of 10 µM 222T-mA (dT**G₃**TT**G₃**AA**G₃**TT**G₃**T) with 20 µM QQQQ in 0.5 mM KCl and 100 mM TMAA. Labeled species are: Unbound DNA (M), 1:1 complex (ML) and 2:1 complex (ML_2), specific K^+ adducts (red), non-specific K^+ adducts (black).*

Ligand affinity from MS

In Figure 63 we can visually compare the mass spectra of two ligands with different binding affinities. Each ligand:DNA stoichiometry (blue/green/orange in Figure 63) is spread across multiple charge states. Note that these charge states do not correspond to the charge state in solution, but to the charges left in the final droplet during electrospray ionization before the analyte enters the gas phase.

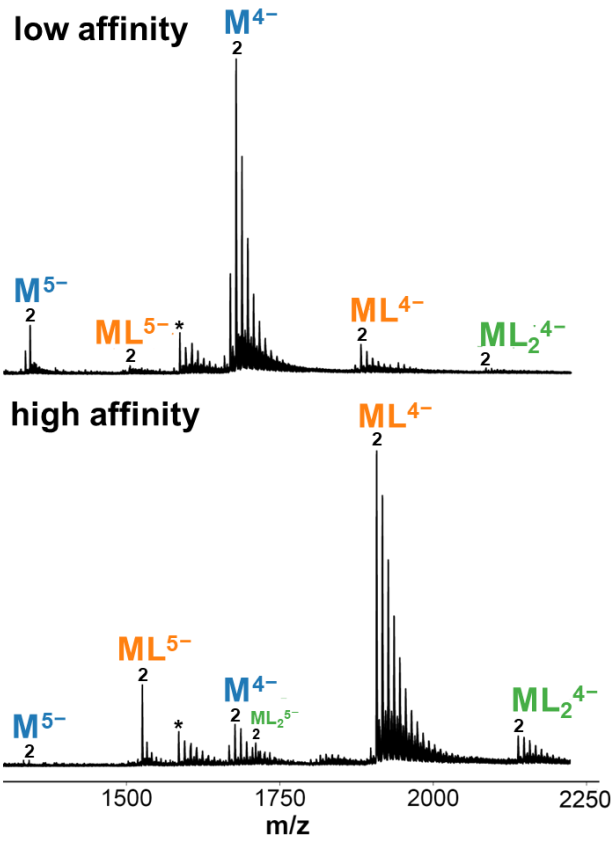


*Figure 63 Mass spectra for 5YEY (d**G₃**TTA**G₃**TTA**G₃**TTT**G₃**) in **Top**: QPPQ. K_{D1} = 75 µM, K_{D2} = 40 µM. **Bottom**: QQPQ. K_{D1} = 1.5 µM, K_{D2} = 22 µM. Sample conditions are: 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8). Labeled species are: Unbound DNA (M), 1:1 complex (ML), 2:1 complex (ML_2), $2K^+$ adduct (2), depurinated DNA (*).*

Visually, little to no complex is visible for QPPQ while the 1:1 complex is the main species with QQPQ. In fact, the K_D values we calculated from these mass spectra (see caption) differ by 1-2 orders of magnitude. When modifications in the oligomer sequence increase ligand affinity by this magnitude, we are optimistic that foldamers can be rationally designed to target specific GQ structures.

The low intensity of 2:1 complex compared to 1:1 complex indicates that the 5YEY GQ only has one accessible binding site. Meanwhile the DNA/ligand pairing from Figure 62 has two accessible binding sites. To properly grasp the foldamer/GQ interaction, we need to screen different foldamers against a selection of GQs.

Full panel screening

Figure 64 presents the ligand screening results in the form of a heatmap. In ‘hot’ zones the majority of DNA is bound by the complex while in ‘cold’ zones the DNA is mostly unbound. The detailed data, including K_D values, are deposited in Table S2, mass spectra are shown in Figure S65 to Figure S100.

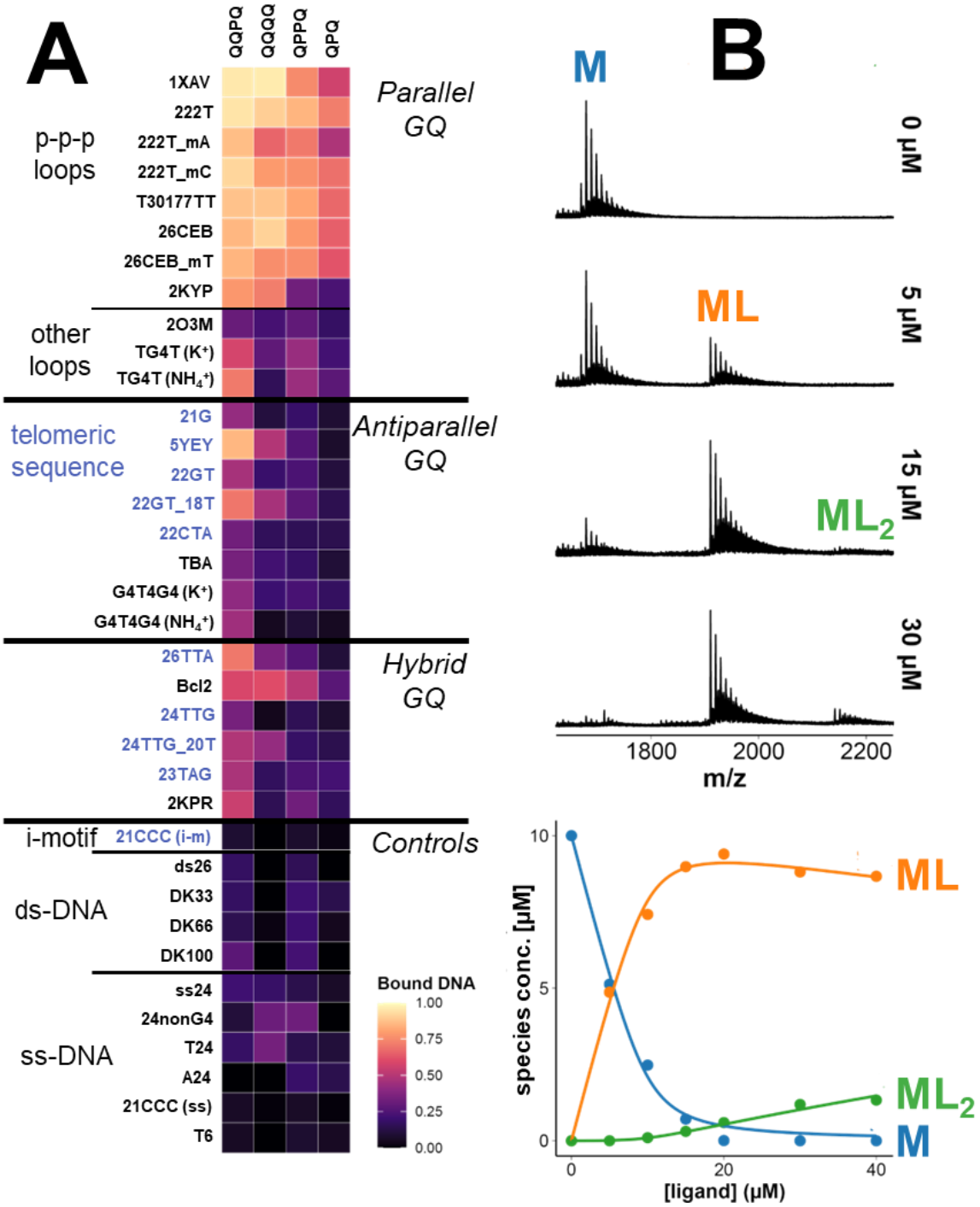


Figure 64 ***A)*** *Heatmap showing the fraction of DNA bound for all DNA/foldamer combinations, calculated from mass spectra. Sample conditions are: 10 µM DNA, 20 µM Ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8). DNA sequences are listed in Table 6. Telomeric-related sequences are highlighted in blue.* ***B)*** *Titration data for 5YEY and QQPQ, showing an extract of the raw mass spectra on top and the processed datapoints on the bottom. The dynamic fitting curves match the experimental datapoints and return* K_{D1} *= 0.54 ± 0.1 µM,* K_{D2} *= 200 ± 70 µM.*

All four foldamers bind parallel GQs with K_D values in the low µM range. Not all parallel GQs are equivalent. 222T_mA and 2KYP are partially unfolded in 0.5 mM K^+, so the GQ has to overcome a higher energetic barrier transitioning from unfolded strand to complex. This is why the fraction of bound DNA is lower compared to the other (stable) sequences. The G-tetrads of 2O3M and TG4T are less exposed than those of common parallel GQs with three propeller loops. 2O3M has a lateral and a snapback loop that

can cover the G-tetrads. TG4T has eight flanking thymines that can form a sparsely populated T4-tetramer, which also protects the GQ from 5'-5'-dimerization.[58] We therefore hypothesize that foldamers bind to the external G-tetrads, but with steric restrictions, so that loops/bases covering the tetrads will impede ligand binding.

We can split the ligands into two groups: QPQ and QPPQ, within our panel, are highly selective to parallel GQs with three propellor loops. We suspect that these ligands are more likely crescent-shaped than helical, because the oligomer chain is not long enough (it takes 5 Q units for 2 helical turns). QQPQ and QQQQ are stable helices; we validated the network of intramolecular H-bonds by NOESY (see NMR section). Their affinity to parallel GQs is 3-4 times higher compared to QPQ and QPPQ. Apart from parallel GQs, the ligand QQPQ remarkably interacts with 5YEY (K_{D1} = 1.5 µM), but also TG4T, 22GT_18T and 26TTA ($K_{D1} \approx 5$ µM). 5YEY and 22GT_18T are telomeric mutants with a TTA→TTT mutation in the third loop, which induces a switch from a 2-tetrad to a 3-tetrad antiparallel GQ.[59] QQPQ and QQQQ are sensitive to this structural switch: The affinities to the mutants is 5-10 times higher than to the 'wild-type' sequences (compare: 21G/5YEY, 22GT/22GT_18T, 24TTG/24TTG_20T). 26TTA is a hybrid-2 GQ that is targeted by QQPQ but no other foldamer in this panel. Nonetheless, we cannot assess whether QQPQ is selective towards hybrid-2 over hybrid-1 topology.

None of the ligands interact well with most antiparallel GQs, where the G-tetrads are usually covered by base triads and/or diagonal loops (21G, 22GT, 22CTA, TBA, G4T4G4). This is further evidence that foldamers selectively target GQs whose configuration allows exposed G-tetrads. The low affinity towards alternative DNA structures (i-motif, duplex, single strand) underlines the GQ specificity of foldamers. The helical shape of foldamers prevents duplex intercalation, as they are too bulky to lodge in between base pairs.[42] Among the controls, the foldamers were irresponsive to changes in G-richness, molecularity or strand length.

ESI-MS titrations

We selected six DNA sequences from the ligand screening panel and titrated them against seven ligands: QPQ, QQPQ, QPPQ, Q_3, Q_4, Q_5 and Q_8. The two main objectives are 1) Get more reproducible K_D values for sequences of interest 2) Study the effect of foldamer length (Q_3 to Q_8) on ligand affinity. Figure 64 features an extract of the titration data for 5YEY and QQPQ. All K_D values are listed in ESI-MS titrations: KD values and response factor estimates
Table S3. Estimated response factors of each DNA:Ligand complex are listed in Table S4. The full titration datasets are in Figure S101 to Figure S141.

The control sequence ss24 is a single-stranded 24mer with 25% of each nucleobase. None of the foldamers target this sequence (K_{D1} > 100 µM) save for Q_3, which has significant populations of 1:1 complex, 2:1 complex and 3:1 complex (Figure S129). QQQ is thus not GQ specific and we do not have an explanation for this observation.

T24 is another single-stranded control sequence that does not interact with QPQ, QQPQ or QPPQ (K_D > 100 µM). However, it forms high-affinity 2:1 complex with Q_3 to Q_5 (K_{D2} < 1 µM), including a 4:1 complex with Q_3 (K_{D4} = 0.01 ± 4 µM). Although our ligands are racemic, there is an induced CD in the presence of T24 that matches the CD signature of a right-handed foldamer helix (Figure S142). Our hypothesis is that T24 and Q_n foldamers associate into a right-handed double helix, wherein the foldamer is dimerized in order to accommodate the full length of the T24 strand. We attempted

to reproduce our results on U24 (Figure S143), because U-rich motifs play a role in RNA expression,[60,61] Alas, there is no induced CD switching from dT to rU.

21G and 5YEY are the same telomeric repeat sequence save for one A→T base switch, which triggers a topology switch from polymorphic (21G) to antiparallel (5YEY). QQPQ, Q_4 and Q_5 are sensitive to this switch and form stable 1:1 complex with 5YEY (K_{D1} < 10 µM). Q_4 is the most sensitive ligand, with a 20x change in K_{D1} between 21G and 5YEY.

1XAV and 222T showed the highest fraction of bound ligand in the foldamer screening (96%). All foldamers form high-affinity 1:1 complex with 222T (K_{D1} < 10 µM). The strongest binding partner for 1XAV is Q_4 (K_{D1} = 75 ± 59 nM), the weakest one is Q_8 (K_{D1} = 28 ± 6.7 µM). All foldamers form 2:1 complex (K_{D2} ≈ 1 µM), except for Q_8 and QPQ. Q_3 is the only ligand to form a 3:1 complex, but based on our results with control sequence ss24, the interaction might not be GQ specific.

Thanks to the titration data, we can place the affinity of foldamers in the low µM range for parallel GQs, as well with a few other specific sequences (5YEY, T24). Foldamer length is important and there seems to be a 'sweet spot' at about 4-5 monomers. Having only 3 units comes with a decline in affinity (QPQ) or GQ specificity (QQQ). Having more than 5 monomers leads to no affinity increase. In fact, the DNA signal loses intensity when ligand is added – and the rate at which it drops is proportional to the number of Q units (Figure S63). The reason is that each Q unit adds a positive charge to the ligand, which help reduce charge repulsion among DNA polyanions, leading to DNA aggregation. We confirm this hypothesis from 1) seeing precipitation in solutions with > 100 µM DNA concentration in excess of ligand and 2) loss of CD signature for a stable (i.e. resistant to disruption) parallel GQ when Q_8 is added (Figure S144).

Investigating the foldamer-GQ interaction

We picked the foldamer QQPQ for structural investigation because 1) It was the foldamer with the highest ligand affinity and 2) It selectively targeted certain telomeric repeat sequences, among which we picked 5YEY and 26TTA. We picked T30177TT as a stable parallel GQ. 222T and 222T-mA are polymorphic in 0.5 mM K^+ with a mix of 3-tetrad parallel and 2-tetrad antiparallel GQ.[10] The equilibrium of 222T in our experimental conditions is shifted towards the parallel form, while 222T-mA is shifted towards the antiparallel form (ref. Figure 65).

Figure 6 shows the interaction of foldamer QQPQ with the five aforementioned DNA sequences by four different methods. 1) Mass spectrometry. Mass spectra provide a quick readout on ligand stoichiometry and the quantity of each species. The number of specific K^+ ions indicates the number of G-tetrads. 2) Ion mobility spectrometry (IMS). This technique lets us track changes in structure even when there is no change in mass. The ions' collisional cross section (CCS) depends on its ion mobility and charge state (within constant experimental settings). Significant conformational changes are expected to lead to significant shifts in ion mobility, resulting in separate CCS distributions. Thus, the shape of the CCS distribution for each ligand stoichiometry indicates how many folded species are present at minimum. 3) Circular dichroism (CD). We follow how the ligand affects GQ topology by comparing CD signatures with and without adding ligand. 4) UV-melting. The difference in melting temperature with and without ligand (ΔT) is a common reference to quantify how potent a ligand is at stabilizing the GQ structure. When there are sufficiently populated

species with significantly different melting temperatures, multiple melting transitions will be visible.

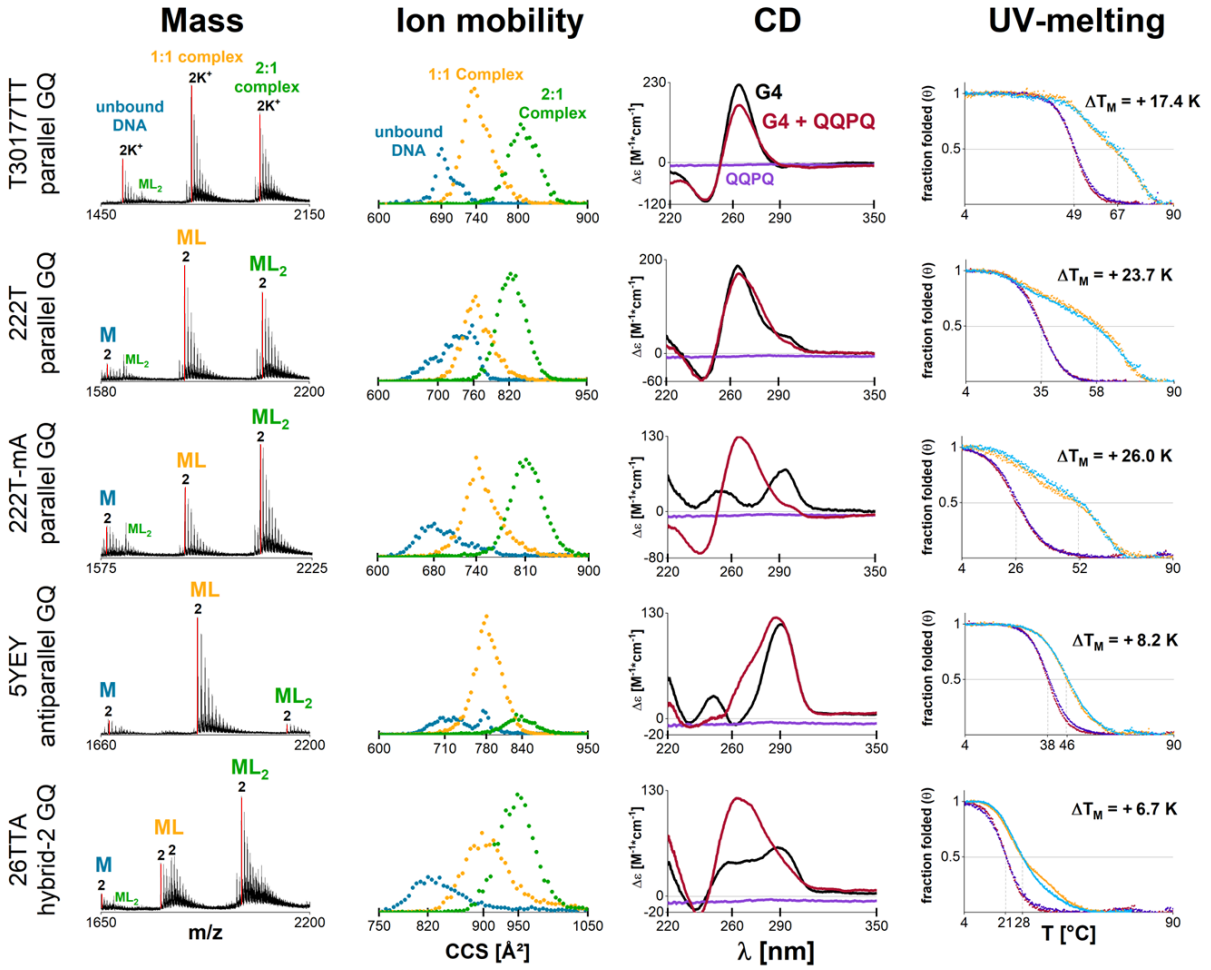


Figure 65 The effect of foldamer QQPQ on five different GQs. Samples contain 10 µM DNA, 20 µM QQPQ (unless specified otherwise), 0.5 mM KCl, 100 mM TMAA (pH 6.8). ***1st column****: Mass spectra showing the 4- charge state (5- for 26TTA).* ***2nd column****: CCS distributions extracted from the 2K⁺ adduct of each species, shown in red in the corresponding mass spectrum.* ***3rd column****: CD spectra of the complex (red) and control spectra with only DNA (black) and only ligand (purple).* ***4th column****: UV-melting curves with ligand (heating: orange; cooling: cyan) and without ligand (heating: crimson, cooling: navy).*

For T30177TT we see: 2 K^+ is the dominant adduct in the mass spectra, a parallel CD signature and a fraction folded of 1 at 25°C. We are thus dealing with a stable, 3-tetrad parallel GQ. Adding ligand does not change the K^+ adduct distribution, the shape of the CCS distribution or the CD signature, meaning that ligand binding does not change the folded GQ structure. The slight drop in CD intensity is likely due to ligand-induced DNA aggregation (Figure S144). The melting temperature increases by 17 K, but two distinct melting transitions appear. We know from previous temperature-controlled MS studies on GQ ligands that the 1:1 complex usually melts before the 2:1 complex.[62] It could be the same here.

222T and 222T-mA are polymorphic in 0.5 mM K^+; the CCS distributions of the unbound DNA are therefore the sum of multiple Gaussian functions, each representing a different set of conformers. Upon ligand binding, the CCS distributions are less convoluted (i.e. they look more like a single Gaussian distribution), which means that the ligand selectively targets certain conformers within the ensemble of our polymorphic sequences. We know that the complex is a 3-tetrad parallel GQ, because adding ligand shifts the K^+ adduct distribution towards the $2K^+$ species (also shown in

Figure 2) and shifts the CD spectra towards a parallel signature (222T-mA has a remarkable signature switch). The ligand targets parallel GQ and stabilizes it: We report ΔT_M values of ca. +25 K. Similar to T30177TT, the DNA/ligand system has two separated, broadened melting transitions which likely correspond to the 1:1 and 2:1 complex.

5YEY only has one accessible binding site, judging by the ratio of 1:1 to 2:1 complex in the mass spectrum (see also the titration in Figure S118). The ligand affects GQ folding, causing a CD band at 260 nm to increase. The CD signature looks like the 1:1 complex could be a hybrid GQ, but the CCS distributions give no conclusive evidence whether there is only one hybrid species or multiple species with different topologies. The affinity of the 1:1 complex between 5YEY and QQPQ (K_{D1} = 0.54 ± 0.1 µM) is comparable to parallel GQs. Nonetheless, the melting curve is steeper and shifts less (ΔT_M = +8.2 K) in comparison. That implies that the complex formation with 5YEY is more enthalpy-driven than with parallel GQs.[45]

26TTA mirrors the behavior of 222T/222T-mA: The K^+ adduct distribution shifts towards $2K^+$, a parallel CD signature emerges and the UV melting transition broadens in the presence of QQPQ. The ligand promotes a structural rearrangement of 26TTA, probably from hybrid to parallel topology.

X-ray crystallography

We then used X-ray crystallography to test the end-stacking hypothesis built from the CD and native MS results. Molecular crowding induces a bias for the parallel topology,[63,64] so we picked the parallel GQ sequences 1XAV and 222T in order to 1) Maintain the solution-phase topology and 2) Facilitate crystallization. We attempted to crystallize both sequences with QQPQ or QQQQ, as well was 5YEY/QQPQ and T24/QQQQ. We only obtained crystals for the 222T/QQPQ complex, whose crystal structure is depicted in Figure 66. Crystal and structure refinement information is deposited in Table S5. Electron density maps are shown in Figure S145 to Figure S147.

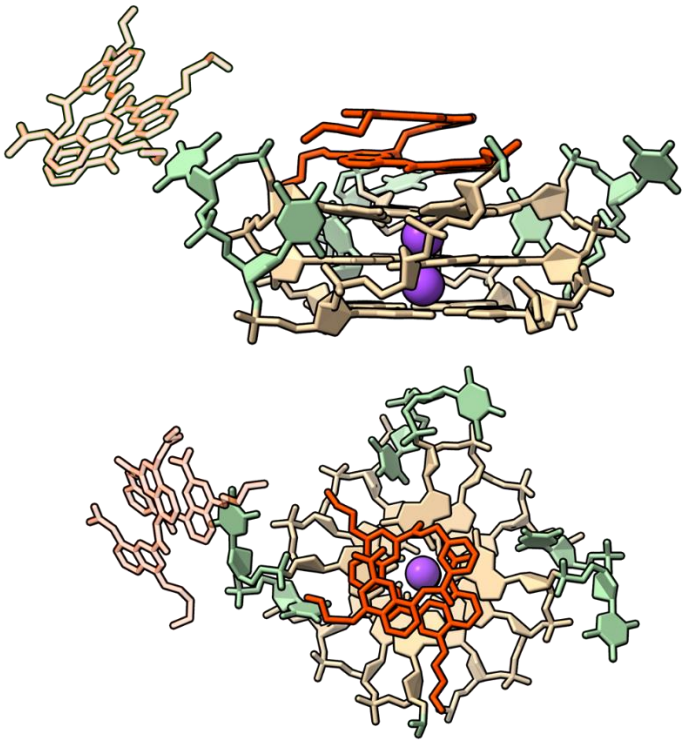


*Figure 66 X-ray crystal structure of GQ 222T (dT**G₃**TT**G₃**TT**G₃**TT**G₃**T) and foldamer QQPQ from the side (**top**) and from above (**bottom**). Color coding: QQPQ in red, guanine in tan, thymine in green, potassium in purple. The crystallization matrix contains 100 mM KCl, 15 mM $MgCl_2$, 50 mM Tris (pH 7.5), 10% PEG-550 and a 9:4 ratio of ligand:DNA.*

The structure validates our previous deductions: Two quinolines are π-stacking onto the G-tetrad at the 3' end. The helix appears to flatten out at the binding side to better accommodate itself. We suspect that the P unit reduces tension caused by the flattening process by enhancing structural flexibility, which comes from introducing an extra methylene group into the helix chain. That is why QQPQ is the ligand with the highest binding affinity: It has two Qs for π-stacking, followed by a P at which the molecule can bend back into its helical shape. The top view illustrates how the diameter of the foldamer helix matches with the size of a G-tetrad. The propellor loops are not limiting the ligand's access to the G-tetrads and therefore do not interfere with ligand binding.

According to our ESI-MS titrations, the complex is saturated in a 3-fold excess of ligand, where it forms a 2:1 complex (Figure S111). We expected to have one ligand on each end of the GQ, however, one ligand is seen stacking onto a thymine base (Figure 66). To understand this observation, we have to take a look beyond the single

unit cell and take crystal packing artifacts into account. The 5' G-tetrad is occupied by another GQ (Figure S148), blocking off a potential ligand binding site. Parallel GQs in particular dimerize in concentrated solution because of their exposed G-tetrads. The preferred interface for dimerization is 5' to 5' end.[65,66] The foldamer sitting atop the thymine establishes a contact with a neighboring unit cell. Such crystal contacts are needed to promote crystal formation and growth, but they can only persist in solid phase.

The electron density map of the GQ is well-resolved (Figure S145), but the electron density of the foldamer is spread more thinly within the binding site (Figure S147), implying that the ligand has a dynamic structure within the binding site. Several factors can contribute to these dynamics: 1) Rotational freedom along the helical axis 2) Exchanges between left- and right-handed helix, since the foldamer is racemic 3) The foldamer latching onto the G-tetrad from its C-terminus or N-terminus 4) Conformational freedom of the foldamer side chains. We tested models where either the C- or N-terminus faces the G-tetrad (Figure S149) and found no significant change in refinement statistics that would suggest any of those four arrangements is more favorable than the others. So even though the crystal structure depicts a static image of the ligand, there are a series of dynamic exchanges in solution that we cannot see in solid phase.

In summary, we have validated the first ligand binding site and the binding mode, but need supporting evidence for the second binding site and the ligand orientation.

NMR spectroscopy for crystal structure refinement

NMR enables us to better interpret the crystal structure by identifying deviations between solid-phase and solution-phase structure. We are unable to produce well-resolved NMR spectra with 222T, probably due to its structural polymorphism. Hence, we substituted 222T (dT**G$_3$**TT**G$_3$**TT**G$_3$**TT**G$_3$**T) with 2LK7 (dTT**G$_3$**T**G$_3$**T**G$_3$**T**G$_3$**T), whose structure is very similar to 222T but more dynamically constrained (1 nt loops) and protected from dimerization (TT motif at 5').[56] We validated that QQPQ targets 2LK7 with sub-µM affinity (Figure S150, ESI-MS titrations: KD values and response factor estimates
Table S3). All 1D and 2D NMR spectra of 2LK7/QQPQ are deposited in the supporting information (Figure S151 to Figure S165).
Our key findings are summarized in Figure 67. Panel A shows a ^{1}H-^{1}H NOESY spectrum of the 2LK7/QQPQ complex. We observe two networks of NOE contacts that correspond to the amide protons of the foldamer (designated *NH1* to *NH4* in panel D). The amide protons are in close contact with each other through the H-bond network that enables the helical foldamer arrangement (Figure 61). The detection of (exchangeable) amide protons in the 1D spectrum (panel C) further indicates that these protons are protected from solvent exchange. These factors are supporting evidence for the helical folding of QQPQ in solution phase.

Panel A shows there are two distinct NOE networks of the same amide protons (lime green and blue). Consequently, there are two significantly different chemical environments for the foldamer, which must be the first and second binding site, since we detected 2:1 complex in the mass spectra. One binding site is the G-tetrad on the 3' end, as seen in the crystal structure. A second ligand is shown stacking onto a thymine in the loop region, but the NMR data show no chemical shift in the loop region upon adding ligand to the sample, proving that this interaction does not take place in

solution. A more suitable candidate for the second binding site is the G-tetrad on the 5’ end, which is inaccessible in solid phase due to GQ dimerization.

We then labeled the 5’ and 3’ end of 2LK7 by switching the flanking dT nucleotides with dU (which lacks the characteristic methyl group of thymine). These chemical changes, albeit small, cause visible chemical shifts in the proximity of each respective thymine base. Figure 67B shows how we assigned a binding site to each NOE network. We are looking at the NOE between the amide group (*NH1*) and the methyl group (*HAc*) at the N-terminus of the foldamer. Labeling the 3’ end (U18) causes a chemical shift in the blue NOE network, while the green NOE network remains unaffected. Likewise, labeling the 5’ end (U2) causes a chemical shift in the green NOE network, while the blue NOE network remains unaffected. Therefore, one foldamer must be close to the 5’ end and another one close to the 3’ end. While the 3’ foldamer is shown in the crystal structure, we can now safely deduce that in solution the second foldamer is π-stacking onto the 5’ G-tetrad.

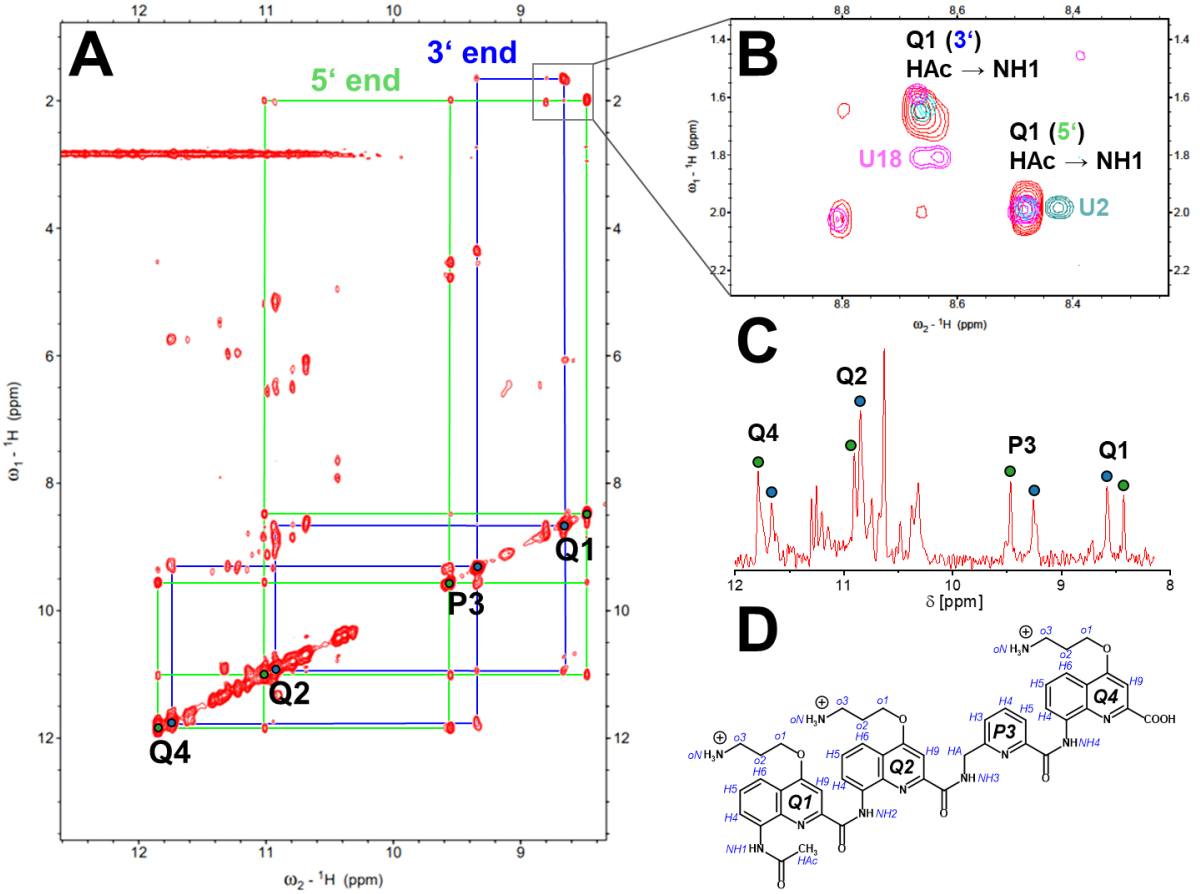


*Figure 67 A) ^{1}H-NOESY of 2LK7 (dTT**G$_3$**T**G$_3$**T**G$_3$**T) in excess of QQPQ, showing two NOE networks connecting the foldamer amide protons (NH), indicating that the foldamer is located in two distinct chemical environments. B) Marking the flanking nucleotides through dT→dU mutations induces chemical shifts in the foldamer on the 5’ end (pos. 1,2) and 3’ end (pos. 18) of the GQ. C) 1D-NMR of 2LK7 in 3-fold excess of QQPQ, showing the guanine H1-proton region. Foldamer amide protons are highlighted. D) Scheme of QQPQ with proton labels shown in blue. Samples contain 100 µM 2LK7, 300 µM QQPQ, 10 mM KP buffer (pH 7.0) in 90/10 H$_2$O/D$_2$O*

Conclusion

We performed a comprehensive study on the interaction between G-quadruplexes and small oligoamide quinoline helices, referred to as foldamers. Foldamers do not target double-stranded DNA, single strands or i-motifs. They selectively target parallel G-quadruplexes, forming 1:1 and 2:1 complex with sub-µM K_D values. We were able to crystallize a foldamer-DNA complex and resolve the crystal structure. Two quinoline units in the foldamer helix interact with an exposed G-tetrad by π-stacking. The 'bulkiness' of the foldamer helix causes its topology selectivity. The helix clashes with nucleobases surrounding the G-tetrad, making it sensitive to the presence of loops and flanking nucleotides. We believe that through rational design, the structure of the foldamer helix can be adjusted to adapt itself to the G-tetrad surrounding nucleobases (except when it is a base triad or a diagonal loop). We also did not explore the effect of the foldamer sidechain, which has great potential to be tinkered with. Overall, we think foldamers are a highly promising ligand scaffold that could be molded to selectively target a G-quadruplex target of interest.

As a perspective, we hope to more closely investigate the interaction between 5YEY and QQPQ, forming a 1:1 complex with sub-µM K_D value. 5YEY is the only antiparallel G-quadruplex targeted by a foldamer – vice versa QQPQ is the only foldamer that targets 5YEY. QQPQ changes the CD signature of 5YEY and 5YEY induces CD in QQPQ. We believe the QQPQ/5YEY pair might be one of the special cases we predicted above, where the foldamer helix and the nucleobases surrounding the G-tetrad match each other's shapes, leading to a strong and selective interaction. We failed to crystallize the complex, so we will study this system further by NMR (work in progress).

References

(1) Parkinson, G. N. The Reality of Quadruplex Nucleic Acids as a Therapeutic Target. In *Guanine Quartets*; 2012; pp 237–247. https://doi.org/10.1039/9781849736954-00237.

(2) Chaudhuri, R.; Bhattacharya, S.; Dash, J.; Bhattacharya, S. Recent Update on Targeting C-MYC G-Quadruplexes by Small Molecules for Anticancer Therapeutics. *J. Med. Chem.* **2021**, *64* (1), 42–70. https://doi.org/10.1021/acs.jmedchem.0c01145.

(3) Patel, D. J.; Phan, A. T.; Kuryavyi, V. Human Telomere, Oncogenic Promoter and 5′-UTR G-Quadruplexes: Diverse Higher Order DNA and RNA Targets for Cancer Therapeutics. *Nucleic Acids Research* **2007**, *35* (22), 7429–7455. https://doi.org/10.1093/nar/gkm711.

(4) Kosiol, N.; Juranek, S.; Brossart, P.; Heine, A.; Paeschke, K. G-Quadruplexes: A Promising Target for Cancer Therapy. *Molecular Cancer* **2021**, *20* (1), 40. https://doi.org/10.1186/s12943-021-01328-4.

(5) Dai, J.; Carver, M.; Yang, D. Polymorphism of Human Telomeric Quadruplex Structures. *Biochimie* **2008**, *90* (8), 1172–1183. https://doi.org/10.1016/j.biochi.2008.02.026.

(6) Phan, A. T.; Mergny, J.-L. Human Telomeric DNA: G-Quadruplex, i-Motif and Watson-Crick Double Helix. *Nucleic Acids Res* **2002**, *30* (21), 4618–4625. https://doi.org/10.1093/nar/gkf597.

(7) Phan, A. T. Human Telomeric G-Quadruplex: Structures of DNA and RNA Sequences: Human Telomeric G-Quadruplex Structures. *FEBS Journal* **2010**, *277* (5), 1107–1117. https://doi.org/10.1111/j.1742-4658.2009.07464.x.

(8) Gao, J.; Pickett, H. A. Targeting Telomeres: Advances in Telomere Maintenance Mechanism-Specific Cancer Therapies. *Nat Rev Cancer* **2022**, *22* (9), 515–532. https://doi.org/10.1038/s41568-022-00490-1.

(9) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Research* **2005**, *33* (9), 2908–2916. https://doi.org/10.1093/nar/gki609.

(10) Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *Journal of the American Chemical Society*, 2016, *138*, 2780–2792. https://doi.org/10.1021/jacs.5b13130.

(11) Marchand, A.; Gabelica, V. Folding and Misfolding Pathways of G-Quadruplex DNA. *Nucleic Acids Research*, 2016, *44*, 10999–11012. https://doi.org/10.1093/nar/gkw970 %J Nucleic Acids Research.

(12) Frelih, T.; Wang, B.; Plavec, J.; Šket, P. Pre-Folded Structures Govern Folding Pathways of Human Telomeric G-Quadruplexes. *Nucleic Acids Research* **2020**, *48* (4), 2189–2197. https://doi.org/10.1093/nar/gkz1235.

(13) Gabelica, V. A Pilgrim's Guide to G-Quadruplex Nucleic Acid Folding. *Biochimie*, 2014, *105*, 1–3. https://doi.org/10.1016/j.biochi.2014.08.007.

(14) Šponer, J.; Bussi, G.; Stadlbauer, P.; Kührová, P.; Banáš, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of Guanine Quadruplex Molecules–Funnel-like Mechanism or Kinetic Partitioning? An Overview from MD Simulation Studies.

Biochimica et Biophysica Acta (BBA) - General Subjects **2017**, *1861* (5, Part B), 1246–1263. https://doi.org/10.1016/j.bbagen.2016.12.008.

(15) Largy, E.; Mergny, J.-L. Shape Matters: Size-Exclusion HPLC for the Study of Nucleic Acid Structural Polymorphism. *Nucleic Acids Research* **2014**, *42* (19), e149. https://doi.org/10.1093/nar/gku751.

(16) Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C. Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **1992**, *31* (3), 833–841. https://doi.org/10.1021/bi00118a028.

(17) Neidle, S.; Balasubramanian, S. The Role of Cations in Determining Quadruplex Structure and Stability. In *Quadruplex Nucleic Acids*; 2006; pp 100–130. https://doi.org/10.1039/9781847555298-00100.

(18) Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *The Alkali Metal Ions: Their Role for Life*, 2016, 203–258. https://doi.org/10.1007/978-3-319-21756-7_7.

(19) Khristenko, N.; Amato, J.; Livet, S.; Pagano, B.; Randazzo, A.; Gabelica, V. Native Ion Mobility Mass Spectrometry: When Gas-Phase Ion Structures Depend on the Electrospray Charging Process. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (6), 1069–1081. https://doi.org/10.1007/s13361-019-02152-3.

(20) Griffeycor, R. H.; Sasmor, H.; Greig, M. J. Oligonucleotide Charge States in Negative Ionization Electrospray-Mass Spectrometry Are a Function of Solution Ammonium Ion Concentration. *J Am Soc Mass Spectrom* **1997**, *8* (2), 155–160. https://doi.org/10.1016/S1044-0305(96)00200-0.

(21) Ferreira, R.; Marchand, A.; Gabelica, V. Mass Spectrometry and Ion Mobility Spectrometry of G-Quadruplexes. A Study of Solvent Effects on Dimer Formation and Structural Transitions in the Telomeric DNA Sequence d(TAGGGTTAGGGT). *Methods* **2012**, *57* (1), 56–63. https://doi.org/10.1016/j.ymeth.2012.03.021.

(22) Marchand, A.; Ferreira, R.; Tateishi-Karimata, H.; Miyoshi, D.; Sugimoto, N.; Gabelica, V. Sequence and Solvent Effects on Telomeric DNA Bimolecular G-Quadruplex Folding Kinetics. *J. Phys. Chem. B* **2013**, *117* (41), 12391–12401. https://doi.org/10.1021/jp406857s.

(23) Ghosh, A.; Largy, E.; Gabelica, V. DNA G-Quadruplexes for Native Mass Spectrometry in Potassium: A Database of Validated Structures in Electrospray-Compatible Conditions. *Nucleic Acids Research* **2021**, *49* (4), 2333–2345. https://doi.org/10.1093/nar/gkab039.

(24) Monchaud, D.; Allain, C.; Bertrand, H.; Smargiasso, N.; Rosu, F.; Gabelica, V.; De Cian, A.; Mergny, J.-L.; Teulade-Fichou, M.-P. Ligands Playing Musical Chairs with G-Quadruplex DNA: A Rapid and Simple Displacement Assay for Identifying Selective G-Quadruplex Binders. *Biochimie* **2008**, *90* (8), 1207–1223. https://doi.org/10.1016/j.biochi.2008.02.019.

(25) Micheli, E.; Lombardo, C. M.; D'Ambrosio, D.; Franceschin, M.; Neidle, S.; Savino, M. Selective G-Quadruplex Ligands: The Significant Role of Side Chain Charge Density in a Series of Perylene Derivatives. *Bioorganic & Medicinal Chemistry Letters* **2009**, *19* (14), 3903–3908. https://doi.org/10.1016/j.bmcl.2009.03.106.

(26) Ma, Y.; Iida, K.; Nagasawa, K. Topologies of G-Quadruplex: Biological Functions and Regulation by Ligands. *Biochemical and Biophysical Research Communications* **2020**, *531* (1), 3–17. https://doi.org/10.1016/j.bbrc.2019.12.103.

(27) Müller, D.; Saha, P.; Panda, D.; Dash, J.; Schwalbe, H. Insights from Binding on Quadruplex Selective Carbazole Ligands. *Chemistry – A European Journal* **2021**, *27* (50), 12726–12736. https://doi.org/10.1002/chem.202101866.

(28) Dickerhoff, J.; Dai, J.; Yang, D. Structural Recognition of the MYC Promoter G-Quadruplex by a Quinoline Derivative: Insights into Molecular Targeting of Parallel G-Quadruplexes. *Nucleic Acids Research* **2021**, *49* (10), 5905–5915. https://doi.org/10.1093/nar/gkab330.

(29) Ritson, D. J.; Moses, J. E. A Fragment Based Click Chemistry Approach towards Hybrid G-Quadruplex Ligands: Design, Synthesis and Biophysical Evaluation. *Tetrahedron* **2012**, *68* (1), 197–203. https://doi.org/10.1016/j.tet.2011.10.066.

(30) Reznichenko, O.; Cucchiarini, A.; Gabelica, V.; Granzhan, A. Quadruplex DNA-Guided Ligand Selection from Dynamic Combinatorial Libraries of Acylhydrazones. *Org. Biomol. Chem.* **2021**, *19* (2), 379–386. https://doi.org/10.1039/D0OB01908A.

(31) Hamon, F.; Largy, E.; Guédin-Beaurepaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J.-L.; Riou, J.-F.; Nguyen, C.-H.; Teulade-Fichou, M.-P. An Acyclic Oligoheteroaryle That Discriminates Strongly between Diverse G-Quadruplex Topologies. *Angewandte Chemie International Edition* **2011**, *50* (37), 8745–8749. https://doi.org/10.1002/anie.201103422.

(32) Lin, C.; Wu, G.; Wang, K.; Onel, B.; Sakai, S.; Shao, Y.; Yang, D. Molecular Recognition of the Hybrid-2 Human Telomeric G-Quadruplex by Epiberberine: Insights into Conversion of Telomeric G-Quadruplex Structures. *Angewandte Chemie International Edition* **2018**, *57* (34), 10888–10893. https://doi.org/10.1002/anie.201804667.

(33) McQuaid, K. T.; Takahashi, S.; Baumgaertner, L.; Cardin, D. J.; Paterson, N. G.; Hall, J. P.; Sugimoto, N.; Cardin, C. J. Ruthenium Polypyridyl Complex Bound to a Unimolecular Chair-Form G-Quadruplex. *J. Am. Chem. Soc.* **2022**, *144* (13), 5956–5964. https://doi.org/10.1021/jacs.2c00178.

(34) Largy, E.; Teulade-Fichou, M.-P. Screening for Quadruplex Binding Ligands: A Game of Chance? **2012**. https://doi.org/10.1039/9781849736954-00248.

(35) Guichard, G.; Huc, I. Synthetic Foldamers. *Chem. Commun.* **2011**, *47* (21), 5933–5941. https://doi.org/10.1039/C1CC11137J.

(36) Jena, P. V.; Shirude, P. S.; Okumus, B.; Laxmi-Reddy, K.; Godde, F.; Huc, I.; Balasubramanian, S.; Ha, T. G-Quadruplex DNA Bound by a Synthetic Ligand Is Highly Dynamic. *J. Am. Chem. Soc.* **2009**, *131* (35), 12522–12523. https://doi.org/10.1021/ja903408r.

(37) Müller, S.; Laxmi-Reddy, K.; Jena, P. V.; Baptiste, B.; Dong, Z.; Godde, F.; Ha, T.; Rodriguez, R.; Balasubramanian, S.; Huc, I. Targeting DNA G-Quadruplexes with Helical Small Molecules, 2014, *15*, 2563–2570. https://doi.org/10.1002/cbic.201402439.

(38) Mandal, P. K.; Baptiste, B.; Langlois d'Estaintot, B.; Kauffmann, B.; Huc, I. Multivalent Interactions between an Aromatic Helical Foldamer and a DNA G-

Quadruplex in the Solid State. *ChemBioChem* **2016**, *17* (20), 1911–1914. https://doi.org/10.1002/cbic.201600281.

(39) Qi, T.; Maurizot, V.; Noguchi, H.; Charoenraks, T.; Kauffmann, B.; Takafuji, M.; Ihara, H.; Huc, I. Solvent Dependence of Helix Stability in Aromatic Oligoamide Foldamers. *Chem. Commun.* **2012**, *48* (51), 6337–6339. https://doi.org/10.1039/C2CC31533E.

(40) Shirude, P. S.; Gillies, E. R.; Ladame, S.; Godde, F.; Shin-ya, K.; Huc, I.; Balasubramanian, S. Macrocyclic and Helical Oligoamides as a New Class of G-Quadruplex Ligands. *J. Am. Chem. Soc.* **2007**, *129* (39), 11890–11891. https://doi.org/10.1021/ja073775h.

(41) Delaurière, L.; Dong, Z.; Laxmi-Reddy, K.; Godde, F.; Toulmé, J.-J.; Huc, I. Deciphering Aromatic Oligoamide Foldamer–DNA Interactions. *Angewandte Chemie International Edition* **2012**, *51* (2), 473–477. https://doi.org/10.1002/anie.201106208.

(42) Summers, P. A.; Thomas, A. P.; Kench, T.; Vannier, J.-B.; Kuimova, M. K.; Vilar, R. Cationic Helicenes as Selective G4 DNA Binders and Optical Probes for Cellular Imaging. *Chem. Sci.* **2021**, *12* (43), 14624–14634. https://doi.org/10.1039/D1SC04567A.

(43) Tataurov, A. V.; You, Y.; Owczarzy, R. Predicting Ultraviolet Spectrum of Single Stranded and Double Stranded Deoxyribonucleic Acids. *Biophysical Chemistry* **2008**, *133* (1), 66–70. https://doi.org/10.1016/j.bpc.2007.12.004.

(44) Mergny, L., Jean-Louis; Lacroix. Analysis of Thermal Melting Curves. *Oligonucleotides*, 2003, 515–537.

(45) Mergny, J. L.; Lacroix, L. UV Melting of G-Quadruplexes. *Curr Protoc Nucleic Acid Chem*, 2009, *Chapter 17*, Unit 17.1. https://doi.org/10.1002/0471142700.nc1701s37.

(46) Gabelica, V.; Livet, S.; Rosu, F. Optimizing Native Ion Mobility Q-TOF in Helium and Nitrogen for Very Fragile Noncovalent Structures. *Journal of the American Society for Mass Spectrometry*, 2018, *29*, 2189–2198. https://doi.org/10.1021/jasms.8b05701.

(47) Marchand, A.; Livet, S.; Rosu, F.; Gabelica, V. Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* **2017**, *89* (23), 12674–12681. https://doi.org/10.1021/acs.analchem.7b01736.

(48) Largy, E.; König, A.; Ghosh, A.; Ghosh, D.; Benabou, S.; Rosu, F.; Gabelica, V. Mass Spectrometry of Nucleic Acid Noncovalent Complexes. *Chem. Rev.* **2021**. https://doi.org/10.1021/acs.chemrev.1c00386.

(49) Rosu, F.; De Pauw, E.; Gabelica, V. Electrospray Mass Spectrometry to Study Drug-Nucleic Acids Interactions. *Biochimie* **2008**, *90* (7), 1074–1087. https://doi.org/10.1016/j.biochi.2008.01.005.

(50) Gabelica, V.; Rosu, F.; De Pauw, E. A Simple Method to Determine Electrospray Response Factors of Noncovalent Complexes. *Analytical Chemistry*, 2009, *81*, 6708–6715. https://doi.org/10.1021/ac900785m.

(51) Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *Journal of Applied Crystallography* **1993**, *26* (6), 795–800. https://doi.org/10.1107/S0021889893005588.

(52) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Cryst D* **2019**, *75* (10), 861–877. https://doi.org/10.1107/S2059798319011471.

(53) Lin, L. Y.; McCarthy, S.; Powell, B. M.; Manurung, Y.; Xiang, I. M.; Dean, W. L.; Chaires, B.; Yatsunyk, L. A. Biophysical and X-Ray Structural Studies of the (GGGTT)3GGG G-Quadruplex in Complex with N-Methyl Mesoporphyrin IX. *PLOS ONE* **2020**, *15* (11), e0241513. https://doi.org/10.1371/journal.pone.0241513.

(54) Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. Grade2 Version 1.4.1, Global Phasing Ltd.

(55) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. *Protein Science* **2021**, *30* (1), 70–82. https://doi.org/10.1002/pro.3943.

(56) Do, N. Q.; Phan, A. T. Monomer–Dimer Equilibrium for the 5′–5′ Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study. *Chemistry – A European Journal* **2012**, *18* (46), 14752–14759. https://doi.org/10.1002/chem.201103295.

(57) Daubenfeld, T.; Bouin, A.-P.; van der Rest, G. A Deconvolution Method for the Separation of Specific Versus Nonspecific Interactions in Noncovalent Protein-Ligand Complexes Analyzed by ESI-FT-ICR Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2006**, *17* (9), 1239–1248. https://doi.org/10.1016/j.jasms.2006.05.005.

(58) Šket, P.; Plavec, J. Tetramolecular DNA Quadruplexes in Solution: Insights into Structural Diversity and Cation Movement. *J. Am. Chem. Soc.* **2010**, *132* (36), 12724–12732. https://doi.org/10.1021/ja104889t.

(59) Lenarčič Živković, M.; Rozman, J.; Plavec, J. Adenine-Driven Structural Switch from a Two- to Three-Quartet DNA G-Quadruplex. *Angewandte Chemie International Edition* **2018**, *57* (47), 15395–15399. https://doi.org/10.1002/anie.201809328.

(60) Akashi, M.; Shaw, G.; Hachiya, M.; Elstner, E.; Suzuki, G.; Koeffler, P. Number and Location of AUUUA Motifs: Role in Regulating Transiently Expressed RNAs. *Blood* **1994**, *83* (11), 3182–3187. https://doi.org/10.1182/blood.V83.11.3182.3182.

(61) Munoz-Tello, P.; Rajappa, L.; Coquille, S.; Thore, S. Polyuridylation in Eukaryotes: A 3′-End Modification Regulating RNA Life. *Biomed Res Int* **2015**, *2015*, 968127. https://doi.org/10.1155/2015/968127.

(62) Marchand, A.; Rosu, F.; Zenobi, R.; Gabelica, V. 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.* **2018**, *140* (39), 12553–12565. https://doi.org/10.1021/jacs.8b07302.

(63) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. Not so Crystal Clear: The Structure of the Human Telomere G-Quadruplex in Solution Differs from That

Present in a Crystal. *Nucleic Acids Research* **2005**, *33* (14), 4649–4659. https://doi.org/10.1093/nar/gki782.

(64) Zhang, D.-H.; Fujimoto, T.; Saxena, S.; Yu, H.-Q.; Miyoshi, D.; Sugimoto, N. Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors. *Biochemistry* **2010**, *49* (21), 4554–4563. https://doi.org/10.1021/bi1002822.

(65) Smargiasso, N.; Rosu, F.; Hsia, W.; Colson, P.; Baker, E. S.; Bowers, M. T.; De Pauw, E.; Gabelica, V. G-Quadruplex DNA Assemblies: Loop Length, Cation Identity, and Multimer Formation. *J. Am. Chem. Soc.* **2008**, *130* (31), 10208–10216. https://doi.org/10.1021/ja801535e.

(66) Kogut, M.; Kleist, C.; Czub, J. Why Do G-Quadruplexes Dimerize through the 5'-Ends? Driving Forces for G4 DNA Dimerization Examined in Atomic Detail. *PLOS Computational Biology* **2019**, *15* (9), e1007383. https://doi.org/10.1371/journal.pcbi.1007383.

(67) Ma, Y.; Tsushima, Y.; Sakuma, M.; Sasaki, S.; Iida, K.; Okabe, S.; Seimiya, H.; Hirokawa, T.; Nagasawa, K. Development of G-Quadruplex Ligands for Selective Induction of a Parallel-Type Topology. *Org. Biomol. Chem.* **2018**, *16* (40), 7375–7382. https://doi.org/10.1039/C8OB01702F.

(68) Ai, T.; Yang, Q.; Lv, Y.; Huang, Y.; Li, Y.; Geng, J.; Xiao, D.; Zhou, C. Insight into How Telomeric G-Quadruplexes Enhance the Peroxidase Activity of Cellular Hemin. *Chemistry – An Asian Journal* **2018**, *13* (14), 1805–1810. https://doi.org/10.1002/asia.201800464.

(69) Wang, Z.-F.; Chang, T.-C. Molecular Engineering of G-Quadruplex Ligands Based on Solvent Effect of Polyethylene Glycol. *Nucleic Acids Research* **2012**, *40* (17), 8711–8720. https://doi.org/10.1093/nar/gks578.

(70) Liu, W.; Lin, C.; Wu, G.; Dai, J.; Chang, T.-C.; Yang, D. Structures of 1:1 and 2:1 Complexes of BMVC and MYC Promoter G-Quadruplex Reveal a Mechanism of Ligand Conformation Adjustment for G4-Recognition. *Nucleic Acids Research* **2019**, *47* (22), 11931–11942. https://doi.org/10.1093/nar/gkz1015.

CONCLUSION

The essence of science lies in the joy of discovery. Asking questions and navigating through a sea of uncertainties, determined to find answers. Our journey started with questioning how the topology of a G-quadruplex is determined by the cation that stabilizes it. The best way to study a rule is to find and analyze the exceptions. We modified DNA sequences with the intention to obtain a 3-tetrad antiparallel G-quadruplex in K^+. Our initial attempts yielded 4-tetrad antiparallel and 3-tetrad hybrid G-quadruplexes. We eventually found what could be the exception: A potential 3-tetrad antiparallel G-quadruplex, produced as the result of kinetic trapping by G-quadruplex formation at low temperature. We were unable to validate the number of G-tetrads, but learned how to be more rational in designing sequences with exceptional conformations. Analyzing these structures would elucidate why 3-tetrad antiparallel G-quadruplex are rarely formed in K^+.

We explored how trivalent lanthanide cations determine G-quadruplex topology and discovered there is neither a G-quadruplex, nor an unfolded strand, but a unique so far unknown secondary structure. The structure can be formed by mixing telomeric repeat DNA with 5-10 equivalents of Tb^{3+}, which coordinates 6-12 guanines. Our attempts to resolve the secondary structure were unsuccessful and warrant further investigation.

We carried out a comprehensive analysis on the topology-selectivity of helical foldamer ligands. We discovered that foldamers selectively target parallel G-quadruplexes with sub-µM K_D values and bind via π-stacking. The selectivity is achieved by the steric exclusion of structures covering the G-tetrads, including base triads, base pairs, lateral loops, diagonal loops or flanking nucleotides. We believe that foldamers can be rationally designed as sequence-specific ligands by combining the steric exclusion of interfering structures with the inclusion of topology-dependent structures (e.g. grooves and loops) by modeling the shape of the foldamer helix and utilizing the sidechains.

Beyond our scientific discoveries, each project shows a unique way in which analytical chemistry pushes the boundaries in G-quadruplex research. The lanthanide project showed that exceptional results are overlooked when data is interpreted in a convenient manner. Assuming that a cation and a G-rich sequence form a G-quadruplex is easy; analytically validating G-quadruplex formation was not that easy, it turns out. The G-extension project taught us that trial-and-error approaches, which are common in G-quadruplex research, may often bring unsatisfactory results. We need analytical chemistry to lay a theoretical foundation that lets us evolve from trial-and-error approaches to more rational methodologies. The foldamer project showed how great opportunities arise from minor inconveniences. Different labs using the same instrument brought together molecules that we would have thought of combining. We did not brush off foldamers as just a contaminant and it led us to discover a never-seen-before scaffold for topology-selective G-quadruplex ligands.

APPENDIX

Chapter 1

Full CD titration datasets of 24TTG, 2LBY and 2LK7

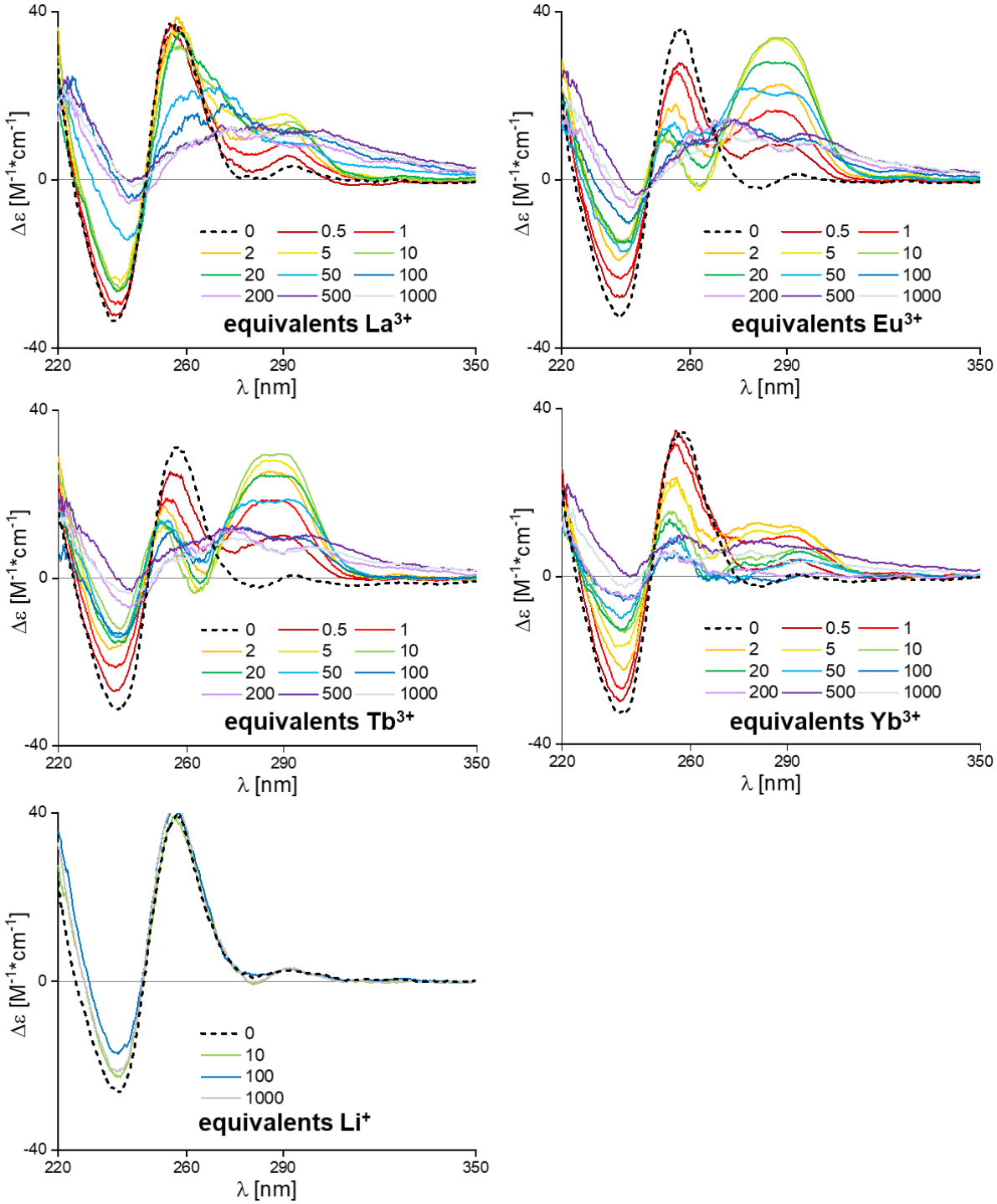


*Figure S1 CD titrations of 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**A) with varying equivalents of cations. Samples contain 10 µM DNA, 50 mM TMAA (pH 6.8), 0-10 mM metal chloride.*

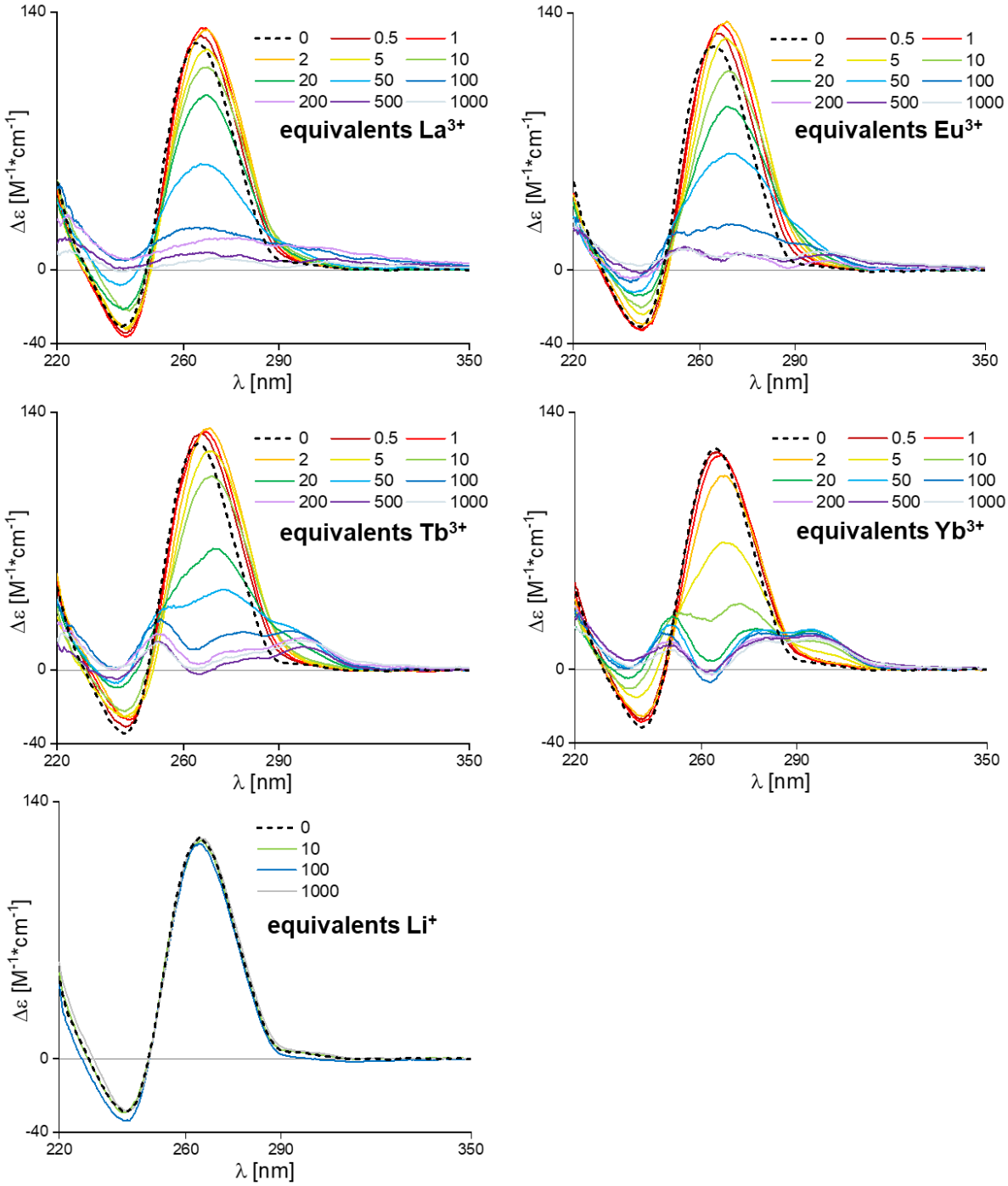


*Figure S2 CD titrations of 2LBY (dTA**G₃**A**G₃**TA**G₃**A**G₃**T) with varying equivalents of cations. Samples contain 10 µM DNA, 50 mM TMAA (pH 6.8), 0-10 mM metal chloride.*

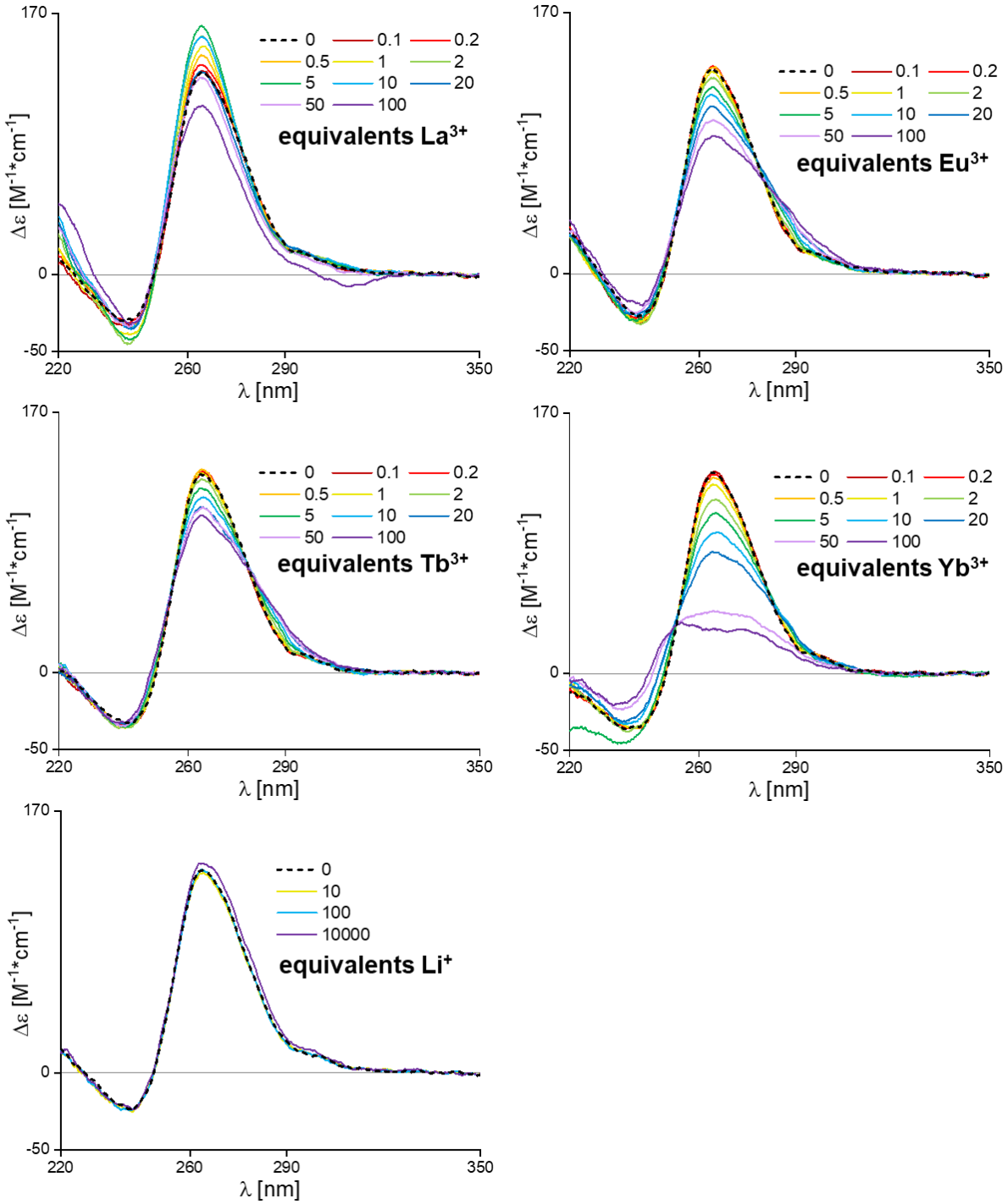


*Figure S3 CD titrations of 2LK7 (dTT**G₃**T**G₃**T**G₃**T**G₃**T) with varying equivalents of cations. Samples contain 10 µM DNA, 50 mM TMAA (pH 6.8), 0-10 mM metal chloride*

Mass spectra of desalted 2LK7 and 2LBY

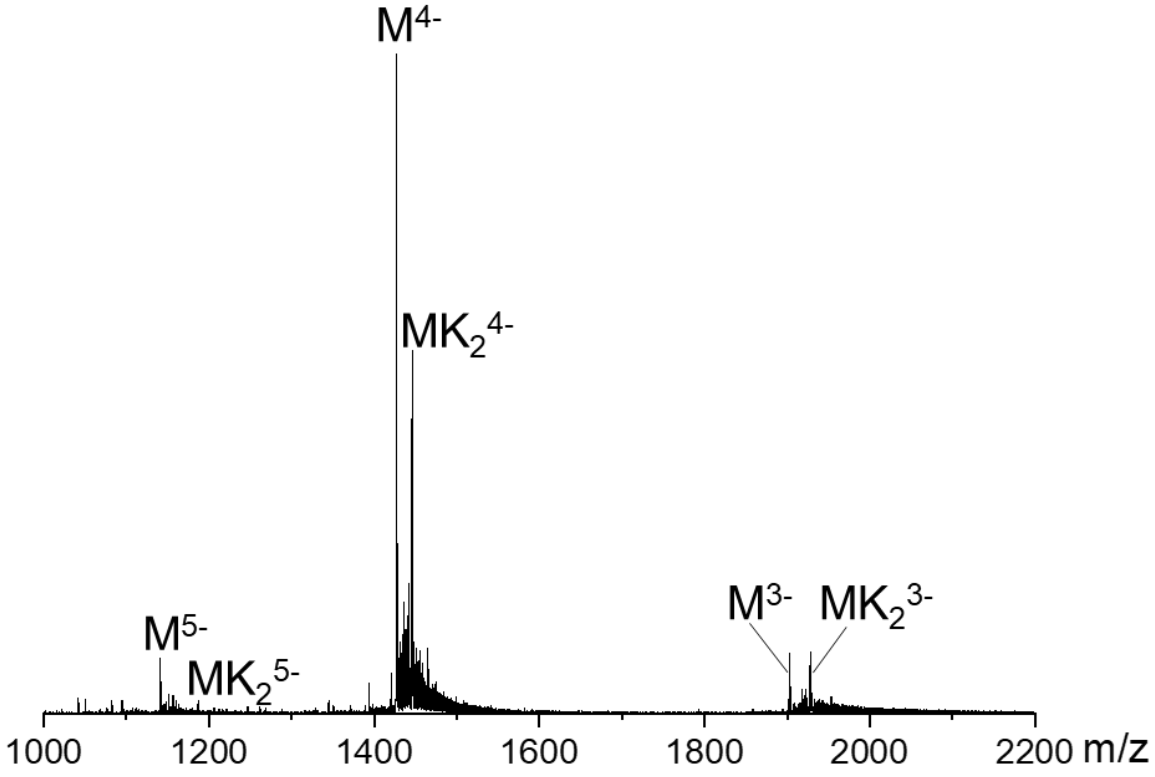


*Figure S4 Mass spectrum of 10 µM 2LK7 (dTT**G₃TG₃TG₃TG₃**T) and 50 mM TMAA (pH 6.8) in H_2O, measured on an Agilent 6560 IMS-QTOF. M = DNA, K = K^+.*

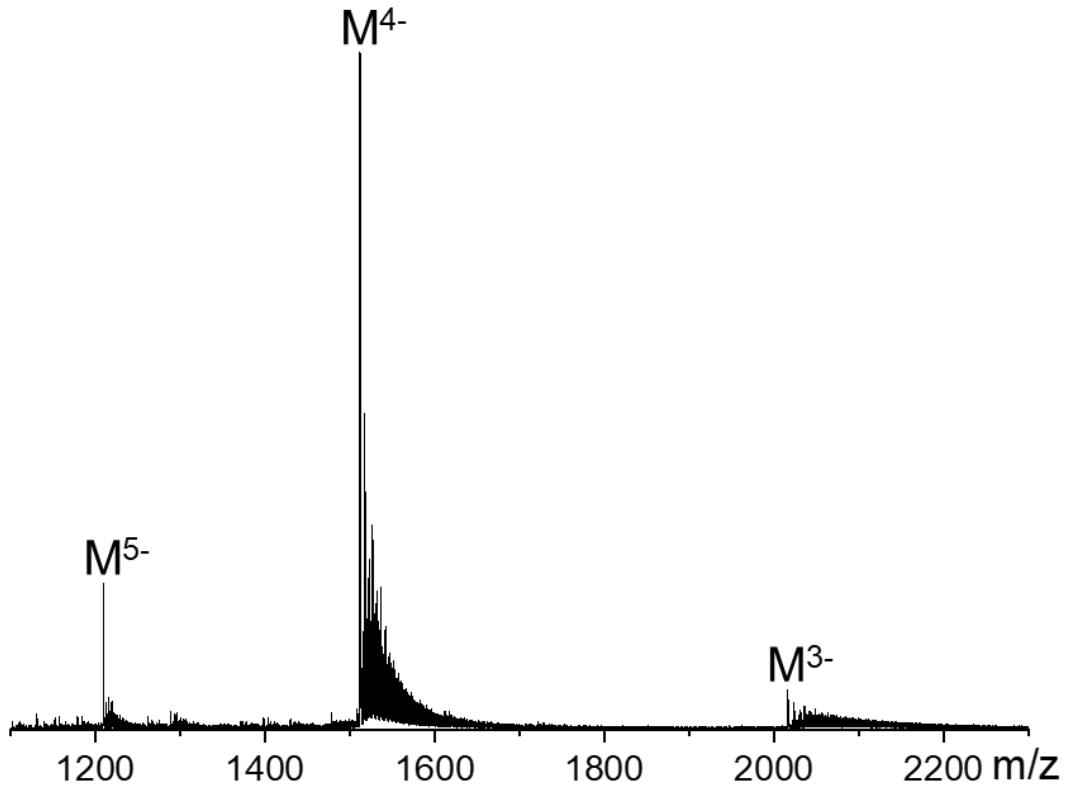


*Figure S5 Mass spectrum of 10 µM 2LBY (dTA**G₃AG₃TAG₃AG₃**T) and 50 mM TMAA (pH 6.8) in H_2O, measured on an Agilent 6560 IMS-QTOF. M = DNA, K = K^+.*

Luminescence

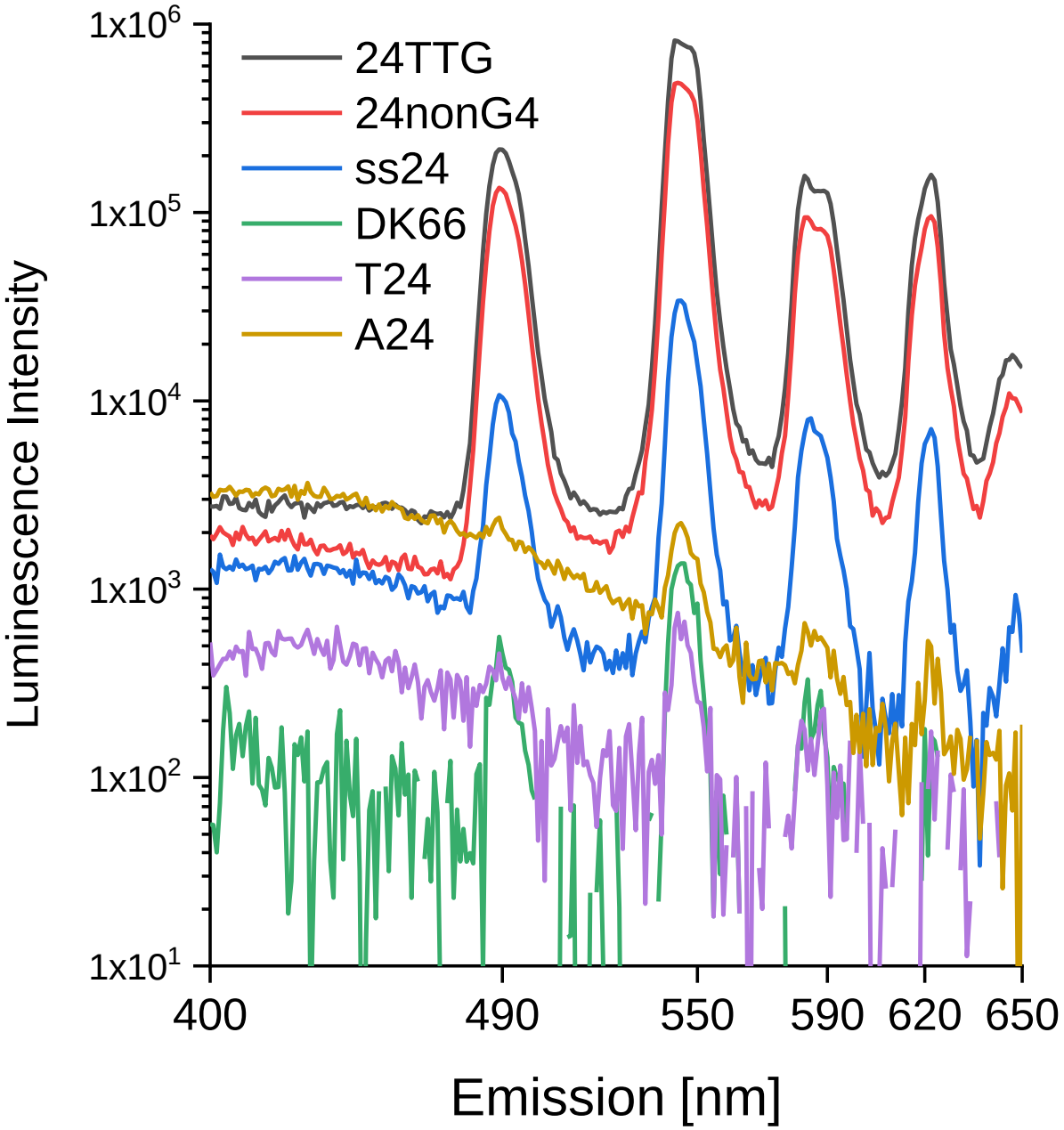


Figure S6 Emission spectra for the Tb^{3+} luminescence with different DNA sequences. Parenthesis show the G-richness of each sequence (DK66 has 33% but is also the only double-stranded species). Samples contain 10 µM DNA, 50 µM $TbCl_3$ and 50 mM TMAA (pH 6.8). Excitation wavelength is 290 nm. T = 20°C.

Full CD titrations of Tb^{3+} onto various 24-mer DNA sequences

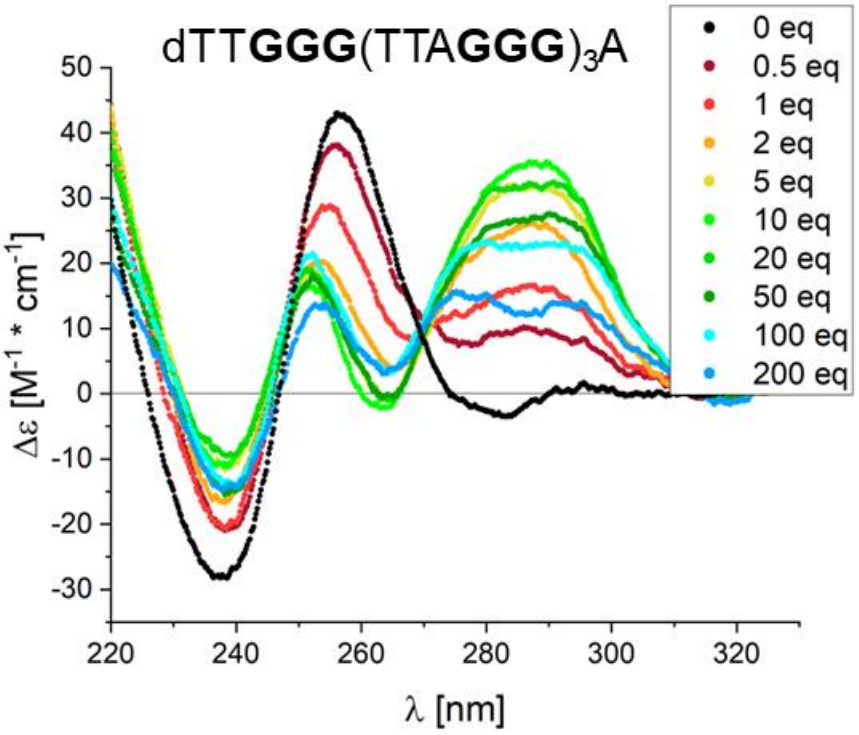


Figure S7 CD titration of Tb^{3+} equivalents (eq) onto 24TTG (sequence in image). Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

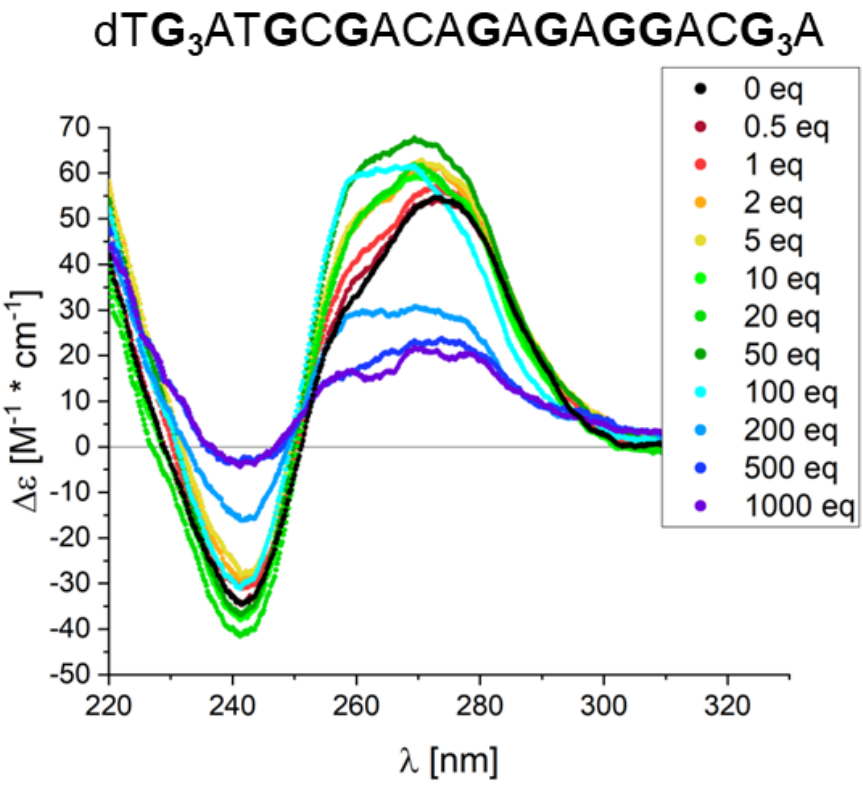


Figure S8 CD titration of Tb^{3+} equivalents (eq) onto 24nonG4 (sequence in image). Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

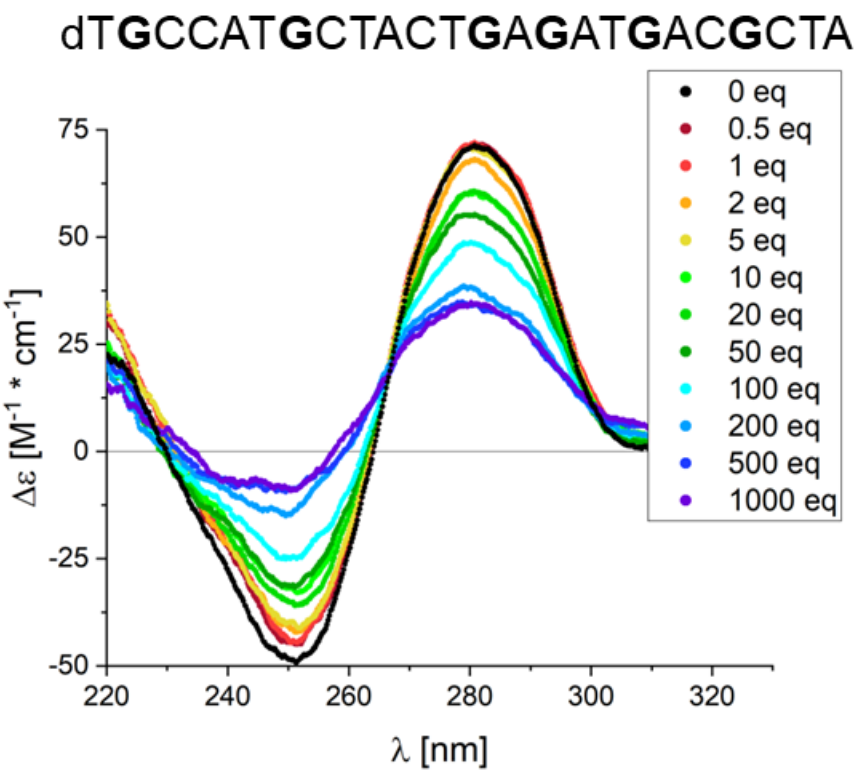


Figure S9 CD titration of Tb^{3+} equivalents (eq) onto ss24 (sequence in image). Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

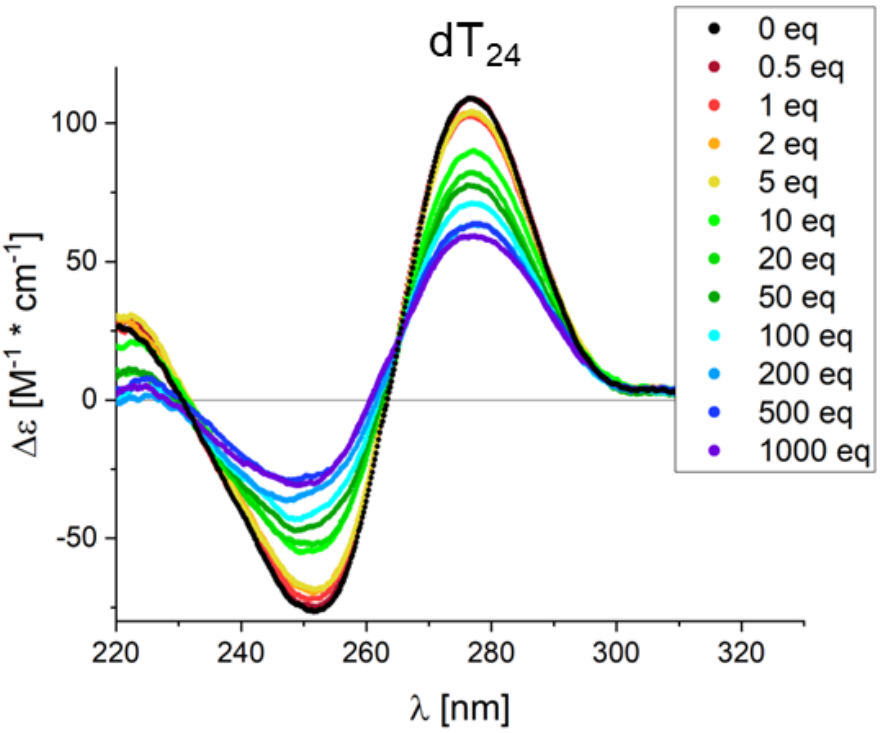


Figure S10 CD titration of Tb^{3+} equivalents (eq) onto T24 (sequence in image). Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

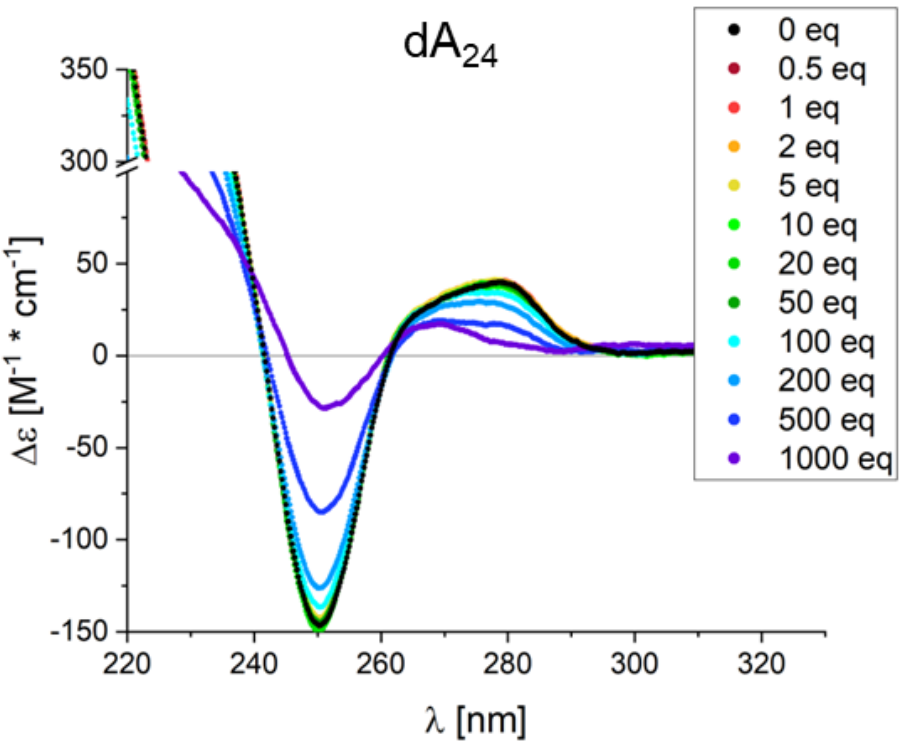


Figure S11 CD titration of Tb^{3+} equivalents (eq) onto A24 (sequence in image). Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

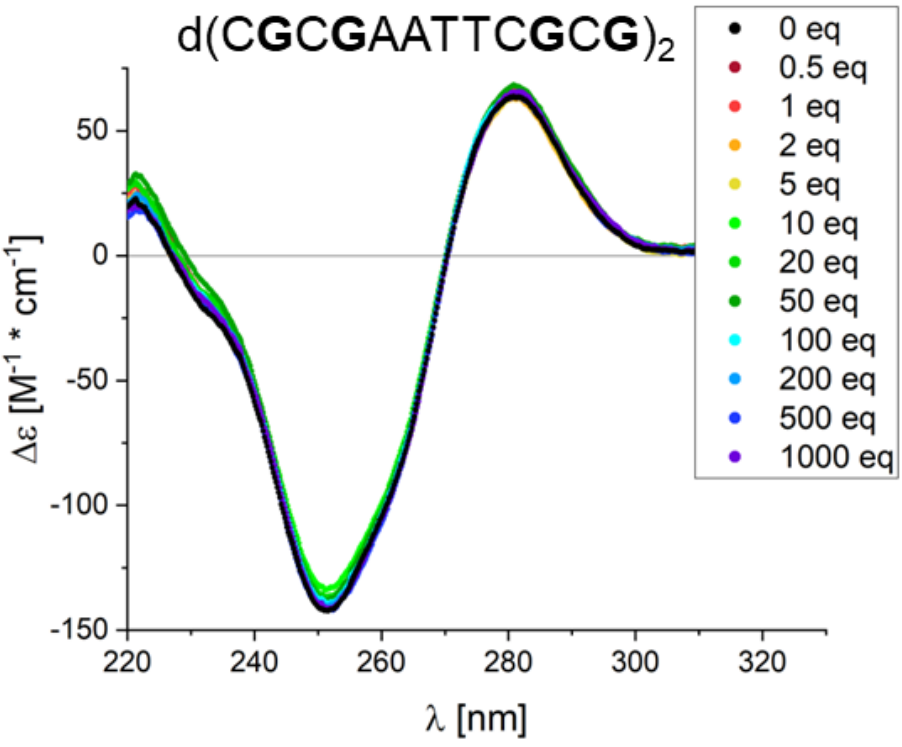


Figure S12 CD titration of Tb^{3+} equivalents (eq) onto DK66 (sequence in image). Samples contain 20 µM DNA, 0-10 mM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

Full MS titrations of Tb^{3+} onto various 24-mer DNA sequences

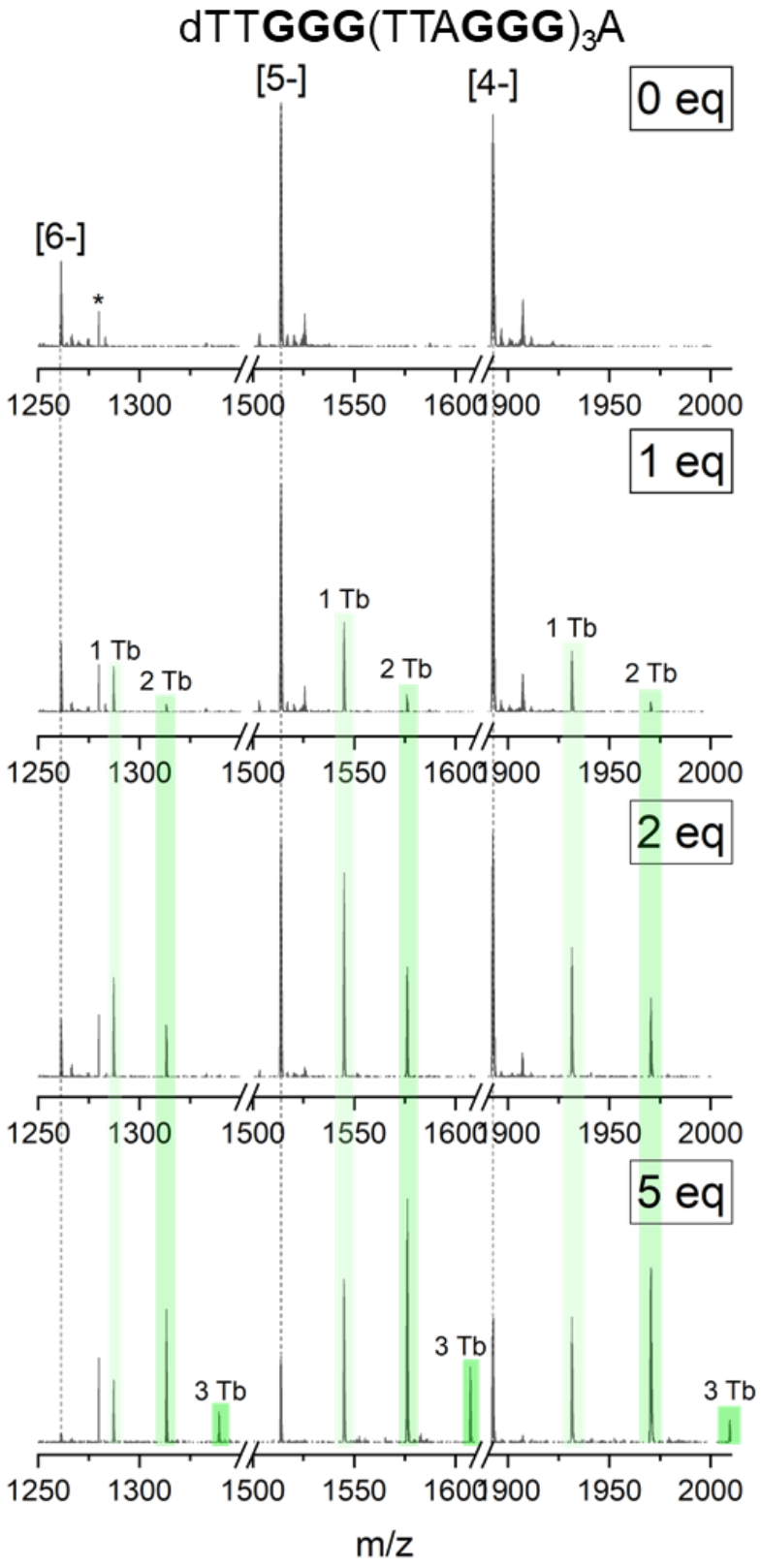


Figure S13 MS titration of Tb^{3+} equivalents (eq) onto 24TTG (sequence in image). Samples contain 10 µM DNA, 0-50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

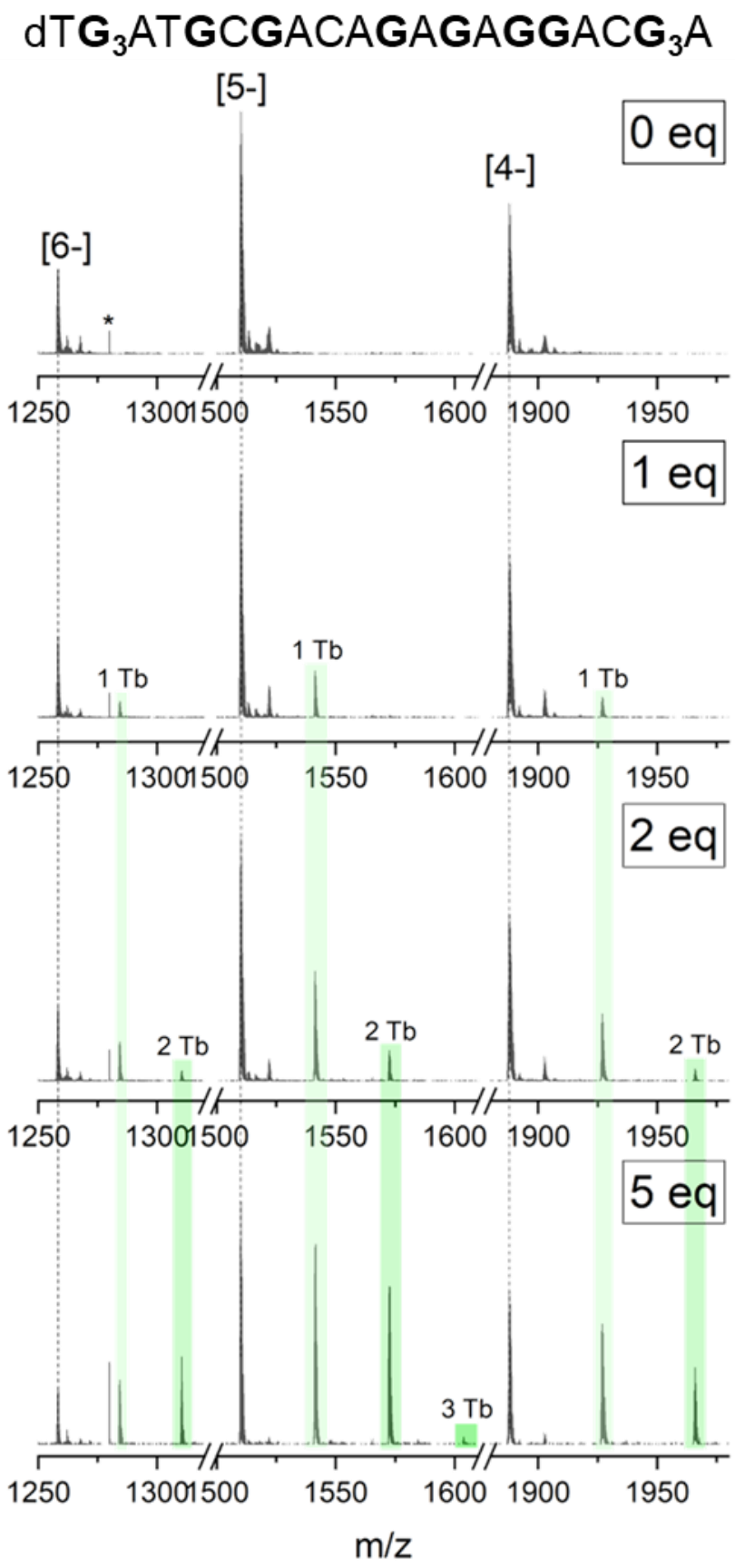


Figure S14 MS titration of Tb^{3+} equivalents (eq) onto 24nonG4 (sequence in image). Samples contain 10 µM DNA, 0-50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

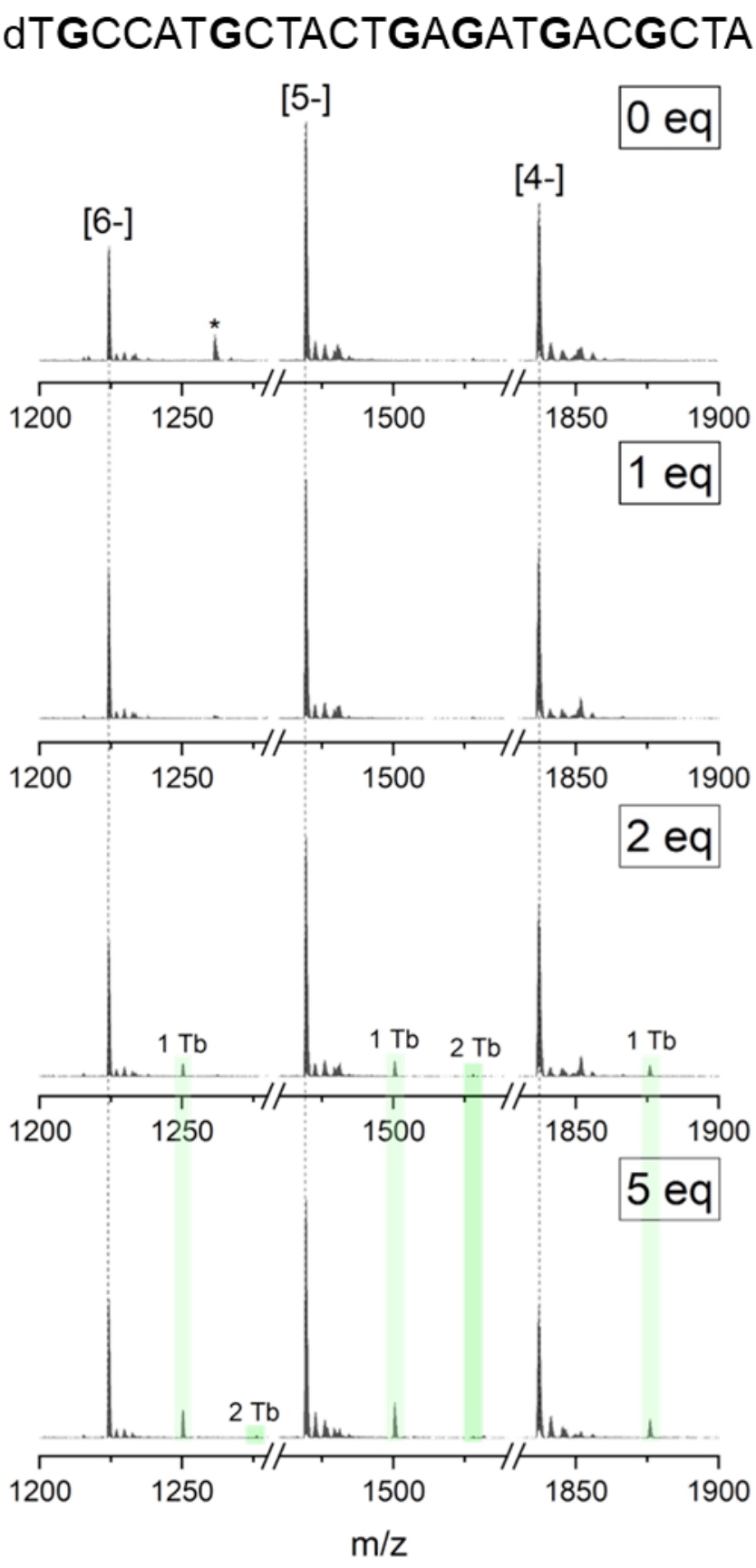


Figure S15 MS titration of Tb^{3+} equivalents (eq) onto ss24 (sequence in image). Samples contain 10 µM DNA, 0-50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

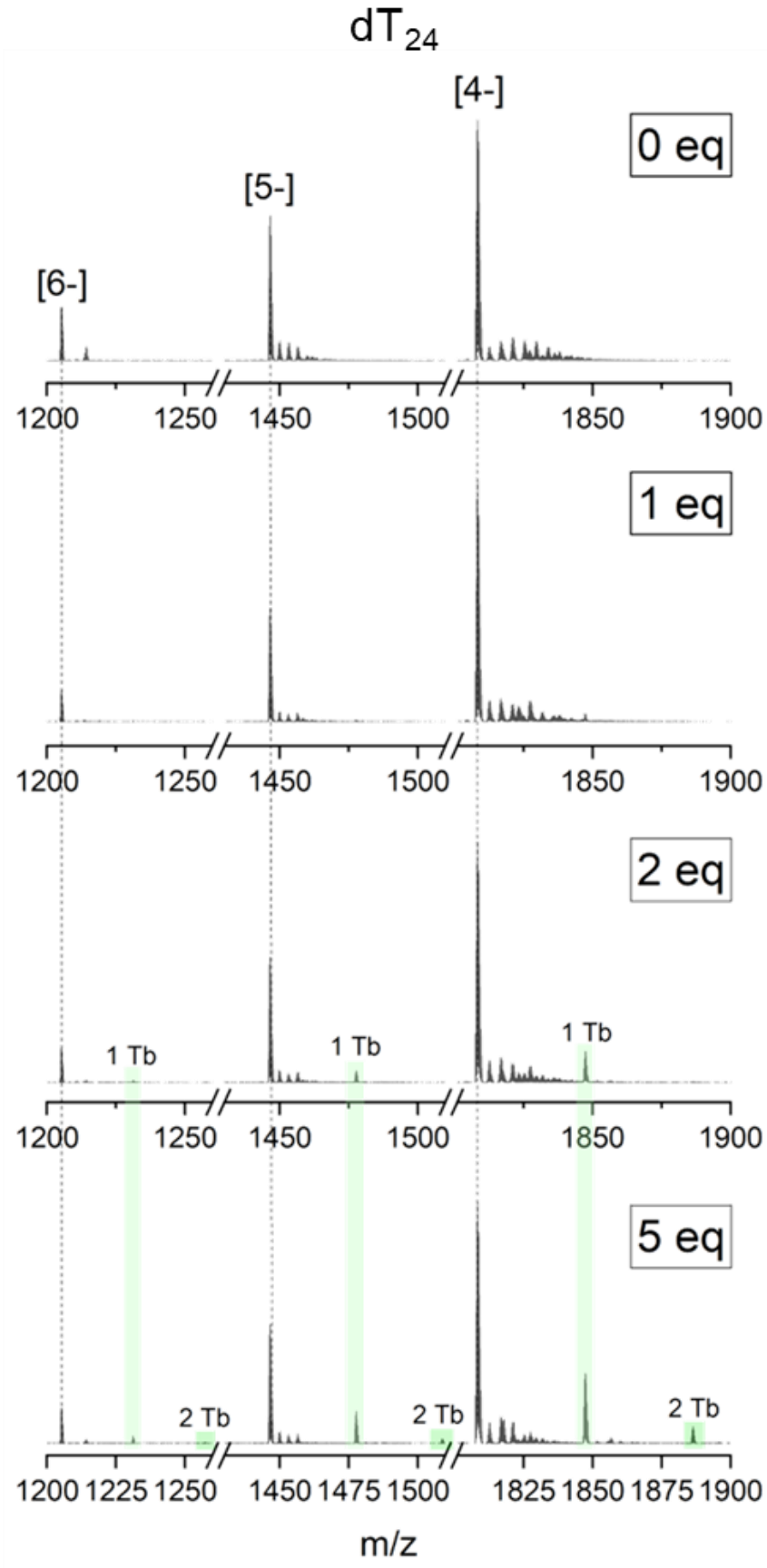


Figure S16 MS titration of Tb^{3+} equivalents (eq) onto T24 (sequence in image). Samples contain 10 µM DNA, 0-50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

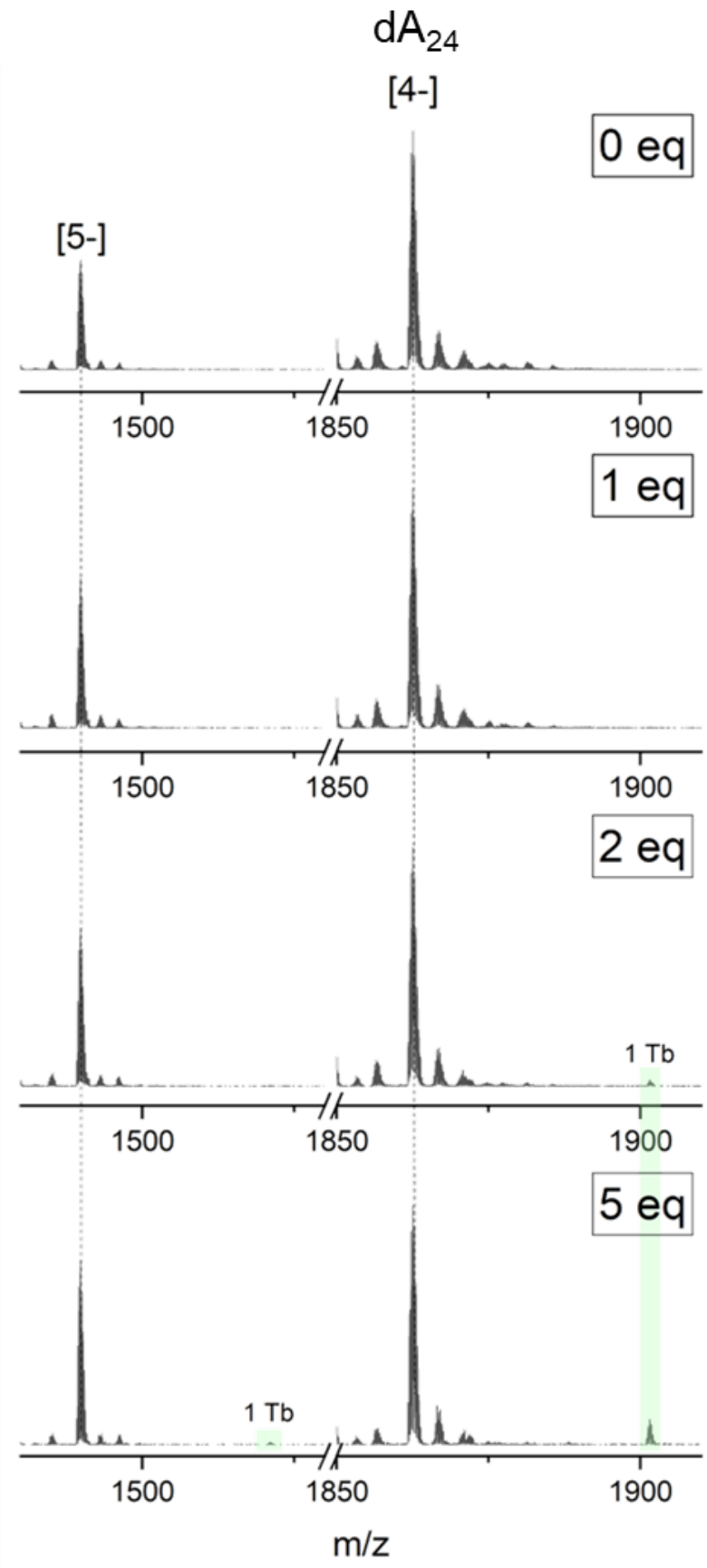


Figure S17 MS titration of Tb^{3+} equivalents (eq) onto A24 (sequence in image). Samples contain 10 µM DNA, 0-50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

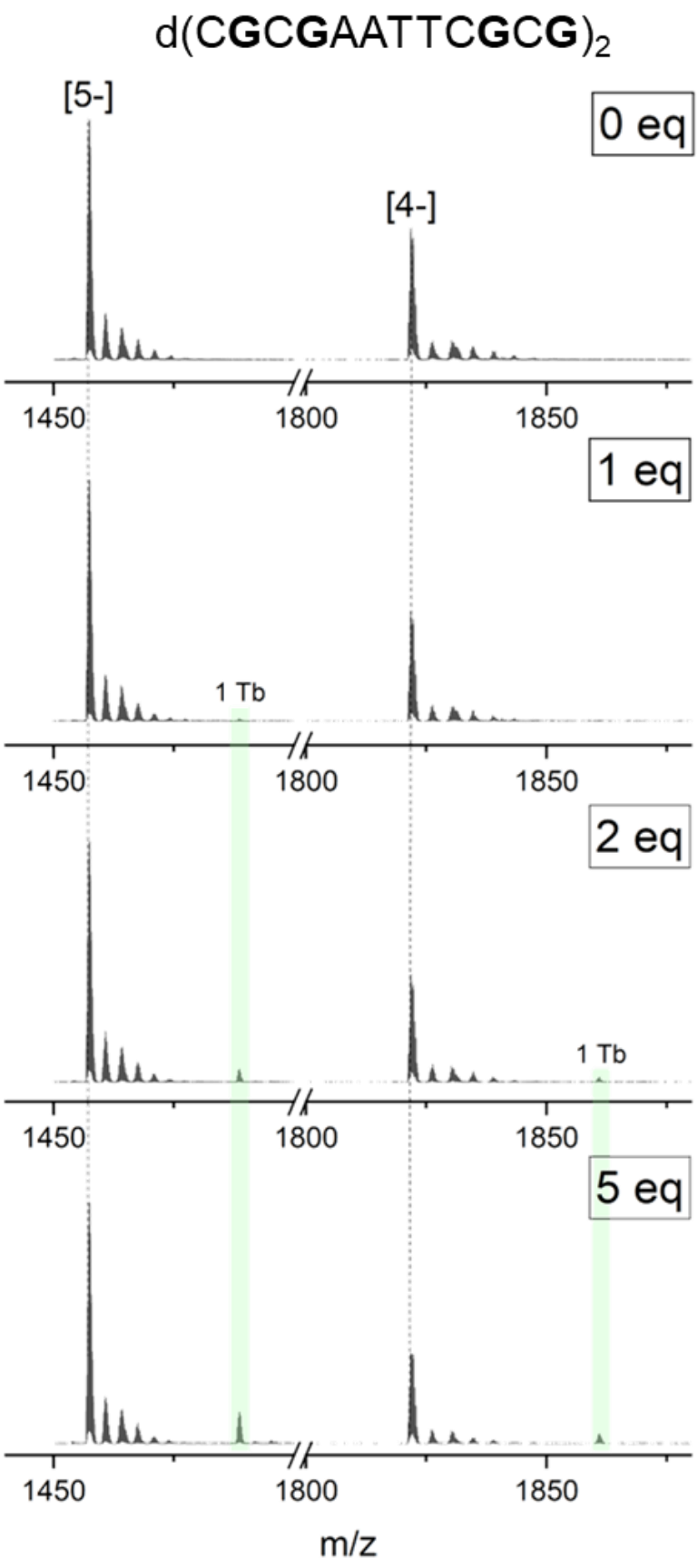


Figure S18 MS titration of Tb^{3+} equivalents (eq) onto DK66 (sequence in image). Samples contain 20 µM DNA, 0-50 µM $TbCl_3$, 50 mM TMAA (pH 6.8) in H_2O.

24TTG competition experiments with K^+ and Tb^{3+}

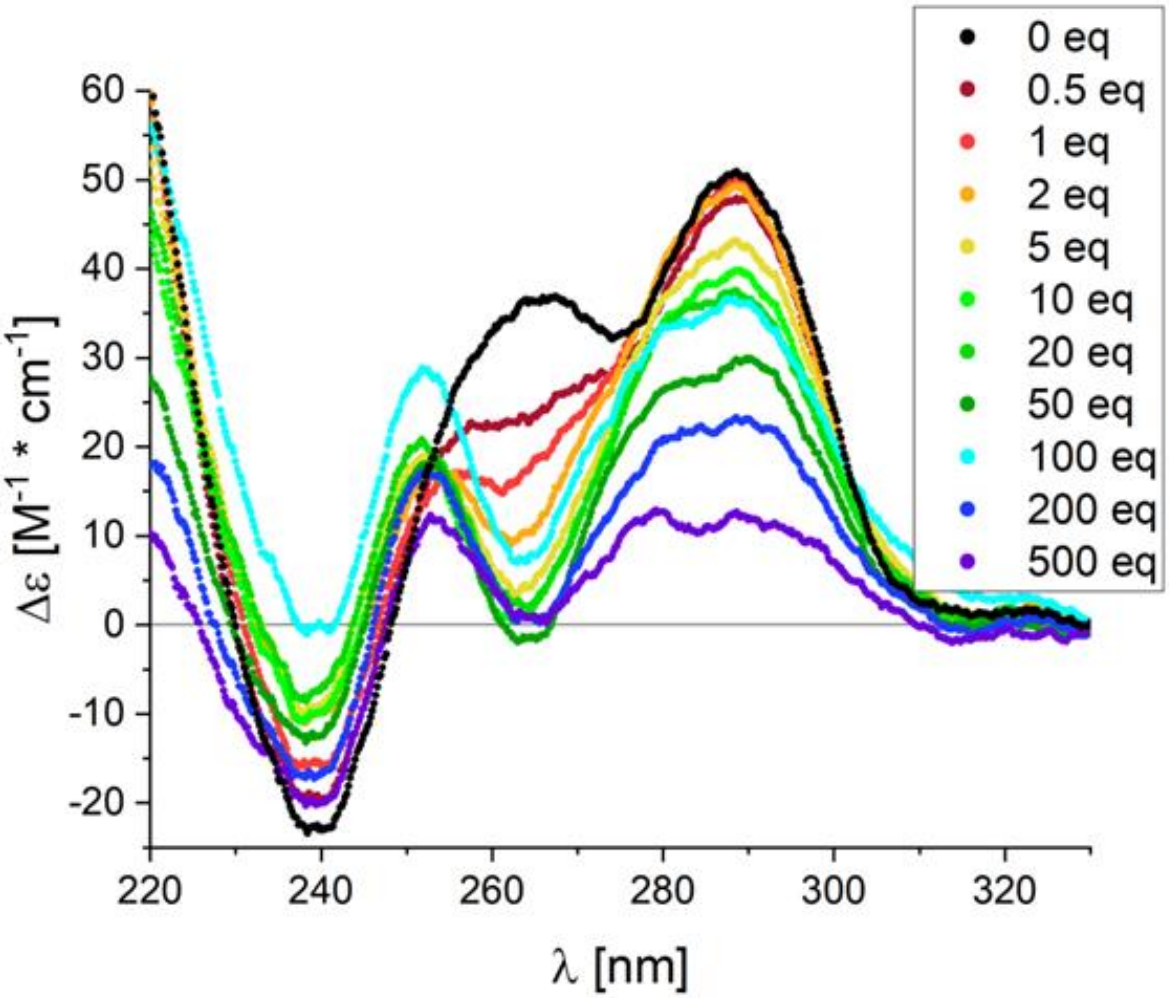


*Figure S19 CD titration of Tb^{3+} equivalents (eq) onto 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A), pre-folded in 100 µM K^+. Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 100 µM KCl, 50 mM TMAA (pH 6.8) in H_2O.*

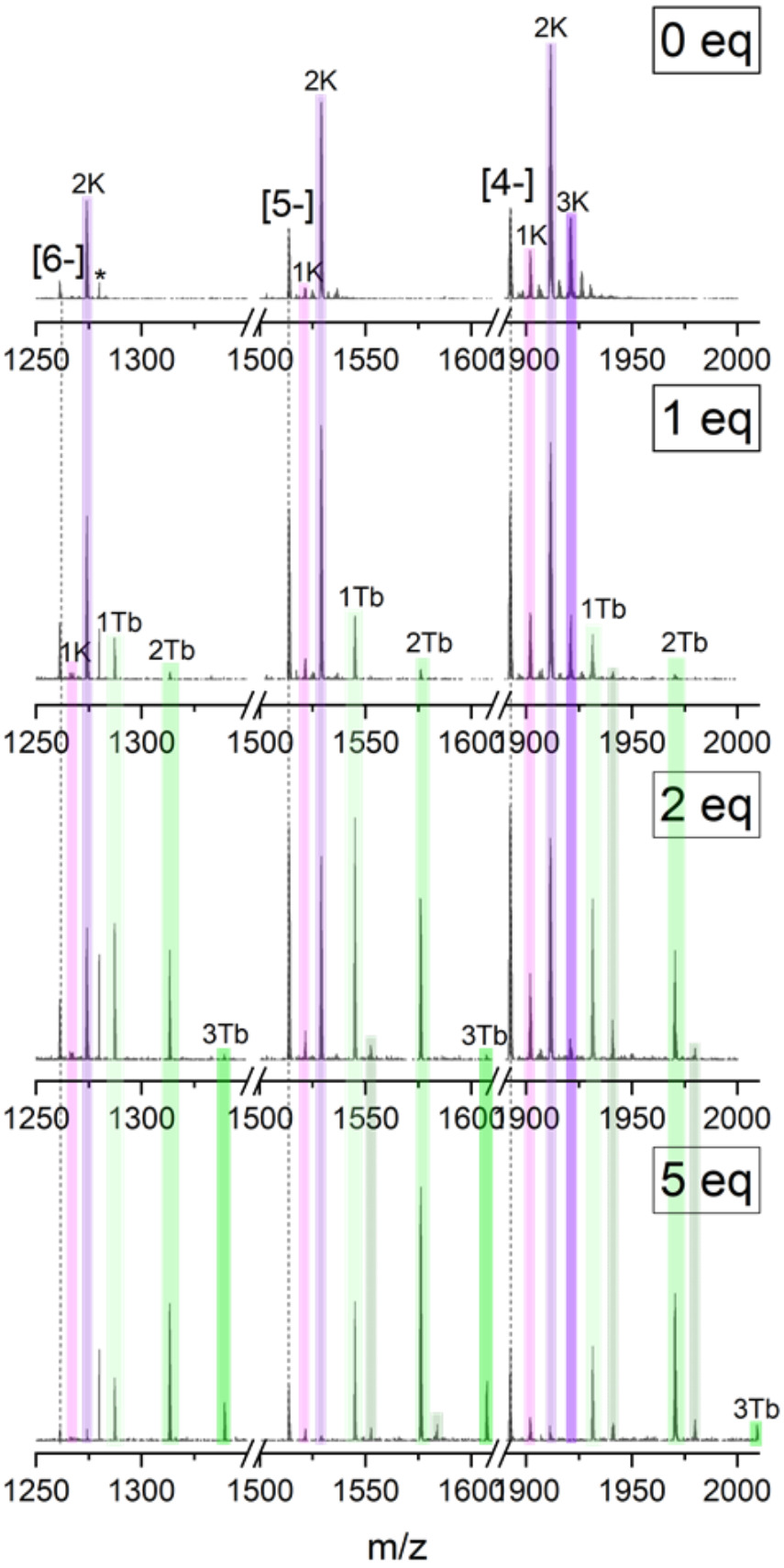


*Figure S20 MS titration of Tb^{3+} equivalents (eq) onto 24TTG (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A), pre-folded in 100 µM K^+. Samples contain 10 µM DNA, 0-10 mM $TbCl_3$, 100 µM KCl, 50 mM TMAA (pH 6.8) in H_2O.*

Chapter 2

CD and MS spectra at 100/0.5/0.05 mM KCl

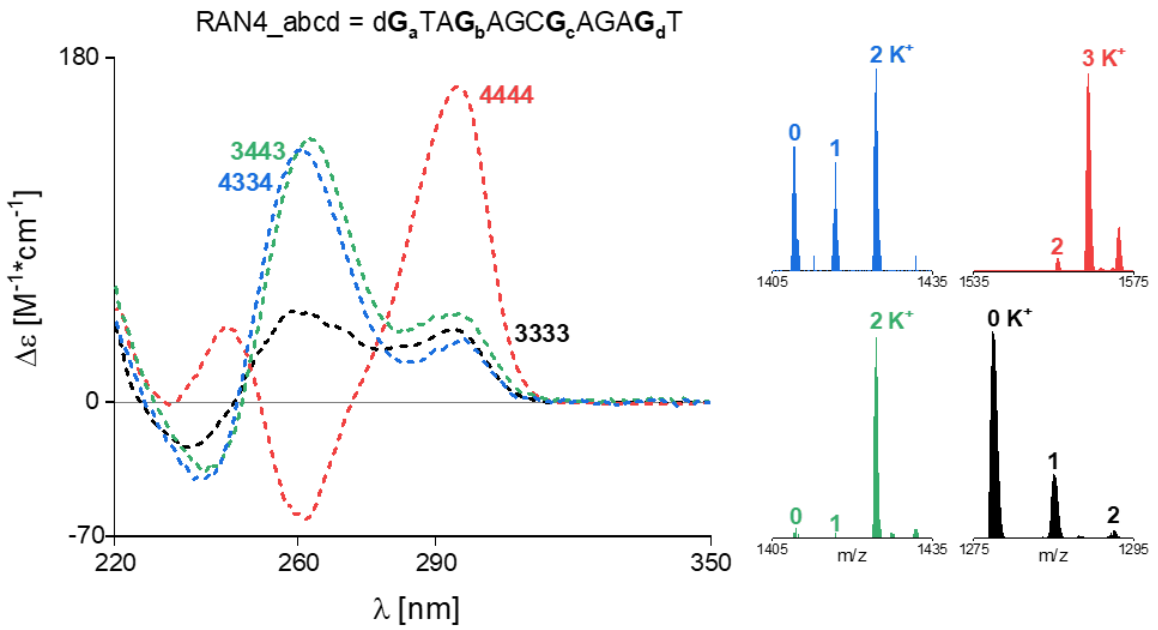


Figure S21 ***Left panel****: CD spectra of the RAN4 sequence motif with varying G-tract lengths.* ***Right panel****: Corresponding mass spectra, showing the 5- charge state. Sample conditions are: 10 µM DNA, 500 µM KCl and 100 mM TMAA (pH 6.8).*

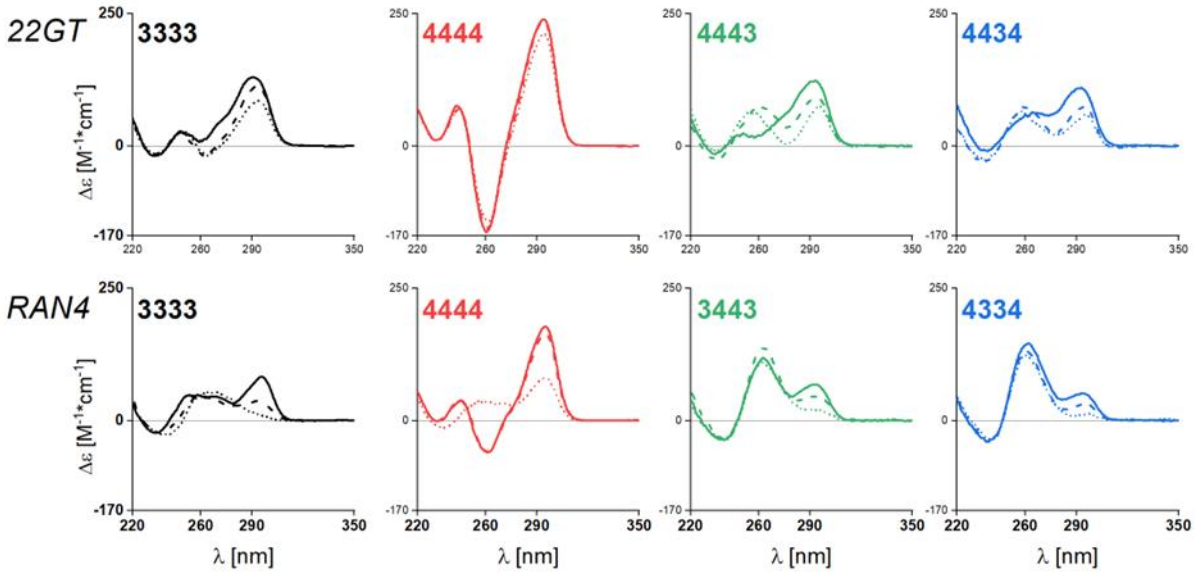


Figure S22 CD signatures of ***Top:*** *telomeric repeat sequence 22GT (d**G**aTTA**G**bTTA**G**cTTA**G**dT) and* ***Bottom:*** *RANKL promoter sequence RAN4 (d**G**aTA**G**bAGC**G**cAGA**G**d) at different KCl concentrations (full: 100 mM, dashed: 500 µM, dotted: 50 µM). The four-digit numbers indicate the length of G-tracts. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8); samples with 100 mM KCl only contain 10 mM TMAA. Optical path length is 10 mm. Temperature is 22°C.*

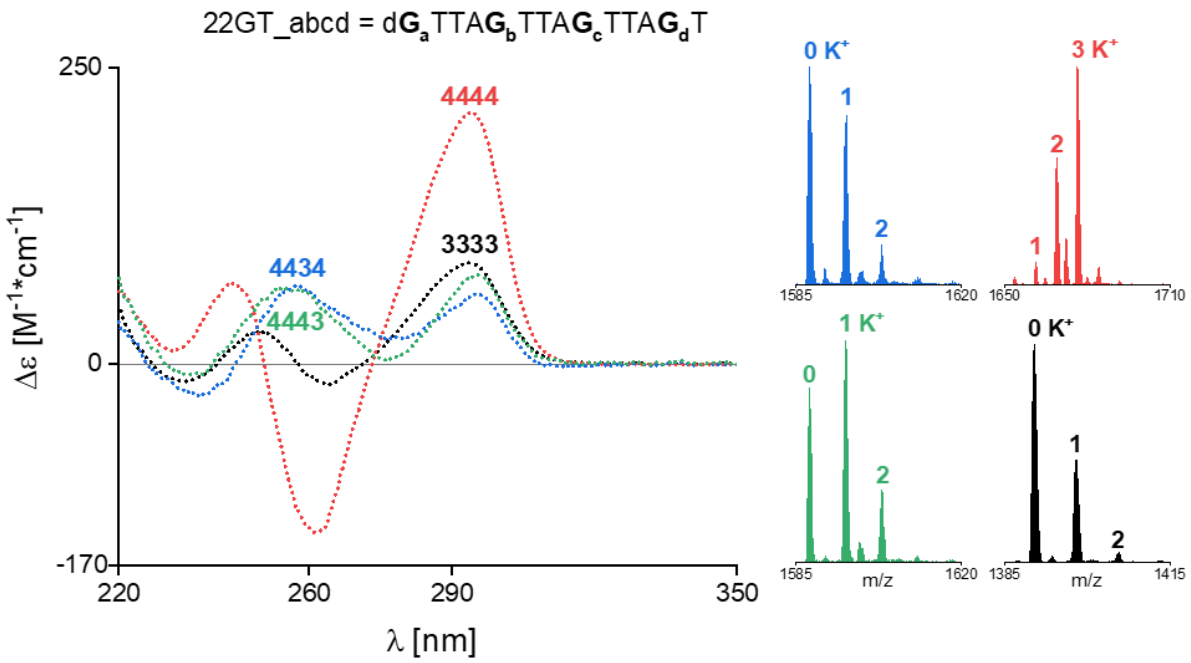


Figure S23 ***Left panel****: CD spectra of the 22GT sequence motif with varying G-tract lengths.* ***Right panel****: Corresponding mass spectra, showing the 5- charge state. Sample conditions are: 10 µM DNA, 50 µM KCl and 100 mM TMAA (pH 6.8).*

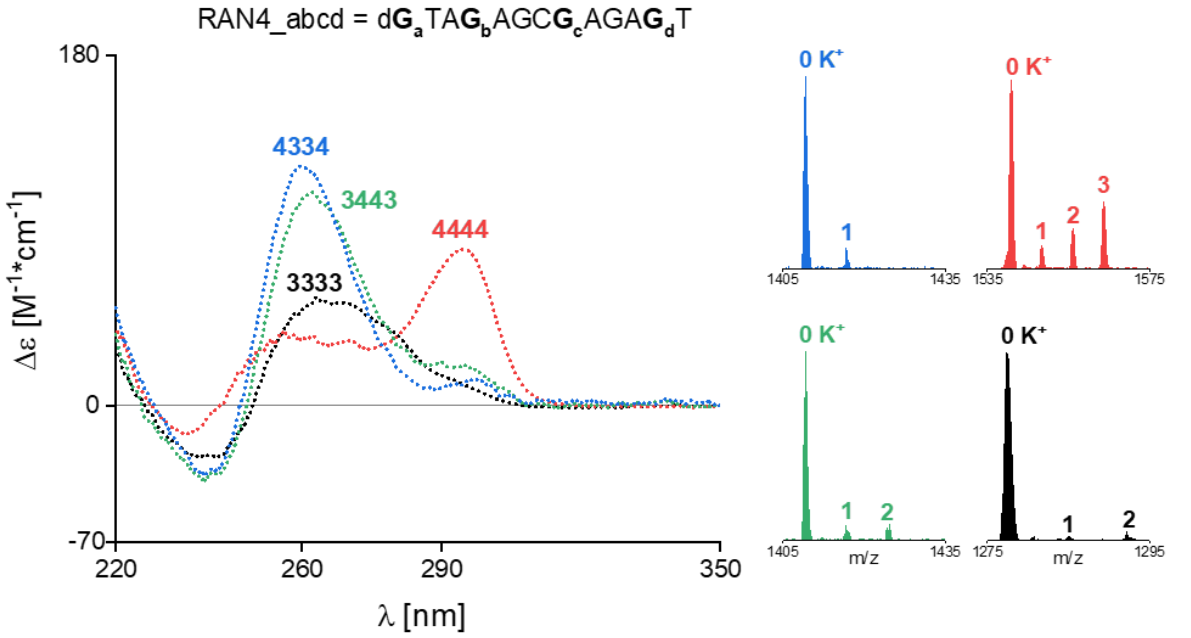


Figure S24 ***Left panel****: CD spectra of the RAN4 sequence motif with varying G-tract lengths.* ***Right panel****: Corresponding mass spectra, showing the 5- charge state. Sample conditions are: 10 µM DNA, 50 µM KCl and 100 mM TMAA (pH 6.8).*

CD melting

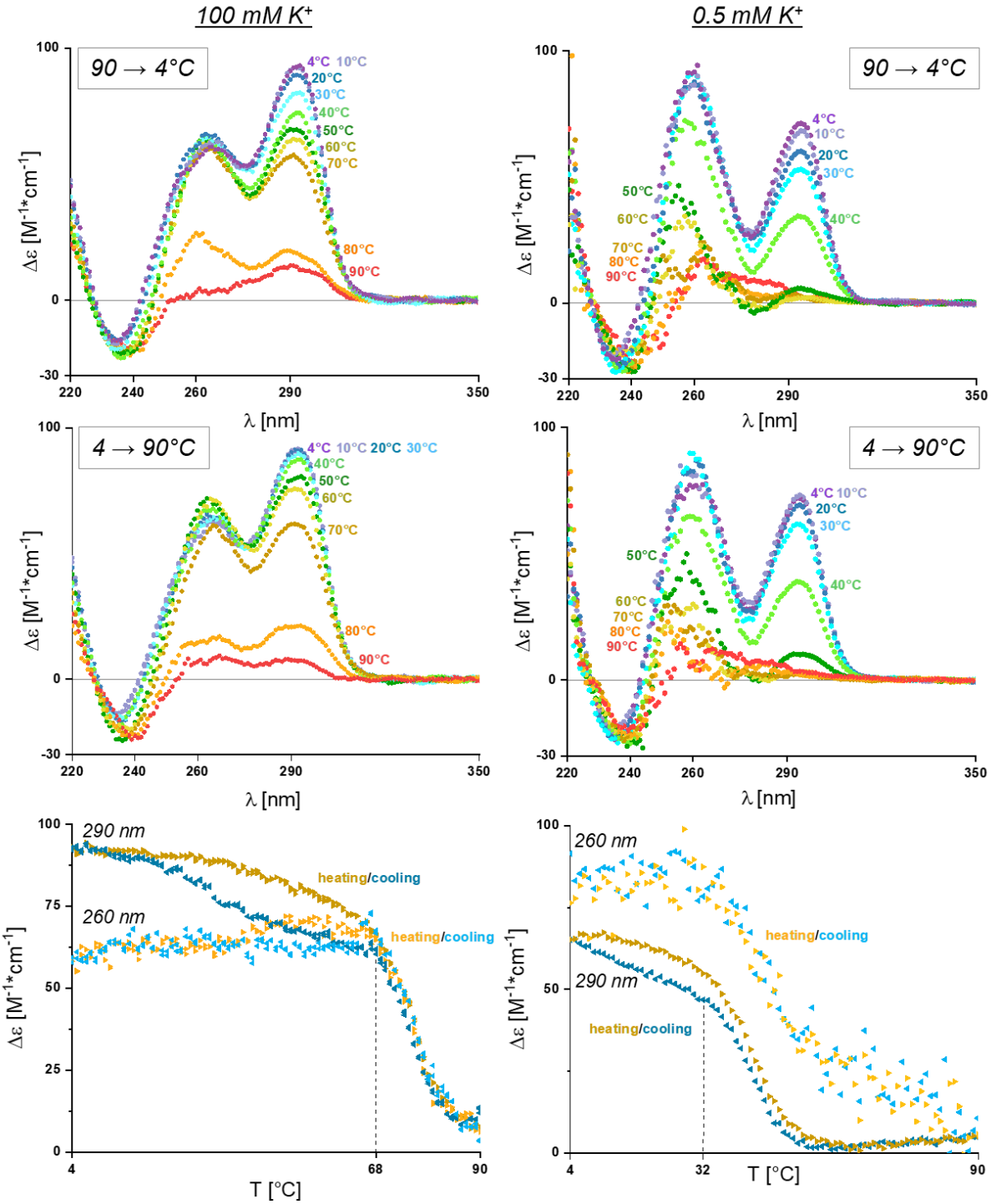


*Figure S25 Temperature-dependent CD spectra of 22GT_4443 (d**G₄**TTA**G₄**TTA**G₄**TTA**G₃**T) for both the cooling (**top**) and heating (**middle**) cycle. Melting curves are shown for maxima of the two CD bands (**bottom**). Sample conditions are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8).*

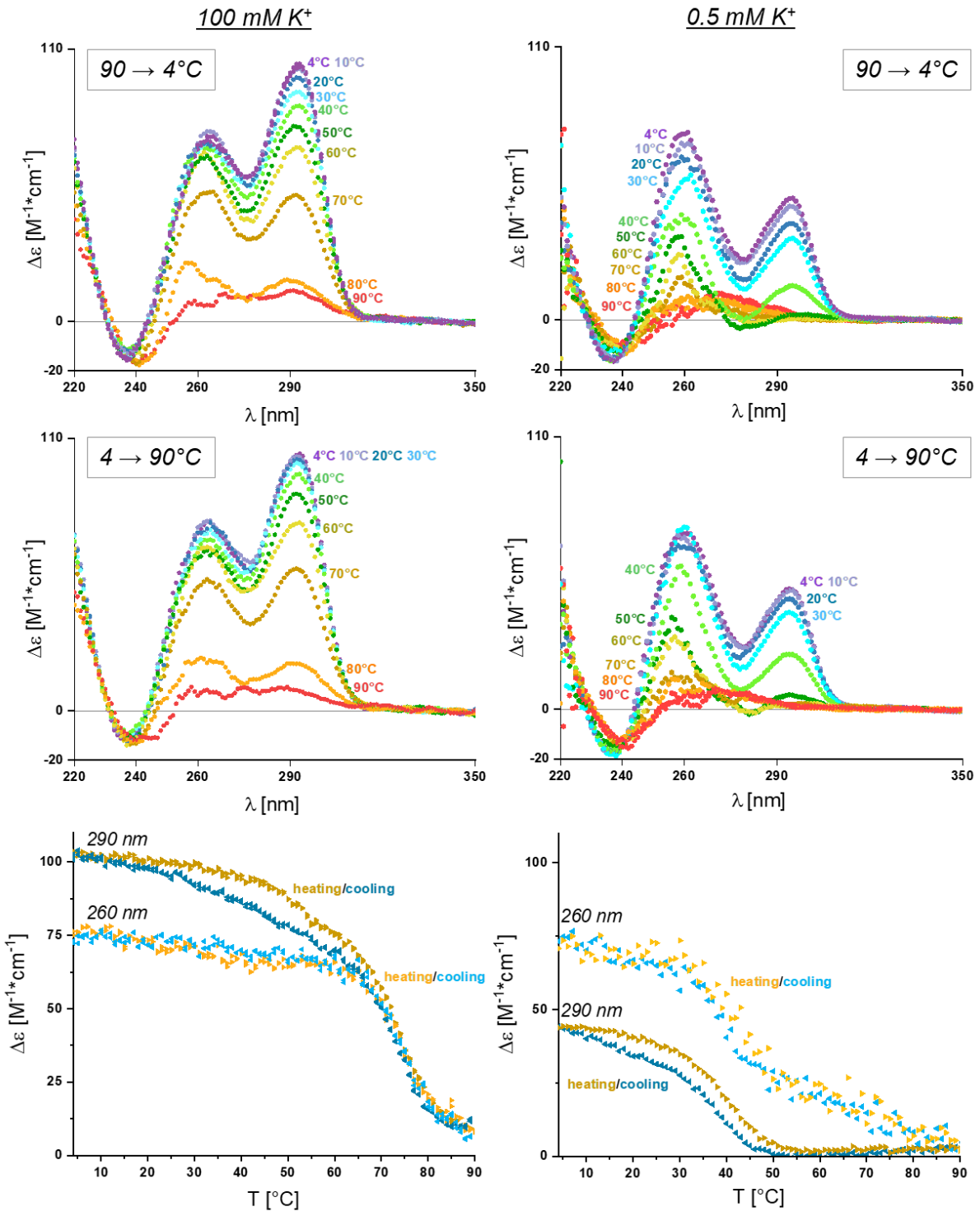


*Figure S26 Temperature-dependent CD spectra of 22GT_4434 (d**G₄**TTA**G₄**TTA**G₃**TTA**G₄**T) for both the cooling (**top**) and heating (**middle**) cycle. Melting curves are shown for maxima of the two CD bands (**bottom**). Sample conditions are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8).*

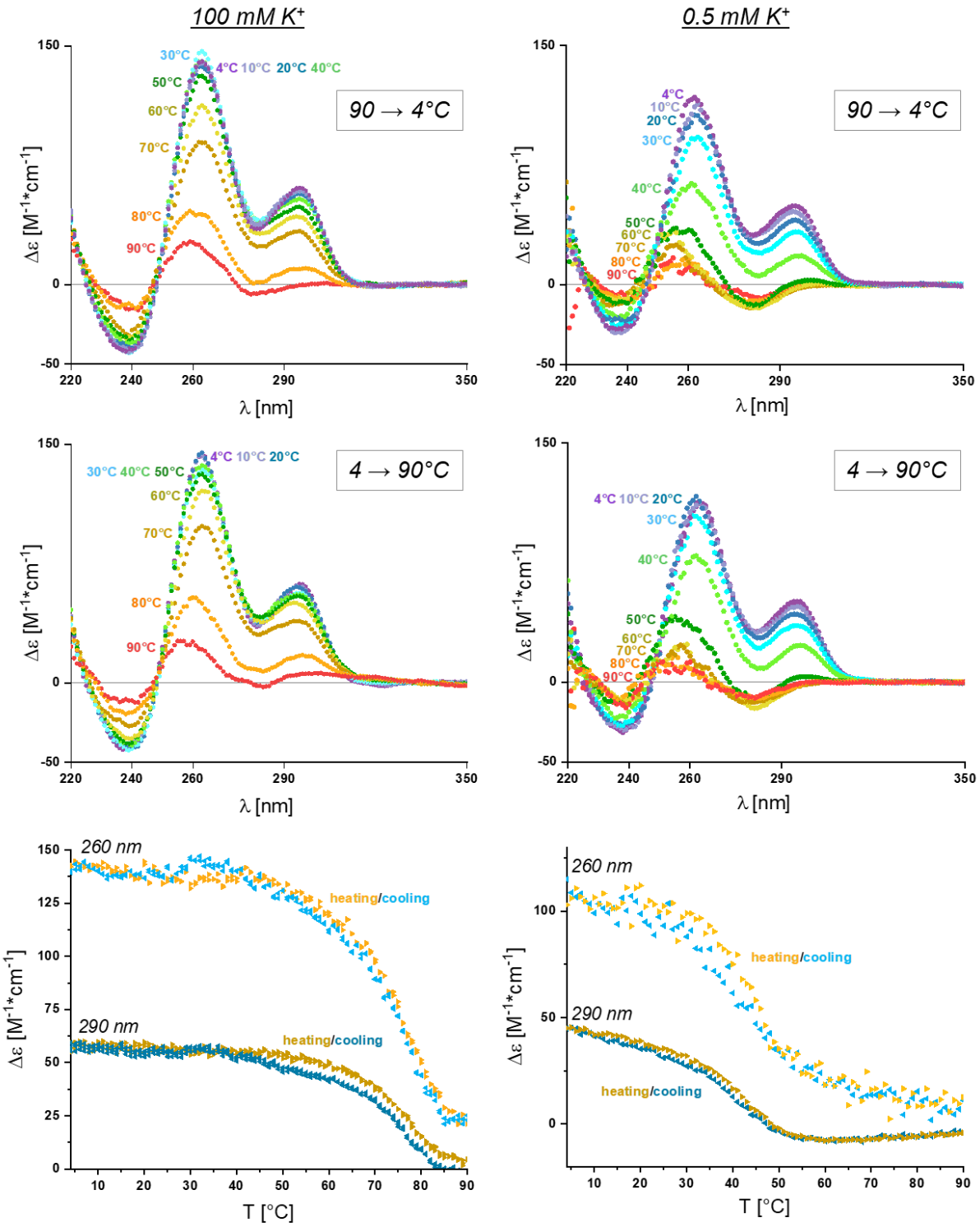


*Figure S27 Temperature-dependent CD spectra of RAN4_3443 (d**G**$_3$TA**G**$_4$AGC**G**$_4$AGA**G**$_3$) for both the cooling (**top**) and heating (**middle**) cycle. Melting curves are shown for maxima of the two CD bands (**bottom**). Sample conditions are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8).*

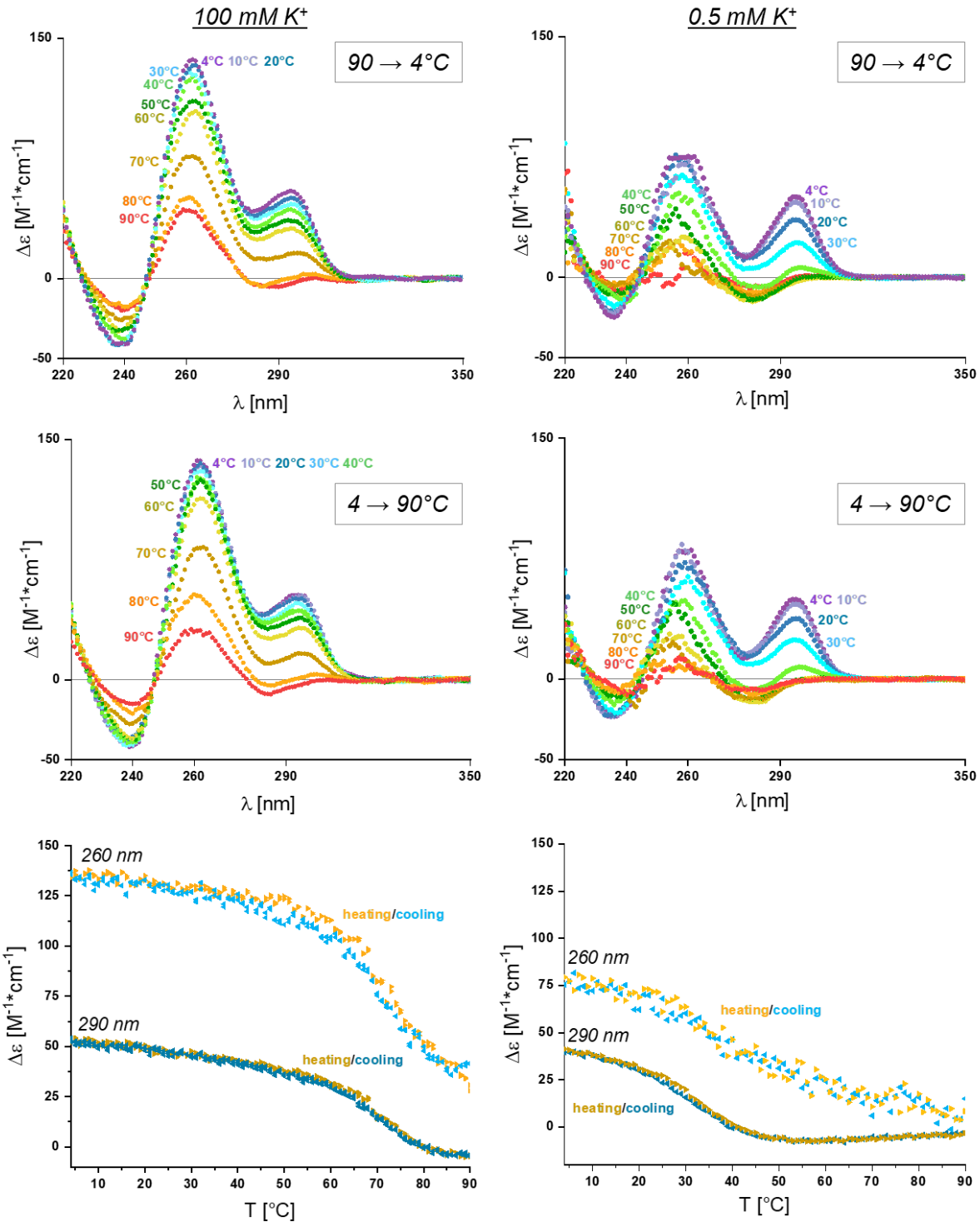


*Figure S28 Temperature-dependent CD spectra of RAN4_4334 (d**G₄**TA**G₃**AGC**G₃**AGA**G₄**) for both the cooling (**top**) and heating (**middle**) cycle. Melting curves are shown for maxima of the two CD bands (**bottom**). Sample conditions are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8).*

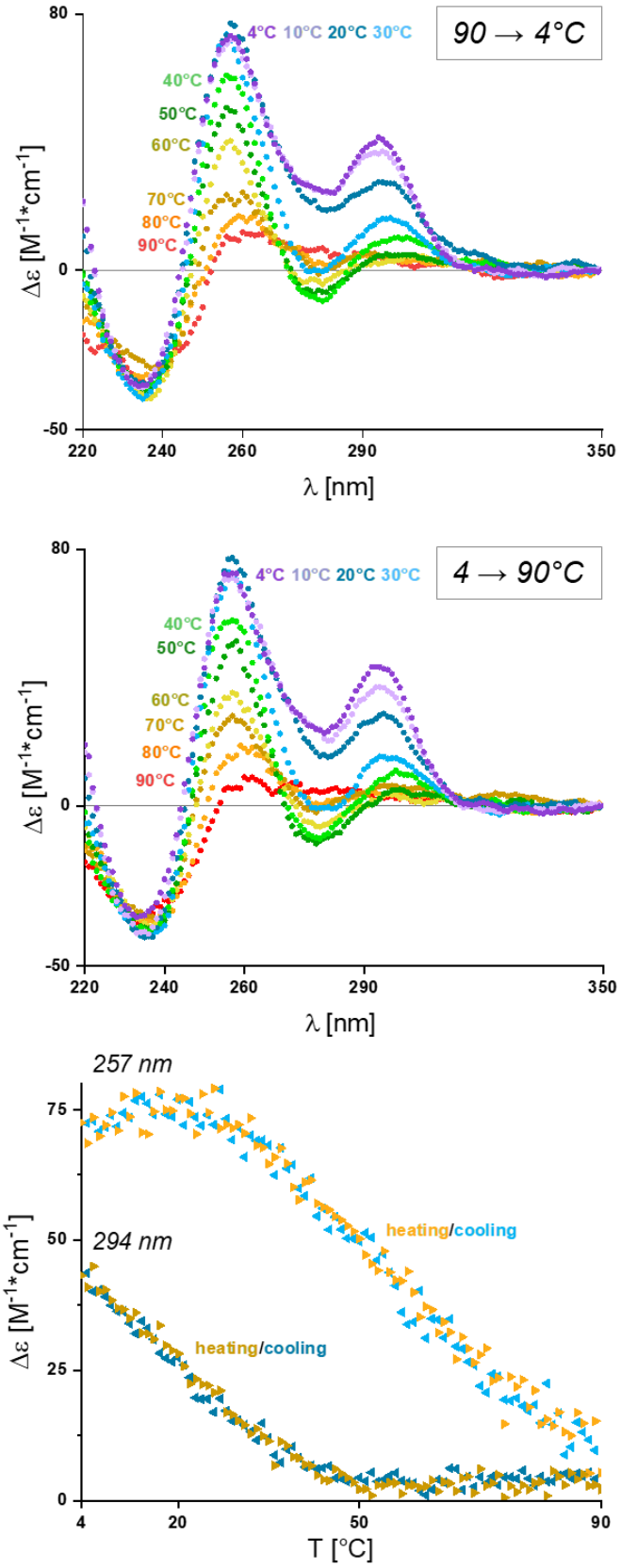


*Figure S29 Temperature-dependent CD spectra of 22GT_4443 (d**G₄**TTA**G₄**TTA**G₄**TTA**G₃**T) for both the cooling (**top**) and heating (**middle**) cycle. Melting curves are shown for maxima of the two CD bands (**bottom**). Sample conditions are: 10 µM DNA, 100 mM LiCl, 10 mM TMAA (pH 6.8).*

Singular value decompositions of CD melting data

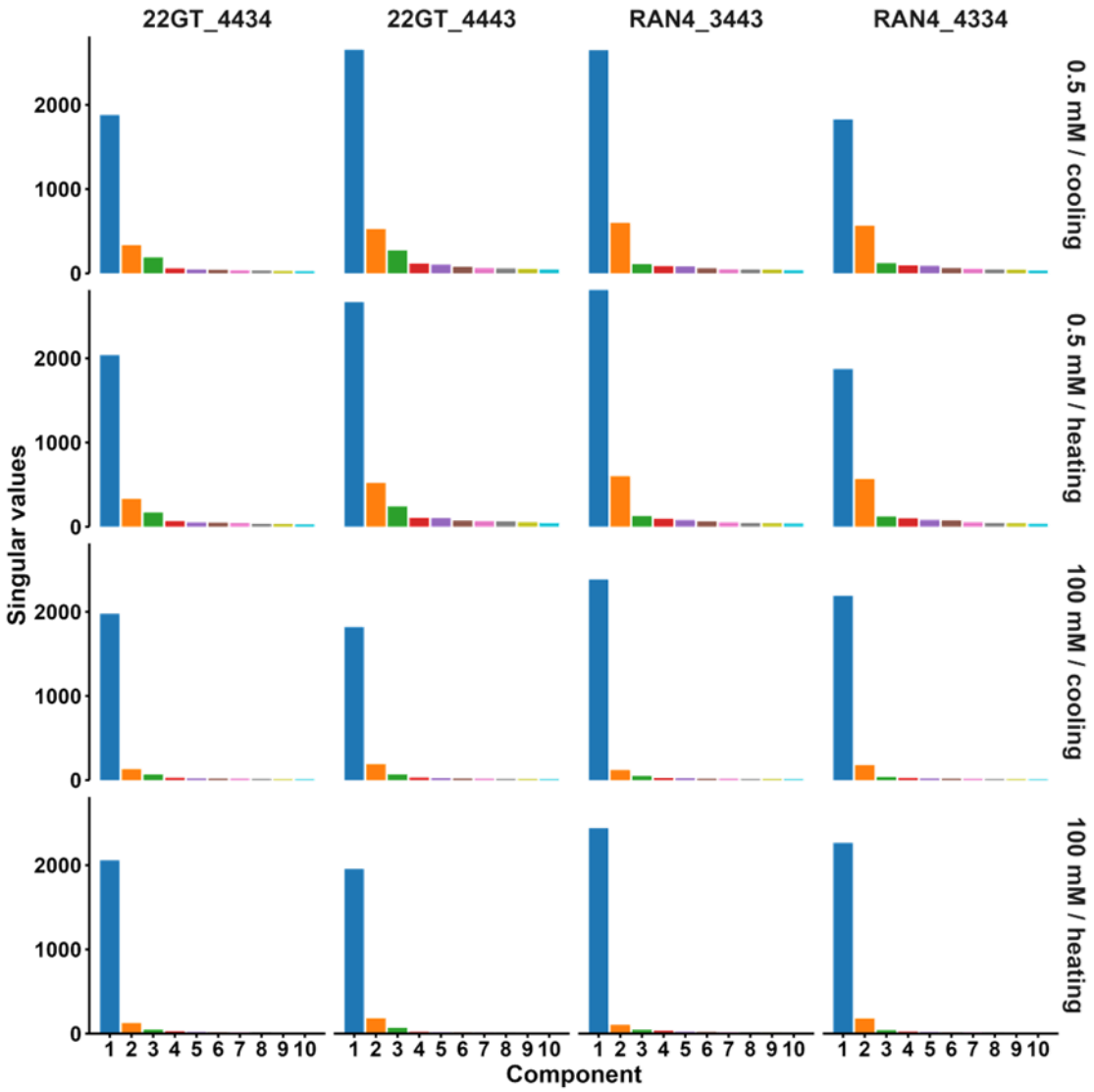


Figure S30 The singular values associated with each component (CD signature) extracted from CD melting data by singular value decomposition. The higher the singular value, the more the component contributes to the overall CD signature. Sample conditions for the CD melting are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8)

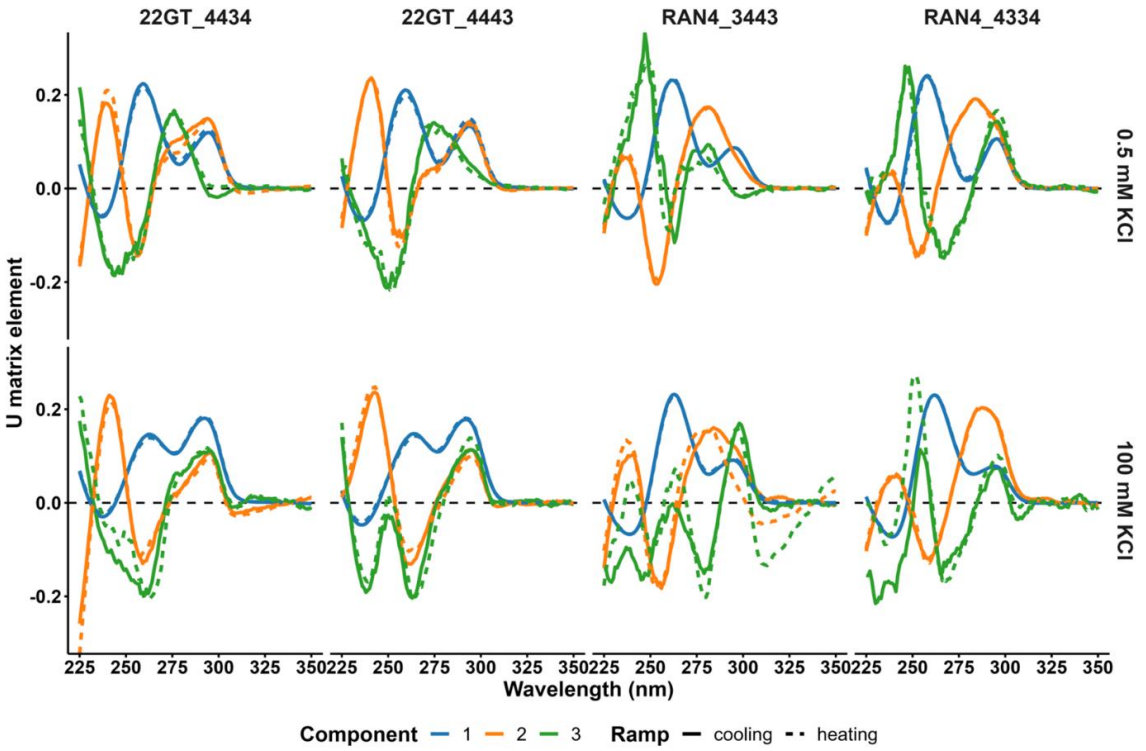


Figure S31 The U-matrices for the first three components, which contain the CD signatures of three putative G-quadruplex species. Sample conditions for the CD melting are: 10 µM DNA and either 100 mM KCl, 10 mM TMAA (pH 6.8) or 0.5 mM KCl. 100 mM TMAA (pH 6.8).

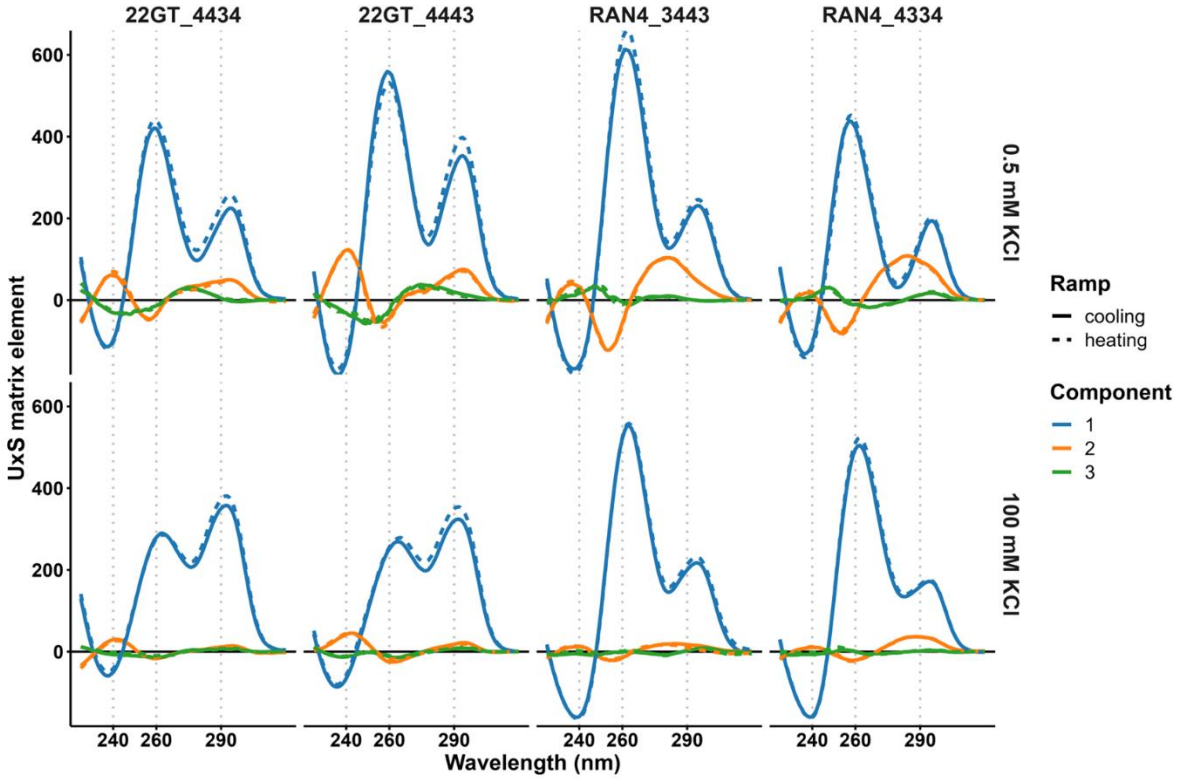


Figure S32 The U-matrix element of the singular value decomposition contains the CD signatures of the major (blue) and minor (orange, green) G-quadruplex species for each putative 3-tetrad G-quadruplex sequence in 0.5 and 100 mM K^+. Signatures are multiplied with the singular values (S) to highlight the statistical significance of each component.

CD spectra in ammonium and sodium

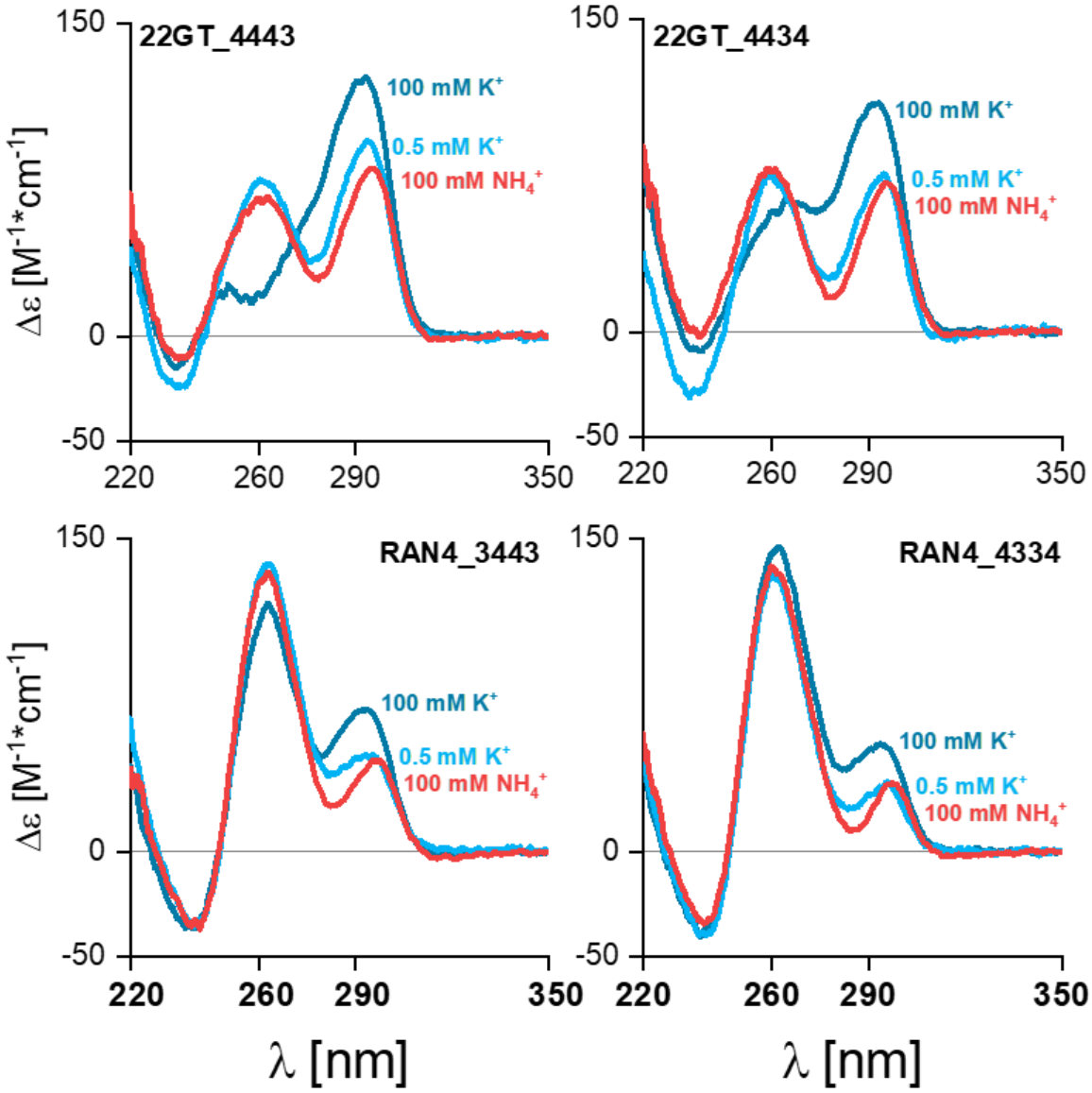


Figure S33 Comparing whether low K^+ concentration or switching from K^+ to NH_4^+ is better suited to replicate the G-quadruplex folding in 100 mM KCl for the sake of MS analysis. ***Dark blue****: 5 µM DNA, 100 mM KCl, 10 mM TMAA (pH 6.8).* ***Light blue****: 5 µM DNA, 0.5 mM KCl, 100 mM TMAA (pH 6.8).* ***Red****: 5 µM DNA, 100 mM ammonium acetate (pH 6.8). Results show that both actions yield a similar effect, but to minimize risk of structural uncertainty, we opted to measure in low K^+ concentration rather than switching to a different cation.*

Chapter 3

Foldamer synthesis and characterization

1. Materials and methods

1.1. Nuclear Magnetic Resonance

1D NMR spectra of oligomers were recorded on a Bruker Avance NEO NMR spectrometer (Bruker BioSpin) operating at 700,15 MHz for ^{1}H observation, equipped with a 5mm TXI probe with a gradient. All NMR experiments were performed at 273 K. Chemical shift values are given in ppm with reference to residual signals of solvent DMSO-d_6 (δ=2.50). All coupling constants (J) are given in Hertz and ^{1}H NMR splitting patterns with observed first order coupling are designated as singlet (s), broad singlet (brs), doublet (d), triplet (t) or multiplet (m).

1.2. High Performance Liquid Chromatography

HPLC analyses and purification were performed on a reverse phase C8 column on Jasco Extrema analytical and preparative systems. Mobile phases were composed of milli-Q water + 0.1% TFA (solvent A) and Acetonitrile + 0.1% TFA (solvent B). Analyses were done using the following gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B. Purifications were performed using the gradient 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 32 min: 0% A, 100% B – 35 min: 0% A, 100% B.

1.3. Mass spectrometry analyses

MS characterizations were performed on an Agilent Technologies 6230 TOF LC/MS spectrometer. The instrument is equipped with an ESI source and experiment were recorded in positive mode. The spray voltage was maintained at 3500 V and capillary temperature set at 300 °C. Samples were introduced by injection through a 20 µL sample loop into a 600 µL.min^{-1} flow of acetonitrile from the LC pump.

2. Methods for chemical synthesis

Commercial reagents were purchased from Sigma-Aldrich, Alfa-Aesar or TCI and used without further purification. Low-loading ProTide resin was purchased from CEM. Chloroform ($CHCl_3$), Triethylamine (TEA) and *N*,*N*-diisopropylethylamine (DIEA) were distilled over calcium hydride prior to use. Dry organic solvents: Tetrahydrofuran (THF) and Dichloromethane (DCM), used for solution and solid phase synthesis, were dispensed from a solvent purification system that passes solvents through packed column of dry neutral alumina. Milli-Q water was delivered from a PureLab Prima 7/15/20 system.

2.1. General method for oligomer synthesis

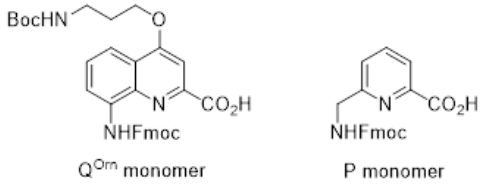


*Figure S34 Structures of monomers **Q**^{Orn} and **P**. These monomers were synthetized as Fmoc-N-protected and acid free form according to the reported procedure.*[5]

[5] Vallade M., Sai Reddy P., Fischer L., Huc I., Enhancing aromatic foldamer helix dynamics to probe interactions with protein surfaces, *Eur. J. Org. Chem*., **2018**, 5489.

Resin grafting: On *LL* ProTide resin, the first **Q^{Orn}** monomer (3 equiv.) was grafted using CsI (5 equiv.) and DIEA (6 equiv.) in dry DMF. Reaction mixture was vigorously shaking overnight. After reaction, the resin was filtered and washed three times with DMF and dichloromethane.

Fmoc deprotection: Grafted resin was washed twice with DMF, suspended in a 20% piperidine in DMF solution (4mL) and slowly stirred for 3 minutes. Resin was then filtered, washed twice with DMF and suspended again in a 20% piperidine in DMF solution and stirred for 7 minutes. The resin was then filtered and washed three times with DMF and dry THF.

***In-situ* coupling procedure**: *For coupling on the aromatic amines of the **Q^{Orn}** monomer.* Resin was suspended in dry THF and collidine (9 equiv.) was added. A solution of monomer (3 equiv.), PPh_3 (8 equiv.) and trichloroacetonitrile (TCAN, 9 equiv.) in dry $CHCl_3$ was added on the resin. The reaction was assisted by micro-waves (25 W, 50°C) for 15 minutes and repeated once. After reaction, the resin was filtered and washed with dry THF and DMF.

HBTU coupling: *For coupling on the aliphatic amines of the **P** monomer*. Resin was suspended in dry DMF. Monomer (3 equiv.) and HBTU (2.9 equiv.) as powder were added followed by DIEA (6 equiv.). The reaction was assisted by micro-waves (50 W, 50°C) for 10 minutes and repeated once. After reaction, the resin was filtered and washed with DMF.

Resin cleavage: Resin was washed three times with DMF and dichloromethane and was suspended in a solution of TFA/TIPS/H_2O 95:2.5:2.5 (*v/v/v*). The mixture was vigorously stirred for 4 hours. The resin was filtered, and the filtrate was evaporated under reduced pressure. The residual solid was suspended in Et_2O and centrifugated at 4°C for 5 minutes. Et_2O was removed and the yellow solid was dried under vacuum and then freeze-dried in water.

Preparative HPLC purifications: Crude compounds were purified using solvents A and B. The following gradient was used: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 22 min: 0% A, 100% B – 27 min: 0% A, 100% B. Collected fractions were analyzed by analytic HPLC and the relevant ones were combined and freeze-dried twice to remove the excess of TFA.

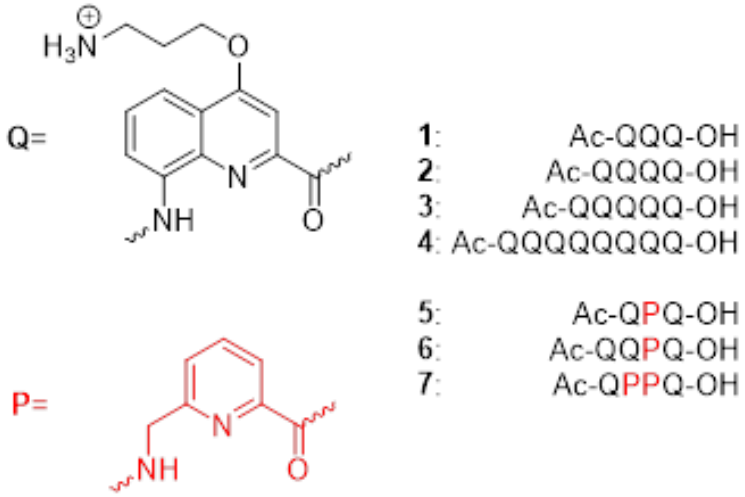


Figure S35 Synthesized oligomers

Synthesis of Q3,Q4,Q5 and Q8

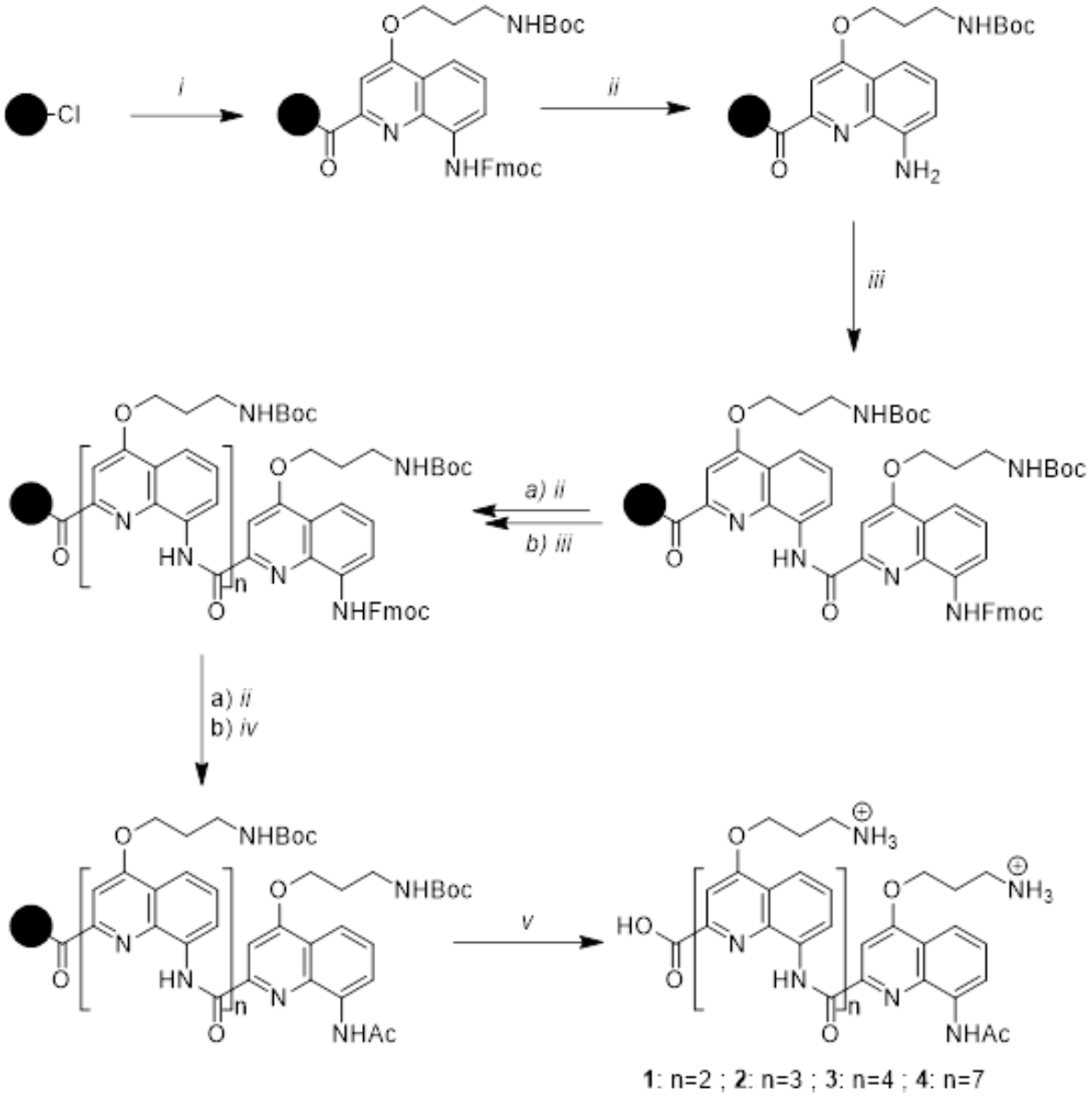


*Figure S36 Solid phase synthesis of oligomers **1** to **4**. i) **Q**^{Orn} monomer, CsI, DIEA, dry DMF. ii) 2 times, piperidine/DMF 2:8 (v/v), iii) 2 times, **Q**^{Orn} monomer, PPh_3, TCAN, collidine, THF/$CHCl_3$. These two last steps are repeated until the desired length is obtained. iv) 2 times, acetyl chloride, DIEA, THF. v) TFA/TIPS/H_2O 95:2.5:2.5 (v/v/v)*

Synthesis of QPQ

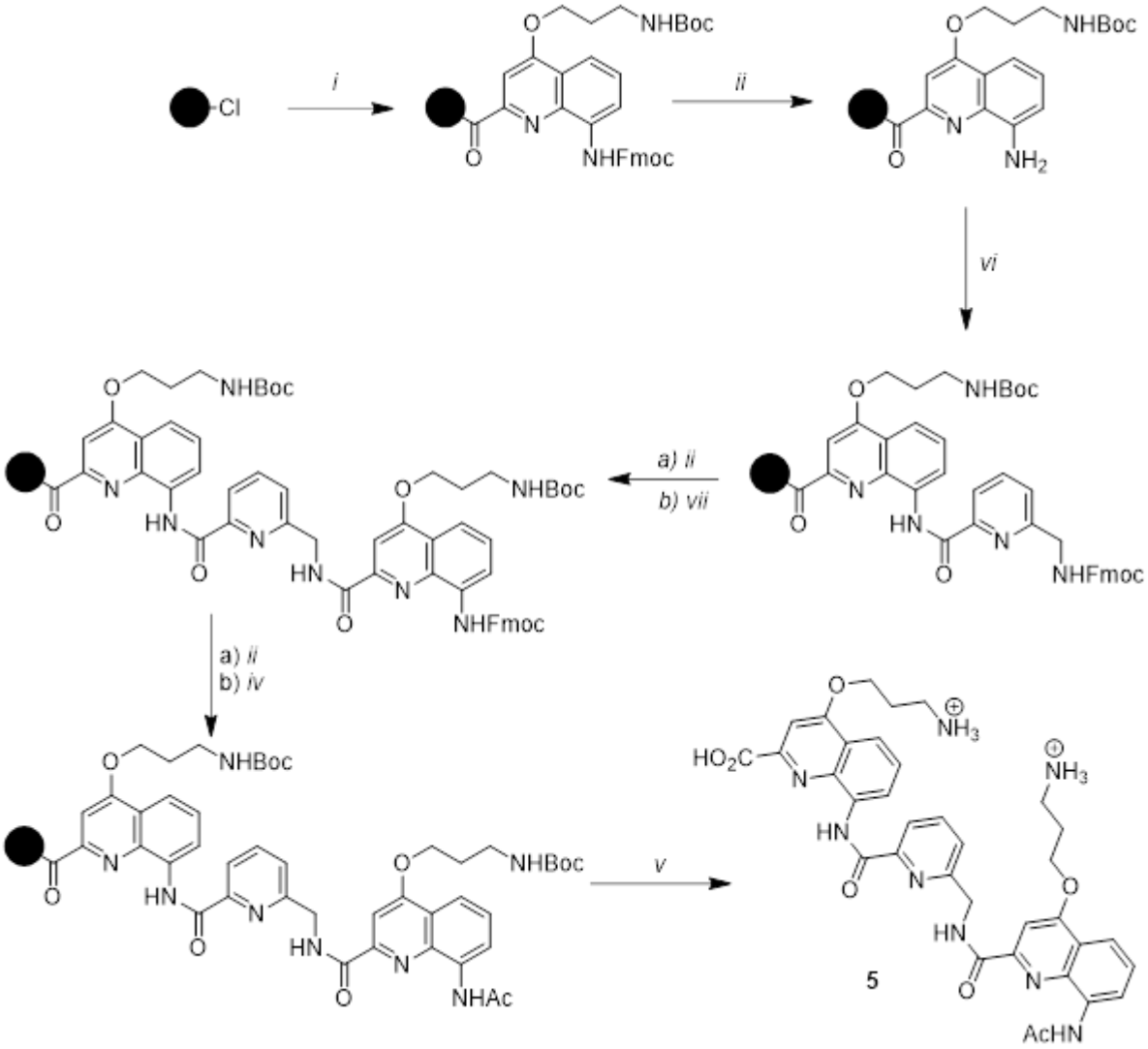


*Figure S37 Solid phase synthesis of oligomer **5**. vi) 2 times, **P** monomer, PPh_3, TCAN, collidine, THF/$CHCl_3$. vii) 2 times, $\mathbf{Q}^{Orn}$ monomer, HBTU, DIEA, dry DMF.*

Synthesis of QQPQ

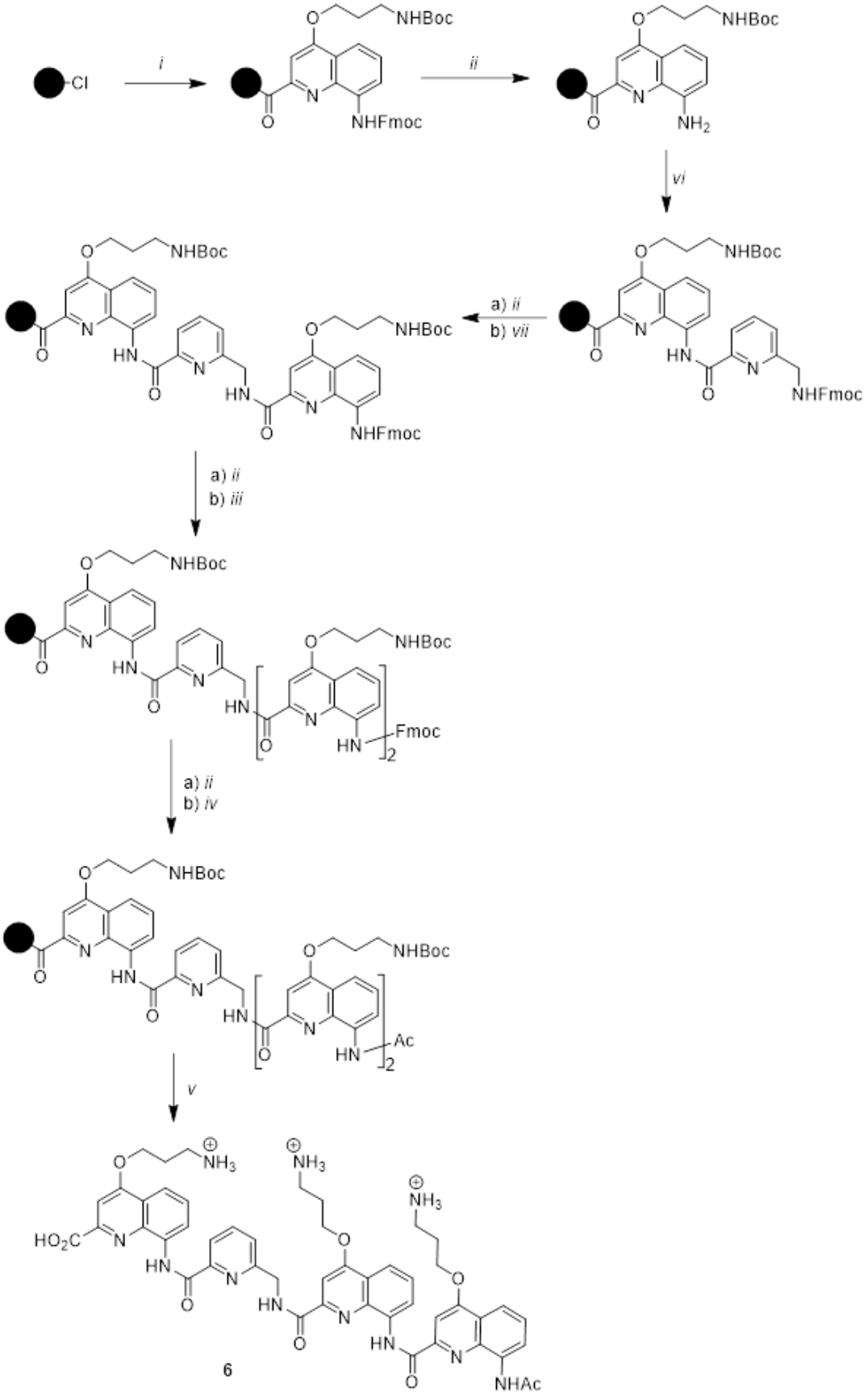


*Figure S38 Solid phase synthesis of oligomer **6**.*

Synthesis of QPPQ

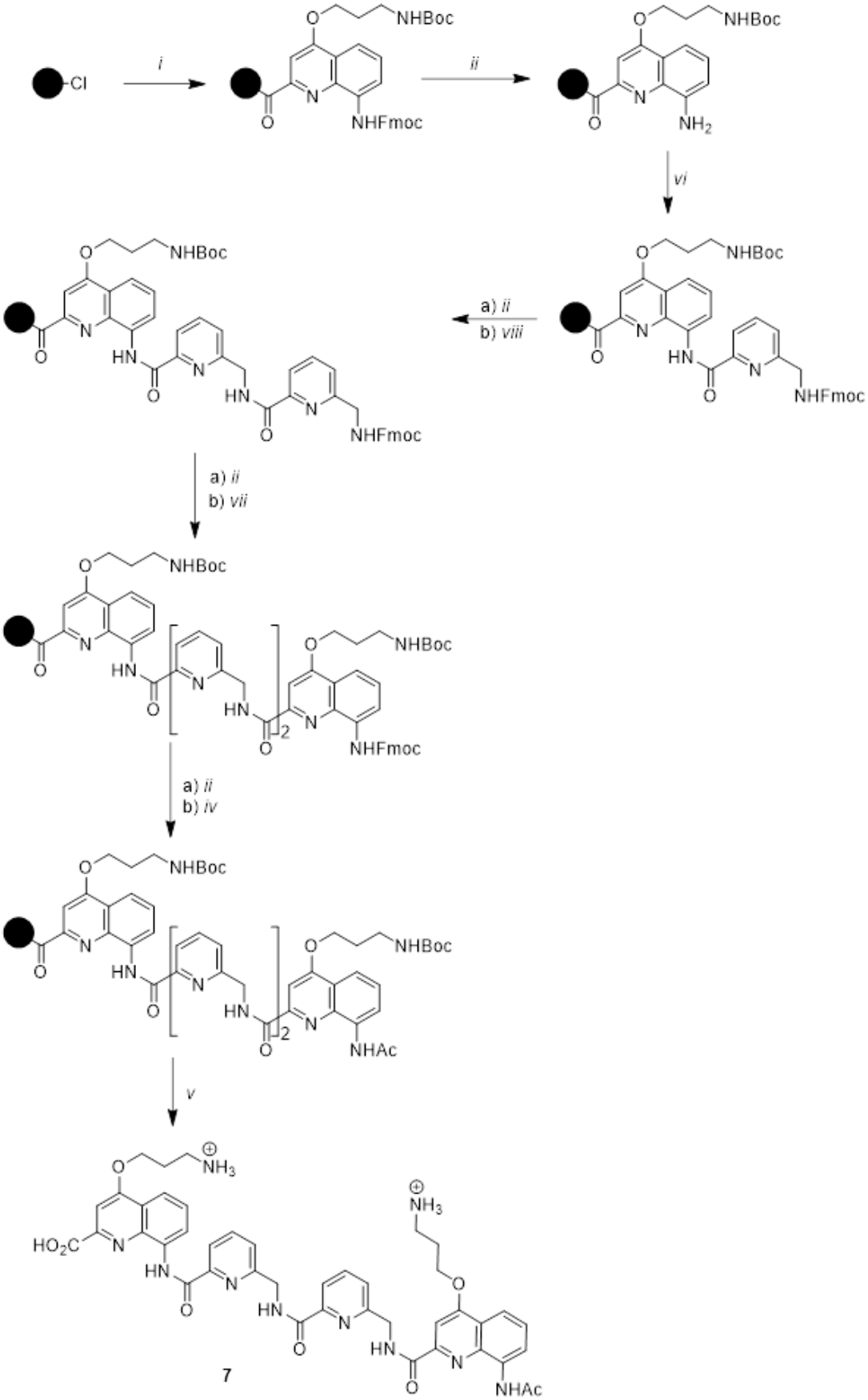
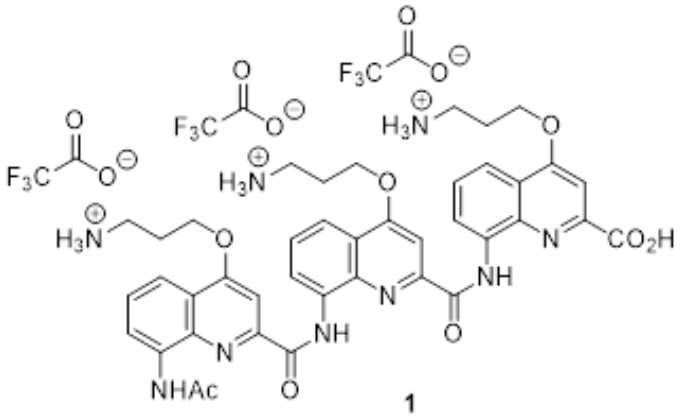


*Figure S39 Solid phase synthesis of oligomer **7**. viii) 2 times, **P** monomer, HBTU, DIEA, dry DMF.*

Characterization of Q_3



1

Trimer **1** was synthetized as a trifluoroacetate salt on 8 µmol scale following the *in-situ* activation procedure. The target compound was obtained as a light-yellow solid after purification by preparative HPLC (4 mg, 63% yield). ^{1}H NMR (700 MHz, DMSO-d6). δ 12.45 (brs, 1H), 12.27 (s, 1H), 12.16 (s, 1H), 9.23 (s, 1H), 8.91 (d, 1H, $^{3}J_{H-H}$=7.5 Hz), 8.85 (d, 1H, $^{3}J_{H-H}$=7.4 Hz), 8.03 (d, 1H, $^{3}J_{H-H}$=8.2 Hz), 7.76-7.87 (m, 9H), 7.71-7.74 (m, 2H), 7.66-7.70 (m, 2H), 7.28 (t, 1H, $^{3}J_{H-H}$=8.0 Hz), 6.58 (s, 1H), 4.57 (t, 2H, $^{3}J_{H-H}$=5.8 Hz), 4.54 (t, 2H, $^{3}J_{H-H}$=5.5 Hz), 4.12 (t, 2H, $^{3}J_{H-H}$=5.5 Hz), 3.04-3.16 (m, 6H), 2.21-2.27 (m, 4H), 2.15-2.20 (m, 2H), 1.68 (s, 2H). HRMS (ESI^{+}) m/z $[M+H]^{+}$ 790.3303 (calc. 790.3307 for $C_{41}H_{44}O_8N_9^{+}$).

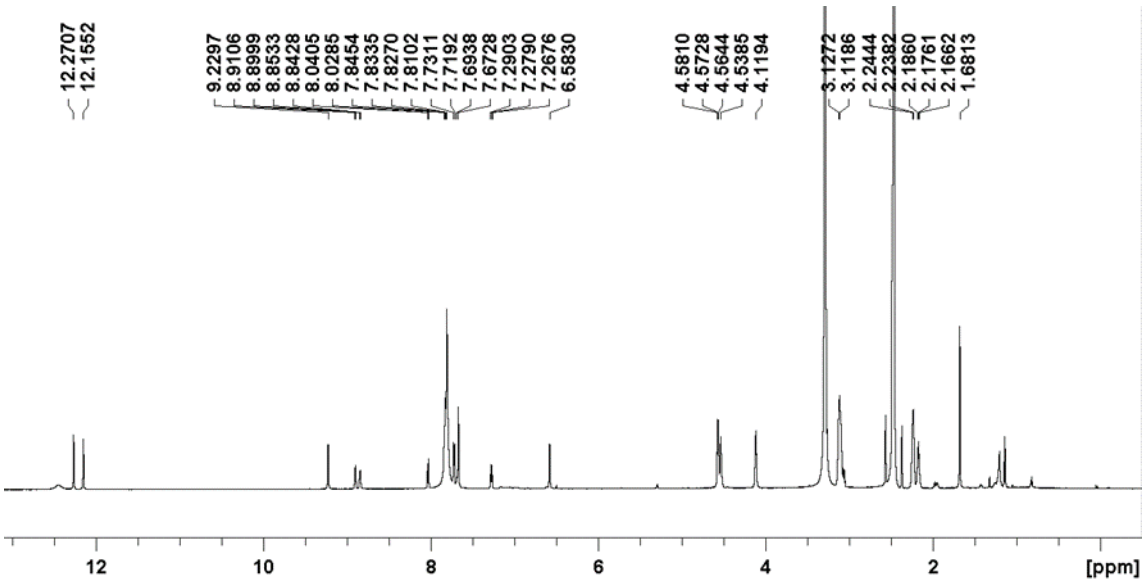


*Figure S40 ^{1}H NMR spectrum of oligomer **1**, measured in DMSO-d6 at 25°C.*

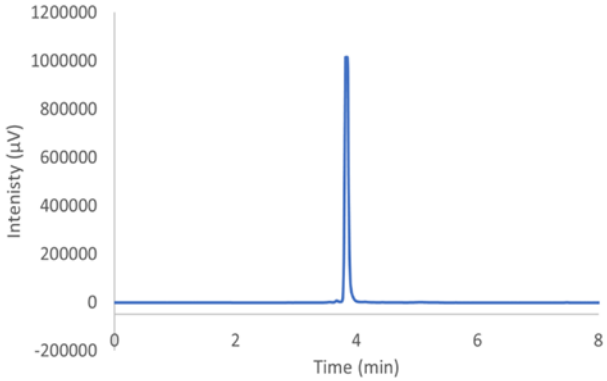
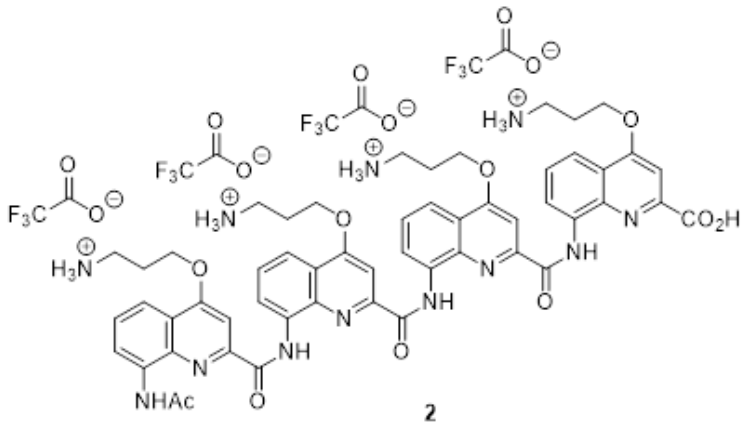


*Figure S41 HPLC trace of oligomer **1**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15min: 0% A, 100% B.*

Characterization of Q_4



2

Tetramer **2** was synthetized as a trifluoroacetate salt on 16 µmol scale following the *in-situ* activation procedure. The target compound was obtained as a light-yellow solid after purification by preparative HPLC (16 mg, 67% yield). ^{1}H NMR (700 MHz, DMSO-d6). δ 12.53 (brs, 1H), 12.20 (s, 1H), 11.77 (s, 1H), 11.69 (s, 1H), 9.08 (s, 1H), 8.97 (d, 1H, 3JH-H=7.3 Hz), 8.43 (brs, 1H), 7.81-7.99 (m, 12H), 7.75-7.80 (m, 2H), 7.74 (s, 1H), 7.44 (brs, 1H), 7.37 (t, 1H, 3JH-H=7.8 Hz), 7.26 (s, 1H), 6.75 (s, 1H), 6.66 (brs, 1H), 4.54-4.59 (m, 4H), 4.34 (brs, 1H), 4.23 (brs, 1H), 4.16 (brs, 2H), 3.07-3.22 (m, 6H), 2.20-2.46 (m, 8H), 1.69 (s, 3H). HRMS (ESI$^+$) m/z 1033,4316 $[M+H]^+$ (calc. 1033,4315 for $C_{54}H_{57}O_{10}N_{12}^+$).

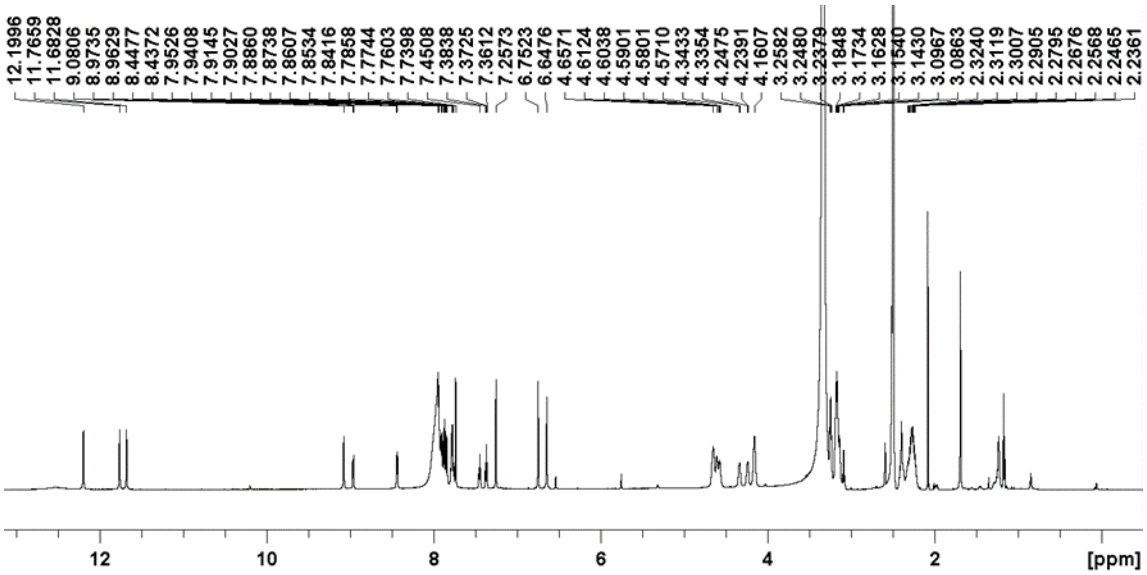


*Figure S42 ^{1}H NMR spectrum of tetramer **2**, measured in DMSO-d6 at 25°C on 700MHz Bruker Avance NEO spectrometer (TXI probe (^{1}H, ^{13}C, ^{15}N, ^{2}H), 5mm, z-gradients).*

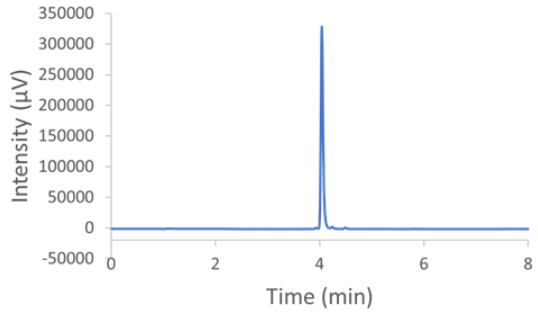
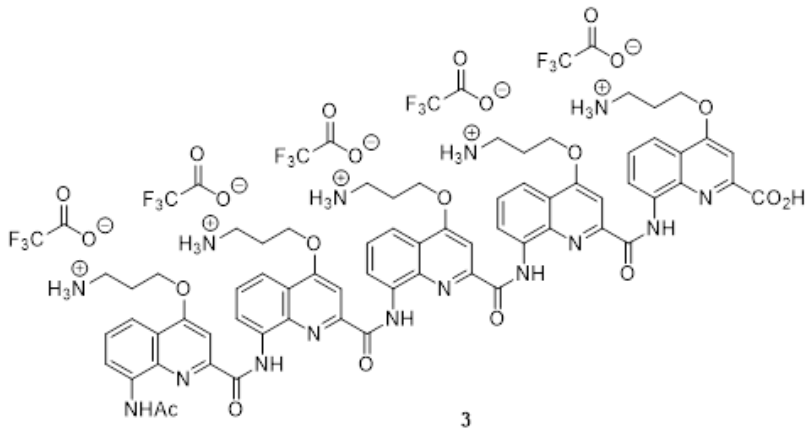


*Figure S43 HPLC trace of oligomer **2**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B.*

Characterization of Q_5



Pentamer **3** was synthetized as a trifluoroacetate salt on 10 µmol scale following the *in-situ* activation procedure. The target compound was obtained as a light-yellow solid after purification by preparative HPLC (7.5 mg, 59% yield). ^{1}H NMR (700 MHz, DMSO-d6). δ 12.21 (brs, 1H), 11.74 (s, 1H), 11.68 (s, 1H), 11.64 (s, 1H), 11.52 (s, 1H), 8.87 (s, 1H), 8.50 (d, 2H, $^{3}J_{H-H}$), 7.4 Hz), 7.88-8.06 (m, 17H), 7.78-7.86 (m, 3H), 7.75 (d, 1H, $^{3}J_{H-H}$=8.2 Hz), 7.67 (d, 1H, $^{3}J_{H-H}$=7.3 Hz), 7.51 (t, 1H, $^{3}J_{H-H}$=7.8 Hz), 7.38 (t, 1H, $^{3}J_{H-H}$=7.8 Hz), 7.33 (t, 1H, $^{3}J_{H-H}$=7.8 Hz), 7.30 (s, 1H), 7.20 (s, 1H), 6.79 (s, 1H), 6.67 (s,

1H), 6.52 (s, 1H), 4.67-4.73 (m, 2H), 3.08-3.28 (m, 13H), 2.17-2.46 (m, 13H), 1.36 (s, 2H). HRMS (ESI$^+$) m/z $[M+H]^+$ 1276.5369 (calc. 1276.5323 for $C_{67}H_{70}O_{12}N_{15}^+$).

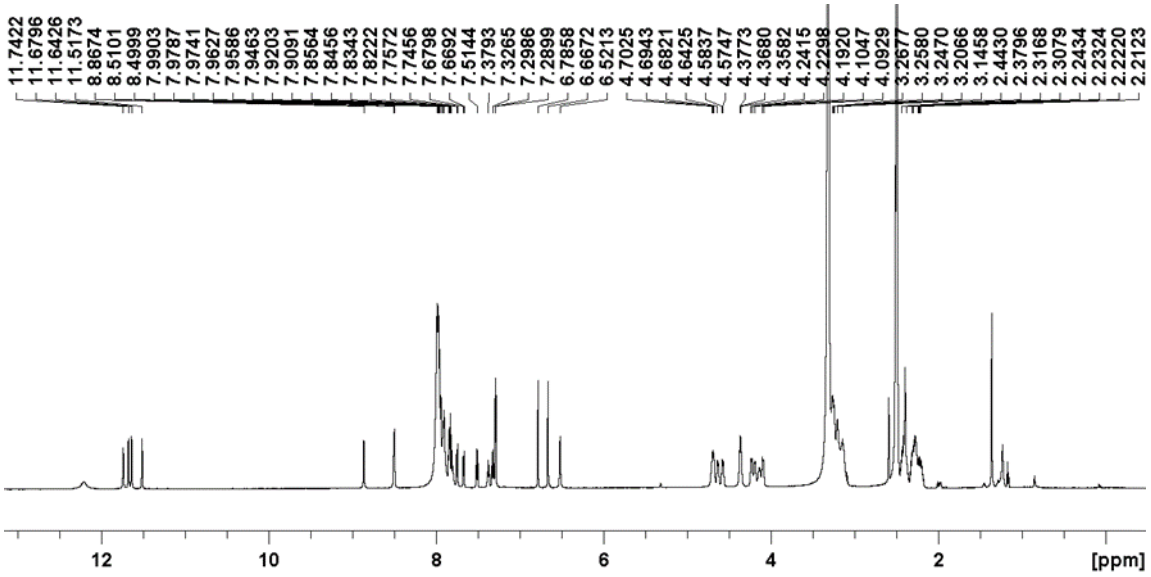


*Figure S44 ^{1}H NMR spectrum of pentamer **3**, measured in DMSO-d6 at 25°C on 700MHz Bruker Avance NEO spectrometer (TXI probe (^{1}H, ^{13}C, ^{15}N, ^{2}H), 5mm, z-gradients).*

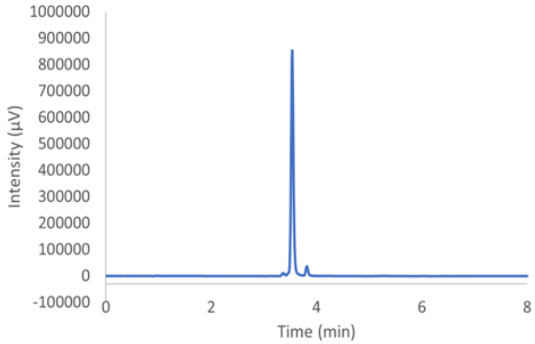
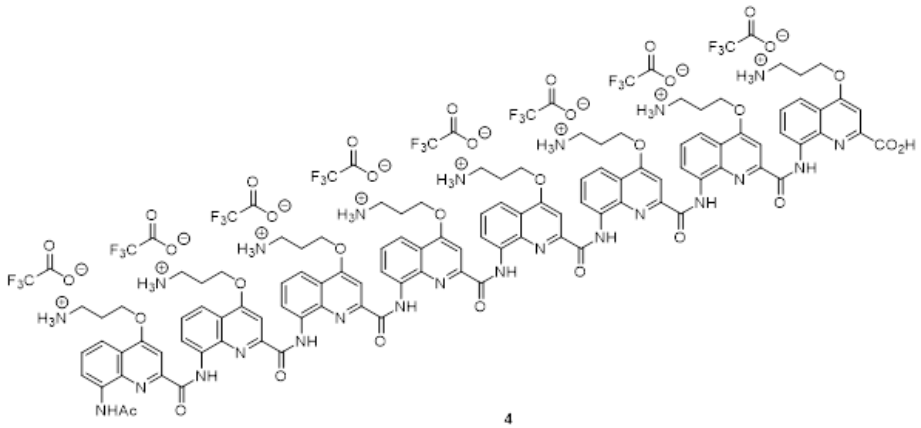


*Figure S45 HPLC trace of oligomer **3**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B.*

Characterization of Q_8



4

Octamer **4** was synthetized as a trifluoroacetate salt on 10 µmol scale following the *in-situ* activation procedure. The target compound was obtained as a light-yellow solid after purification by preparative HPLC (11.5 mg, 57% yield). ^{1}H NMR (700 MHz, DMSO-d6). δ 11.93 (brs, 1H), 11.28 (s, 1H), 11.25 (s, 1H), 11.14 (s, 1H), 11.04 (s, 1H), 11.03 (s, 1H), 10.94 (s, 1H), 10.88 (s, 1H), 8.91 (brs, 1H), 8.44 (s, 1H), 8.07-8.27 (m,

11H), 7.83-8.04 (m, 12H), 7.74-7.81 (m, 3H), 7.67 (d, 2H, $^{3}J_{H\text{-}H}$=7.9 Hz), 7.44-7.51 (m, 3H), 7.31-7.43 (m, 4H), 7.17 (t, 1H, $^{3}J_{H\text{-}H}$=8.3 Hz), 7.12 (t, 1H, $^{3}J_{H\text{-}H}$=8.1 Hz), 7.02 (d, 1H, $^{3}J_{H\text{-}H}$=7.3Hz), 6.96 (s, 1H), 6.85 (s, 1H), 6.58 (s, 1H), 6.53 (s, 1H), 6.36 (s, 1H), 6.30 (s, 1H), 6.17 (s, 1H), 5.91 (s, 1H), 4.36-4.51 (m, 5H), 4.12-4.34 (m, 9H), 3.97-4.07 (m, 3H), 3.85-3.90 (m, 1H), 2.96-3.19 (m, 16H), 1.95-2.35 (m, 15H), 1.24 (s, 3H). HRMS (ESI^{+}) m/z $[M+H]^{+}$ 2005.8351 (calc. 2005.8346 for $C_{106}H_{109}O_{18}N_{24}^{+}$).

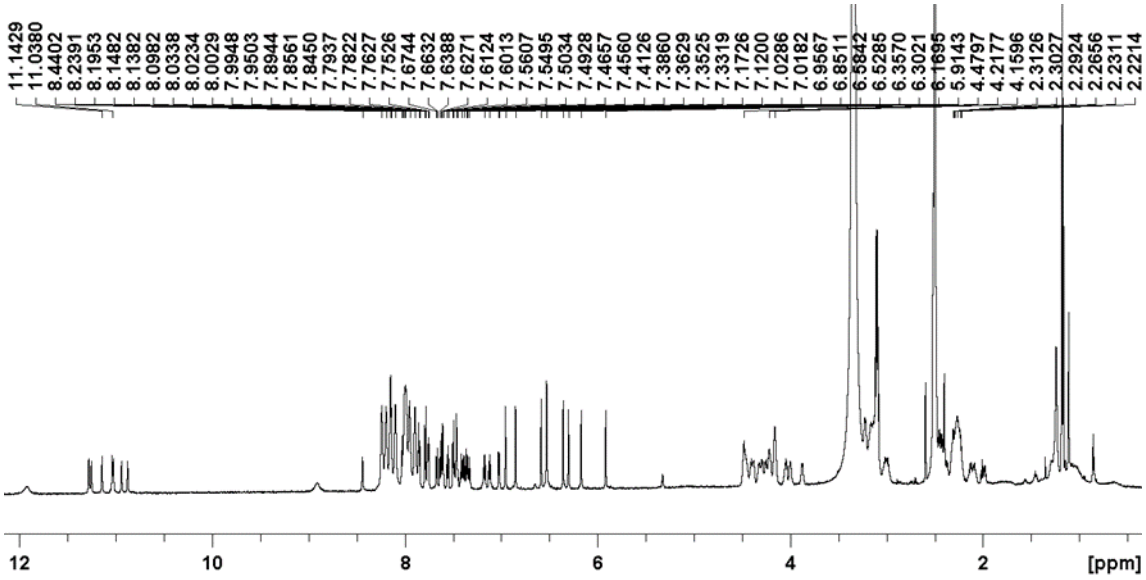


*Figure S46 ^{1}H NMR spectrum of octamer **4**, measured in DMSO-d6 at 25°C on 700MHz Bruker Avance NEO spectrometer (TXI probe (^{1}H, ^{13}C, ^{15}N, ^{2}H), 5mm, z-gradients).*

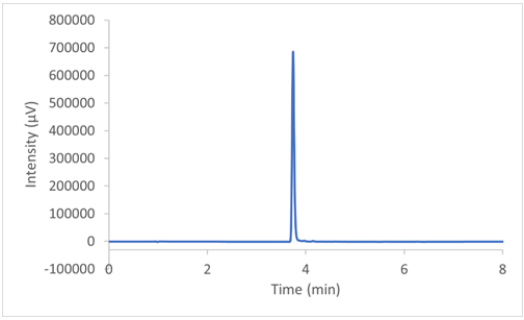
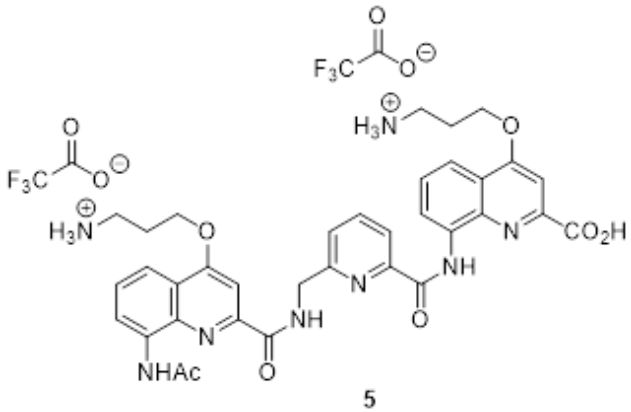


*Figure S47 HPLC trace of oligomer **4**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B.*

Characterization of QPQ



5

Trimer **5** was synthetized as a trifluoroacetate salt on 16 µmol scale following *in-situ* activation procedure for coupling on **Q^{Orn}** monomer and HBTU coupling procedure for

coupling on **P** monomer. The target compound was obtained as a white solid after purification by preparative HPLC (7mg, 48% yield). ^{1}H NMR (700 MHz, DMSO-d6). δ 13.61 (brs, 1H), 12.60 (s, 1H), 10.35 (t, 1H, 3JH-H=6.2 Hz), 10.25 (s, 1H), 8.91 (d, 1H, 3JH-H=7.2 Hz), 8.74 (d, 1H, 3JH-H=7.7 Hz), 8.08-8.17 (m, 2H), 7.69-7.97 (m, 8H), 7.60-7.67 (m, 3H), 6.55 (s, 1H), 4.99 (d, 2H, 3JH-H=6.1 Hz), 4.43-4.50 (m, 4H), 3.06-3.15 (m, 5H), 2.29 (s, 2H), 2.16-2.25 (m, 4H), 1.24 (s, 3H). HRMS (ESI^{+}) m/z $[M+H]^{+}$ 681.2777 (calc. 681.2780 for $C_{35}H_{37}O_7N_8^{+}$).

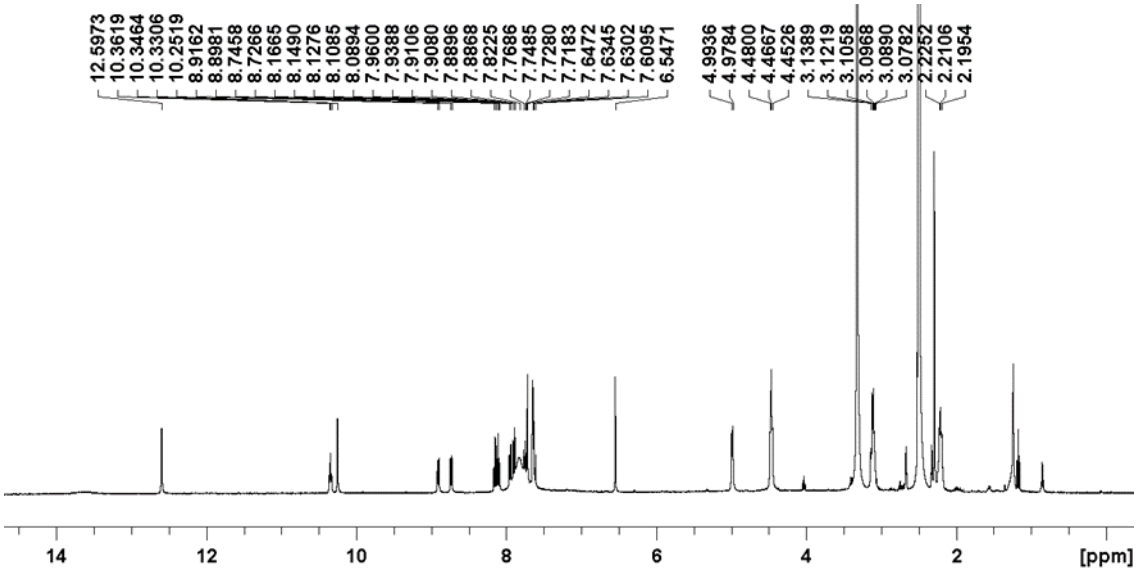


*Figure S48 ^{1}H NMR spectrum of trimer **5**, measured in DMSO-d6 at 25°C on 700MHz Bruker Avance NEO spectrometer (TXI probe (^{1}H, ^{13}C, ^{15}N, ^{2}H), 5mm, z-gradients).*

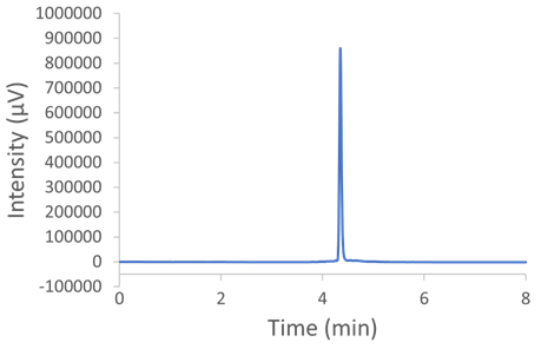
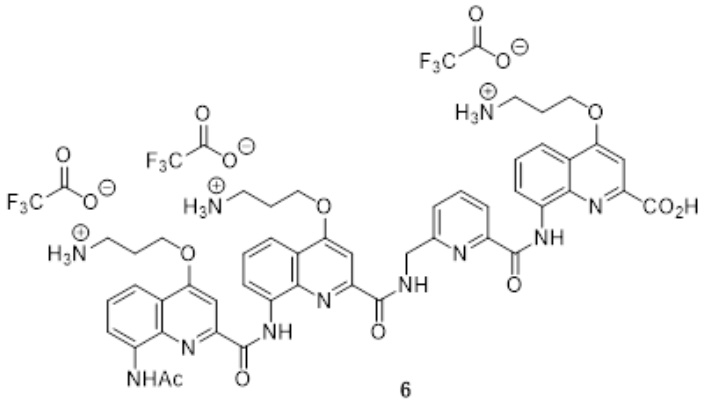


*Figure S49 HPLC trace of oligomer **5**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B.*

Characterization of QQPQ



6

Tetramer **6** was synthetized as a trifluoroacetate salt on 16 µmol scale following *in-situ* activation procedure for coupling on **Q^{Orn}** monomer and HBTU coupling procedure for coupling on **P** monomer. The target compound was obtained as a white solid after purification by preparative HPLC (13 mg, 64% yield). ^{1}H NMR (700 MHz, DMSO-d6). δ 11.82 (brs, 1H), 11.56 (s, 1H), 9.52 (brs, 1H), 9.30 (s, 1H), 8.31 (d, 1H, $^{3}J_{H-H}$=7.4 Hz), 7.77-8.14 (m, 10H), 7.61-7.71 (m, 2H), 7.75 (t, 1H, $^{3}J_{H-H}$=7.9 Hz), 7.43 (t, 1H, 3JH-H=8.0 Hz), 7.31 (d, 1H, $^{3}J_{H-H}$=7.4 Hz), 7.11 (s, 1H), 6.74 (s, 1H), 6.57 (s, 1H), 5.72 (s, 1H), 4.78 (brs, 2H), 4.53-4.56 (m, 2H), 4.48-4.52 (m, 2H), 4.18-4.22 (m, 2H), 3.06-3.25 (m, 7H), 2.16-2.38 (m, 5H). HRMS (ESI^{+}) m/z [M+H]$^{+}$ 924.3048 (calc. 924.3787 for $C_{48}H_{50}O_9N_{11}^{+}$).

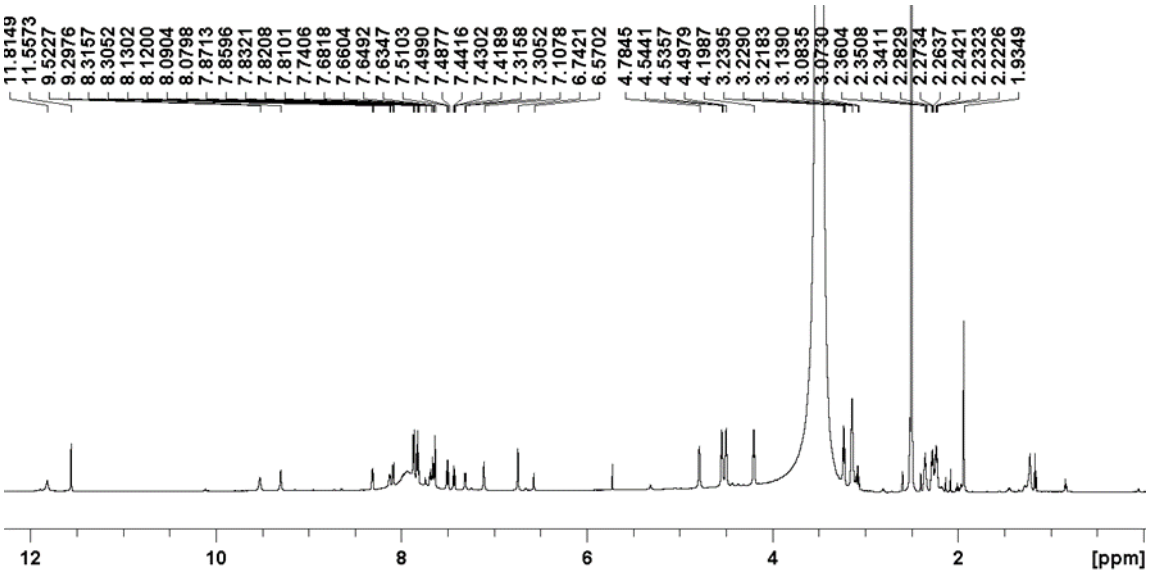


*Figure S50 ^{1}H NMR spectrum of tetramer **6**, measured in DMSO-d6 at 25°C on 700MHz Bruker Avance NEO spectrometer (TXI probe (^{1}H, ^{13}C, ^{15}N, ^{2}H), 5mm, z-gradients).*

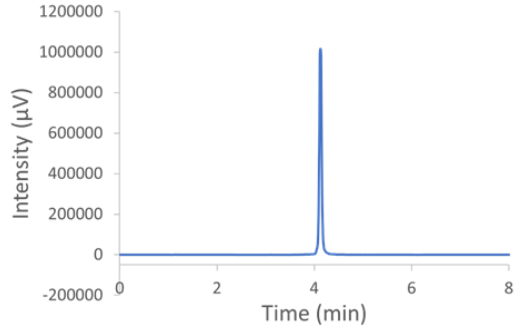
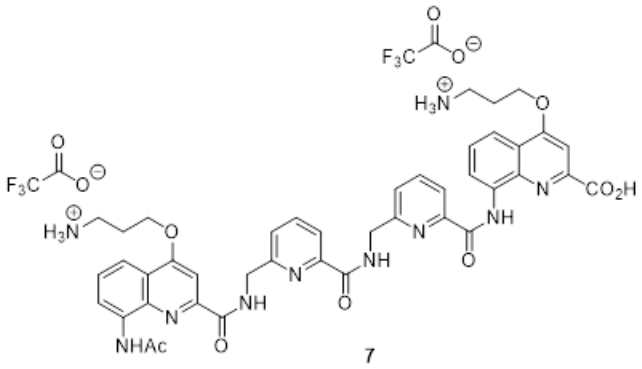


*Figure S51 HPLC trace of oligomer **6**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B.*

Characterization of QPPQ



7

Tetramer **7** was synthetized as a trifluoroacetate salt on 16 µmol scale following *in-situ* activation procedure for coupling on **Q^{Orn}** monomer and HBTU coupling procedure for coupling on **P** monomer. The target compound was obtained as a white solid after purification by preparative HPLC (11 mg, 66% yield). ^{1}H NMR (300 MHz, DMSO-d6). δ 12.60 (s, 1H), 10.09 (brs, 1H), 9.61 (brs, 1H), 8.91 (d, 1H, $^{3}J_{H-H}$=7.7 Hz),7.92-8.15 (m, 7H), 7.64-7.82 (m, 9H), 7.48-7.60 (m, 4H), 7.21-7.36 (m, 4H), 6.84 (d, 1H, $^{3}J_{H-H}$=7.2 Hz), 4.87 (d, 2H, $^{3}J_{H-H}$=6.2 Hz), 4.73 (d, 2H, $^{3}J_{H-H}$=6.1 Hz), 4.43-4.52 (m, 2H), 4.34-4.42 (m, 3H), 2.96-3.15 (m, 6H), 1.50-2.69 (m, 8H). HRMS (ESI^{+}) m/z [M+H]$^{+}$ 815.3253 (calc. 815.3260 for $C_{42}H_{43}O_8N_{10}^{+}$).

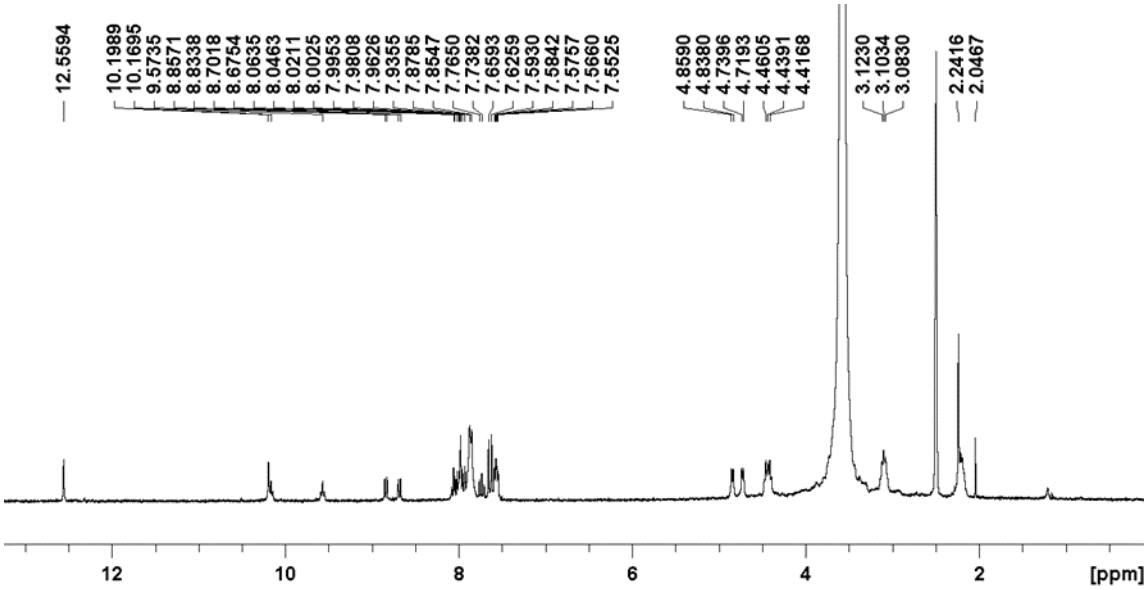


*Figure S52 ^{1}H NMR spectrum of tetramer **7**, measured in DMSO-d6 at 25°C on 300MHz Bruker Avance NEO spectrometer (TXI probe (^{1}H, ^{13}C, ^{15}N, ^{2}H), 5mm, z-gradients).*

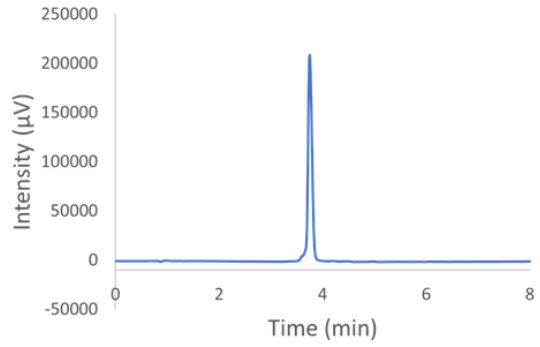


*Figure S53 HPLC trace of oligomer **7**. Analysis was performed using solvents A and B with the gradient: 0 min: 100% A, 0% B – 2 min: 100% A, 0% B – 12 min: 0% A, 100% B – 15 min: 0% A, 100% B.*

Desalting procedure

Traces of Na^+ in the oligonucleotide stock give rise to visible Na^+ adducts in the mass spectrum, causing loss of resolution and S/N ratio. We therefore strip our stock solutions off of Na^+ using centrifugal filters (Amicon Ultracel 3K, Millipore). The filters contain a cellulose matrix that will hold back the DNA macromolecule, while allowing the washing solution to pass through via centrifugation.

Prior to desalting, the oligonucleotides are dissolved in water and annealed at 85°C for 2-3 min. Then, the first segment of the desalting process is exchanging the non-volatile Na^+ ion with volatile NH_4^+. The filter unit is filled with a washing solution of 500 mM Ammonium acetate and placed in the centrifuge for 15 minutes at 15,000 rpm. This process is repeated four times. The second segment is diluting with pure water to gradually flush out the NH_4^+ ions. Some NH_4^+ ions will remain electrostatically bound to the DNA strand, but they will detach during the MS ionization process, given their volatility. This washing process is repeated six times. The third segment is the recovery of the desalted DNA stock solution by placing the filter upside-down in a fresh tube and pushing it out at 1,000 rpm for 3 minutes.

A proof of concept for the desalting method is illustrated in Figure S54.

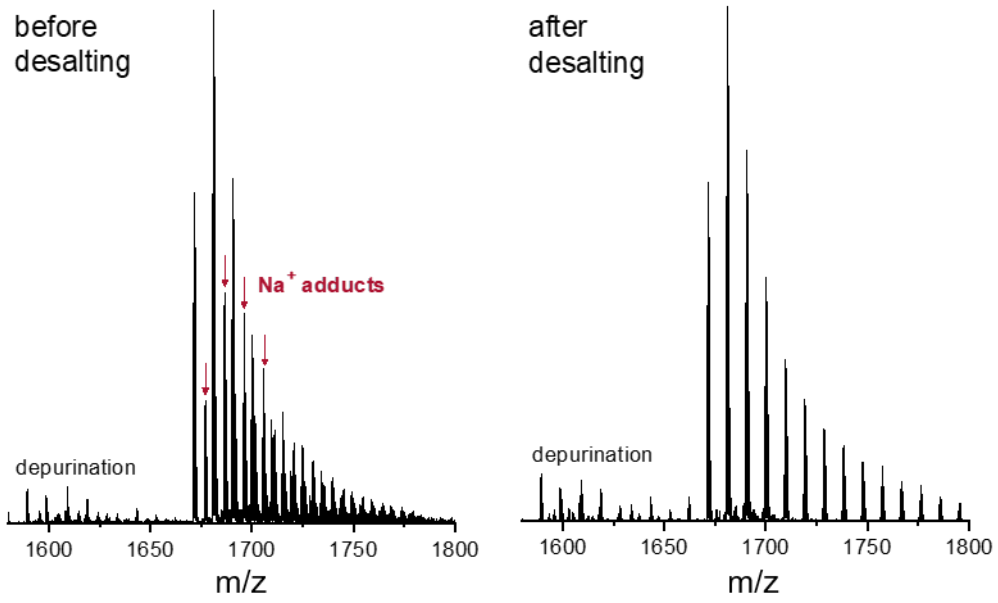


*Figure S54 Mass spectrum of DNA sequence 21G (d**G₃**TTA**G₃**TTA**G₃**TTA**G₃**) before (left) and after (right) desalting, zoomed in on the 4- charge state. Sample conditions: 10 µM DNA, 1 mM KCl, 100 mM TMAA in H_2O.*

CD/UV-melting curves of mutated G-quadruplex sequences

This section is for the mutant sequences 222T-mA, 222T-mC, 26CEB-mT, 22GT-18T and 24TTG-20T that were introduced into the foldamer screening panel to probe the foldamers' sensitivity to loop-induced G4 topology switches. 22GT-18T and 24TTG-20T are telomeric repeats where the third loop has a TTA to TTT mutation. Removing the adenine disrupts the formation of AGA triads and induces a topology switch.[59] 222T-mA, 222T-mC and 26CEB-mT are mutated in the middle loop. The effect of mutating the middle loop is unknown for these sequences. We also provide spectra for 'bcl2' since we were unsure of its G-quadruplex stability in low K^+ concentration.

For our six undocumented sequences we provide CD spectra and UV melting curves. One dataset in 1 mM KCl, to be orthogonal with our previously established database.[23] Another dataset in 0.5 mM KCl, which is the concentration for the ligand screening. In second approach, this dataset gives insight for how much cutting the K^+ concentration in half affects G-quadruplex stability.

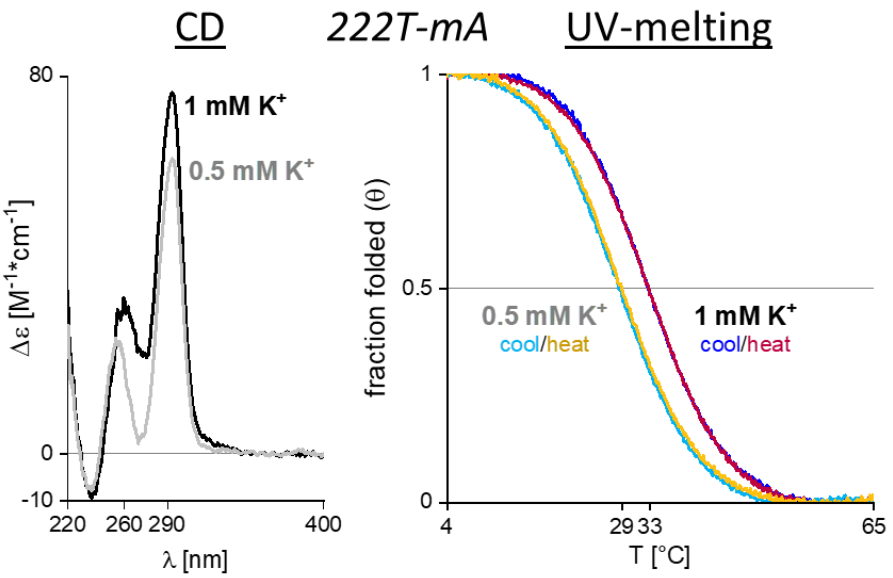


Figure S55 ***Left****: CD spectra of 222T-mA (dT**G₃**TT**G₃**AA**G₃**TT**G₃**T) in 1 mM (black) and 0.5 mM KCl (grey).* ***Right****: UV-melting curves in 1 mM (blue/red) and 0.5 mM (cyan/orange) KCl. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8).*

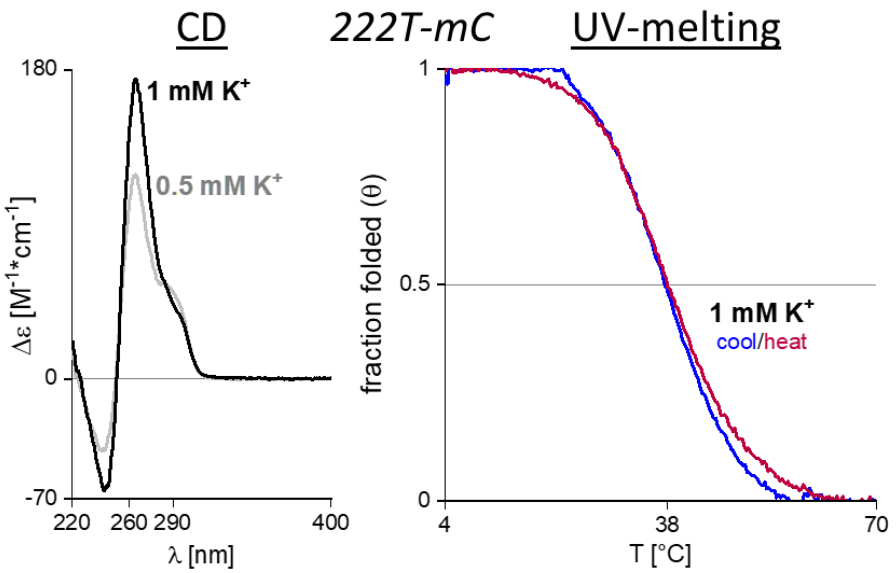


Figure S56 ***Left****: CD spectra of 222T-mC (dT**G₃**TT**G₃**CC**G₃**TT**G₃**T) in 1 mM (black) and 0.5 mM KCl (grey).* ***Right****: UV-melting curve in 1 mM (blue/red) KCl. No data could be obtained in 0.5 mM KCl. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8).*

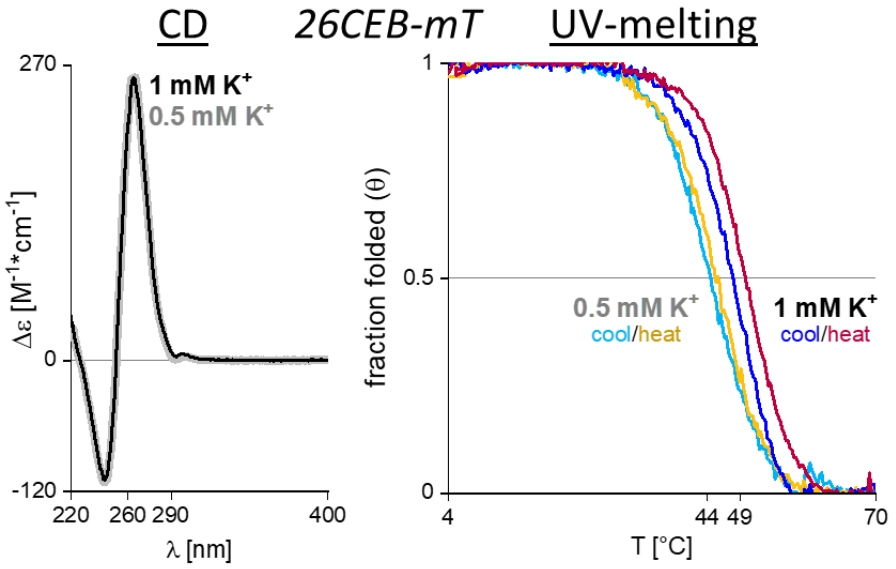


*Figure S57 CD spectra of 26CEB-mT (dAA**G₃**T**G₃**TTTTTTT**GT**G₃**T**G₃**T) in 1 mM (black) and 0.5 mM KCl (grey).* ***Right****: UV-melting curve in 1 mM (blue/red) KCl. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8).*

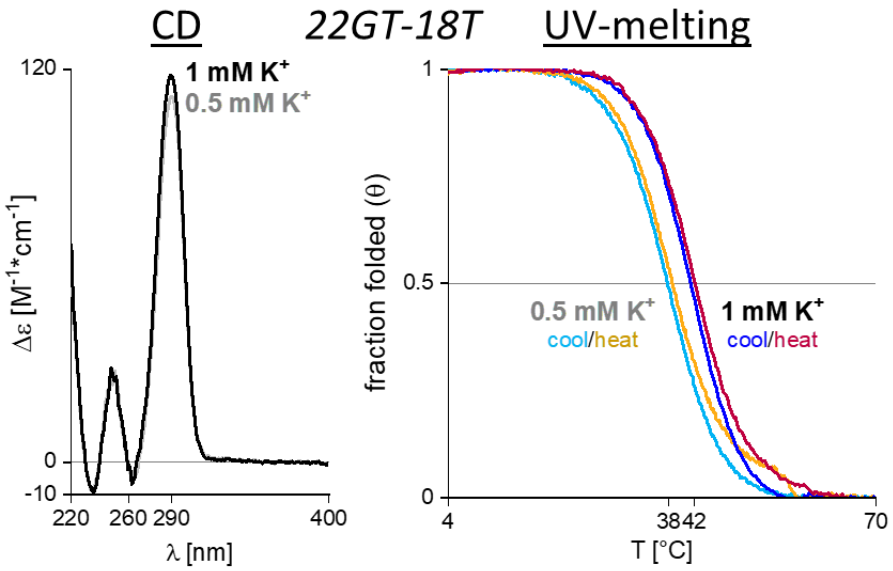


*Figure S58 CD spectra of 22GT-18T (d**G₃**TTA**G₃**TTA**G₃**<u>TTT</u>**G₃**T) in 1 mM (black) and 0.5 mM KCl (grey). **Right**: UV-melting curves in 1 mM (blue/red) and 0.5 mM (cyan/orange) KCl. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8).*

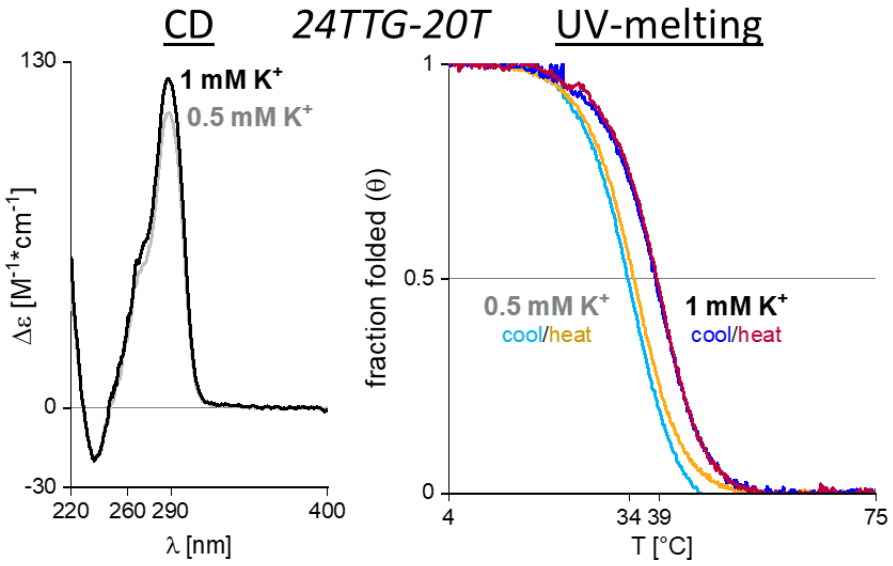


*Figure S59 CD spectra of 24TTG-20T (dTT**G₃**TTA**G₃**TTA**G₃**<u>TTT</u>**G₃**A) in 1 mM (black) and 0.5 mM KCl (grey). **Right**: UV-melting curves in 1 mM (blue/red) and 0.5 mM (cyan/orange) KCl. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8).*

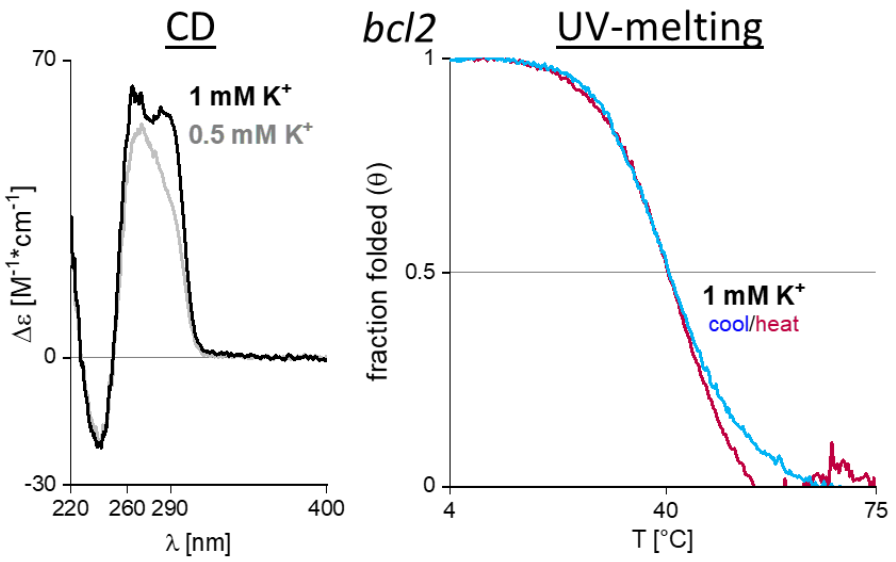


*Figure S60 CD spectra of bcl2 (d**G₃**C**G**C**G₃**A**GG**AATT**G₃**C**G₃**) in 1 mM (black) and 0.5 mM KCl (grey). **Right**: UV-melting curve in 1 mM (blue/red) KCl. No data could be obtained in 0.5 mM KCl due to lack of a low temperature baseline. Samples contain 10 µM DNA and 100 mM TMAA (pH 6.8).*

Figure S61 lets us directly compare the CD signatures of the mutated sequences with their 'wild-type' versions.

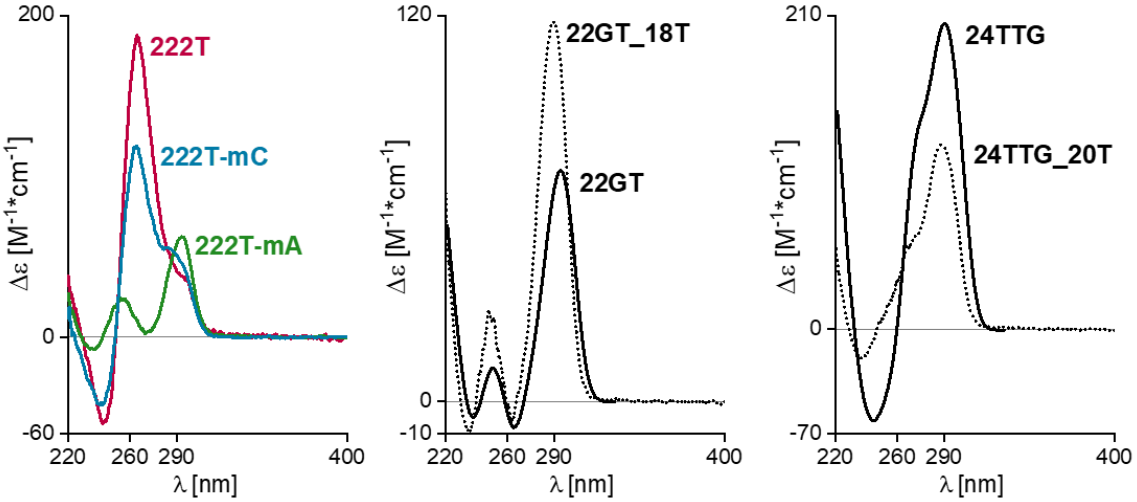


*Figure S61 Comparing CD signatures of 'wild-type' and mutant sequences. **Left**: The 222T sequence motif (dT**G₃**TT**G₃**TT**G₃**TT**G₃**T) in 0.5 mM KCl **Middle**: the 22GT sequence motif (d**G₃**TTA**G₃**TTA**G₃**TTA**G₃**T) in 1 mM KCl **Right**: The 24TTG sequence motif (dTT**G₃**TTA**G₃**TTA**G₃**TTA**G₃**A) in 1 mM KCl. Samples contain 10 µM DNA, 0.5 or 1 mM KCl and 100 mM TMAA (pH 6.8). 'wild-type' signatures of 22GT and 24TTG are taken from previously published data.[23]*

CD curves of multi-stranded G-quadruplexes in NH_4^+ and K^+

The screening panel contains tetramolecular G-quadruplex $(dTG_4T)_4$ and bimolecular G-quadruplex $(dG_4T_4G_4)_2$. We know that these quadruplexes take weeks/months to fold completely in solution. We expected the formation to be notoriously slow in 0.5 mM K^+, which is why we measured another sample in 150 mM NH_4^+, hoping to speed up the formation kinetics and obtain quantitative amounts of G-quadruplex.

To our surprise, we noticed slight changes in ligand binding when the cation was switched (see Figure 64). We gathered CD spectra of both sequences in K^+ and NH_4^+ (Figure S62) after having let the samples rest for 10/11 months. The CD signatures do not change enough to indicate that switching from K^+ to NH_4^+ induces a change in topology. Since the true concentration of folded G-quadruplex is uncertain, we did not convert the ellipticity to Δε.

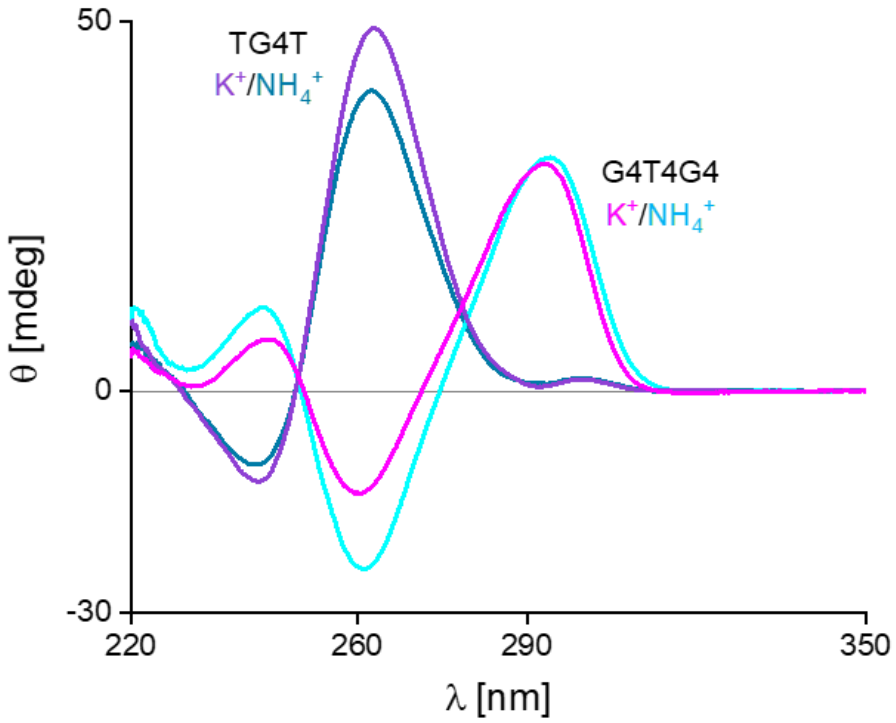


Figure S62 CD spectra of multi-stranded G-quadruplexes TG4T and G4T4G4 in different cation buffers. Samples contain: 10 µM G4T4G4/20 µM TG4T, 150 mM ammonium acetate (pH 6.8) or 0.5 mM KCl and 100 mM TMAA (pH 6.8).

Parameters of the ESI-IMS-QTOF instrument

Table S1 Instrumental settings of the Agilent 6560 IMS-QTOF, optimized for native MS on oligonucleotides.

AREA	PARAMETER	VALUE	UNIT	AREA	PARAMETER	VALUE	UNIT
Cell	Col. Cell Flow	20	psig	Ion Funnel	Funnel RF HP	200	Volts
	Col. Cell Energy	0	Voltage		Funnel Delta LP	100	Volts
	Cell Exit	12.1	Volts		Funnel Delta HP	150	Volts
	Cell Entrance	-20	Volts		Funnel DC	50	Volts
	Hex dV	3	Volts	Optics 1	Lens 2	-16.5	Volts
	Hex DC Entrance	-20	Volts		Lens 1	-23	Volts
	Hex RF	550	Volts		Oct1 DC	-25	Volts
	Col. Cell Gas	Nitrogen		Optics 2	Bot Slit	41.35	Volts
IM	Trap Funnel Pressure	3.725	Torr		Top Slit	41.2	Volts
IM Drift Tube	DT Voltage	-210	Volts		Vertical Q	-12.9	Volts
	DT Entrance Voltage	-650	Volts		Horizontal Q	1.5	Volts
	DT Pressure	3.895	Torr		Slicer	10	Volts
	DT Temperature	24.3	°C		Ion Focus	-10	Volts
IM Front Funnel	HP Funnel Pressure	3.1	Torr		Oct 2 RF Vpp	600	Volts
	Trap Funnel RF	89	Volts		Oct2 DC	-14.6	Volts
	Trap Funnel Exit	-10.8	Volts	Source	Corona	65.9	uA
	HP Funnel Delta	-118	Volts		Gas Temp	280	°C
	HP Funnel RF	90	Volts		Nebulizer	12	psig
	Trap Funnel Delta	-121	Volts		Drying Gas	2	l/min
	HP Funnel Exit	0	Volts		Sheath Gas Flow	0.8	l/min
IM Rear Funnel	Rear Funnel RF	179	Volts		Ion Polarity	(negative)	
	Rear Funnel Exit	-35.4	Volts		Nozzle Voltage	0	Volts
	Rear Funnel Entrance	-199	Volts		Vcap	3500	Volts
	IM Hex RF	599	Volts		Oct 1 RF Vpp	750	Volts
	IM Hex Entrance	-31.8	Volts		Skimmer	-27	Volts
IM Trap Funnel	Exit Grid 2 Low	-72	Volts		Fragmentor	320	Volts
	Exit Grid 2 Delta	-63	Volts	TOF	Mirror Back	-1250	Volts
	Exit Grid 1 Low	-69	Volts		Mirror Mid	1675	Volts
	Exit Grid 1 Delta	-64	Volts		Mirror Front	7000	Volts
	Exit	-67	Volts		Acc Focus	1950	Volts
	Entrance Grid Low	-72	Volts		Puller Offset	-32	Volts
	Entrance Grid Delta	-69.9	Volts		Puller	700	Volts
	Entrance	-69	Volts		Pusher	-1200	Volts
Ion Funnel	Funnel RF LP	100	Volts		Min Range	62080	ns

ESI-MS titrations: Data processing

Using dT_6 as an internal calibrant we quantify the change in DNA response.[50] Figure S63 shows how the ratio of collective DNA signal vs. dT_6 signal evolves as a function of ligand concentration. The estimated response factors are listed in table. Based on the estimates we allowed the response factors for the complex signals to fluctuate between 0.8 and 1.1 during dynamic fitting.

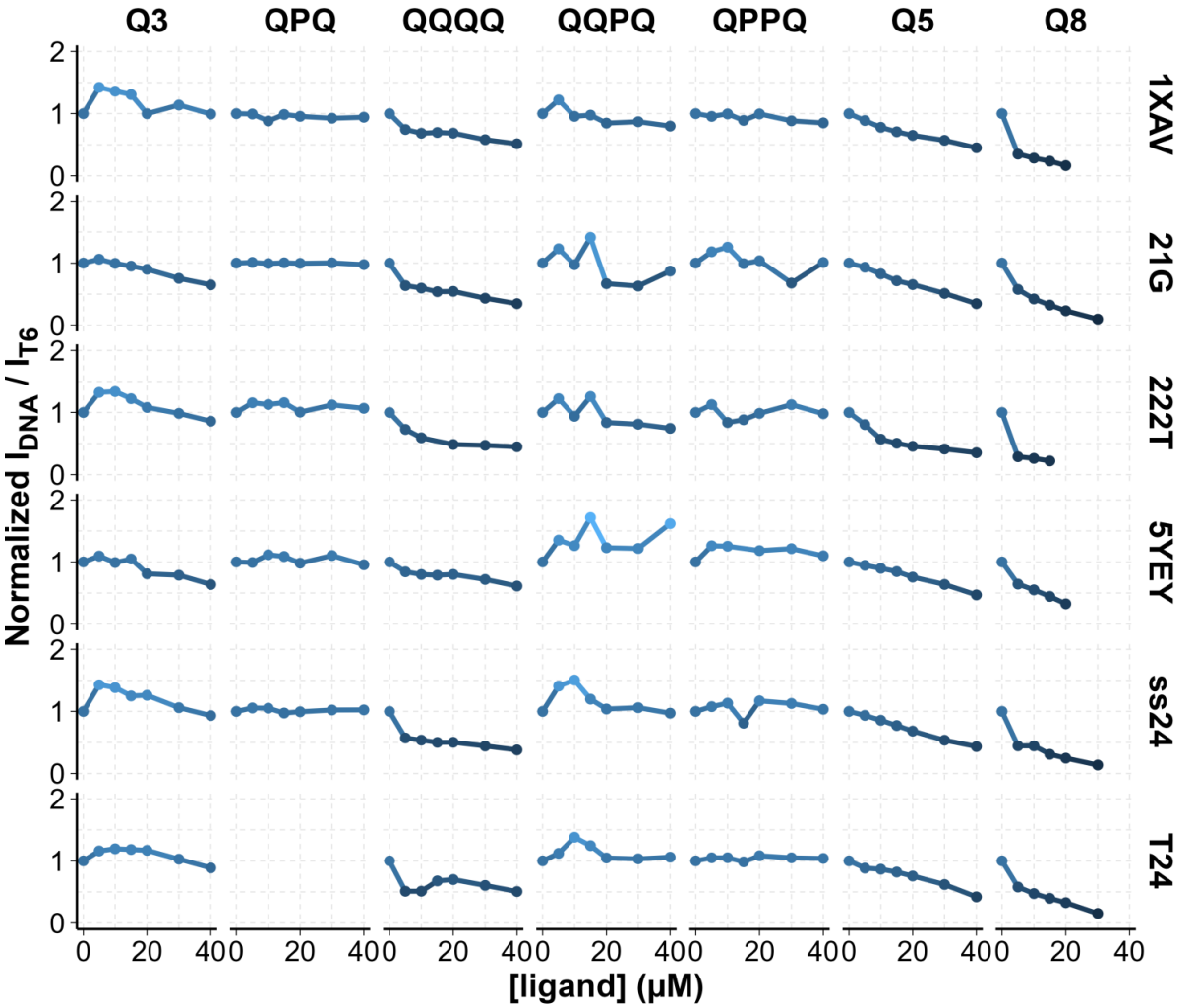


Figure S63 The sum of all DNA signal intensities vs. the signal of calibrant dT_6. The first datapoint is normalized to the value 1.

A declining curve means that adding ligand negatively impacts the detection of DNA ions. This leads to an ever-decreasing S/N ratio until the DNA species fall under the limit of detection. The latter happened with Q_8, which is why the titration datasets with Q_8 are incomplete.

The titration data points, corrected for noise and response, are shown in Figure S64, alongside the dynamic fits from which we obtained refined K_D values. All K_D values are listed in Table S3.

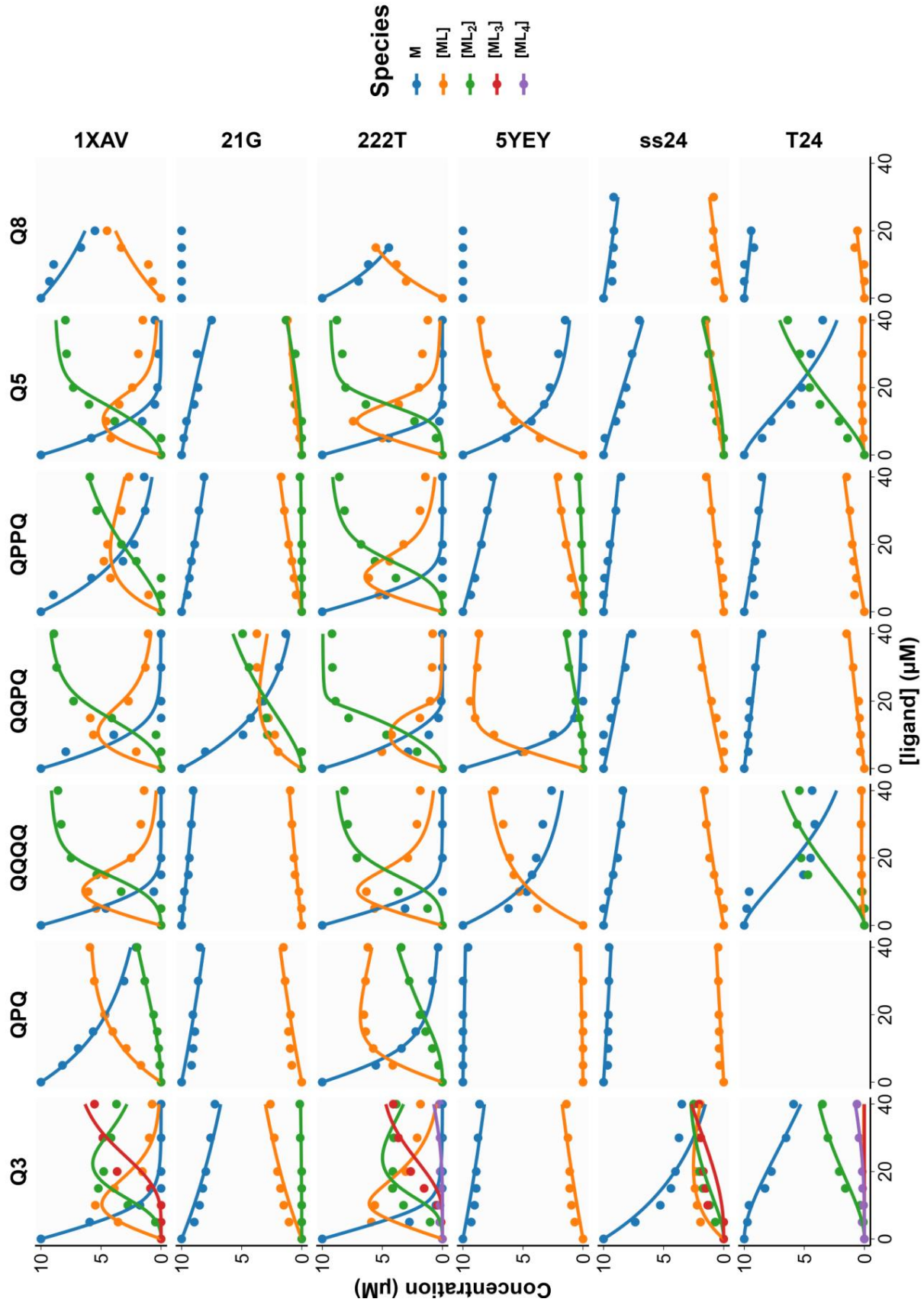


Figure S64 Data points obtained from processing the ESI-MS titration mass spectra. K_D values are calculated by dynamically fitting the experimental data based on the complex formation equilibria. Curves show the dynamic fits. Samples contain 10 μM DNA, 0-40 μM ligand, 4 μM dT_6, 0.5 mM KCl and 100 mM TMAA (pH 6.8).

Ligand screening: Comprehensive list of DNA/Ligand species concentrations and their K_D values

Table S2 Data extracted from the ligand screening, featuring: Concentration of unbound DNA (M), 1:1 complex (ML) and 2:1 complex (ML_2), fraction of DNA bound (fb), complex dissociation constants K_{D1} and K_{D2} as well as ligand cooperativity. Sample conditions were: 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

DNA	Ligand	M [µM]	ML [µM]	ML_2 [µM]	fb	K_{D1} [µM]	K_{D2} [µM]	Coop.[6]
1XAV	QQPQ	0.4	3.7	5.9	*0.96*	**0.5**	**1.5**	positive
1XAV	QQQQ	0.3	3.2	6.5	*0.97*	**0.3**	**0.9**	positive
1XAV	QPPQ	2.5	4.6	3.0	*0.75*	**5.2**	**7.4**	positive
1XAV	QPQ	4.7	4.2	1.1	*0.53*	**15**	**27**	positive
222T	QQPQ	0.4	4.3	5.3	*0.96*	**0.5**	**2.1**	negative
222T	QQQQ	0.8	3.9	5.4	*0.92*	**1.1**	**2.0**	positive
222T	QPPQ	1.4	5.0	3.6	*0.86*	**2.3**	**5.4**	positive
222T	QPQ	2.6	5.9	1.5	*0.74*	**4.8**	**22**	negative
222T_mA	QQPQ	1.2	2.6	6.3	*0.88*	**2.2**	**1.0**	positive
222T_mA	QQQQ	3.6	3.0	3.4	*0.64*	**12**	**4.5**	positive
222T_mA	QPPQ	3.0	3.0	4.0	*0.70*	**8.9**	**3.4**	positive
222T_mA	QPQ	5.5	3.0	1.5	*0.45*	**25**	**14**	positive
222T_mC	QQPQ	0.5	2.9	6.5	*0.95*	**0.7**	**0.9**	positive
222T_mC	QQQQ	1.9	4.0	4.1	*0.81*	**3.7**	**3.7**	positive
222T_mC	QPPQ	2.2	3.7	4.1	*0.78*	**4.9**	**3.6**	positive
222T_mC	QPQ	3.0	5.1	1.9	*0.70*	**6.6**	**15**	positive
T30177TT	QQPQ	0.9	4.7	4.4	*0.91*	**1.2**	**3.4**	positive
T30177TT	QQQQ	0.7	4.5	4.7	*0.93*	**1.0**	**2.9**	positive
T30177TT	QPPQ	1.5	6.2	2.4	*0.85*	**2.1**	**12**	negative
T30177TT	QPQ	2.9	6.2	0.8	*0.71*	**5.7**	**47**	negative
26CEB	QQPQ	1.2	8.2	0.5	*0.88*	**1.6**	**83**	negative
26CEB	QQQQ	0.7	8.4	0.9	*0.93*	**0.8**	**45**	negative
26CEB	QPPQ	1.8	7.7	0.5	*0.82*	**2.6**	**95**	negative
26CEB	QPQ	3.3	6.4	0.3	*0.67*	**6.6**	**136**	negative
26CEB_mT	QQPQ	1.2	7.9	0.9	*0.88*	**1.5**	**43**	negative
26CEB_mT	QQQQ	2.2	6.2	1.6	*0.78*	**3.8**	**21**	negative
26CEB_mT	QPPQ	2.1	7.3	0.7	*0.79*	**3.3**	**63**	negative
26CEB_mT	QPQ	3.6	6.0	0.4	*0.64*	**8.1**	**103**	negative
2KYP	QQPQ	2.2	2.4	5.4	*0.78*	**6.2**	**1.5**	positive
2KYP	QQQQ	2.7	6.6	0.7	*0.73*	**4.9**	**57**	negative
2KYP	QPPQ	7.1	1.7	1.2	*0.29*	**66**	**11**	positive
2KYP	QPQ	8.0	1.7	0.3	*0.20*	**84**	**47**	positive
2O3M	QQPQ	7.2	1.6	1.2	*0.28*	**71**	**11**	positive
2O3M	QQQQ	7.8	2.2	0	*0.22*	**64**	**0**	
2O3M	QPPQ	7.5	2.2	0.3	*0.25*	**58**	**58**	positive
2O3M	QPQ	8.6	1.4	0	*0.14*	**119**	**0**	

[6] Positive, if $4*K_{D1} > K_{D2}$. Negative, if $4*K_{D1} < K_{D2}$. Assuming independent and equivalent binding sites.[49]

TG4T_K	QQPQ	4.2	5.0	0.8	*0.58*	**11**	**43**	positive
TG4T_K	QQQQ	7.3	2.7	0	*0.27*	**46**	**0**	
TG4T_K	QPPQ	5.7	3.8	0.5	*0.43*	**23**	**60**	positive
TG4T_K	QPQ	8.0	2.0	0	*0.20*	**70**	**0**	
TG4T_NH4	QQPQ	2.8	6.2	1.0	*0.72*	**5.3**	**37**	negative
TG4T_NH4	QQQQ	8.9	1.1	0	*0.11*	**149**	**0**	
TG4T_NH4	QPPQ	5.8	3.9	0.3	*0.42*	**23**	**101**	negative
TG4T_NH4	QPQ	7.5	2.3	0.2	*0.25*	**58**	**114**	positive
21G	QQPQ	5.9	2.6	1.5	*0.41*	**32**	**13**	positive
21G	QQQQ	9.1	0.9	0	*0.09*	**197**	**0**	
21G	QPPQ	8.4	1.3	0.2	*0.16*	**116**	**53**	positive
21G	QPQ	9.3	0.7	0	*0.07*	**250**	**0**	
5YEY	QQPQ	1.2	7.3	1.6	*0.88*	**1.5**	**22**	negative
5YEY	QQQQ	5.0	5.0	0	*0.50*	**15**	**0**	
5YEY	QPPQ	7.8	1.8	0.4	*0.22*	**75**	**40**	positive
5YEY	QPQ	9.4	0.6	0	*0.06*	**309**	**0**	
22GT	QQPQ	5.5	3.1	1.4	*0.45*	**25**	**16**	positive
22GT	QQQQ	8.5	1.5	0	*0.15*	**104**	**0**	
22GT	QPPQ	8.0	1.6	0.4	*0.20*	**89**	**36**	positive
22GT	QPQ	9.1	0.9	0	*0.09*	**196**	**0**	
22GT_18T	QQPQ	2.6	6.5	0.9	*0.74*	**4.6**	**41**	negative
22GT_18T	QQQQ	5.4	4.6	0	*0.46*	**18**	**0**	
22GT_18T	QPPQ	7.7	1.9	0.4	*0.23*	**69**	**45**	positive
22GT_18T	QPQ	8.8	1.2	0.0	*0.12*	**143**	**0**	
22CTA	QQPQ	7.0	2.0	1.0	*0.30*	**57**	**16**	positive
22CTA	QQQQ	8.6	1.4	0	*0.14*	**114**	**0**	
22CTA	QPPQ	8.6	1.4	0	*0.14*	**116**	**0**	
22CTA	QPQ	8.8	1.2	0	*0.12*	**135**	**0**	
TBA	QQPQ	6.7	3.3	0	*0.33*	**34**	**0**	
TBA	QQQQ	8.1	1.9	0	*0.19*	**78**	**0**	
TBA	QPPQ	8.6	1.4	0	*0.14*	**118**	**0**	
TBA	QPQ	9.3	0.7	0	*0.07*	**252**	**0**	
G4T4G4_K	QQPQ	6.1	3.3	0.6	*0.39*	**28**	**39**	positive
G4T4G4_K	QQQQ	8.2	1.8	0	*0.18*	**81**	**0**	
G4T4G4_K	QPPQ	7.9	2.1	0	*0.21*	**67**	**0**	
G4T4G4_K	QPQ	8.4	1.6	0	*0.16*	**99**	**0**	
G4T4G4_NH4	QQPQ	5.8	3.7	0.5	*0.42*	**24**	**53**	positive
G4T4G4_NH4	QQQQ	9.7	0.3	0	*0.03*	**562**	**0**	
G4T4G4_NH4	QPPQ	9.2	0.8	0	*0.08*	**225**	**0**	
G4T4G4_NH4	QPQ	9.6	0.4	0	*0.04*	**470**	**0**	
26TTA[7]	QQPQ	2.1	3.8	3.3	*0.71*	**4.1**	**4**	positive
26TTA	QQQQ	6.8	2.2	1.0	*0.32*	**48**	**18**	positive
26TTA	QPPQ	7.7	1.9	0.4	*0.23*	**70**	**40**	positive

[7] This is the only instance where we observed a 3:1 complex. [ML_3] = 0.8 µM, K_{D3} = 22 µM (negative coop.).

26TTA	QPQ	9.1	0.9	0	*0.09*	**204**	**0**	
Bcl2	QQPQ	4.4	3.4	2.1	*0.56*	**16**	**9.8**	positive
Bcl2	QQQQ	4.0	4.0	1.9	*0.60*	**12**	**13**	positive
Bcl2	QPPQ	5.3	3.6	1.1	*0.47*	**21**	**24**	positive
Bcl2	QPQ	7.5	2.5	0	*0.25*	**52**	**0**	
24TTG	QQPQ	7.1	1.1	1.8	*0.29*	**102**	**4.6**	positive
24TTG	QQQQ	9.8	0.2	0	*0.02*	**809**	**0**	
24TTG	QPPQ	8.8	1.2	0	*0.12*	**145**	**0**	
24TTG	QPQ	9.4	0.6	0	*0.06*	**284**	**0**	
24TTG_20T	QQPQ	5.2	3.6	1.2	*0.48*	**20**	**22**	positive
24TTG_20T	QQQQ	6.0	4.0	0	*0.40*	**24**	**0**	
24TTG_20T	QPPQ	8.3	1.7	0	*0.17*	**92**	**0**	
24TTG_20T	QPQ	8.9	1.1	0	*0.11*	**153**	**0**	positive
23TAG	QQPQ	5.7	3.4	1.0	*0.43*	**25**	**25**	positive
23TAG	QQQQ	8.9	1.1	0	*0.11*	**148**	**0**	
23TAG	QPPQ	8.1	1.9	0	*0.19*	**75**	**0**	
23TAG	QPQ	8.3	1.7	0	*0.17*	**93**	**0**	
2KPR	QQPQ	4.6	1.9	3.5	*0.54*	**26**	**3**	positive
2KPR	QQQQ	8.7	1.3	0	*0.13*	**131**	**0**	
2KPR	QPPQ	7.0	2.0	1.0	*0.30*	**56**	**17**	positive
2KPR	QPQ	8.6	1.1	0.3	*0.14*	**144**	**37**	positive
21CCC (i-m)	QQPQ	9.4	0.6	0	*0.06*	**294**	**0**	
21CCC (i-m)	QQQQ	10	0	0	*0.00*	**0**	**0**	
21CCC (i-m)	QPPQ	9.4	0.6	0	*0.06*	**290**	**0**	
21CCC (i-m)	QPQ	9.8	0.2	0	*0.02*	**1280**	**0**	
ds26	QQPQ	8.7	1.3	0	*0.13*	**120**	**0**	
ds26	QQQQ	10	0	0	*0.00*	**0**	**0**	
ds26	QPPQ	8.8	0	1.2	*0.12*	**0**	**0**	
ds26	QPQ	10	0	0	*0.00*	**0**	**0**	
DK33	QQPQ	8.5	1.5	0	*0.15*	**107**	**0**	
DK33	QQQQ	10	0	0	*0.00*	**0**	**0**	
DK33	QPPQ	8.1	1.9	0	*0.19*	**80**	**0**	
DK33	QPQ	8.8	1.2	0	*0.12*	**141**	**0**	
DK66	QQPQ	8.7	1.3	0	*0.13*	**123**	**0**	
DK66	QQQQ	10	0	0	*0.00*	**3981**	**0**	
DK66	QPPQ	8.3	1.7	0	*0.17*	**88**	**0**	
DK66	QPQ	9.7	0.3	0	*0.03*	**586**	**0**	
DK100	QQPQ	7.7	2.3	0	*0.23*	**58**	**0**	
DK100	QQQQ	10	0.0	0	*0.00*	**0**	**0**	
DK100	QPPQ	8.1	1.9	0	*0.19*	**77**	**0**	
DK100	QPQ	10	0	0	*0.00*	**0**	**0**	
ss24	QQPQ	8.2	1.8	0	*0.18*	**83**	**0**	
ss24	QQQQ	8.7	1.3	0	*0.13*	**126**	**0**	
ss24	QPPQ	9.1	0.9	0	*0.09*	**182**	**0**	
ss24	QPQ	9.7	0.3	0	*0.03*	**626**	**0**	
24nonG4	QQPQ	9.2	0.8	0	*0.08*	**208**	**0**	

24nonG4	QQQQ	7.2	2.8	0	*0.28*	**44**	**0**	
24nonG4	QPPQ	7.5	2.5	0	*0.25*	**51**	**0**	
24nonG4	QPQ	10	0	0	*0.00*	**0**	**0**	
T24	QQPQ	8.3	1.5	0.2	*0.17*	**102**	**65**	positive
T24	QQQQ	6.8	0.4	2.8	*0.32*	**235**	**1.0**	positive
T24	QPPQ	8.8	1.1	0.1	*0.12*	**149**	**195**	positive
T24	QPQ	9.2	0.8	0	*0.08*	**210**	**0**	
A24	QQPQ	10	0	0	*0.00*	**0**	**0**	
A24	QQQQ	10	0	0	*0.00*	**0**	**0**	
A24	QPPQ	8.3	1.7	0	*0.17*	**92**	**0**	
A24	QPQ	8.8	1.2	0	*0.12*	**138**	**0**	
21CCC (ss)	QQPQ	9.5	0.5	0	*0.05*	**354**	**0**	
21CCC (ss)	QQQQ	10	0	0	*0.00*	**4064**	**0**	
21CCC (ss)	QPPQ	9.5	0.5	0	*0.05*	**359**	**0**	
21CCC (ss)	QPQ	9.9	0.1	0	*0.01*	**3365**	**0**	
T6	QQPQ	9.7	0.3	0	*0.03*	**722**	**0**	
T6	QQQQ	10	0	0	*0.00*	**0**	**0**	
T6	QPPQ	9.7	0.3	0	*0.03*	**725**	**0**	
T6	QPQ	9.8	0.2	0	*0.02*	**1150**	**0**	

Ligand screening – Mass spectra

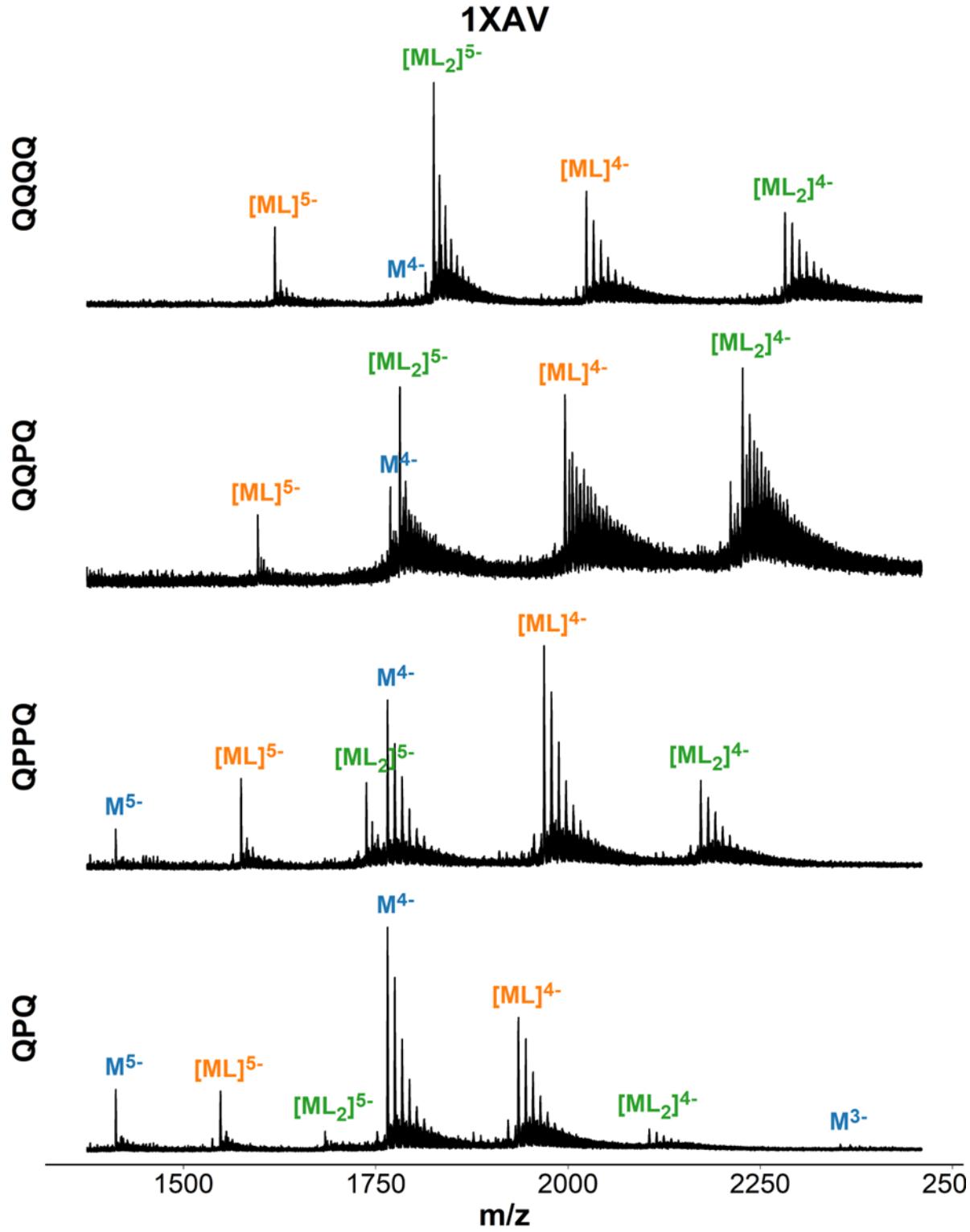


*Figure S65 Mass spectra of 1XAV (dT**GA**GGG**T**GGG**TA**GGG**T**GGG**TAA**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

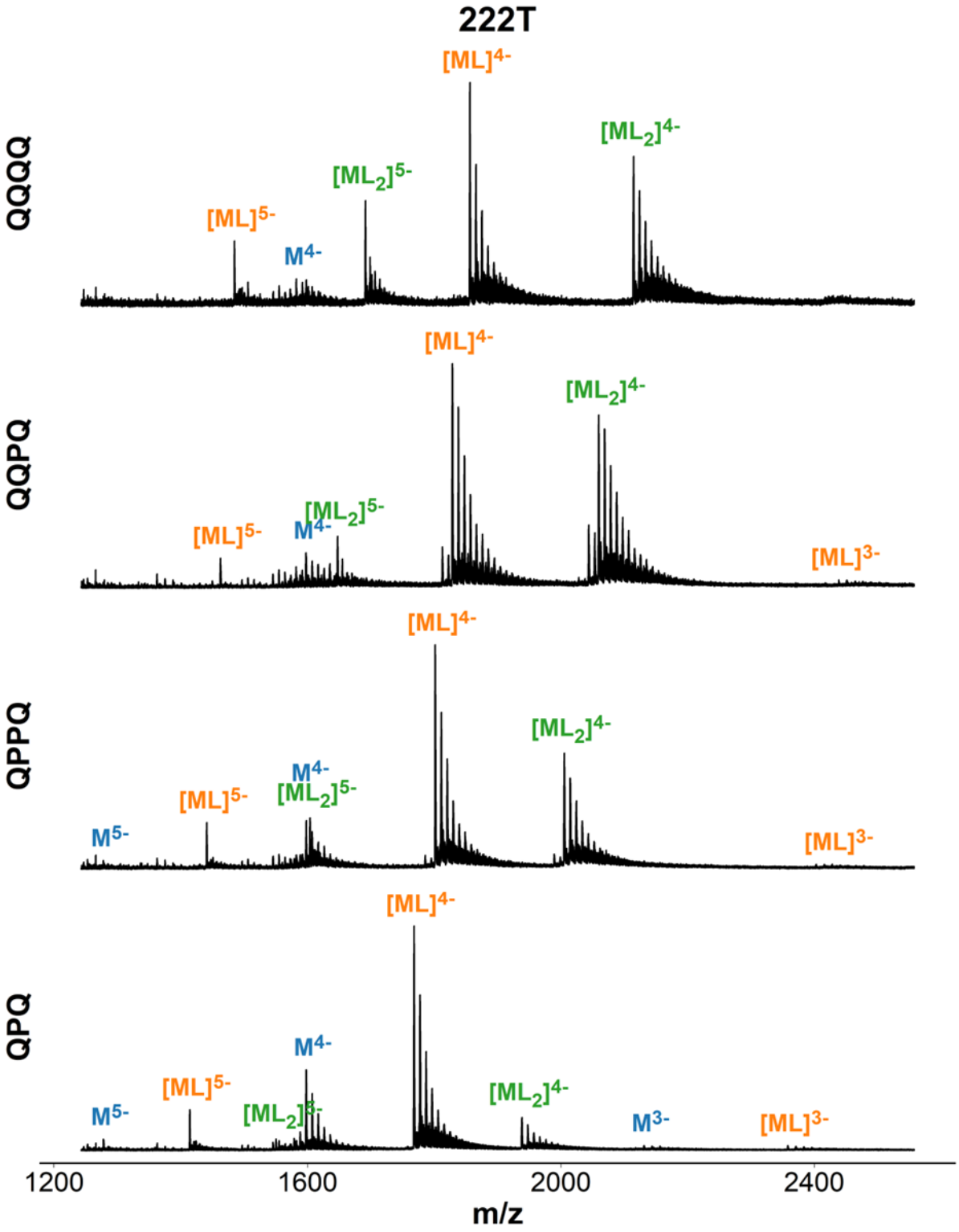


*Figure S66 Mass spectra of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

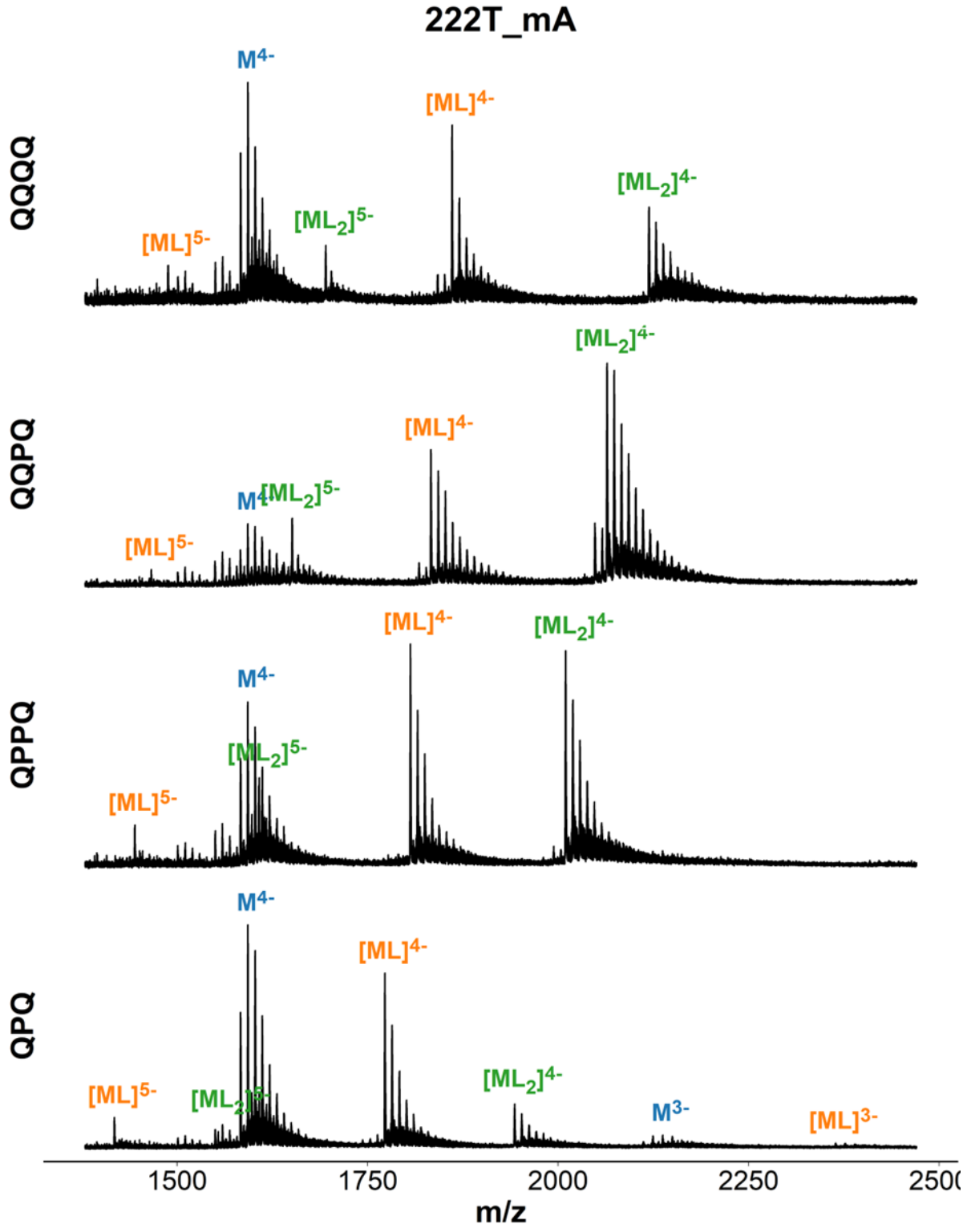


*Figure S67 Mass spectra of 222T-mA (dT**GGG**TT**GGG**AA**GGG**TT**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

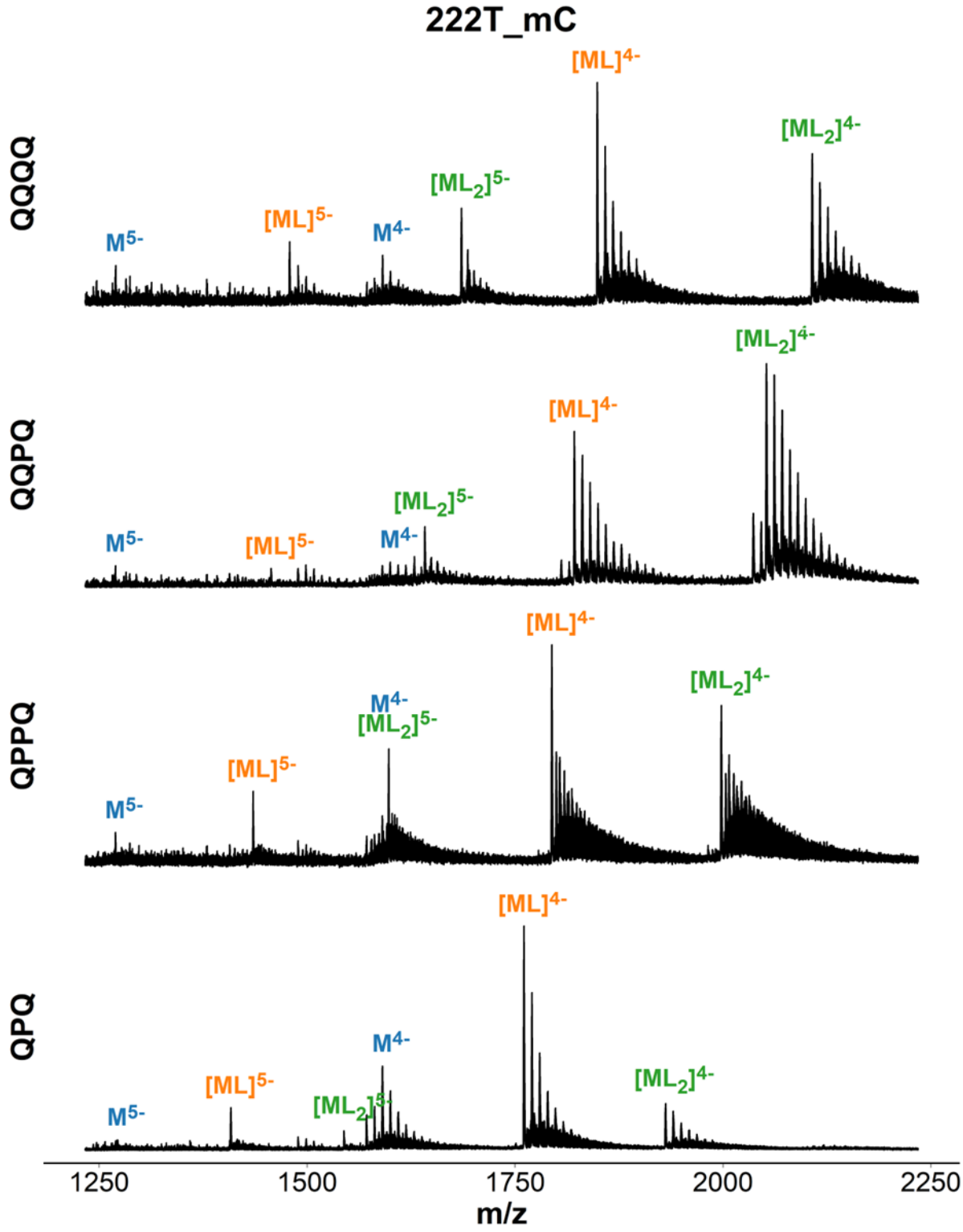


*Figure S68 Mass spectra of 222T-mC (dT**GGG**TT**GGG**CC**GGG**TT**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

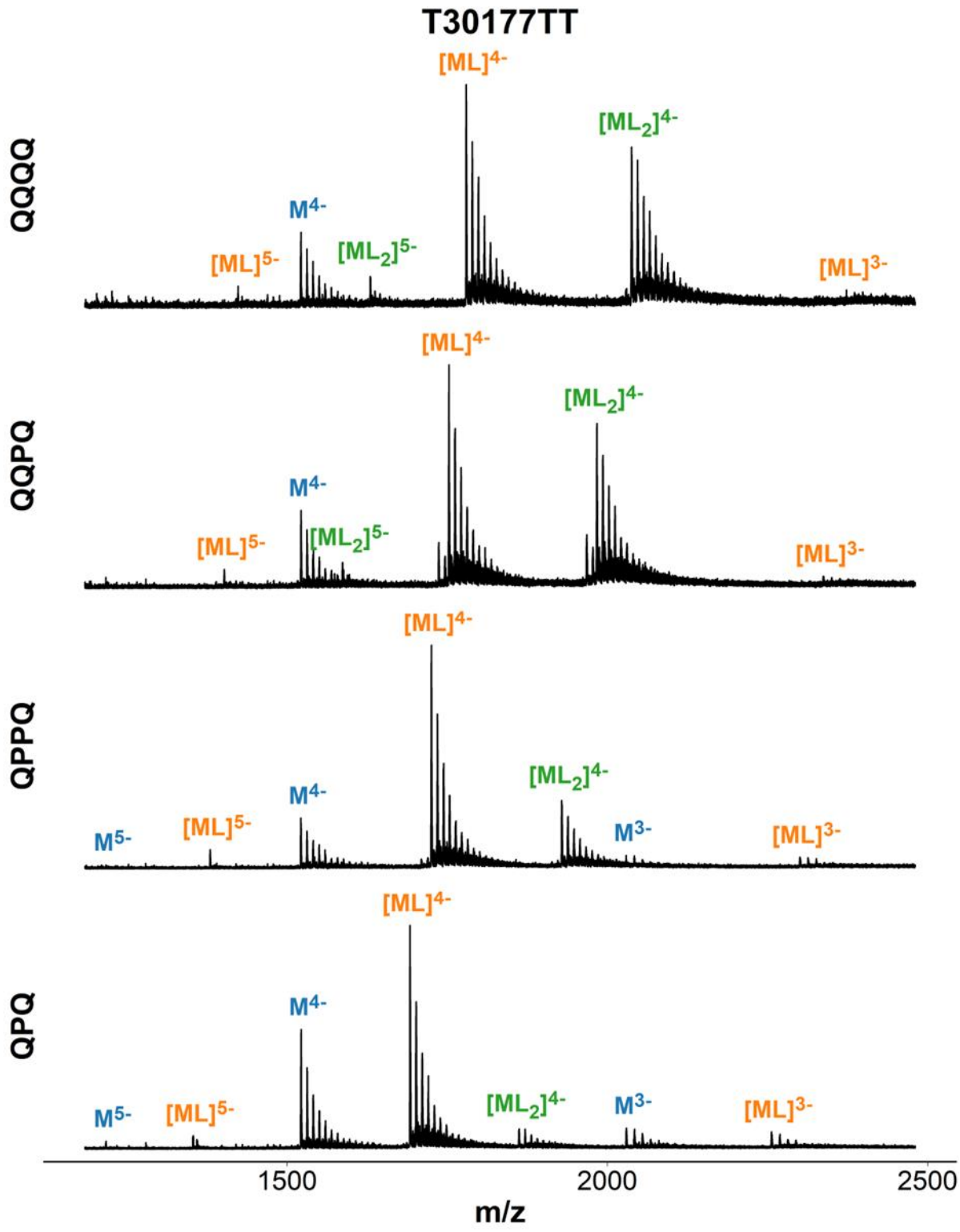


*Figure S69 Mass spectra of T30177TT (dTT**G**T**GG**T**GGG**T**GGG**T**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

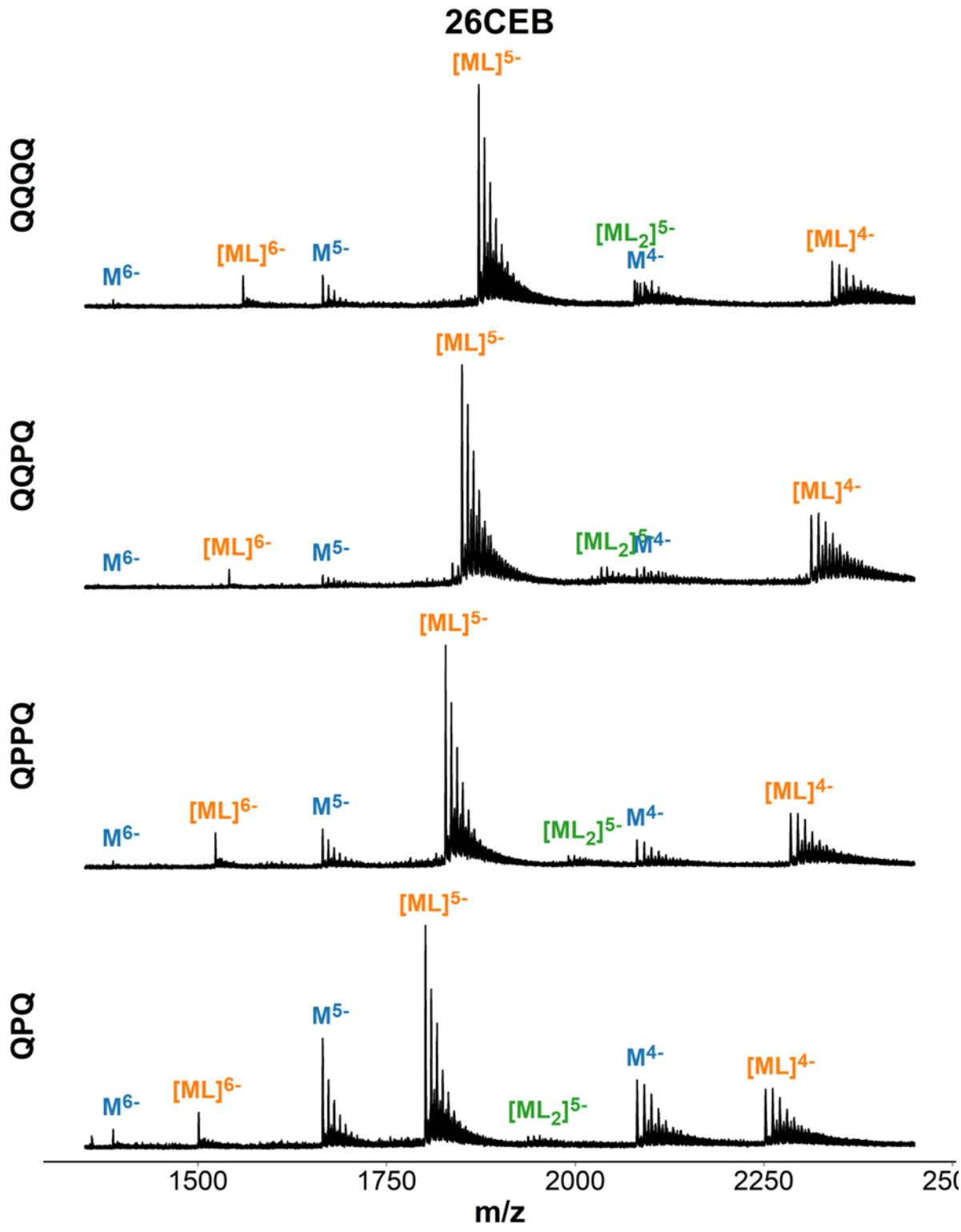


*Figure S70 Mass spectra of 26CEB (dAA**GGG**T**GGG**T**G**TAA**G**T**G**T**GGG**T**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

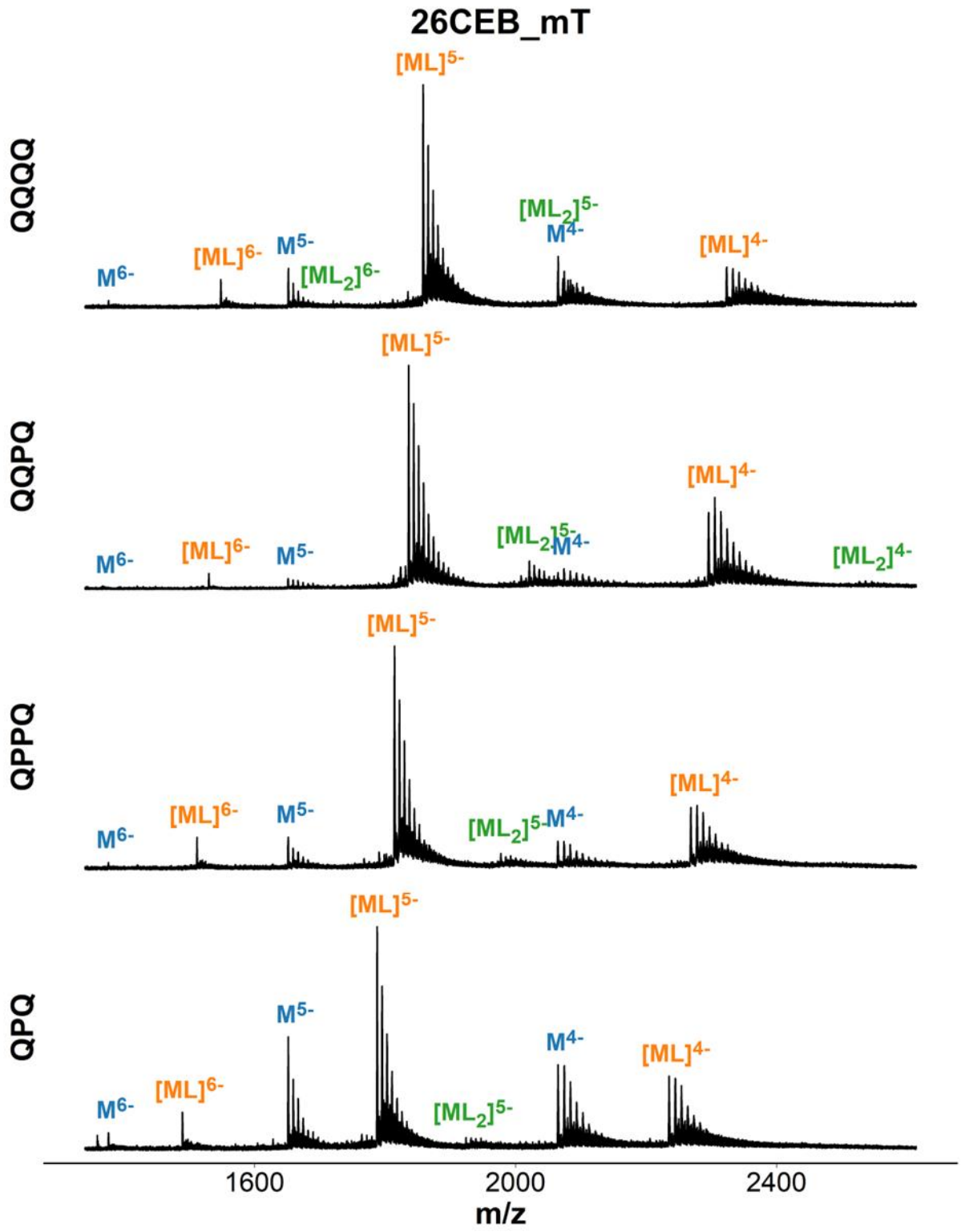


*Figure S71 Mass spectra of 26CEB-mT (dAA**GGG**T**GGG**TTTTTTT**G**T**GGG**T**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

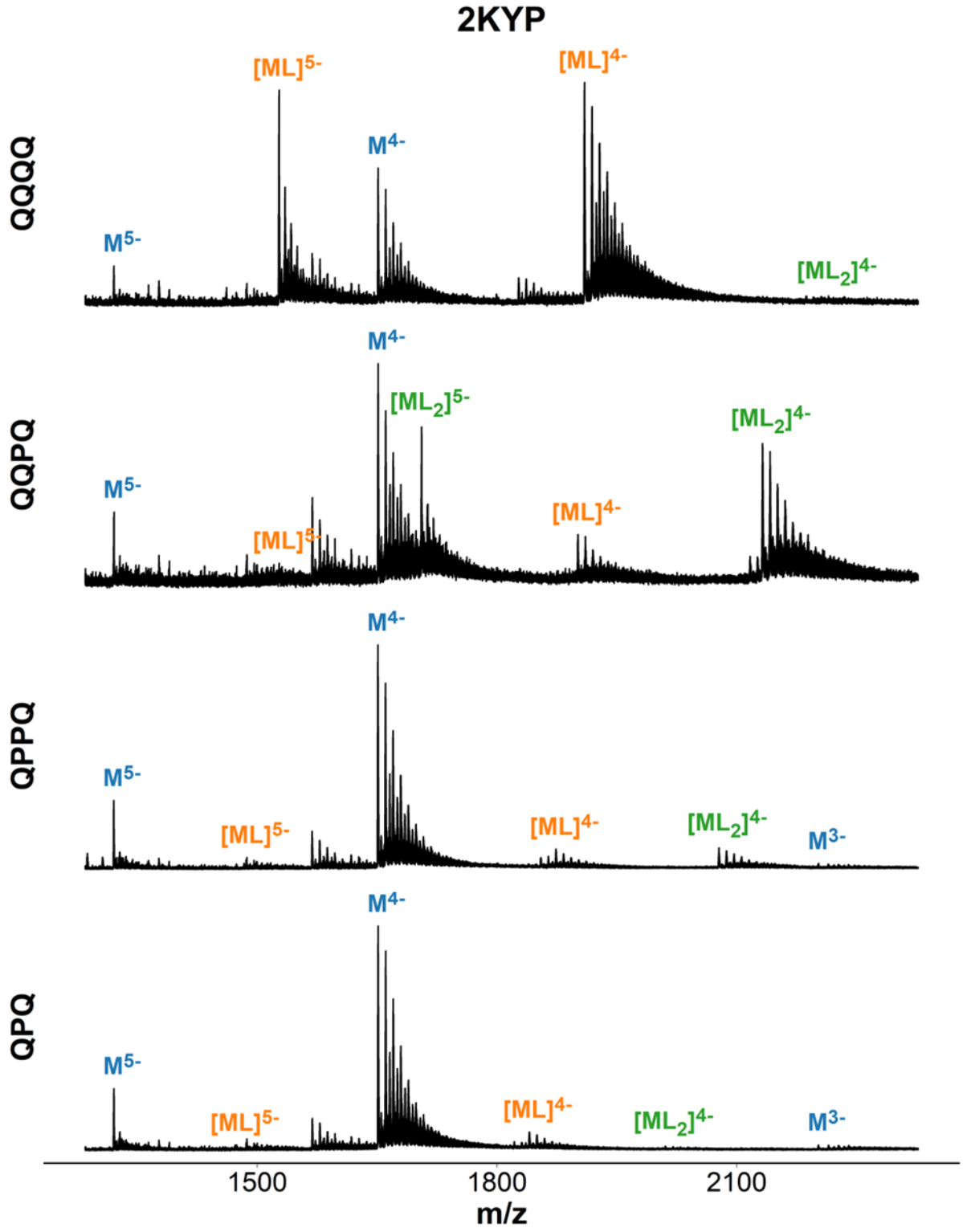


*Figure S72 Mass spectra of 2KYP (dC**GGG**C**GGG**C**G**CTA**GGG**A**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

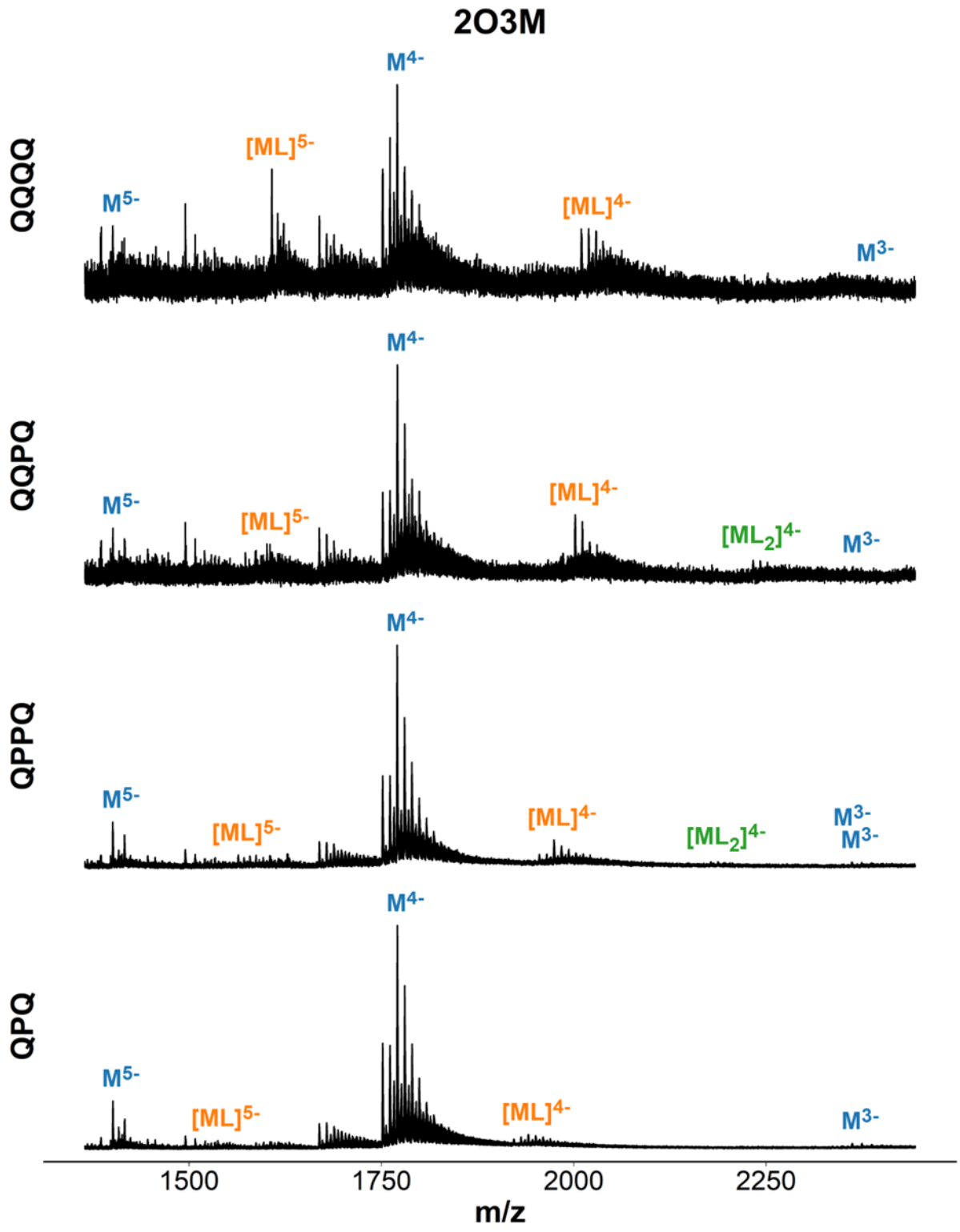


*Figure S73 Mass spectra of 2O3M (dA**GGG**A**GGG**C**G**CT**GGG**A**GG**A**GGG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

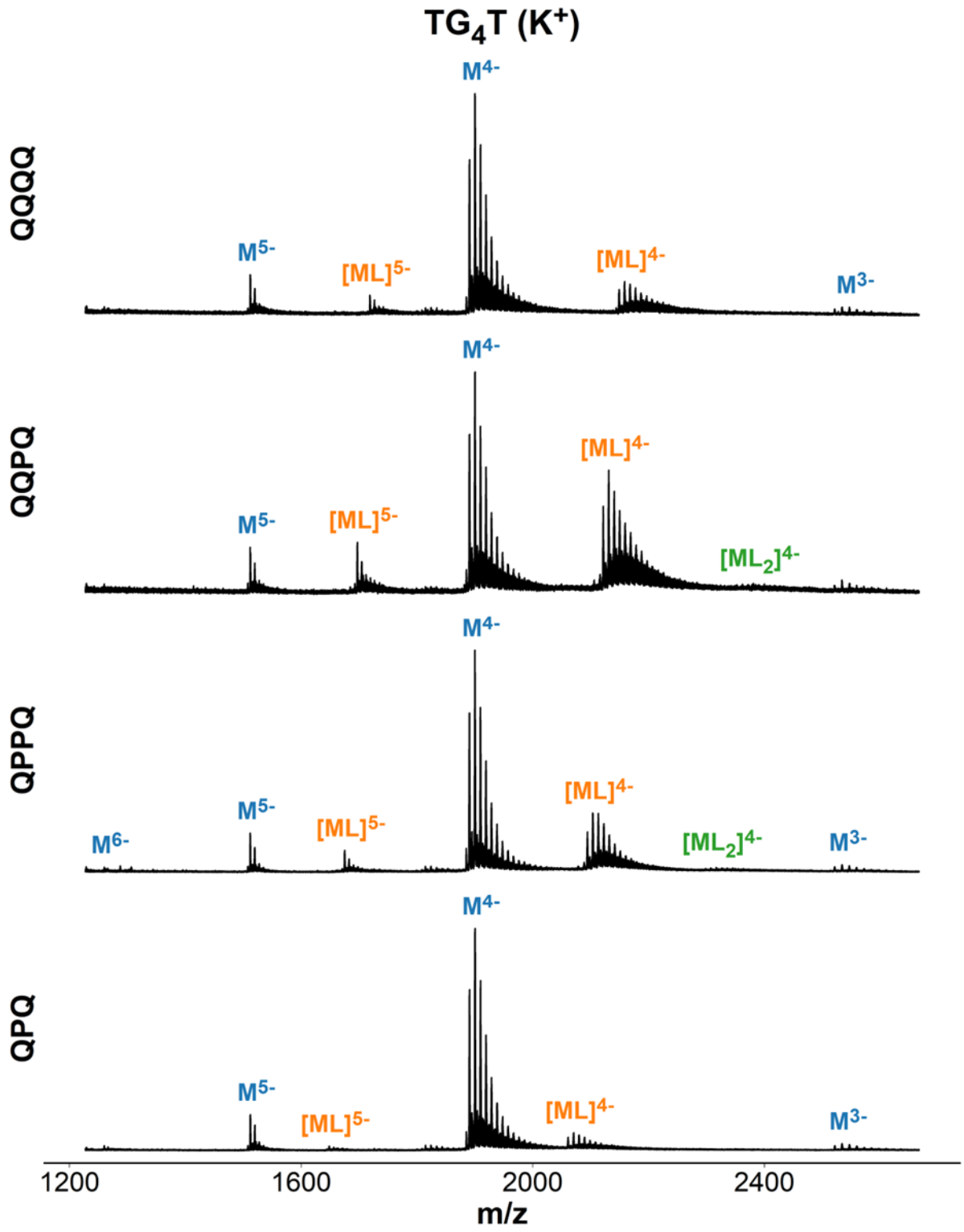


*Figure S74 Mass spectra of TG4T ([dT**GGGG**T)$_4$) in presence of ligand. Samples contain 40 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

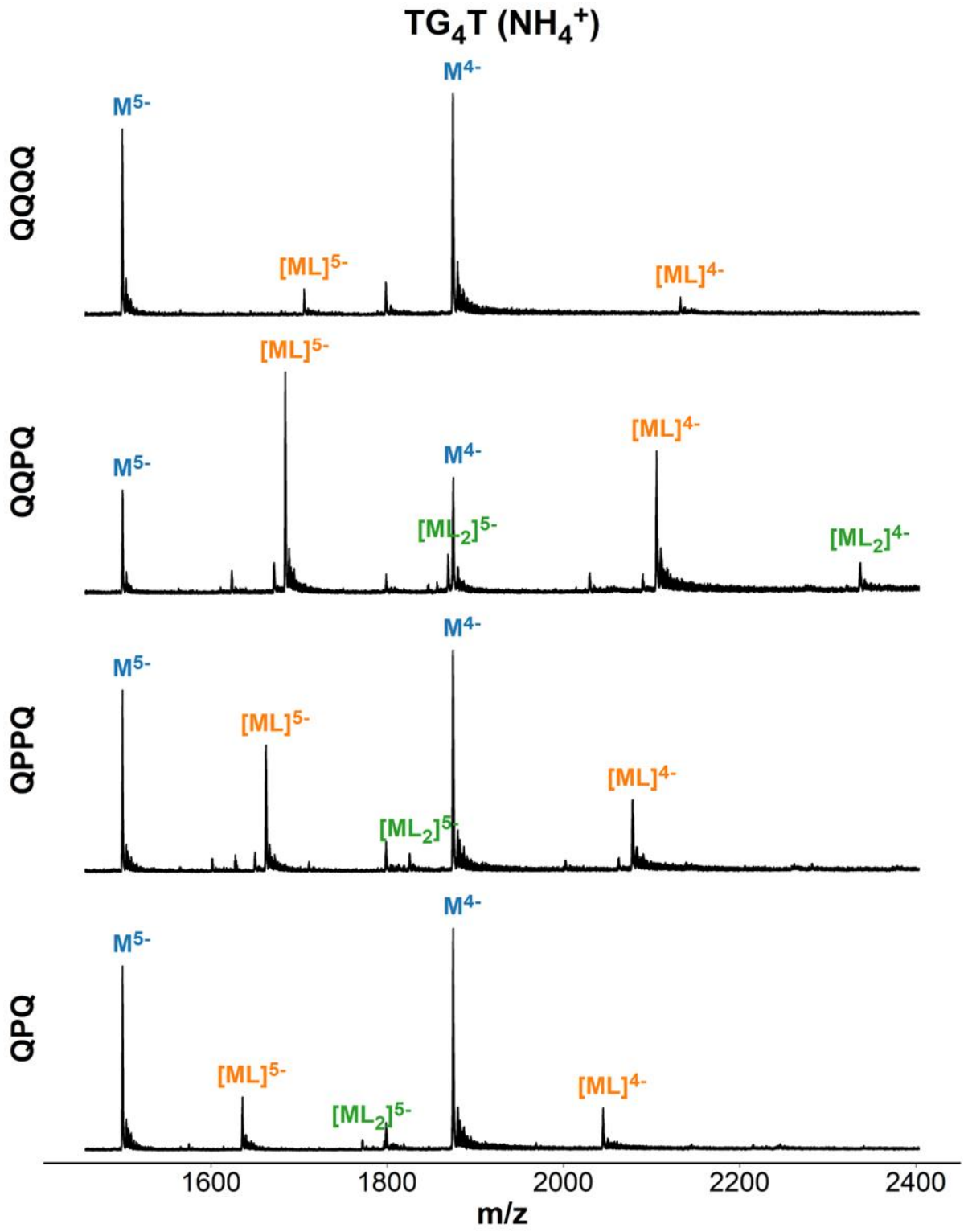


*Figure S75 Mass spectra of TG4T ([dT**GGGG**T)$_4$) in presence of ligand. Samples contain 40 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).*

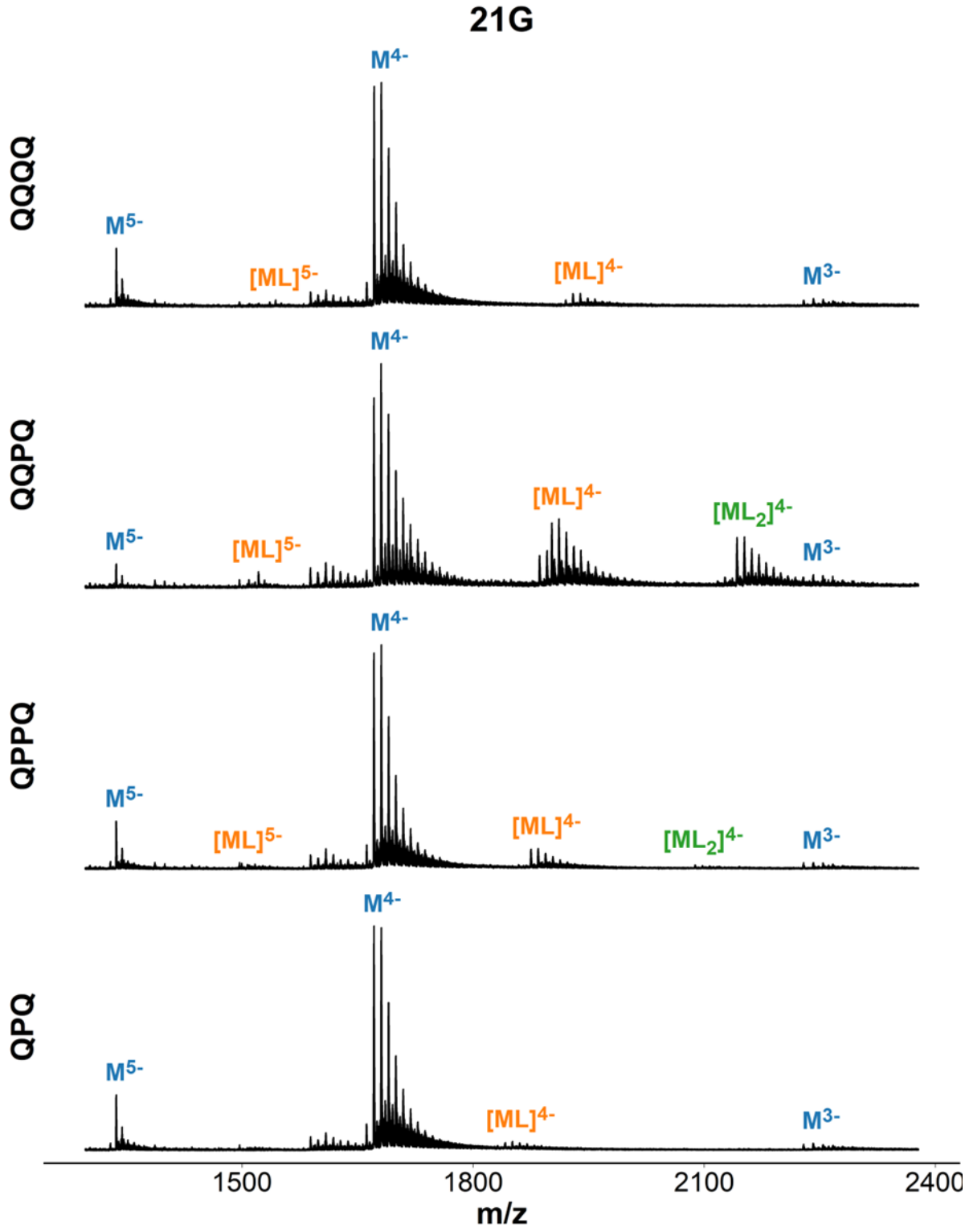


*Figure S76 Mass spectra of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

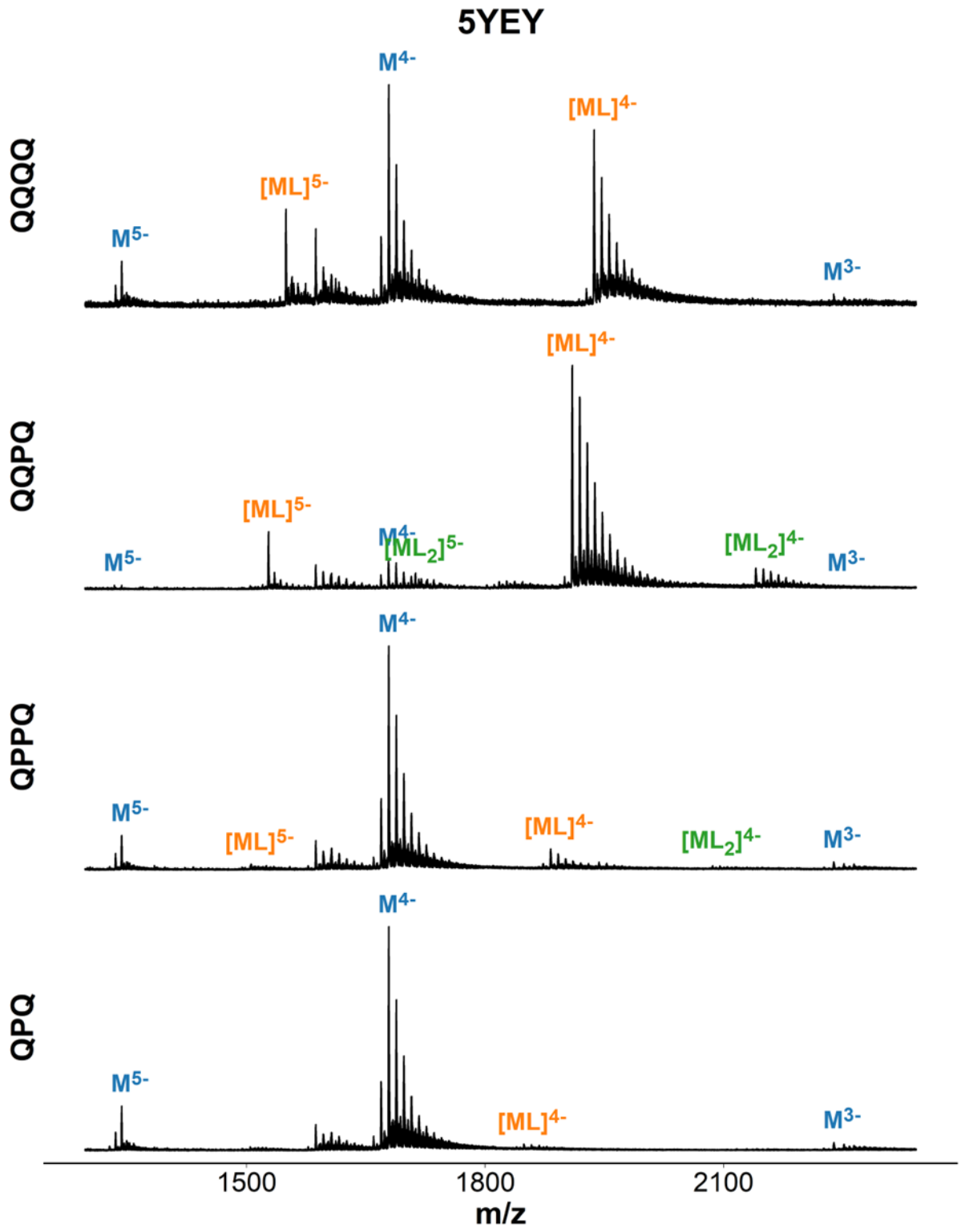


*Figure S77 Mass spectra of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

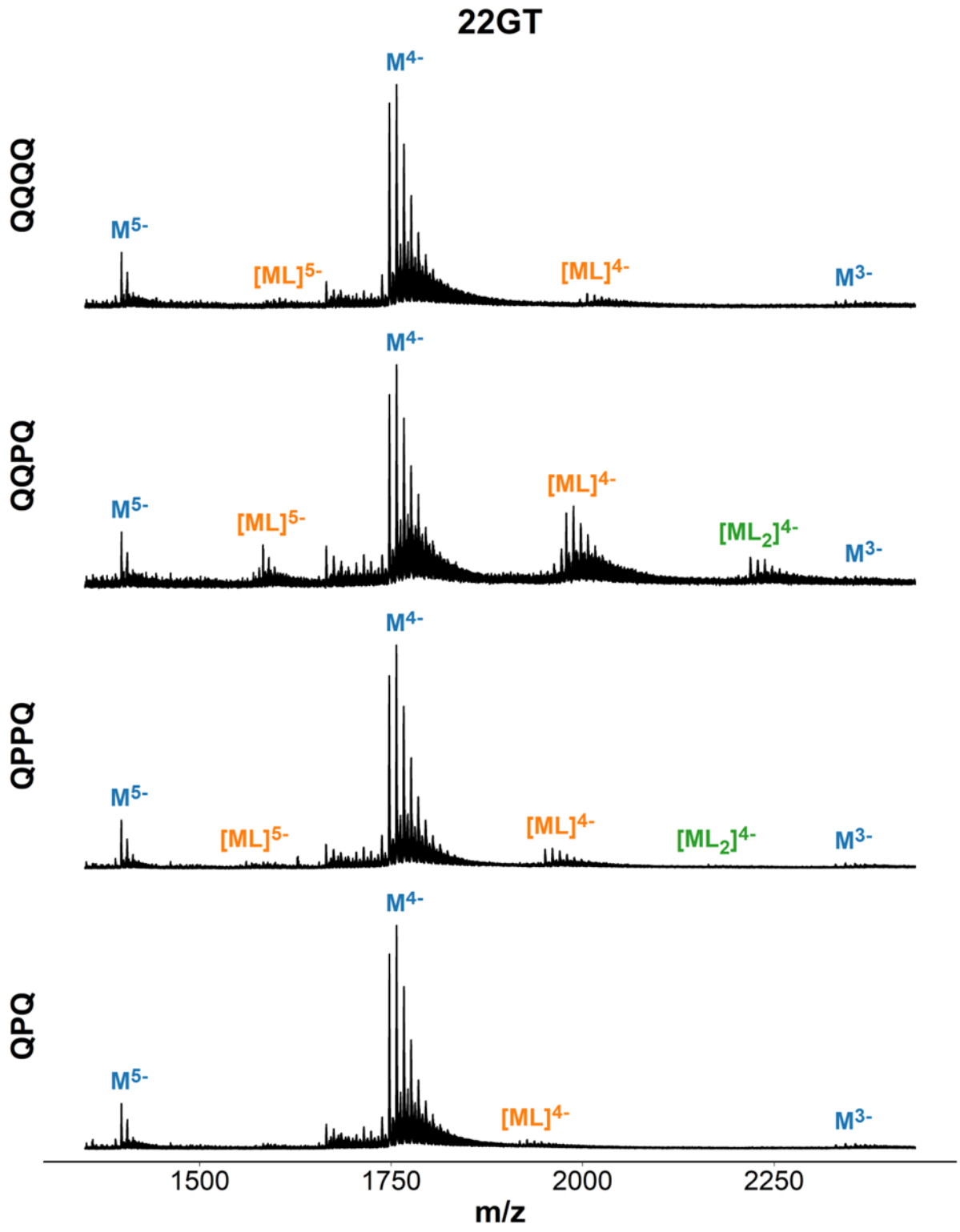


*Figure S78 Mass spectra of 22GT (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

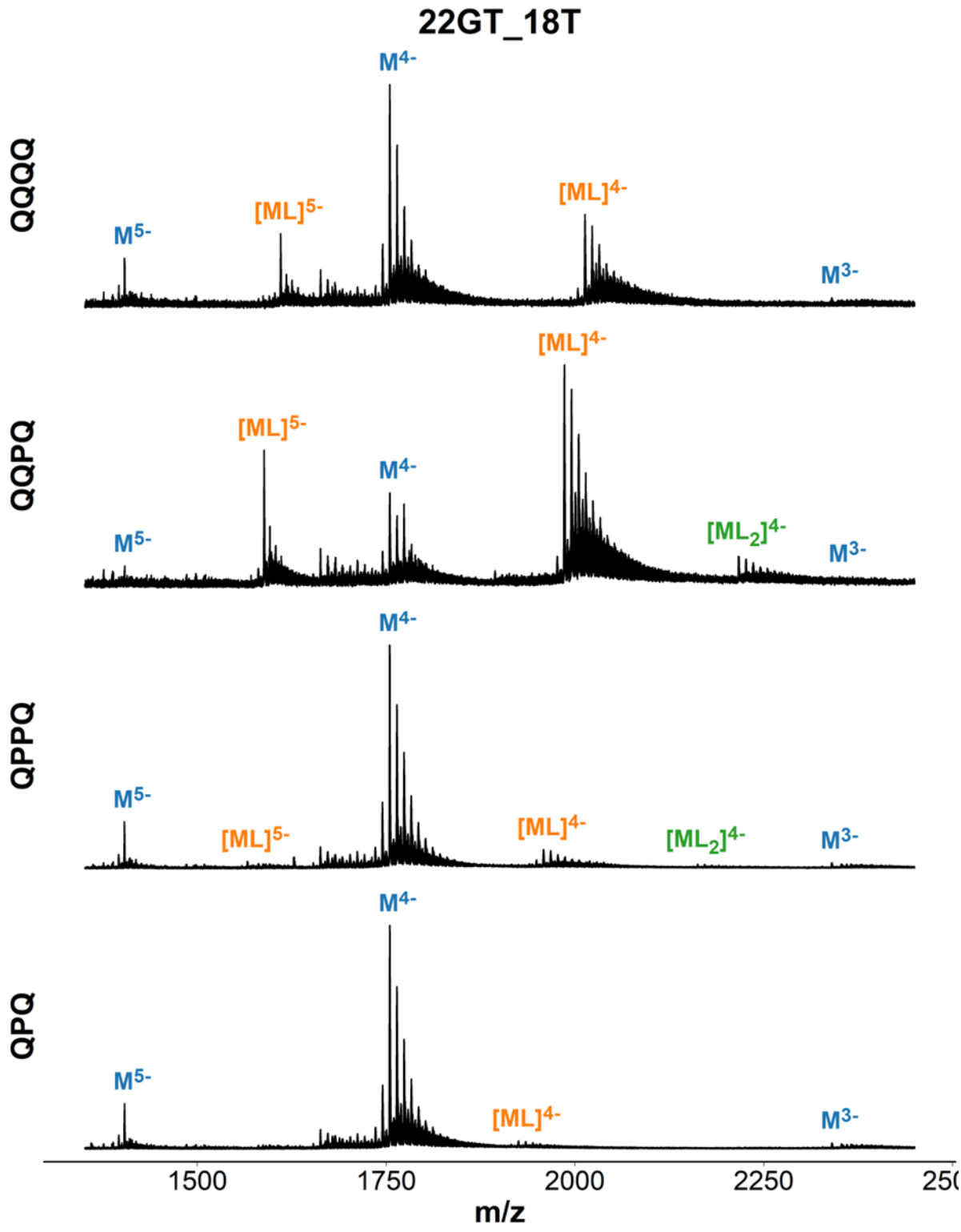


*Figure S79 Mass spectra of 22GT-18T (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

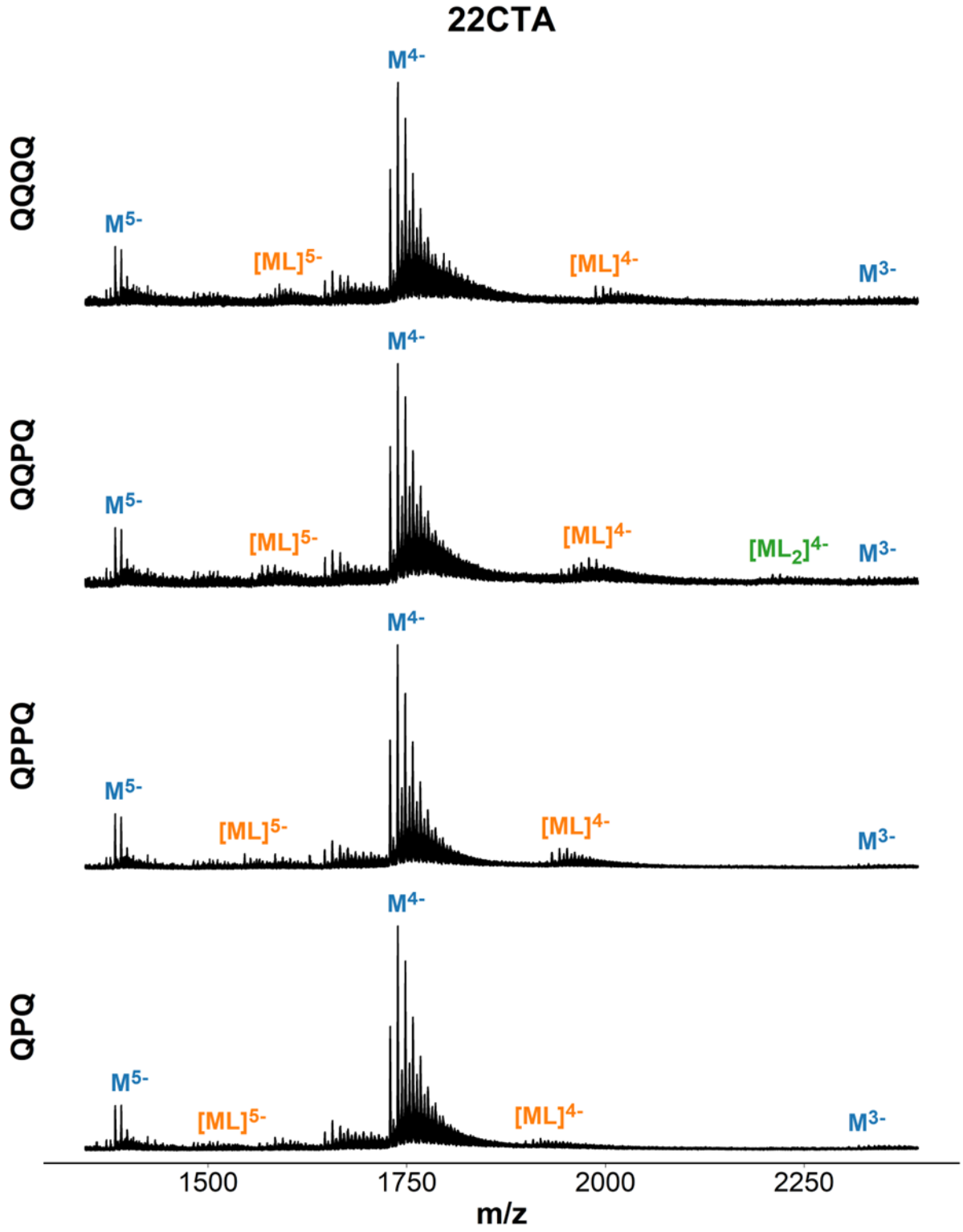


*Figure S80 Mass spectra of 22CTA (dA**GGG**CTA**GGG**CTA**GGG**CTA**GGG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

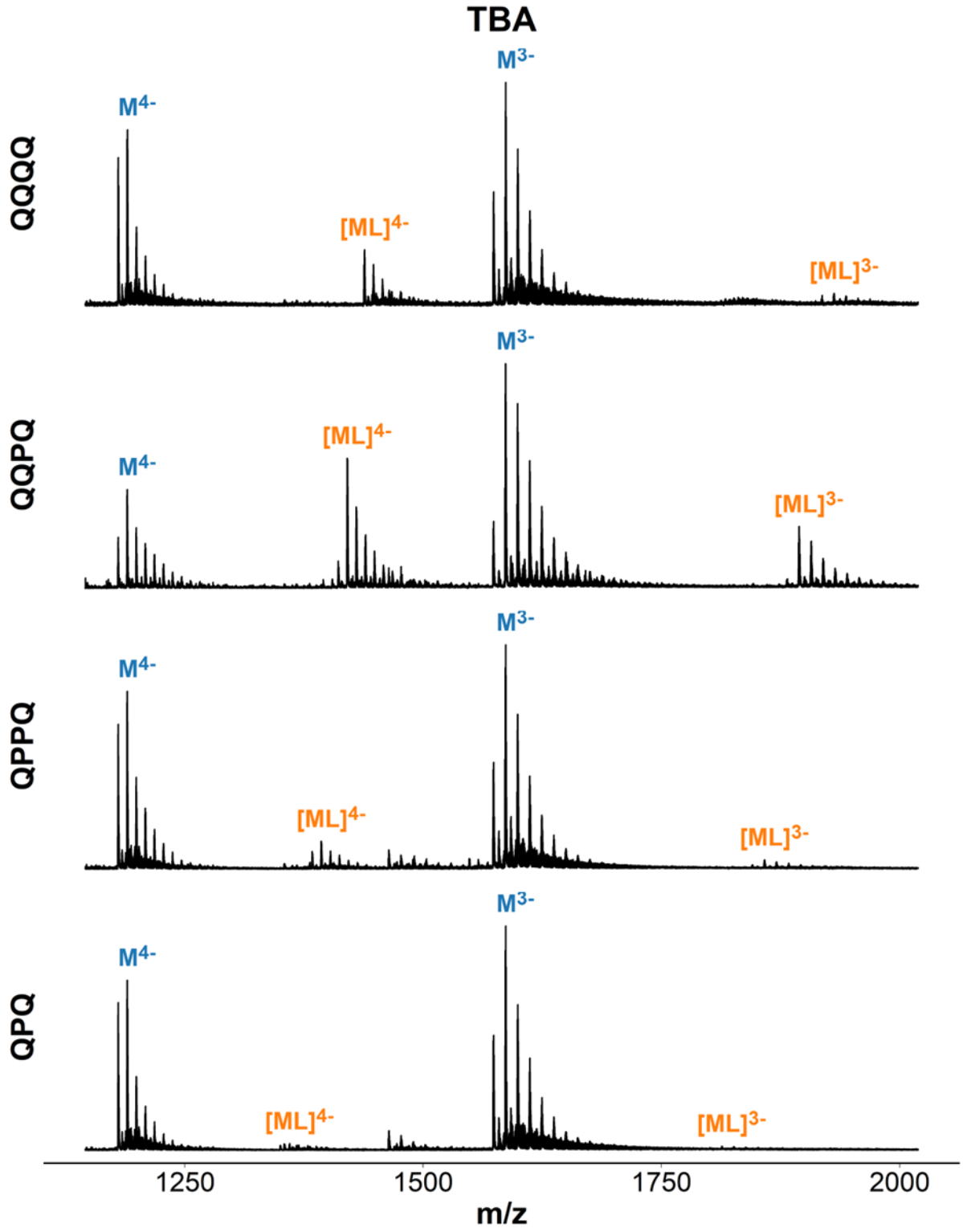


*Figure S81 Mass spectra of TBA (d**GG**TT**GG**T**G**T**GG**TT**GG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

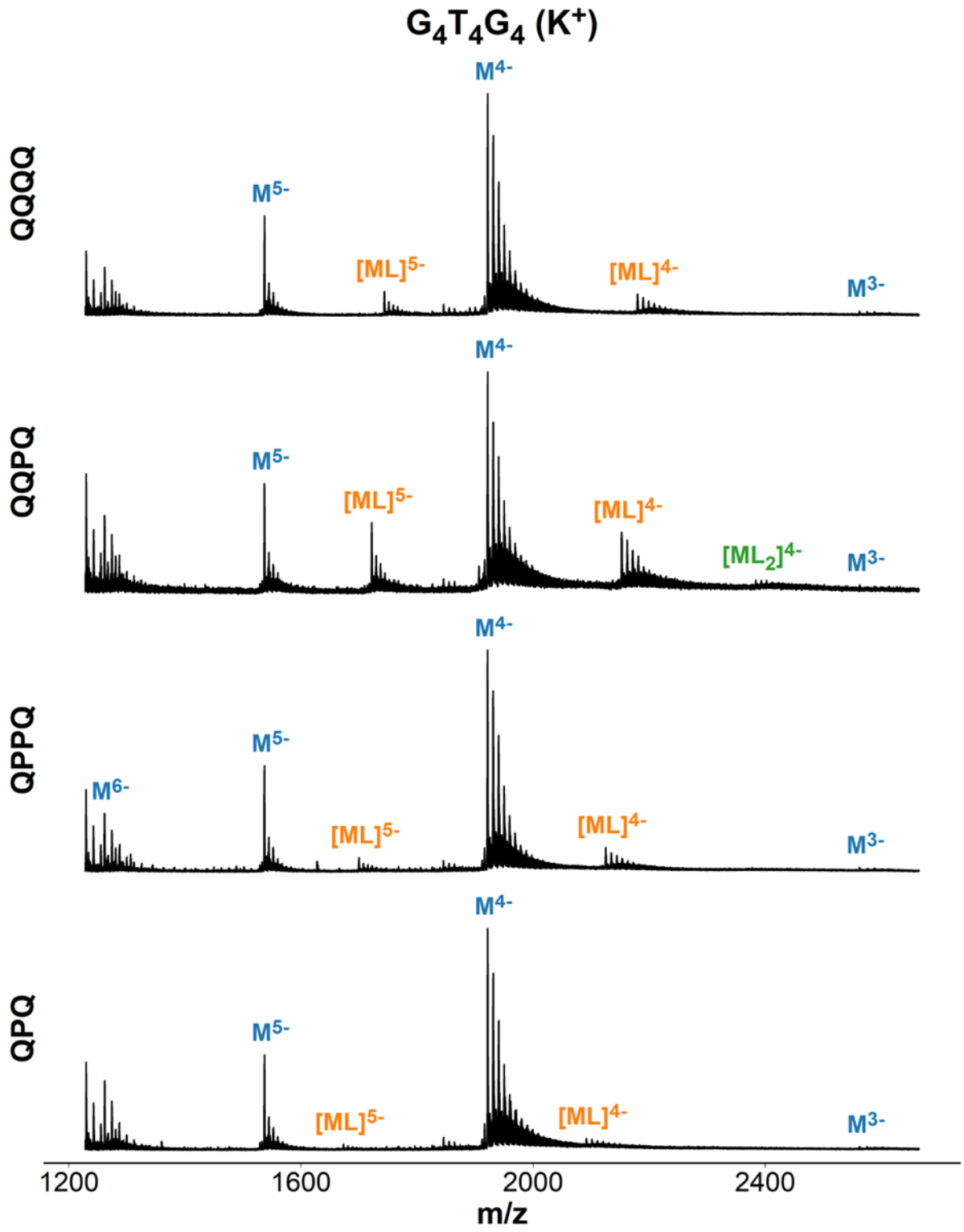


*Figure S82 Mass spectra of G4T4G4 ([d**GGGG**TTTT**GGGG**]$_2$) in presence of ligand. Samples contain 20 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

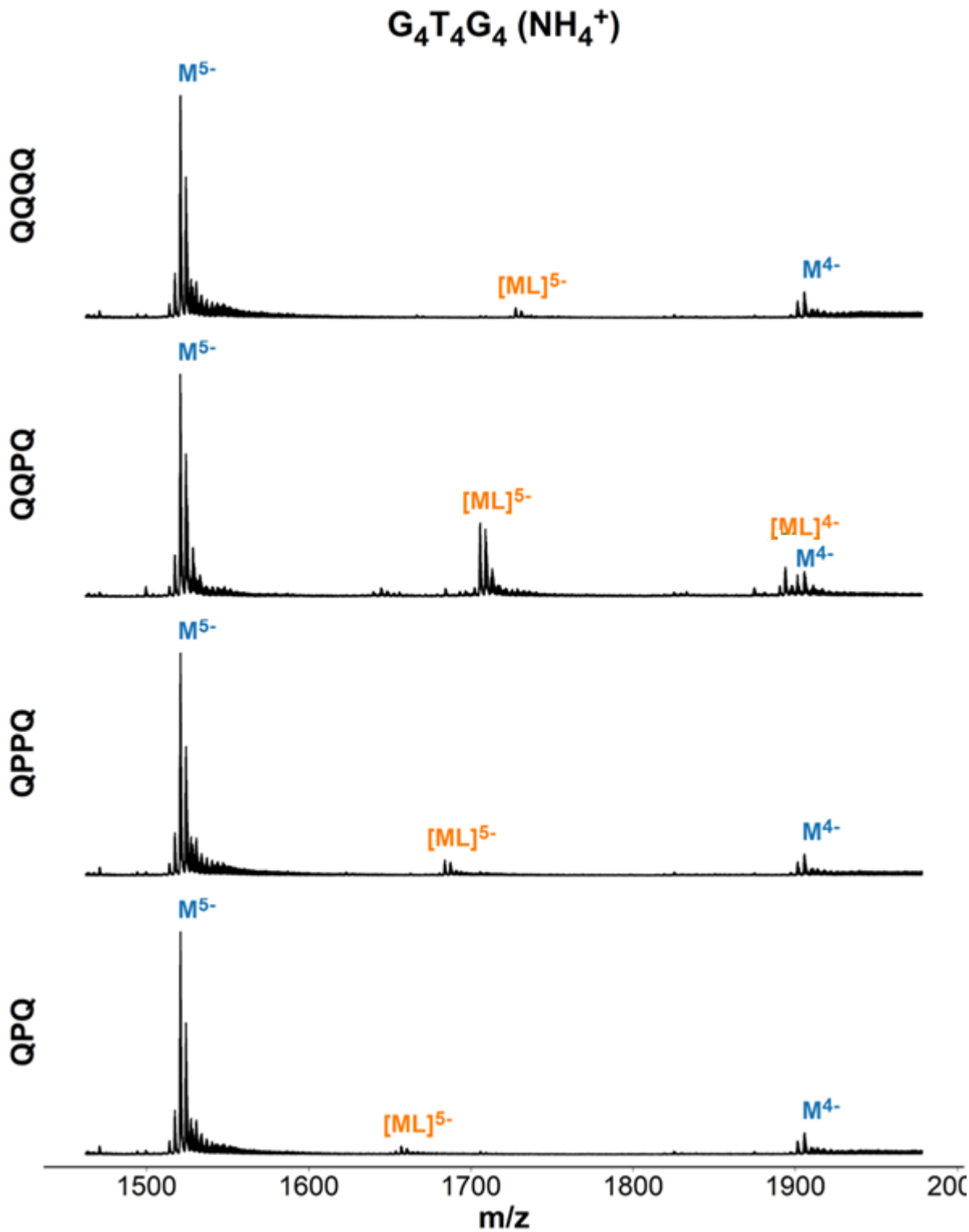


*Figure S83 Mass spectra of G4T4G4 ([d**GGGG**TTTT**GGGG**]$_2$) in presence of ligand. Samples contain 20 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).*

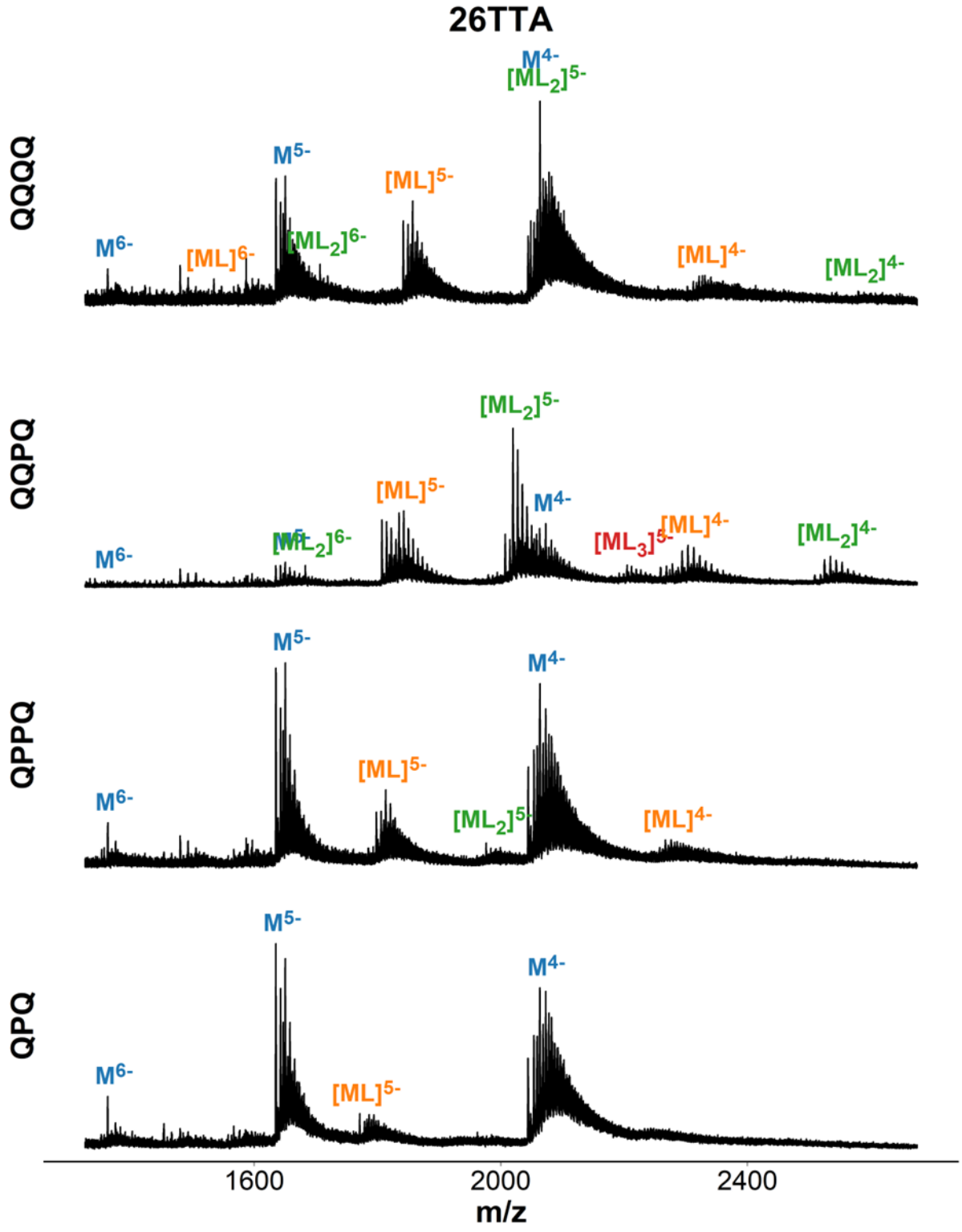


*Figure S84 Mass spectra of 26TTA (dTTA**GGG**TTA**GGG**TTA**GGG**TTA**GGG**TT) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

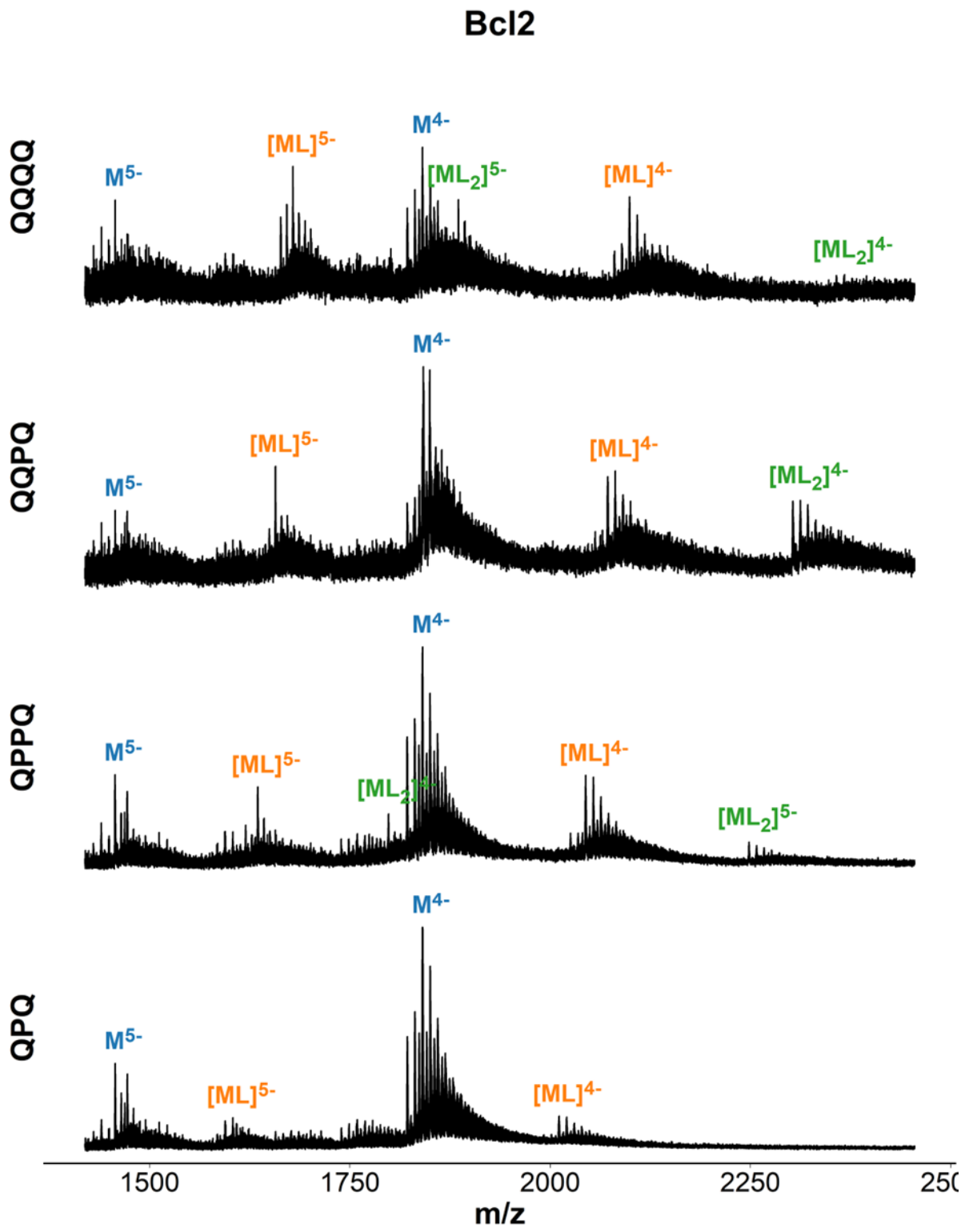


*Figure S85 Mass spectra of bcl2 (d**GGG**C**G**C**GGG**A**GG**AATT**GGG**C**GGG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

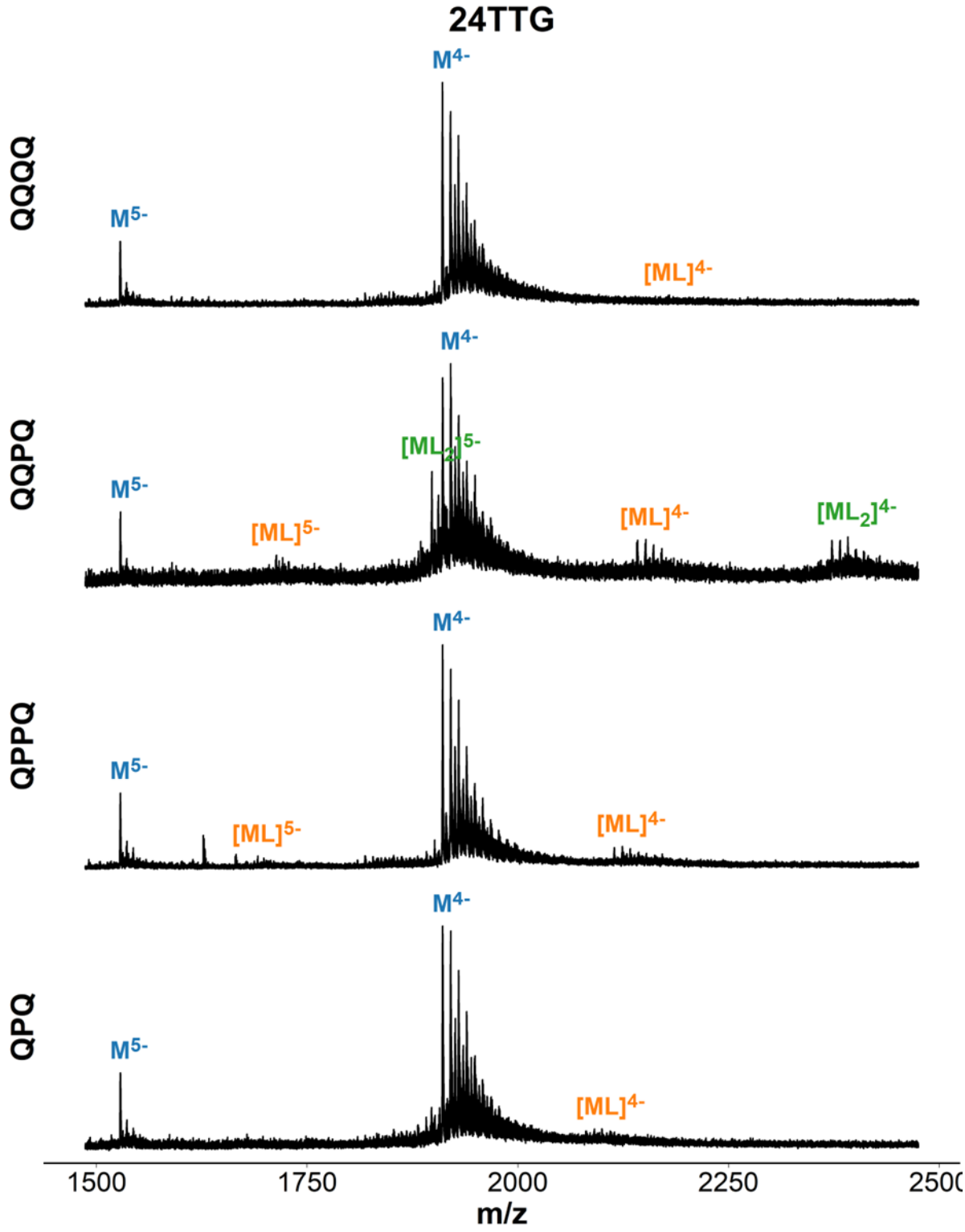


*Figure S86 Mass spectra of 24TTG (dTT**GGG**TTA**GGG**TTA**GGG**TTA**GGG**A) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

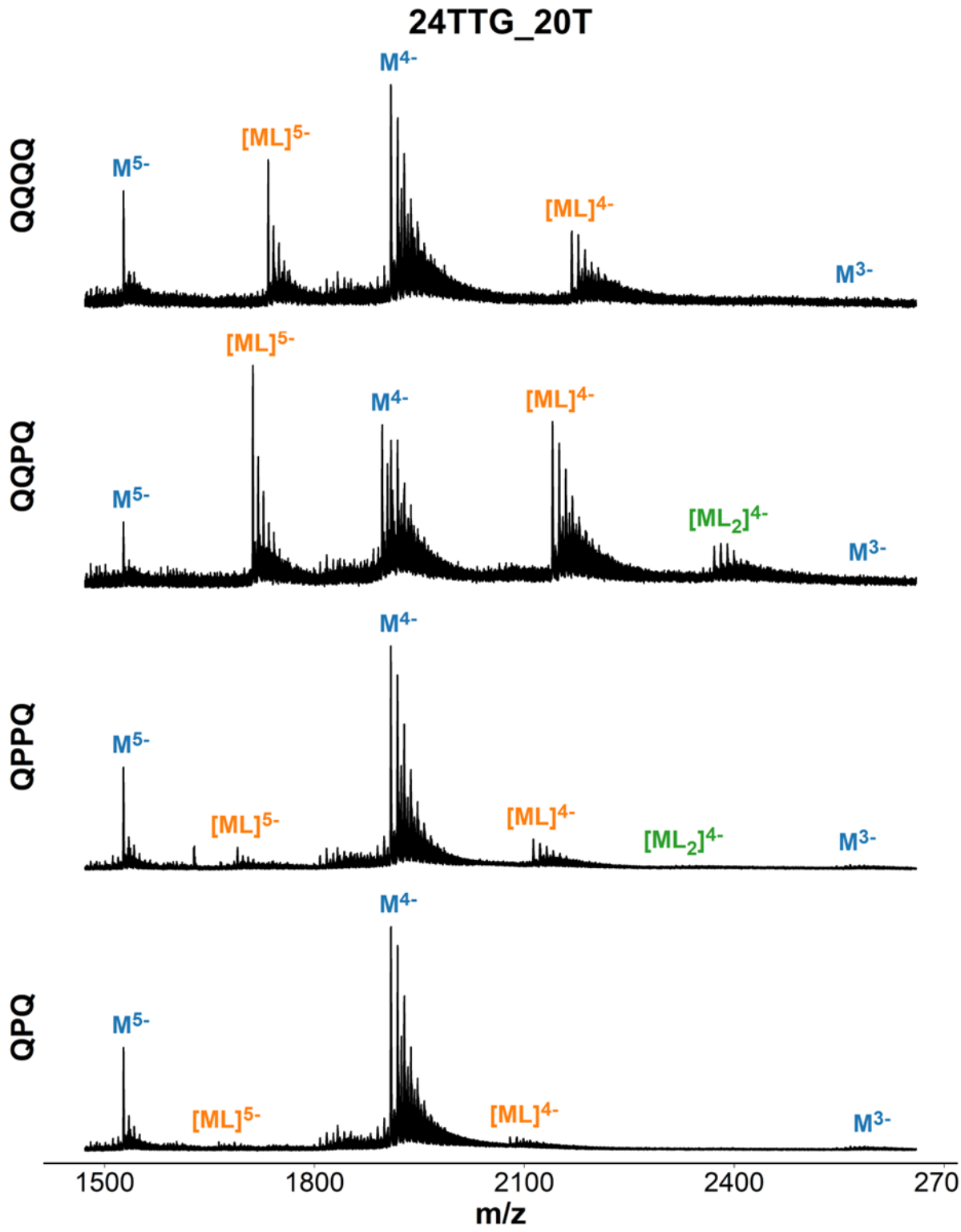


*Figure S87 Mass spectra of 24TTG-20T (dTT**GGG**TTA**GGG**TTA**GGG**TTT**GGG**A) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

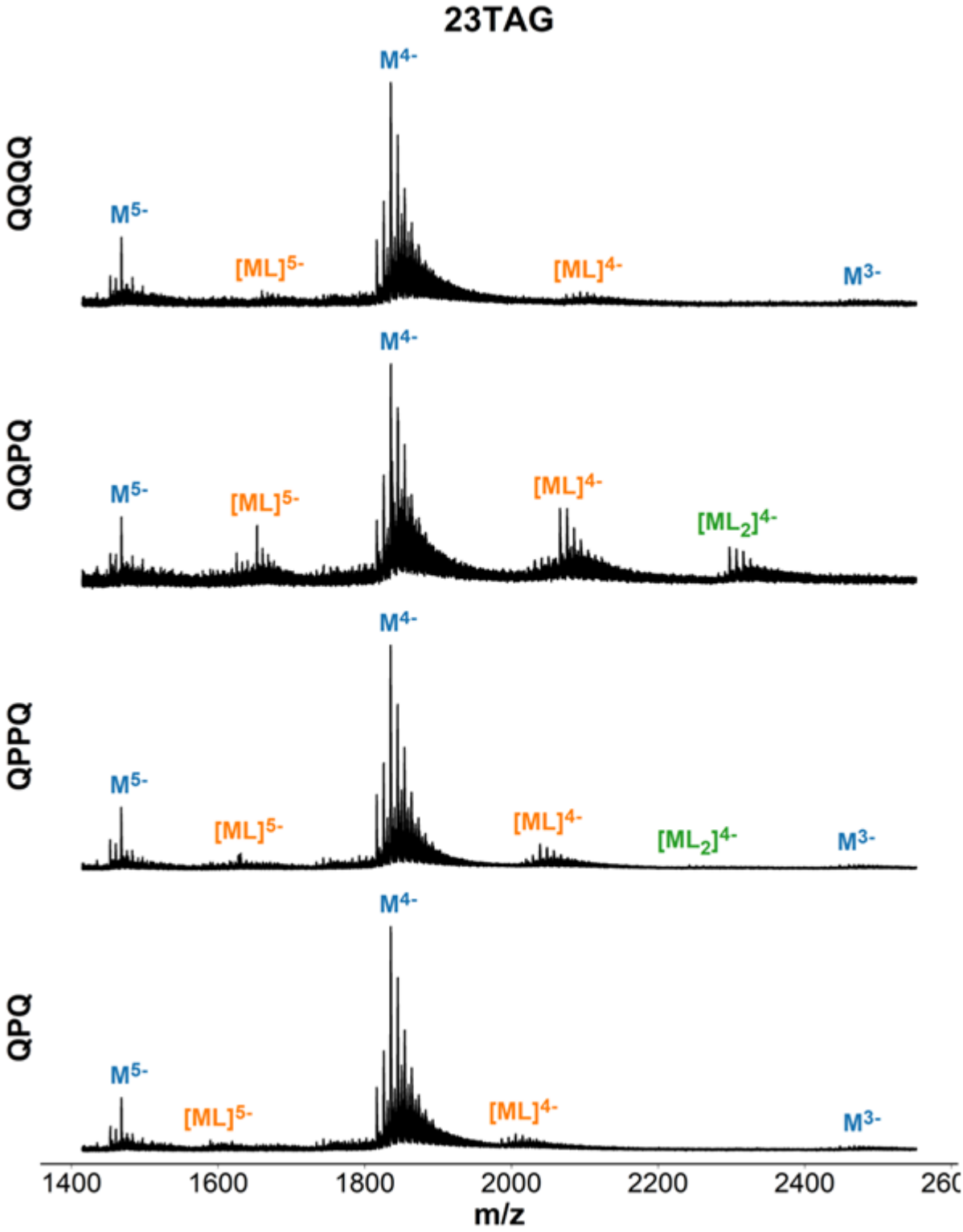


*Figure S88 Mass spectra of 23TAG (dTA**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

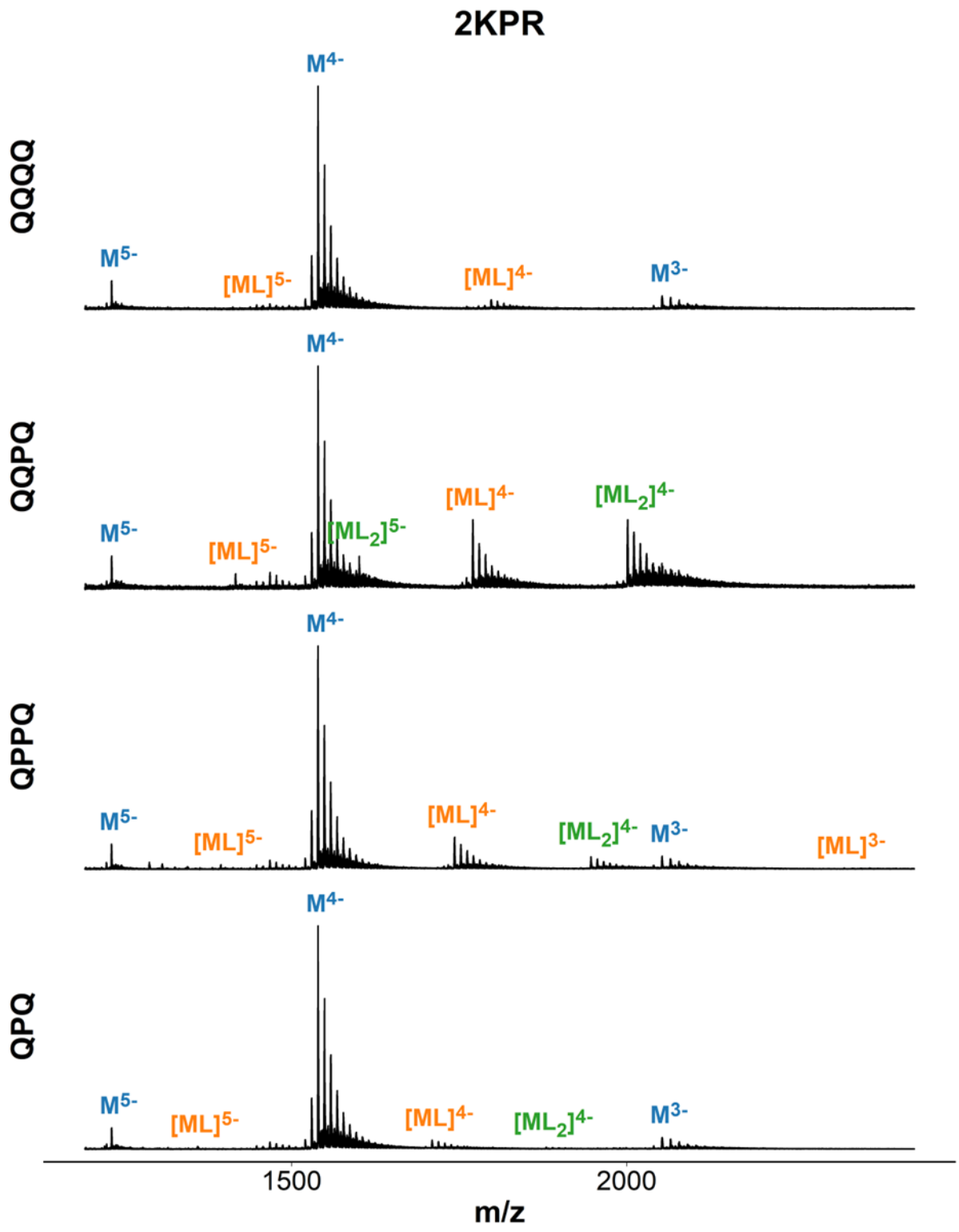


*Figure S89 Mass spectra of 2KPR (d**GGG**T**GGGG**AA**GGGG**T**GGG**T) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

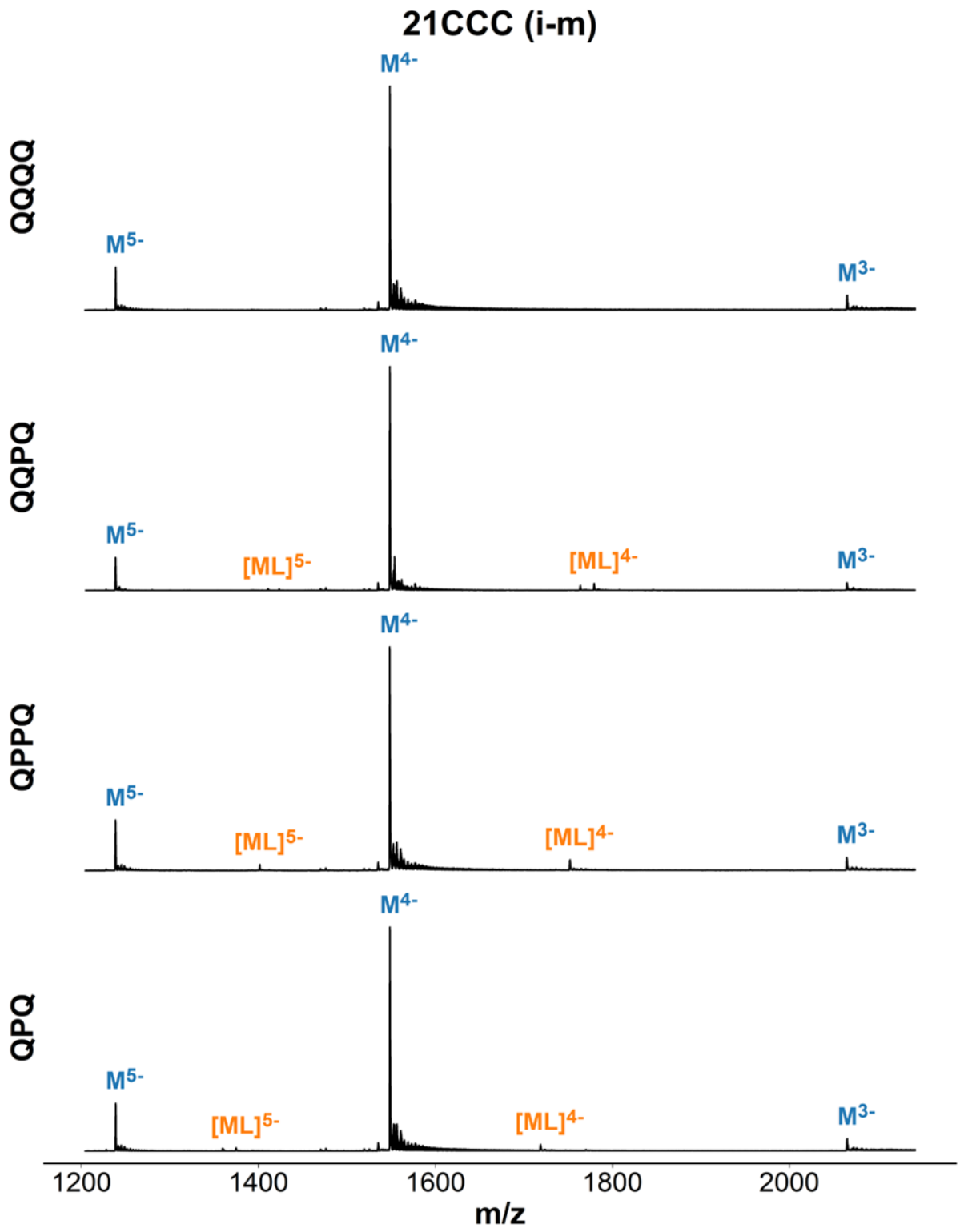


*Figure S90 Mass spectra of 21CCC (d**CCC**TAA**CCC**TAA**CCC**TAA**CCC**) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 5.5).*

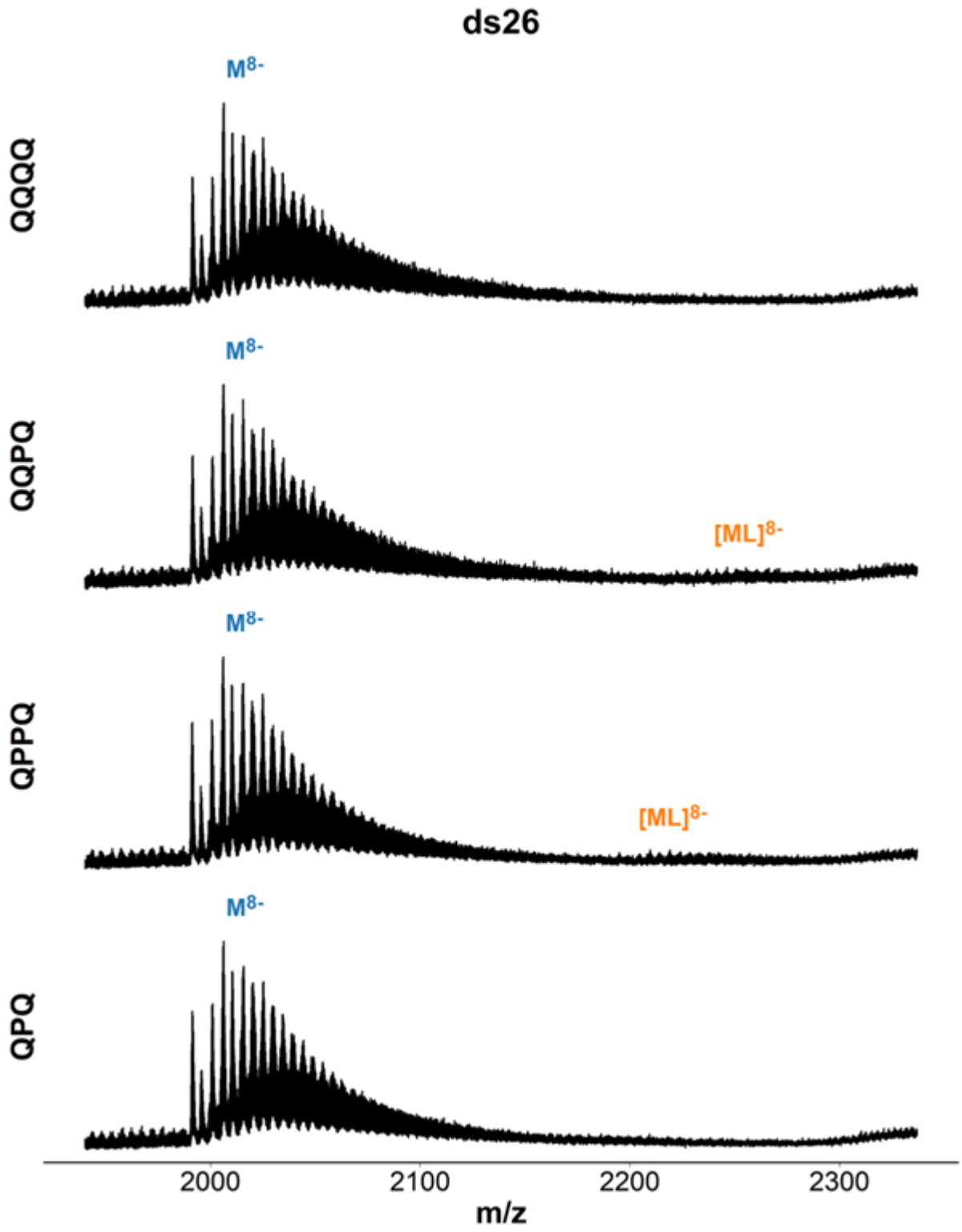


Figure S91 Mass spectra of ds26 ([dCAATCGGATCGAATTCGATCCGATTG]$_2$) in presence of ligand. Samples contain 20 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

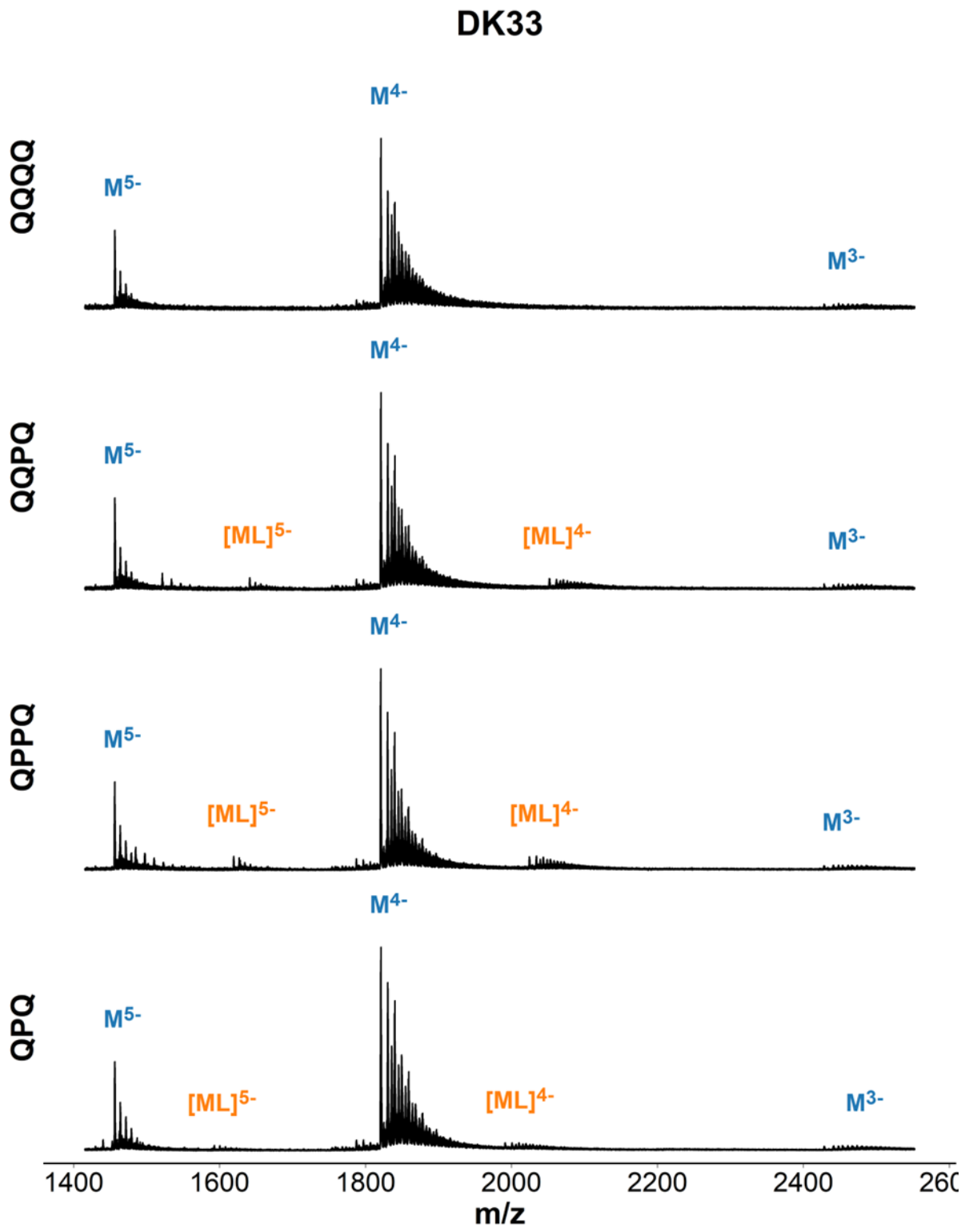


Figure S92 Mass spectra of DK-33 ([dCGTAAATTTACG]$_2$) in presence of ligand. Samples contain 20 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

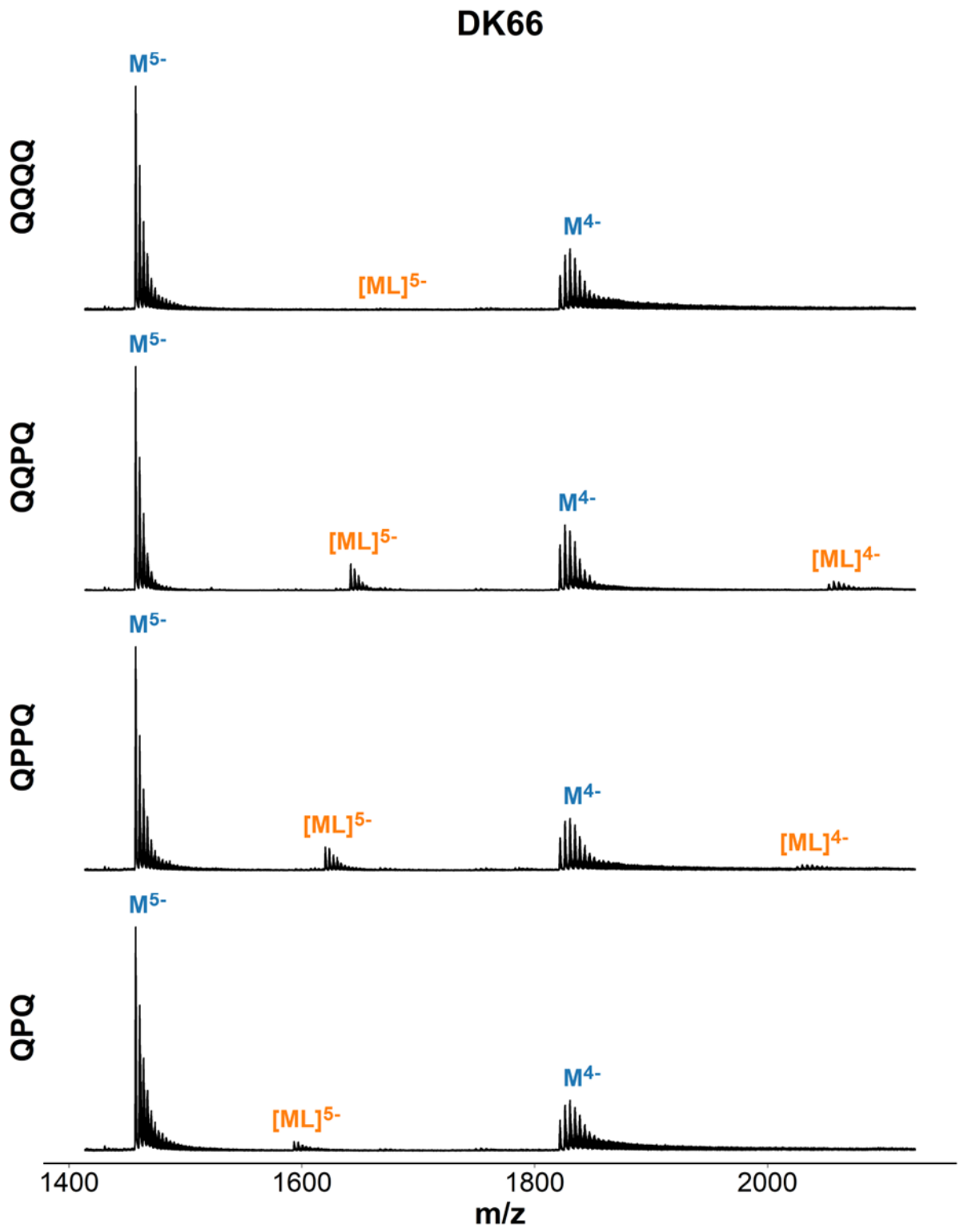


Figure S93 Mass spectra of DK-66 ([dCGCGAATTCGCG]$_2$) in presence of ligand. Samples contain 20 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).

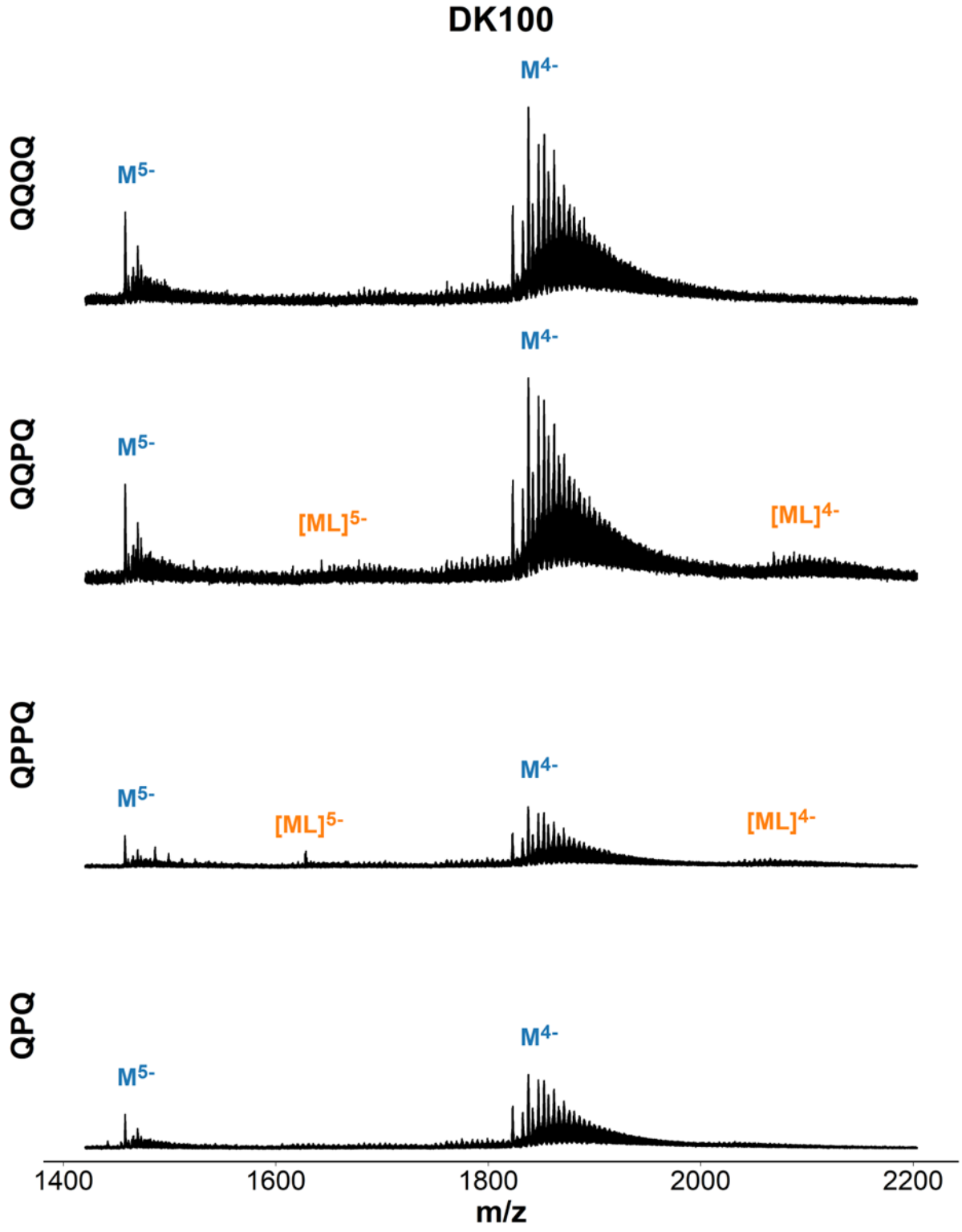


Figure S94 Mass spectra of DK-100 ([dCGCGGGCCCGCG]$_2$) in presence of ligand. Samples contain 20 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

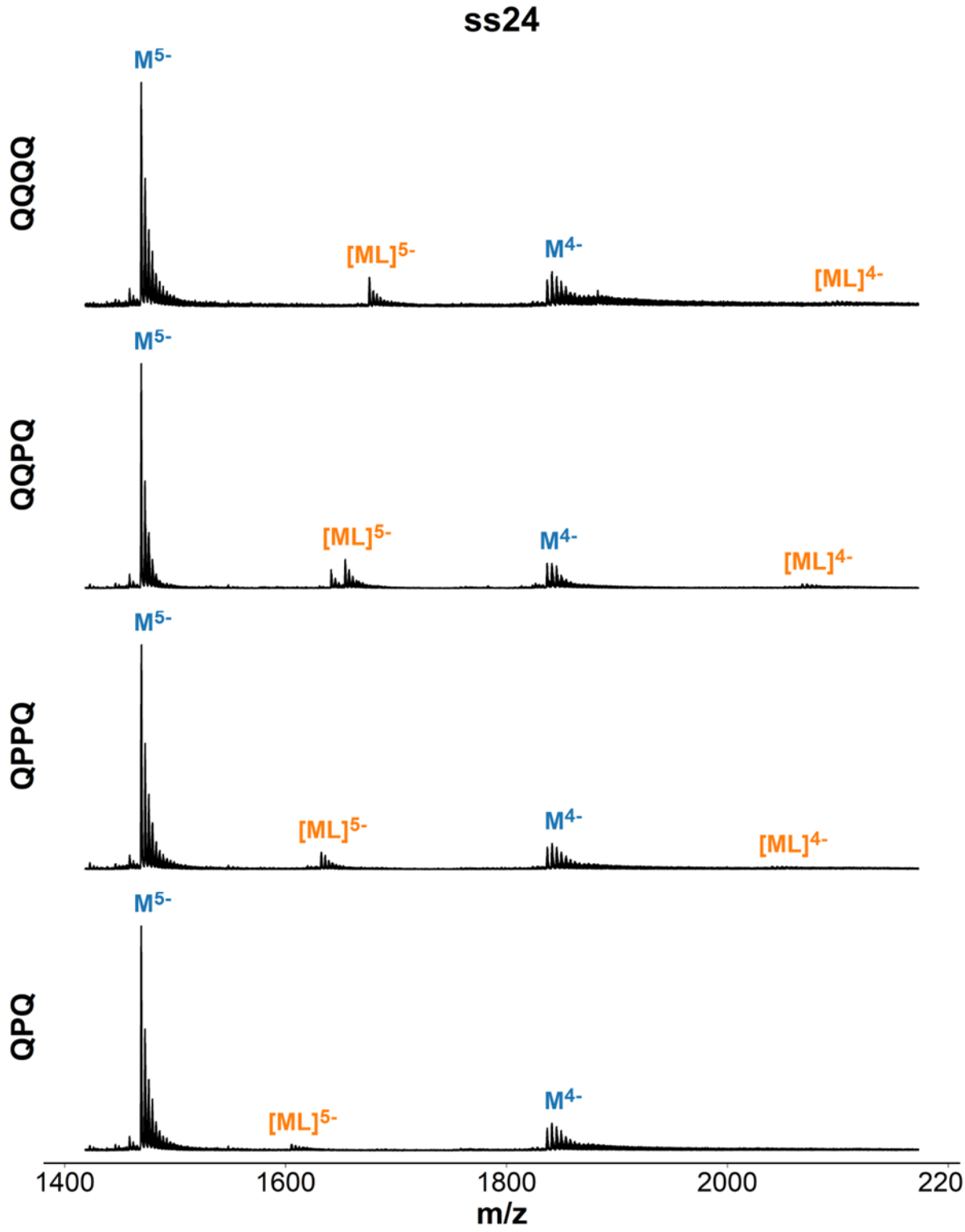


Figure S95 Mass spectra of ss24 (dTGCCATGCTACTGAGATGACGCTA) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).

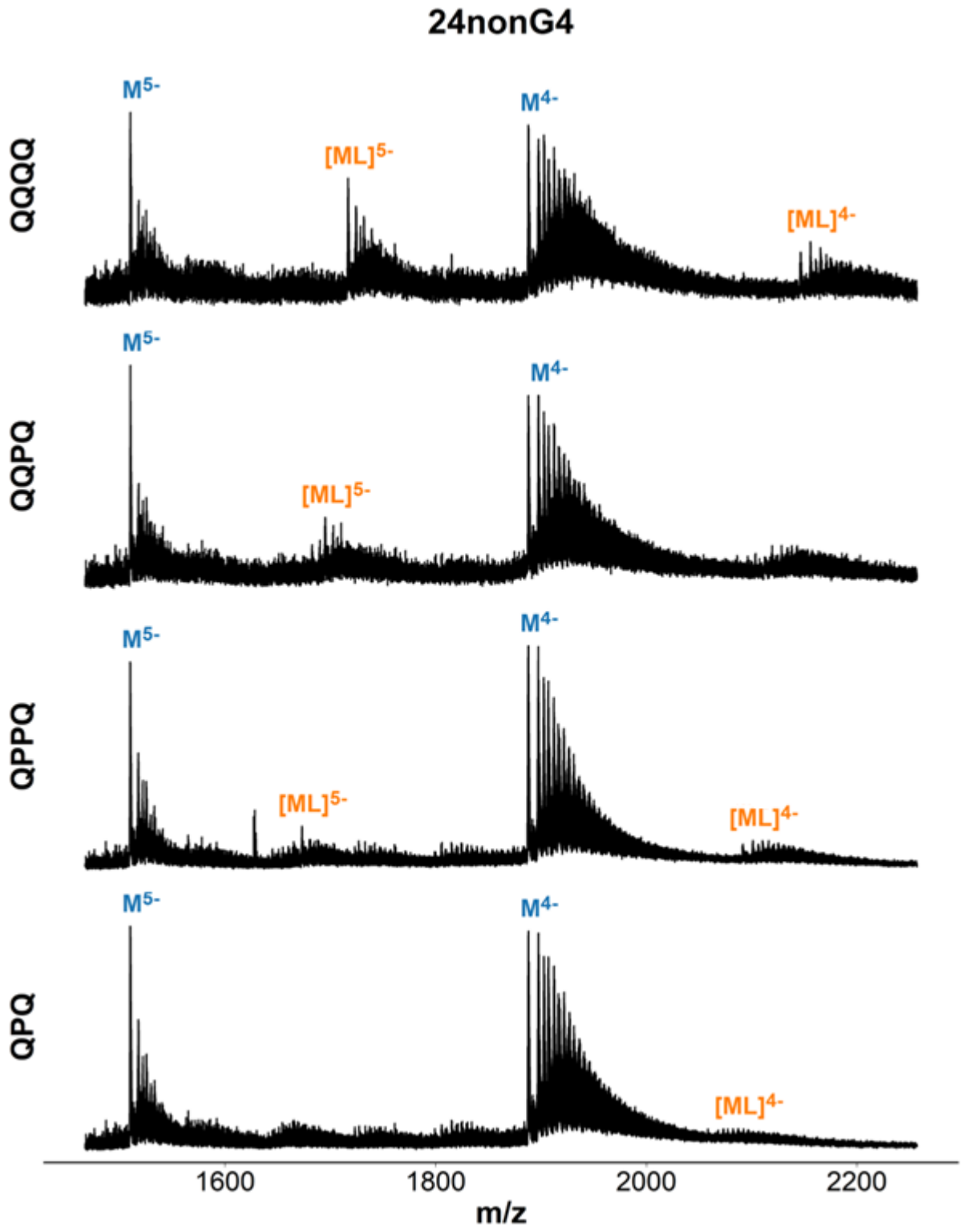


Figure S96 Mass spectra of 24nonG4 (dTGGGATGCGACAGAGAGGACGGGA) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

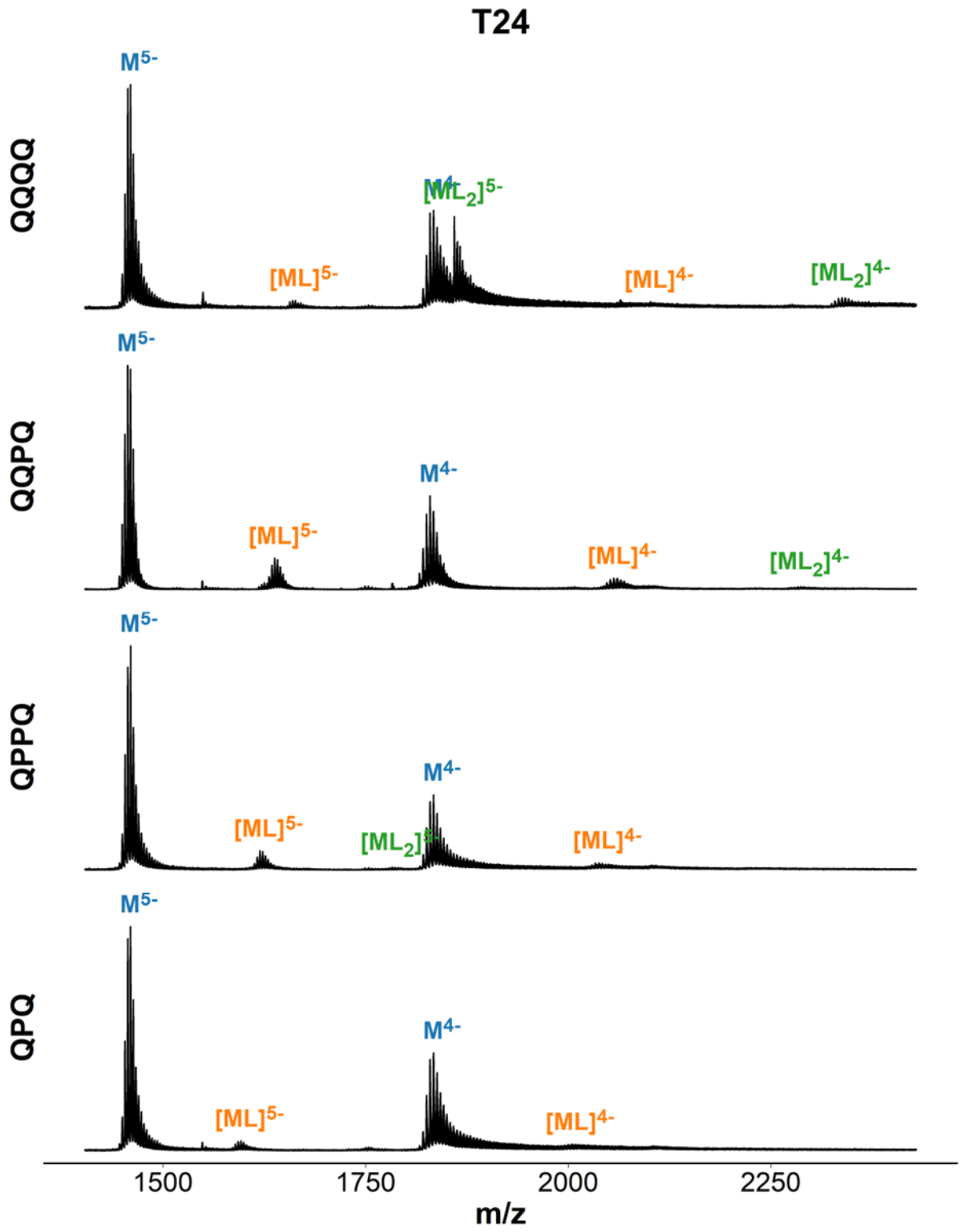


Figure S97 Mass spectra of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).

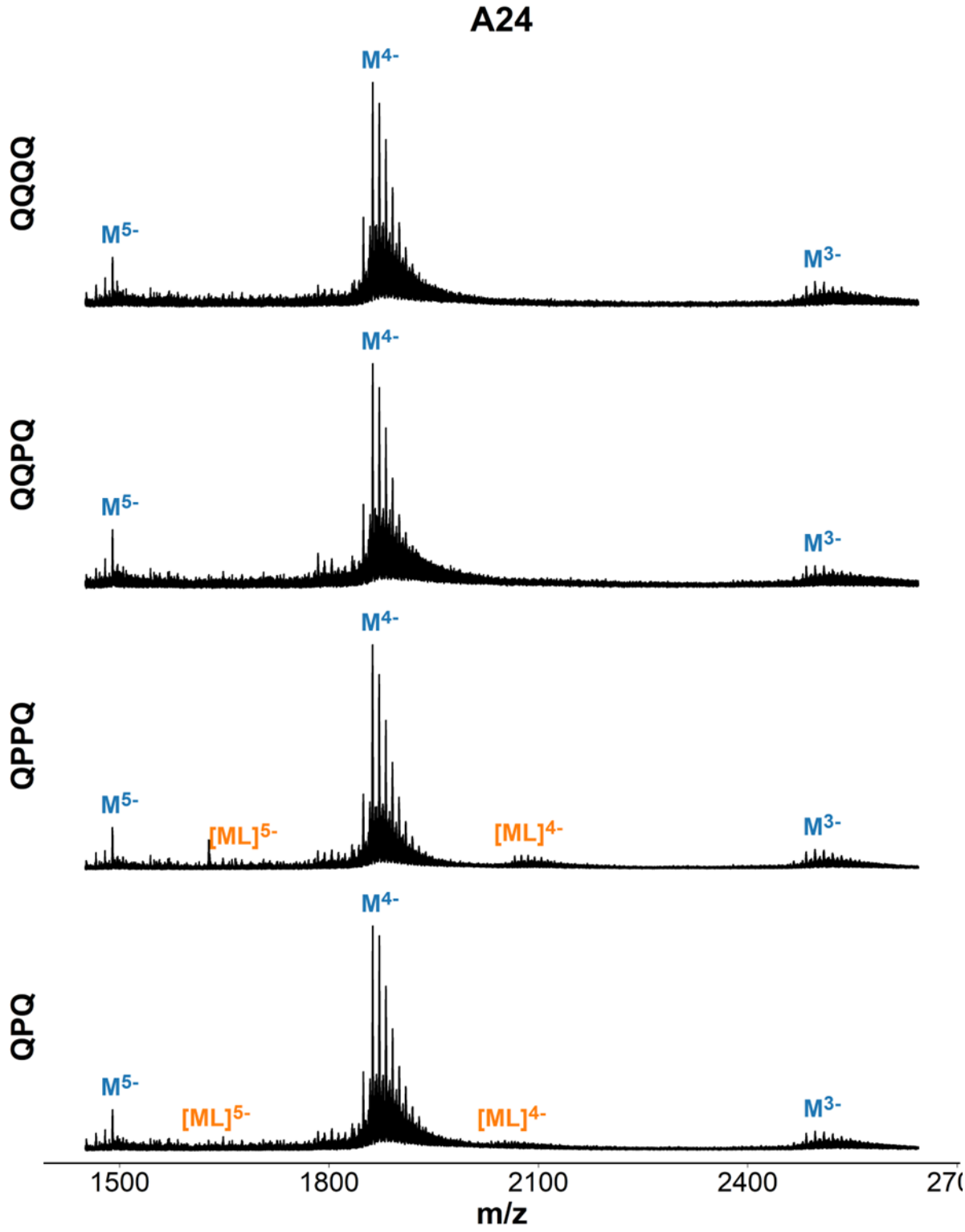


Figure S98 Mass spectra of A24(dAAAAAAAAAAAAAAAAAAAAAAAA) in presence of ligand. Samples contain 10 μM DNA, 20 μM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

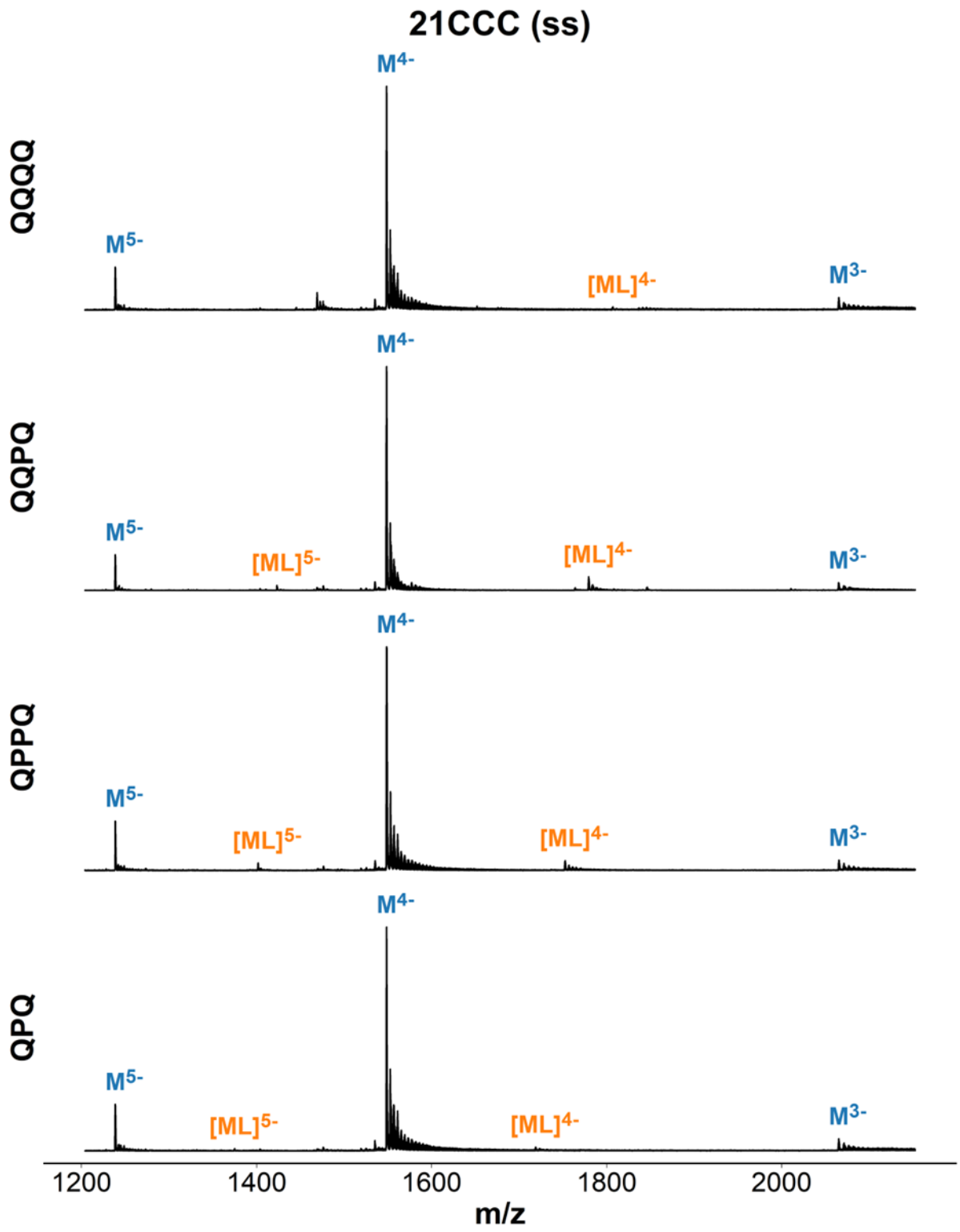


Figure S99 Mass spectra of 21CCC (dCCCTAACCCTAACCCTAACCC) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).

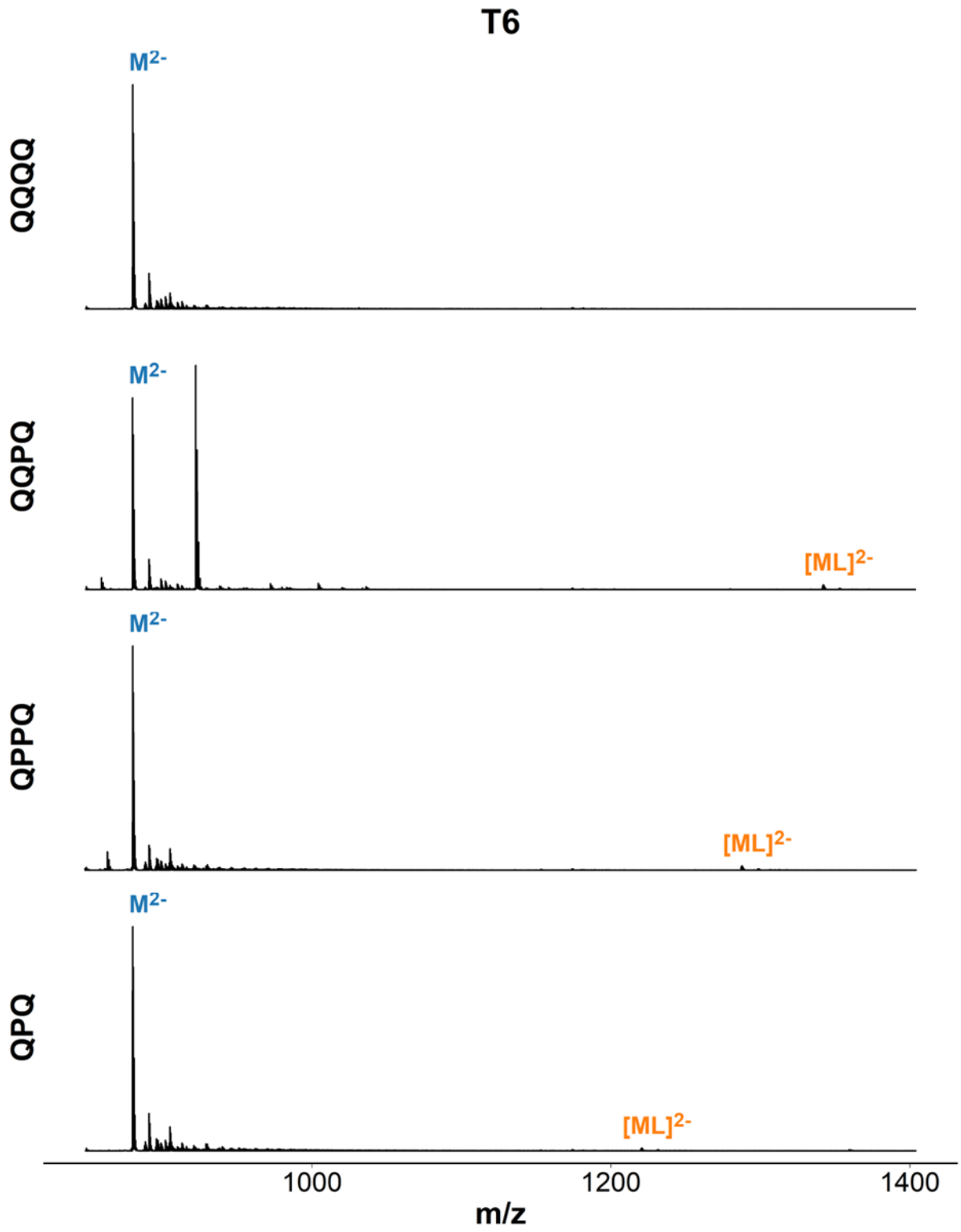


Figure S100 Mass spectra of T6 (dTTTTTT) in presence of ligand. Samples contain 10 µM DNA, 20 µM ligand, 150 mM ammonium acetate (pH 6.8).

ESI-MS titrations: K_D values and response factor estimates

Table S3 All K_D values obtained from ESI-MS titration of 6 DNA sequences with 7 foldamer ligands. K_D < 5 µM are highlighted in blue, K_D > 100 µM are faded out.

Dissociation constant estimates (µM)

	K_{D1}		K_{D2}		K_{D3}		K_{D4}	
	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.
2LK7								
QQPQ	0.18	*0.079*	2.1	*0.69*				
1XAV								
Q3	0.11	*0.074*	1.2	*0.94*	4.1	*2.5*		
QPQ	13	*0.75*	110	*43*				
QQQQ	0.075	*0.059*	0.83	*0.49*				
QQPQ	0.41	*0.29*	1.9	*1.2*				
QPPQ	6.9	*2*	9.7	*5.7*				
Q5	0.25	*0.2*	0.7	*0.53*				
Q8	28	*6.7*						
222T								
Q3	0.1	*0.1*	2.8	*3.1*	6.1	*11*	64	*440*
QPQ	1.8	*0.25*	50	*13*				
QQQQ	0.088	*0.11*	1.6	*1.1*				
QQPQ	0.027	*0.081*	0.064	*0.19*				
QPPQ	0.1	*0.093*	1.4	*0.85*				
Q5	0.011	*0.014*	0.4	*0.35*				
Q8	7.8	*1.1*						
5YEY								
Q3	180	*28*						
QPQ	1900	*780*						
QQQQ	6.6	*1.3*						
QQPQ	0.54	*0.098*	200	*70*				
QPPQ	110	*23*	310	*760*				
Q5	3.9	*0.37*						
21G								
Q3	78	*19*	1500	*21000*				
QPQ	170	*26*						
QQQQ	330	*23*						
QQPQ	8.6	*2.3*	14	*9.2*				
QPPQ	160	*20*	1000	*4500*				
Q5	210	*74*	47	*45*				
SS24								
Q3	12	*6.2*	28	*87*	15	*63*		
QPQ	580	*110*						
QQQQ	180	*14*						
QQPQ	150	*21*						
QPPQ	270	*30*						
Q5	130	*42*	44	*41*				
Q8	210	*45*						
T24								
Q3	3400	*8200*	0.36	*0.91*	260000	*1.2e+08*	0.0092	*4.2*
QQQQ	350	*3400*	0.53	*5.5*				
QQPQ	270	*24*						
QPPQ	190	*25*						
Q5	180	*490*	1	*3*				
Q8	290	*87*						

Table S4 Estimated response factors of complex species ML, ML_2, ML_3, ML_4 relative to the unbound DNA (M with R = 1) using dT_6 as an internal calibrant. The calculation method was previously described in [50].

Response factors estimates

	K_{D1}		K_{D2}		K_{D3}		K_{D4}	
	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.	ESTIMATE	STD.DEV.
2LK7								
QQPQ	0.987	0.099	0.968	0.071				
1XAV								
Q3	0.79	0.14	1.2	0.39	0.85	0.15		
QPQ	0.985	0.058	1.2	0.37				
QQQQ	1	0.15	0.95	0.067				
QQPQ	1	0.22	1	0.11				
QPPQ	1.1	0.31	0.91	0.2				
Q5	1.1	0.24	0.905	0.073				
Q8	1	0.2						
222T								
Q3	0.85	0.18	1.2	0.87	0.8	1.1	0.6	3.5
QPQ	0.944	0.052	1.1	0.2				
QQQQ	1	0.24	0.94	0.11				
QQPQ	1	0.36	0.998	0.088				
QPPQ	1	0.17	0.974	0.092				
Q5	0.95	0.13	0.948	0.065				
Q8	0.997	0.077						
5YEY								
Q3	0.97	0.18						
QPQ	1	0.58						
QQQQ	0.944	0.069						
QQPQ	1	0.028	1.2	0.39				
QPPQ	0.96	0.21	1	2.7				
Q5	0.974	0.026						
21G								
Q3	0.96	0.27	1	17				
QPQ	0.98	0.18						
QQQQ	0.997	0.087						
QQPQ	0.91	0.26	0.99	0.28				
QPPQ	0.99	0.13	1	5.3				
Q5	0.84	0.32	1.2	0.73				
SS24								
Q3	0.71	0.41	1	3.2	0.8	1.2		
QPQ	0.99	0.26						
QQQQ	0.997	0.088						
QQPQ	1	0.16						
QPPQ	1	0.14						
Q5	0.76	0.28	1.2	0.69				
Q8	0.98	0.26						
T24								
Q3			1.1	0.2			0.64	0.44
QQQQ	1	12	0.9	0.32				
QQPQ	1	0.11						
QPPQ	0.98	0.16						
Q5	0.6	1.9	0.95	0.18				
Q8	1	0.39						

ESI-MS titrations: Mass spectra

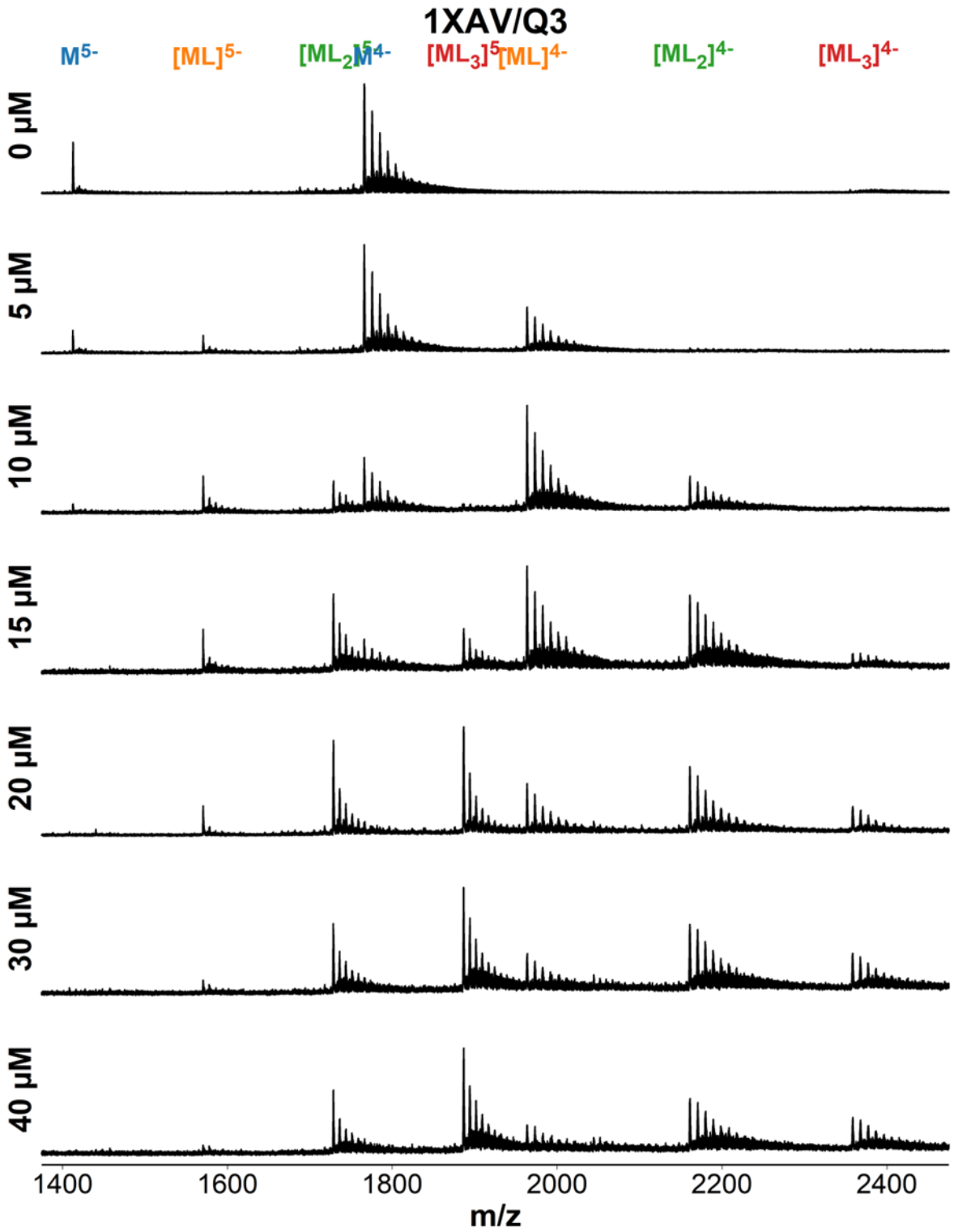


*Figure S101 ESI-MS titration of 1XAV (dT**GA**GGG**T**GGG**TA**GGG**T**GGG**TAA**) with foldamer QQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

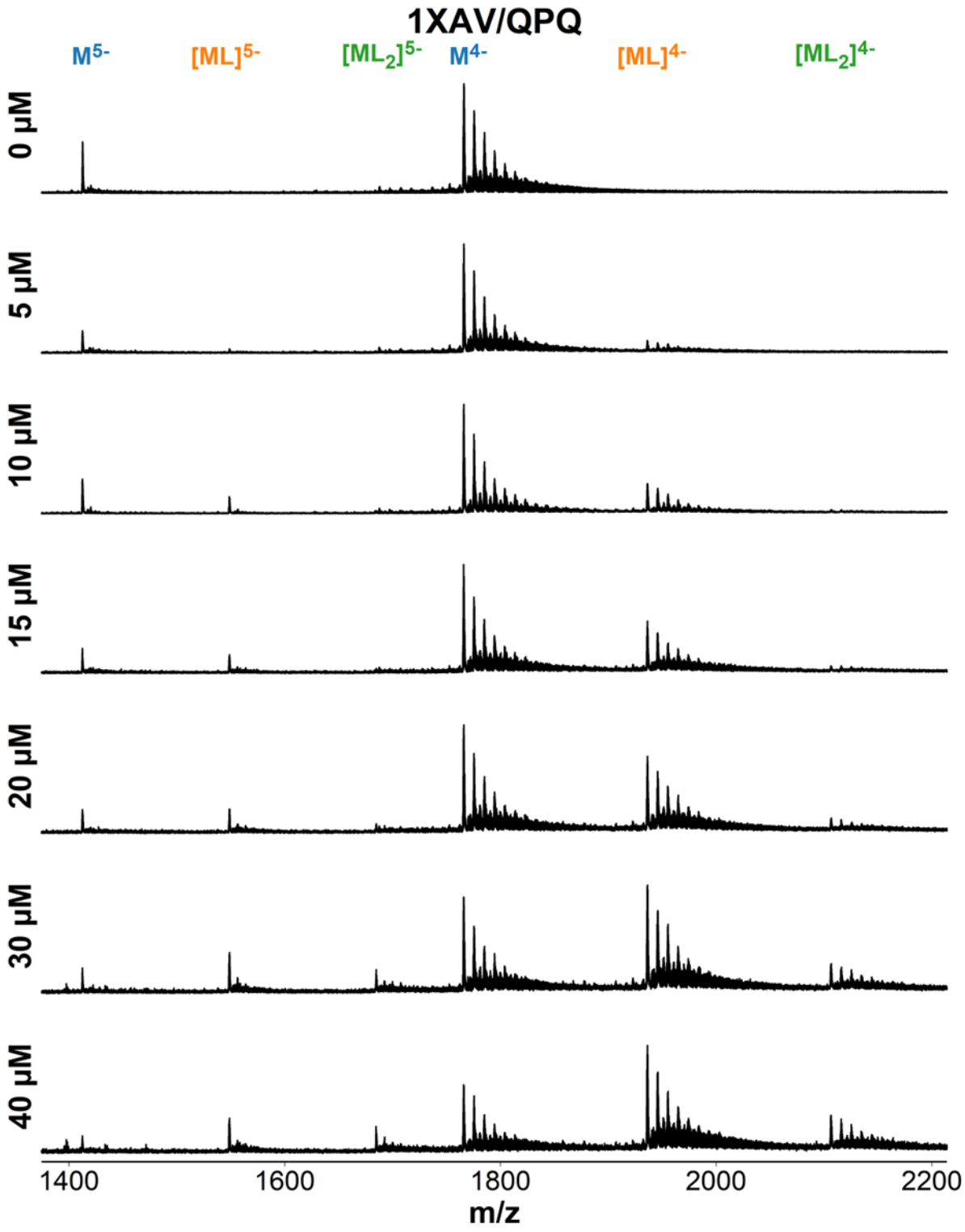


*Figure S102 ESI-MS titration of 1XAV (dT**G**A**GGG**T**GGG**TA**GGG**T**GGG**TAA) with foldamer QPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

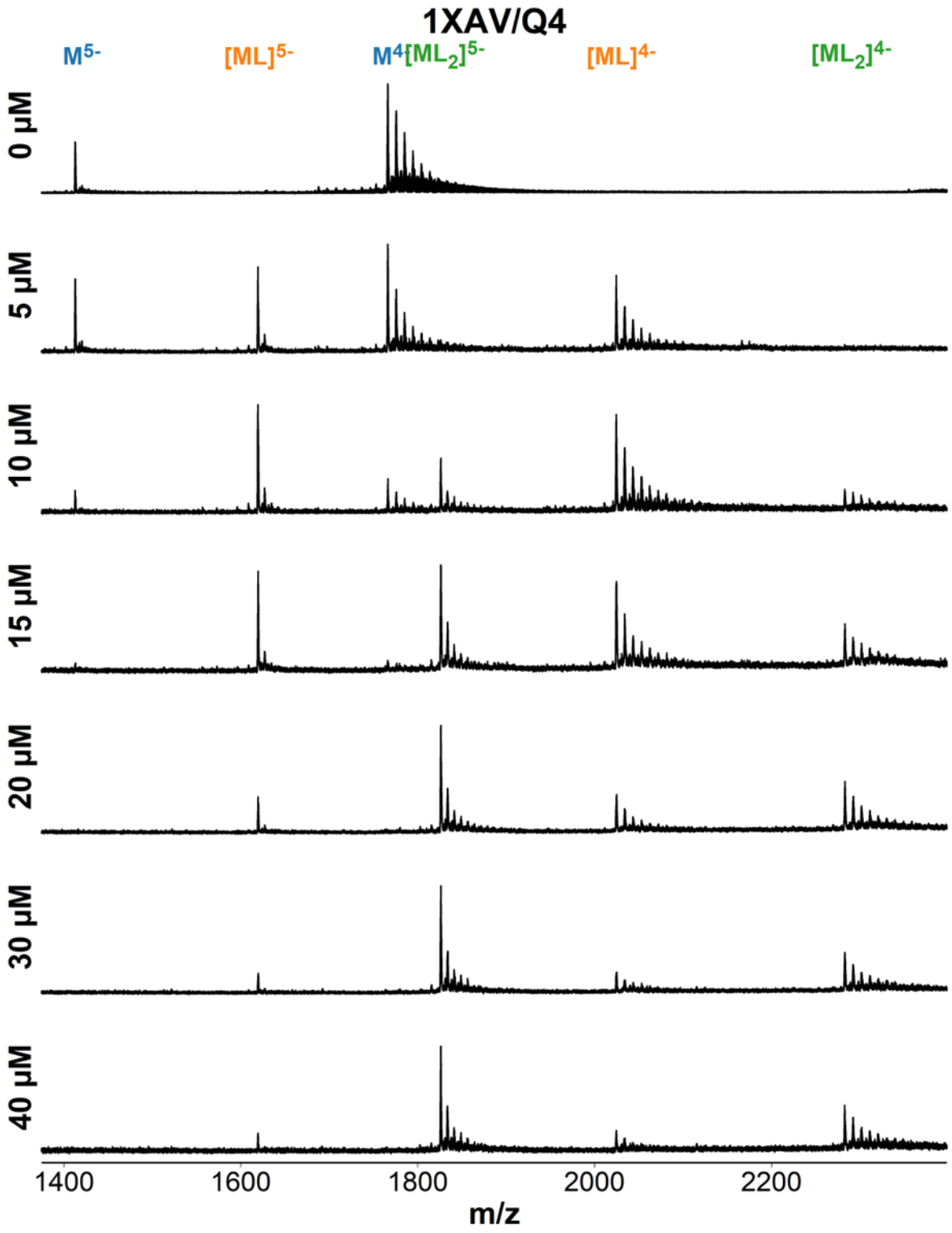


*Figure S103 ESI-MS titration of 1XAV (dT**G**A**GGG**T**GGG**TA**GGG**T**GGG**TAA) with foldamer QQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

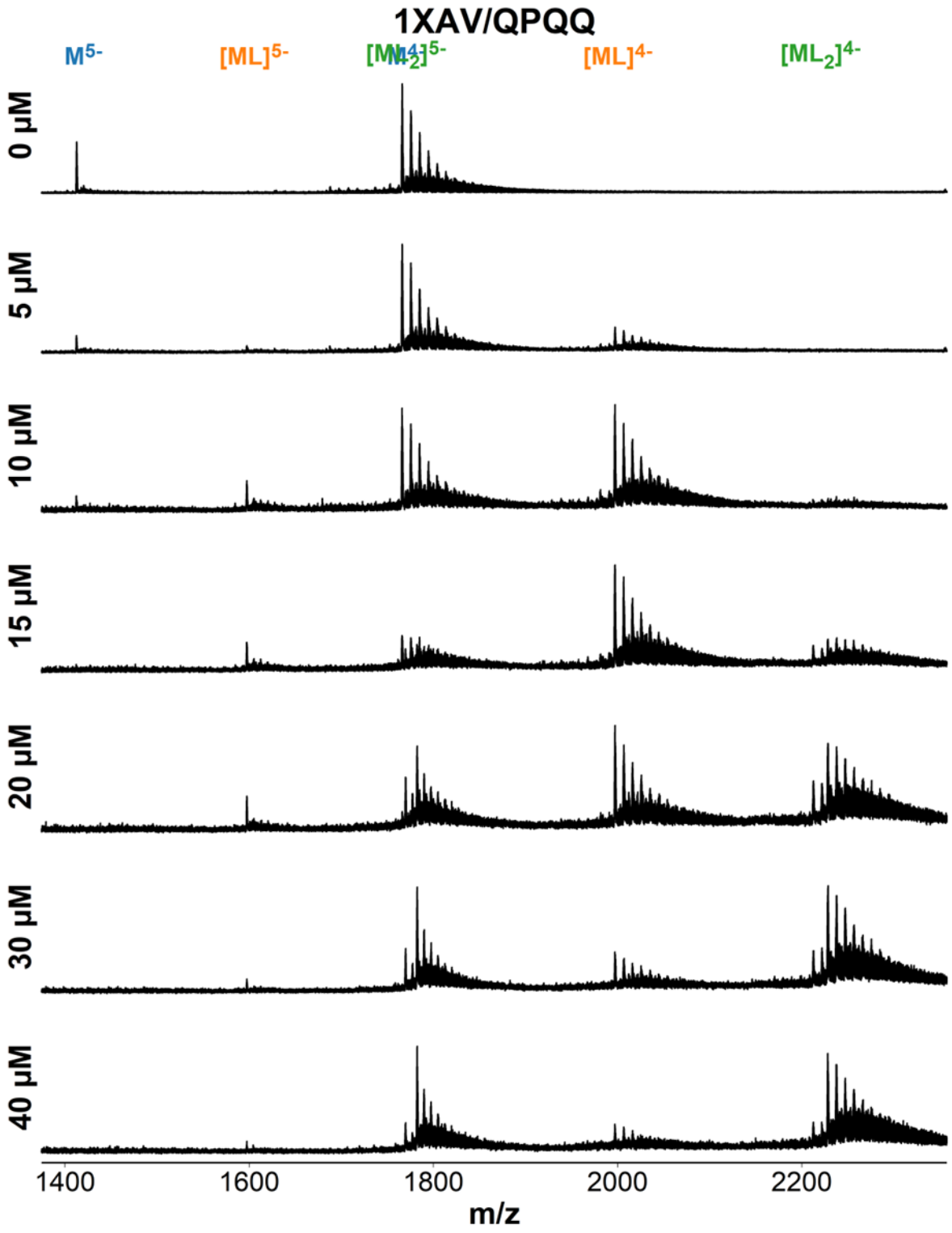


*Figure S104 ESI-MS titration of 1XAV (dT**G**A**GGG**T**GGG**TA**GGG**T**GGG**TAA) with foldamer QQPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

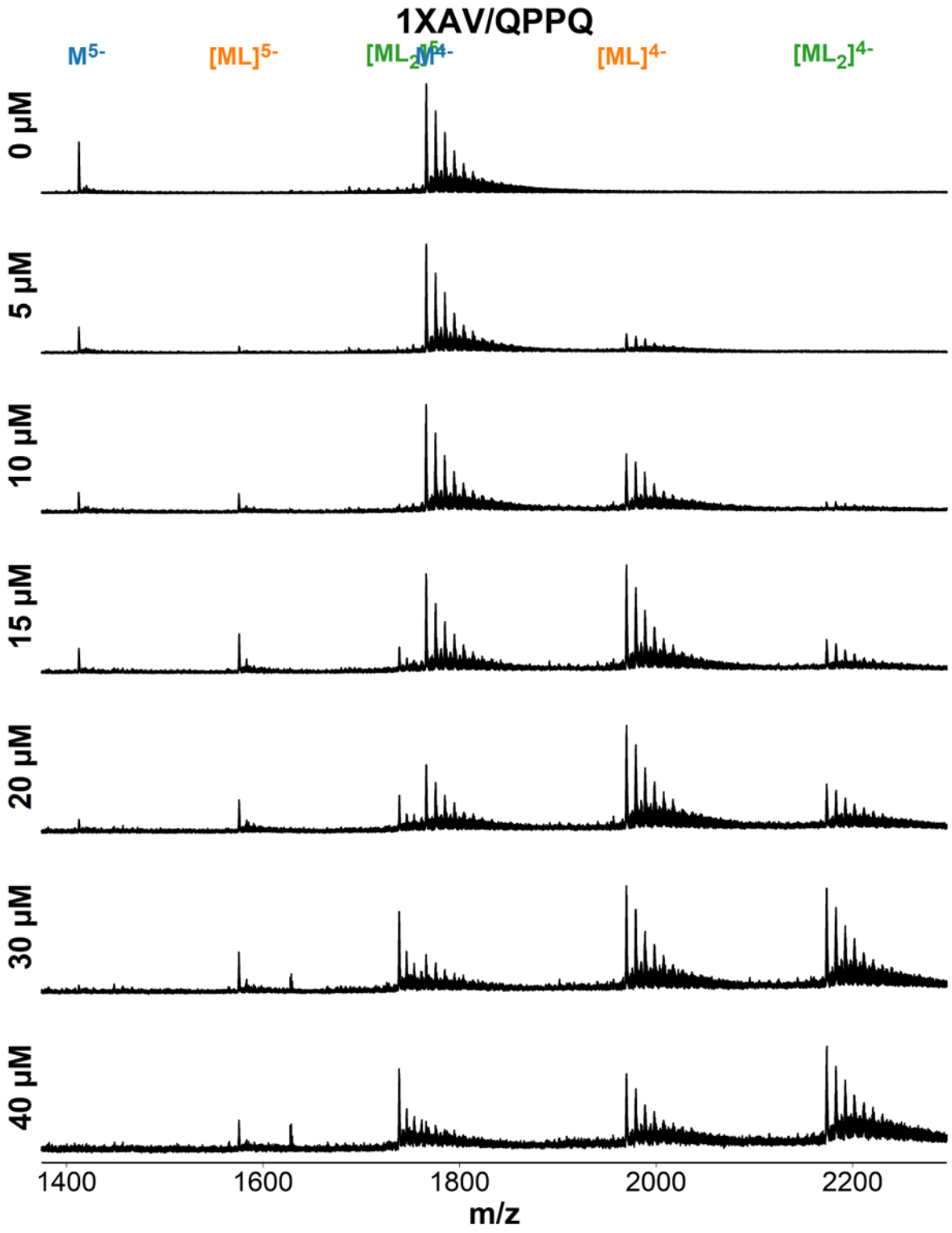


*Figure S105 ESI-MS titration of 1XAV (dT**G**A**GGG**T**GGG**TA**GGG**T**GGG**TAA) with foldamer QPPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

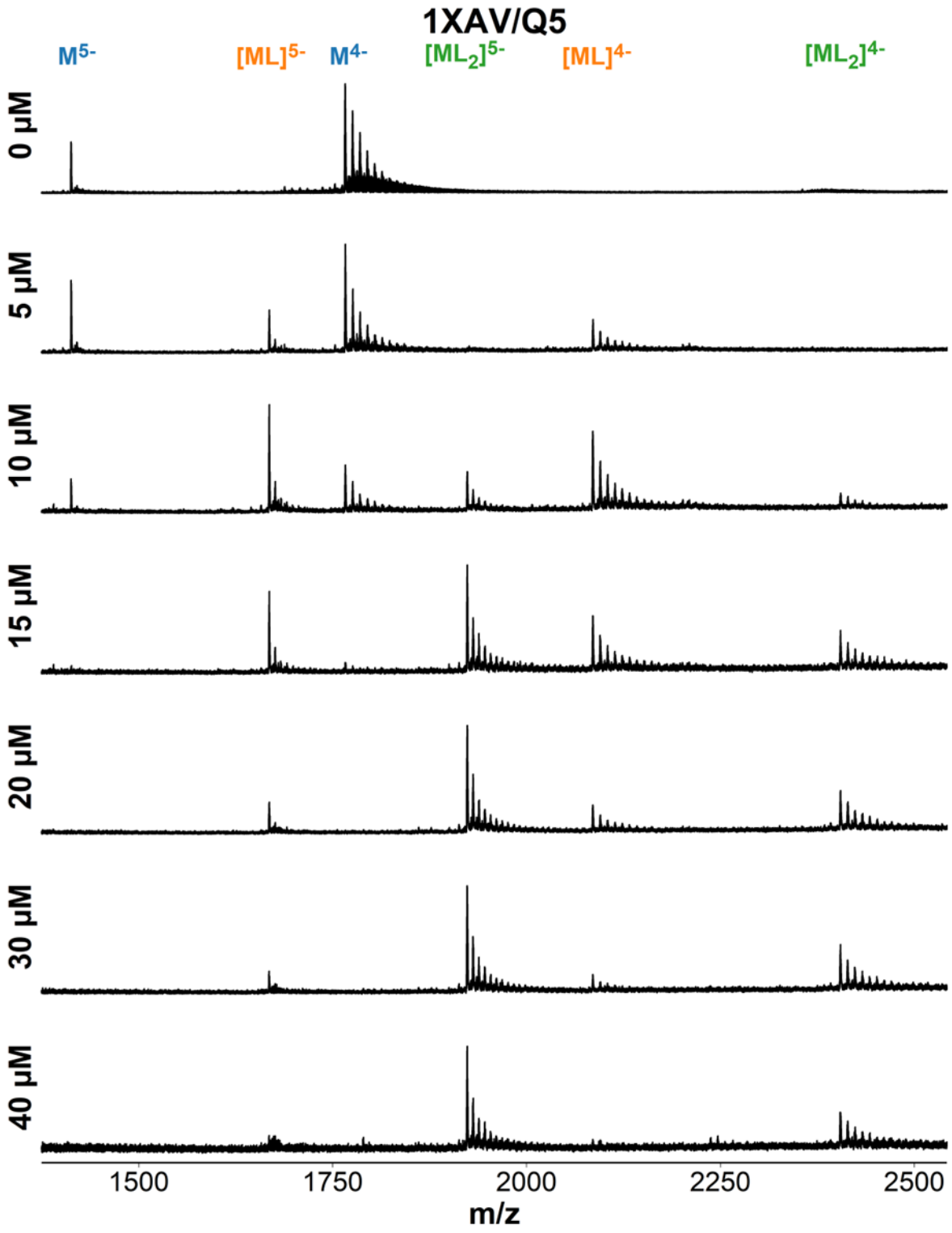


*Figure S106 ESI-MS titration of 1XAV (dT**G**A**GGG**T**GGG**TA**GGG**T**GGG**TAA) with foldamer QQQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

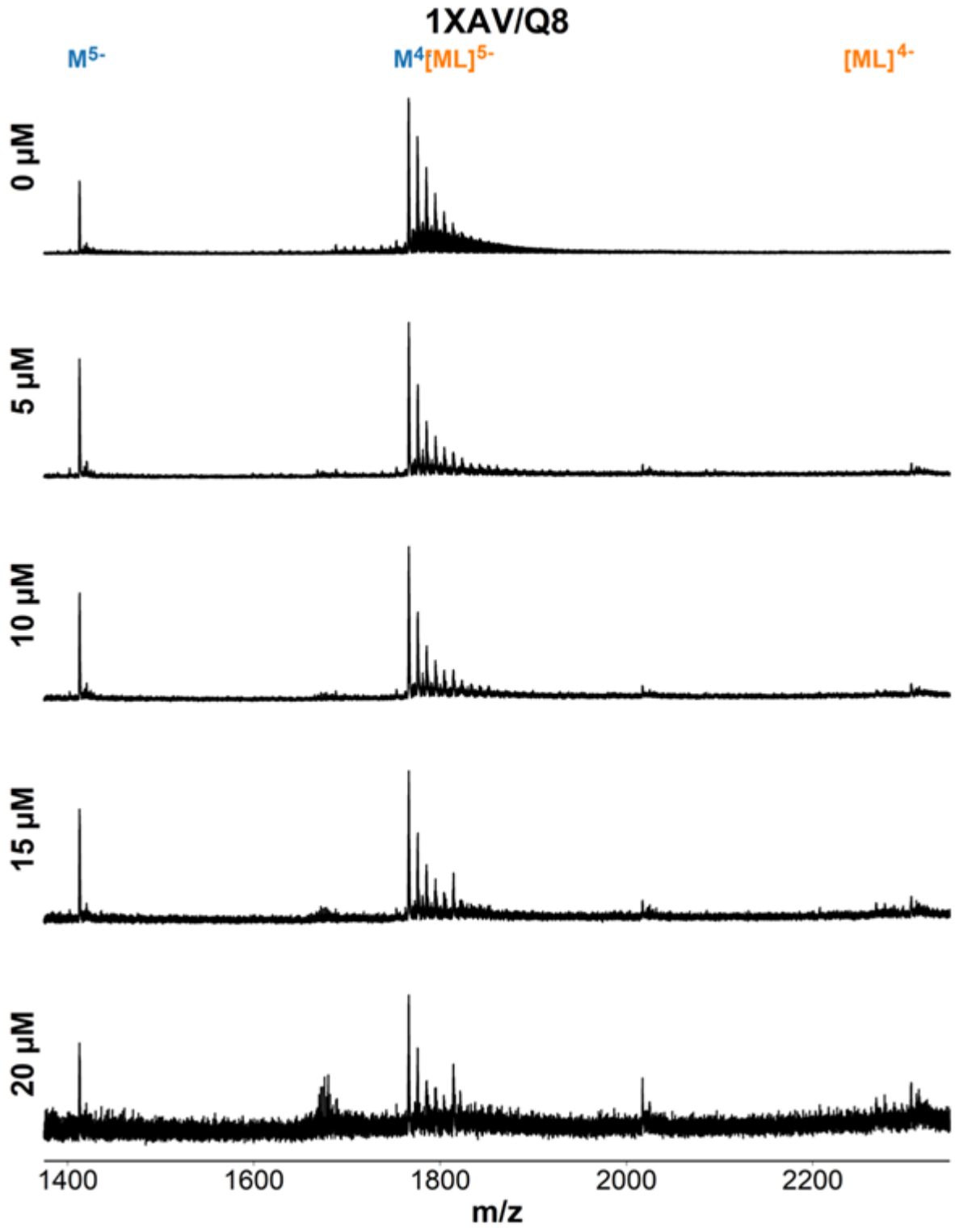


*Figure S107 ESI-MS titration of 1XAV (dT**G**A**GGG**T**GGG**TA**GGG**T**GGG**TAA) with foldamer QQQQQQQQ. Samples contain 10 µM DNA, 0-20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

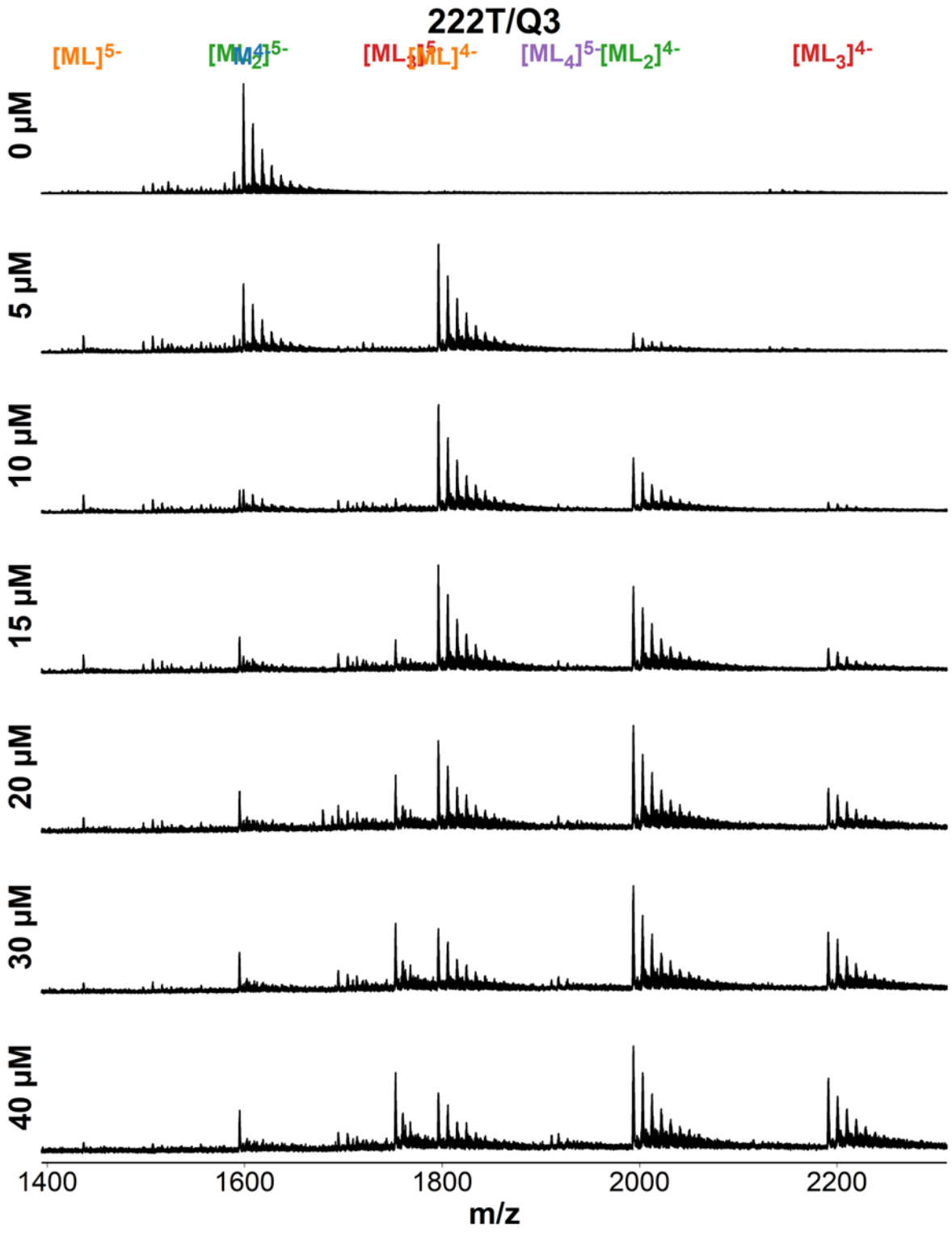


*Figure S108 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

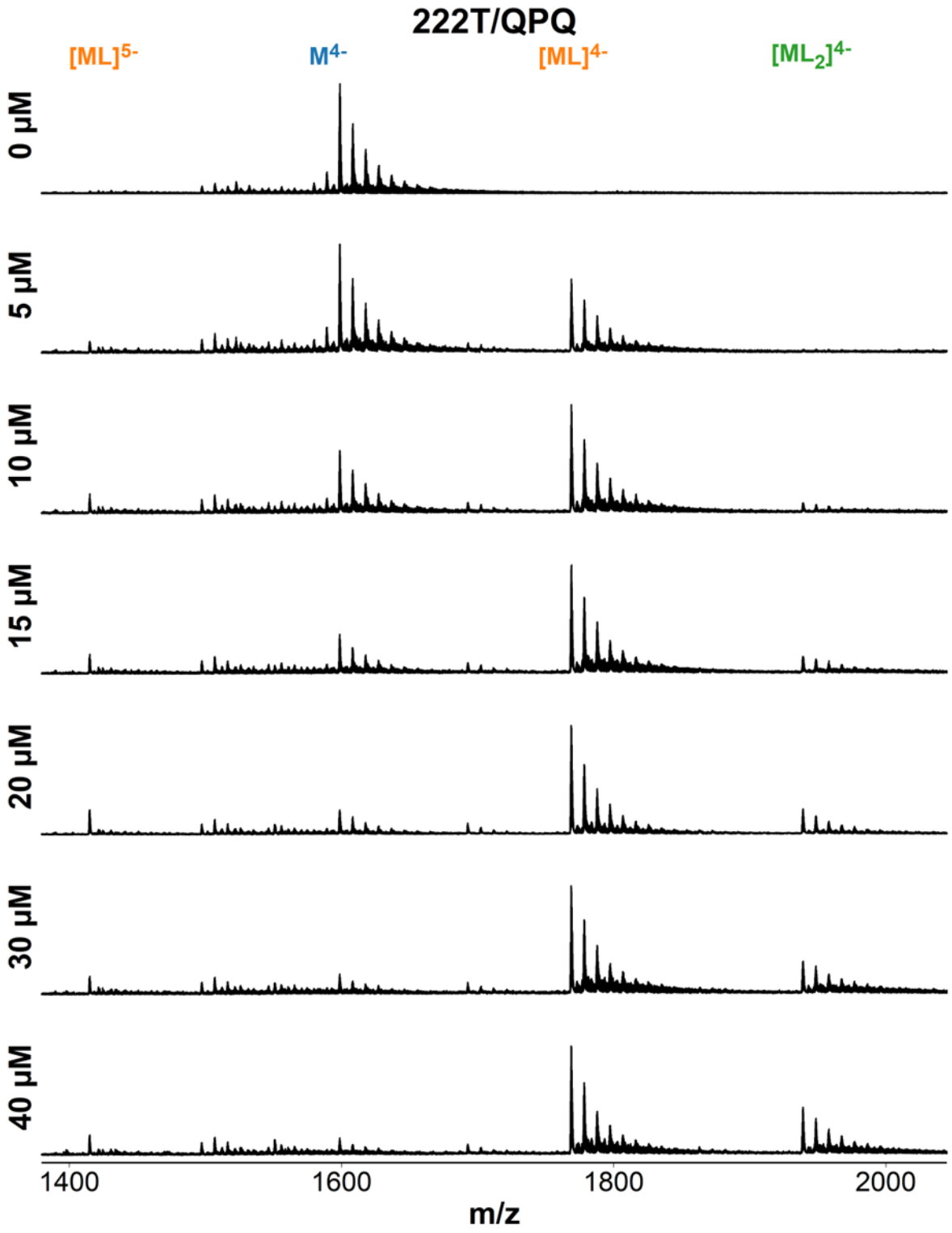


*Figure S109 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

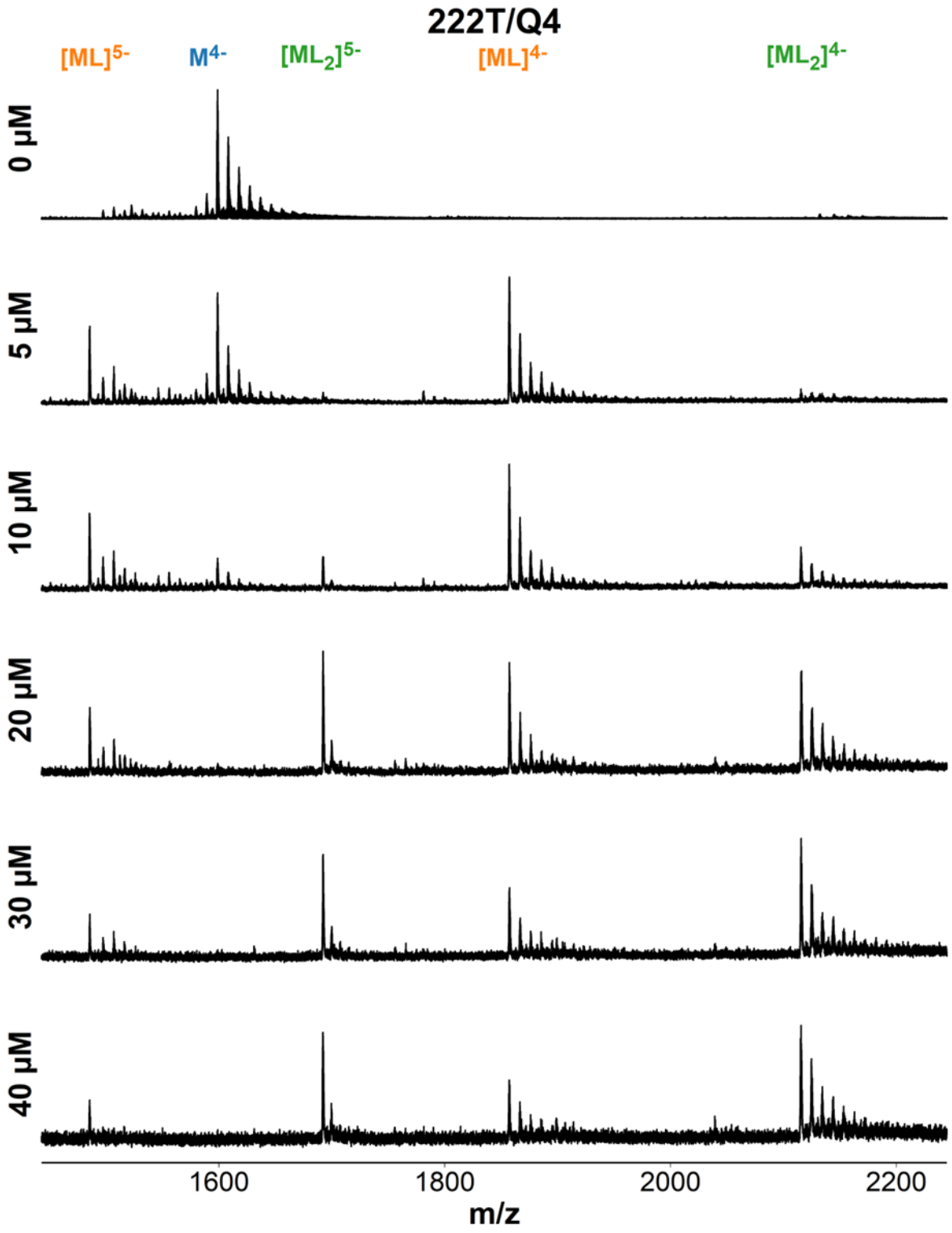


*Figure S110 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

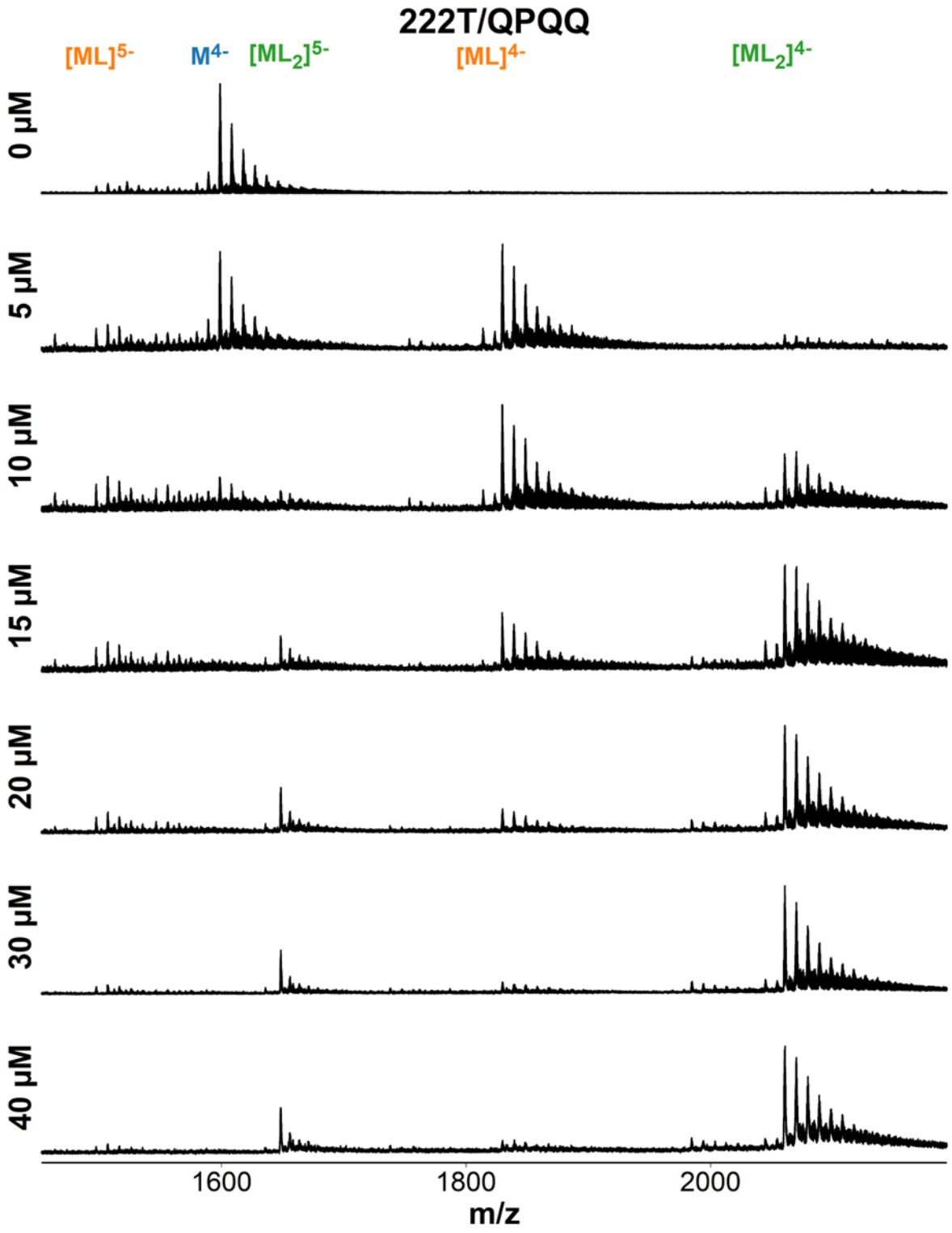


*Figure S111 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QQPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

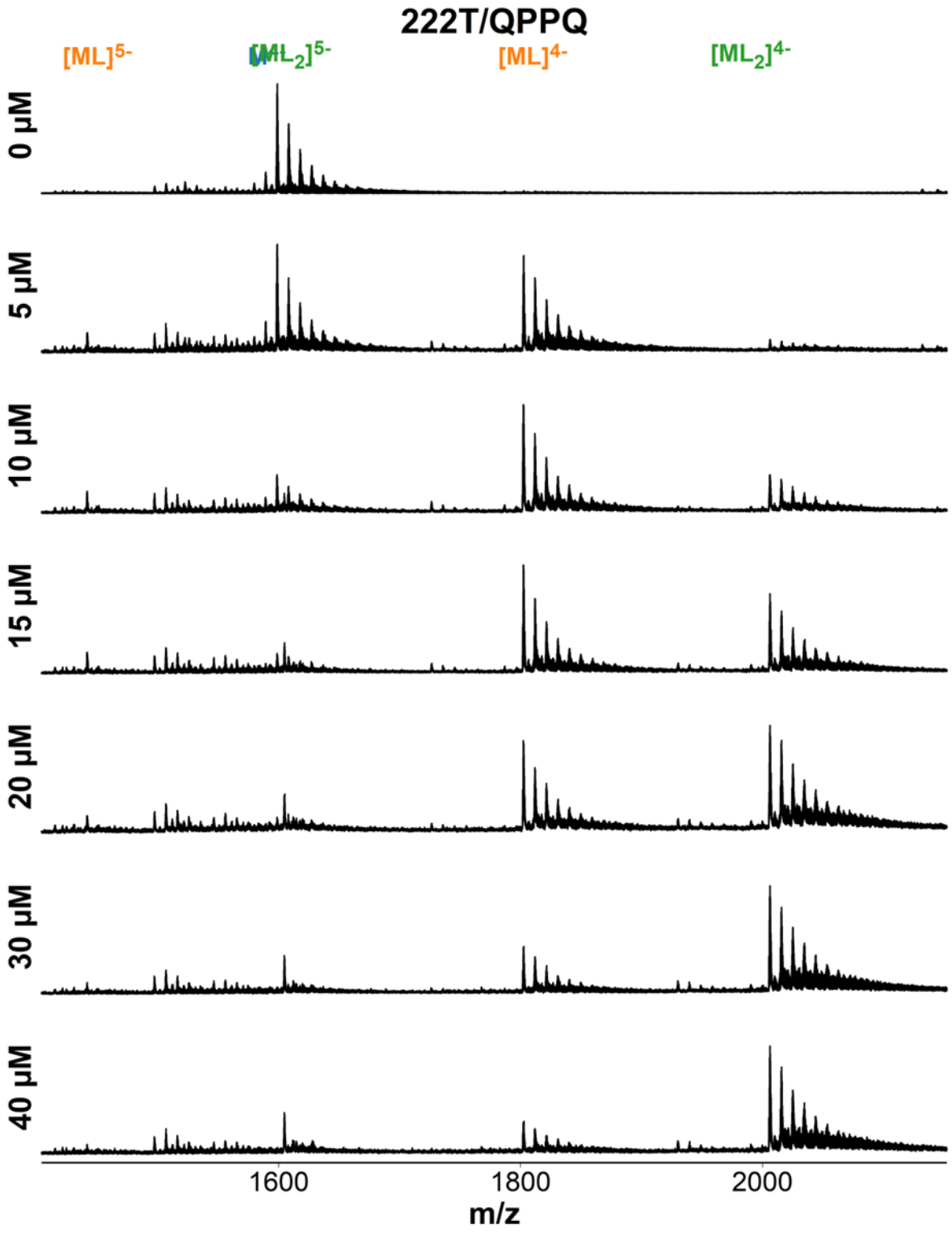


*Figure S112 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QPPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

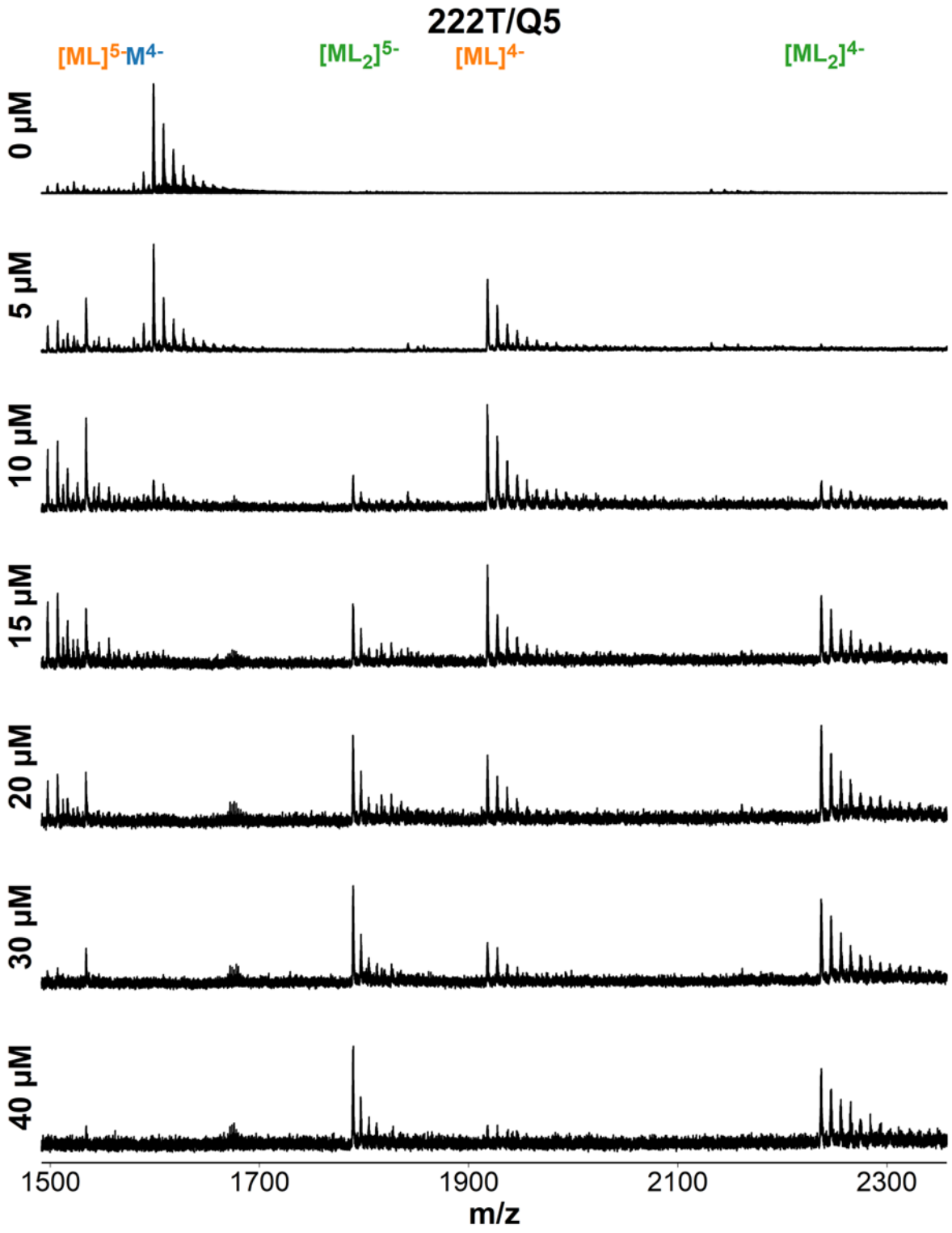


*Figure S113 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QQQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

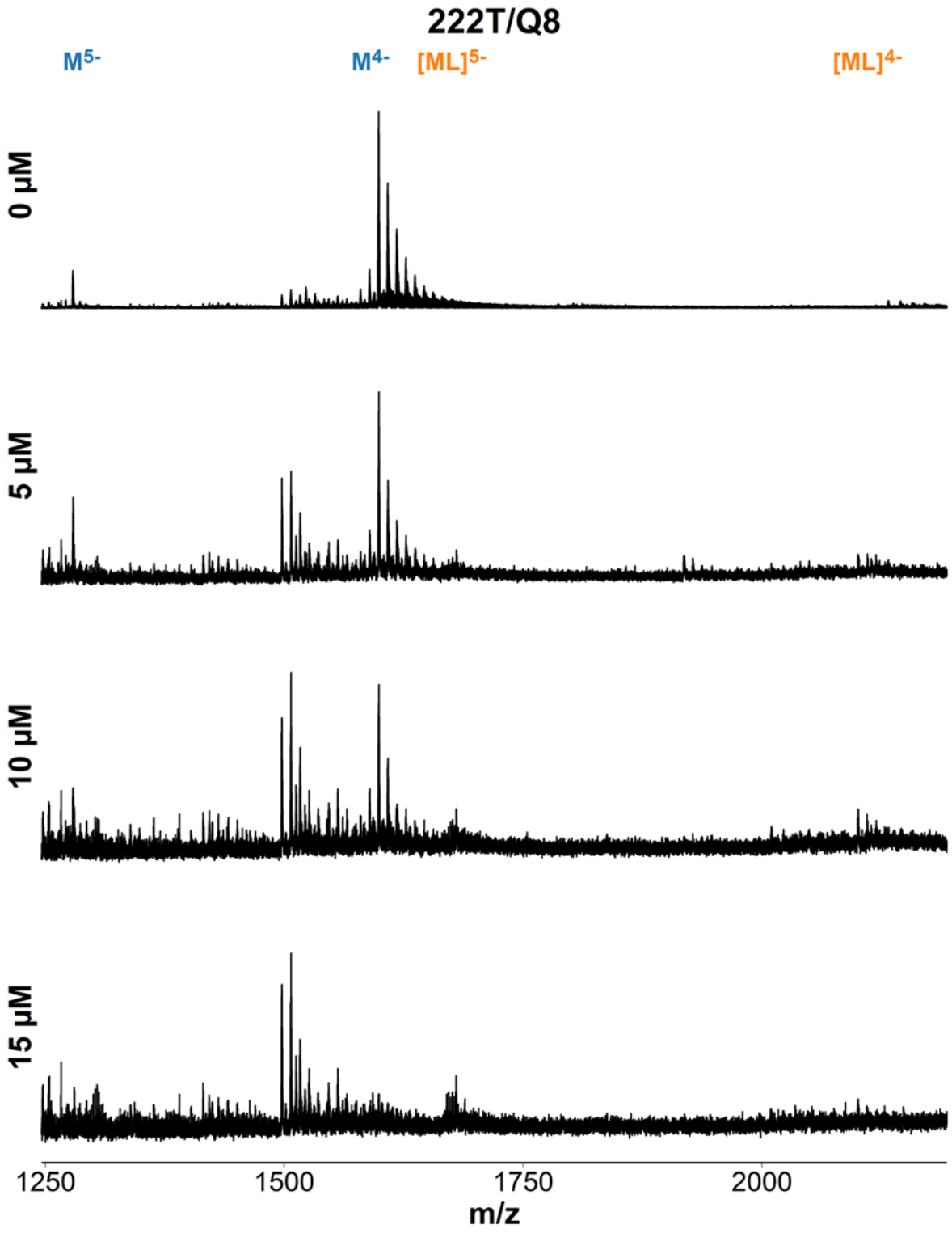


*Figure S114 ESI-MS titration of 222T (dT**GGG**TT**GGG**TT**GGG**TT**GGG**T) with foldamer QQQQQQQQ. Samples contain 10 µM DNA, 0-15 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

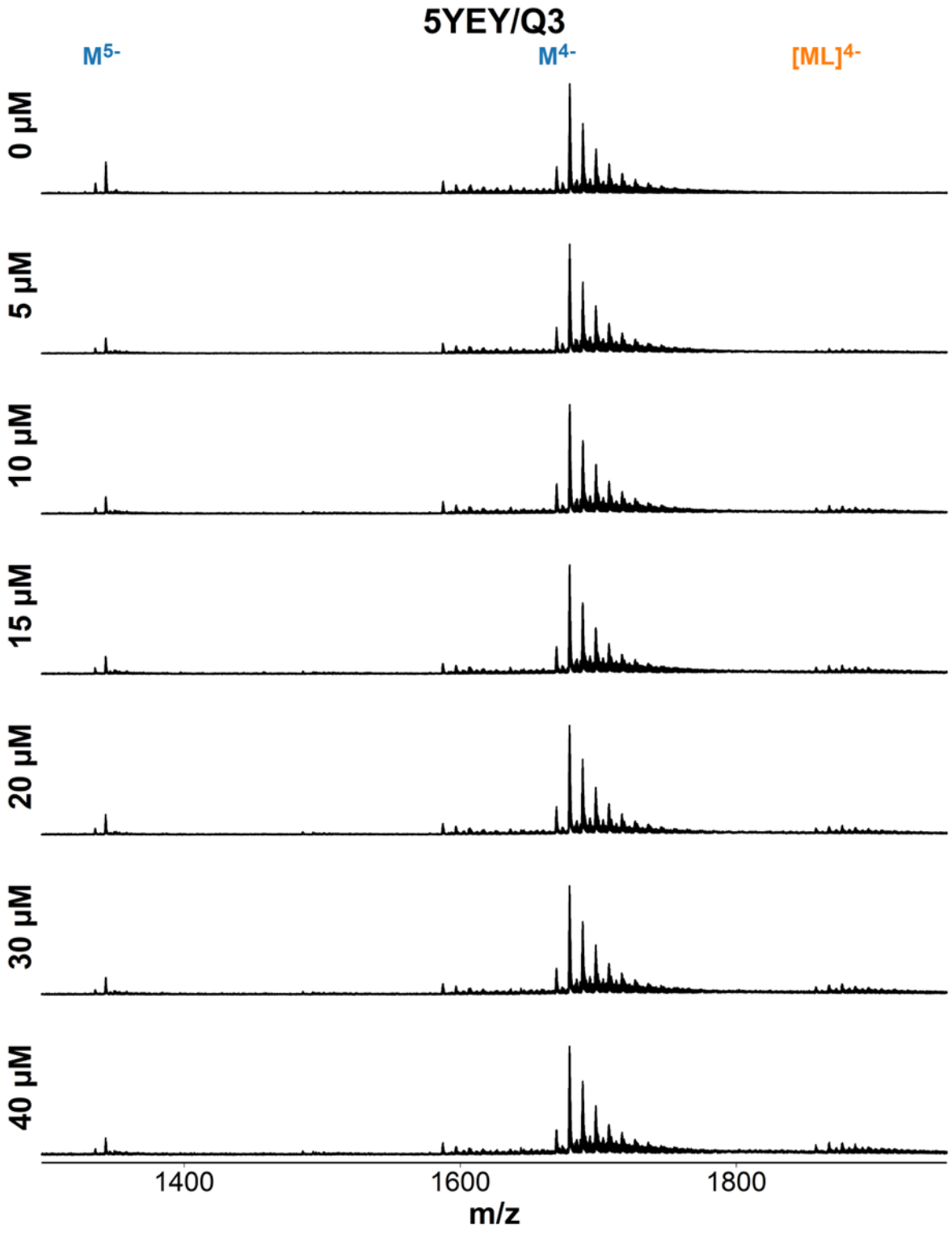


*Figure S115 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

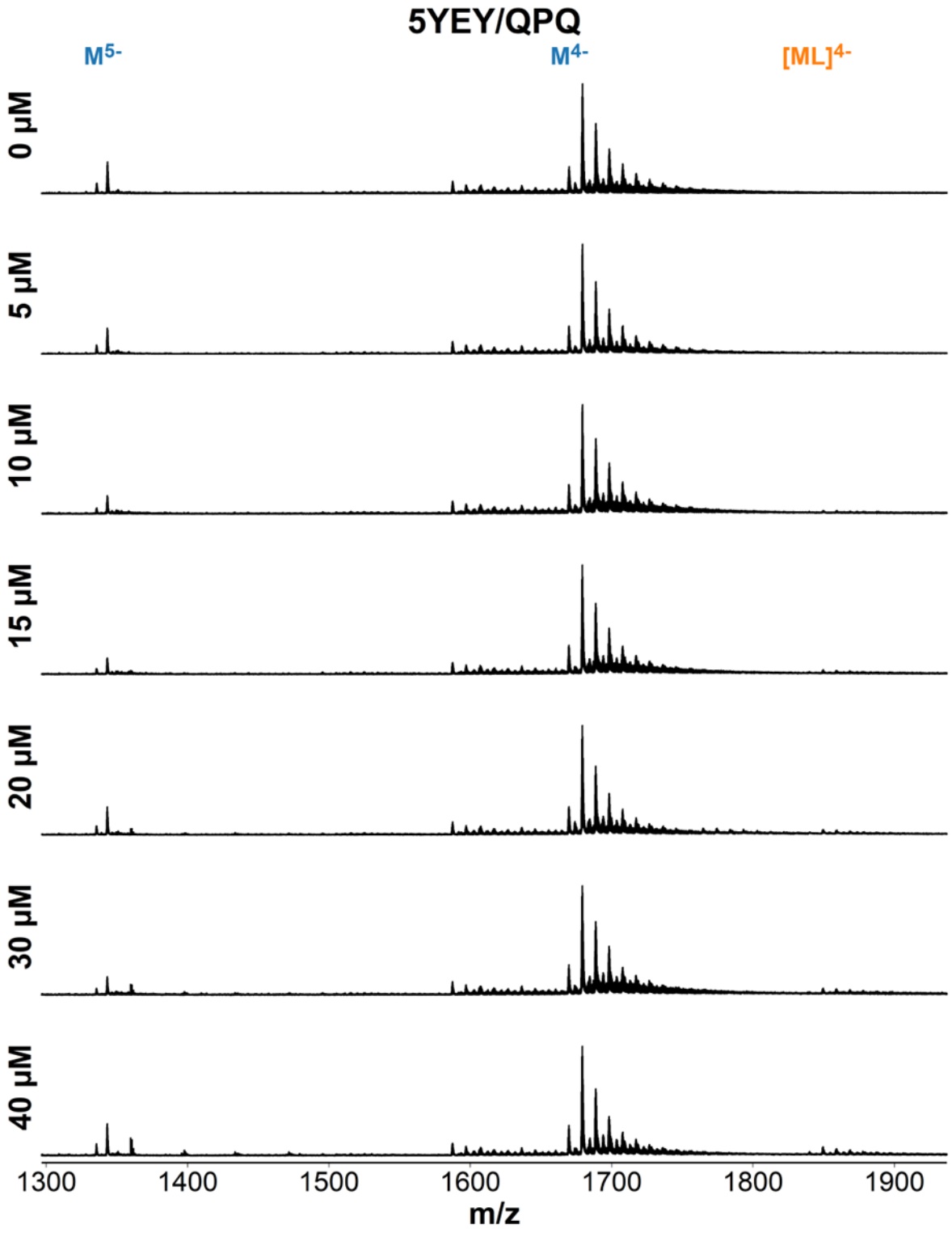


*Figure S116 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

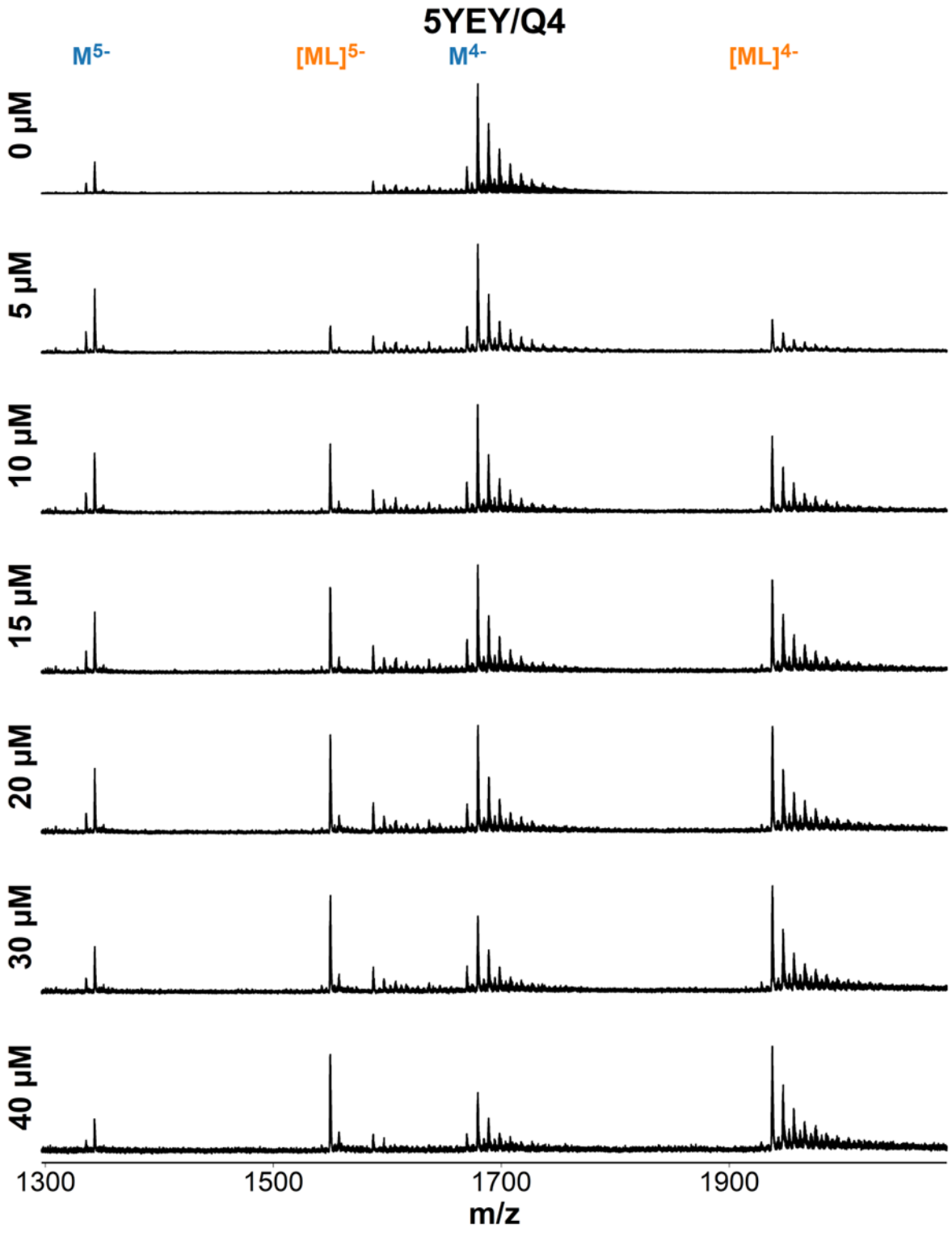


*Figure S117 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

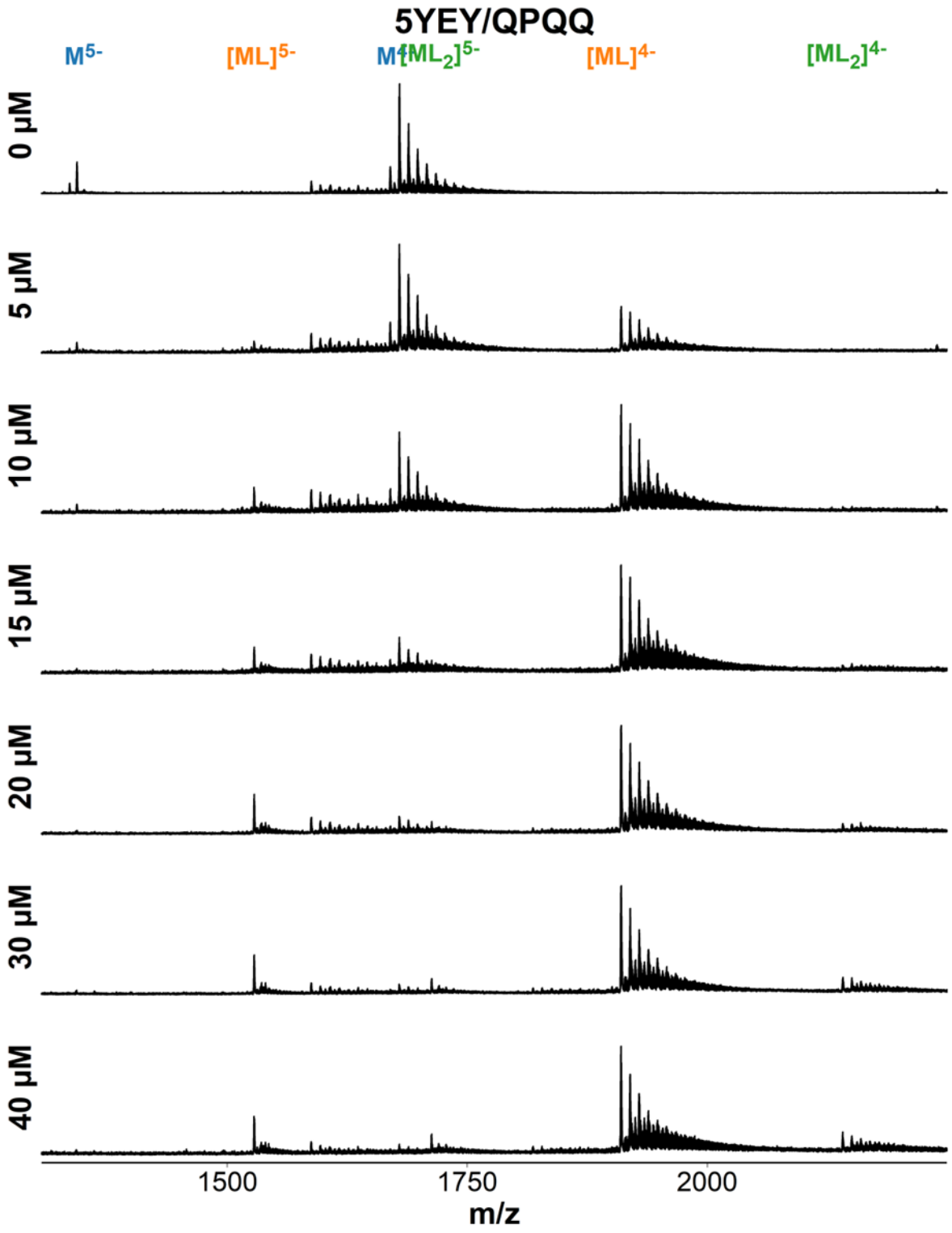


*Figure S118 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QQPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

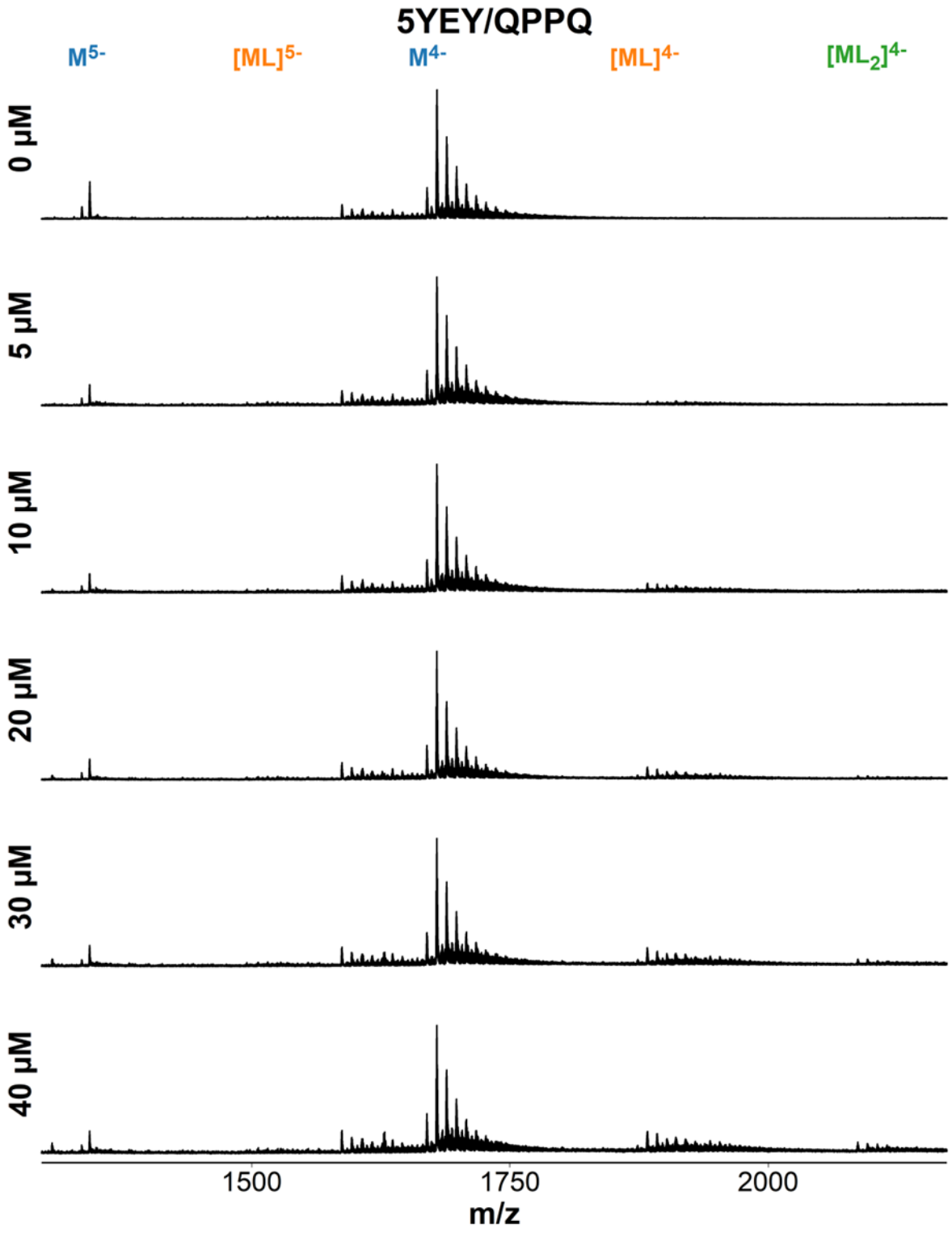


*Figure S119 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QPPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

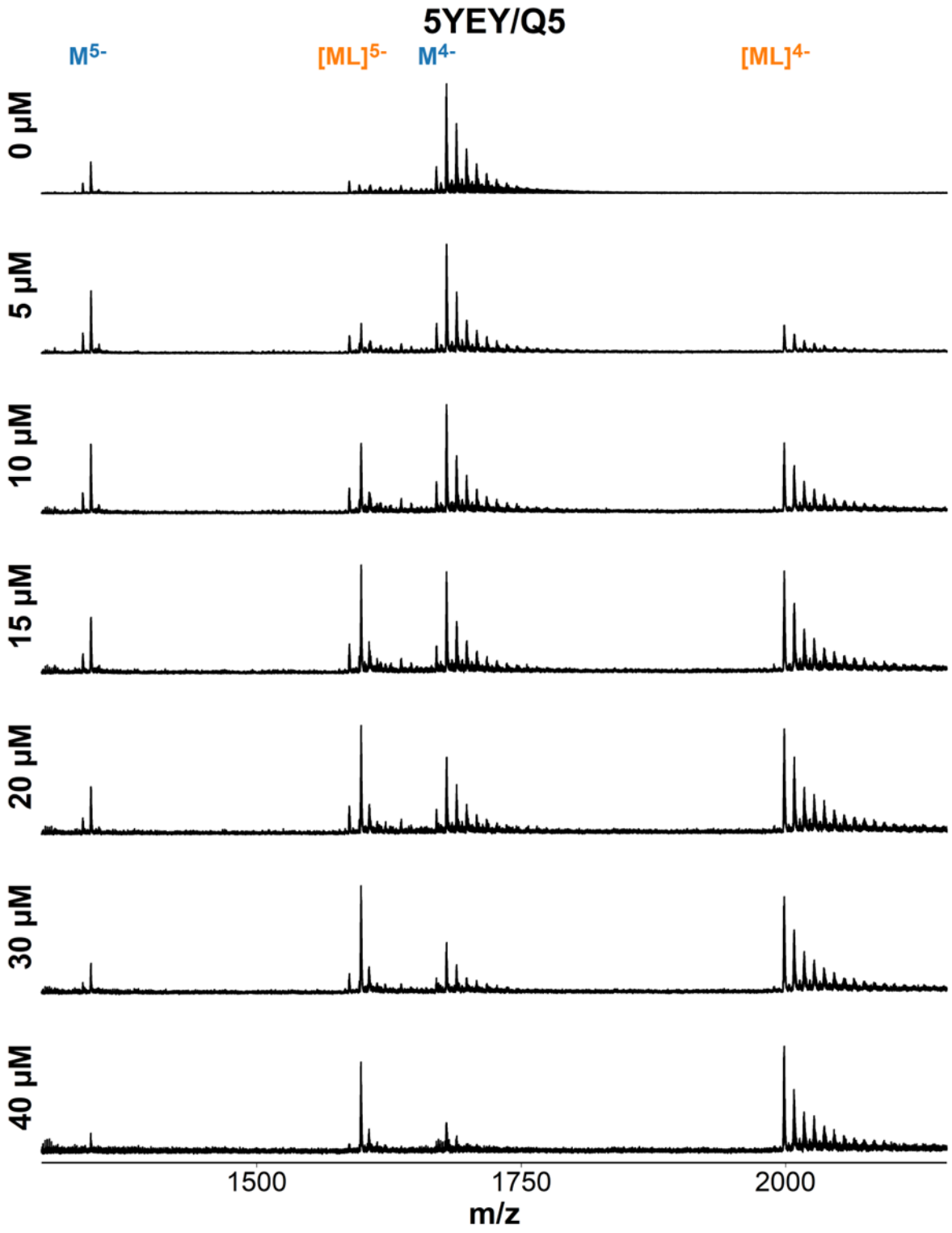


*Figure S120 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QQQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

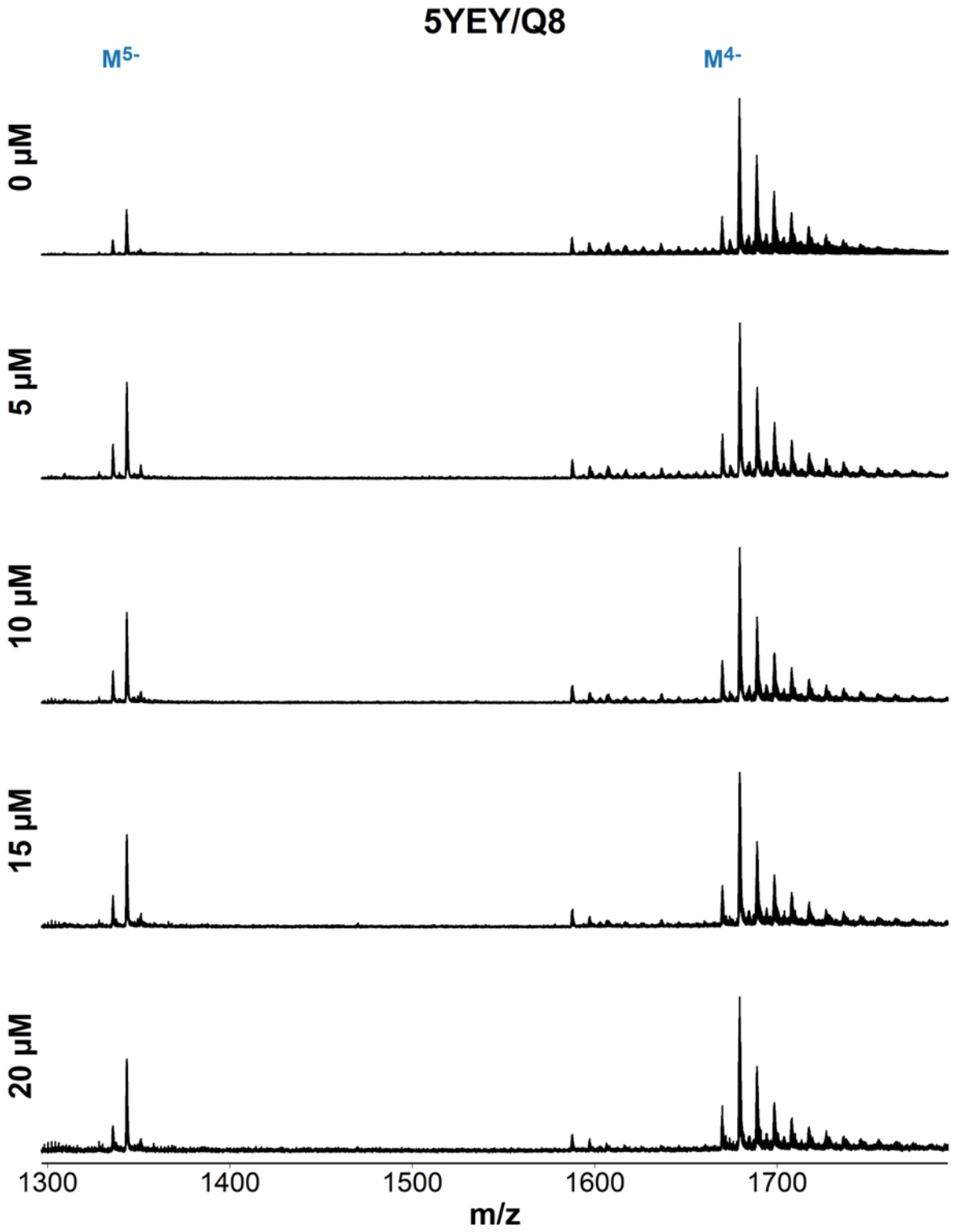


*Figure S121 ESI-MS titration of 5YEY (d**GGG**TTA**GGG**TTA**GGG**TTT**GGG**) with foldamer QQQQQQQQ. Samples contain 10 µM DNA, 0-20 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

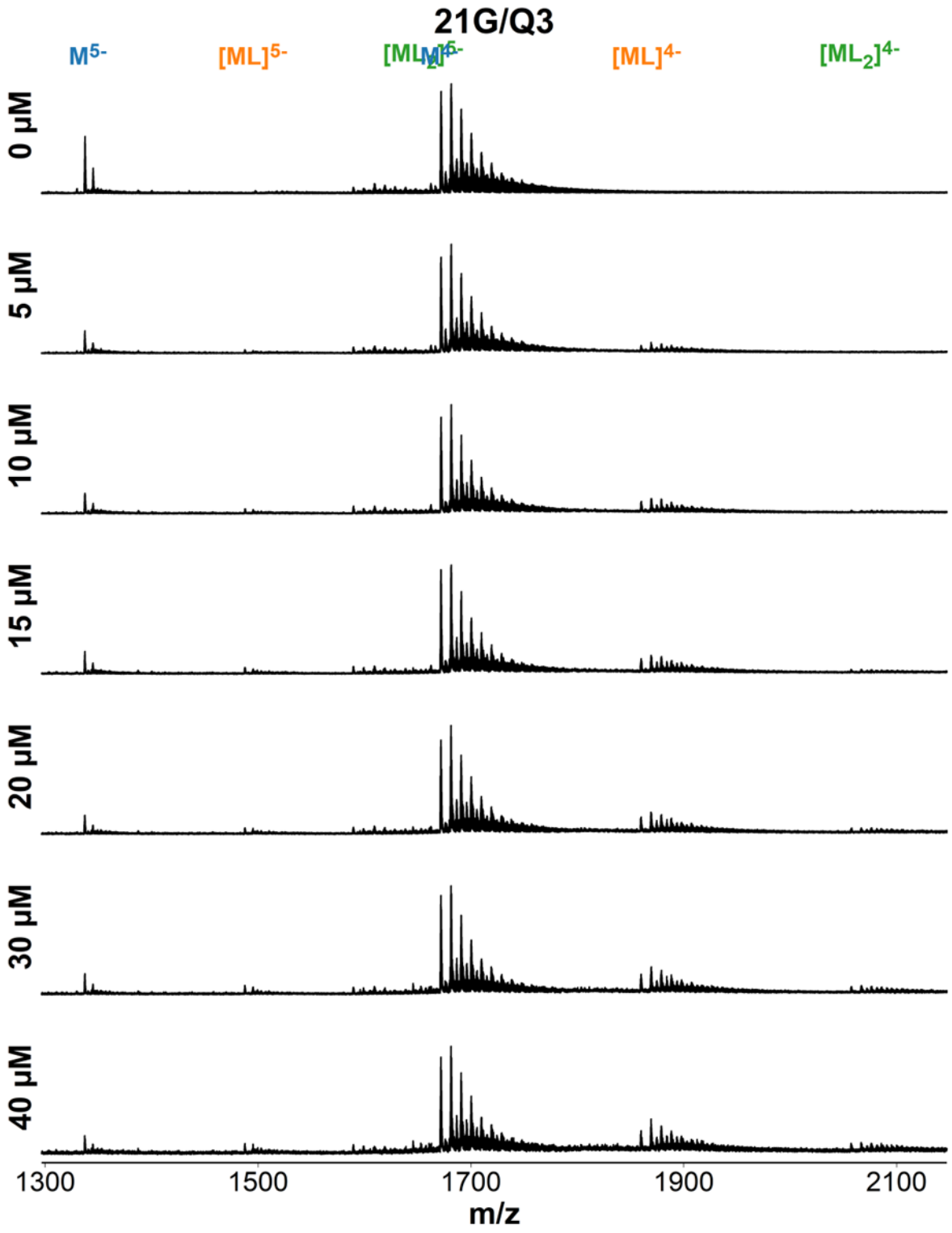


*Figure S122 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

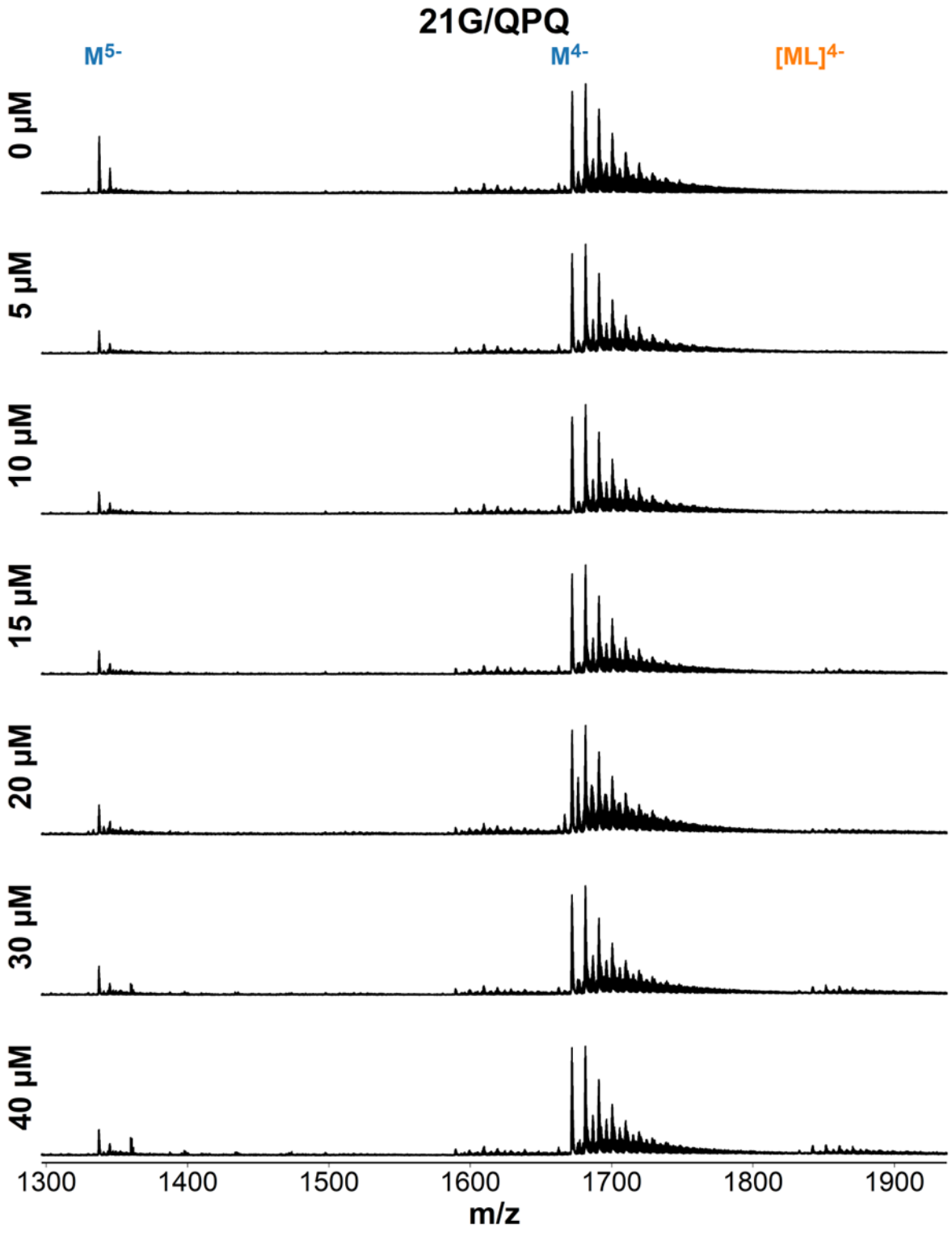


*Figure S123 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

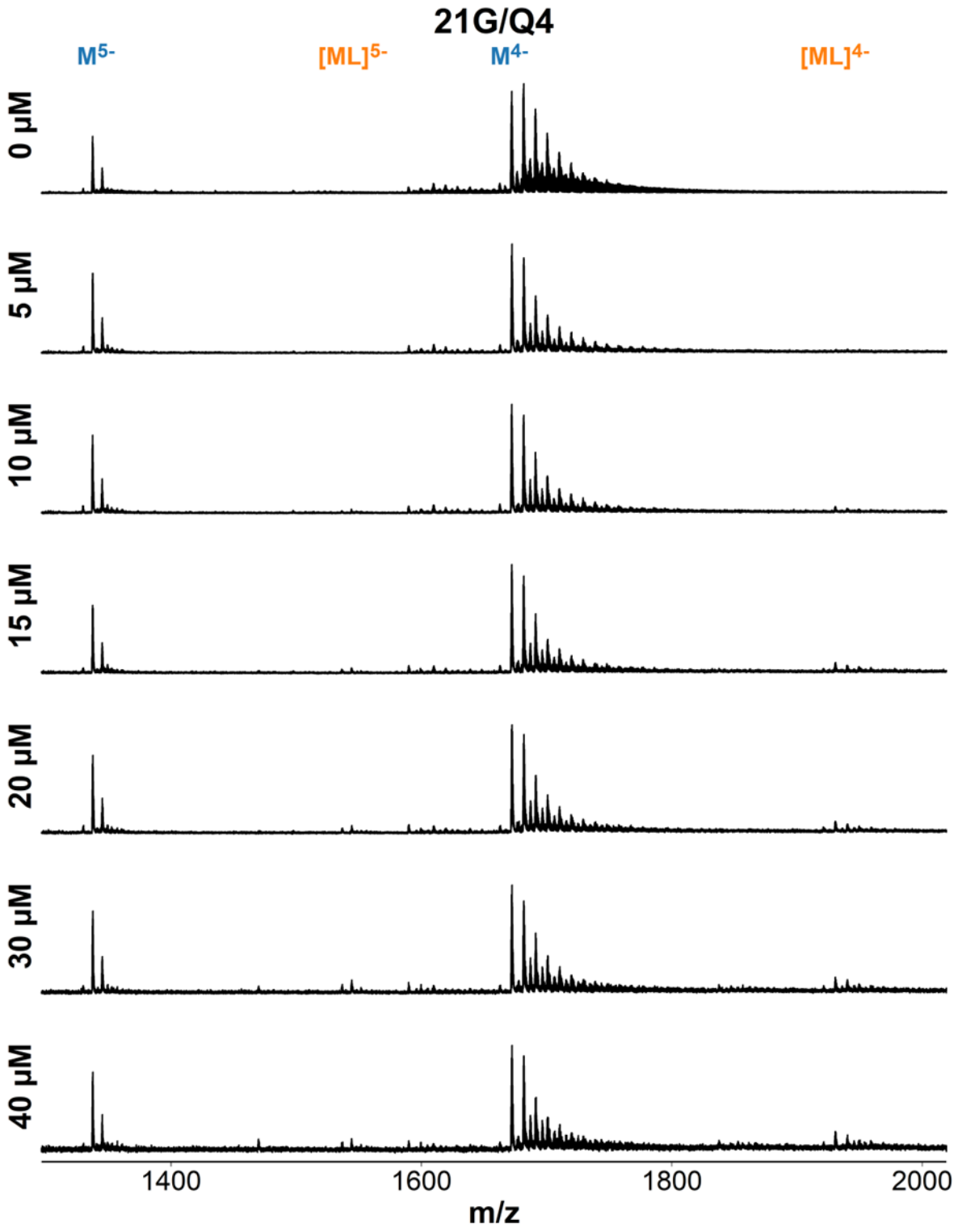


*Figure S124 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

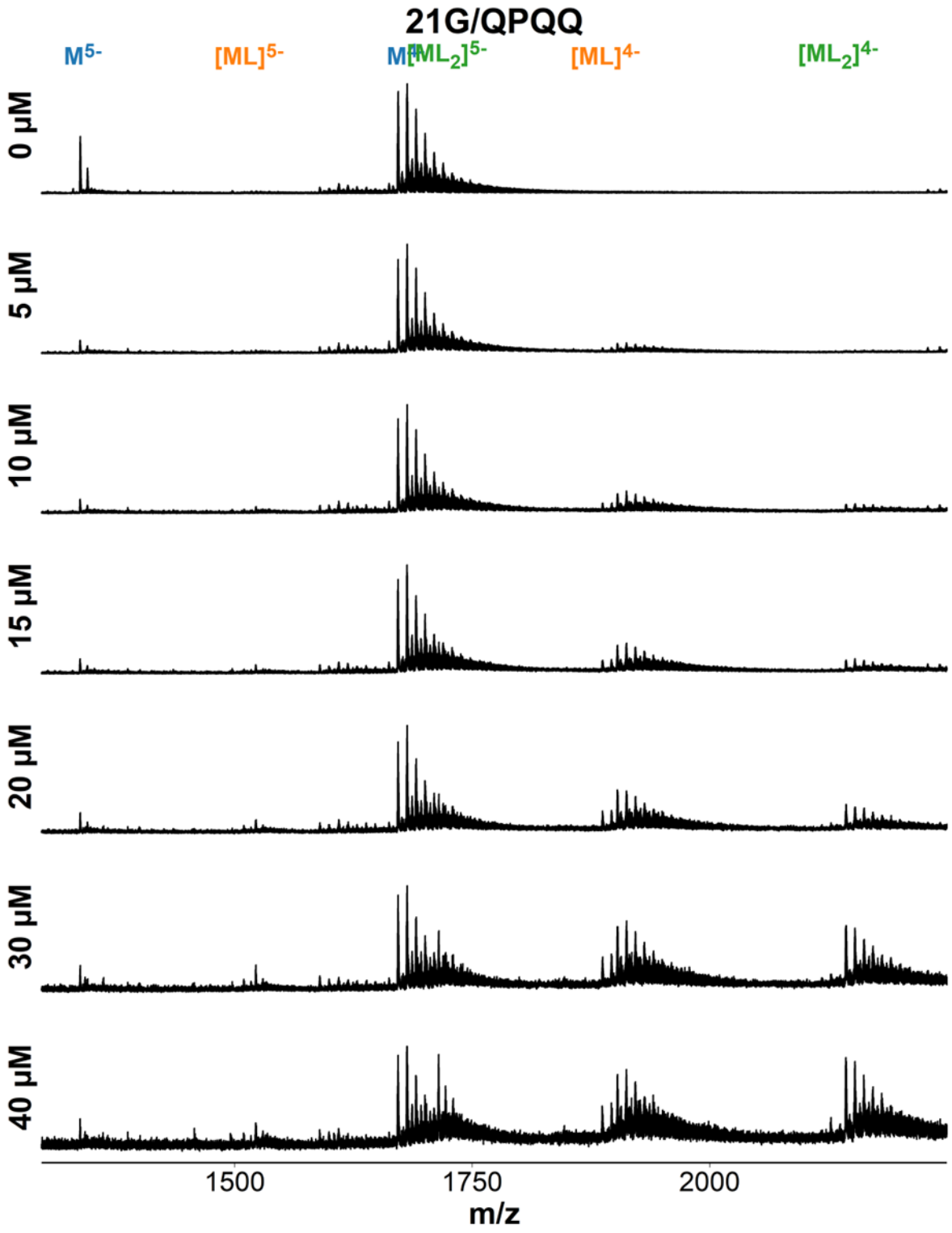


*Figure S125 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QQPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

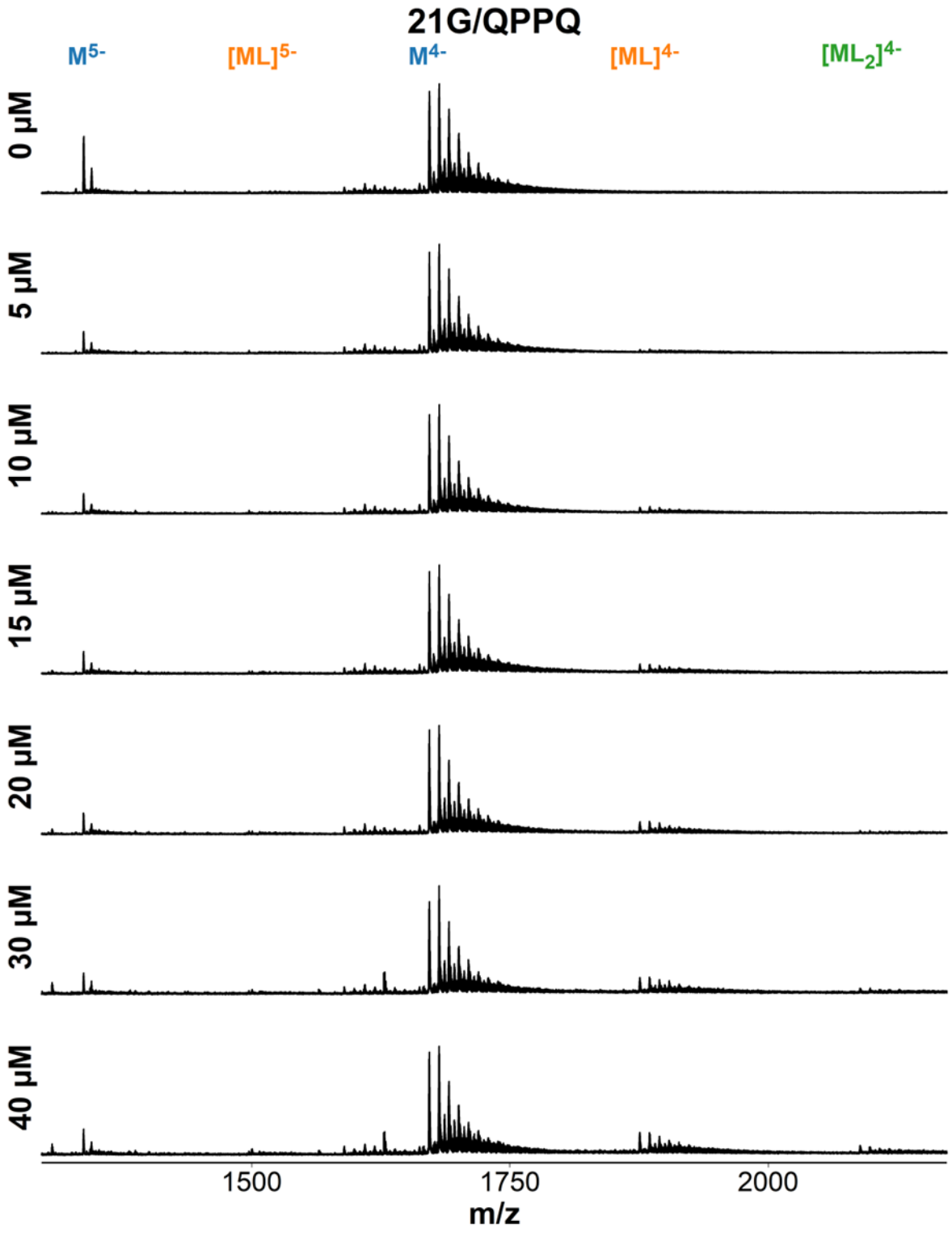


*Figure S126 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QPPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

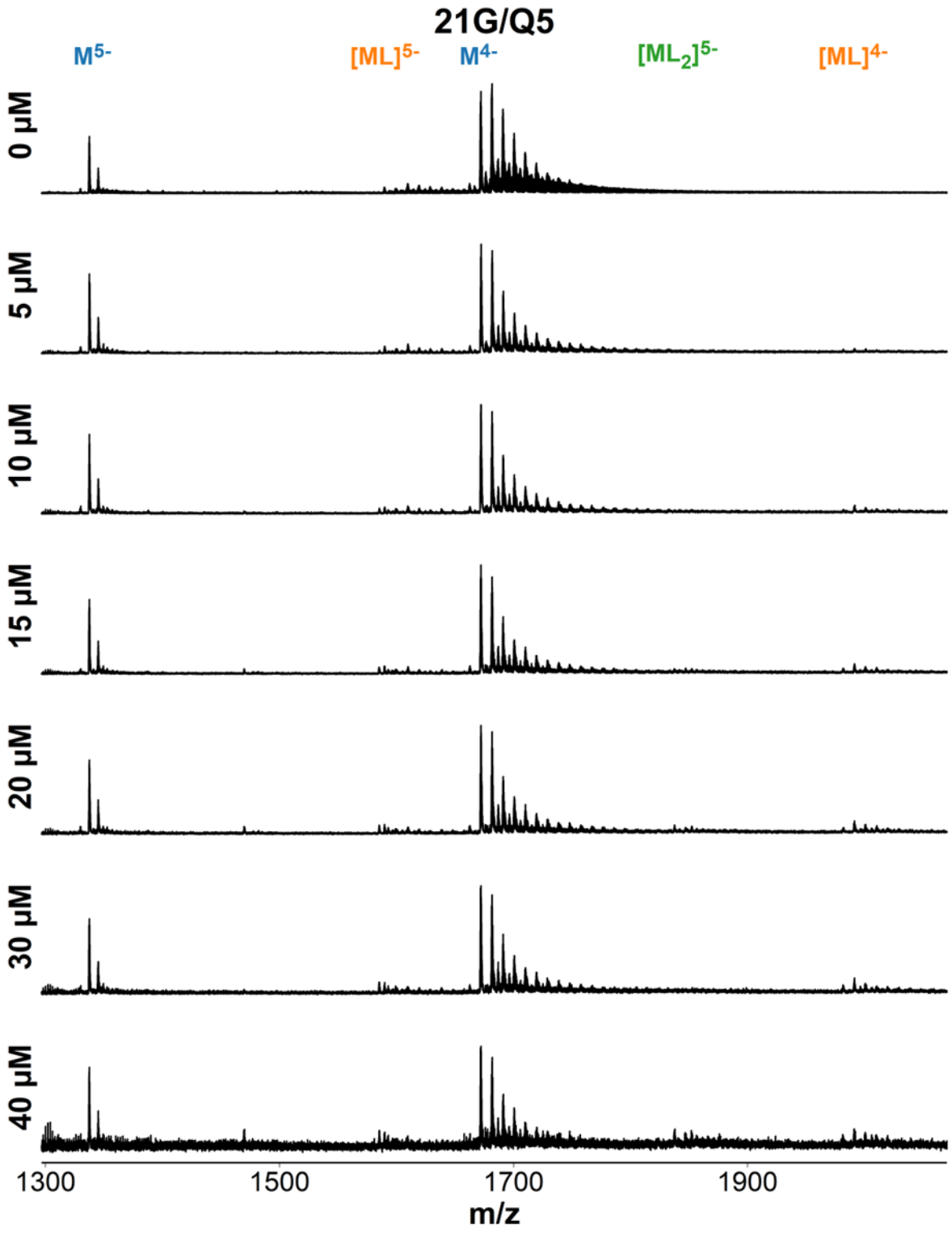


*Figure S127 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QQQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

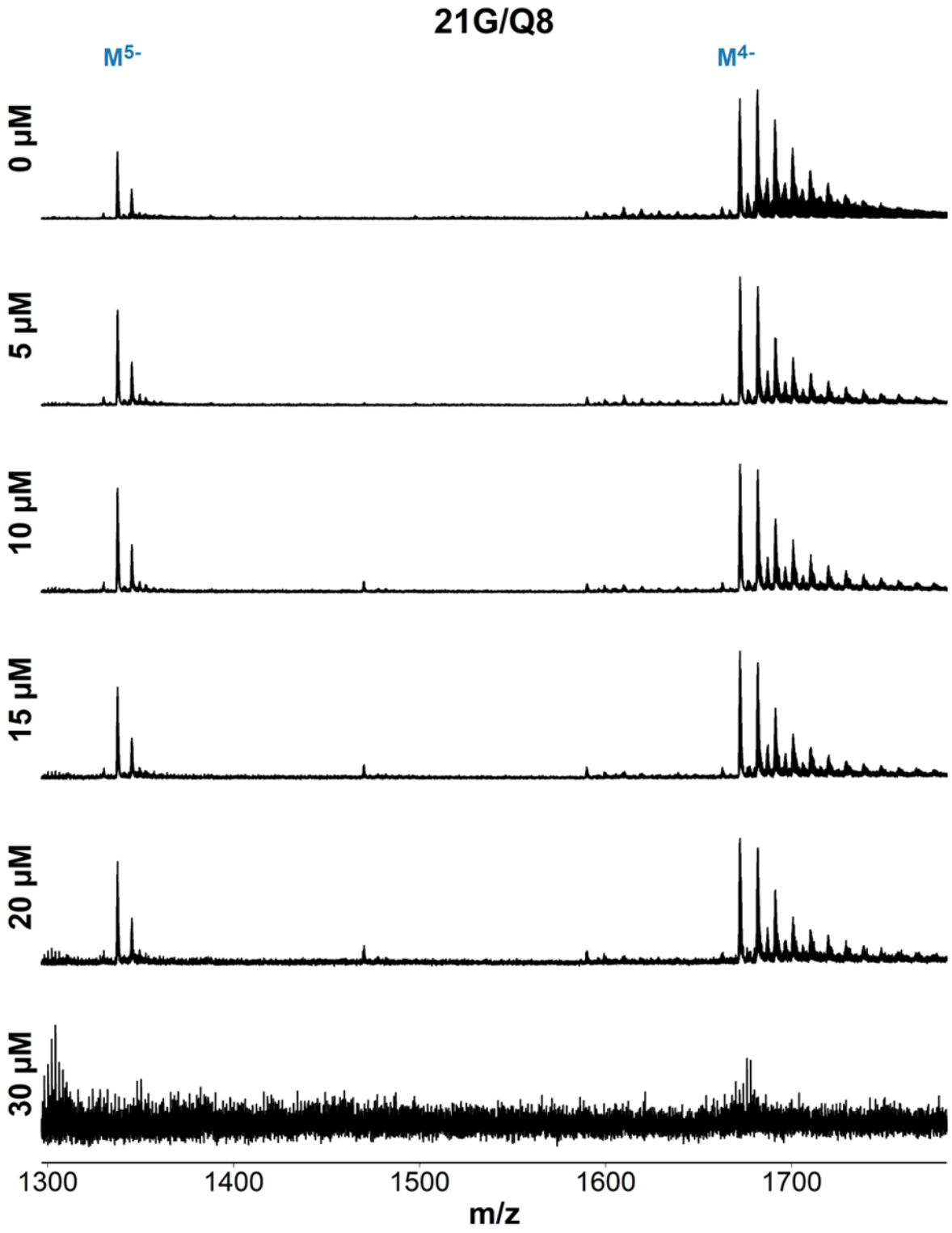


*Figure S128 ESI-MS titration of 21G (d**GGG**TTA**GGG**TTA**GGG**TTA**GGG**) with foldamer QQQQQQQQ. Samples contain 10 µM DNA, 0-30 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).*

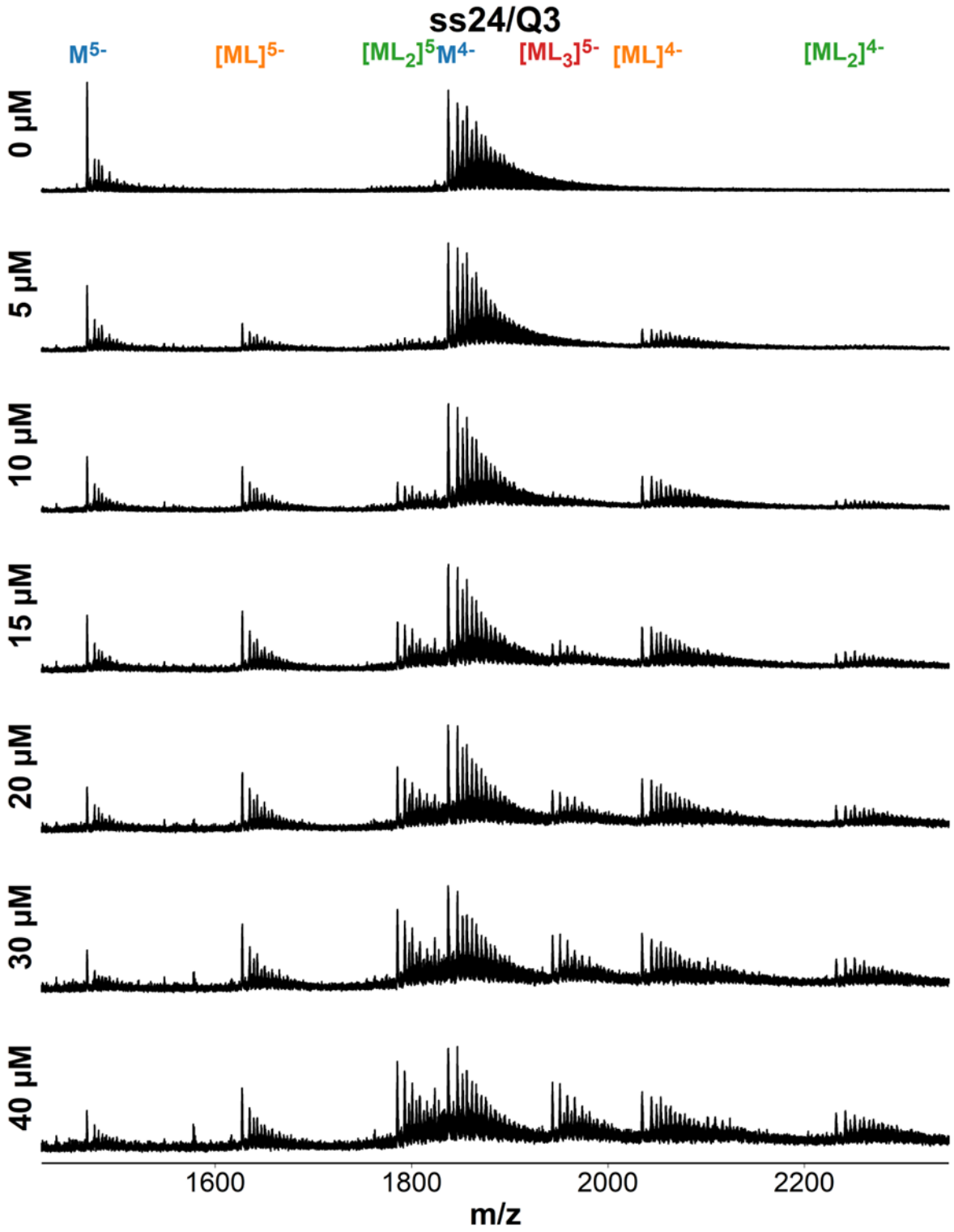


Figure S129 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

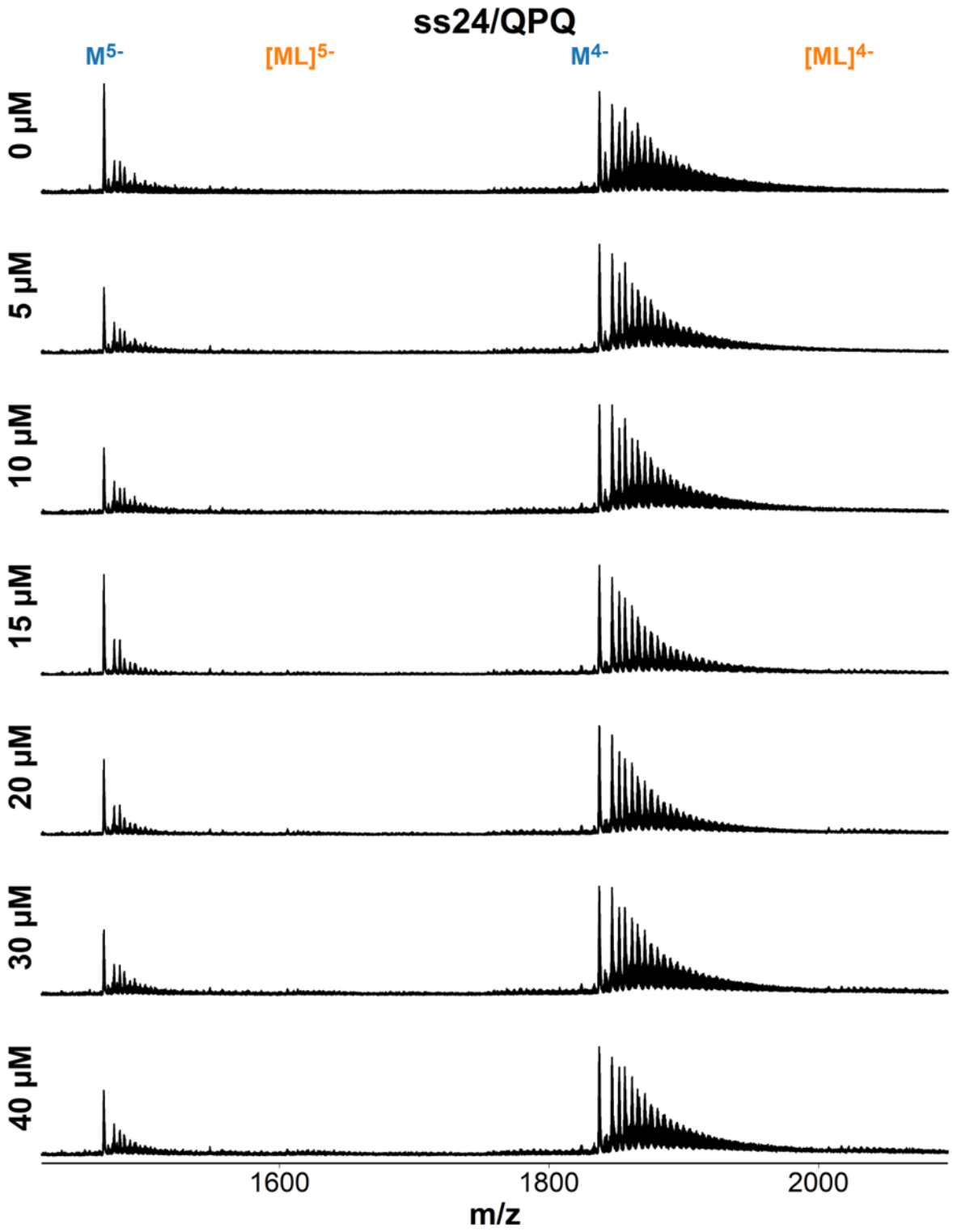


Figure S130 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

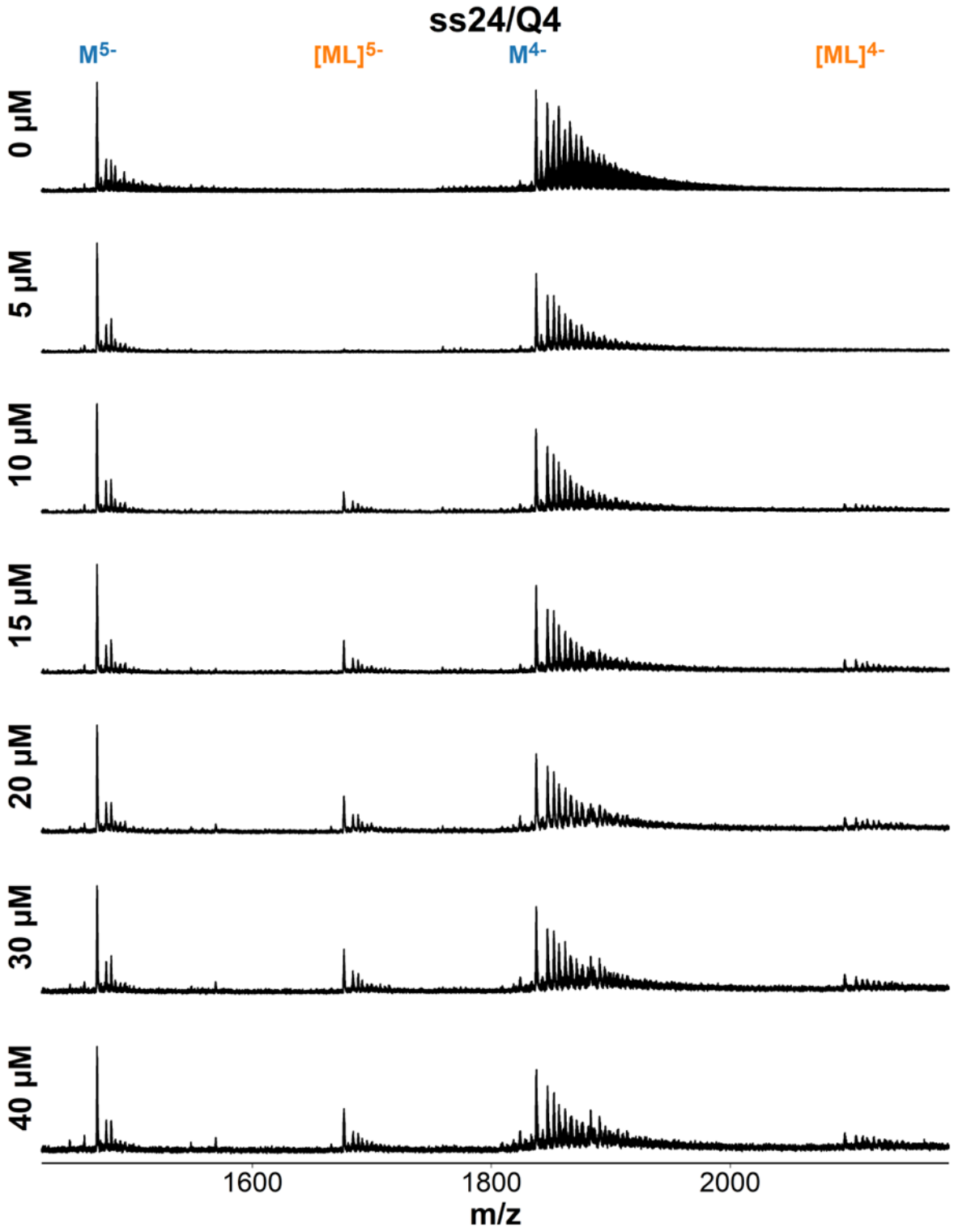


Figure S131 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

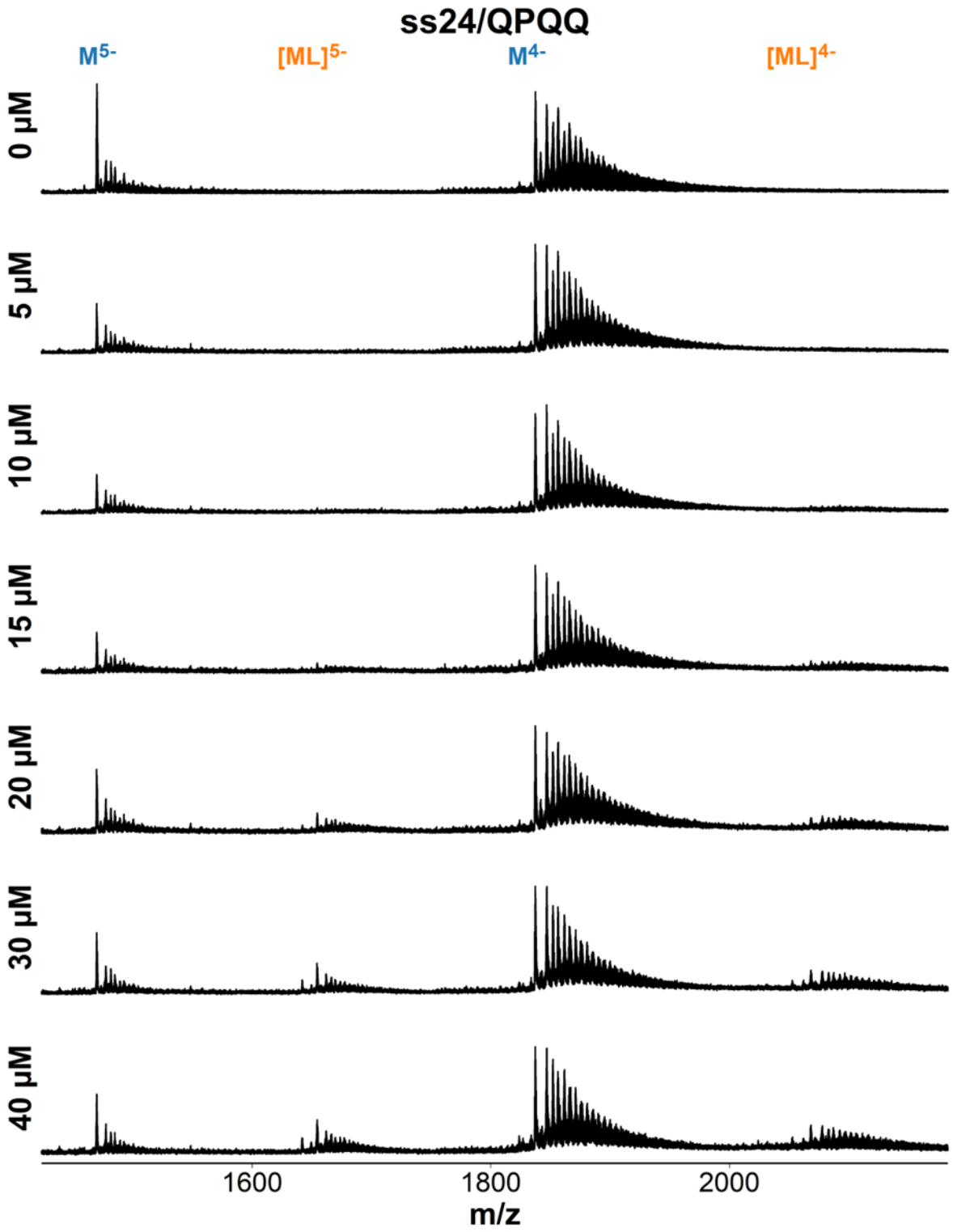


Figure S132 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QQPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

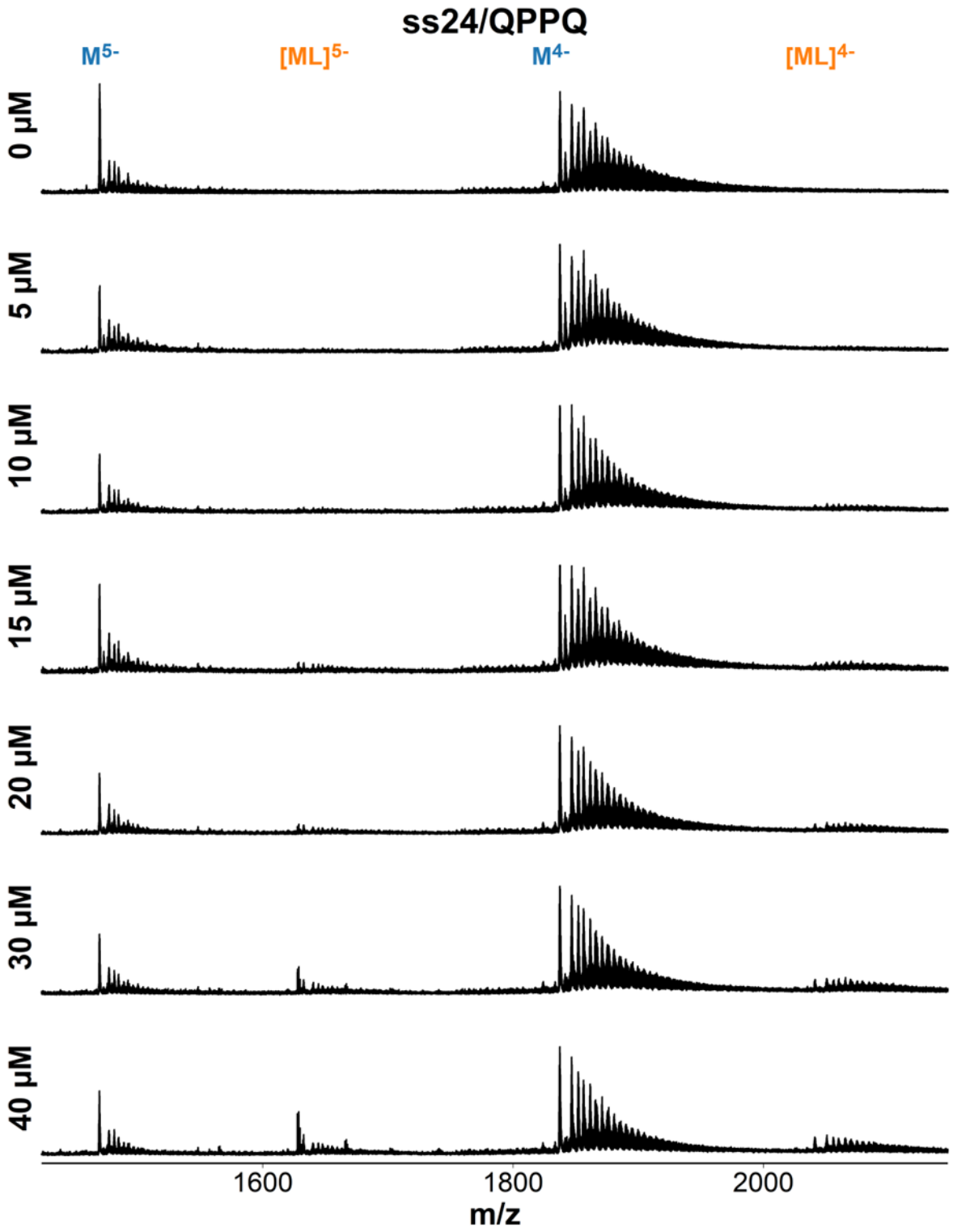


Figure S133 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QPPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

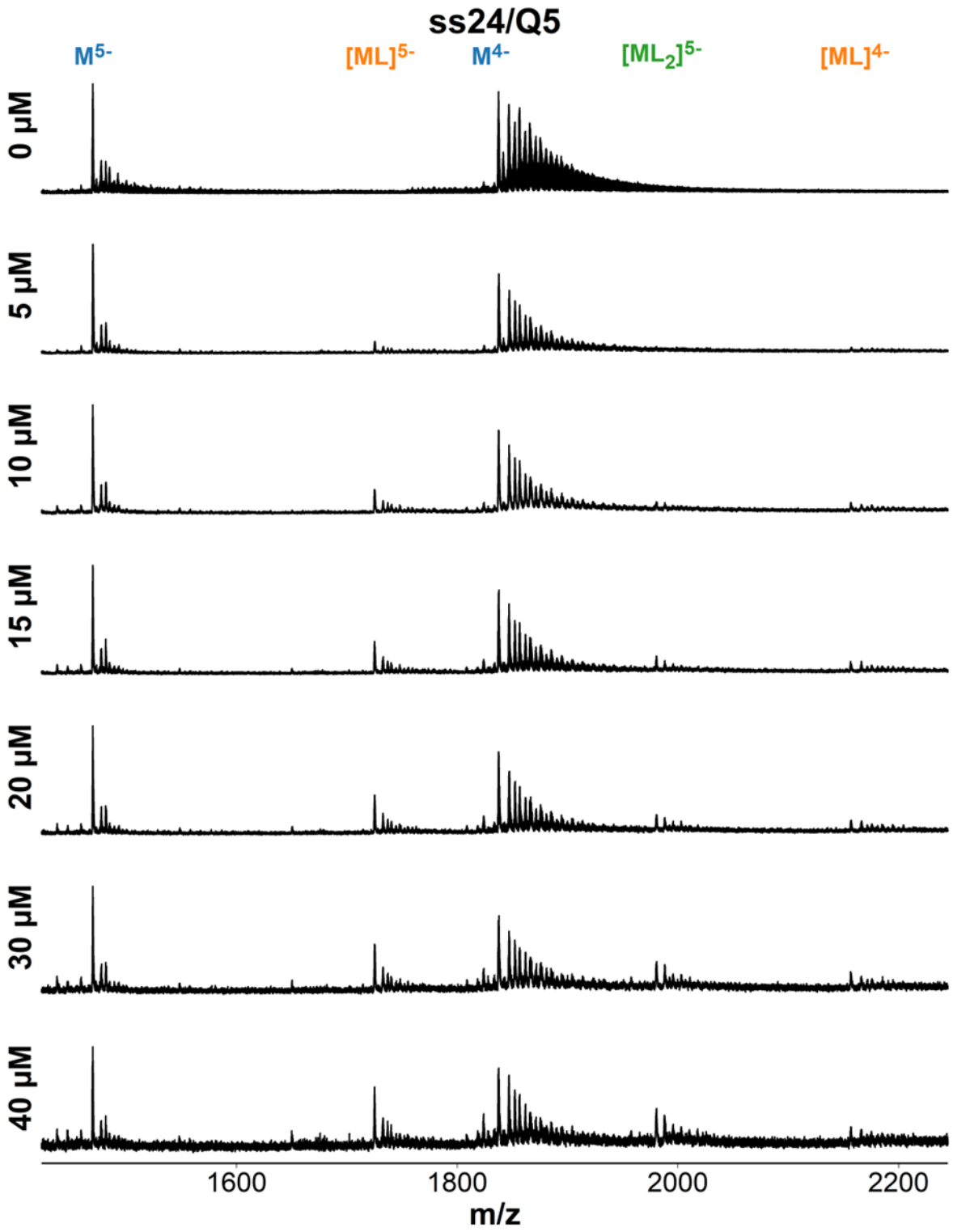


Figure S134 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QQQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

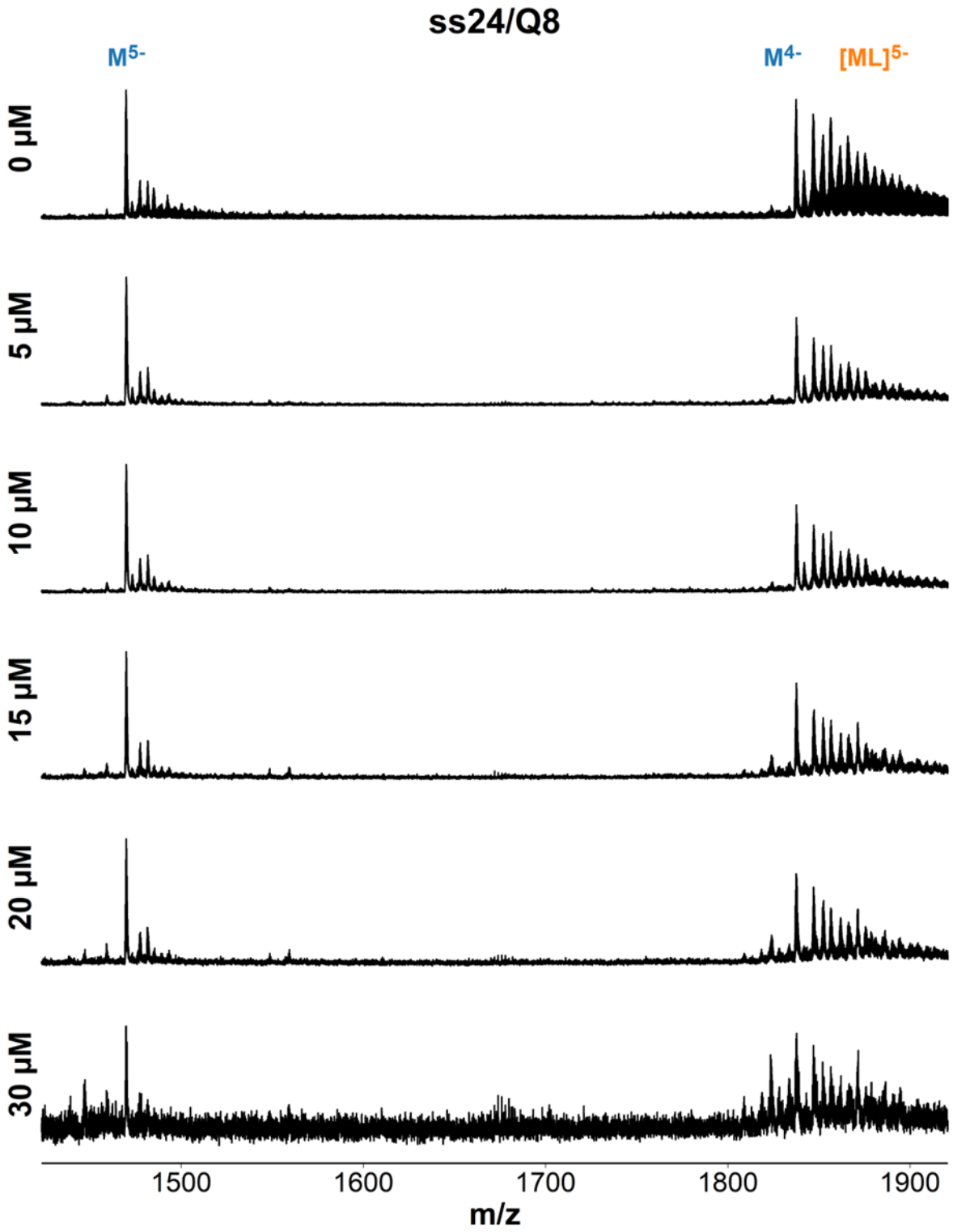


Figure S135 ESI-MS titration of ss24 (dTGCCATGCTACTGAGATGACGCTA) with foldamer QQQQQQQQ. Samples contain 10 µM DNA, 0-30 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

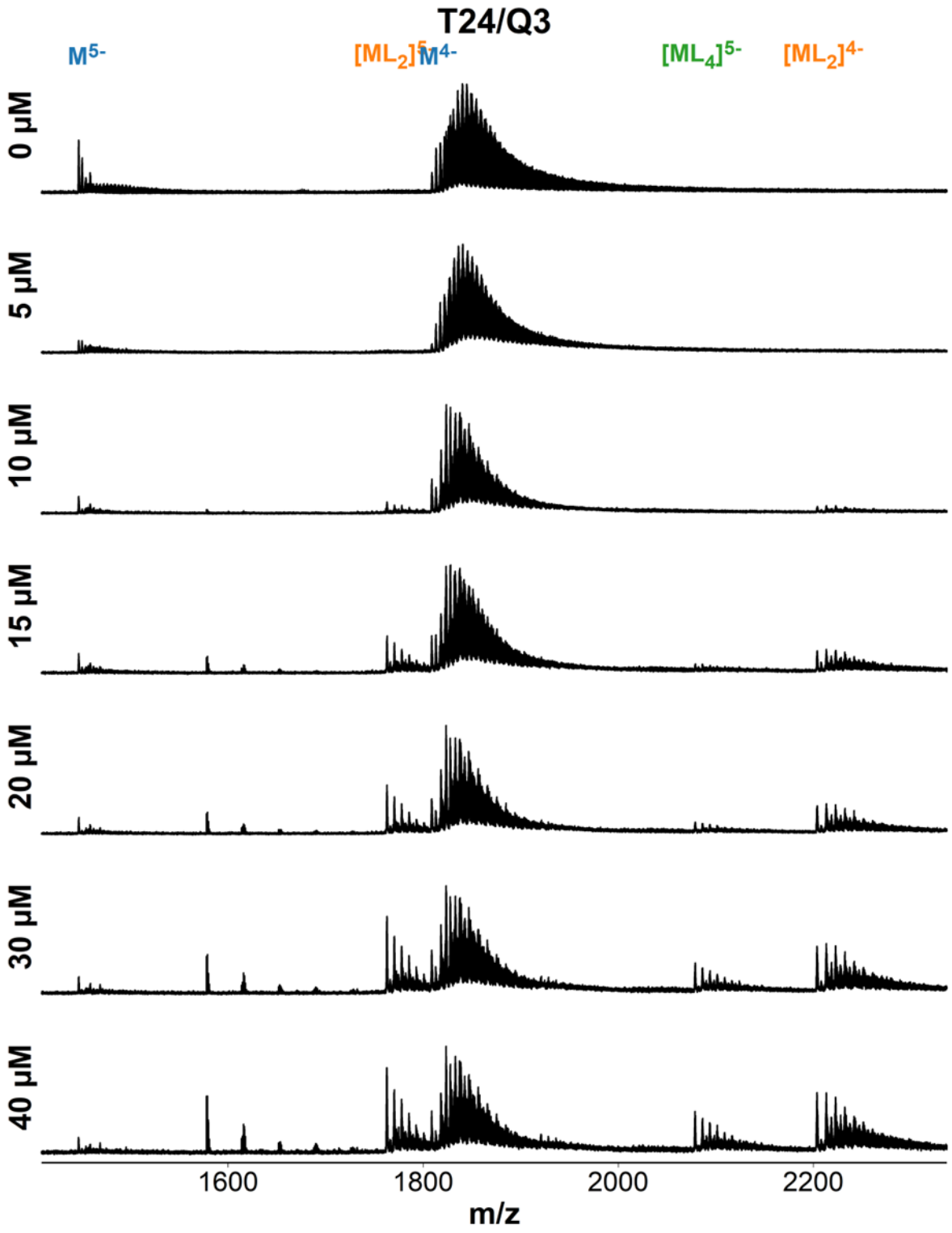


Figure S136 ESI-MS titration of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) with foldamer QQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

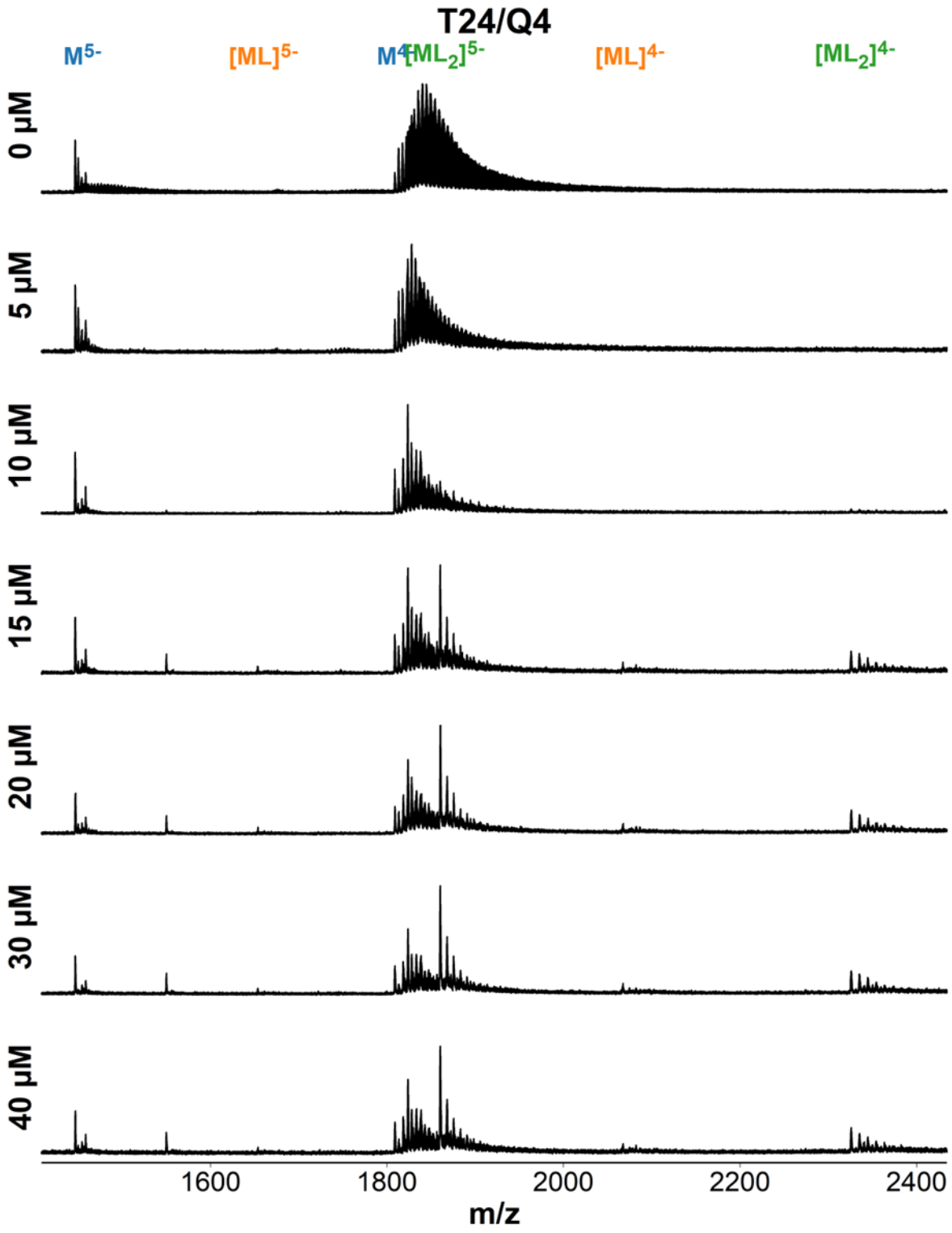


Figure S137 ESI-MS titration of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) with foldamer QQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

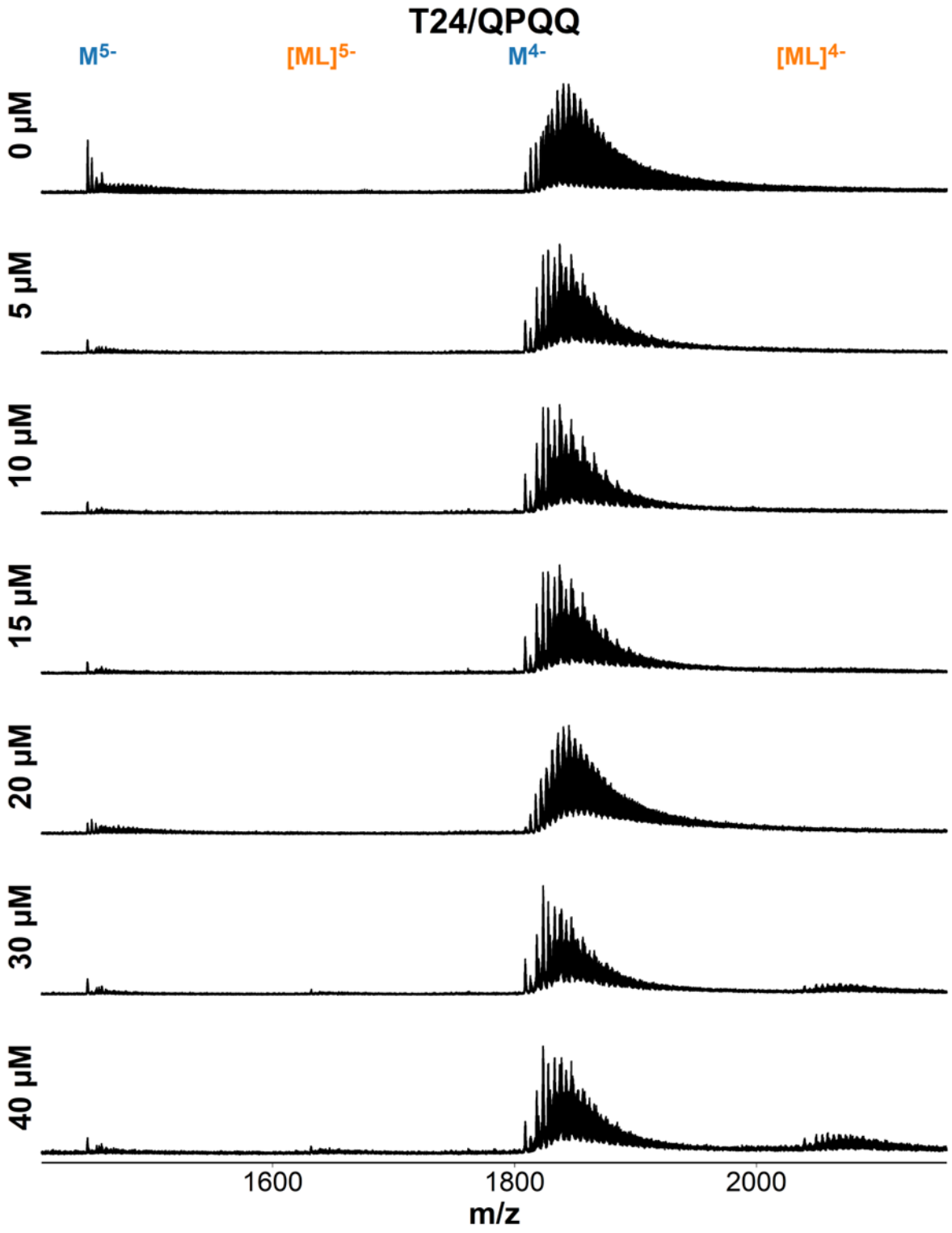


Figure S138 ESI-MS titration of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) with foldamer QQPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

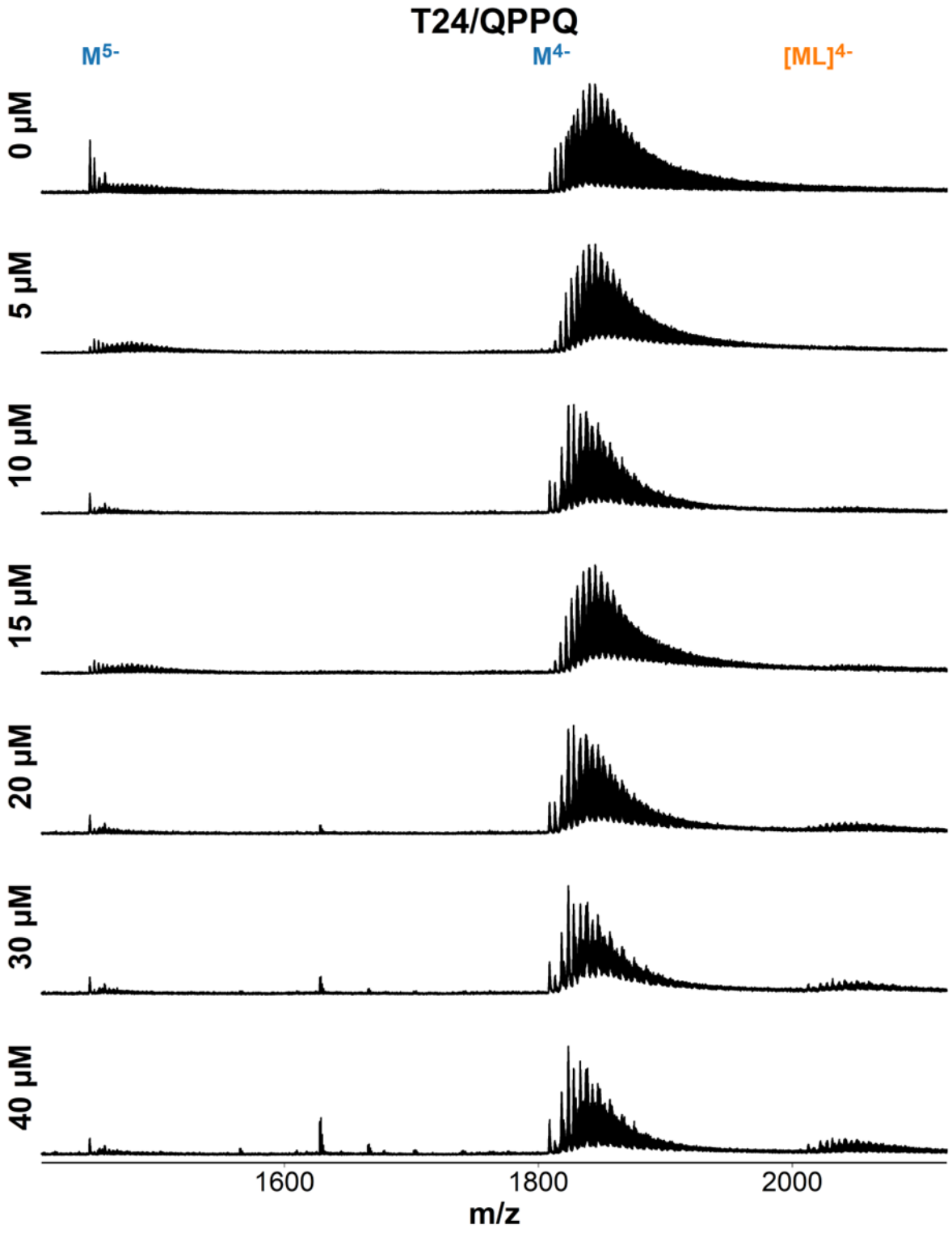


Figure S139 ESI-MS titration of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) with foldamer QPPQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

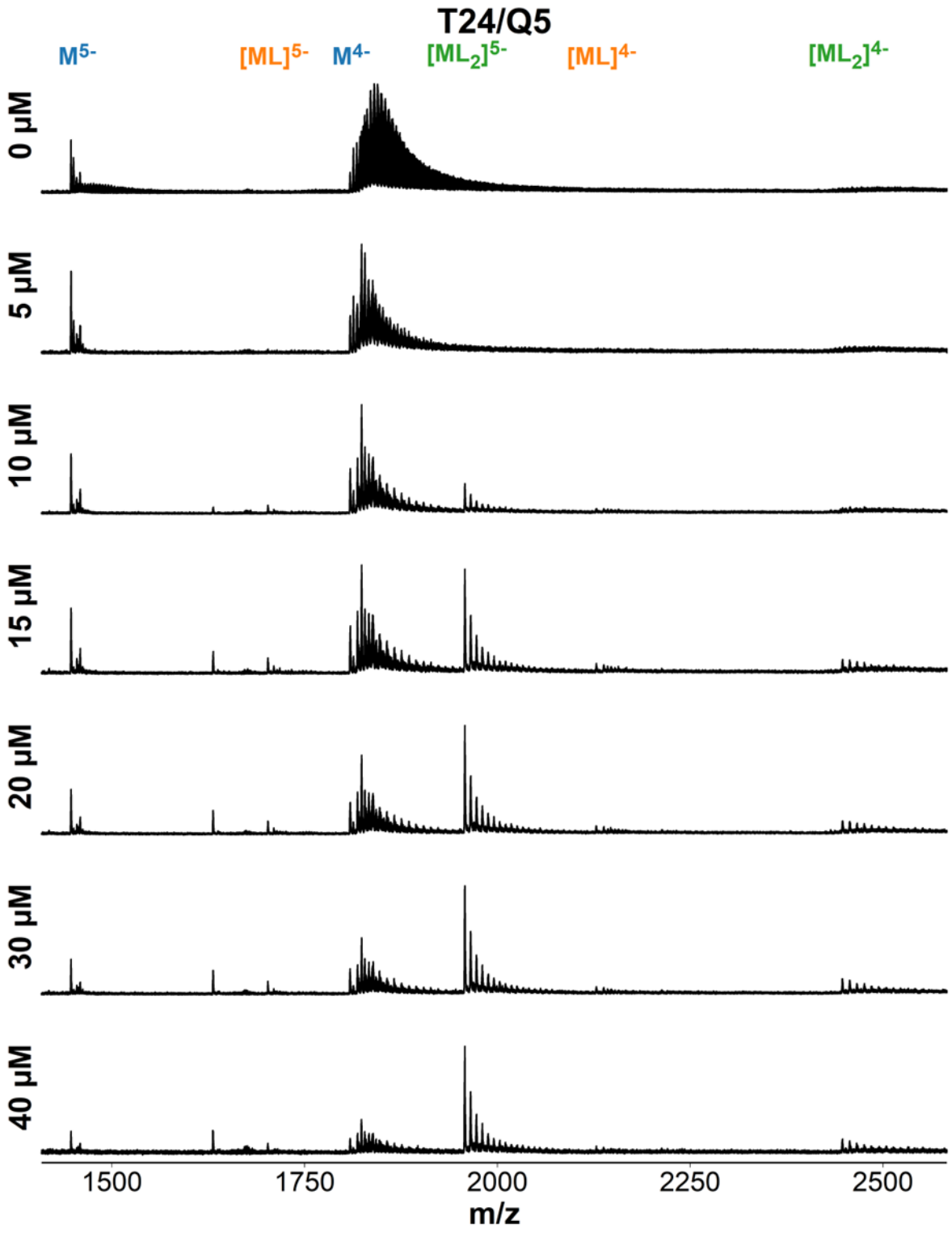


Figure S140 ESI-MS titration of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) with foldamer QQQQQ. Samples contain 10 µM DNA, 0-40 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

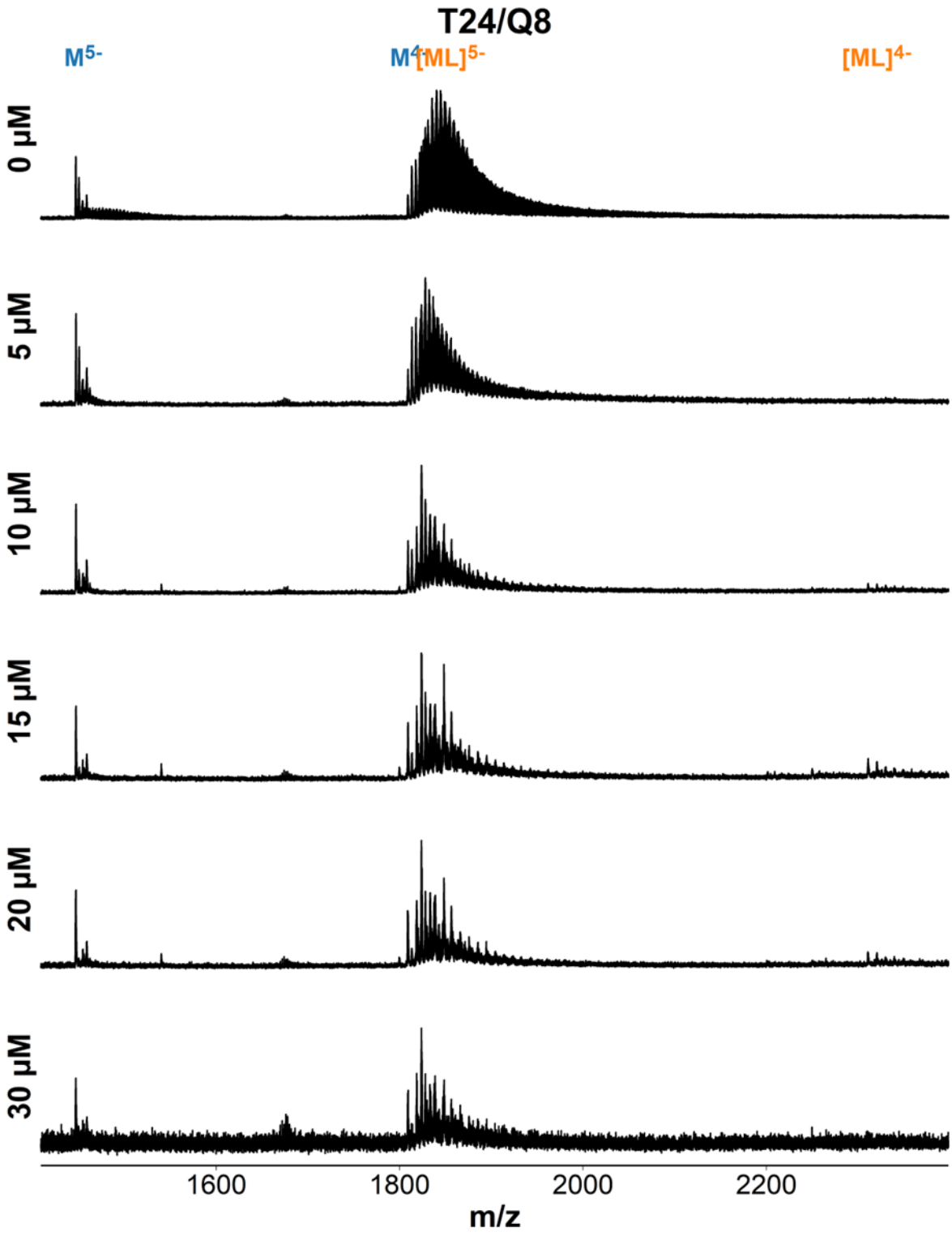


Figure S141 ESI-MS titration of T24 (dTTTTTTTTTTTTTTTTTTTTTTT) with foldamer QQQQQQQQ. Samples contain 10 µM DNA, 0-30 µM ligand, 0.5 mM KCl, 100 mM TMAA (pH 6.8).

T24 induces CD on Q_n-type foldamers

Among our ligand screening results, we noticed that Q-mers (i.e. Q_3, Q_4, Q_5, Q_8) form high-affinity 2:1 complexes (low µM K_D) with single-stranded sequence T24. We investigated this phenomenon via CD (Figure S142).

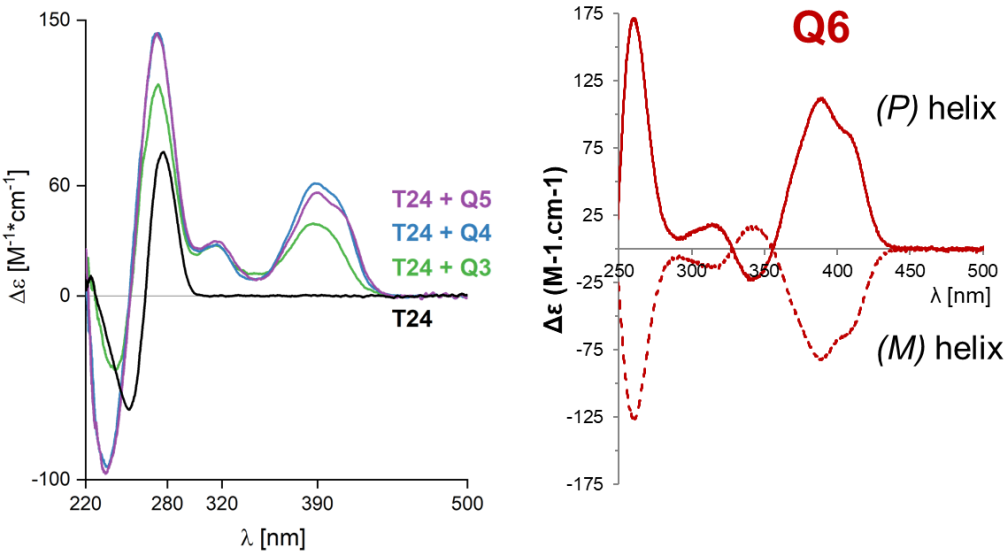


Figure S142 ***Left****: CD signatures of dT_{24} with different Q-mer foldamers. Samples contain: 10 µM T24, 40 µM Q_3/30 µM Q_4/24 µM Q_5 (normalized to 120 µM Q monomer) and 100 mM TMAA (pH 6.8).* ***Right****: Enantiomer-separated CD signatures of the (M) and (P) helix for Q6.*

In absence of T24, Q_3/Q_4/Q_5 have no CD signature, since they are racemic. But with T24, we see CD bands that are consistent with the *(P)*-helix. We do not know that kind of structure T24 and Q-mer foldamers form; our attempt of crystallizing a T24/Q_4 complex was unsuccessful. Based on the high 2:1 cooperativity and the induced CD on the (right-handed) *(P)*-helix we speculate that T24 and Q_n foldamers associate into some sort of double helix, with the foldamer dimerizing to accommodate the full length of the DNA sequence.

U-rich motifs play a role in RNA expression,[60,61] which is why we attempted to reproduce our results on U24 (Figure S143). Alas, we could not see the same induced CD when switching from dT to rU.

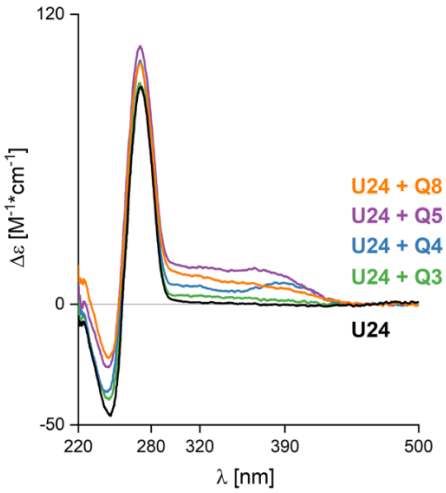


Figure S143 CD signatures of rU_{24} with different Q-mer foldamers. Samples contain: 10 µM U24, 40 µM Q_3/30 µM Q_4/24 µM Q_5/15 µM Q_8 (normalized to 120 µM Q monomer) and 100 mM TMAA (pH 6.8).

Foldamer-induced disruption of G-quadruplex

Figure S63 shows that the decline in DNA signal response is proportional to the number of Q units in the foldamer. Each Q unit carries a positive charge, so we suspect that the loss of signal is due to ligand-induced DNA aggregation, with the ligand acting as a sort of flocculant that helps overcome charge repulsions between DNA polyanions. To test this hypothesis, we picked T30177TT, a parallel G-quadruplex that is stable even in low K^+ and Q8, the ligand with the fastest decrease in response factor. If the ligand causes DNA aggregation, the CD signature of T30177TT should decrease the more ligand is added.

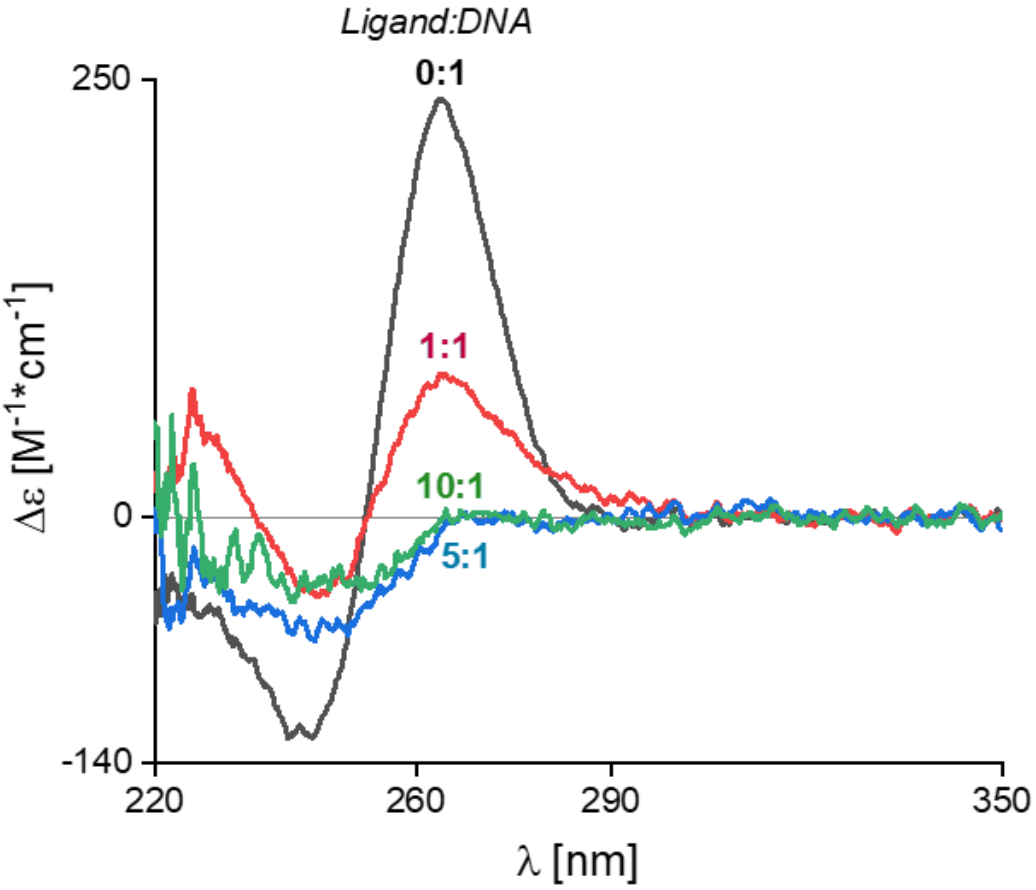


*Figure S144 Tracking the ligand-induced aggregation of T30177TT (dTT**GT**GG**T**GGG**T**GGG**T**GGG**T**) by adding more and more Q8. Samples contain 1 µM DNA, 0/1/5/10 µM Q_8, 1 mM KCl and 100 mM TMAA (pH 6.8).*

The experimental results are consistent with our hypothesis, indicating that G-quadruplexes with too many positive charges run the risk of causing DNA aggregation, or even precipitation (aka 'salting out' a biomolecule by neutralizing all of its charges).

X-ray crystallography

Table S5 Data acquisition and refinement parameters for the 222T/QQPQ crystal. Parenthesis show statistics for the highest-resolution shell.

Space group	P42
Unit cell (Å)	a = b = 32.71, c = 61.01, α = β = γ = 90.000°
Wavelength (Å)	0.979
Resolution range (Å)	30.5-2.512 (2.602-2.512)
Completeness (%)	99.10 (100.0)
Total reflections	29748 (2985)
Unique reflections	2226 (218)
Multiplicity	13.4 (13.7)
Average I/σ	21.61 (4.09)
Wilson B-Factor	58.79
R-merge	0.1469 (1.054)
R-means	0.1531 (1.095)
R-pim	0.0426 (0.2952)
CC ½	0.999 (0.872)
CC*	1 (0.965)
Reflections used in refinement	2207 (218)
Reflections used for R-free	221 (22)
R-work	0.2563 (0.2933)
R-free	0.3037 (0.4455)
CC (work)	0.902 (0.773)
CC (free)	0.990 (0.336)
r.m.s. bond deviation (Å)	0.018
r.m.s. angle deviation (°)	1.84
Clashscore	23.19
Number of non-hydrogen atoms	531
DNA atoms	388
QQPQ atoms	243
Other atoms (K^+ and Mg^{2+})	2
Average B factors	66.24
DNA atoms	61.87
QQPQ atoms	78.21
Other atoms (K^+ and Mg^{2+})	69.63

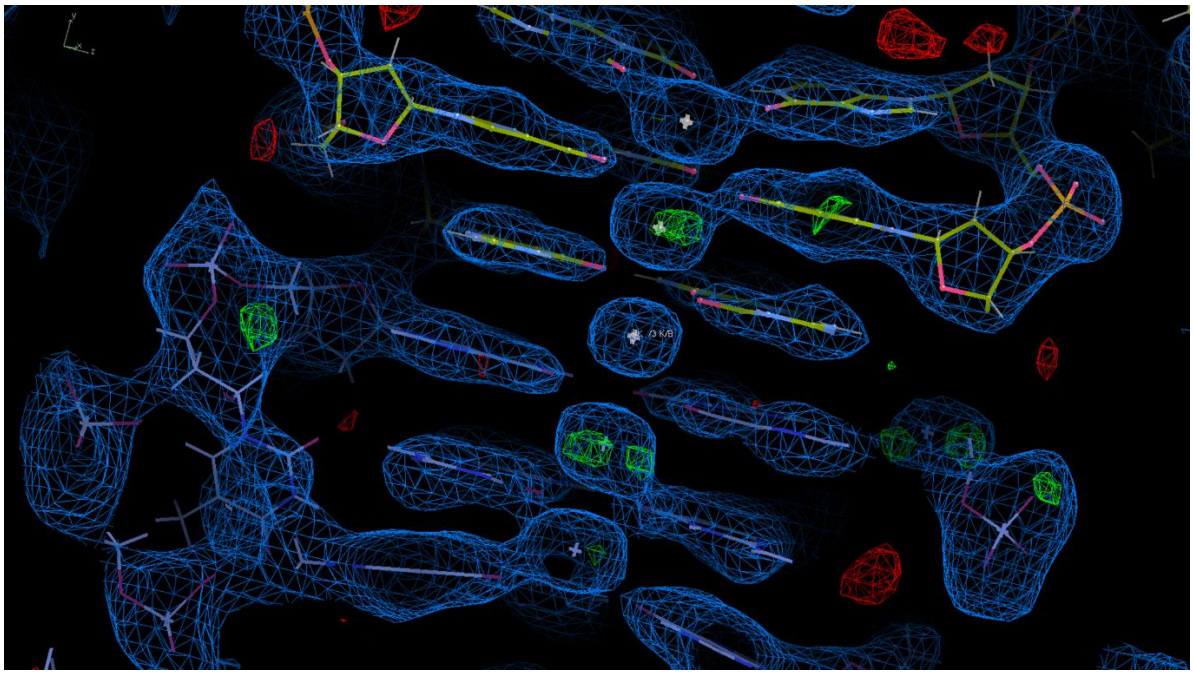


Figure S145 Electron density map of the G-quadruplex core structure

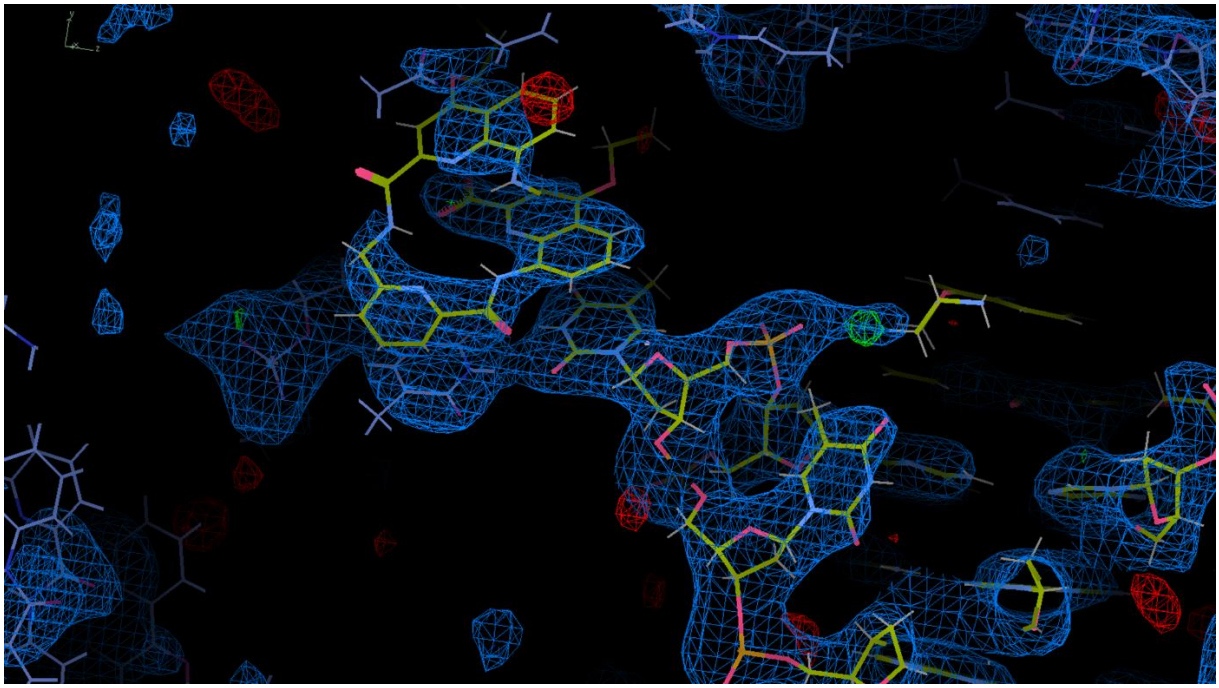


Figure S146 Electron density map of the QQPQ foldamer stacked on top of a thymine in the 222T loop region.

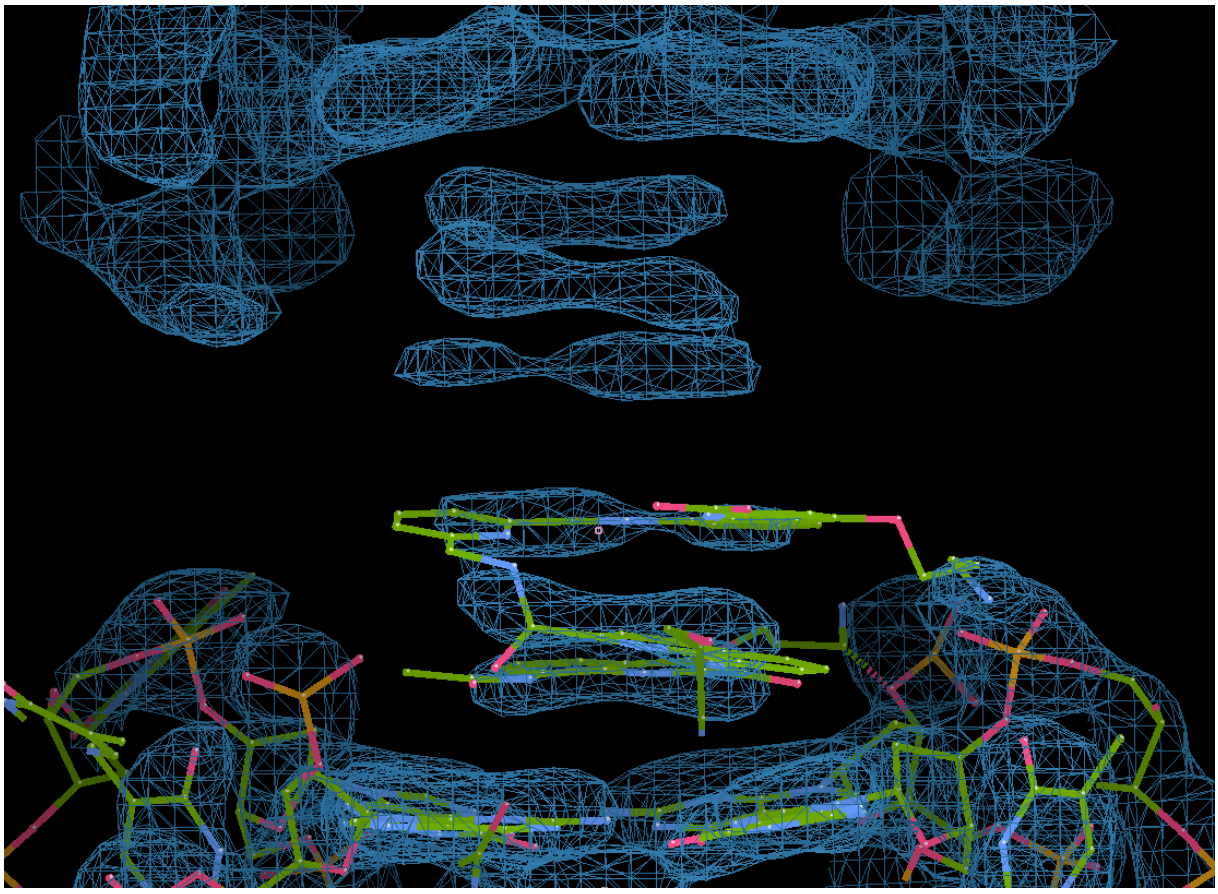


Figure S147 Electron density map of the QQPQ foldamer stacked onto the G-tetrad of 222T, showing two adjacent unit cells on top and on the bottom.

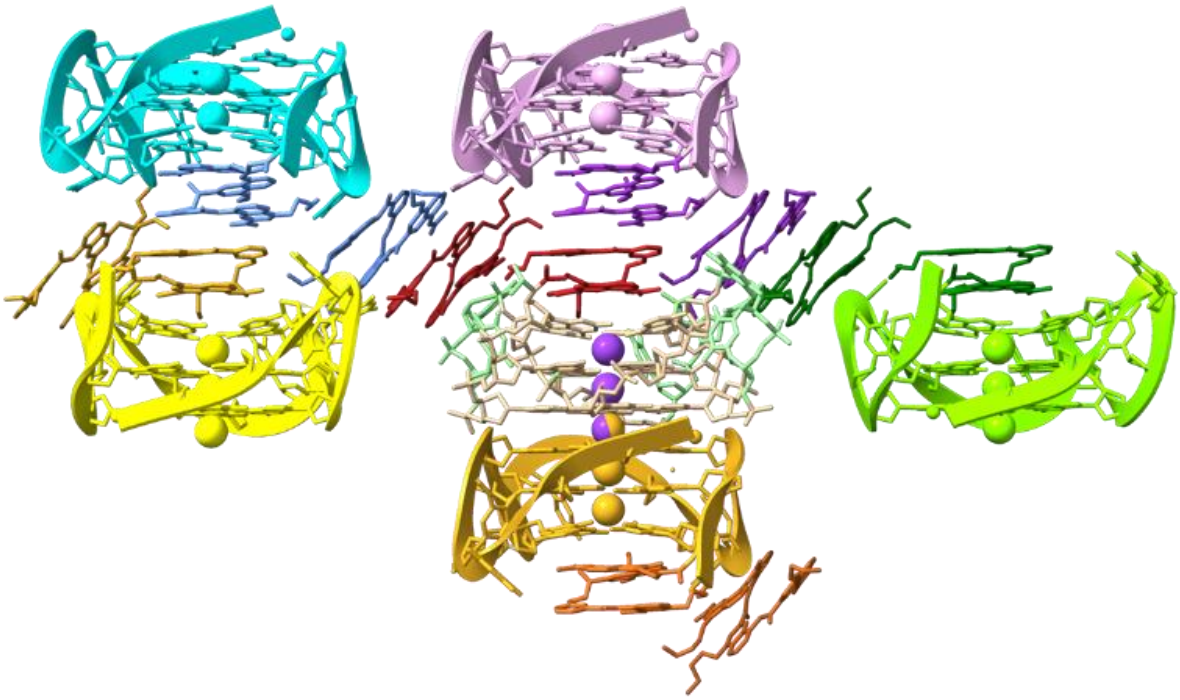


Figure S148 Crystal packing of the 222T/QQPQ crystal. Different colors show different unit cells.

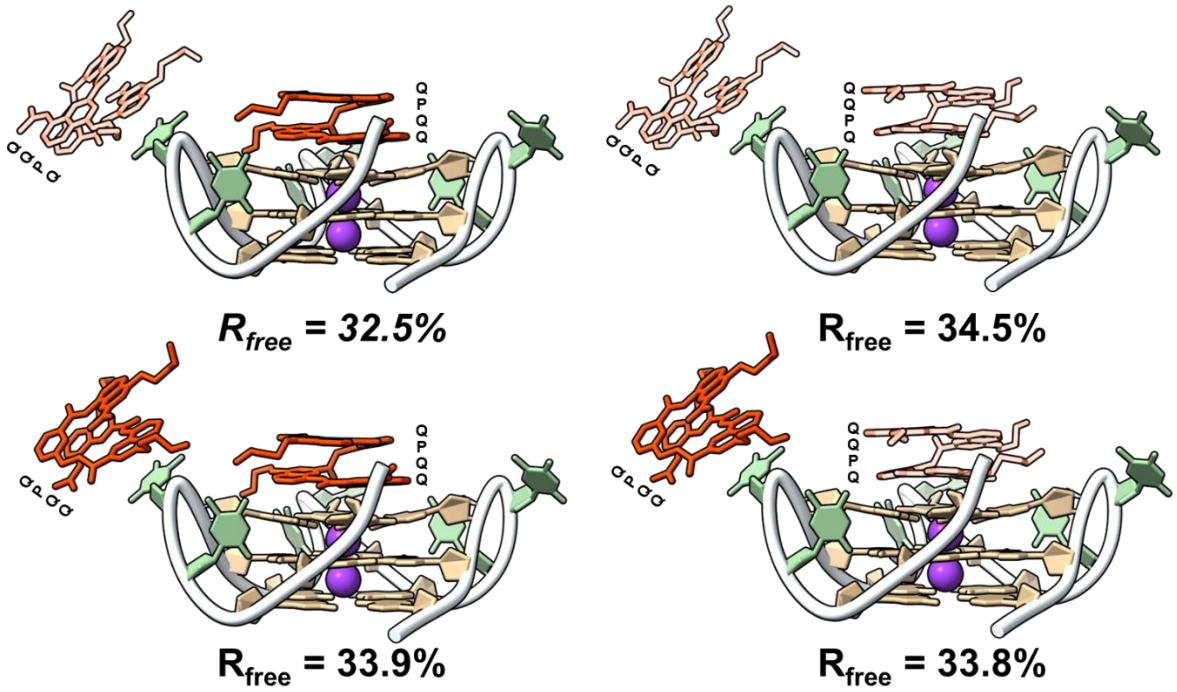


Figure S149 Flipping the foldamer between C- and N-terminal binding to test whether any conformation is statistically preferential. Schemes represent the four different models and their resulting R_{free} values.

1D and 2D NMR spectra

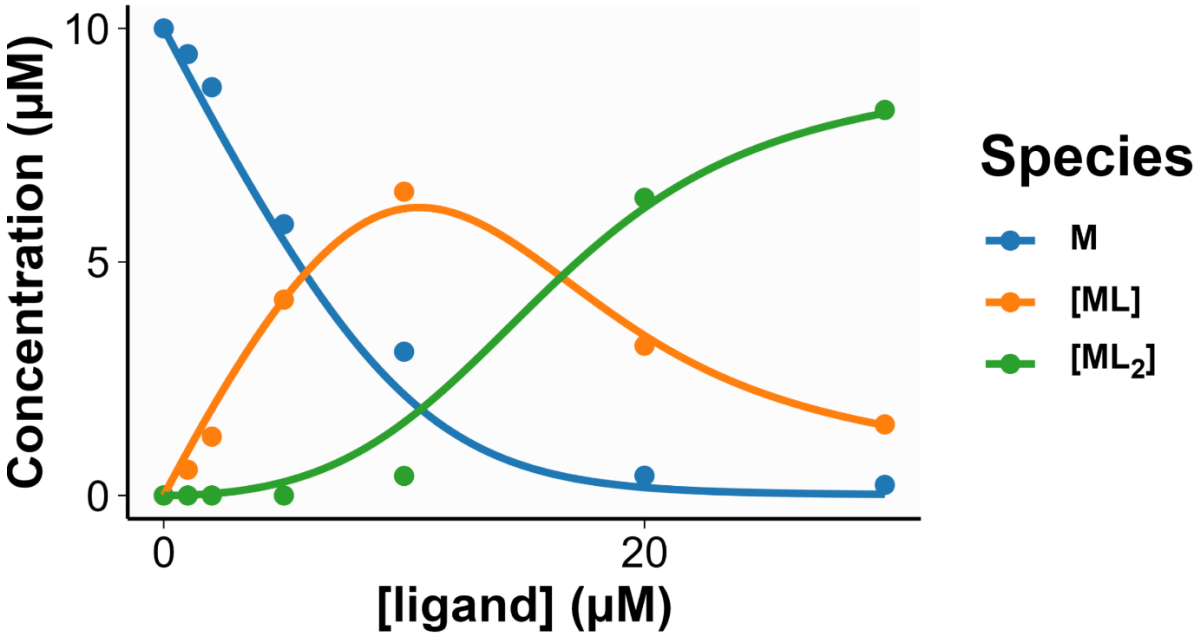


*Figure S150 Dynamic fitting of MS titration data for 2LK7 (dTT**G₃**T**G₃**T**G₃**T**G₃**T) and foldamer QQPQ with K_{D1} = 0.18 ± 0.08 μM and K_{D2} = 2.1 ± 0.7 μM. Samples contain 10 μM DNA, 0/1/2/5/10/20/30 μM QQPQ, 0.5 mM KCl and 100 mM TMAA (pH 6.8).*

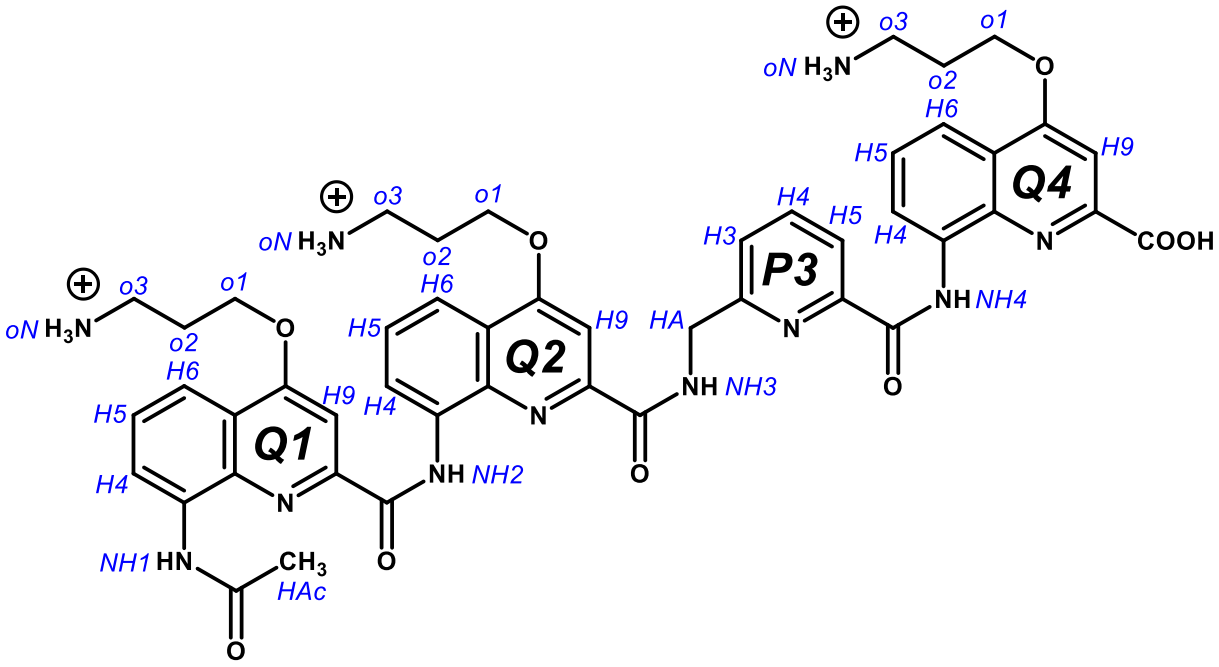


Figure S151 Proton labels in the QQPQ molecule.

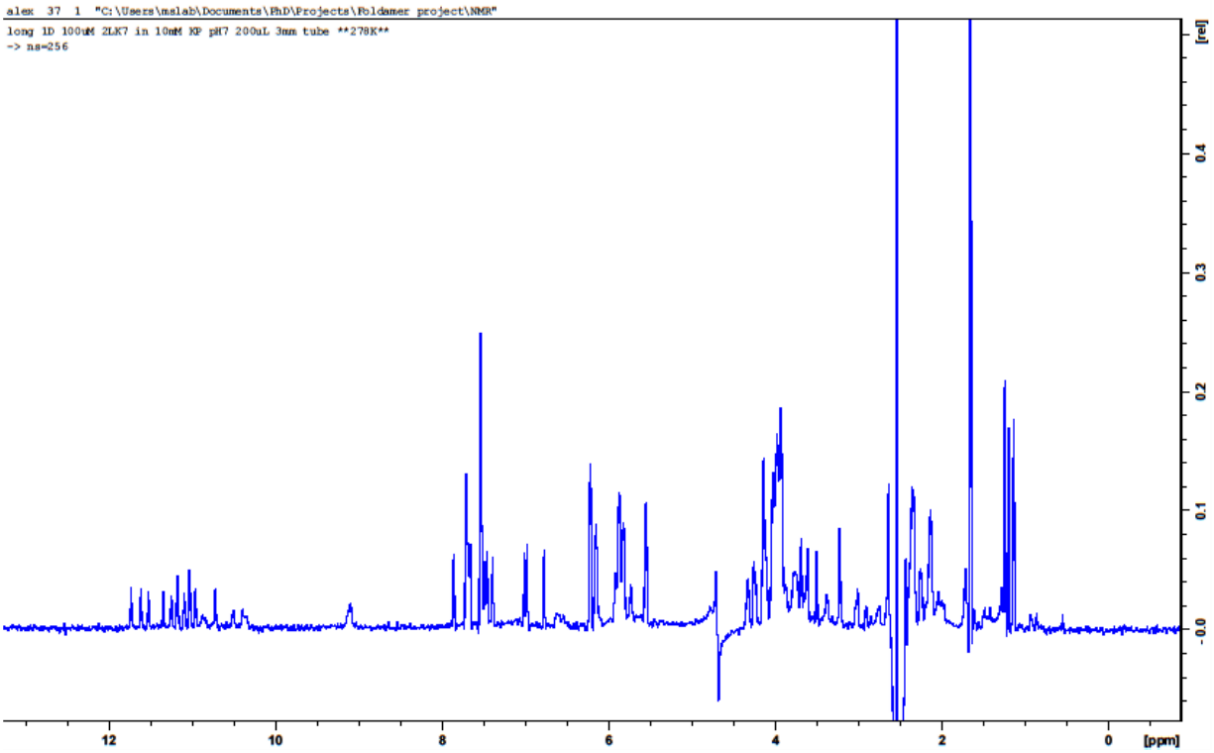


*Figure S152 Full 1D NMR spectrum of 2LK7 (dTT**G**$_3$T**G**$_3$T**G**$_3$T**G**$_3$T) on a Bruker Avance 700 MHz at 278 K. Sample matrix: 100 µM DNA, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O. Water signal was suppressed by excitation sculpting.*

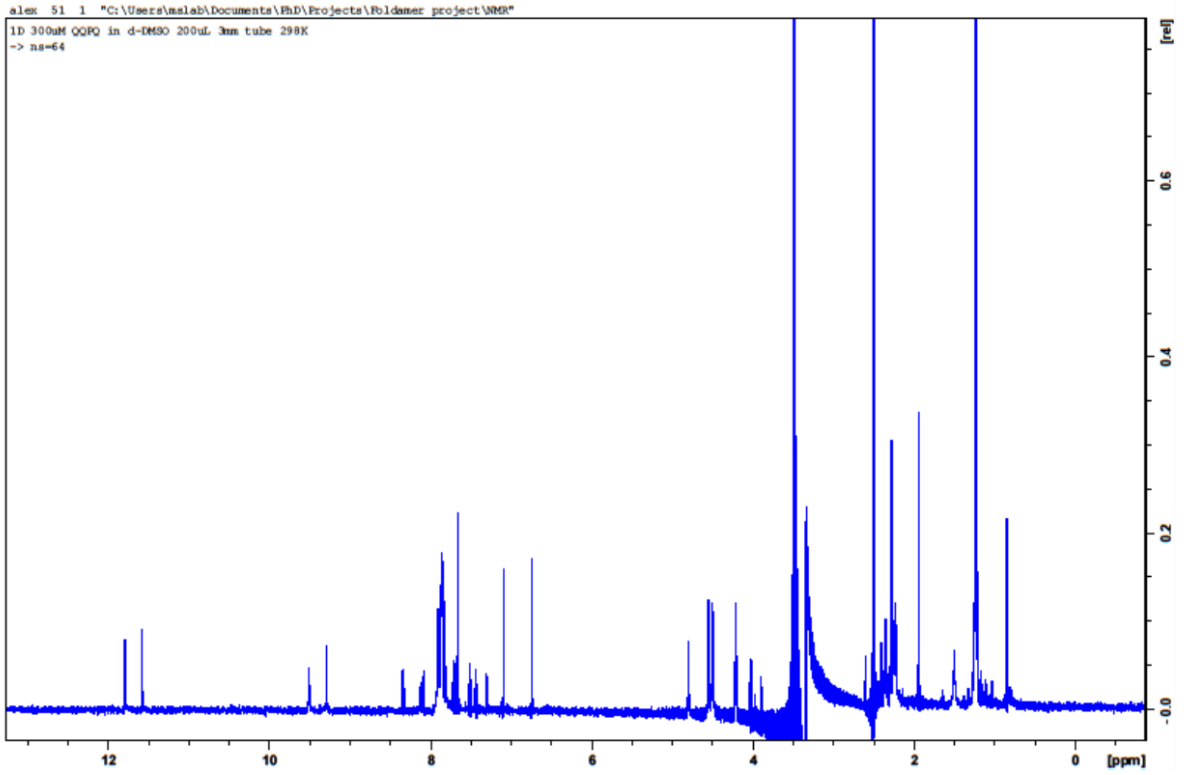


Figure S153 Full 1D NMR spectrum of Foldamer QQPQ on a Bruker Avance 700 MHz at 298 K. Sample matrix: 300 µM ligand in d6-DMSO.

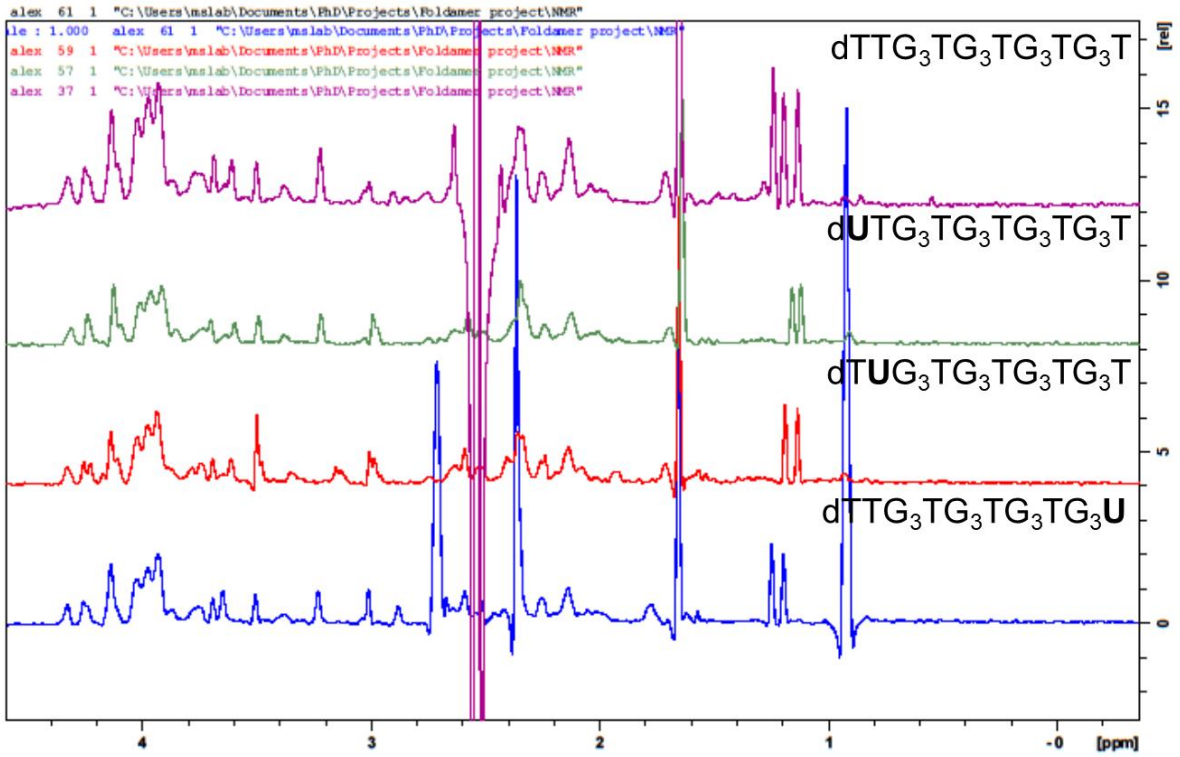


Figure S154 1D NMR spectra of 2LK7 and labeled derivatives on a Bruker Avance 700 MHz at 278 K, showing the low shift region. Sample matrix: 100 µM DNA, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O. Water signal was suppressed by excitation sculpting.

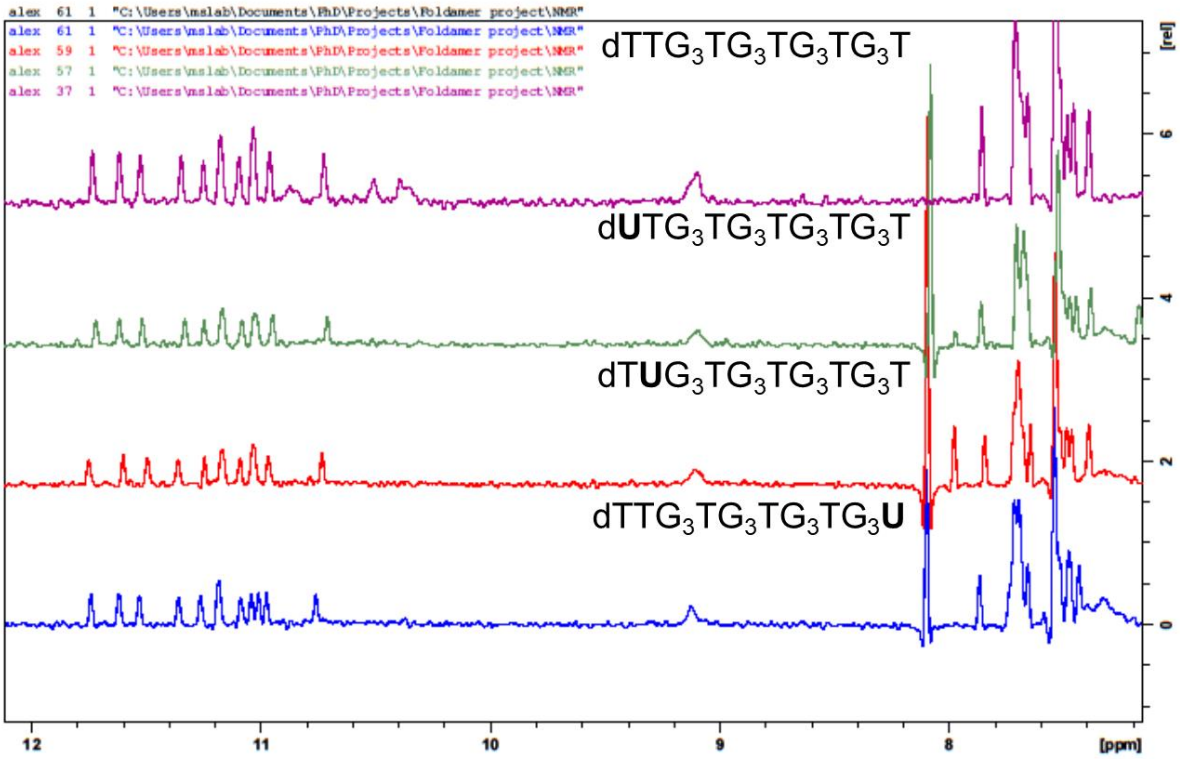


Figure S155 1D NMR spectra of 2LK7 and labeled derivatives on a Bruker Avance 700 MHz at 278 K, showing the high shift region. Sample matrix: 100 µM DNA, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O. Water signal was suppressed by excitation sculpting.

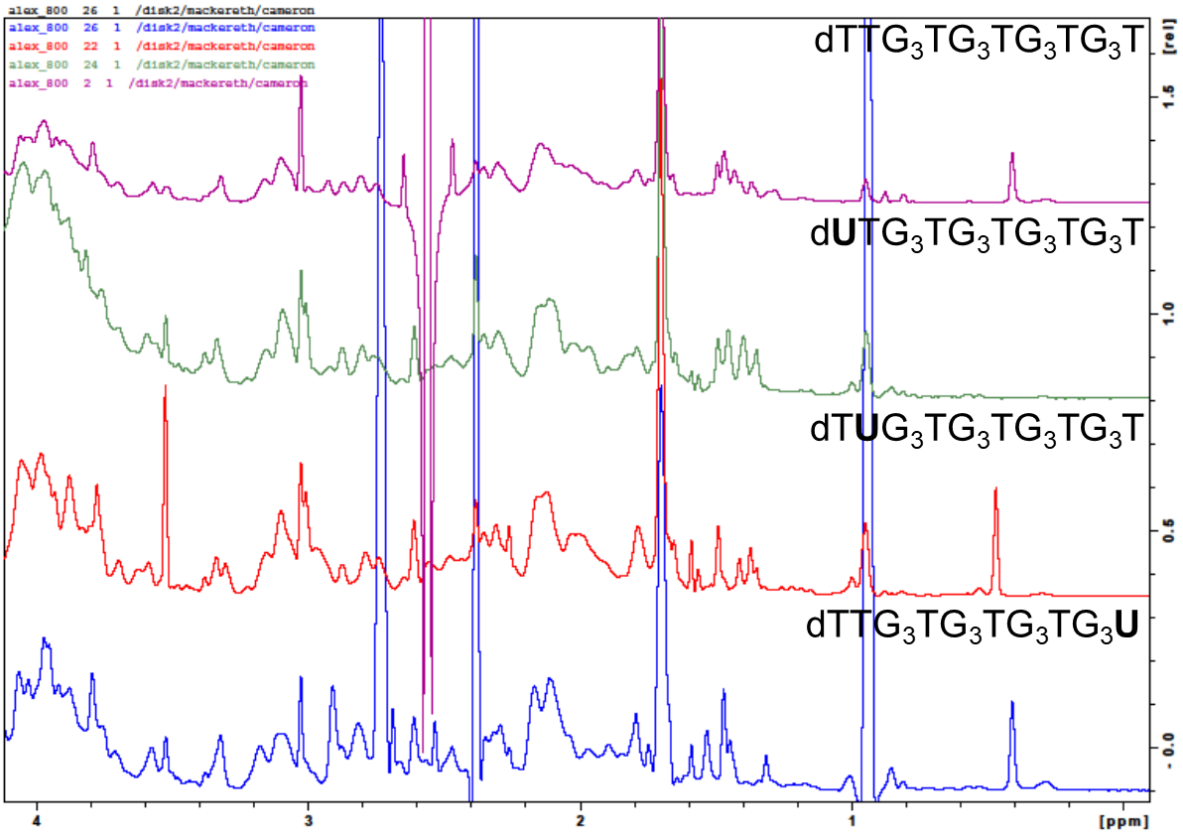


Figure S156 1D NMR spectra of 2LK7 and labeled derivatives with QQPQ on a Bruker Avance 700 MHz at 278 K, showing the low shift region. Sample matrix: 100 µM DNA,300 µM ligand. 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O.

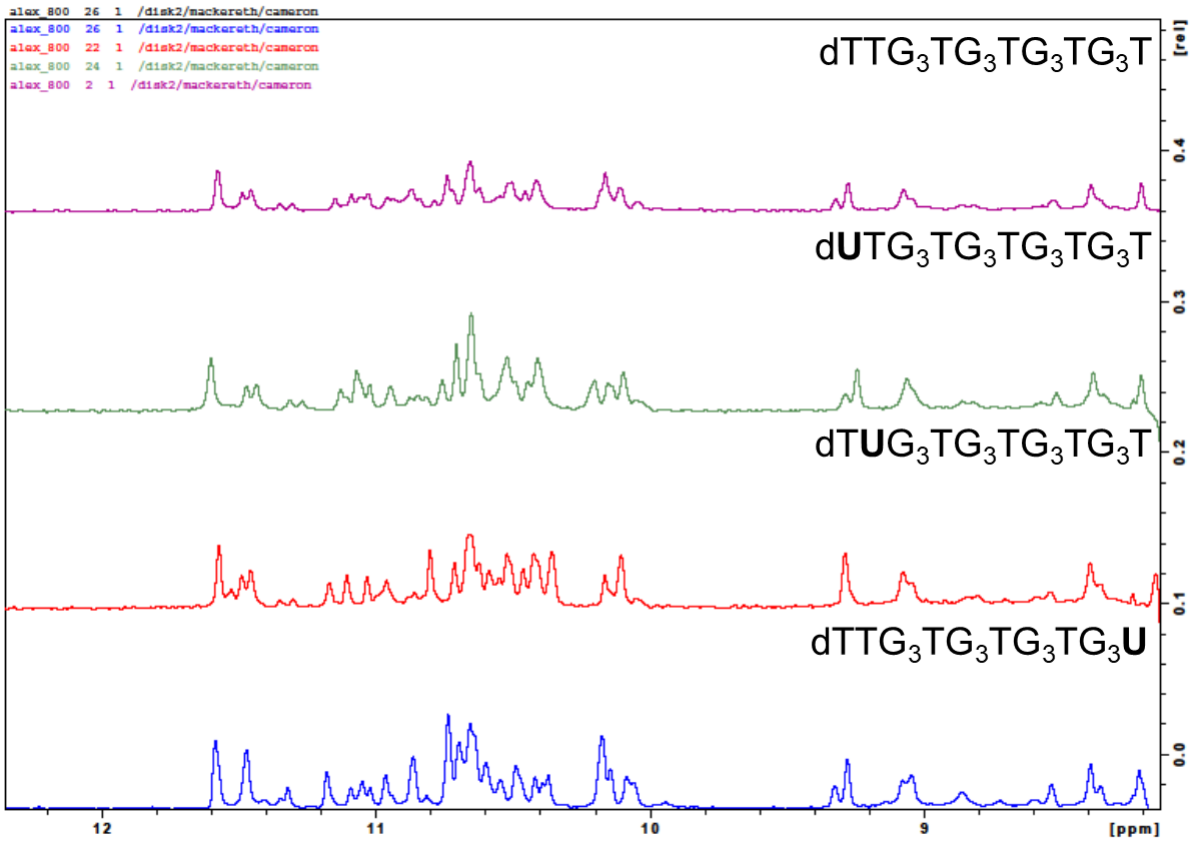


Figure S157 1D NMR spectra of 2LK7 and labeled derivatives with QQPQ on a Bruker Avance 700 MHz at 278 K, showing the high shift region. Sample matrix: 100 µM DNA,300 µM ligand. 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O.

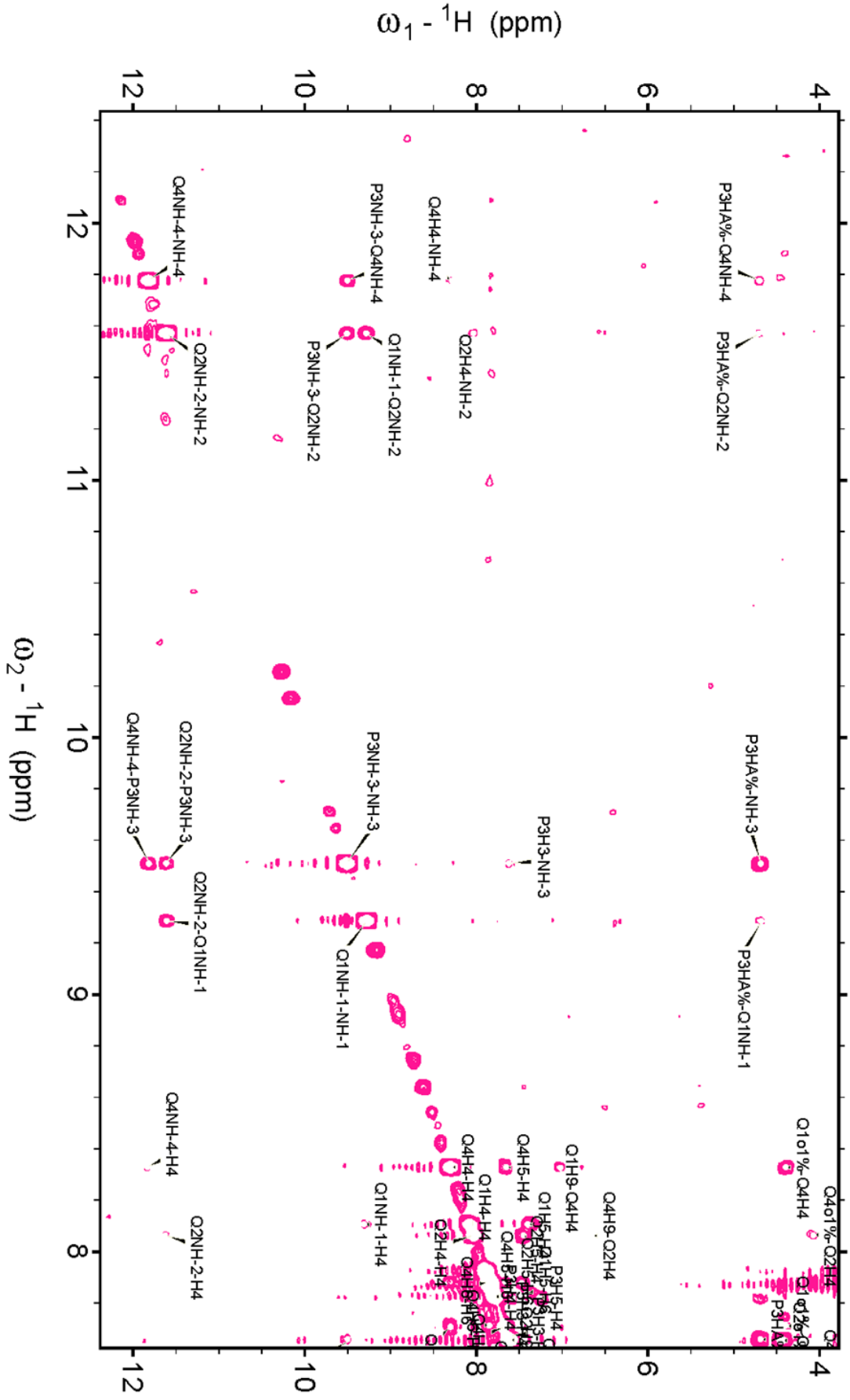


Figure S158 ^{1}H-^{1}H NOESY of QQPQ on a Bruker Avance-III 800 MHz; high shift region containing amide protons (NH). Sample matrix: 300 µM QQPQ in d6-DMSO.

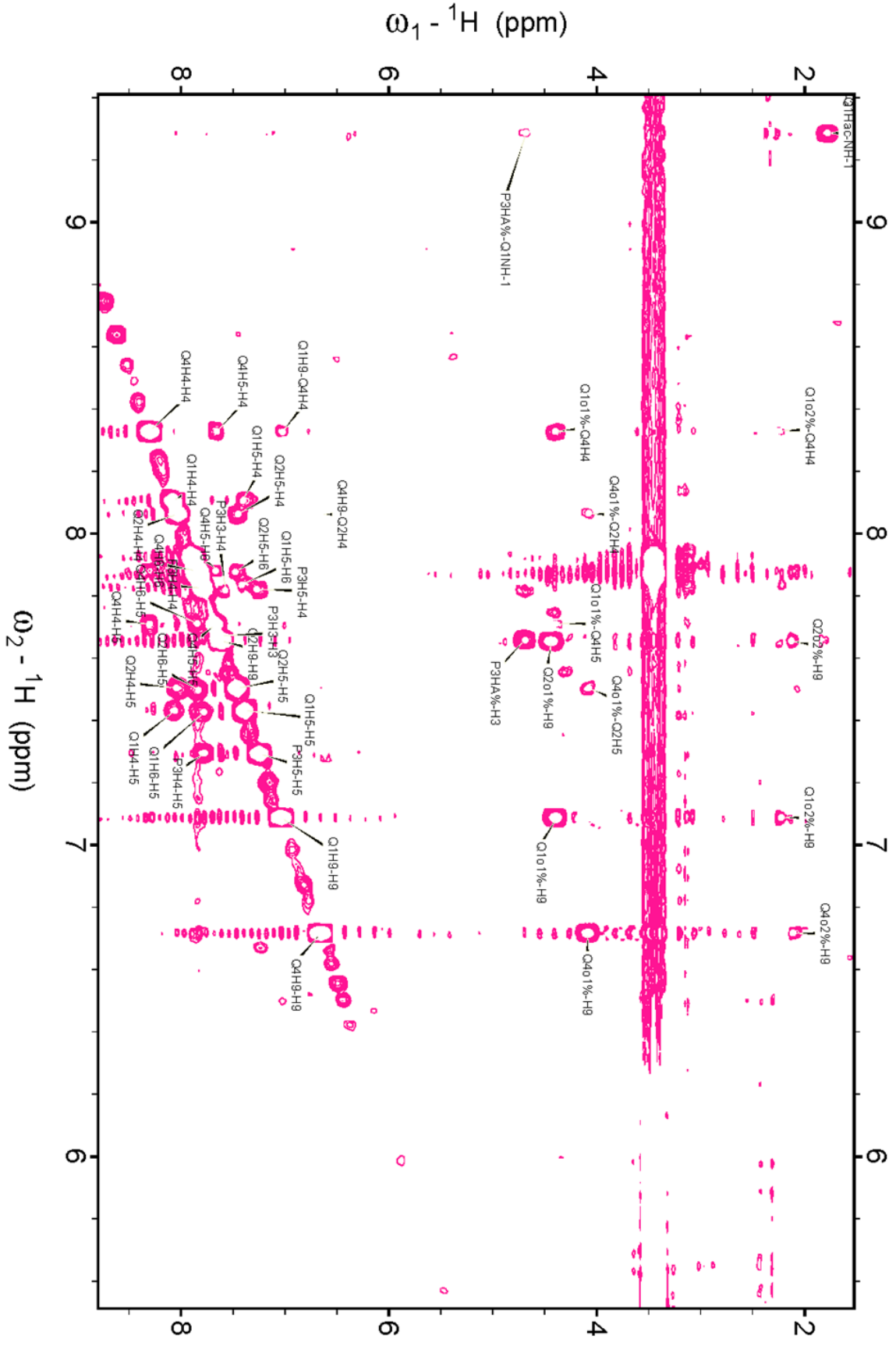


Figure S159 ^{1}H-^{1}H NOESY of QQPQ on a Bruker Avance-III 800 MHz; middle region containing the aromatic protons in the foldamer core (H4, H5, H6, H9). Sample matrix: 300 µM QQPQ in d6-DMSO.

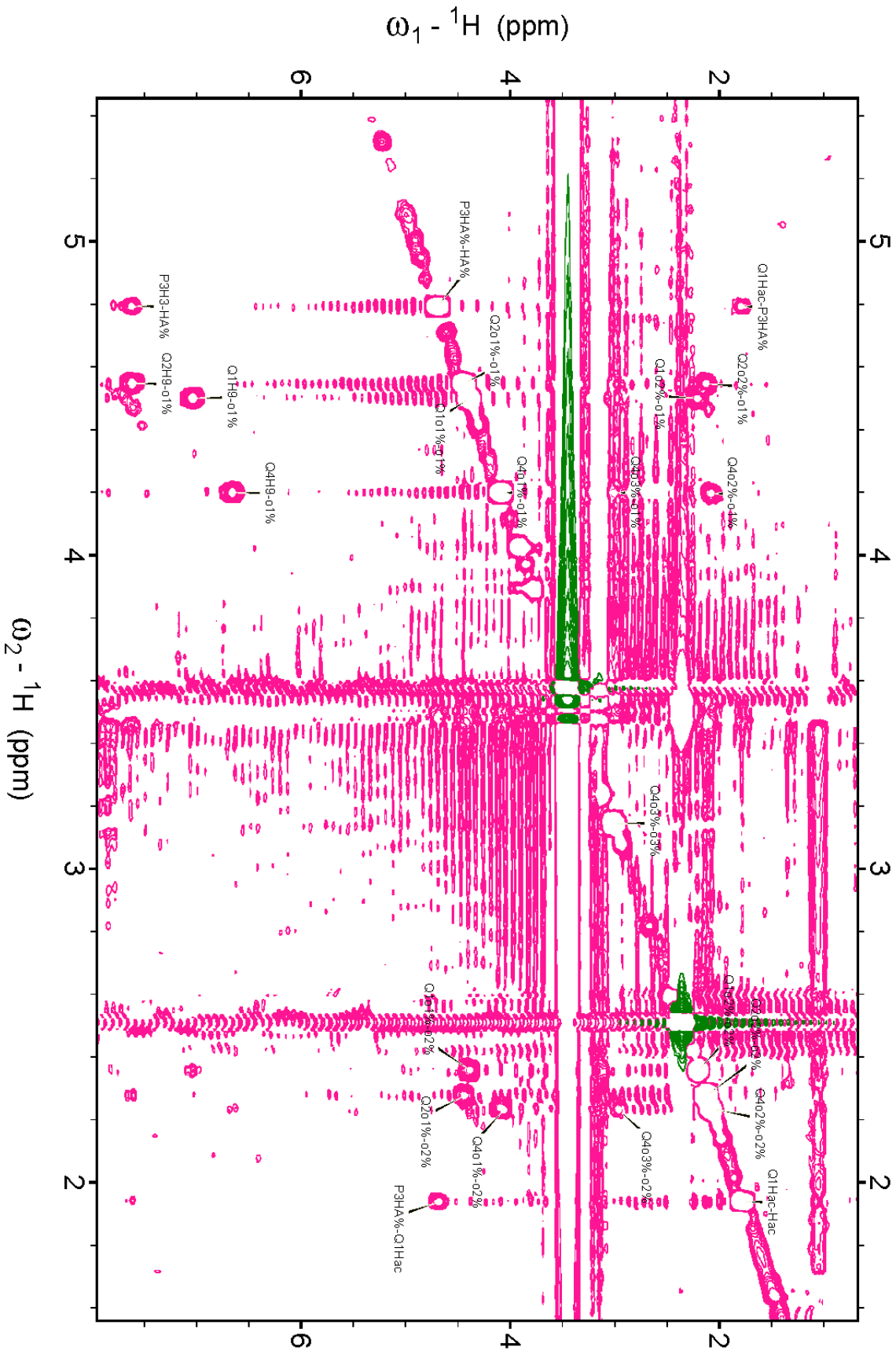


Figure S160 ^{1}H-^{1}H NOESY of QQPQ on a Bruker Avance-III 800 MHz; low shift region containing the aliphatic protons, which are mostly in the side chains(o1, o2, o3, HA, HAc). Sample matrix: 300 µM QQPQ in d6-DMSO.

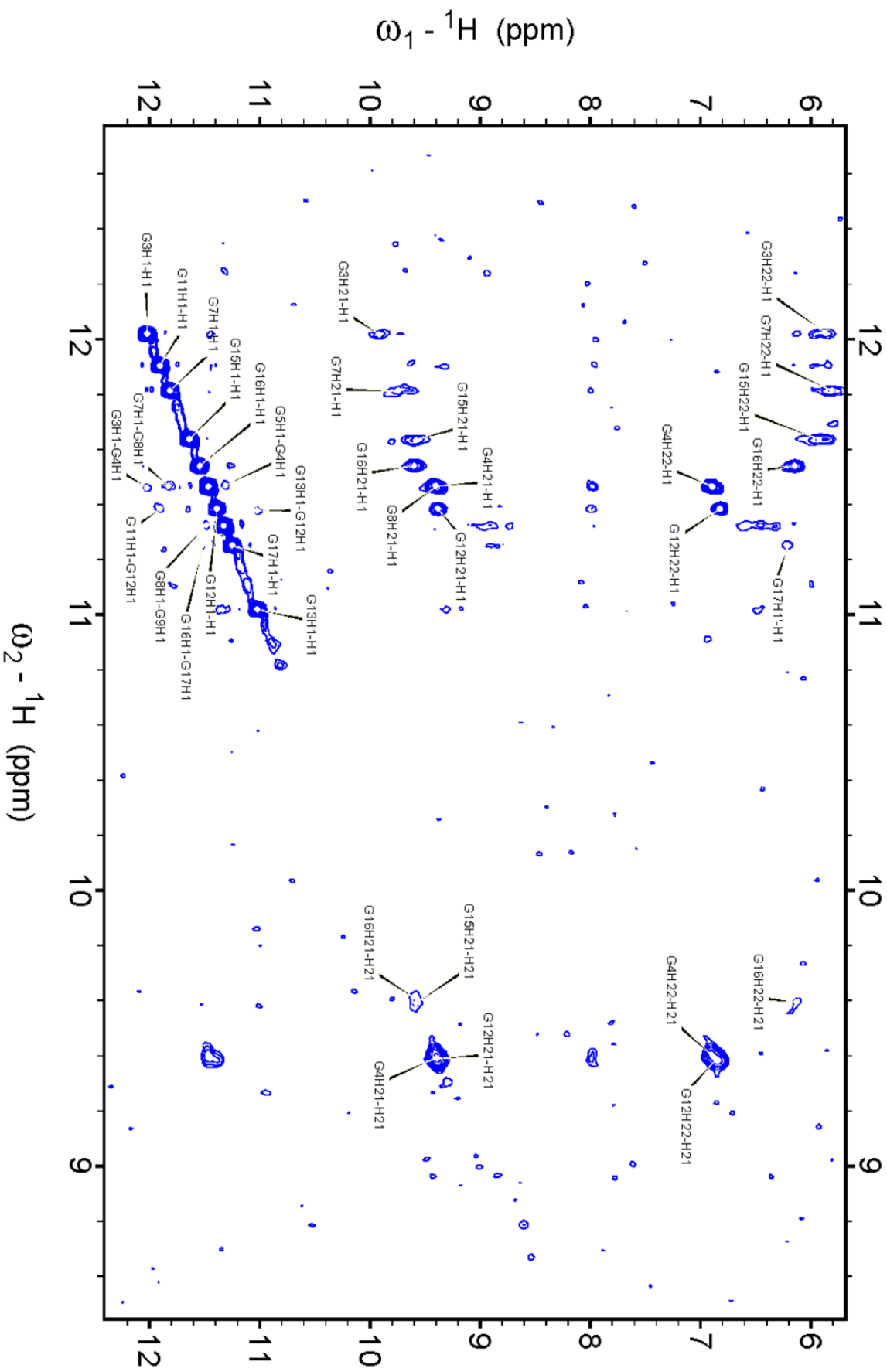


Figure S161 ^{1}H-^{1}H NOESY of 2LK7 on a Bruker Avance-III 800 MHz; high shift region showing the guanine H1/H2 protons. Sample matrix: 100 µM 2LK7, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O.

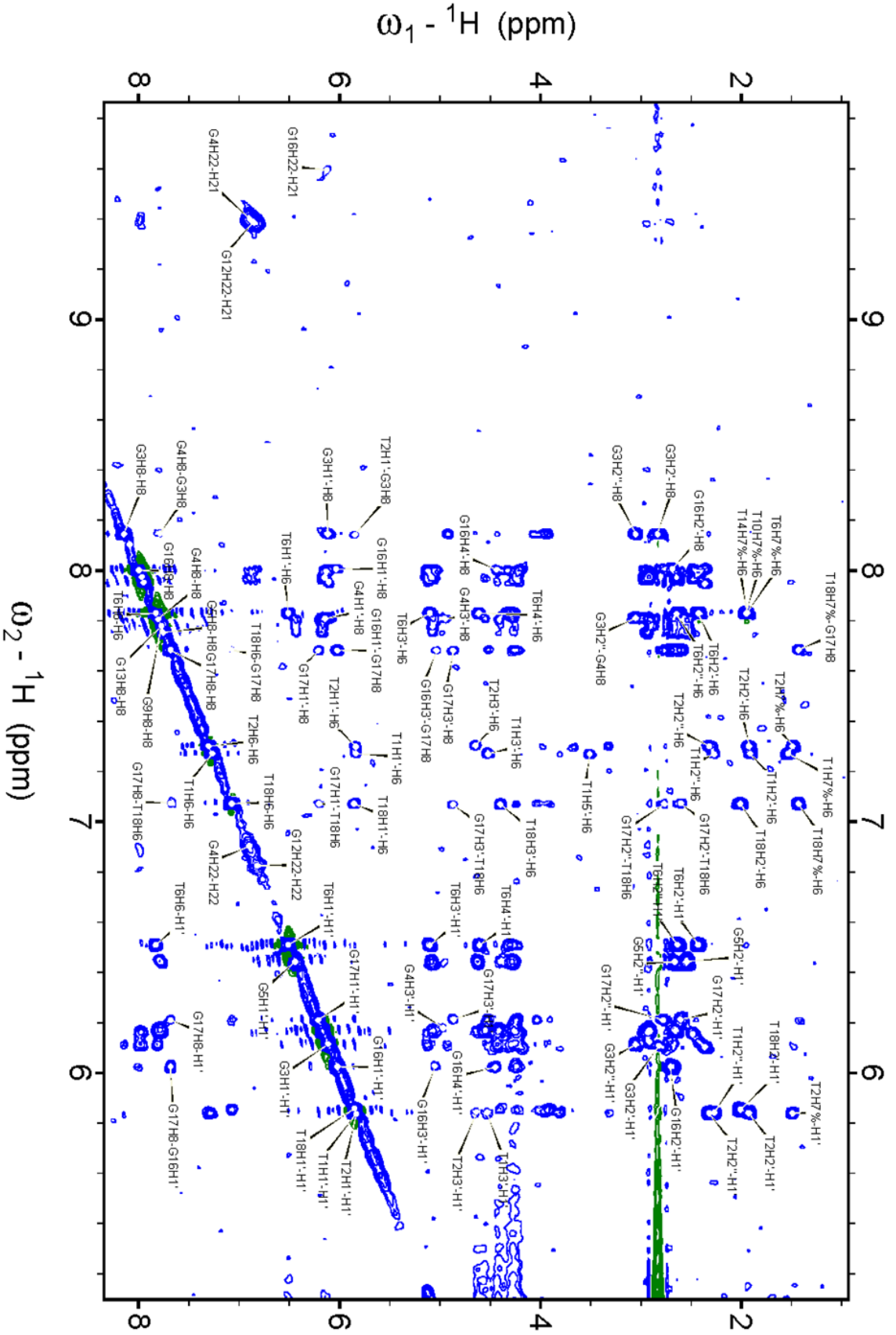


Figure S162 ^{1}H-^{1}H NOESY of 2LK7 on a Bruker Avance-III 800 MHz; middle region containing the guanine H8, thymine H6 and sugar H1' protons. Sample matrix: 100 µM 2LK7, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O.

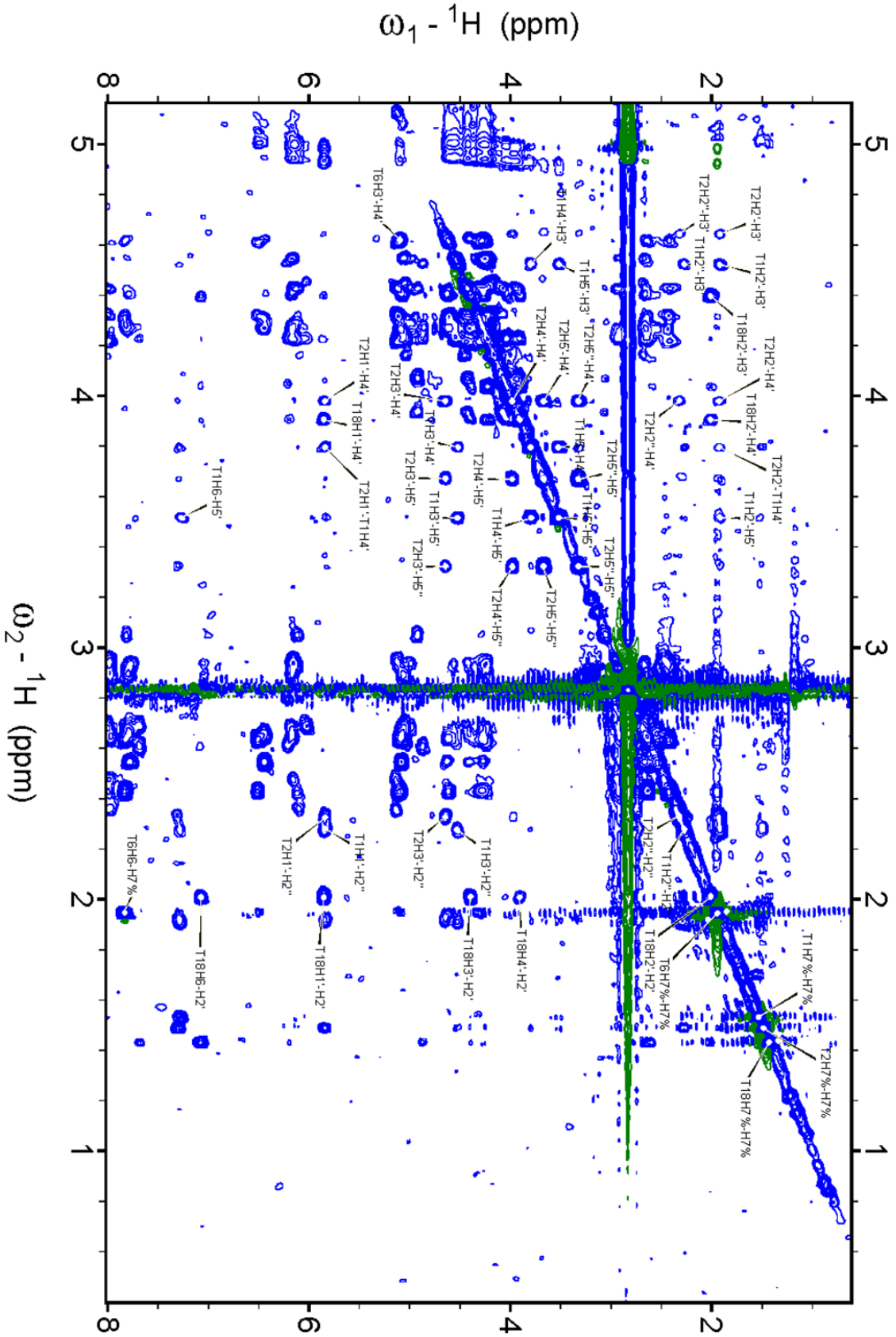


Figure S163 ^{1}H-^{1}H NOESY of 2LK7 on a Bruker Avance-III 800 MHz; low shift region containing sugar protons (H2' to H5') and the thymine methyl group (H7). Sample matrix: 100 µM 2LK7, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O.

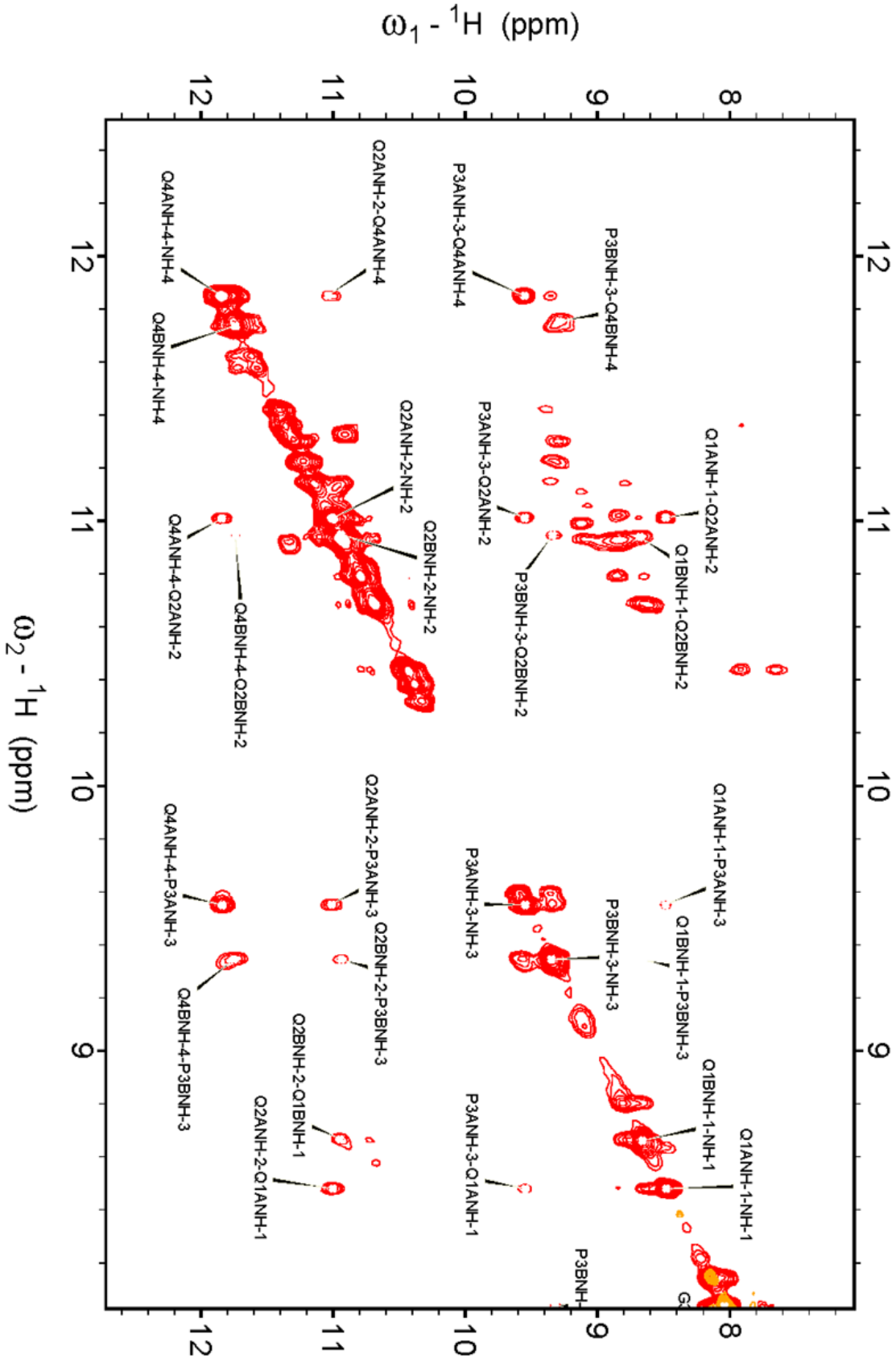


Figure S164 ^{1}H-^{1}H NOESY of 2LK7/QQPQ on a Bruker Avance-III 800 MHz; high shift region containing guanine H1 protons and foldamer NH protons. Sample matrix: 100 µM 2LK7, 300 µM QQPQ, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O.

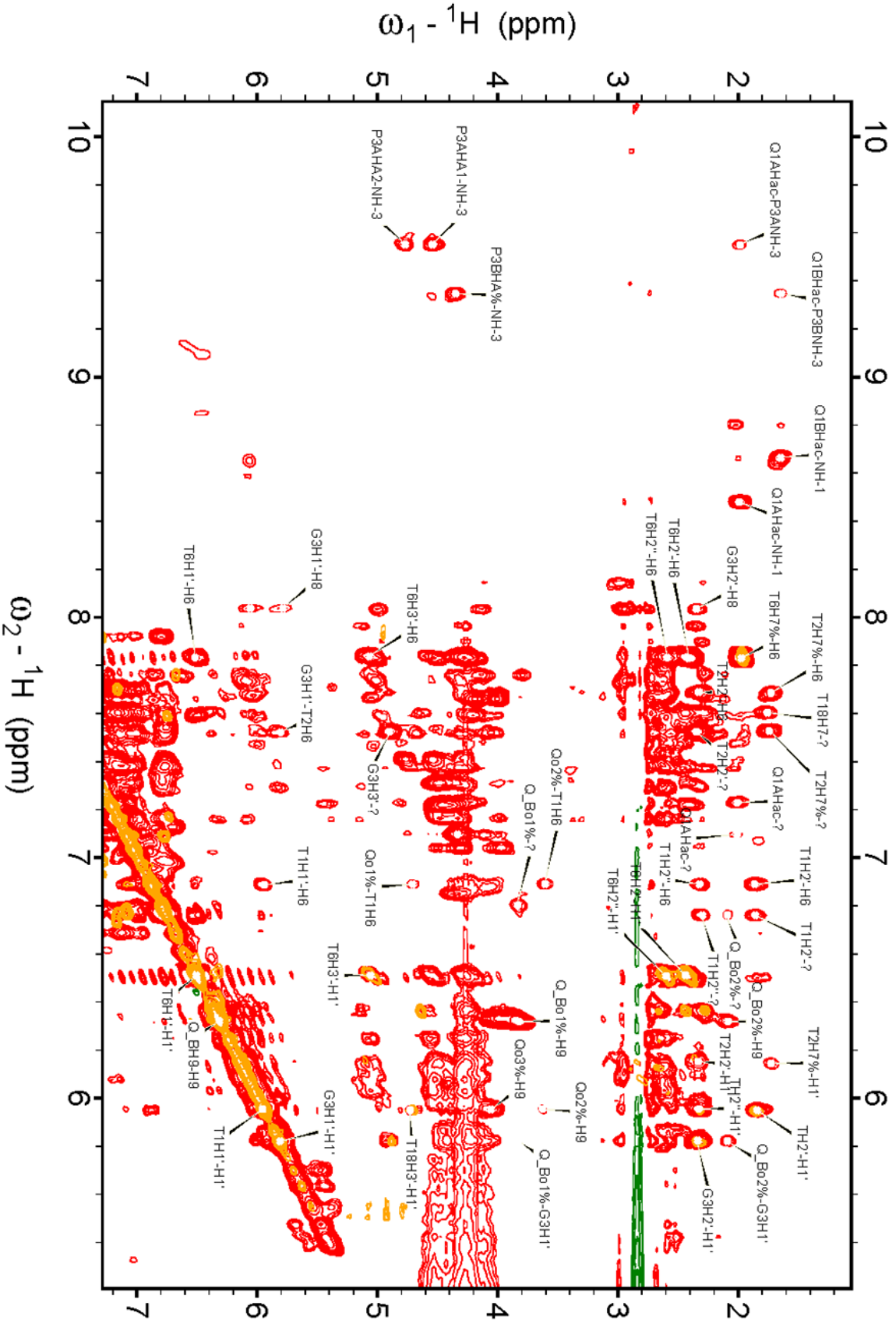


Figure S165 ^{1}H-^{1}H NOESY of 2LK7/QQPQ on a Bruker Avance-III 800 MHz; middle region containing aromatic protons of nucleobases and foldamer (as well as H1' sugar protons). Sample matrix: 100 µM 2LK7, 300 µM QQPQ, 10 mM potassium phosphate buffer (pH 7), 90/10 H_2O/D_2O. The orange overlay is a TOCSY spectrum of the same system.

Comparing Foldamers with established G-Quadruplex ligands

In order to define foldamers as a new class of G-quadruplex ligands, we need to differentiate them from already established G-quadruplex ligands. Figure S166 displays the ligands we picked for experimental comparison. There are two foldamers: QQPQ as the most promising ligand and QPPQ as a form of negative control, as it binds G-quadruplexes with lower affinity and specificity compared to QQPQ. For the established ligands we chose 4 ligands from 3 families of G-quadruplex ligands: PhenDC3 and PDS as two variants of acylhydrazones, Cu-ttpy as an organometallic complex and NMM as a porphyrin.

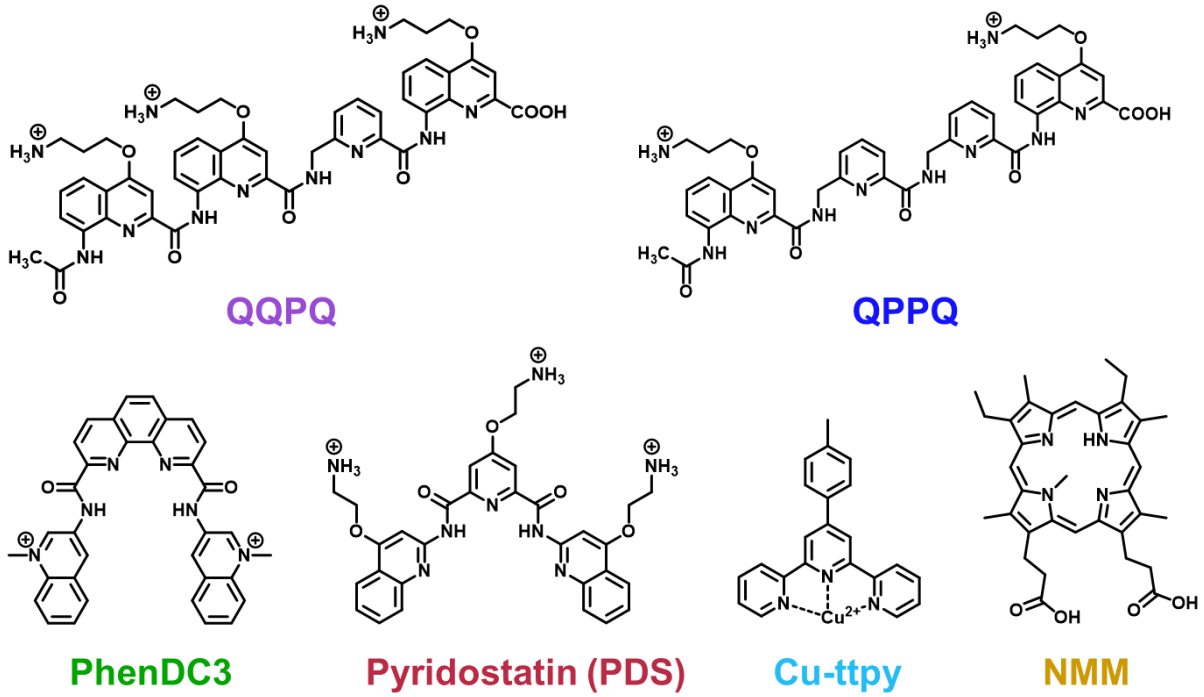


Figure S166 Foldamer ligands (top) and established G-quadruplex ligands (bottom) used for comparison.

We compared the mass and CD spectra of the six ligands with three different DNA sequences that represent the three main topology classes: parallel (Figure S167), antiparallel (Figure S168) and hybrid (Figure S169). In the case of 222T, we reduced the K^+ concentration to 0.15 mM in order to destabilize the parallel G-quadruplex and therefore better outline the G-quadruplex stabilizing effect of the ligands.

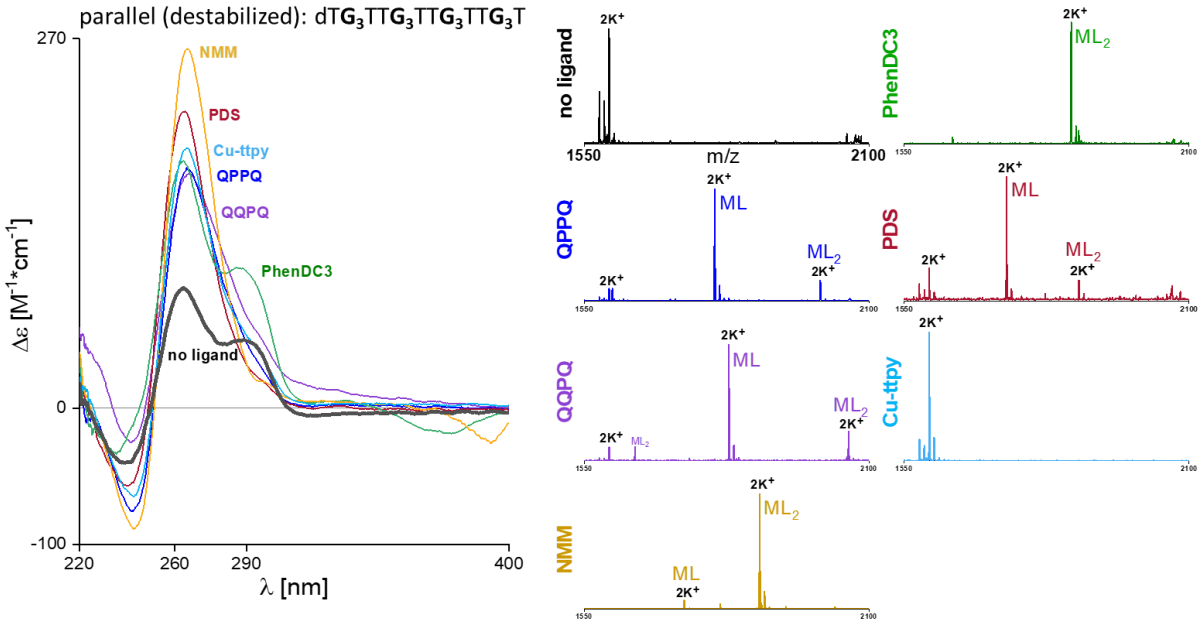


*Figure S167 CD spectra (left) and mass spectra (right) of 222T (dT**G₃**TT**G₃**TT**G₃**TT**G₃**T) with different ligands. Samples contain 10 µM DNA, 20 µM ligand, 0.15 mM KCl and 100 mM TMAA (pH 6.8). Mass spectra show the 4-charge state.*

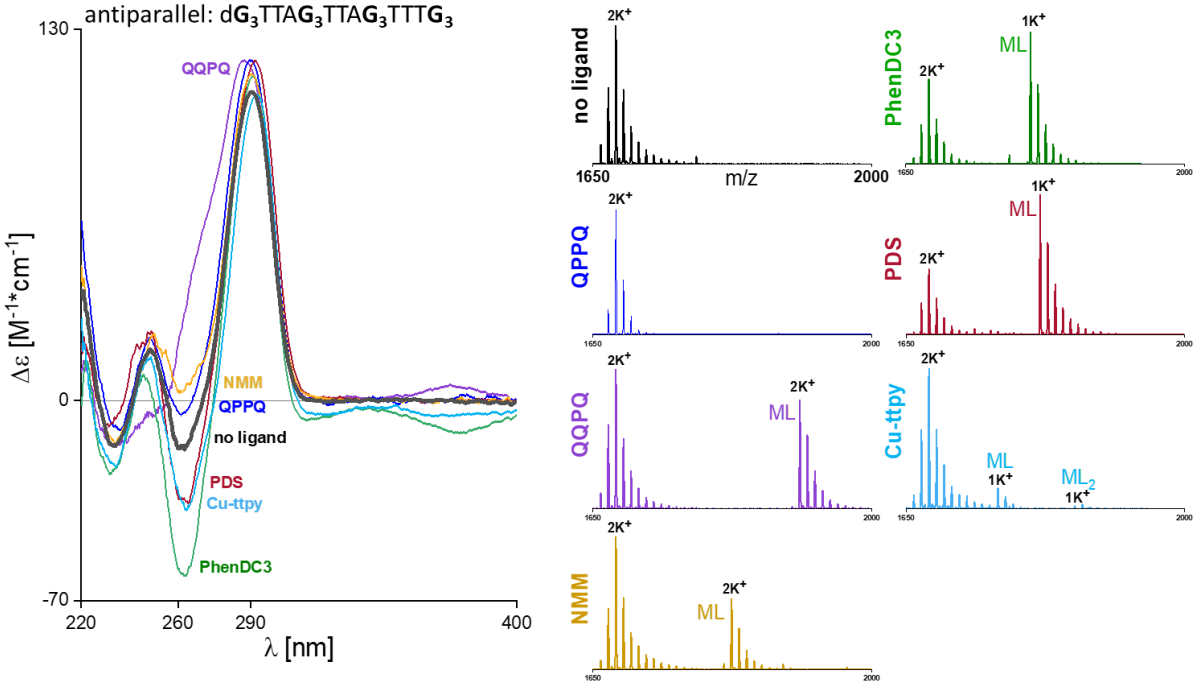


*Figure S168 CD spectra (left) and mass spectra (right) of 5YEY (d**G₃**TTA**G₃**TTA**G₃**TTT**G₃**) with different ligands. Samples contain 10 µM DNA, 10 µM ligand, 0.5 mM KCl and 100 mM TMAA (pH 6.8). Mass spectra show the 4-charge state.*

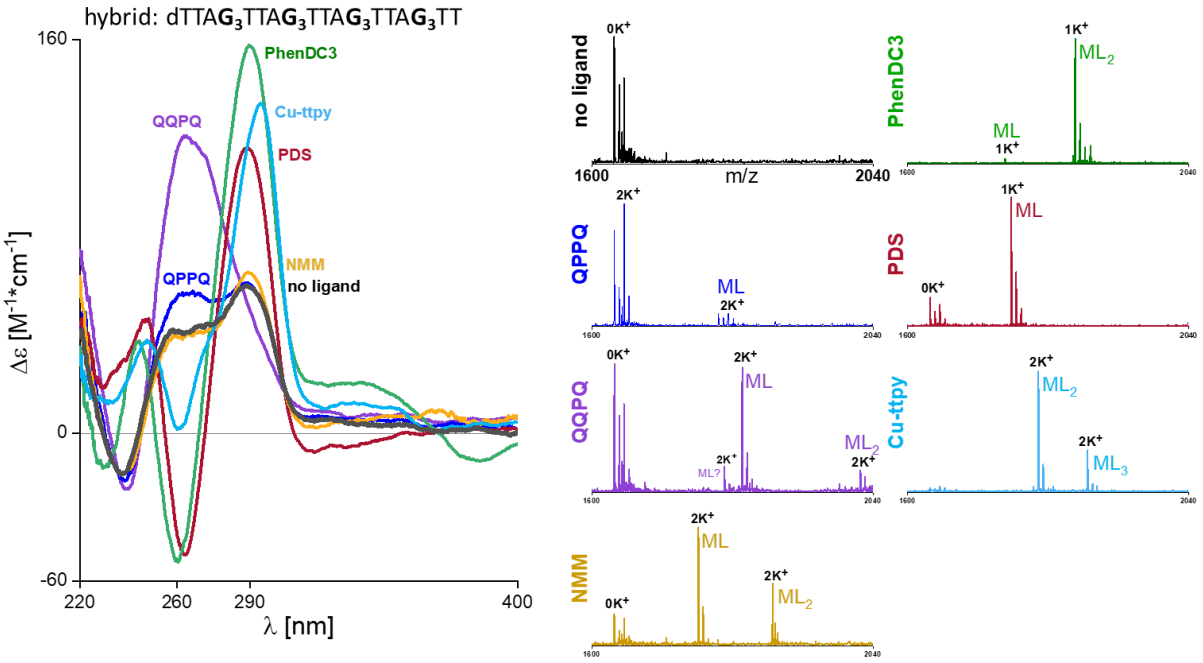


*Figure S169 CD spectra (left) and mass spectra (right) of 26TTA (dTTA**G₃**TTA**G₃**TTA**G₃**TTA**G₃**TT) with different ligands. Samples contain 10 µM DNA, 20 µM ligand, 0.5 mM KCl and 100 mM TMAA (pH 6.8). Mass spectra show the 5- charge state.*