Des outils moléculaires toujours plus multifonctionnels pour l'étude de structures alternatives d'acides nucléiques.

Francesco Rota Sperti

DOCTORAL THESIS OF THE UNIVESRITY OF BURGUNDY FRANCHE-COMTÉ

Doctoral School n°553

Doctoral School Carnot - Pasteur

PhD manuscript, Doctor in Chemistry

of

Francesco ROTA SPERTI

Ever smarter molecular tools for studying alternative nucleic acid structures in human cells.

PhD defense on 20th October 2022, Dijon.

Jury composition:

Daniela VERGA	CR CNRS, Institut Curie	Reviewer
Filippo DORIA	Associated Professor, University of Pavia	Reviewer
Franck DENAT	Professor, University of Burgundy	Examiner
Frédéric BOSCHETTI	CEO, CheMatech (Dijon)	Examiner
David MONCHAUD	DR CNRS, University of Burgundy	Director
Ibai VALVERDE	CR CNRS, University of Burgundy	Co-Director

Abstract

Title : Ever smarter molecular tools for studying alternative nucleic acid structures in human cells.

Keywords : alternative structures, nucleic acid, molecular tools.

Abstract: G-quadruplexes (G4s) are non-canonical secondary structures found throughout the human genome and transcriptome in guanine (G)-rich sequences. Accumulating evidence now points towards a strong involvement of G4s in many biological processes (*e.g.*, transcription and translation) but their roles are still not fully elucidated. To investigate this, we focused on the development of biomimetic G4-ligands that could be used as tools to either isolate G4s by affinity precipitation prior to their identification by sequencing or visualize them directly in cells by optical imaging. These tools referred to as TASQs (for Template-Assembled Synthetic G-Quartets) are composed of four G moieties linked to a central template thanks to spacers of different chemical nature. In the presence of G4s, the four Gs assemble in a synthetic G-quartet that, by stacking atop the native G4, stabilize its structure. This bioinspired interaction ensures an exquisite level of selectivity. The central template can be either functionalized (*i.e.*, by addition of a useful appendage) or selected on the basis of its properties (*e.g.*, a fluorescent template), thus allowing for a high level of versatility. In this Ph.D. work, we aimed at improving the chemical accessibility to these unique molecular tools and developing new TASQs. We first focused on the development of TASQ bearing a biotin appendage designed for G4-capture purposes. We improved the chemical accessibility of **BioTASQ**, used in the G4-RNA precipitation and sequencing (**G4RP-seq**) protocol, and developed two closely related TASQs, **BioCyTASQ** and **BioTriazoTASQ**. With excellent *in vitro* G4-interacting properties, **BioCyTASQ** soon became the gold standard for biotinylated TASQs, while the bulkier **BioTriazoTASQ** showed a certain level of preference towards G4-RNA isolation. To further improve the TASQ versatility we then replaced the biotin moiety by fully orthogonal appendages: **MultiTASQ** and **^{Az}MultiTASQ**, bearing an alkyne and azide appendages, respectively, were used to isolate and visualize G4s in cells, using click chemistry (CuAAC and SPAAC); **photoMultiTASQ,** functionalized with both an alkyne and a diazirine appendage, was successfully synthesized in the aim of being used to covalently trap G4s and its surrounding cellular partners (*i.e.*, G4-binding proteins) upon photoactivation. We then shifted our focus on fluorescent TASQs designed for cellular imaging. We first focus on *twice-as-smart* naphthalene-based probes, whose emission is governed by G4-interaction. We revisited the **N-TASQ** synthesis and developed a new derivative, referred to as **^{Tz}N-TASQ**, comprising triazole linkers, which displayed an interesting bias towards the fluorescent detection of RNA G4s. We finally focused on near infrared (NIR) emitting probes based on a squaraine template. Preliminary results show how the resulting **Square-TASQ** exhibited strong G4-binding along with desirable spectroscopic properties (with an emission centered at 640 nm), which makes it a promising NIR G4 probe.

By expanding the TASQ portfolio with new molecular tools to better isolate, visualize G4s, or both, we provide new tools that could shed a bright light on the role(s) that G4s play in cells.

Abstract

Titre : Des outils moléculaires toujours plus multifonctionnels pour l'étude de structures alternatives d'acides nucléiques.

Mots clés : structures alternatives, acides nucléiques, outils moléculaires.

Résumé : Les quadruplexes d'ADN et d'ARN, ou G-quadruplexes (ou G4s), sont des structures secondaires d'acides nucléiques qui se forment au sein de séquences riches en guanines (Gs) de notre génome et transcriptome. Les rôles biologiques des G4s sont actuellement au cœur d'un engouement scientifique intense, mais ils ne sont pas encore compris avec exactitude. Pour gagner des informations précieuses quant à ces rôles, l'utilisation des petites molécules (ou ligands) est cruciale car elle permet d'étudier la biologie des G4s en milieu cellulaire. Notre recherche s'inscrit dans cette dynamique et tend à développer une classe d'outils originaux appelés TASQ (pour *template-assembled synthetic G-quartets*). Les TASQs sont composés de quatre Gs assemblées autour d'un gabarit (*template*) central, qui peuvent donc former un G-quartet intramoléculaire. Ce faisant, les TASQs interagissent avec les G4s selon un processus bioinspiré, reposant sur la forte interaction entre deux G-quartets, l'un synthétique (TASQ) et l'autre natif (G4). Dans le cadre de ce travail de thèse, nous avons développé une série de nouveaux TASQs, répondant à des besoins chémobiologiques précis.

Nous avons tout d'abord amélioré la synthèse du **BioTASQ**, une des molécules phare de notre laboratoire, utilisée comme hameçon moléculaire pour l'isolation et l'identification des G4s *in vitro* et *in vivo*. Nous avons ensuite développé des dérivés plus facile d'accès et plus performants, les **BioCyTASQ** et **BioTriazoTASQ**. Grâce à ses propriétés de reconnaissance de G4s optimisées, le **BioCyTASQ** est devenu la molécule de choix pour nos études.

Le **BioTriazoTASQ**, assemblé par chimie click a, quant à lui, montré une affinité particulière pour les G4s d'ARN, qui méritera d'être exploitée plus avant. Nous avons ensuite développé des outils plus versatiles, disposant d'une fonction supplémentaire permettant des manipulations bioorthogonales en cellules. Les **MultiTASQ** et **^{az}MultiTASQ** sont en effet équipés de chaine alcyne et azide leur permettant d'être fonctionnalisés par chimie click (avec ou sans catalyseur, *i.e.*, CuAAC ou SPAAC) *in situ*. Cela peut être exploité pour la visualisation de G4s en cellules humaines (*in situ click imaging*) et leur isolation/identification (*click-seq*). Un composé encore plus versatile, le **photoMultiTASQ**, a été synthétisé dans le but de conduire de concert des études génomiques et protéomiques.

Nous avons ensuite retravaillé la synthèse d'une sonde fluorescente TASQ nommée **N-TASQ**, et développé un nouvel analogue, le **^{tz}N-TASQ**. Cette molécule est plus facilement accessible car son assemblage repose sur la chimie click. A nouveau, ce TASQ assemblé par chimie click a montré une affinité particulière pour les G4s d'ARN, qui méritera également d'être exploitée plus avant. Nous avons aussi conçu puis synthétisé une nouvelle sonde, le **Square-TASQ**, dont les propriétés photophysiques font d'elles une sonde NIR (*near infrared*) prometteuse. Nous avons donc, au cours de cette thèse, étendu le portefeuille d'outils TASQs qui permettent de voir, d'isoler et d'identifier les G4s dans un contexte cellulaire, dans le but final de gagner des informations précieuses quant à leur biologie.

Acknowledgments:

I would first like to thank my two supervisors David Monchaud and Ibai Valverde for guiding and helping me during these three years, trusting my work and making me a better scientist.

I am extremely grateful to each member of the jury, Daniela Verga and Filippo Doria for accepting to review this manuscript, to Frank Denat and Frédérick Boscetti to have taken the time to evaluate my work.

I also would like to thank the whole GATTACA teams past and present; Jérémie for the incredible help with the biophysical and cellular assays, Angie for the great work with the cells and Marc to endure my tortures and each one of my crazy requests.

I am extremely grateful to Baptiste and Jiri for always listening, challenging me and pushing me to improve.

I also would like to thank Claire Bernard for the help with the patent, Marie-Jo Penouilh and Quentin Bonnin for the high-resolution mass analysis, Antony Romieu, Victor Goncalves, Michael Claron and all the people that spent time helping me. I finally would like to thank all the other student for the support, especially Thibaud for sharing the office with me.

Vorrei infine ringraziare la mia famiglia, i miei genitori, i miei nonni, i miei fratelli per il supporto dimostratomi, Morgane per essere sempre stata al mio fianco, i miei amici per avermi sempre sostento.

Table of Contents

Annex 1 : Oligonucleotide list

Status	Nature	Name	Sequence
Double Labeled	DNA	F21T	FAM-d[$^{5'}$GGGTTAGGGTTAGGGTTAGGG$^{3'}$]-TAMRA
		F-Myc-T	FAM-d[$^{5'}$GAGGGTGGGGAGGGTGGGGAAG$^{3'}$]-TAMRA
		F-kit-T	FAM-d[$^{5'}$CGGGCGGGCGCGAGGGAGGGG$^{3'}$]-TAMRA
		F-duplex-T	FAM-d[$^{5'}$TATAGCTATATTTTTTTATAGCTATA$^{3'}$]-TAMRA
	RNA	F-TERRA-T	FAM-r[$^{5'}$GGGUUAGGGUUAGGGUUAGGG$^{3'}$]-TAMRA
		F-NRAS-T	FAM-r[$^{5'}$GGGAGGGGCGGGUCUGGG$^{3'}$]-TAMRA
		F-VEGF-T	FAM-r[$^{5'}$GGAGGAGGGGAGGAGGA$^{3'}$]-TAMRA
Single Labeled	DNA	F-Myc	FAM-d[$^{5'}$GAGGGTGGGGAGGGTGGGGAAG$^{3'}$]
		Cy5-Myc	Cy5-d[$^{5'}$GAGGGTGGGGAGGGTGGGGAAG$^{3'}$]
		F-22AG	FAM-d[$^{5'}$GGGTTAGGGTTAGGGTTAGGG$^{3'}$]
		F-SRC	FAM-d[$^{5'}$GGGAGGGAGGGCTGGGGG$^{3'}$]
		F-duplex	FAM-d[$^{5'}$TATAGCTATATTTTTTTATAGCTATA$^{3'}$]
	RNA	F-TERRA	FAM-r[$^{5'}$GGGUUAGGGUUAGGGUUAGGG$^{3'}$]
		F-VEGF	FAM-r[$^{5'}$GGAGGAGGGGAGGAGGA$^{3'}$]
		F-NRAS	FAM-r[$^{5'}$GGGAGGGGCGGGUCUGGG$^{3'}$]
Unlabeled	DNA	ds17	*strand 1*: d[$^{5'}$CCAGTTCGTAGTAACCC$^{3'}$]
			strand 2: d[$^{5'}$GGGTTACTACGAACTGG$^{3'}$]
		SRC	d[$^{5'}$GGGAGGGAGGGCTGGGGG$^{3'}$]
		22AG	d[$^{5'}$GGGTTAGGGTTAGGGTTAGGG$^{3'}$]
		97nt seq	d[$^{5'}$TTAGATACCATCAAACACCATTAGGTTCTACTGTAC-TGCCCTGCCCTGCCCTGCCCTGTCTGTCTCTCTTCCACT-GCCTGTTACGGCTGAATGGCTA$^{3'}$]
		Primer	d[^{5}TTAGATACCATCAAACACCATTAGG$^{3'}$]
	RNA	VEGF	r[$^{5'}$GGAGGAGGGGAGGAGGA$^{3'}$]
		TERRA	r[$^{5'}$GGGUUAGGGUUAGGGUUAGGG$^{3'}$]

Annex 2: List of abbreviations

AMC Aminomethylcyclen
Az Azide
BCN bicyclo[6.1.0]non-4-yne
Boc tert-butyloxycarbonyl
BG4-IP-Seq BG4-immunoprecipitation sequencing
CacoK potassium-rich cacodylate buffer
Cbz benzyloxycarbonyl
Chip-Seq chromatin immunoprecipitation and sequencing
Cy Cyclen
CT-DNA calf thymus DNA
CuAAC copper(I)-catalyzed alkyne–azide cycloaddition
DAPI 4',6-diamidino-2-phenylindole
DBCO Dibenzocyclooctyne
DBU 1,8-diazabicyclo[5.4.0]undec-7-ene
DCM dichloromethane
DNA Deoxyribonucleic Acid
DIPEA N,N-diisopropylethylamine
DMAP 4-dimethylaminopyridine
DMF N,N-dimethylformamide
DMSO dimethylsulfoxide
DPPA diphenylphosphoryl azide
dsDNA double stranded DNA
EDC 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
Et ethyl
FAM 6-carboxyfluorescein
FRET Förster (or fluorescence) resonance energy transfer
G guanine
G4 guanine quadruplex
G4DP-seq G4-DNA-specific precipitation sequencing
G4RP-seq G4-RNA-specific precipitation sequencing
HB hydrogen bond
HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt hydroxybenzotriazole
HPLC high-performance liquid chromatography
Me methyl
MD molecular dynamics
MS mass spectrometry
NMR nuclear magnetic resonance
Ph phenyl
PNA peptide nucleic acid
qPCR quantitative polymerase chain reaction
QFS quadruplex forming sequences
REE red edge effect
RNA ribonucleic acid
RP reverse phase
RPM rotation per minute
SA streptavidin
SPAAC Strain-Promoted Azide-Alkyne Cycloaddition
TAMRA 5-carboxytetramethylrhodamine
TASQ template-assembled synthetic guanine quartet
TBTA tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
Tf trifluoromethanesulfonyl
TFA trifluoroacetic acid
THPTA tris(3-hydroxypropyltriazolylmethyl)amine
Ts p-toluenesulfonyl
TSTU N,N,N',N'-Tetramethyl-O-(N-succinimidyl)uronium tetrafluoroborate
Tz Triazole
γH2AX phosphorylated H2A histone family member X

G-Quadruplexes (G4s) and their biological roles: state of the art

G-quadruplexes (or G4s) are non-canonical secondary structures adopted by single-stranded nucleic acid sequences with high guanine (G) content[1-3]. The ability of guanylic acid (a guanine derivative) to self-assemble into highly ordered suprastructures (a "semi-liquid jelly") was firstly noticed in 1910[4-5], more than forty years prior to the discovery of the DNA double helix by Watson, Crick, Wilkins and Fraklin[6-8]. The exact nature of this supramolecular association was deciphered in 1962, when Gellert, Lipsett and Davies resolved the X-ray structure of a crystal formed from a highly concentrated solution of guanylic acid, which unambiguously demonstrated that Gs self-assembled into G-quartets, which stacked on top of each other[9]. The authors identified the 8 hydrogen bonds (HBs) responsible for the G-quartet assembly and stability, which are created from both the Watson-Crick face (both the nitrogen atoms N1 and N2, which interact with their cytosine partner) and the Hoogsteen[10-11] face (in which the N1 and N2 of one G interact with the oxygen O6 and the nitrogen N7 of a second G) of the Gs (*Figure 1*). The interaction of four Gs creates an 8-HB network that results in the formation of a stable G-quartet. This molecular structure is further stabilized by the presence of cations such as potassium (K^+) or sodium (Na^+) that counterbalance the electrostatic repulsion of the 4 inwardly pointing carbonyl groups[12-13]. Two or more G-quartets can be further stabilized by π-stacking interactions between quartet layers, with cations either sandwiched in between planes (*i.e.*, K^+) or along the inner channel of the G-quartet stem (*i.e.*, Na^+), creating the very heart of a G4-architecture (*Figure 1*). The stability of a G4 is directly correlated to the number of its constitutive G-quartets: the more, the better[14].

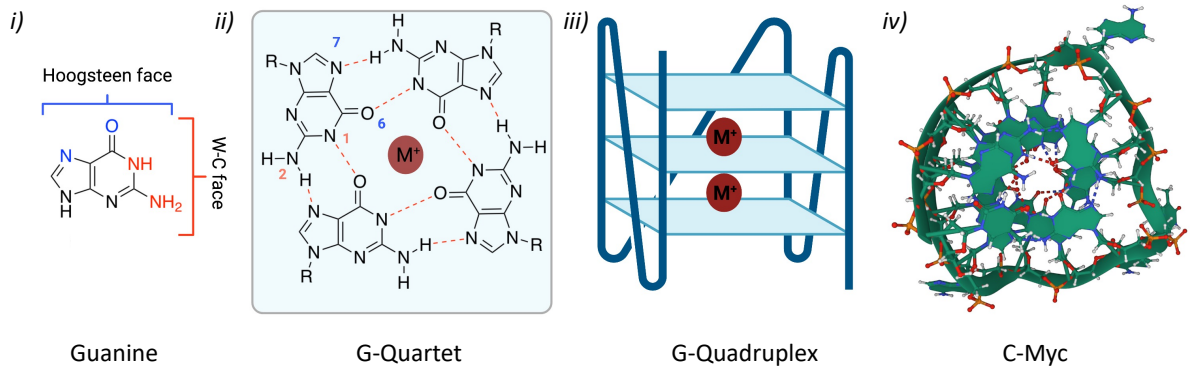


Figure 1: i) Guanine Hoogsteen and Watson-Crick faces. ii) Formation of the G-quartet highlighting HBs and the central monovalent cation. iii) tridimensional structure of a quadruplex (G4) composed of three stacked quartets iv) c-Myc G4 structure from Protein Data Bank (PDB ID: 2LBY).

G4 can be formed either intermolecularly or intramolecularly, with different strand orientations (parallel, antiparallel or hybrid, determined by the formalized orientation of the nucleobases, *i.e*., 5'-> 3', that is, from the 5' OH group to the 3' OH group), and nucleobase orientations[15] (*syn* or *anti*, determined by glycosidic bond angle, that is, the angle between the guanine plane and the ribose sugar) resulting in highly polymorphic structures[16-19]. This polymorphism is encoded by the sequences the G4s fold from but is also dependent on the environmental conditions they are studied in[20] (*i.e.*, cations, pH, temperature, additional stabilizing elements, etc.). This is especially true for G4-DNA, not for G4-RNA which tend to adopt a parallel conformation only[21].

The study of biologically relevant G4s started by the end of the 1980's, more than twenty years after their discovery, when their formation in naturally occurring G-rich sequences was postulated (at telomeres)[22-24]. G4s thus evolved from *in vitro* oddities to biological players which deserved to be further investigated to better understand their role(s) in living cells. The search for genomic G4s outside the telomeric regions started with *in silico* investigations[25-26]: the complete mapping of the human genome by *ad hoc in silico* tools (*e.g.,* QuadParser, G4Hunter) led to the identification on hundreds of thousands G4-forming sequences (QFSs), using algorithms such as $G_{3+}N(_{1-7})G_3+N(_{1-7})G_{3+}N(_{1-7})G_{3+}$ where Gs are guanines and N any other possible nucleobases (forming the so-called 'loops' that connect G-quartets)[27]. More recently, a series of machine learning-based tools were developed for both tracking G4-genome/transcriptome-wide and predicting their stability[28]. These investigations highlighted that the QFSs are not randomly distributed in the genome/transcriptome but particularly abundant in key regulatory regions such as telomeres and gene promoters for G4-DNA[29], and untranslated regions (5'- and 3'-UTR) for G4-RNA[30]. These calculations were not only confirmed but also extended with sequencing-based techniques such as **G4-seq**[31] and **G4 ChIP-seq**[32] (for G4-DNA) and **rG4-seq**[33] and **G4RP-seq**[34] (for G4 RNA); this will be further discussed in the following chapters.

I. Putative roles of G4s:

1. G4-DNA at telomeres:

The highest concentration of QFSs occurs at the telomeres, the terminal and capping part of the chromosomes[35], which consist of 5 to 10 kilobase-pair (kbp) of the $^{5'}$TTAGGG$^{3'}$/$^{3'}$AATCCC$^{5'}$ repeats. They are extended by a single-stranded G-rich protrusion (*ca*. 200-nucleotide (nt) long) comprising $^{5'}$TTAGGG$^{3'}$ repeats only (known as the "G-overhang"), which is thus prone to fold into G4-structures. The very first evidence of the formation of G4s in cells was obtained using an antibody (Sty49) raised against the G-rich sequence found in the telomeres of *S. Lemnae*[36]. Since then, evidence of G4-formation in mammalian cells, at telomeres but also elsewhere, was obtain *via* a series of orthogonal techniques, ranging from sequencing-based techniques (*e.g.*, G4 **ChIP-seq**)[32] to the immunodetection of G4 with the specific antibodies[37], and the direct detection of G4s with fluorescent small molecules in either fixed or living cells[38-39] (this will be also further discussed in the following chapters).

Telomeric G4s were the first structures to be studied in detail due to their link with cancer development (*vide infra*)[40]. Telomeres are thus long double-stranded sequences that end with a G-overhang. The higher-order structure of telomeres is currently actively scrutinized: it is for example hypothesized that the telomeres fold upon themselves in order to allow for the overhang to invade upstream double-stranded DNA, a suprastructure referred to as the T-loop (for telomeric loop)[41]. Telomeres do not encode genetic information but protect chromosome ends, and thus participate to the genomic stability. Telomeres are also considered as the "mitotic clock" of the cell, being shortened after every replication[42] (due to the so-called "end replication problem"[43], or "marginotomy"[44-45]), allowing for a precise countdown before cell death (the so-called "Hayflick limit", discovered by Leonard Hayflick in the 1960's)[46]. This process provides cell with a limited lifetime (normal conditions); it can be circumvented, however, by an enzyme, the telomerase[47], which fight against the telomeric attrition and thus, effectively immortalize cells[48]. This enzyme is not active in normal somatic cells (residually active in stem cells) but highly upregulated in cancer cells (>80%), making it an invaluable cancer marker and an interesting anticancer target. This was attempted by direct targeting[49] (*e.g.,* GRN163 and GRN163L), with a modest success; an alternative approach is to target telomeric G4s, in the aim of sequestering the telomerase substrate

under a shape that is not recognized by the enzyme. This can be achieved using G4-ligands, which will be further discussed in the following chapters.

The biology of telomeric G4s is interesting: the folded G4 conformation protects the overhang by disrupting the T-loop[41] and modifies the its structure, which makes it unrecognizable by the telomerase enzyme[50]; therefore; promoting and/or stabilizing G4s with G4-ligands could modulate the telomerase processivity both *in vitro* and *in vivo*[51]. G4s can also participate to the formation of a higher-order structure known as the G-loop[52], which was shown to be central to alternative lengthening of telomeres (ALT)[53-54], another cancerogenic mechanism, non-dependent on telomerase upregulation. The assembly and disassembly of the telomeric structure is controlled by several nucleoproteins (shelterin complex) including POT1 and TPP1 that act as regulatory agents for telomere homeostasis[55]. Telomeric sequences can also be transcribed to generate a long RNA sequence known as telomeric repeat-containing RNA (TERRA), which was also shown to fold into G4-structures[56-57]. Modulation of the interaction between TERRA and the basic N-terminal Gly/Arg-rich (GAR) domain of telomeric repeat-binding factor 2 (TRF2), by the addition of G4-ligand for instance, disrupts the RNA/protein complex, which results in telomere shortening, aberration and instability, thus inferring roles for TERRA G4s in telomere biology[58].

These different involvements explain why telomeric G4s were soon studied as targets for anticancer treatment, with the hope of decreasing telomerase activity *via* the sequestration of its substrate under a secondary structure the enzyme cannot cope with[51]. It was indeed shown that G4 stabilization by *ad hoc* ligands induces selective cell senescence of tumor cells, thus deeming G4s as targets (and G4-ligands as tools) for anticancer therapy[59]. However, these studies were performed with first generation G4-ligands, not highly specific for telomeric G4s, thus raising questions about the actual origins of the cellular activity observed. Also, no G4-ligands have yet reached clinical use, which keeps on casting doubts as the actual relevance of these ligands as new therapeutic tools.

2. **G4-DNA in intragenomic regions:**

Beyond telomeres[60], G4s found in intragenomic regions were also suspected to control DNA replication initiation[61], replication fork progression[62], transcription[63], translation[64] and genome instability[65] (*Figure 2*). To this end, QFS must escape from their duplex constraint to fold, making their formation linked to DNA transactions (replication, transcription and repair)[66]. Indeed, the polymerases in charge of these transactions must read the DNA code and thus, trigger local melting of the duplex, which offers a window of opportunity for G4s to fold. For instance, it was shown that QFS at replication origins might control replication firing[61]: two independent studies showed that mutations in these G-rich regions that affect the G4 stability result in a strong decrease in the origin activity[61, 67]. It was also demonstrated that exogenously added G4s could compete for replication factors, showing that G4s are likely to be docking sites for replication factors[61]. *Via* genome-wide techniques (*e.g.*, **G4-ChIP-Seq**)[32] the prevalence of G4-forming sequences was confirmed, as well as the link between G4s and transcriptional activity, again indicating that G4s recruit the transcription machinery and/or chromatin remodeling factors, but they can also restrain/reinitiate transcription of highly active genes[68]; further studies are thus required to decipher the complex biology of G4s. The case study is certainly the G4-forming sequence found in the promoter region of the oncogene Myc[69-70]: the Myc genes encode for transcriptional factors that regulate the expression of 10-15% of the human genes and are found to be upregulated in a vast number (>80%) of solid tumors[71], making them interesting cancer markers. Mutations and/or treatment with G4-ligands that target G4s in the Myc promoter repress its transcription and downregulate its expression, indicating a direct correlation between G4 formation and the transcriptional repression[72]. Myc is only one of the many examples of genes whose expression is tightly controlled by upstream G4s, which again warrants further studies to decipher the very complex G4-biology.

Interesting insights into G4 biology were obtained through the study of a special class of enzymes that unwind G4s, known as G4 helicases[73]. Deficiency in G4 helicases often results in an elevated level of genetic instability (throughout different species, indicating how G4-mediated regulation mechanisms are conserved through evolution)[74]. These results thus imply that G4s might act as a threat for genome stability and that deficiency in enzymes in charge of G4-resolution triggers genomic instability and aging[75]; this was demonstrated by

knock-down studies of G4-helicases such as FANCJ, but also by the reversal of the genetic instability triggered by a G4-ligand (here, PDS, *vide infra*) by the over-expression of another helicase Pif1[76].

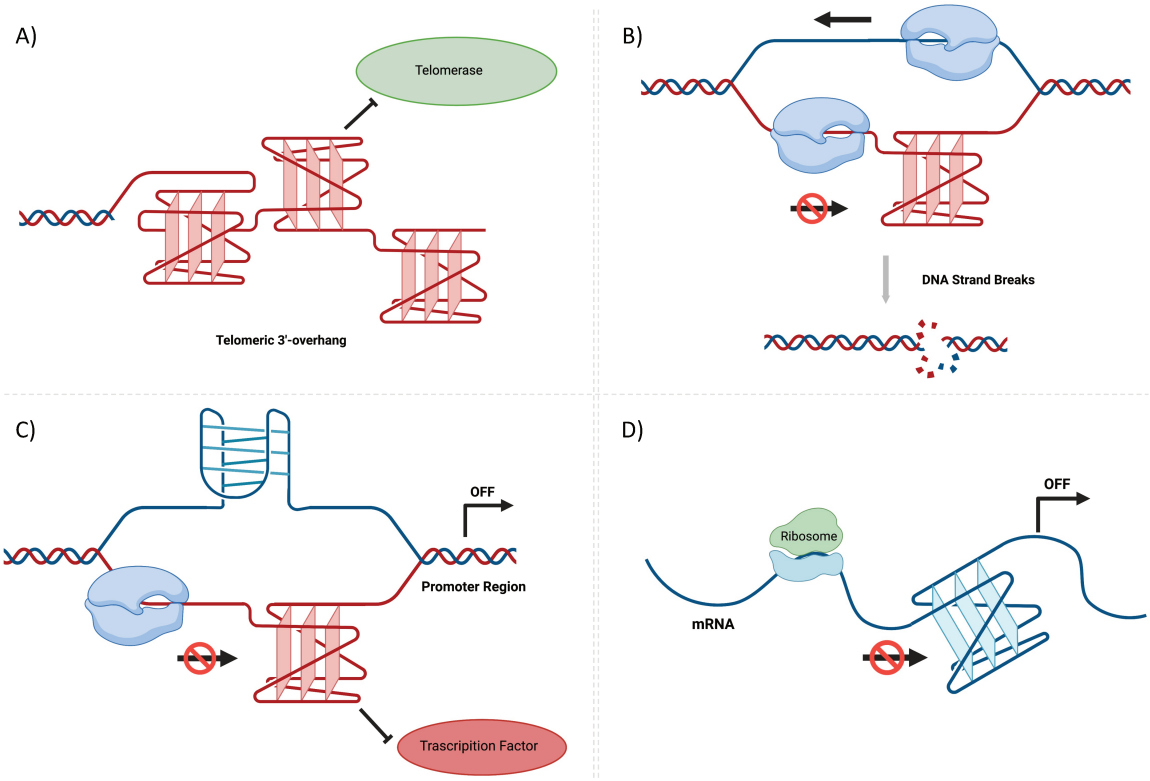


Figure 2: A) Telomeric G4s formed in the G-overhang found at the end of the chromosomes that interfere with the telomerase processivity. B) G4 can stall the polymerase processivity, which is recognized as a DNA damage (double strand break, DSB), in the presence of a G4-ligand or in absence of specific G4 helicases. C) G4s inhibit transcription factor binding, which alters gene expression at transcriptional level. D) G4 folding in messenger RNA (mRNA) inhibits ribosome scanning and thus, gene expression at transational level.

3. G4s in RNA:

Despite their apparent similarities, G4-DNA and RNA are, in fact, rather different[30]. While DNA mostly exists in the double-stranded form, RNA is single-stranded by nature, allowing for the formation of a broad variety of secondary structures including hairpins, bulge, looped structures and, of course, G4s (G4-RNA is also referred to as rG4). Moreover, while DNA is found in the nucleus, RNA can be found in both the nucleus and the cytoplasm. Even at molecular level, some differences between G4 and rG4 can be noticed, the latter being less polymorphic and more thermodynamically stable. These differences actually result in drastically different biological roles[1].

RNA rG4-forming sequences are commonly found in regulatory regions of functional RNAs including messenger RNA (mRNA), microRNA (miRNA) and non-coding RNA (ncRNA)[77-78]. The most studied class of rG4-containing RNAs is mRNA: rG4 can be found in as 3'-UTR[79] (where they can control mRNA lifetime), 5'-UTR (where they control mRNA translation)[80] and in introns of pre-mRNAs[81]. When found in 5'-UTRs[82], the actual roles of rG4s depend on the translation modality: they can act as physical obstacles to ribosome entry in the case of cap-dependent translation, which thus inhibits gene expression, and they can facilitate the docking of ribosomes in the case of cap-independent translation, which thus favors gene expression[83]. This generates a real conundrum about the actual roles of rG4s in RNA translation and the relevance of targeting rG4s in the aim of harnessing gene expression at translational level. When found in 3' UTRs, rG4s are shown to inhibit translation through a complex (and not completely understood) mechanism involving the binding of a series of 3'-processing protein. Finally, rG4s present in pre-mRNA introns and/or exons enhance, inhibit or promote alternative splicing *via* the recruitment of specific G4-interacting proteins or steric interference with regulatory elements in the vicinity of the rG4 sites[84]. To date, the knowledge about rG4s still needs to be deepened and interesting yet uncomplete insights have been gained through transcription-wide techniques such as **rG4-seq**[33] and **G4RP-seq**[34].

II. G4 and G4-ligands.

1. G4-ligands as therapeutic agents

Beyond their poorly understood and somewhat contradictory roles in gene expression control[85], G4s are also associated with DNA damage. indeed, they can act as roadblocks to the motion of enzymes in charge of DNA transactions, particularly when they are stabilized by G4-ligands[86]. This crisis situation is then recognized as a DNA damage as it generates double strand DNA breaks (DSBs) which eventually leads to cell apoptosis[66]. This was exploited by combining G4-ligands (to create G4-mediated DNA damage) with inhibitors of DNA repair (to preclude damage repair) in an approach referred to as chemically induced synthetic lethality[87]. Due to their intrinsically fast replication pace, tumor cells often display an impaired DNA repair machinery[88] which makes them more sensitive than normal, healthy cells to G4-mediated DNA damage, especially upon synthetic lethality treatment. This approach thus suggests that using G4-ligands might have interesting therapeutic outcomes[59].

2. G4-ligands as chemical biology tools

G4-ligands are also invaluable tools as they are quite often cell-permeable, which allows them to interrogate G4-biology in a functional cellular context. In this approach, ligands are not used as therapeutic agents but as chemical biology tools[89]. This is particularly important with regard to the current interrogations concerning the cellular roles of G4/rG4 and their formation *in vivo* throughout our genome and transcriptome. To this end, the development of G4-selective ligands combined with the recent advances in sequencing techniques offer a unique opportunity to accurately portray the roles of G4/rG4 during the key steps of cell life (transcription, translation and replication). In this context, ligands can allow for G4 isolation and identification (as it will be thoroughly discussed in the following chapters) and also for localizing G4s in cells when used as imaging agents[90]. Ideally, a single G4-targeting molecular tool could be used for performing imaging, purifying G4 by affinity capture for identification purposes and identifying their cellular partners by proteomics; the development of such a highly valuable molecular tools is precisely the goal of the present work. Our aim is indeed to develop and characterize multivalent G4-ligands to accurately portray the "G4ome" and make them available to all in order to boost research on G4-biology.

3. The structures of classical G4-ligands:

The interactions between G4 and ligands rely on the ability of the small molecules to interact with G4 and/or promote their formation. Till the very first prototype developed in 1997 by Neidle, Hurley and co-workers[51] (*vide infra*), hundreds of candidates have been tested as specific ligands; they generally possess several common features, including a planar aromatic core surrounded by positively charged side-chains (*Figure 3*)[86]. The planar aromatic core is important to create π-stacking interactions with the accessible external G-quartet of a G4 (and its important size generally precludes unwarranted intercalation in between two base pairs of a duplex DNA). The positively charged sidechains aim at creating electrostatic interactions with the negatively charged phosphate backbone of the DNA and RNA. The first ligand ever reported was an acridone developed by Neidle and Hurley in 1997[51]. Several years later, closely related acridine derivatives were crystallized with G4s, thus providing invaluable

structural insights into how a G4 can accommodate a ligand[91-92]. These results gave strong impetus to the identification of new G4-ligands with exquisite G4-affinity and selectivity.

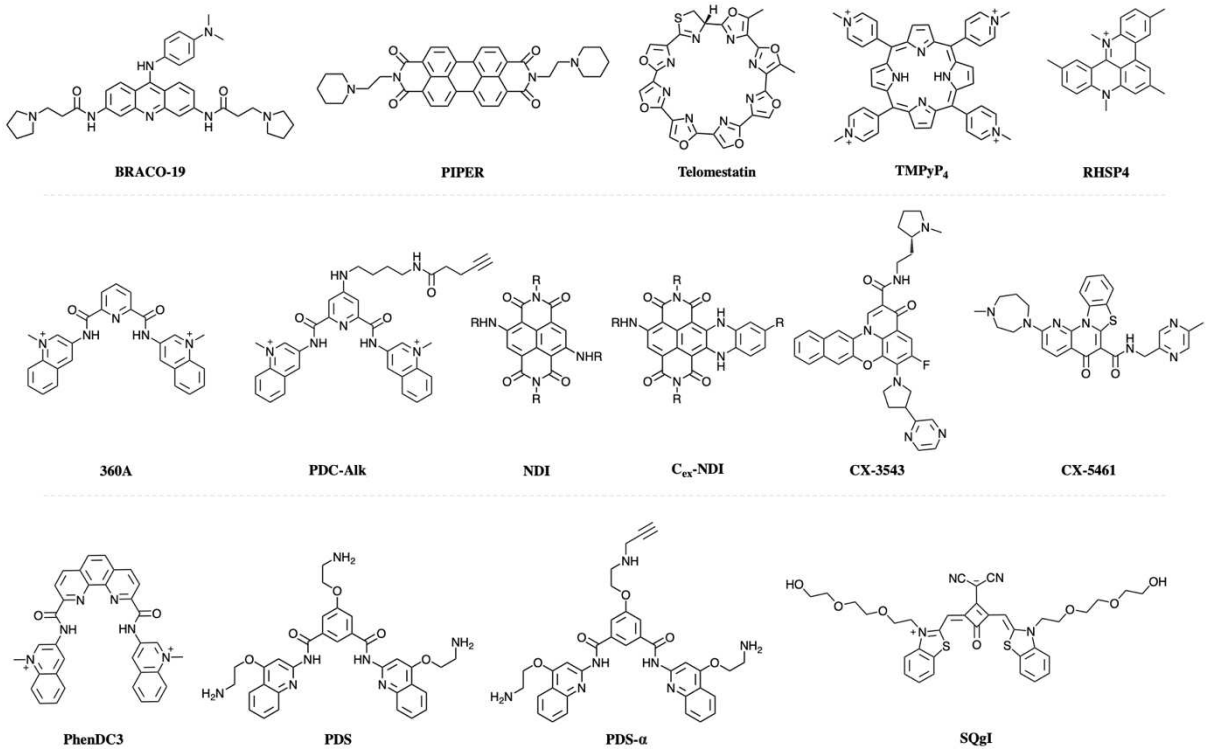


Figure 3: Chemical structure of some of the most popular G4-ligands.

The most popular ligands are: **RHPS4**, developed by Searle and co-workers, who also provided the first NMR structure of a G4/ligand complex[93]; the tetrasubstituted porphyrin **TMPyP4**, developed by Hurley and co-workers, which binds with a high affinity to G4s but with a very low selectivity[94]; the 3,6,9-trisubstituted acridine ligand **BRACO-19**, developed by Neidle and co-workers, which was the first ligand to be rationally designed[95] and found to be an efficient anticancer agent against tumor xenografts[96]; the perylene diimide ligand **PIPER** also developed by Hurley and co-workers[97]; some naphthalene diimide (**NDI**) derivatives, also in their core-extended form developed by Neidle[98] and Freccero and co-workers[99], respectively; a natural compound named telomestatin, isolated by Shin ya's group, which is one of the few examples of neutral G4-ligands with exquisite G4-interacting properties[100]; some bisquinolinium derivates developed by Riou[101] (**360A**), Teulade-Fichou[102] (**PhenDC3**) and Balasubramanian and co-workers[103] (pyridostatin, **PDS**); and two fluoroquinolone derivates (**CX-3543**[104] and **CX-5461**[105]) that reached (and unfortunately stopped at) phases II and I respectively[106]. **PhenDC3** and **PDS** are, by far, the most used ligands nowadays, as their

excellent G4-interacting properties were confirmed *via* multiple techniques and their cell permeability allows them for being conveniently used in cells[107]. Biorthogonal versions of **PDS** (PDS-α) and **360A** were also developed to exploit its excellent selectivity by click chemistry [108-109] (see *Chapter 3*). A wide range of fluorescent G4-probes was also developed to visualize G4s in cells, including several functionalized **NDIs**[99], which display fluorescence in the visible spectrum and close to the near infrared (NIR) region and squaraine-based ligands (**SQgl**) with remarkable selectivity and turn-on properties (see *Chapters 5 and 6*)[110].

4. Biomimetic Ligands

The thorough studies performed with these ligands have cast a bright light on the major challenges that must be addressed to design an 'ideal' G4-ligand. It has to interact with G4s (high affinity) but not too much (to avoid promoting their formation, which could introduce a bias during cellular investigations), and must be exquisitely selective. The rational design described above aims at transforming DNA intercalators (*e.g.,* acridine) into G4-ligands (*e.g.,* **BRACO-19**), but we believe that this approach suffers from a major drawback: it is based on duplex-interacting compounds, and the duplex interaction is precisely what must be avoided. We thus tried to think different: we decided to develop a new class of biomimetic G4-ligands that could mimic the strong interaction between G-quartets. It is indeed known for more than two decades now that the stability of a G4 depends mostly on the number of its constitutive G-quartets (a two-quartet G4 being less stable than a three-quartet G4, etc.). Therefore, using a synthetic G-quartet as a ligand would lead to a stabilized G4, without the caveat of trying to conceal the initial duplex interaction.

The design of this type of new, water-soluble and biomimetic ligands followed a step-by-step progression (*vide infra*). This design was based on the use of a templated synthetic G-quartet (that is, gathering four Gs within the same chemical scaffold) to make the intramolecular G-quartet assembly kinetically and thermodynamically favored. Once assembled, this synthetic G-quartet favors the π-stacking interactions with the accessible G-quartet of a G4 (a 'native' G-quartet), on the basis of both shape recognition (a *'like-likes-like'* approach), electrostatic interactions (using protonable G moieties) and chelation of biologically relevant cations (Na^+ or K^+) in between the synthetic and native G-quartets[111].

The first report on a synthetic G-quartet (SQ) was provided by Sherman and co-workers in 2008[112]. They linked Gs to a calixarene template and termed these new compounds TASQs for template-assembled synthetic G-quartets. These TASQs were intended to be used in organic solvents only (as a supramolecular 'box' for solubilizing cations initially not soluble in organic media), precluding their use in biological media. A water-soluble version of this TASQ was developed in 2013, replacing the hydrophobic calixarene with a water soluble one (decorated with phosphate groups)[113]. Two water-soluble TASQ were reported in 2011: one (from our group) based on a DOTA-like polyazamacrocycle template functionalized with four Gs[114], the second one based on a macrocyclopeptide template (RAFT)[115]. Only the DOTA-based TASQ, consequently termed **DOTASQ**, was used as a biomimetic G4-ligand; the RAFT-based TASQ was used for NMR studies, and the water-soluble calixarene-based TASQ was used to evaluate the ability of G4-ligands, *e.g.,* **BRACO-19** and **$TMPyP_4$**, to interact with an isolated G-quartet.

The distinguishing feature of **DOTASQ** is its conformational plurality, as it can adopt an "open" conformation (where the guanines are independent from each other) or a "closed" conformation (where the intramolecular G-quartet is formed). **DOTASQ** by itself had a rather modest affinity for G4s. It had an acceptable G4-stabilizing ability only when terbium was chelated by the DOTA template[114]. However, **Tb·DOTASQ** showed an exquisite selectivity for G4s over double-stranded DNA (dsDNA). For this reason, its scaffold was gradually modified to strengthen its binding properties, firstly by using protonable Gs to promote electrostatic interactions with G4s (**^{PNA}DOTASQ**[116]), and then, by modifying the central scaffold. The introduction of positively charged Gs in the case of **^{PNA}DOTASQ** was achieved thanks to the use of a peptide nucleic acid guanine (^{PNA}G); it ensured a high level of G4-stabilization (without the need of a terbium) while maintaining the exquisite G4-selectivity of the parent compound **DOTASQ**. ^{PNA}Gs enable the π-π stacking between the G4 and the synthetic G-quartet and further strengthen the binding *via* electrostatic interaction between the positively charged amine (at pH 7.2) and the negatively charged phosphate backbone of nucleic acids. The G4-binding mode of **^{PNA}DOTASQ** was further demonstrated by NMR titration (*Figure 4*): the tetrameric $(T_2AG_3T)_4$ was chosen as its symmetry provides only three signals observed in the Hoogsteen region (from 10 to 12 ppm) corresponding to its three constitutive G-quartets. Upon addition of increasing concentrations of **^{PNA}DOTASQ**, the appearance of an additional

signal at 10.35 indicated the formation of an additional G-quartet layer (belonging to the TASQ; of course, **^{PNA}DOTASQ** alone does not present any signal in this region). These investigations were further validated by the control experiment performed with dsDNA where no changes were observed, demonstrated the '*like-likes-like*' binding modality of TASQs. This G4-promoted folding of **^{PNA}DOTASQ** into its closed conformation makes it a smart G4-ligand; this conformational switch only occurs in the presence of the G4-target, thus ensuring a fully unique *active* selectivity for G4s that, coupled with the increased G4-stabilization, is at the very heart of every TASQ design.

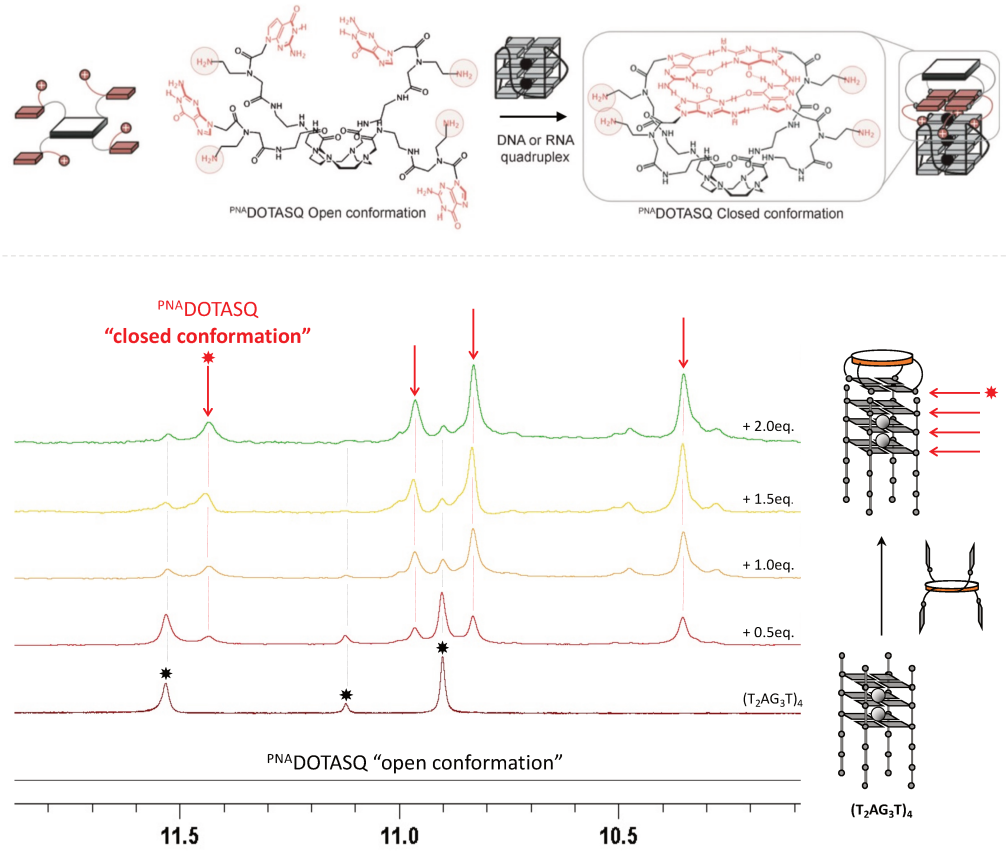


*Figure 4: i) Schematic representation of the 'open' and 'close' conformations of **^{PNA}DOTASQ** upon G4 addition. ii) NMR titration of $(T_2AG_3T)_4$ with increasing concentration of **^{PNA}DOTASQ**. In black **^{PNA}DOTASQ** alone; in wine, the G4 alone. The introduction of 2 mol. equiv. (green) of TASQ results in the formation of an additional peak at 11.3 ppm (indicated with a *), corresponding to the formation of an additional G-quartet, which belongs to the closed conformation of the TASQ.*

The scaffold of TASQs was modified to endow them with fluorescent properties (*Figure 5*): the pyrene-based **PyroTASQ**[117] exhibited G4-specific binding and showed also very interesting fluorescence turn-on properties, triggered by the presence of G4s only, thus allowing for the specific detection of G4s *in vitro*. **PyroTASQ** was the first example of twice-as-smart G4-

ligands, being both a smart G4-ligand and a smart fluorescent probe. However, its tendency to self-aggregate when used in cell-based assays precluded a wider use. A second twice-as-smart ligand based on a naphthalene core (**N-TASQ**)[118] was thus developed to be suited to be used in cells, owing to the decreased tendency of naphthalene to self-associate. **N-TASQ** was successfully used to track G4 in cancer cells[52, 118-119], providing the first direct visualization of G4 in live cells, but was also used in central nervous system cells[76]. More recently, a biotin 'bait' was added to the **^{PNA}DOTASQ** scaffold and the resulting **BioTASQ**[120] allowed for isolating G4s by affinity capture (*in vitro* and from cells). The properties of **BioTASQ** (and derivatives) will be further discussed hereafter.

Globally speaking, the G4-interacting properties (affinity, selectivity) of TASQs are very interesting but their chemical accessibility is troublesome (low yields, small quantities, non-commercial starting materials); the aim of this PhD thesis was thus to enhance the TASQ chemical access and to develop new TASQs aimed at addressing new challenges.

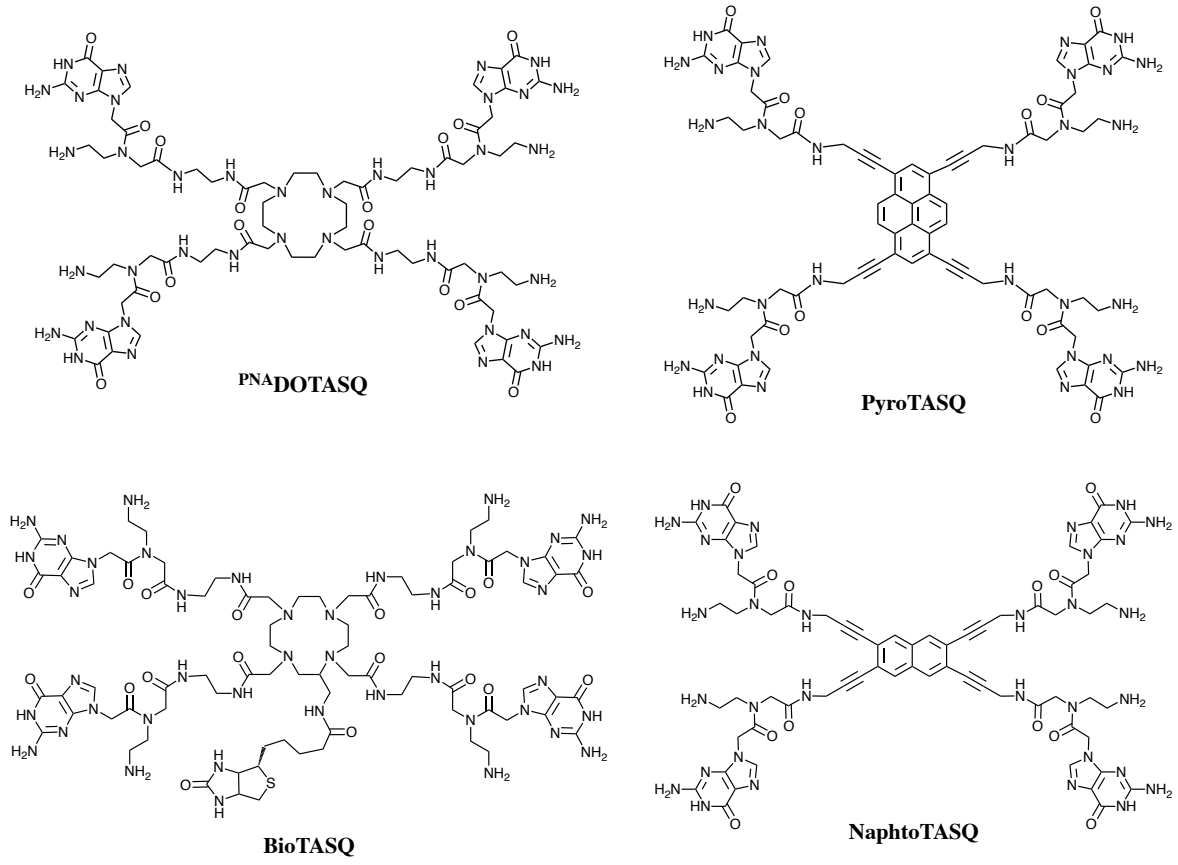


Figure 5: Chemical structure of some of the TASQs developed before the present Ph.D. work.

Therefore, in the following chapters, we will discuss the improvements made with regards to the chemical access of TASQ, the validation of their G4-interacting properties *in vitro* and *in vivo.* Although TASQ are not 'ideal' G4-ligands (high molecular weight, complicated synthesis, etc.), they elicit a series of desirable features:

i) A good affinity and an exquisite selectivity achieved thanks to their bioinspired design;

ii) A very low cytotoxicity;

iii) A high level of versatility that can be achieved by modifying the central template;

iv) A very low ability to promote G4 formation (thus avoiding bias during in vivo investigations).

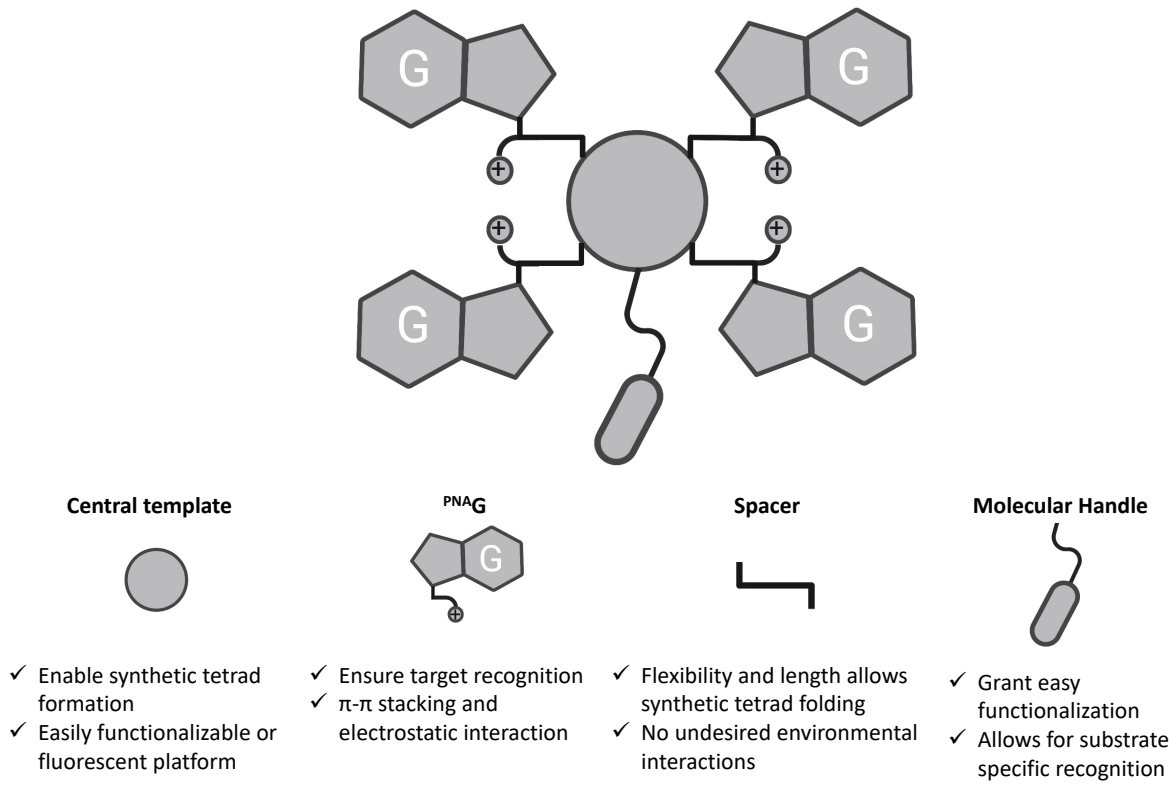


Figure 6: Schematic representation of a TASQ ligand and its structural specificities, i.e., a central platform, a molecular spacer, PNAG and an appendix to introduce new functionalities.

Aim of this PhD thesis:

Three structural components must be taken into account when designing TASQs: i) the central platform, which can be either a simple molecular template to allow for the intramolecular G-quartet folding, or a template equipped with specific appendages (*e.g.,* biotin, alkyne groups, etc.) or a fluorescent platform to perform direct visualization; ii) a spacer that links ^{PNA}G to the central template; and iii) the Gs themselves, which govern G4 recognition (*Figure 6*). The choice and optimization of these three parameters will be discussed in the following chapters:

- In *Chapter 1* we will discuss about the optimization of the synthesis of the central platform aminomethylcyclen (AMC);
- In *Chapter 2* we will discuss about two biotinylated TASQs, whose synthesis was either optimized (**BioTASQ**) or rethought to be more straightforward and reliable (**BioCyTASQ**). The G4-interacting properties (both *in vitro* and *in vivo*) will be presented and compared;
- In *Chapter 3* we will discuss about the new generation biotinylated TASQs (**BioTriazoTASQ**), obtained by modifying the linker between ^{PNA}G and the central platform. We will also detail the relationship between G4s, ligand-stabilized G4s and DNA damage;
- In *Chapter 4* we will discuss about the introduction of different bioorthogonal appendages (alkyne for **MultiTASQ**, azide for **azidoMultiTASQ**, diazirine for **photoMultiTASQ**) and their use for a new series of bioorthogonal investigations (*in situ* click chemistry).
- In *Chapter 5* we will discussed about the optimization of the synthesis of the twice-as-smart **N-TASQ**, as well as the development of a new G4 probe ^{Tz}**N-TASQ** with triazole-based linkers. Again, their G4-interacting properties (both *in vitro* and *in cella*) will be presented and compared;
- Finally in *Chapter 6* we will introduce a squaraine-based TASQs that represent the very first prototypes of near infrared (NIR) G4-probe.

Chapter 1:

Revisiting the synthesis of AMC

The choice of an adequate template is crucial for the development of a good biomimetic G4-ligand. **Aminomethylcyclen** (**AMC**, *Figure 7*)[121] was identified as an ideal candidate for this task owing to its unique features. Cyclen-based TASQs:

- are water soluble;
- are assembled around a small-sized cyclen-based macrocycle, which keeps in close proximity the arms bearing the G moieties thus helping the formation of the intramolecular synthetic G-quartet;
- benefit from an electronic organization that allows the nitrogen's electron pairs to point toward the same side (due to the cyclic strain and charge repulsion), which might conformationally pre-organize the TASQ to form an intramolecular G-quartet;
- can be functionalized on its four nitrogen atoms, thus allowing for the introduction of four G arms, and also *C*-functionalized, exhibiting a primary amine that can be selectively acylated to allow for the introduction of different moieties. The use of **AMC** as a central template thus opens a wide field of possible applications (further discussed hereafter).

The mono-functionalization with different appendixes that could be effectively used as a chemical handle makes **AMC** an ideal scaffold for versatile G4-ligands but its synthesis, even if relatively straightforward at a first glance (*see below*), is cumbersome and poorly efficient. We thus decided to revisit the **AMC** chemical access and optimize it to make it compatible with large-scale production.

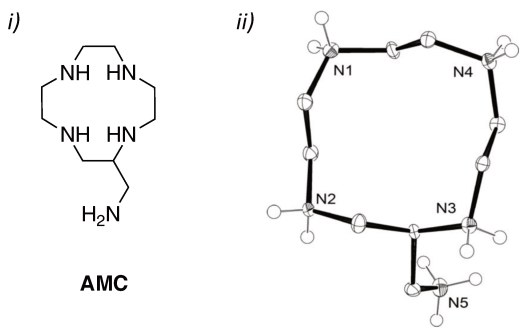
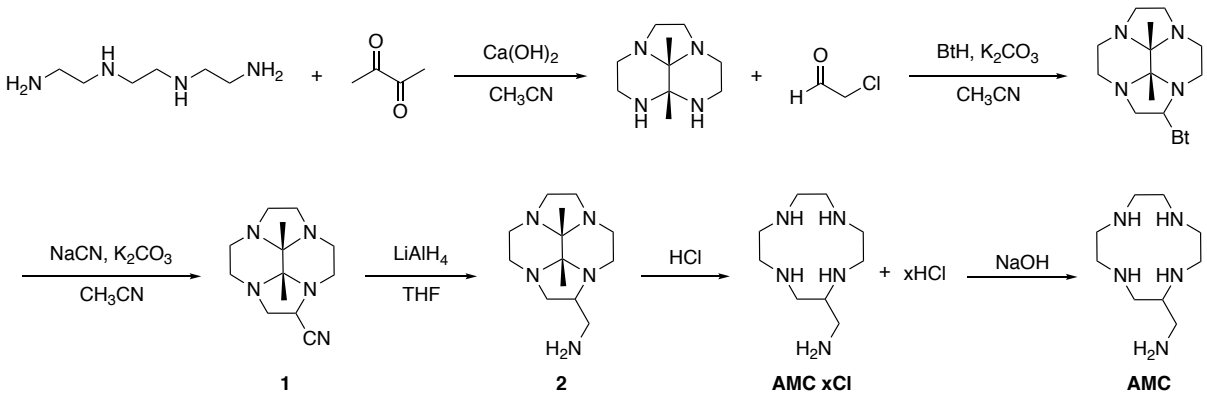


Figure 7 : i) AMC chemical structure, ii) ORTEP view of the penta-protonated AMC, showing thermal ellipsoids at the 50 % probability level. Hydrogen atoms attached to carbon atoms of the cyclic backbone, water molecules, and counterions have been omitted for clarity

I.1. The original synthesis of AMC.

The synthetic approach (*Scheme 1*) is based on the formation of a removable organic staple that allows for the formation of the macrocycle and avoids polymerization.



Scheme 1 : Synthetic pathway of aminomethylcyclen

The reported **AMC** synthesis[121] is the following: the linear polyamine 2,2,2-tetraamine was condensed with 2,3-butanedione in acetonitrile in presence of calcium hydroxide to afford a tricyclic *bis*-aminal. Chloroacetaldehyde and benzotriazole (Bt) were added to form the cyclen scaffold from which Bt was displaced by sodium cyanide providing nitrile **1**, which was then purified prior to its reduction with $LiAlH_4$ into a primary amine to provide **2**. As further discussed below, this purification step is crucial. The remaining *bis*-aminal bridge is then cleaved under aqueous acidic conditions to afford **AMC** as a hydrochloride salt. The free amine form is obtained after addition of sodium hydroxide and extraction with chloroform, which is usually performed freshly prior to its use. This synthesis is considered as poorly efficient since only 15 g of **AMC** are usually obtained from 140 g of linear tetraamine (13 % chemical yield). However, the reactants are unexpensive, making this synthetic route particularly attractive despite its global low yield.

I.2. The optimized synthesis of AMC.

This synthetic approach presents two major pitfalls which are detrimental for both a large-scale production and a possible commercialization. Firstly, the use of a hazardous and toxic chemicals (such as sodium cyanide and alkylating agents) in large quantities, and secondly a long and poorly reproducible purification step of compound **1** over an alumina column. This

time-consuming and cumbersome purification has to be carefully monitored by NMR to obtain the nitrile **1** in high purity, which is required to guarantee a precipitation of the **AMC** hydrochloride at the very last step of the reaction sequence.

While the first drawback (toxic reactants) is impossible to circumvent without changing the synthetic pathway, we decided to tackle the second drawback and explore a different purification technique. Based on preliminary observations, pentane was shown to solubilize only the targeted nitrile, thus, to obtain the nitrile **1** in high purity, we decided to consider an isolation of **1** *via* a Soxhlet purification and compared its efficiency with the classical alumina column purification. The extraction was carried out in pentane with the crude material adsorbed on alumina, this process was continued for 48 hours and the desired compound was obtained with a high degree of purity, as illustrated by NMR (*see experimental part*).

The Soxhlet purification turned out to be a far better option than the column chromatography. It was accompanied by: 1- an improved yield (from 26% to 38%), 2- an increased reproducibility, 3- a greatly reduced volume of used solvents (650 mL of pentane *vs.* 15L), and 4- a far lower number of NMR analyses (1 *vs.* 1 per eluted fraction). This purification method could be considered as time-consuming, but it does not require the presence of an operator (which is, in the context of a possible commercialization, a major advantage). The only drawback of this approach is the stability of the intermediate, which must be purified immediately after adsorption on alumina.

Overall, the global yield from the starting material (the tetraamine) is still low (13% for the pathway involving a purification on alumina column *vs.* 19% for the Soxhlet-based pathway), but the low cost of the starting materials, the reduced volumes required for the purification of compound **1**, the decreased number of analyses, and the increased reproducibility of the process pave the way towards a viable commercial exploitation

Chapter 2:

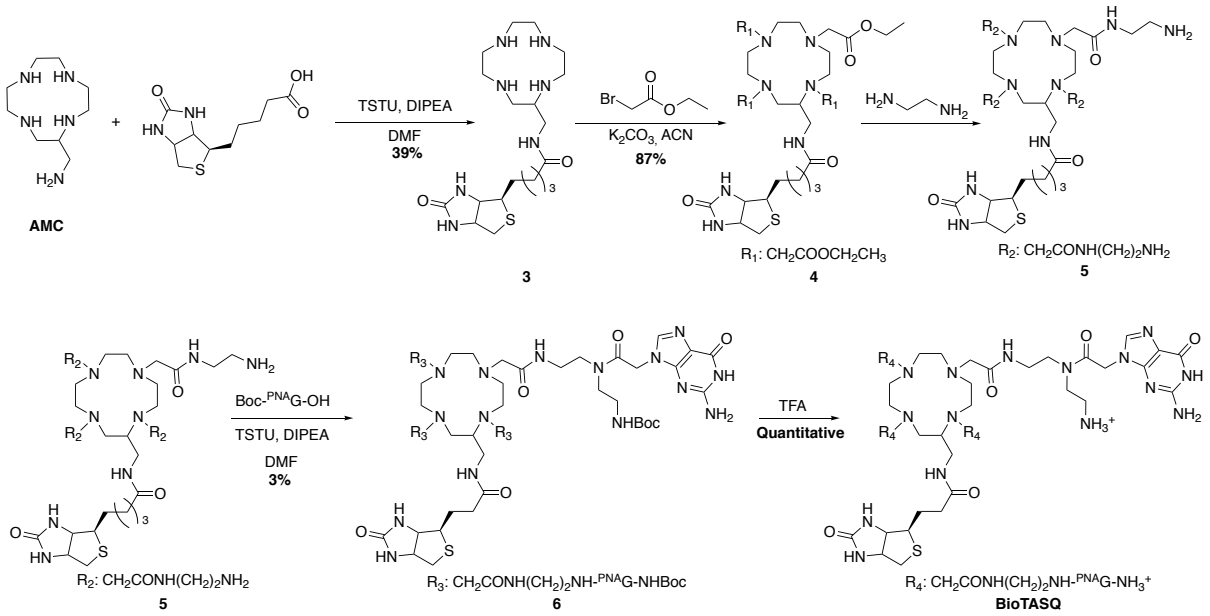
BioTASQ, CyTASQ & BioCyTASQ

Once the issues related to the chemical access of **AMC** were addressed, we decided to focus on how we could exploit the **AMC**'s unique feature, that is, the presence of a pendant primary amine. We first introduced a biotin on this amine to exploit the well-established biotin/avidin interaction for the subsequent manipulations of the resulting biotinylated TASQ. Indeed, the biotin/avidin interaction is among the strongest non-covalent interactions[122], which is thoroughly used in biochemistry and biology, notably for isolation purposes *via* streptavidin coated magnetic beads, their magnetic immobilization allows for a straightforward and easy separation of biotinylated molecules. This strategy was applied only marginally to the G4-field, and only a handful of biotinylated G4-ligands have been described in the literature[123-124]. Among them, our group developed in 2018 the first biotinylated TASQ named **BioTASQ**.[120] This molecule showed an excellent biophysical behavior and was used to isolate G4s both *in vitro* (the so-called fluorescence pull-down assay[125], *vide infra*) and from cell lysates (the so-called **G4RP-seq**[120, 126] protocol, *vide infra*). One limitation remains, related to the chemical accessibility of **BioTASQ** which is very troublesome, despite (again) an apparent simplicity. We thus decided to improve its synthesis, which was done according to two parallel approaches: 1- the optimization of the already developed synthetic scheme, thanks to the identification of the bottlenecks of the synthesis; and 2- the development of a closely related but new **BioTASQ** derivatives, which could display the same G4-interacting properties while being more chemically accessible.

II.1. The original synthesis of BioTASQ.

The first synthesis of **BioTASQ** (*Scheme 2*) was the following: **AMC** was reacted with the *N*-hydroxysuccinimide ester of the biotin in the presence of *N,N*-diisopropylethylamine (DIPEA) as base in anhydrous dimethylformamide (DMF) to provide the biotinylated **AMC** (**3**) in 39% chemical yield. Of note, it was necessary to work in diluted conditions (0.01 M) and with 0.9 equivalent of biotin to avoid *bis*-functionalization of **AMC**. Next, compound **3** was reacted with ethyl bromoacetate to obtain **4** which was purified by preparative reverse phase high pressure liquid chromatography (RP-HPLC) (87% chemical yield). The ethyl esters were then

reacted with ethylenediamine to obtain **5**, which displays four primary amines that can be easily reacted with ^{PNA}G. Guanine monomers were prepared in a one-step catalytic hydrogenation (Pd/C) from the commercially available Boc-^{PNA}G(Z) (72% yield).



*Scheme 2 : **BioTASQ** synthetic pathway*

Of note, the ethylenediamine step is time consuming (up to 8 days), difficult to monitor due to the polarity of the resulting product (not retained in a classical C18 RP-HPLC column) and the intermediate **5** could not be purified. The next step has thus to be performed on a crude mixture, activating the ^{PNA}G arms *via* the addition of *N,N,N',N'*-Tetramethyl-*O*-(*N*-succinimidyl)uronium tetrafluoroborate (TSTU) and DIPEA followed by the addition of the tetraamine **5**. The Boc-protected **BioTASQ** (**6**), which was purified by semi-preparative RP-HPLC, was obtained in a very low chemical yield (3%). The final step (deprotection of the Boc groups by a TFA treatment) being quantitative, the overall yield of the **BioTASQ** synthesis is very low (*ca*. 1%).

II.2. The optimized synthesis of BioTASQ.

The major bottleneck of this reaction scheme is the reaction between the intermediate **4** and ethylenediamine. The lack of isolation/purification of compound **5** negatively impacts the subsequent reactions, resulting in very low yield over two steps (3%). While the isolation of compound **5** was not possible (due to its very high polarity), we decided to improve the

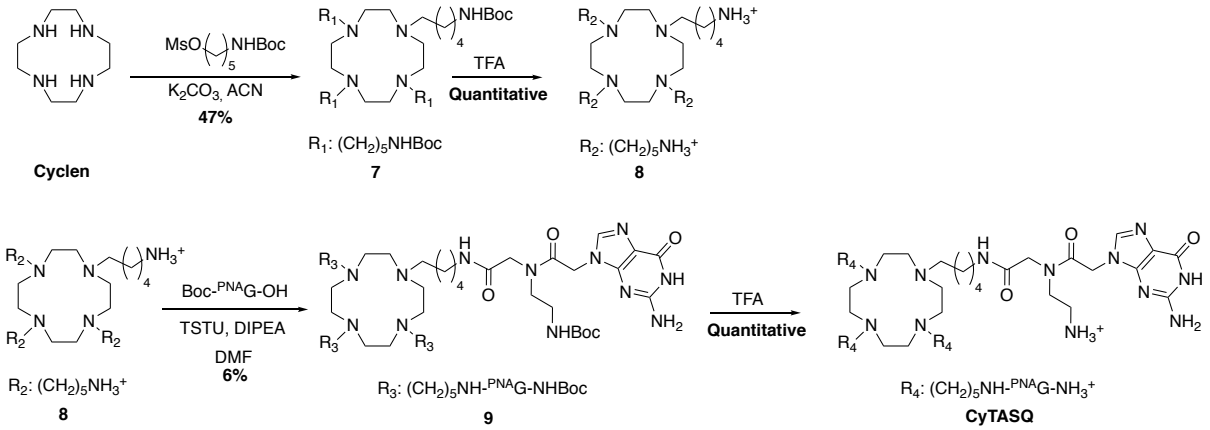
coupling reaction between **5** and ^{PNA}Gs modifying the coupling conditions. The initially used TSTU/DIPEA mixture was selected as it was efficient for the **AMC** mono-acylation (first step) and allowed for an easy monitoring of the formation of the activated ester by RP-HPLC. However, in order to improve the yield of the quadruple amide coupling, *N,N,N',N'*-Tetramethyl-*O*-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU)[127] was preferred. This coupling agent did improve the overall yield over two steps, and compound **6** was obtained with a 37% yield. Globally speaking, changing TSTU for HBTU improves the overall yield of the **BioTASQ** synthesis from 1 to 12%.

While this improvement was promising, the bottleneck of the synthesis was not solved, as compound **5** was still not isolated and its formation was still slow and difficult to monitor. We therefore decided to change the chemical nature of the spacer, swapping amide arms for lipophilic alkyl arms, whose introduction could be (in theory) monitorable and the intermediate isolated. To this end, we first had to show that the nature of the spacer did not modify the G4-interacting properties of the resulting biotinylated TASQ. We thus synthesized **CyTASQ** (*Scheme 3*), to investigate the influence of the arm on its ability to interact with G4s, prior to synthesize **BioCyTASQ** (*Scheme 4*), as a simplified **BioTASQ** derivative.[128]

II.3. The synthesis of CyTASQ.

The synthesis of **CyTASQ** (*Scheme 3*) is straightforward as it relies on the commercially available cyclen starting material. Cyclen was reacted with an excess of 5-(Boc-amino)pentylmesylate (prepared in two steps from the commercially available 5-aminopentan-1-ol, 93% chemical yield) in the presence of potassium carbonate (K_2CO_3) in acetonitrile (ACN). A five sp^3 carbon spacer was chosen between the central core and the ^{PNA}G, since the alkyl chain contains the same number of atoms as the aminoethylacetamide arms of **^{PNA}DOTASQ** and, as a consequence, it was estimated that it would preserve the self-assembly properties of the Gs. The introduction of the spacers provided compound **7** (47% chemical yield) that could be isolated by column chromatography and was subsequently deprotected with trifluoroacetic acid (TFA, quantitative) to afford compound **8** bearing four free amino groups (100% yield). These amines were then coupled with Boc-^{PNA}G monomers (4.4 mol. equiv) in presence of TSTU and DIPEA to afford the compound **9** in 6% chemical yield. Even though the yield increased up to two-fold (compared to compound **6**) thanks to

the isolation and characterization of compound **7**, its low yield can be attributed to the choice of the coupling agent, thus underlining the need of finding the most suited reagent to achieve the most efficient conversion. Inspired by the results achieved with **BioTASQ** (*vide infra*), an HBTU/DIPEA mixture was selected to yield to **BioCyTASQ** (*Scheme 4*).



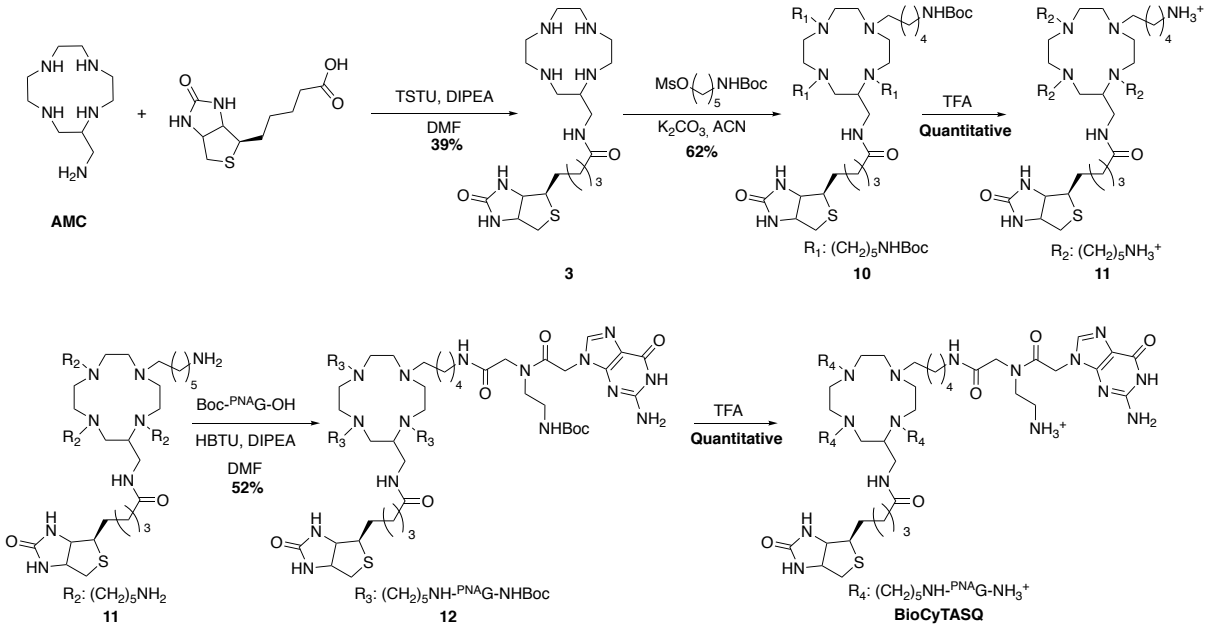
*Scheme 3 : **CyTASQ** synthetic pathway*

The final compound **CyTASQ** was obtained in a final, quantitative deprotection step (TFA). As further discussed below, the biophysical investigations performed with **CyTASQ** gave highly satisfying results in terms of G4-affinity and selectivity. On this basis, we decided to launch the synthesis of **BioCyTASQ** but, for sake of clarity, all biophysical results are presented in §II.5.

II.4. The synthesis of BioCyTASQ.

The synthesis of **BioCyTASQ** (*Scheme* 4) starts as that of **BioTASQ**, with the coupling of biotin and **AMC** in highly diluted conditions with TSTU and DIPEA. The biotinylated **AMC** was then alkylated with 5-(Boc-amino)pentylmethanesulfonate to obtain **10** (62% chemical yield). After Boc deprotection (TFA, quantitative), compound **11** was coupled to the HBTU-activated Boc-^{PNA}G-OH to obtain the Boc-protected **BioCyTASQ 12** in 52% chemical yield. The combination of the HBTU/DIPEA coupling mixture with the chemical properties of compound **10** that allows for its isolation by RP-HPLC, results in a major improvement of the quadruple coupling between **11** and Boc-^{PNA}G. After a final step of Boc deprotection by TFA (quantitative), **BioCyTASQ** was obtained with an overall yield of 12%, which represents a significant improvement when compared with the original synthesis of **BioTASQ**.

The final ^{PNA}G coupling to achieve biotinylated ligands was improved drastically: yields were increased from 3% for compound **6** using a TSTU/DIPEA mixture to 37% by swapping the latter for HBTU, and even up to 52% for compound **12** (thanks to the purification of compound **10** and the use of an HBTU/DIPEA mixture).



Scheme 4 : ***BioCyTASQ*** *synthetic pathway*

Therefore, the conceptually simple modification of the linker between the cyclen/AMC core and the ^{PNA}G arms results not only in a more efficient chemical access to biotinylated TASQ derivatives but also, as it will be further described hereafter, in an improvement in all the G4-interacting properties, which makes **BioCyTASQ** the new benchmark for biotinylated TASQs.

II.5. Evaluation of the G4-interacting properties of CyTASQ and BioCyTASQ.

In order to investigate the G4-binding ability of these newly developed ligands we used FRET-melting experiments (*see below*)[129]. This high-throughput screening assay allows for a side-by-side comparison of the four TASQs (*Figure 8*): **^{PNA}DOTASQ**, composed of amide spacers and no biotin appendage, **BioTASQ**, assembled with amide spacers and with a biotin handle, **CyTASQ**, with alkyl spacers and no biotin appendage, and **BioCyTASQ,** which bears both alkyl spacers and a biotin handle. FRET-melting experiments give insights into the thermal

stabilities of both DNA and RNA G4s in the presence and in the absence of ligands. The results of these experiments thus provide an estimation of the stability of the TASQ/G4-complex, that is, of the apparent G4-affinity for studied ligands. When performed in the presence of an excess of unlabeled duplex-DNA competitor, competitive FRET-melting assay also gives information about the selectivity of the TASQ for G4s over dsDNA.

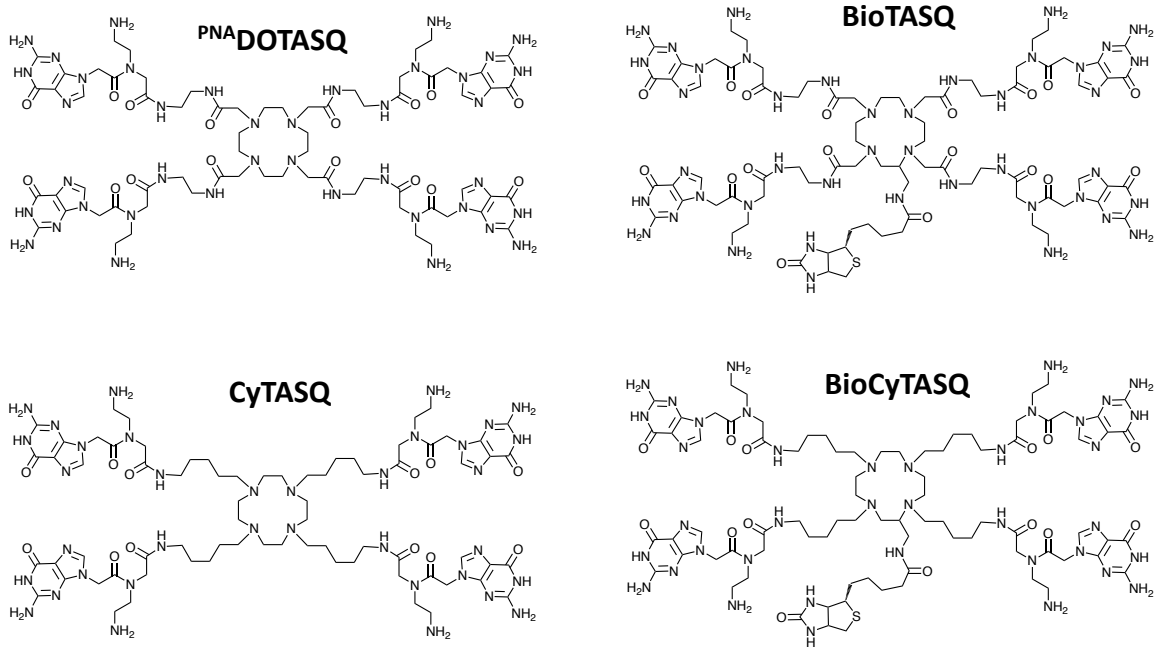


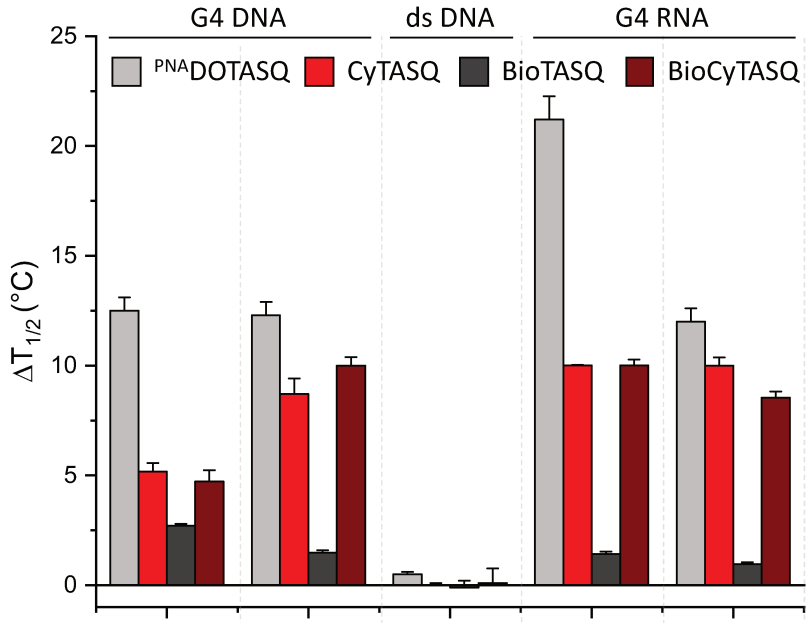
Figure 8 : Chemical structures of TASQs whose G4-interacting properties were analyzed by FRET-melting experiment

This technique relies on the use of QFSs labelled with FAM (a 6-carboxy-fluorescein amidite) at their 5'-end and with a TAMRA (a 6-carboxy-tetramethylrhodamine amidite) at their 3'-end. FAM and TAMRA belong to a so-called 'FRET pair' as the FAM emission spectrum overlaps with the TAMRA excitation spectrum, making an energy transfer possible between the two fluorophores *via* a FRET effect (Förster (or fluorescence) Resonance Energy Transfer). When a DNA and/or RNA G4 is folded, the spatial proximity of the two fluorophores results in the energy transfer of the FAM emission to the TAMRA through nonradiative dipole-dipole coupling, thus quenching the FAM fluorescence. As the temperature increases, DNA/RNA G4s progressively unfold (or 'melt'), which triggers the spatial separation of the two fluorophores and abolishes the energy transfer. The stability of the DNA/RNA G4s can thus be readily quantified monitoring the FAM fluorescence (normalized between 0 and 1) as a function of the temperature (25 to 90 °C). This stability of the DNA/RNA G4 alone is expressed as a mid-

transition temperature, or $T_{1/2}$ (°C). The TASQ ligands can stack on folded G4s, which results in the stabilization of the secondary structures and leads to an increase of the $T_{1/2}$ values. The apparent affinity of a ligand for a G4 can thus be quantified by the shift in $T_{1/2}$ values ($\Delta T_{1/2}$, °C), calculated as follows: $\Delta T_{1/2} = T_{1/2}$ (G4 + ligand) - $T_{1/2}$ (G4 alone).

II.5.a. Comparative FRET-melting experiments.

As seen in *Figure 9*, all TASQs (1 µM, 5 mol. equiv.) display moderate to good affinity for G4 ($\Delta T_{1/2}$ being comprised between 1.5 ± 0.1 and 12.5 ± 0.6 °C for G4-DNA, and between 1.4 ± 0.1 and 21.2 ± 1.1 °C for G4-RNA) along with an excellent selectivity over duplex-DNA (as no stabilization is observed with the control hairpin DNA F-duplex-T).



	F-21-T	F-Myc-T	F-duplex-T	F-TERRA-T	F-VEGF-T
PNADOTASQ	12.5 ± 0.6°C	12.3± 0.6°C	0.5± 0.1°C	21.2± 1.1°C	12.0± 0.6°C
CyTASQ	5.2± 0.4°C	8.7± 0.7°C	0.1± 0.1°C	10.0± 0.2°C	10.0± 0.4°C
BioTASQ	2.7± 0.1°C	1.5± 0.1°C	-0.1± 0.3°C	1.4± 0.1°C	1.0± 0.1°C
BioCyTASQ	4.7± 0.5°C	10.0± 0.4°C	0.1± 0.6°C	10.0± 0.2°C	8.6± 0.3°C

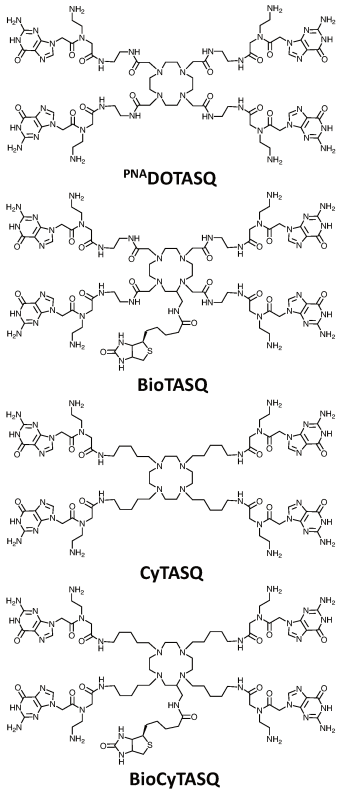


Figure 9 : FRET melting results with a panel of oligonucleotides ODN and 5 equiv. of ***PNADOTASQ, CyTASQ, BioTASQ*** *and* ***BioCyTASQ*** *(n=3).*

In more detail, it is interesting to directly compare ligands to better understand the role of the spacer and of the biotin moiety:

- **^{PNA}DOTASQ *vs.* CyTASQ:** a decrease in the apparent G4 affinity could be noted when comparing **^{PNA}DOTASQ** *vs.* **CyTASQ**, especially for F-21-T and F-TERRA-T, from 12.5 ± 0.6 to 5.2 ± 0.4 °C and from 21.2 ± 1.1 to 10 °C, respectively. This decrease can be explained by the ability of the amide linkers to create intramolecular hydrogen bonds (HB). This network ensures an increased level of pre-organization that results in a favored folding of the intramolecular synthetic G-quartet (*Figure 10*), which thus increases its ability to interact with G4s (apparent affinity).

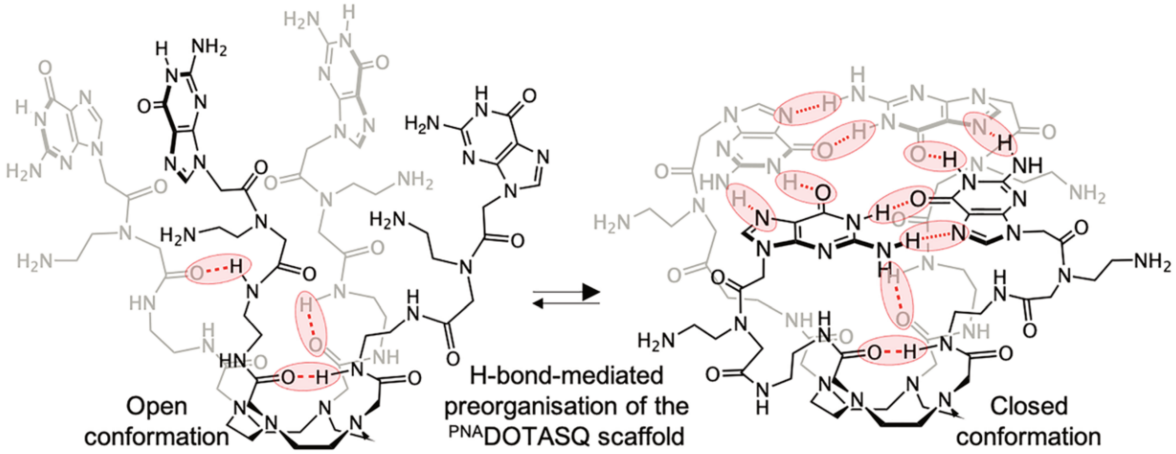


Figure 10 : Schematic representation of possible intramolecular H-bonds in the ^{PNA}DOTASQ scaffold

- **^{PNA}DOTASQ *vs.* BioTASQ:** The difference in apparent G4-affinity is quite drastic when comparing **^{PNA}DOTASQ** *vs.* **BioTASQ**. While the former shows a good G4-affinity, **BioTASQ** displays a much lower affinity for G4s, as exemplified by the $\Delta T_{1/2}$ obtained with F-21-T and F-TERRA-T, which is only of 2.7 ± 0.1 and 1.4 ± 0.1 °C only. This low affinity could originate in the HB donating/accepting ability of the amide bonds. The biotin can interact with the amide linkers (*Figure 11*) which could in turn impede the formation of the synthetic G-quartet and decrease its binding abilities. The favorable preorganization features observed with **^{PNA}DOTASQ** are thus completely lost in **BioTASQ** due to the introduction of a biotin appendage.

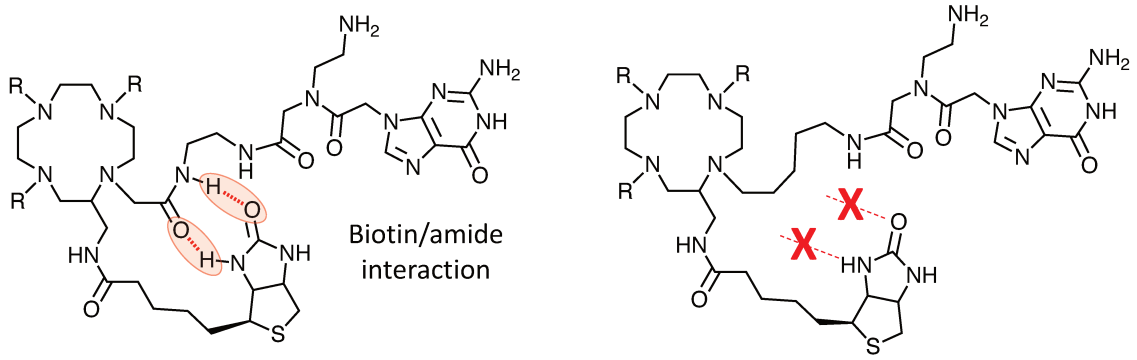


*Figure 11 : **BioTASQ** and **BioCyTASQ** hydrogen bond interaction between the biotin and the spacer.*

- BioTASQ *vs*. BioCyTASQ: Finally, the comparison of the two biotinylated probes, *i. e.*, **BioTASQ** with the amide linkers and **BioCyTASQ** with the alkyl linkers, clearly shows that **BioCyTASQ** systematically outperforms **BioTASQ**, which indicates that the lack of HB donating/accepting ability of the **BioCyTASQ** alkyl arms precludes interactions with biotin, which thus allows for avoiding the internal poisoning that was observed for **BioTASQ**. This hypothesis is further confirmed by the direct comparison of **BioCyTASQ** and **CyTASQ**: the lack of amides (and thus HB network) for both TASQs results in similar apparent affinity values for the two probes, with and without biotin. These results are of utmost importance as they show that changing the nature of the arm allows for the design of a biotinylated TASQ which displays high affinity for G4s, making it an ideal molecular tool to investigate G4s *in vitro*. However, its selectivity for G4s over duplex-DNA must be established prior to launch cell-based assays.

II.5.b. Competitive FRET-melting experiments.

The G4-selectivity can be evaluated through competitive FRET-melting assays. These experiments are performed incubating labeled G4s (0.2 µM) in the presence of unlabeled double stranded DNA competitor (ds17: 15 and 50 mol. equiv.) and **CyTASQ** or **BioCyTASQ** (5 mol. equiv.). The $\Delta T_{1/2}$ values obtained without competitor were normalized to 100 to provide a direct visualization of modification of the affinity in the presence of the competitor.

	CyTASQ		
Eqv ds17	0	15	50
F-Myc-T ^{FRET}S	1	0.99	0.96
F-TERRA-T ^{FRET}S	1	0.95	0.94
	BioCyTASQ		
Eqv ds17	0	15	50
F-Myc-T ^{FRET}S	1	0.98	1.07
F-TERRA-T ^{FRET}S	1	1.03	1.03

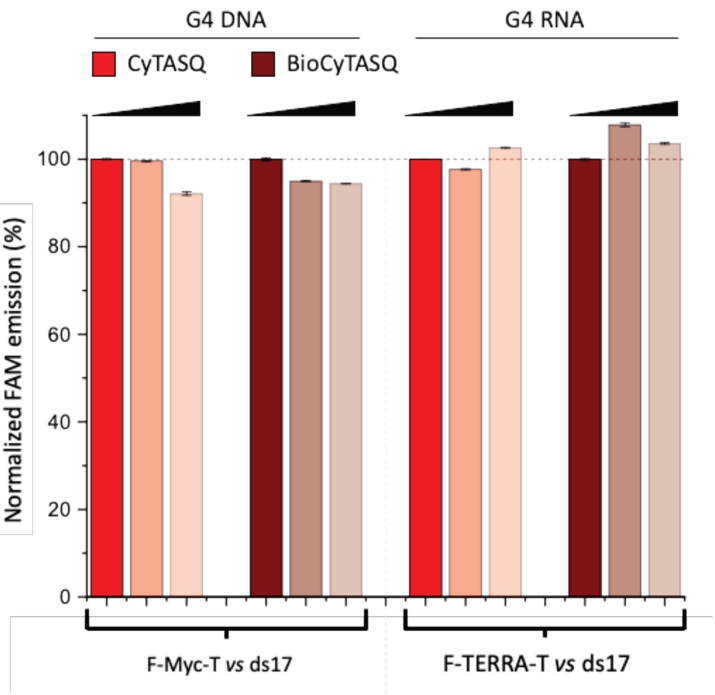


*Figure 12: Results of competitive FRET-melting experiments performed with doubly labeled G4-DNA (F-Myc-T) or G4-RNA (F-TERRA-T) in the presence of TASQ (**CyTASQ** in red, **BioCyTASQ** in wine) and increasing amounts of unlabeled duplex-DNA competitor (ds17, 15, and 50 molar equiv). Selectivity values (^{FRET}S) calculated for **CyTASQ** and **BioCyTASQ** in the following way: $^{FRET}S = \Delta T_{1/2(with\ competitor)}/\Delta T_{1/2(without\ competitor)}$); values between 0.94 and 1.07 showed excellent selectivity. (n=3)*

Both TASQs are highly G4-specific (*Figure 12*) with a G4-stabilization that is >92% (for **CyTASQ**) and >98% maintained (for **BioCyTASQ**) in the presence of a large excess of competitor. This exquisite selectivity makes them ideally suited to be used to investigate G4s in cells (further discussed below).

In conclusion, this first series of *in vitro* experiments demonstrated that the chemical simplification of the TASQ's linkers led to a biotinylated TASQs that is more efficient than **BioTASQ**. The increased flexibility as well as the non-HB donating/accepting nature of the alkyl spacer resulted in a probe, **BioCyTASQ**, with a higher apparent affinity and an excellent

selectivity for G4s over dsDNA. The efficiency of the biotin handle that allows for using biotinylated TASQs as baits for G4s will be discussed in the following chapters.

II.5.c. Fluorescence G4-pull-down assay:

We developed a fluorescence G4-pull-down assay to quantify the ability of biotinylated TASQs to isolate G4s using streptavidin-coated magnetic beads. The pull-down efficiency is monitored by fluorescence measurements is as follows (*Figure 13*):

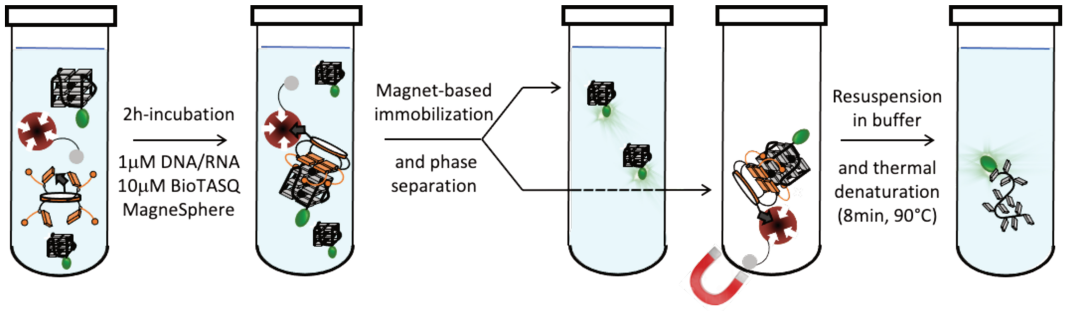
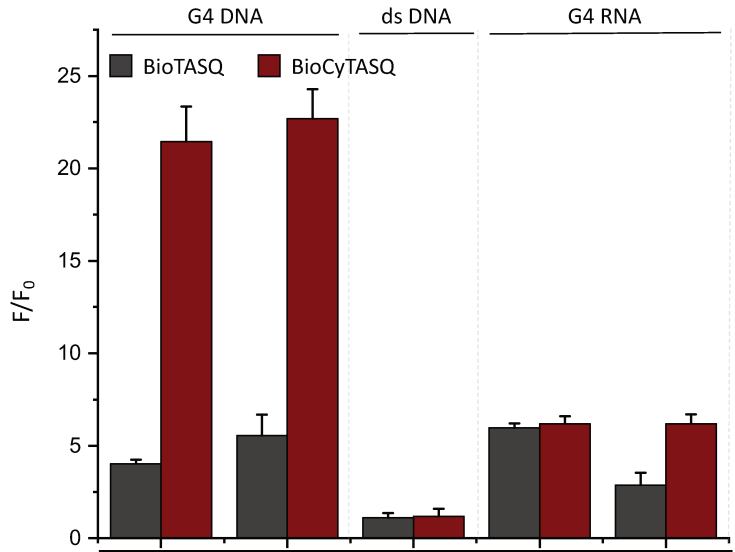


Figure 13 : Schematic representation of the fluorescence pull-down protocol.

FAM-labeled G4-forming DNA/RNA sequences (FAM-G4, 1 µM) were incubated for 2 h at room temperature with biotinylated TASQs (**BioTASQ** or **BioCyTASQ**, 10 µM) in the presence of streptavidin-coated magnetic beads in TrisHCl buffer (20 mM, pH 7.2) containing 1 mM KCl, 99 mM LiCl and 10 mM $MgCl_2$ (this KCl/LiCl ratio, which corresponds to the FRET-melting buffers, ensures proper G4 stability at 25 °C ; the Mg^{2+} decreases the non-specific electrostatic interactions between DNA and TASQs). The FAM-G4/TASQ/beads assemblies were isolated (magnetic immobilization), the supernatant was removed and the FAM-G4 was resuspended in solution by a thermal denaturation step (8 min at 90 °C). The capture/release efficiency of TASQ was quantified by the FAM emission of the resulting solution, normalized to the controls performed without TASQ.

The G4-capture ability of both **BioTASQ** and **BioCyTASQ** was compared against a panel of fluorescently labeled DNA (the G4s F-Myc and F-SRC, the dsDNA F-duplex as control) and G4-RNA (F-TERRA, F-VEGF). As seen in *Figure 14*, both ligands show an excellent ability to capture G4s, with better performances for **BioCyTASQ** over **BioTASQ**. For instance, the final fluorescence level obtained of F-Myc is 21.4 ± 1.9-fold enhanced with **BioCyTASQ**, while it is

only 4.0 ± 0.2-fold with **BioTASQ** (value normalized to 1 on the fluorescence of the experiment performed without TASQ). Interestingly, no enrichments were obtained with F-duplex. Regarding RNA G4s, both TASQs showed similar enrichments factor (5.7 ± 0.2 *vs.* 6.2 ± 0.4 for F-TERRA with **BioTASQ** and **BioCyTASQ** respectively, 2.9 ± 0.6 *vs.* 6.2 ± 0.5 for F-VEGF), these values are lower than that of DNA G4s, indicating that both probes are more effective for isolating DNA G4s than RNA G4s (further discussed in *Chapter 3*).



	F-Myc	F-SRC	F-duplex	F-TERRA	F-VEGF
BioTASQ	4.0 ± 0.2	5.5 ± 1.1	1.1 ± 0.2	5.7 ± 0.2	2.9 ± 0.6
BioCyTASQ	21.4 ± 1.9	22.7 ± 1.6	1.2 ± 0.4	6.2 ± 0.4	6.2 ± 0.5

Figure 14 : Fluorescence pull-down results: normalized change in the fluorescence values after the denaturation step (n=3).

In conclusion, **BioCyTASQ** has proven not only to be far more chemically accessible than **BioTASQ**, but it also displays better G4-interacting properties than **BioTASQ**, with a higher apparent affinity, a similar selectivity and better pull-down abilities. These excellent *in vitro* results thus open ways towards the study of their G4-isolating in a more biologically relevant context, notably investigating the way both **BioTASQ** and **BioCyTASQ** fish G4s out of cell lysates, and to compare these properties to the well-established antibody **BG4**[37].

II.5.d. G4-pull-down assay in cells:

The ability of **BioTASQ** to isolate G4s from cell lysates was proven thanks to the development of the **G4RP-seq**[120, 126] protocol (for G4-RNA precipitation and sequencing). Human cancer cells were first crosslinked with formaldehyde to halt biological processes and stabilize

transient structural interactions. Harvested cells were then sonicated briefly to release cellular content, cell lysates were incubated overnight with a high concentration of **BioTASQ** (100 µM) before affinity purification with magnetic streptavidin beads. After extraction with TRIzol, the RNA fragments underwent reverse transcriptase quantitative PCR (RT-qPCR) with gene-specific primers to confirm the efficiency of the protocol. The results showed that **BioTASQ** is able to selectively isolate known RNA QFSs such as VEGFA and NRAS.

Prompted by the better G4-DNA isolation results of both **BioTASQ** and **BioCyTASQ** (*Figure 14*) and with the help from Wenli Zhang group (Nanjing University, China), we set out to investigate the abilities of both **BioTASQ** and **BioCyTASQ** to isolate G4-DNA and compared these results side-by-side with **BG4**. We focused on rice genome since G4-DNAs were already isolated and identified by W. Zhang *et al*. thanks to the use of **BG4** *via* the so-called **BG4-IP-seq**[130]. Our goal was thus to develop the **G4DP-seq**[131] protocol (for G4-DNA precipitation and sequencing). To make the investigation as thorough as possible, two parallel approaches were pursued: an *in vitro* approach, with the goal of detecting G4s in purified and fragmented genomic DNA (as already performed with **BG4** (**BG4-IP-seq**)); and an *in vivo* approach, in which G4s were fixed in their native context before being isolated by precipitation with either **BG4** (immunoprecipitation, IP) or TASQ (chemoprecipitation).

II.5.d.1. *in vitro* G4DP-seq:

BG4-IP-Seq was performed as follows: after fragmentation of rice genomic DNA (sonication) into 100- to 500-bp sequences in size, DNA segments were thermally treated (10 min at 90 °C, followed by a slow return to 25 °C in G4-stabilizing conditions, 150 mM K^+) in order to maximize G4 folding. **BG4-IP-Seq** relies on the use of a FLAG-tagged **BG4**, a BG4 modified with a FLAG-TAG that allows specific recognition and isolation (expressed from the pSANG10-3F-BG4 plasmid, 3 µg for 5 µg of DNA), followed by the addition of an anti-**BG4** Ab (anti-FLAG, 3 µg) and then, protein G-coupled magnetic beads for pulling down the G4/BG4/anti-FLAG/beads complexes (*Figure 15)*. The **BG4-IP-Seq** protocol was then adapted to the TASQ molecular tools: **G4DP-seq** relies on the treatment of the prepared DNA fragments (G4-optimized) with TASQ (either **BioTASQ** or **BioCyTASQ**, 100 µM), followed by the addition of streptavidin-coated magnetic beads for pulling down the G4/TASQ/bead complexes. In both instances (**BG4-IP-seq** and **G4DP-seq**), thermal elution steps (65 °C twice, 15 min each)

followed by DNA purification (PhOH/$CHCl_3$ extraction, then EtOH precipitation) provide samples ready for library preparation.

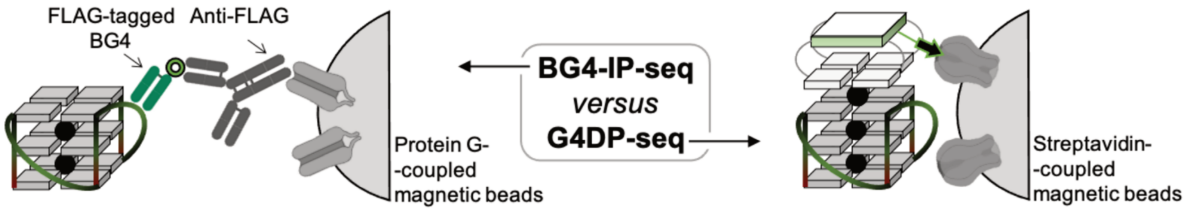


*Figure 15 : Schematic representation of the BG4-IP-seq performed with the antibody **BG4** and the G4DP-seq performed with two biotinylated small molecules, **BioTASQ** and **BioCyTASQ.***

The reproducibility of both **BG4-IP-seq** and **G4DP-seq** protocols was assessed by sequencing two independent libraires for each condition. Collected results (*Figure 16A*) show an elevated ratio of high-quality common reads: 71% for **BG4**, 94% for **BioTASQ** and 74% for **BioCyTASQ**. These reads were used for G4-peak calling (using QuadParser, using $G_{\geq 2}N_xG_{\geq 2}N_xG_{\geq 2}N_xG_{\geq 2}$ synthax), and results seen in *Figure 16B* show that the two TASQs lead to the identification of a higher number of G4-peaks (59,791 and 49,898 for **BioTASQ** and **BioCyTASQ** respectively) as compared to **BG4** (29,469). This could be ascribed to the G4-stabilizing properties of TASQs (while that of **BG4** remains undescribed); it is however important to note that TASQs do not promote G4 folding (as further discussed below), which could introduce a bias in the analysis. This first set of data indicates that both Ab and ligands capture G4s efficiently, with a somehow better performance of TASQs *vs.* BG4 in terms of number of detected peaks. The genomic distributions of these peaks are globally comparable (*Figure 16C*), the most abundantly detected G4s belonging to promoters, distal intergenic regions and exons, although with some variations in the distribution between probes. Pairwise comparisons were performed (*Figure 16D*) and indicate that the 8,877 common G4s (detected by both Ab and TASQs, *vide infra*) are primarily distal intergenic G4s (53%), while those enriched specifically by each bait are exonic G4s with TASQs.

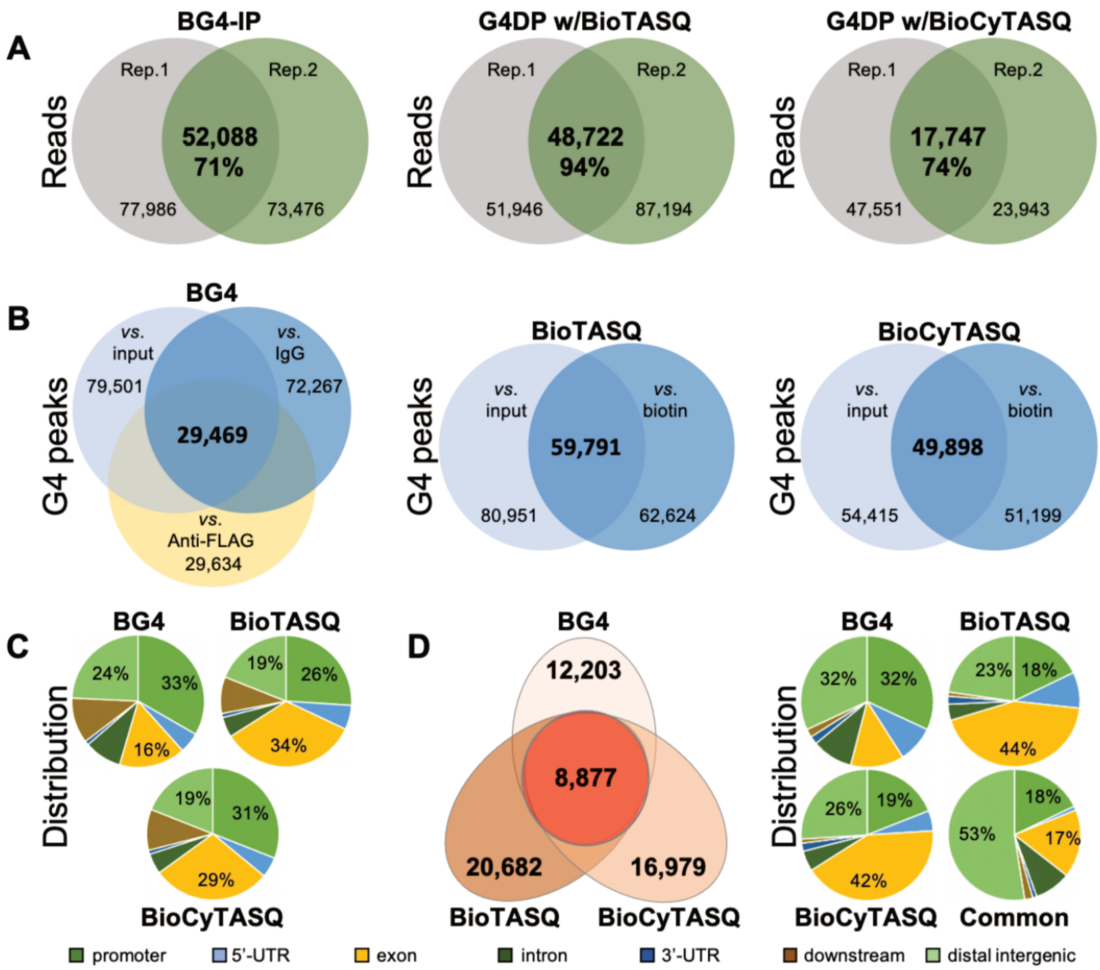


Figure 16: ***A.*** *Venn plots illustrating the reproducibility of* ***BG4-IP-seq*** *(performed with* ***BG4****) and* ***G4DP-seq*** *(performed with either* ***BioTASQ*** *or* ***BioCyTASQ****) protocols.* ***B****. G4 peak calling performed against controls (input, IgG and anti-FLAG for* ***BG4-IP-seq****; input and biotin for* ***G4DP-seq****) that leads to the identification of high- confidence G4 peaks.* ***C****. Genomic distribution of the high-confidence G4 peaks.* ***D****. Pairwise comparison of G4 peaks and genomic distribution of the 12,203* ***BG4****-specific peaks, 20,682* ***BioTASQ****-specific peaks, 16,979 BioCyTASQ-specific peaks and 8,877 common peaks.*

The 8,877 common G4s can be considered as high-confidence G4s as they are trapped and enriched by two orthogonal techniques (antibody- and ligand-based affinity capture), which is quite unique and leaves no doubt about their G4-nature. This was further established by a qualitative analysis of the isolated oligonucleotide (multiple expectation-maximizations for motif elicitation analyses, MEME), which shows that the common G4-motifs correspond mostly to two-quartet G4s of general sequence $G_2N_1G_2N_1G_2N_{1\text{-}2}G_2$. The pairwise comparison also identifies G4s motifs specifically enriched by **BioTASQ** (20,682), **BioCyTASQ** (16,979) and **BG4** (12,203) that were similarly analyzed without revealing any bias for particular G4 structural types. This observation lends credence to the interchangeable use of either capture tools.

Finally, it was of interest to further compare the properties of the two TASQs. As indicated above, the G4 peaks enriched by both TASQs contain a high number of QFS. The overlap between detected G4 peaks reveals a quite homogeneous distribution with 82.3% common peaks. However, some discrepancies can be noted: a closer look to the G4 peaks indicates that the distribution of length is wider for **BioCyTASQ** (average length 550 bp) than for **BioTASQ** (average length 370 bp), which underlines that difference in capture efficiency exist between the two TASQs, in line with data presented above (*Figure 14*).

II.5.d.2. In vivo G4DP-seq:

These results led us to investigate whether TASQs could be suited for the capture of G4s *in vivo*. The **G4DP-seq** protocol was thus adapted using **BioTASQ** as bait since it was used for the **G4RP-seq**, which similarly traps G4-RNA in *in vivo*-like conditions. Rice seedlings were cross-linked using formaldehyde (1%) before being ground in liquid nitrogen to disrupt cell walls and isolate nuclei (several washing steps), which were subsequently lysed for isolating genomic DNA. Cross-linked chromatin was thus fragmented (100- to 500-bp), which was followed by **BioTASQ** incubation (4 °C, overnight) and addition of streptavidin-coated beads for pulling down the cross-linked G4/TASQ/bead complexes. After thermal elution and cross-link reversal steps, DNA was precipitated for preparation of the sequencing libraries as described above. The comparison of the *in vitro*/*in vivo* conditions was performed by sequencing two replicated libraries for each condition. Results seen in *Figure 17A* show that very consistent results were obtained in both conditions, with 59,791 and 57,500 high-confidence G4-peaks identified for *in vitro* and *in vivo* G4DP-seq, respectively. The overlap between the G4-peaks is important, with >60% common G4s, as well as the agreement between the two **G4DP-seq** data sets, with >70% common peaks. However, a significant difference was found when analyzing the genomic distributions of these peaks: while the distribution for *in vitro* G4-peaks was in line with those previously obtained, G4s identified by i*n vivo* **G4DP-seq** were particularly enriched in exons (62%), which is in line with the recent involvement G4s in alternative splicing events at the transcripts level[53-54].

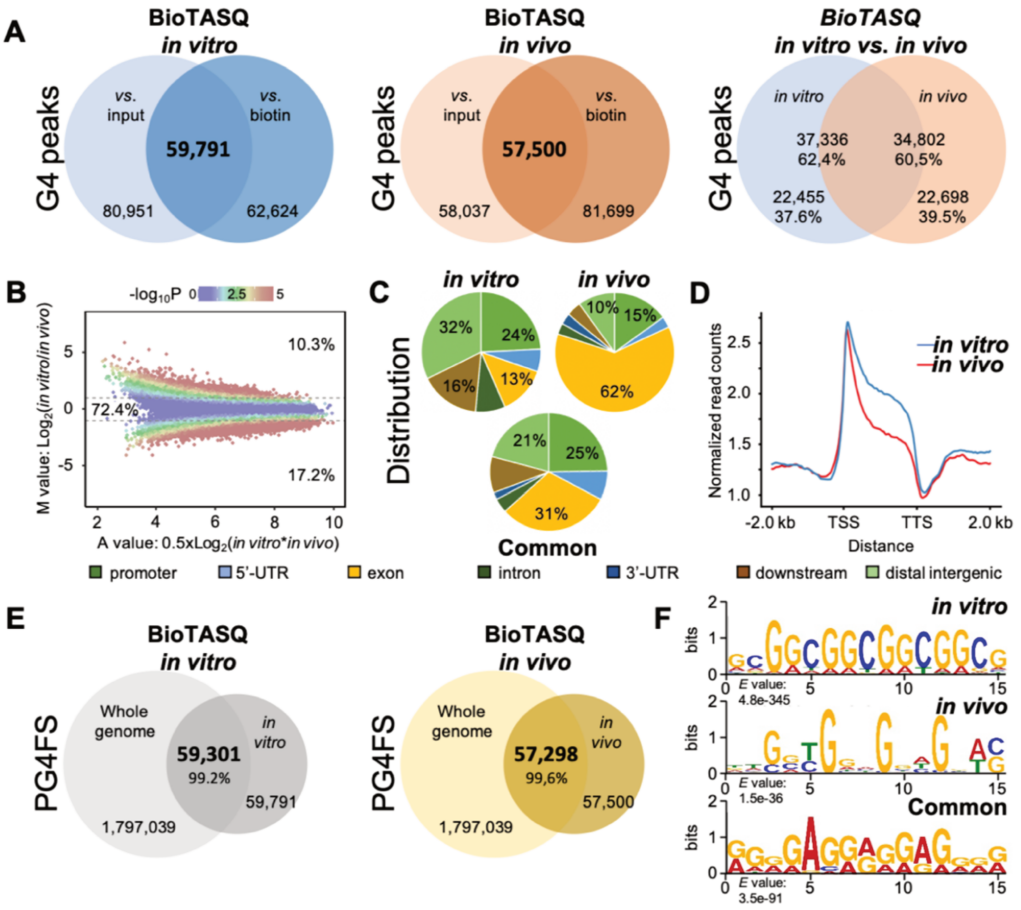


*Figure 17: **A.** Venn plots illustrating the G4-peak calling against two controls (input and biotin) for **G4DP-seq** performed in vitro or in vivo, leading to the identification of high-confidence G4-peaks, which are next compared (right plot). **B.** MA-plot illustrating the significant overlap (>70%) between G4-peak data sets of **G4DP-seq** performed in vitro or in vivo. **C.** Genomic distribution of the high-confidence G4 peaks detected in vitro, in vivo or in both conditions (common). **D.** Profiling of in vivo and in vitro **G4DP** read counts across ± 2 kb from the transcription start sites (TSSs) to the transcription terminate sites (TTSs) of genes. **E.** Number of PG4FSs identified by QuadParser for **G4DP-seq** performed in vitro or in vivo, compared to genome-wide PG4FS. **F.** Motifs discovery using MEME for G4DP-seq peak data set, performed with the 22,455 in vitro-specific peaks, 22,698 in vivo-specific peaks and the 34,802 common peaks, which reveal the presence of repetitive, G-rich motifs.*

The nature of the identified G4-peaks was investigated *in silico*: a QuadParser analysis of the high-confidence peaks (59,791 for *in vitro* **G4DP-seq**; 57,500 for *in vivo* **G4DP-seq**) revealed a very high G4-motif content, with >99% of these peaks containing the consensus $G_{\geq 2}N_xG_{\geq 2}N_xG_{\geq 2}N_xG_{\geq 2}$ sequence. This high content represents only a minor fraction (*ca*. 3%) of the QFS detected genome-wide (*ca*. 1,800,000 putative G4-motifs). The very low difference (<0.1%) between the numbers of G4-peaks identified with **BioTASQ** *in vivo* and *in vitro* (57,500 and 59,791, respectively) confirms that the TASQ does not promote G4-folding in our operating conditions.

Finally, the G4-motif occurrence within the 22,455 peaks detected *in vitro*, 22,698 peaks detected *in vivo* and the 34,802 common peaks was investigated *via* MEME analyses: the most enriched motifs were again found to correspond to two-quartet G4s; the particular G-richness observed for the common peaks is interesting as these G4s were trapped and enriched by two orthogonal techniques (*in vitro vs in vivo* **G4DP-seq**), which, again, leaves no doubt about their G4-nature.

In conclusion, the performance of **BG4** was compared to that of two TASQs under strictly similar experimental conditions. Both Ab and TASQs efficiently fish G4s out from cell lysates both in *in vitro* conditions and in *in vivo*-like conditions. When compared in terms of mapping quality (clean reads), G4-peaks calling and genomic distribution, both techniques provide an accurate portrayal of G4-biology, and this series of results show how **BioTASQ** and **BioCyTASQ** can thus successfully compete with antibodies (**BG4**) for affinity purification.

II.5.e. Pre-targeted G4 imaging:

The results described above show how the biotin/streptavidin interactions can be used to investigate the genomic (**G4DP-seq**) and transcriptomic (**G4RP-seq**) prevalence of G4s. We reasoned that this interaction could also be exploited for different purposes, notably for visualizing G4s in cells if we used fluorescently labeled streptavidin derivatives.

We thus decided to investigate whether both **BioTASQ** and **BioCyTASQ** could help localize G4s in human cancer cells (MCF7) using a Cy3-labeled streptavidin (SA-Cy3). Two strategies were implemented, with a systematic comparison of the properties of **BioTASQ** and **BioCyTASQ**. In the first, so-called 'post-fixation strategy', MCF7 cells are first fixed and permeabilized (cold MeOH) and incubated sequentially with either **BioTASQ** or **BioCyTASQ** (1 µM, 1 h), SA-Cy3 (1 µg/mL) and DAPI (2.5 µg/mL). In the second, so-called 'live-cell strategy', MCF7 cells are first treated live with either **BioTASQ** or **BioCyTASQ** (1 µM, 24 h; IC_{50} > 100 µM) before being fixed and permeabilized (cold MeOH), and labeled with SA-Cy3 and DAPI.

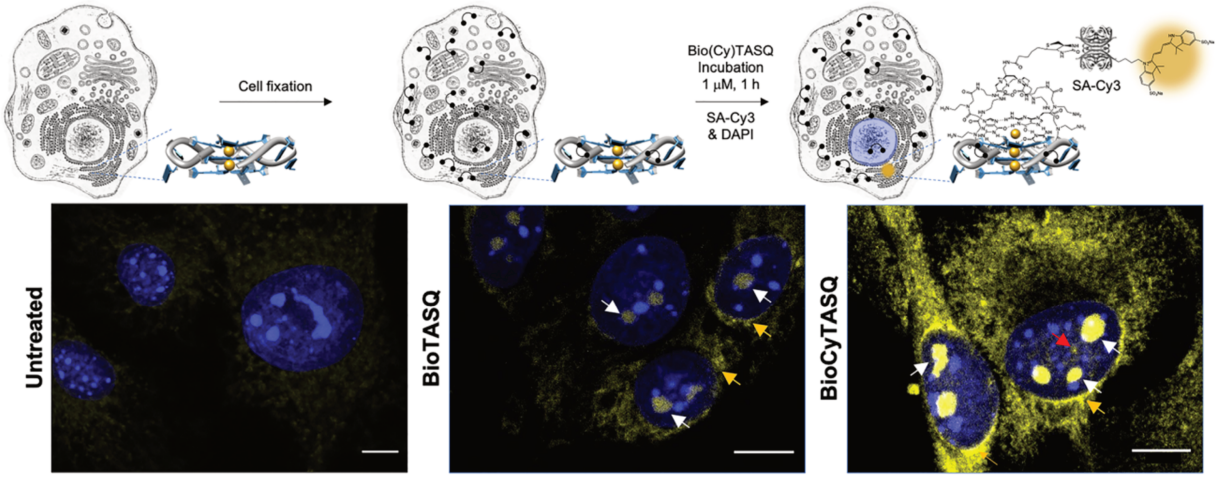


*Figure 18 : Schematic representation (and microscopy images) of the **BioTASQ/BioCyTASQ**-based, two-step strategies implemented to label G4s in MCF7 cells with post fixation labeling. Images are collected in the range of 518–562 nm for SA-Cy3 (yellow) and <425 nm for DAPI (blue channel); scale bar 10 µM; yellow, white and red arrows indicate G4-foci found in perinuclear regions, nucleoli and nucleoplasm, respectively.*

The first strategy leads to a diffuse labeling in the cytoplasm (*Figure 18*) along with stronger labeling in perinuclear regions (yellow arrows), nucleoli (white arrows) and non-nucleoli nucleoplasmic sites (red arrows). Both probes have the same distribution, with the **SA-Cy3/BioCyTASQ** providing a more intense signal. The second strategy provides a similar perinuclear/nucleoli/nuclear distribution (*Figure 19*) but with a higher contrast, along with abundant G4-RNA foci, mostly found in the ribosome sites (nucleoli (white arrows) and rough endoplasmic reticulum (yellow arrows) and sparser G4-DNA sites (located in the nucleoplasm (red arrows)).

Globally speaking, these images (both post-fixation and live-cell incubation) provide rather low-quality image compared to picture obtained with TASQ specifically designed for G4-visualization (*Chapter 5*) but turn out to be useful anyway as they 1- demonstrate the feasibility of the pre-targeting G4-imaging, and 2- allow for envisioning a global approach in which the same molecular tools could be used for visualizing (pre-targeting G4 imaging) and identifying (**G4RP-seq** and **G4DP-seq**) G4s in cells.

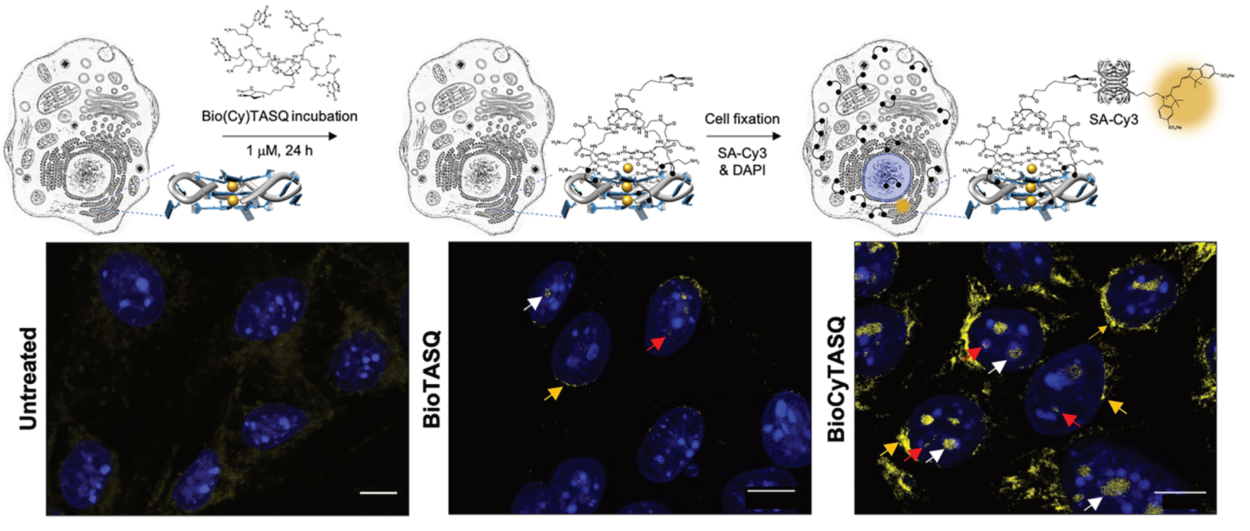


*Figure 19: Schematic representation (and microscopy images) of the **BioTASQ/BioCyTASQ**-based, two-step strategies implemented to label G4s in MCF7 cells with live cell labeling. Images are collected in the range of 518–562 nm for SA-Cy3 (yellow) and <425 nm for DAPI (blue channel); scale bar 10 µm; yellow, white and red arrows indicate G4 foci found in perinuclear regions, nucleoli and nucleoplasm, respectively.*

Of note, both strategies are intended to provide complementary outputs as the former leads to a snapshot of G4-landscapes as captured by chemical fixation while the latter highlights cellular sites where TASQs accumulate in live cells. In both instances, **BioCyTASQ** provided better images than **BioTASQ** thanks to an improved uptake that results in a better signal-to-background ratio. These results, in full agreement with the direct G4-labeling pattern observed with **N-TASQ** (*Chapter 5*), complies with the known G4-richness of ribosomal sites. We also observe that the live-cell treatment shows a similar staining pattern to post-fixation labeling, indicating that the G4-landscape is not drastically redistributed by TASQ treatment (mostly due to their low ability to artificially promote G4-formation (*vide supra*), in line with their smart ligand nature). Moreover, live cell imaging proved to be more accurate, likely due to the fact that cellular G4s are not chemically altered by fixation. Unfortunately, in both cases the signal to background fluorescence level is far from ideal, probably due to the natural presence biotin in cell, which makes this imaging technique less effective than the use of specifically developed TASQs (**N-TASQ** and ^{Tz}**N-TASQ**, *vide infra*).

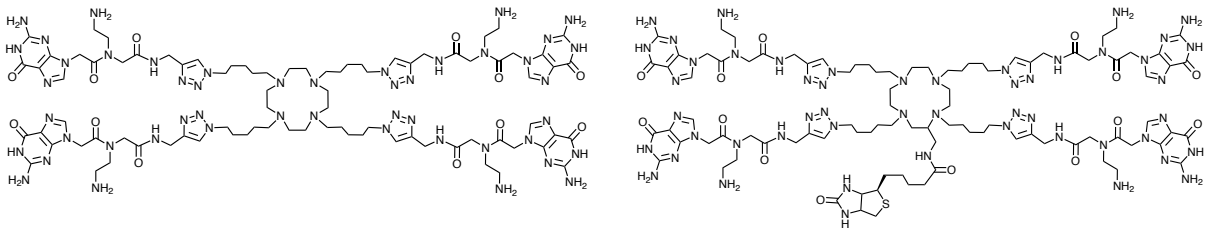
II.6. Conclusion:

A simple chemical modification of the nature of the linker between the central template and the ^{PNA}G arms of the TASQs enabled the development of a new ligand with improved G4-interacting properties. As assessed by FRET-melting, the non-HB-donating/accepting nature of the alkyl linker minimized the undesired interaction between the biotin appendage and the linker, which avoids the internal poisoning observed within the **BioTASQ**. **BioCyTASQ** was also found to display better performances in both fluorescence and cellular pull-down assays. Both TASQs were shown to outperform the well-established **BG4** for isolating G4-DNA *in vitro* and *in vivo*, with **BioCyTASQ** providing a better signal-to-background ratio. The versatility of the biotin appendage was also demonstrated by the development of the pre-targeted G4-imaging protocol, which allows for visualizing G4s by incubation of the TASQs before or after cell fixation. Therefore, we strongly believe that **BioCyTASQ** outperforms **BioTASQ** in all biophysical and cellular essays, which makes it the new benchmark for biomimetic ligands dedicated to the isolation of G4s.

Chapter 3:

TriazoTASQ & BioTriazoTASQ

As described above, we successfully improved the tedious chemical access to **BioTASQ** by designing **BioCyTASQ,** that satisfyingly displays similar (even better) G4-interacting properties. We thus sought to design an even shorter chemical access: inspired by reports in which G4-ligands in general[132], and a synthetic G-quartet in particular[115], were assembled by click chemistry, we decided to investigate the design of new biotinylated TASQs. This was further motivated by the description of possible beneficial effects of the triazole rings on the G- quartet stability (*further discussed below*)[115]. We thus designed two new TASQs: **TriazoTASQ** and its biotinylated counterpart **BioTriazoTASQ**, the former being synthesized to both validate this novel chemical pathway and assess whether the important structural modification of the linkers affects the G4-interacting properties of the resulting TASQ (*Figure 20*)[133].



TriazoTASQ **BioTriazoTASQ**

*Figure 20 : Chemical structures of **TriazoTASQ** and **BioTriazoTASQ**.*

III.1. Click Chemistry: CuAAC

The most popular “click chemistry” reaction is undoubtedly the Cu(I)-catalyzed version of the Huisgen 1,3-dipolar cycloaddition, so called the copper-catalyzed alkyne-azide cycloaddition (CuAAC)[134-135]. This coupling reaction occurs between azides and terminal alkynes in the presence of a copper (I) catalyst and leads to the formation of 1,4-disubstituted 1,2,3-triazoles. Originally performed without a catalyst, the concerted cycloaddition reaction, described by Huisgen in 1963[136], required elevated temperatures or pressures and led to two different products (*Figure 21*). The introduction of a copper(I) catalyst by Tornøe and Medal

in 2001, led to a major improvement in both rate and regioselectivity *via* a stepwise metal-catalyzed reaction[135].

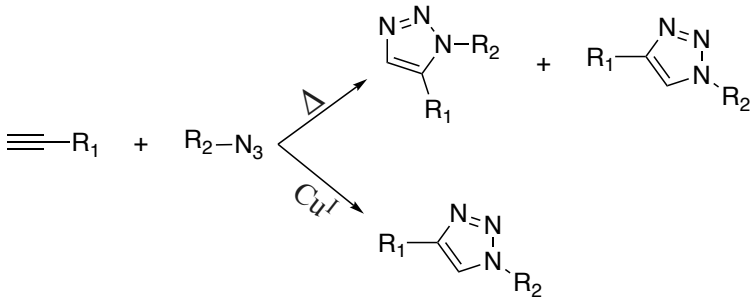


Figure 21 : Two azide-alkyne cycloaddition reactions. Huisgen 1,3 dipolar cycloaddition leading to two products (up), and CuAAC leading to a single isomer[135].

The scope of the CuAAC reaction is well reviewed and many applications have been reported [137-140]. Notably, this approach can be biorthogonal, as click chemistry is suited to biomolecule modification and *in vivo* tagging (*further discussed below*)[141-142]. The reason for this success lies in the practical convenience of click chemistry, being efficient, easy to perform, in simple reaction conditions, with a broad variety of starting materials, solvents, and reagents. Additionally, it is unlikely that the switch of amides to triazoles could disturb the G4-binding conformation of TASQs to G4s since triazoles can be used as amide bond bioisosters[143-144].

We chose this reaction for the assembly of TASQs as CuAAC is completely selective regarding the formation of the 1,4-disubstitued 1,2,3-triazoles, is insensitive to steric hindrance, and is extremely efficient and chemospecific. The pre-catalyst can be a Cu(II) salt (usually $CuSO_4$) together with a reducing agent (usually sodium ascorbate), which leads to the active species Cu(I) *in situ*, in either an aqueous solvent or water/alcohol mixtures. Efforts were invested to decipher the actual mechanism of this reaction (*Figure 22*): a mechanism involving a bimetallic intermediate was first proposed[145], in which the alkynyl moiety was coordinated to one copper center while the azide was coordinated to a second one[134]; another mechanism was suggested in which both azide and alkynyl moieties were attached to the same copper center, with a second copper positioned closely to the alkynyl moiety to positively influence the formation of a metallatricycle. This latter mechanism was recently supported by DFT studies[146]. After cycloaddition and transient formation of a metallotricycle, the intermediate undergoes a reductive elimination that affords a 1,4,5 trisubstituted 1,2,3 triazole metalated

in position 5. The last step consists in a metal-proton exchange to afford a 1,4 disubstituted 1,2,3 triazole[147-148].

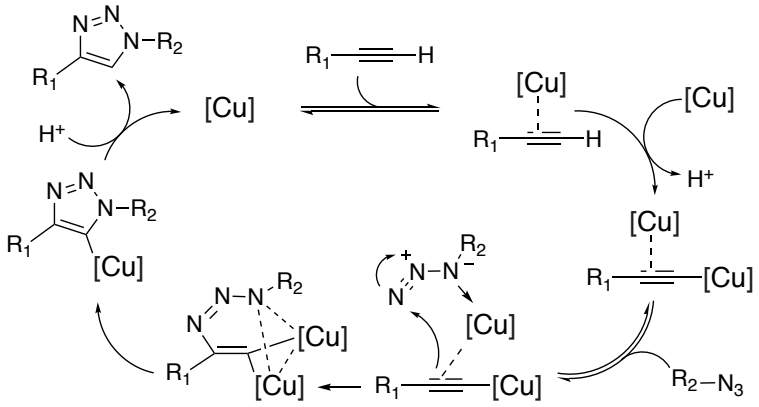
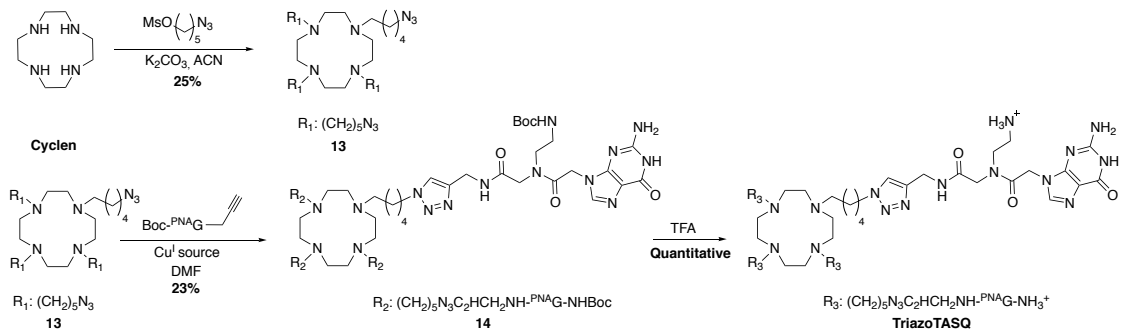


Figure 22: Proposed mechanism for the Cu(I)-catalyzed version of the Huisgen 1,3-dipolar cycloaddition [149].

III.2. TriazoTASQ synthetic pathway:

TriazoTASQ (*Scheme 5*) was initially designed in order to demonstrate that changing classical coupling techniques for CuAAC could afford derivatives in higher yields and in a less time-demanding manner. When designing the retrosynthetic approach for **TriazoTASQ,** the alkyne moiety was introduced directly on the ^{PNA}G carboxylic acid *via* a peptide coupling with propargylamine. The azide function can easily be introduced on the cyclen building block *via* mesylated 5-azidopentan-1-ol, an intermediate that can be obtained from aminopentanol, the starting material for the alkyl spacer. This approach results in the most simple and direct synthetic strategy possible using readily available compounds. It is interesting to note that the spacer, while still having five sp^3 carbons, as for **CyTASQ** and **BioCyTASQ**, will be slightly longer and bulkier due to the presence of the triazole ring.



*Scheme 5: **TriazoTASQ** synthetic pathway*

5-azidopentyl-mesylate, prepared in 2 steps from the commercially available 5-aminopentan-1-ol in 42% chemical yield, was reacted with cyclen to afford the azido-substituted cyclen **13** with 25% yield. The alkynylated ^{PNA}G was prepared from the commercially available Boc^{PNA}G with 78% yield. These two intermediates were then coupled using different sources of copper (*Figure 23*): we firstly used a copper sulfate and sodium ascorbate mixture (the most common Cu(I) source for water-soluble substrates) and assessed the efficiency of the reaction in a DMF/water mixture by monitoring the conversion rate by HPLC-MS. Unfortunately, this classical mixture provided a poor and slow conversion. To improve the reaction efficiency, we decided to use another commercial copper source, the tetrakis(acetonitrile)copper(I) hexafluorophosphate ($(CH_3CN)_4CuPF_6$) with and without the copper ligand TBTA (*tris*[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine), which is known to improve the stability of Cu(I)[150]. This approach resulted in an almost complete conversion of the starting material (by HPLC-MS) in a shorter time (< 2 h).

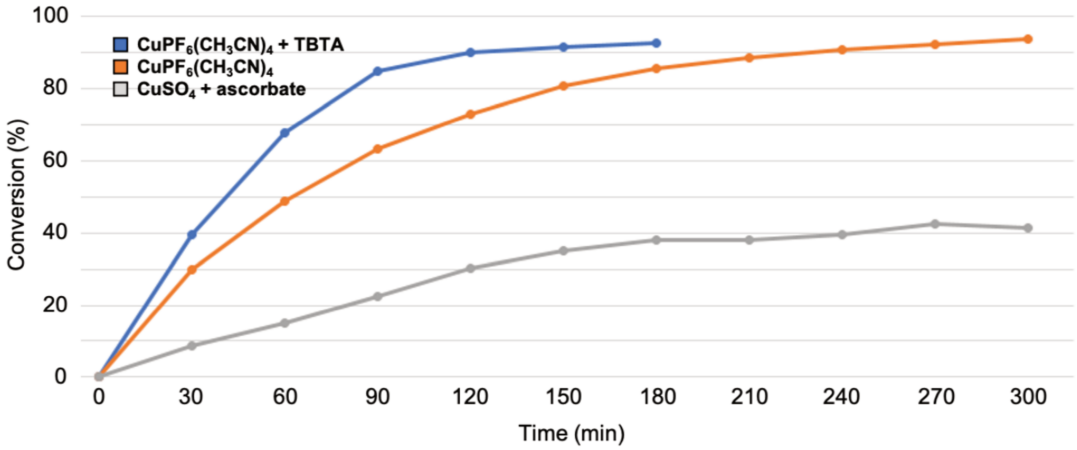


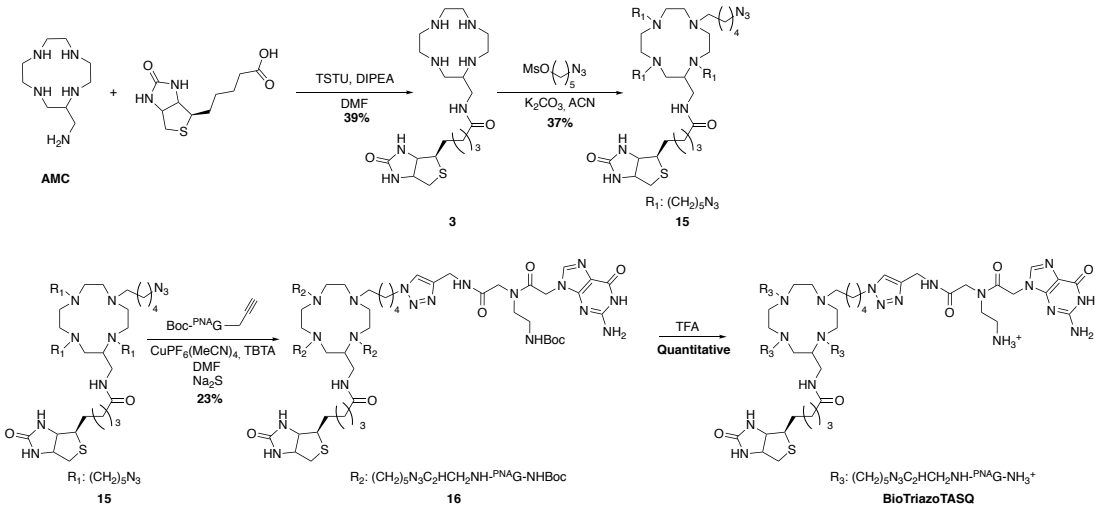
*Figure 23: Kinetics of CuAAC between the functionalized cyclen and the alkynylated PNA-Guanine in the presence of different Cu(I) sources, monitored with HPLC-MS at 260 nm. In blue **Tetrakis**(acetonitrile)copper(I) hexafluorophosphate ($(CH_3CN)_4CuPF_6$), with **TBTA** (tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine), in orange **Tetrakis**(acetonitrile)copper(I) hexafluorophosphate ($(CH_3CN)_4CuPF_6$) and in grey **$CuSO_4$** with **Sodium ascorbate**.*

Of note, it was not possible to perform this reaction with a substoechiometric amount of copper. The cyclen template readily chelates the metal ion, which makes the first molar equivalent of copper unavailable for the coupling reaction. It was thus necessary to add an additional 0.5 equivalents of Cu(I) to initiate the conversion of the starting material. After purification of the main product of the reaction, **14** was isolated as a copper salt (an ICP analysis showed that 2% w/w of Cu were present in the compound, which is consistent with

an equimolar content of Cu per TASQ). We exploited the strong affinity of Cu for sulfur to remove it from the cyclen core, adding four equivalents of sodium sulfide (Na_2S) in water to the solution of **14·Cu^{2+}** in acetonitrile[151]. The demetallation was monitored by HPLC and was completed in less than 10 min. This reaction leads to the formation of a precipitate of CuS which can be easily removed with a simple centrifugation step. Compound **14** was obtained and fully characterized by HPLC and ICP (the final estimated a content of Cu after demetalation was 0.025 %). **TriazoTASQ** was obtained after a last Boc-deprotection step performed with TFA. As above, the biophysical investigations performed with **TriazoTASQ** gave highly satisfying results in terms of G4-affinity and selectivity. On this basis, we decided to launch the synthesis of **BioTriazoTASQ** but, for the sake of clarity, all biophysical results are gathered in §III.4.

III.3. BioTriazoTASQ synthetic pathway:

Thanks to the synthesis of the model molecule **TriazoTASQ,** the synthesis of its biotinylated counterpart **BioTriazoTASQ** was straightforward. The biotinylated **AMC** (see Chapter II) was reacted with 5-azidopentyl mesylate to afford compound **15** in 37% chemical yield. Next, the CuAAC was performed with the alkynylated ^{PNA}G to afford **(Boc)$_4$BioTriazoTASQ·Cu(II)** complex (23% chemical yield), which is subsequently demetalated with Na_2S (quantitative)[152] to provide compound **16**, which is then deprotected with TFA (quantitative) to yield to the final molecule **BioTriazoTASQ** (*Scheme 6*).



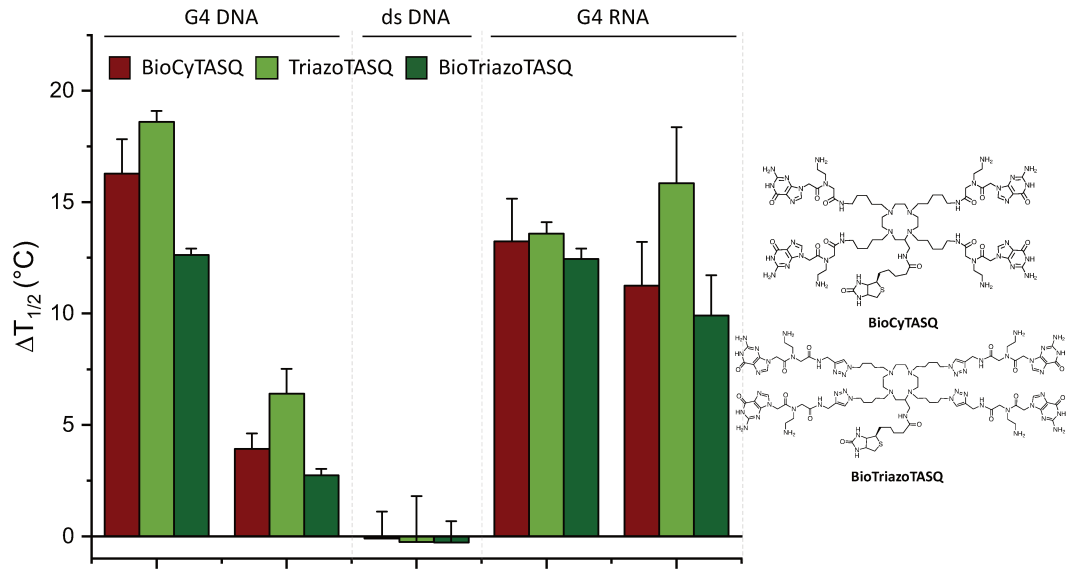
*Scheme 6: **BioTriazoTASQ** synthetic pathway*

The global yield for the synthesis of **BioTriazoTASQ** (2.2%) is lower than that of **BioCyTASQ** (12%), which originates in the low conversion of the alkylation step and in the difficulties to isolate the **(Boc)$_4$BioTriazoTASQ·Cu(II)** complex. For the former, we believe that the problem is in the lower temperature (50 °C *vs.* 70 °C for **BioCyTASQ** synthesis) which is mandatory to avoid degradation of 5-azidopentyl mesylate. For the latter, while the HPLC-MS monitoring showed a complete conversion of the starting material, the isolation of the compound revealed to be much more difficult than the **(Boc)$_4$BioCyTASQ**, for unknown reasons. Despite these lower yields, the G4-interacting properties of both **TriazoTASQ** and **BioTriazoTASQ** were evaluated as above, notably with the aim of investigating the influence of the triazole linkers on their G4-affinity and selectivity.

III.4. The G4-interacting properties of TriazoTASQ and BioTriazoTASQ.

III.4.a. Comparative & competitive FRET-melting experiments

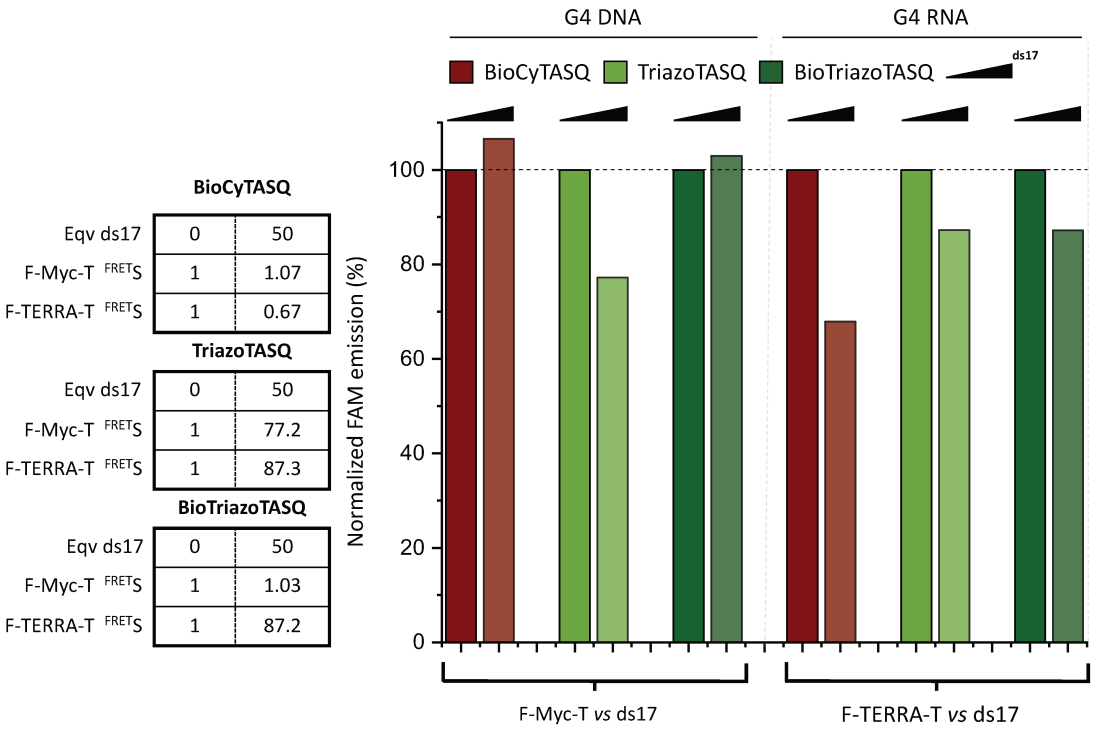
The apparent affinity of **TriazoTASQ** and **BioTriazoTASQ** for both G4-DNA and -RNA was evaluated by FRET-melting assay using the same doubly labeled oligonucleotides than our previous investigations (the DNA F-Myc-T and F21T[129]; the RNA sequence F-TERRA-T and F-NRAS-T; the duplex control F-duplex-T)[83].



	F-Myc-T	F-21-T	F-duplex-T	F-NRAS-T	F-TERRA-T
BioCyTASQ	16.3± 1.5°C	3.9± 0.7°C	-0.1± 1.2°C	13.2± 1.9°C	11.2± 1.9°C
TriazoTASQ	18.6± 0.5°C	6.4± 1.1°C	-0.2± 2.0°C	13.6± 0.5°C	15.8± 2.5°C
BioTriazoTASQ	12.6± 0.3°C	2.7± 0.3°C	-0.2± 0.9°C	12.4± 0.4°C	9.9± 1.8°C

*Figure 24: FRET-melting results ($\Delta T_{1/2}$, °C) for experiments performed with **TriazoTASQ**, **BioCyTASQ** and **BioTriazoTASQ** (1 µM, 5 mol. equiv.) and the G4-DNA, G4-RNA and the control F-duplex-T (0.2 µM), (n=6).*

As seen in *Figure 24*, the results showed that the thermal stabilization imparted by **TriazoTASQ** (expressed as $\Delta T_{1/2}$, in °C) is improved in comparison with the parent compound **CyTASQ** (+23 to +209 %, *Table 1*). These results were thus in line with a possible positive effect of the presence of a ring of four triazoles in the close proximity to the synthetic G-quartet described with RAFT-based TASQ[115]. Both **TriazoTASQ** and **CyTASQ** were found equally selective over duplex-DNA, as demonstrated by the lack of interaction with F-duplex-T ($\Delta T_{1/2}$ < 0.1 °C for both TASQs). This was confirmed by competitive FRET-melting experiments (*Figure 25*) performed with an excess (50 equiv.) of a 17-bp duplex-DNA (ds17), with selectivity factor ^{FRET}S (calculated as $^{FRET}S = \Delta T_{1/2(\text{with competitor})} / \Delta T_{1/(\text{without competitor})}$) highlighting exquisite selectivity: S = 0.77 and 0.87 for **TriazoTASQ** with F-Myc-T and F-VEGF-T respectively.



BioCyTASQ

Eqv ds17	0	50
F-Myc-T ^{FRET}S	1	1.07
F-TERRA-T ^{FRET}S	1	0.67

TriazoTASQ

Eqv ds17	0	50
F-Myc-T ^{FRET}S	1	77.2
F-TERRA-T ^{FRET}S	1	87.3

BioTriazoTASQ

Eqv ds17	0	50
F-Myc-T ^{FRET}S	1	1.03
F-TERRA-T ^{FRET}S	1	87.2

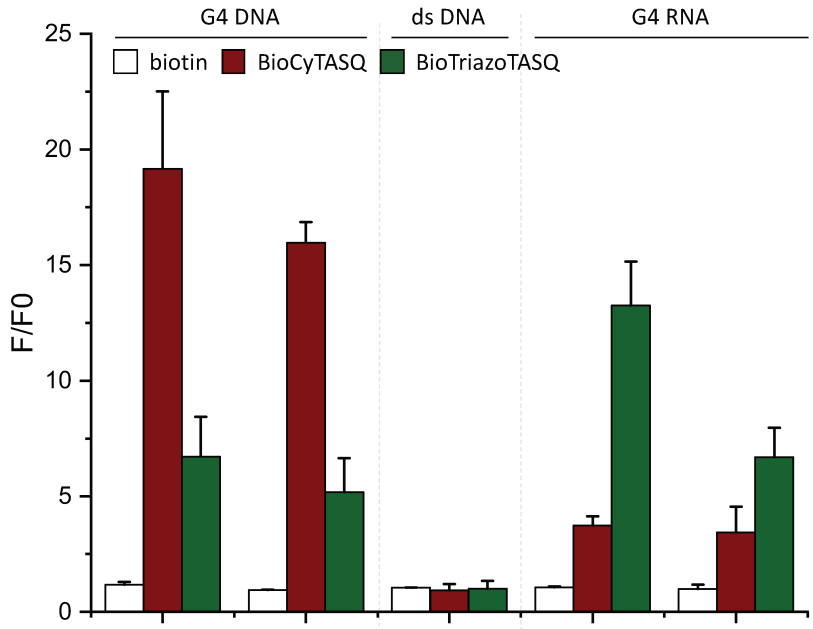
*Figure 25: competitive FRET-melting results, selectivity values (^{FRET}S) calculated for **BioCyTASQ**, **TriazoTASQ** and **BioTriazoTASQ** (n=6).*

The apparent G4-affinity of **BioTriazoTASQ** was also assessed by FRET-melting experiments performed with the same panel of oligonucleotides. It was found that the G4-stabilization imparted by **BioTriazoTASQ** was decreased as compared to the parent compound **BioCyTASQ**, more markedly with G4-DNA (-23 and -31%, respectively) than with G4-RNA (-9 to -11%, *Table 1*). These results indicated that the presence of the triazoles near the synthetic

G-quartet did not improve G4-binding in the biotinylated series. Satisfyingly, the G4-selectivity of **BioTriazoTASQ** remained very high ($\Delta T_{1/2}$ < 0°C with F-duplex-T; ^{FRET}S > 0.87).

III.4.b. Fluorescent G4-pull down assay

In line with the experiments described in §II.5.c, the oligonucleotides used for FRET-melting assays were used here labeled by FAM only (F-Myc, F-22AG, F-duplex, F-NRAS and F-TERRA); these oligonucleotides were incubated (1 µM) for 2 h at room temperature with either **BioCyTASQ** or **BioTriazoTASQ** (10 µM, along with biotin as a control) in the presence of streptavidin-coated magnetic beads. Then, the complexes of nucleic acids, TASQ and beads were magnetically immobilized, the supernatant removed and the FAM-labeled DNA/RNA resuspended in solution after a thermal denaturation step (8 min at 90 °C). The capture efficiency of the two TASQs was quantified by the FAM emission of the resulting solution (normalized to the control performed without TASQ).



	F-Myc	F-22AG	F-duplex	F-NRAS	F-TERRA
biotin	1.2 ± 0.1	0.9 ± 0.1	1.0 ± 0.1	1.0 ± 0.1	0.9 ± 0.2
BioCyTASQ	19.1 ± 3.3	15.9 ± 0.9	0.9 ± 0.3	3.7 ± 0.4	3.4 ± 1.1
BioTriazoTASQ	6.7 ± 1.7	5.2 ± 1.4	1.0 ± 0.3	13.2 ± 1.9	6.7 ± 1.3

*Figure 26: Fluorescence pull-down results (FAM fluorescence enhancement, fold- change) for experiments performed with biotin, **BioCyTASQ** and **BioTriazoTASQ** (10 µM, 10 mol. equiv.) and the G4-DNA F-Myc and F-22AG, the G4-RNA F-NRAS and F-TERRA, and the control F-duplex (1 µM) (n=6).*

As seen in *Figure 26*, both **BioCyTASQ** and **BioTriazoTASQ** were found to efficiently capture G4s although with a noticeable discrepancy, as **BioCyTASQ** was found more efficient with G4-DNA (19.1 ± 3.3 and 15.9 ± 0.9-fold enrichment with F-Myc and F-22AG, respectively, *vs.* 6.7 ± 1.7 and 5.2 ± 1.4-fold for **BioTriazoTASQ**), while **BioTriazoTASQ** was more efficient with G4-RNA (*ca*. 13.2 ± 1.9 and 6.7 ± 1.3-fold enrichment with F-NRAS and F-TERRA, respectively, *vs. ca*. 3.7 ± 0.3 and 3.4 ± 1.1- fold for **BioCyTASQ**). One hypothesis to explain the preferential interaction of **BioTriazoTASQ** with G4-RNA could be that the stacking interaction between the triazole ring and the intramolecular G-quartet brings them closer and pushes the connecting linkers outward. This would make the TASQ's quartet more sterically demanding, which thus could drive the interaction towards G4-RNA, as their external G-quartet is more accessible than that of the G4-DNA.

III.4.c. Molecular dynamics:

To investigate this, a series of preliminary molecular dynamics (MD) simulations were undertaken by Dr. N. Chéron (ENS Paris) with **CyTASQ** and **TriazoTASQ** (*Figure 27*). These calculations did not confirm the coplanar assembly of the triazoles but indicated that:

- **TriazoTASQ** is more sterically demanding than **CyTASQ** (solvent-accessible surface area = 21.4 *vs.* 18.5 nm^2);
- The distance between the cyclen ring (defined as the plane formed by the four nitrogen atoms) and the guanines (defined as the plane formed by the four oxygen atoms) was found to be of 6.0 Å for **CyTASQ**, while it was 7.0 Å for **TriazoTASQ**;
- **TriazoTASQ** was found to be much more flexible with 943 clusters found *vs.* 50 for **CyTASQ** only;
- The nature of the guanine arms does influence the steric hindrance of external G-quartet, that of **CyTASQ** being hindered by surrounding amide connectors while that of **TriazoTASQ** being more freely accessible.

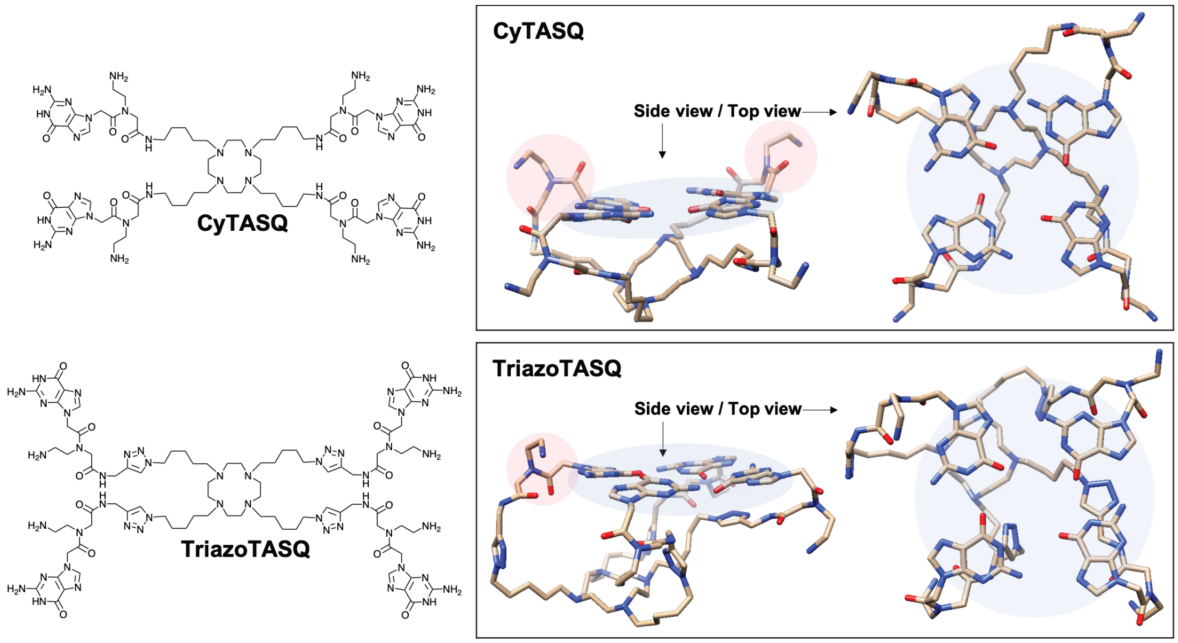


Figure 27: Minimized structures of CyTASQ and TriazoTASQ, highlighting the difference of accessibility to the G-quartet.

Moreover, the accessibility of the two G-quartets was assessed in the following way: *i)* we defined a virtual atom positioned 3 Å above the center of the four oxygen of the guanines; *ii)* we computed the distance between this virtual atom and the oxygens from the amides closest to the guanine; and *iii)* we computed the probability that this distance is below a given value d_0, *i.e.,* the probability that the arm could form a barrier. We found that such a barrier exists more with **CyTASQ** (4.03% of the simulation) than with **TriazoTASQ** (3.87%). Altogether, even if preliminary, these results concur in demonstrating that **TriazoTASQ** displays a more accessible G-quartet than **CyTASQ**, which thus could justify the DNA *vs.* RNA selectivity observed above.

III.4.d. G4-chemo-precipitation monitored by qPCR:

To investigate the ability of TASQ to pull G4s down in more biologically relevant conditions, we performed capture experiments with G4s embedded in longer DNA sequences[153]. As for the *in vitro* fluorescent pull-down, a 97-nucleotide long G4-forming oligonucleotide (d[N_{41}-$G_3(CAG_3)_3$-N_{38}) (4 µM) was incubated for 2 h at room temperature with biotinylated TASQs (10 µM) in presence of streptavidin-coated magnetic beads in G4RP buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, RNase inhibitor (Roche), homebrew protease inhibitor cocktail). The G4/TASQ/beads assemblies were magnetically precipitated,

the supernatant removed and the captured G4s released by heating the TBS 1X solution for 8 minutes at 90°C. The supernatant containing the oligonucleotide was distributed in a 96 well-plate with the *ad hoc* primer and iTaq™ Universal SYBR® Green Supermix in aqueous KCl (380mM) buffer. The polymerase elongated the oligonucleotide through 33 cycles and the SYBR® Green fluorescence was monitored (for all sample the fluorescence was normalized at 2200 after the first cycle). This technique thus relies on the ability of biotinylated TASQ to selectively precipitate G4s which is then detected by polymerase amplification. Compared to the *in vitro* fluorescent pull-down this technique presents two major advantages: i) it is not necessary to use FAM-functionated oligonucleotides, which thus discards any issues or biases related to the presence of a bulky fluorophore in the close proximity of the G4-quartet; and ii) these conditions are closer to *in vivo*-like conditions, as G4-forming sequences are not isolated but included in longer (97-nt long) DNA fragments.

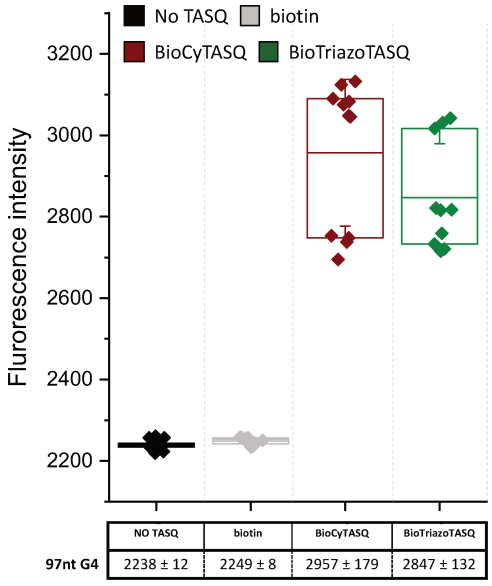


*Figure 28: qPCR pull-down results (SYBR green fluorescence intensity) for experiments performed without molecule (control) or with biotin, **BioCyTASQ** and **BioTriazoTASQ** (40 µM, 10 mol. equiv.) and the G4-strand (4 µM) (n=9).*

BioTriazoTASQ and **BioCyTASQ** (10 µM) were thus incubated with the aforementioned 97 nucleotide long G4-forming sequence, along with biotin as control (10 µM); as seen in *Figure 28* and *Table 1*, the fluorescence signals collected after 33 cycles demonstrate a good amplification with both TASQs, while no modification of fluorescence was monitored with the biotin control. The results clearly show the efficiency of both **BioCyTASQ** and **BioTriazoTASQ** to capture G4s in biologically-relevant conditions.

Table 1 : Summary of the **CyTASQ**, **BioCyTASQ**, ***TriazoTASQ*** and **BioTriazoTASQ** biophysical properties

FRET-melting assay ($\Delta T_{1/2}$, °C)					
	G4-DNA		**duplex-DNA**	**G4-RNA**	
	F-Myc-T	F21T	F-duplex-T	F-NRAS-T	F-TERRA-T
CyTASQ	8.7 ± 0.7	5.2 ± 0.4	0.1 ± 0.1	4.4 ± 0.4	10.0 ± 0.1
TriazoTASQ	18.6 ± 0.5	6.4 ± 1.1	-0.2 ± 2.0	13.6 ± 0.5	15.8 ± 2.5
BioCyTASQ	16.3 ± 1.5	3.9 ± 0.7	-0.1 ± 1.2	13.2 ± 1.9	11.2 ± 1.9
BioTriazoTASQ	12.6 ± 0.3	2.7 ± 0.3	-0.2 ± 0.9	12.4 ± 0.4	9.9 ± 1.8
Fluorescence pull-down assay (Normalized change in FAM fluorescence intensity)					
	G4-DNA		**duplex-DNA**	**G4-RNA**	
	F-Myc	F22AG	F-duplex	F-NRAS	F-TERRA
Biotin	1.2 ± 0.1	0.9 ± 0.1	1.0 ± 0.1	1.0 ± 0.1	0.9 ± 0.2
BioCyTASQ	19.1 ± 3.3	15.9 ± 0.9	0.9 ± 0.3	3.7 ± 0.4	3.4 ± 1.1
BioTriazoTASQ	6.7 ± 1.7	5.2 ± 1.4	1.0 ± 0.3	13.2 ± 1.9	6.7± 1.3

qPCR pull-down assay (SYBRGreen fluorescence intensity)			
No TASQ	**biotin**	**BioCyTASQ**	**BioTriazoTASQ**
2239 ± 12	2249 ± 8	2957 ± 179	2847 ± 132

III.4.e. Pre-targeted G4 imaging

As we showed above, biotinylated TASQs can be used to track G4s in human cells thanks to a pre-targeted G4-imaging protocol using the fluorescently-labeled streptavidin Cy3-SA. To go a step further, we decided to implement this protocol to demonstrate that G4s can act as roadblocks to polymerases, thereby creating a crisis situation recognized as a DNA damage (*schematically represented in Figure 29*)[66]. Examples of the cellular colocalization of G4 and DNA damage sites are not only sparse in the literature but mostly qualitative in nature, associating DNA damage and G4-formation, but failing in providing a quantitative analysis. Neidle *et al.* first reported on the colocalization of the G4-specific antibody[154] **BG4** and an antibody raised against the tumor suppressor p53 binding protein 1 (53BP1)[155], a protein that is recruited to repair double-strand breaks (DSBs). Beyond this G4-immunodetection

approach, the use of a G4-probe and a DNA damage marker are even sparser: the sole example is a qualitative study involving a combination of a naphthalene diimide[156] (NDI, a known fluorescent G4-ligand) and an antibody raised against γH2AX, a phosphorylated histone that signals DSBs[157-158].

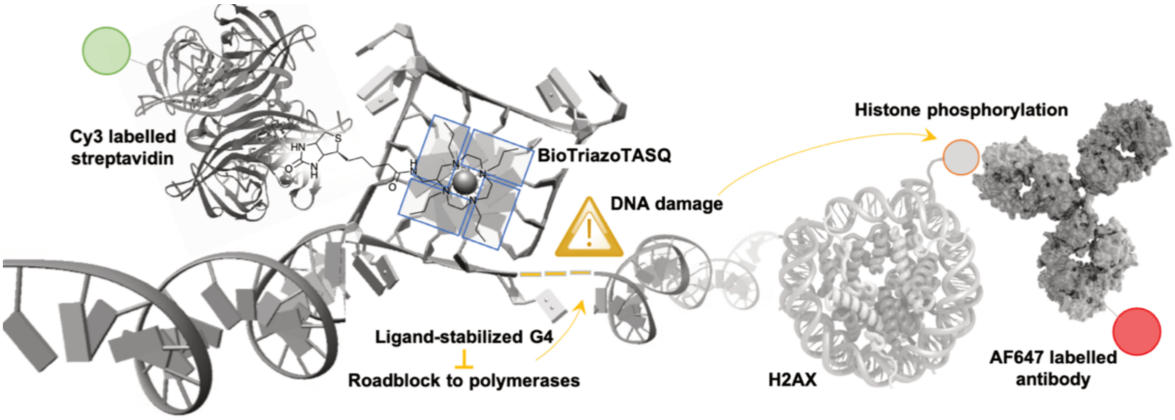


Figure 29: Schematic representation of the TASQ-stabilized G4-mediated DNA damage and related signaling pathway used to colocalize G4 and DNA damage sites (of note, the primary, anti-γH2AX antibody is removed for clarity).

In this study, we focused on the development of a quantitative analysis using **TASQ**-based pre-targeting G4-imaging combined with the anti-γH2AX antibody, which could uniquely allow for the quantification of the level of DNA damage associated to the G4-mediated stalling of the polymerase. To this end, we treated MCF7 cells with either **BioCyTASQ** or **BioTriazoTASQ** (1 µM, 24 h; IC_{50} >100 µM) before fixation and permeabilization with cold MeOH (10 min) followed by the sequential incubation (in a dark chamber) with SA-Cy3 (1 µg/mL, 1 h), γH2AX-specific antibody (1/2000, 2 h), fluorophore-labeled secondary antibody (IgG-AF647, 1/500, 45 min) and DAPI (1 µg/mL, 5 min). The collected images (*see examples in Figure 30*) were quantified for both G4-sites (TASQ, that is, SA-Cy3 *foci*) and DNA damage sites (that is γH2AX *foci*). These results first indicate that the number of G4-sites detected by TASQ (*ca*. 30 *foci*/nucleus) is far lower than that detected by BG4 (*ca*. 400 *foci*/nucleus), and second, that live incubation of MCF7 cells with a subtoxic concentration of TASQs triggered a slight induction of DNA damage only, quantified as the average number of γH2AX *foci* per cell (*ca*. 5.3 *foci*/nucleus for TASQ-treated cells, *vs.* 4.0 *foci*/nucleus for untreated cells, >160 cells/conditions (n = 3)).

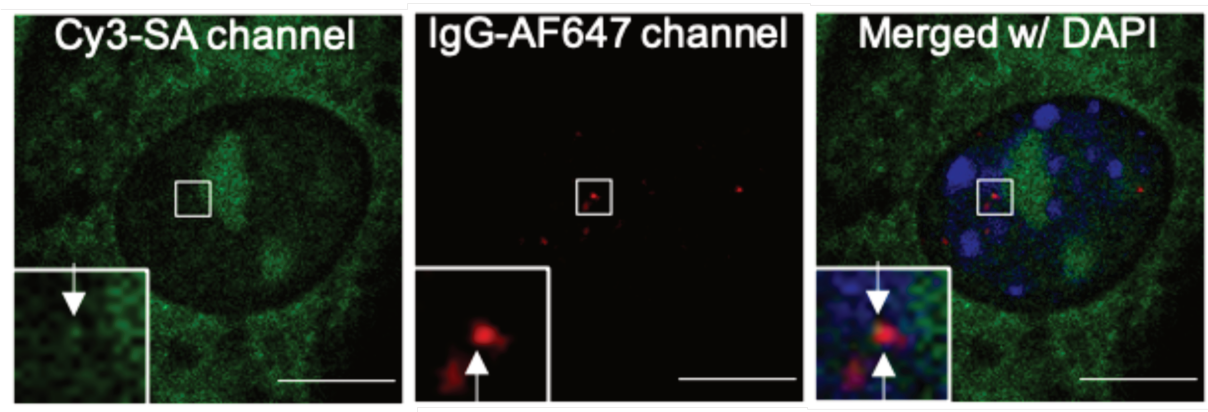


*Figure 30: High-resolution optical images of MCF7 cells treated with **BioTriazoTASQ** (1 µM, 24 h), tagged with Cy3-SA (green) after cell fixation (scale bar = 10 µm), which is followed by immunodetection of DNA damage foci (γH2AX, red) and chromatin staining by DAPI (blue). Inset and arrows highlight of the nucleoplasmic sites where both DSBs and TASQ colocalize.*

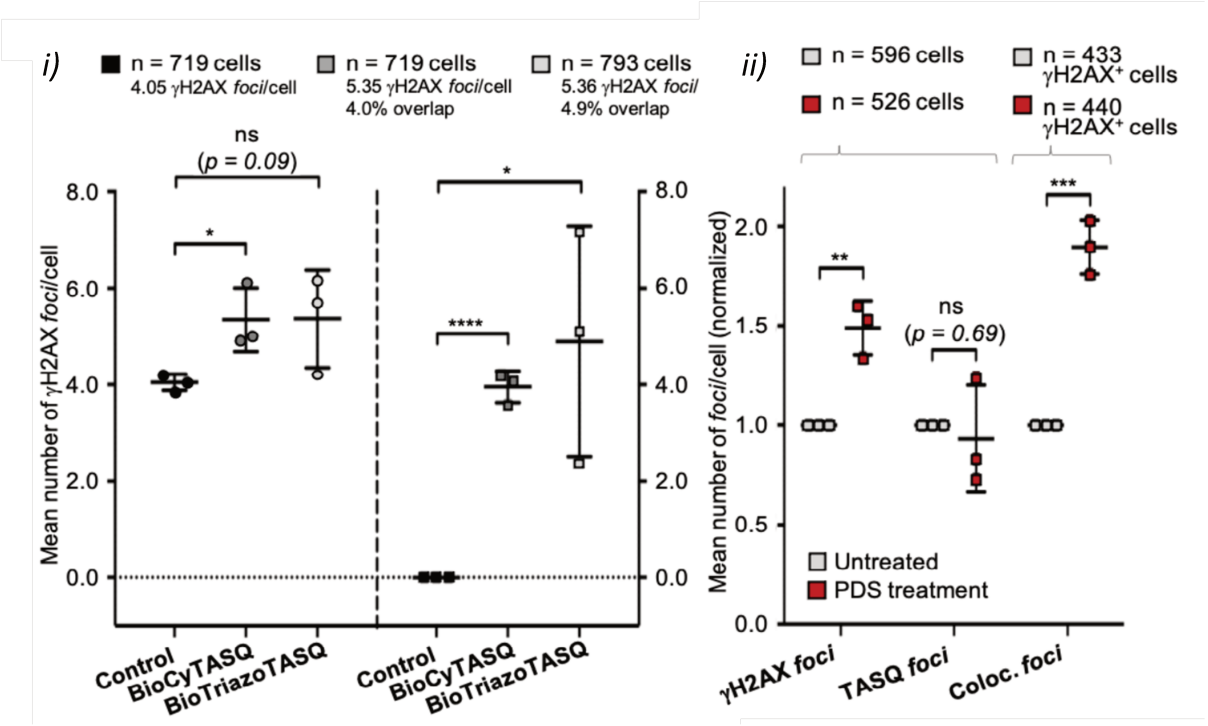


*Figure 31: i) Quantification of γH2AX foci/cell, with or without TASQ treatment, and of the overlaps between green (TASQ) and red signals (γH2AX). ii) Modulation of the γH2AX, **BioTriazoTASQ** and common foci upon incubation of MCF7 cells by PDS (5 µM, 2 h) and G4-labelling (n =3).*

In terms of colocalization, it was found that 4.0 and 4.9% of **BioCyTASQ** and **BioTriazoTASQ** sites, respectively, colocalize with γH2AX *foci* (*Figure 31*), a value higher than the values obtained by Neidle *et al*. with BG4 (1.4%), which likely results from the far lower number of G4-sites detected by TASQ as compared **BG4** (*ca*. 23-fold less) that has to be compared to the 4-fold difference (only) in damage site count (*ca*. 5 γH2AX *vs.* 20 53BP1 *foci*/nucleus. In order to increase the number of DNA damage sites, cells were pre-incubated with the G4-stabilizer

PDS (5 µM, 6 h) known to induce DNA damage. Pre-treatment with an efficient G4-ligands resulted in a raise in the number of γH2AX *foci* (1.5-fold), which originates in the ability of PDS to stabilize G4s in cells, and thus foster G4-mediated DNA damage. To avoid competition between the two ligands (PDS and TASQ) for the cellular G4-targets, **BioTriazoTASQ** was used here as a post-fixation agent (10 µM, overnight): despite a slight decrease of the number of TASQ *foci* (0.9-fold), the number of common TASQ/γH2AX *foci* significantly increased (1.9-fold), which indicates both the suitability of this approach for the study of G4-targeting DNA damaging agents and the versatility of TASQ that can be used either in live cells or after cell fixation. Altogether, these results provided a direct demonstration that ligand-stabilized G4-sites can indeed colocalize with DSBs in cells.

III.5. Conclusion:

The design of **BioTriazoTASQ** aimed at both simplifying the chemical access to biotinylated TASQ but also increasing the G4-binding properties of **BioCyTASQ** by the addition of a triazole ring in the close vicinity of the TASQ's quartet. Unfortunately, the chemical accessibility of both **TriazoTASQ** (the model compound to validate the synthetic scheme) and **BioTriazoTASQ** (that target biotinylated TASQ) turns to be less efficient than their respective parent compounds **CyTASQ** and **BioCyTASQ**. However, and quite satisfyingly, **BioTriazoTASQ** was found to outperform **BioCyTASQ** in terms of RNA G4-capture ability, while retaining its exquisite G4-selectivity. We tried to rationalize these results by molecular dynamics and the collected results show that **TriazoTASQ** is more sterically demanding than **CyTASQ**, which could explain the better G4-RNA interaction. The use of this new molecular bait has now to be validated in G4RP-seq and G4DP-seq protocols. A direct comparison of **BioCyTASQ** and **BioTriazoTASQ** in both isolation protocols aims at validating their unique features, using **BioCyTASQ** for G4-DNA precipitation and **BioTriazoTASQ** for G4-RNA precipitation, maximizing the probes potential.

Chapter 4:

MultiTASQ & azidoMultiTASQ (^{Az}MultiTASQ)

In the previous chapters, we have investigated the role of the linker between the ^{PNA}G arm and the template of TASQs on their overall G4-interacting properties, replacing the original amide linker (**BioTASQ**) by an aliphatic one (**BioCyTASQ**), then introducing a triazole ring (**BioTriazoTASQ**) to allow for a more straightforward synthesis (by CuAAC). We now focus on the modification of the biotin handle: the good performance of **BioCyTASQ** inspired the design of two new ligands, **MultiTASQ** and **azidoMultiTASQ** (**^{Az}MultiTASQ**) in which biotin is replaced by an alkyne or an azide moiety, respectively (these two molecules were patented)[159]. While biotin has proven useful thanks to its strong and selective interaction between the avidin and good level of versatility, it has also some drawbacks as it is sterically demanding and it might divert target engagement in cells due to its ability to create/accept HBs. More importantly, its natural presence in cells results in a poor signal-to-background ratio during pre-targeted G4-imaging and leads to poor quality images.

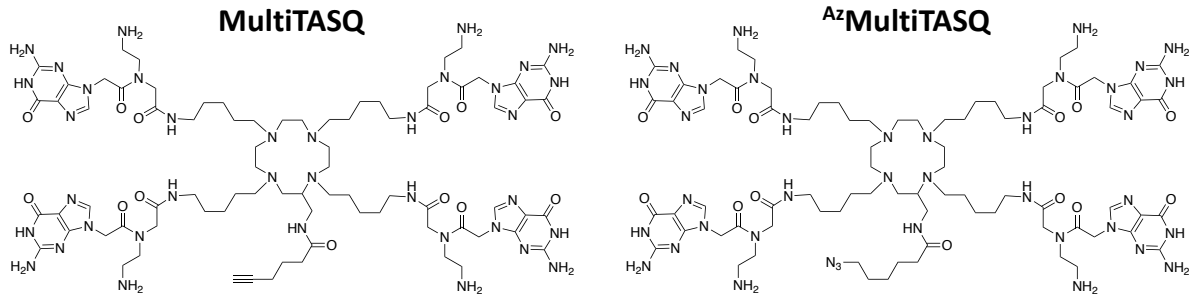


*Figure 32: Chemical structure of **MultiTASQ** and **azidoMultiTASQ**.*

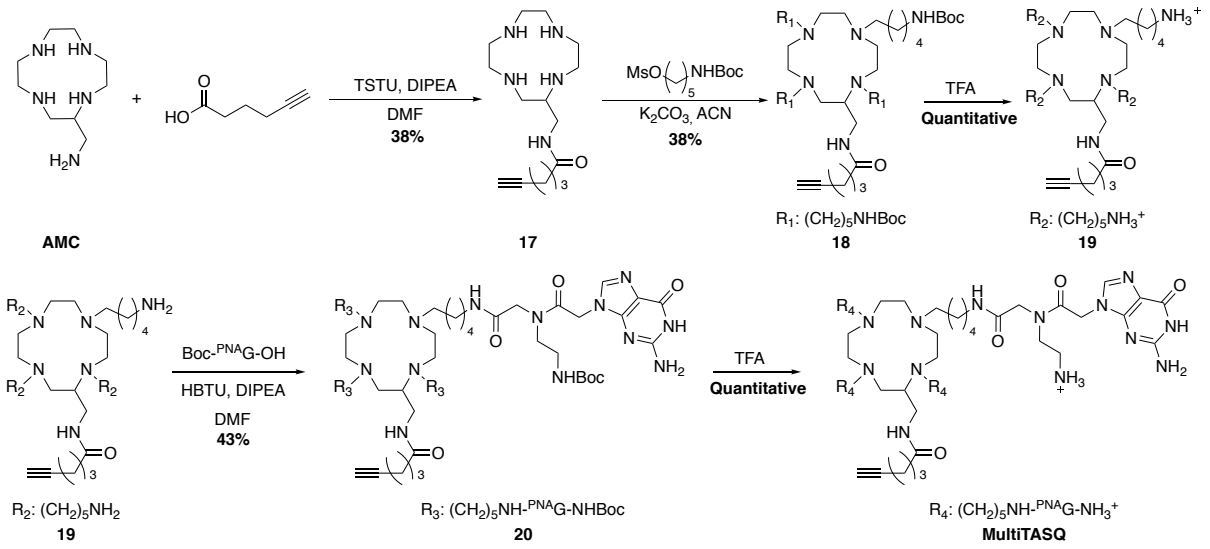
We thus focused on designing probes including two molecular handles that would not interact with their environment (no HB, small size) but could react with great specificity with a wide variety of chemical moieties to achieve specific function. Bioorthogonal reactions, a small subset of the so-called click reactions, can efficiently fulfill our needs, as such reactions are very efficient (even in high dilution), highly chemospecific and (for some of them) can be performed *in vivo*. After a careful study of the variety of reactions available to achieve our goals, CuAAC[134-135] and Strain-Promoted Alkyne-Azide Cycloaddition (SPAAC)[141-142] appeared

to fulfill our requirements. We thus set out to find if **MultiTASQ** and **^{Az}MultiTASQ** (*Figure 32*) could be functionalized chemospecifically *in cella* before or after cell fixation. While **BioTASQ** and **BioCyTASQ** were specifically designed to isolate G4s by affinity capture, and were later exploited for G4-visualization, we envision that these two new probes, **MultiTASQ** and **^{Az}MultiTASQ**, could perform a wider array of tasks thanks to the versatility of the alkyne/azide moiety.

In the previous chapter, we described the CuAAC reactions and conditions; we intend here to further exploit its features (quick, reliable and efficient) in a bioorthogonal manner, *i.e.*, the reaction being performed once the ligand is in its cellular binding site (the so-called *in situ* click chemistry). We also envisioned SPAAC as it is also quick, reliable and efficient, and can be performed without a metal catalyst using strained alkynes such as dibenzocyclooctyne (DBCO) or bicyclo[6.1.0]-nonyne (BCN). While the introduction of a very large moiety such as DBCO or BCN directly on the TASQ structure must be avoided (as it may negatively impact the G4-interacting properties of the resulting TASQ), the addition of the small azide group provides an elegant approach towards SPAAC-compatible TASQs. The relevance of this approach is further bolstered by the commercial availability of many strained alkyne partners (fluorophores, biotin derivatives), which makes SPAAC-compatible TASQs of high interest as potentially usable in living cells.

IV.1. The synthesis of MultiTASQ

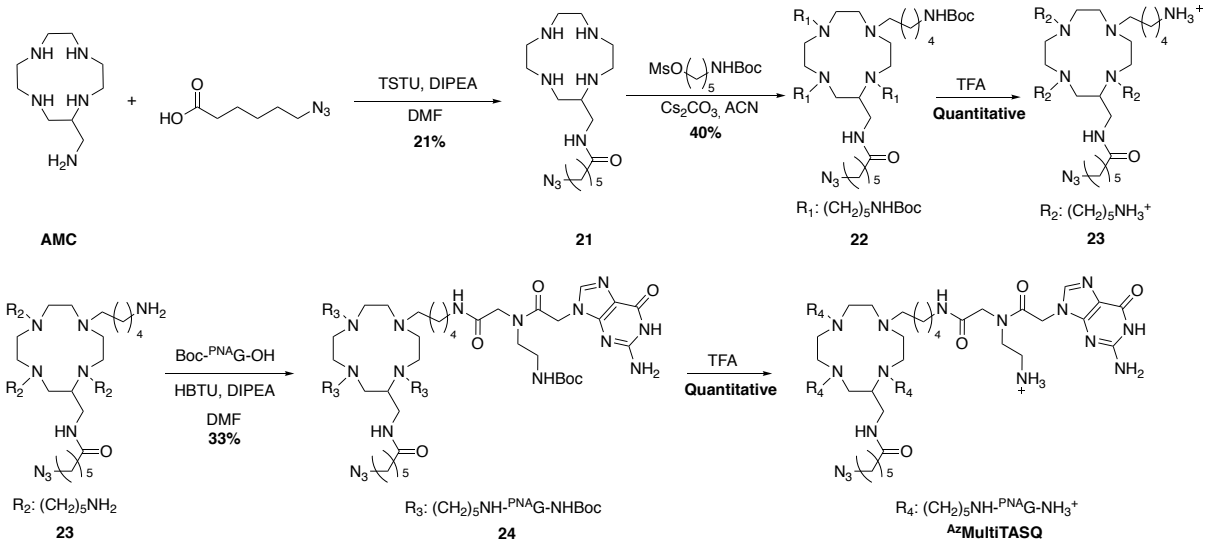
The synthesis of **MultiTASQ** (*Scheme 7*) followed the classical strategy: **AMC** was coupled with 5-hexynoic acid to obtain the derivatized macrocycle **17** (38% chemical yield) using TSTU as coupling reagent and DIPEA as a base. The alkynylated **AMC** was thus reacted with an excess of 5-(Boc-amino)pentylmethanesulfonate in acetonitrile in the presence of K_2CO_3 to provide compound **18** with 38% chemical yield. After a Boc-deprotection step (TFA, quantitative), compound **19** was coupled to Boc-^{PNA}G monomer (4.4 mol. equiv.) in the presence of HBTU and DIPEA, to lead to Boc-protected **MultiTASQ** in 43% chemical yield. The final **MultiTASQ** was obtained after a last Boc-deprotection step (TFA, quantitative), with an overall chemical yield of 6% (over 5 steps).



*Scheme 7: **MultiTASQ** synthetic pathway*

IV.2. The synthesis of azidoMultiTASQ

In a similar approach, the synthesis of **^{Az}MultiTASQ** (*Scheme 8)* started with the functionalization of the **AMC** with 6-azidohexanoic acid, using TSTU as coupling reagent and DIPEA as a base (21% chemical yield). This derivative was then reacted with 5-(Boc-amino)pentylmethanesulfonate in acetonitrile in the presence of Cs_2CO_3 to lead to compound **22** in 40% chemical yield. After a deprotection step (TFA, quantitative), compound **23** was coupled with Boc-^{PNA}G monomer (4.4 mol. equiv.) to obtain the Boc-protected **^{Az}MultiTASQ** in 33% chemical yield. The final **^{Az}MultiTASQ** was obtained after a last deprotection step (TFA, quantitative), with an overall chemical yield of 3% (over 5 steps).

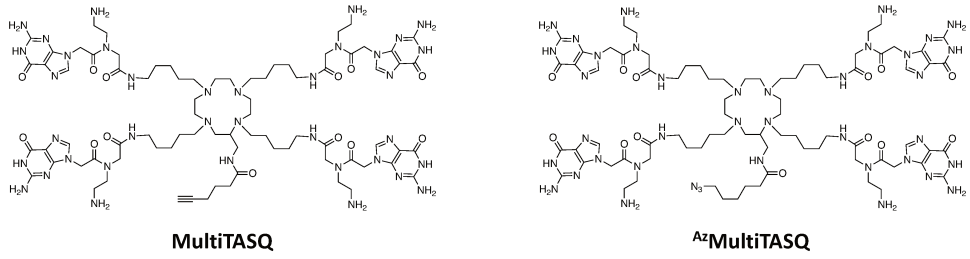


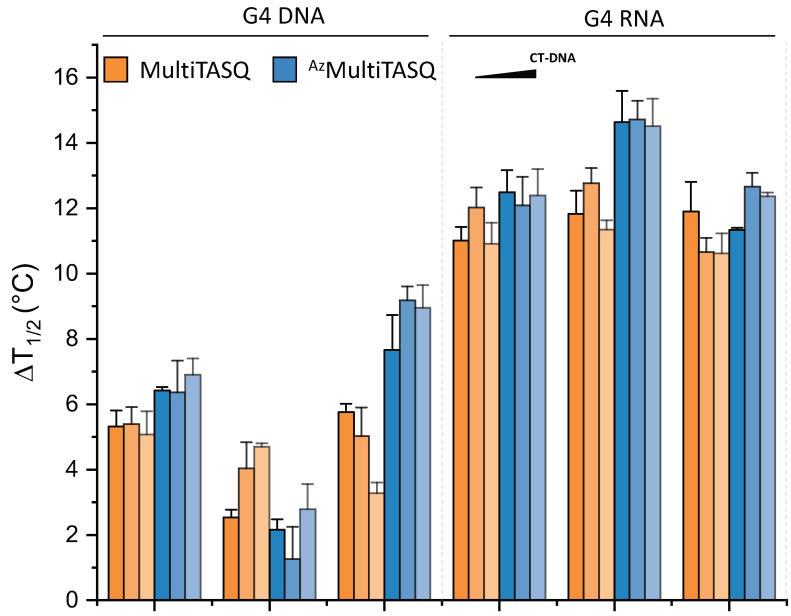
*Scheme 8: ^{Az}**MultiTASQ** synthetic pathway*

IV.3. The G4-interacting properties of MultiTASQ and AzMultiTASQ.

IV.3.a. Comparative & competitive FRET Melting experiments

The apparent affinity of **MutiTASQ** and **AzMultiTASQ** for both G4-DNA and -RNA was evaluated by FRET-melting assay[129], as above, using biologically relevant G4-forming sequences: the G-rich DNA sequences found in the promoter regions of Myc gene (F-Myc-T) and at the human telomeres (F21T); the oncogene promoter c-kit2 (F-kit2-T); the human telomeric transcript (F-TERRA-T) and the 5'-UTR of the mRNA coding for VEGF (F-VEGF-T). The selectivity of both ligands was also investigated by competitive experiments performed in the presence of an excess of unlabeled calf thymus DNA (CT-DNA, 15 to 50 mol. equiv.).





	F-Myc-T	F-21-T	F-kit2-T	F-TERRA-T	F-NRAS-T	F-VEGF-T
MultiTASQ	5.3±0.5°C	2.5±0.2°C	5.8±0.2°C	11.0±0.4°C	11.8±0.7°C	11.9±0.9°C
AzMultiTASQ	6.4±0.2°C	2.2±0.3°C	7.7±1.1°C	12.5±0.4°C	14.6±1.0°C	11.3±0.6°C

MultiTASQ

Eqv CT-DNA	15	50
F-Myc-T ^{FRET}S	1.01	0.95
F-21-T ^{FRET}S	1.60	1.85
F-kit2-T ^{FRET}S	0.87	0.57
F-TERRA-T ^{FRET}S	1.09	0.99
F-NRAS-T ^{FRET}S	1.07	0.95
F-VEGF-T ^{FRET}S	0.90	0.89

AzMultiTASQ

Eqv CT-DNA	15	50
F-Myc-T ^{FRET}S	0.99	1.08
F-21-T ^{FRET}S	0.58	1.29
F-kit2-T ^{FRET}S	1.20	1.17
F-TERRA-T ^{FRET}S	1.02	0.99
F-NRAS-T ^{FRET}S	1.00	0.99
F-VEGF-T ^{FRET}S	1.12	1.09

*Figure 33: FRET-melting results ($\Delta T_{1/2}$, °C) for experiments performed with **MultiTASQ** and **AzMultiTASQ** (1 µM, 5 mol. equiv.) and the G4-DNA F-Myc-T, F21T and F-kit2-T, the G4-RNA F-TERRA-T, F-NRAS-T and F-VEGF-T with increasing concentration of CT-DNA (15 and 50 equivalents). Selectivity values (^{FRET}S) calculated for both **MultiTASQ** and **AzMultiTASQ**. (n=3).*

The $\Delta T_{1/2}$ values obtained with both **MultiTASQ** and **^{Az}MultiTASQ** indicate their good (for DNA) to very good (for RNA) G4-stabilizing properties**.** With F-Myc-T and F-21-T, **MultiTASQ** and **^{Az}MultiTASQ** highlighted very similar stabilization temperature (5.3 ± 0.5 and 6.4 ± 0.2 °C with F-Myc-T, and 2.5 ± 0.2 and 2.2 ± 0.3 °C with F-21-T, for **MultiTASQ** and **^{Az}MultiTASQ**, respectively), while **^{Az}MultiTASQ** showed a better performance with F-kit2-T as compared to **MultiTASQ** (7.7 ± 1.1 *vs.* 5.8 ± 0.2 °C). G4-RNAs showed globally higher melting temperatures, but with no significant difference between **MultiTASQ** and **^{Az}MultiTASQ** could be noticed (F-TERRA-T: 11.0 ± 0.4 and 12.4 ± 0.4 °C; F-VEGF-T: 11.9 ± 0.9 and 11.3 ± 0.6 °C; F-NRAS-T: 11.8 ± 0.7 °C and 14.6 ± 1.1 °C for **MultiTASQ** and **^{Az}MultiTASQ**, respectively). These excellent and comparable G4-interacting properties confirme that the nature of the appendage (alkyne *vs.* azide) does not influence their G4-interacting properties. The results obtained in the presence of an excess of CT-DNA highlight that the two ligands exhibit excellent G4-selectivity, with ^{FRET}S values > 0.57 for **MultiTASQ** with G4-DNA, > 0.89 for **MultiTASQ** with G4-RNA and > 0.57 for **^{Az}MultiTASQ** with G4-DNA, > 0.99 for **^{Az}MultiTASQ** with G4-DNA (see *Figure 33*).

IV.3.b. Apparent K_D measurements

To further investigate the G4-binding abilities of TASQs, we adapted a method developed by the group of S. Balasubramaniam to calculate apparent constant of dissociation ($^{APP}K_D$)[160]. This binding assay is based on the fluorescence quenching of a Cy5-functionalized G4 (Cy5-Myc) upon addition of increasing amounts of a G4-ligand[161] (schematically represented in *Figure 34*). The increasing amounts of ligand trigger the fluorescence quenching of the cyanine in a dose-response fashion, thus enabling the calculation of an apparent constant of dissociation.

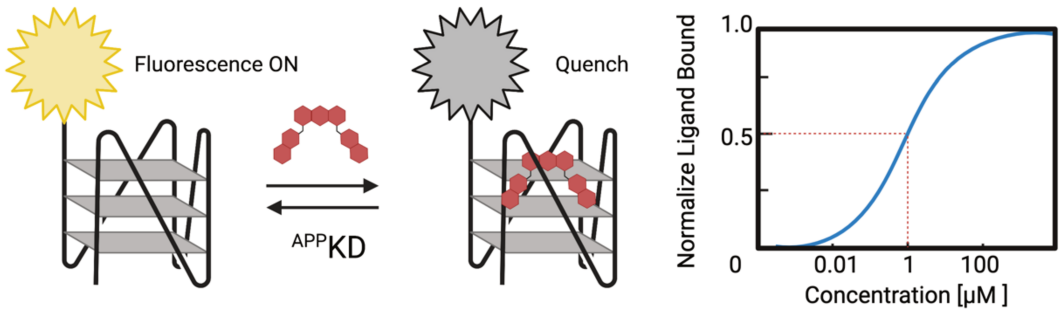


Figure 34:Schematic representation of the fluorescence quench by a G4-ligand with it dose-response sigmoidal curve.

To a solution of folded Cy5-Myc in 50 mM TrisHCl, pH 7.2, 150 mM KCl, 0.05 v/v % Triton X-100 were added various concentrations (serial dilution from 10 to 10^{-5} µM) of four ligands including the firmly established G4-stacker **PhenDC3** (as control) along with **BioCyTASQ**, **MultiTASQ** and **^{Az}MultiTASQ**. After 1-h incubation at room temperature (in the dark), the fluorescence signal of the Cy5 was measured and the collected values normalized from 0 to 100. The values were fitted (sigmoidal fit) and the concentration corresponding to 50% fluorescence intensity was used to define the $^{app}K_D$ values.

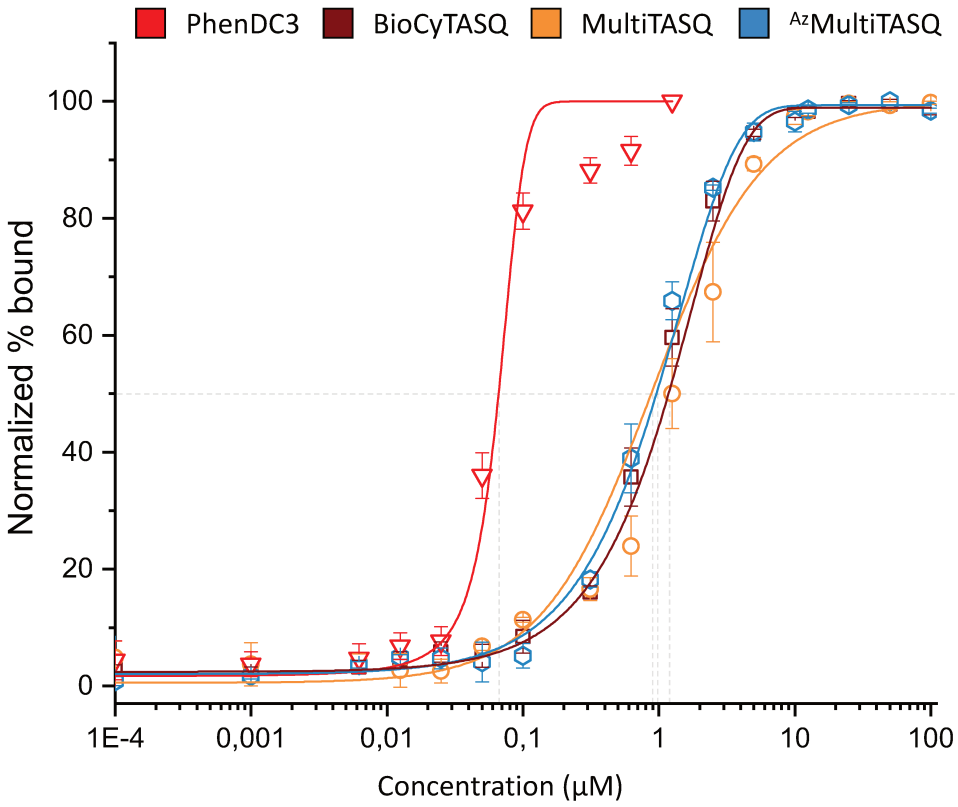


*Figure 35: Titration curves of Cy5-Myc upon addition of increasing amounts of **PhenDC3** (red), **BioCyTASQ** (wine), **MultiTASQ** (yellow) and **^{Az}MultiTASQ** (blue), (n=6).*

We firstly validated the method by calculating the $^{app}K_D$ for **PhenDC3**, already determined in the literature[160]. As previously described, **PhenDC3** was found to have two $^{APP}K_D$, due to the possibility to bind with a 1:1 or a 2:1 stoichiometry with the G4. In our case, only the 1:1 $^{APP}K_D$ will be considered since TASQs have been shown to interact with the same ligand stoichiometry. The value obtained for **PhenDC3** (57.7 ± 0.2 nM) is in line with the literature (29.6 ± 0.9 nM), thus validating our protocol. We then calculated the $^{APP}K_D$ of TASQs: the obtained values of three ligands were found to be roughly an order of magnitude higher than

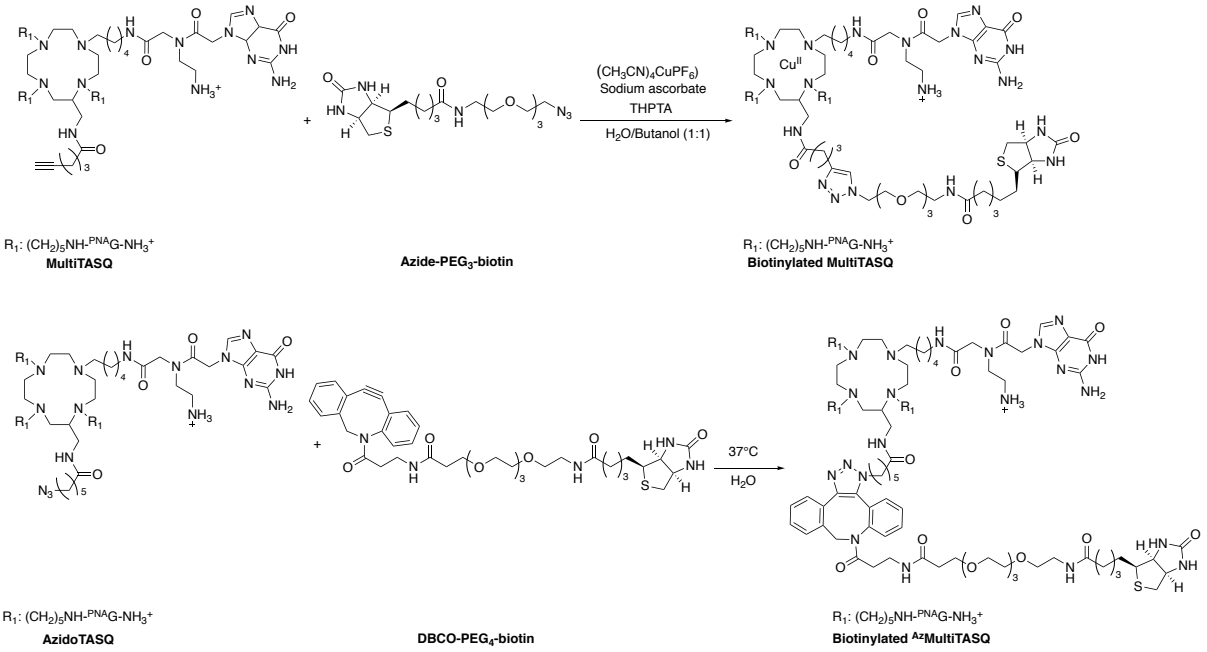
that of **PhenDC3**, with **BioCyTASQ** displaying an $^{APP}K_D$ of 1.01 ± 0.11, **MultiTASQ** of 1.44 ± 0.14 µM and **^{Az}MultiTASQ** of 0.90 ± 0.09 µM *(Figure 35*). This difference were in line with the difference monitored during FRET-melting experiments (**PhenDC3** triggers a stabilization >30°C). These fair $^{APP}K_D$ values, coupled with an exquisite G4-selectivity, make **MultiTASQ** and **^{Az}MultiTASQ** interesting ligands suited to be assessed in fluorescence pull-down experiments. To this end, we will firstly validate the click reactivity of the alkyne and azide moieties with the adequately functionalized biotin derivative.

IV.3.c. Fluorescence G4-pull-down experiments

We first checked that the alkyne/azide moieties can be used for both CuAAC and SPAAC, using either **azido-PEG$_4$-Biotin** conjugate (for CuAAC) or **dibenzocyclooctyne-PEG$_4$-biotin** conjugate (for SPAAC) (*Scheme 9*). To functionalize **MultiTASQ** with a biotin handle, an aqueous solution of tetrakis(acetonitrile)copper(I) tetrafluoroborate was mixed with THPTA ((tris(3-hydroxypropyltriazolylmethyl)amine, a water soluble copper ligand) which makes the solution turn to a dark blue, indicating the presence of chelated Cu(II). Then, sodium ascorbate was added to lead to a colorless solution indicating the formation of the Cu(I)·THPTA complex. Of note, the copper source was originally chosen not to be reduced, but upon dissolution in water and addition of THPTA the presence of the oxidized form was clear, introduction of sodium ascorbate aimed at restoring the active catalyst. **MultiTASQ** was mixed with a slight excess of **azido-PEG$_4$-Biotin** (1.1 mol. equiv.) in a 1:1 mixture of water and 1-butanol (of note, the proportion of 1-butanol is 2% only in the final solution)[108]. The solutions of reactants and catalyst were then mixed and stirred at room temperature for 1 h (with HPLC-MS monitoring to follow the reaction progress). As for **TriazoTASQ** and **BioTriazoTASQ**, the HPLC-MS tracks showed the presence of the copper chelated specie (biotinylated **Cu·MultiTASQ**); we thus decided to investigate whether the presence of the copper was detrimental for the use of the clicked **MultiTASQ**: the solution of the biotinylated and metalated TASQ was split in two, half of this solution was demetallated (Na_2S treatment)[151], the other half was not. The analysis of the fluorescence pull-down results did not show any significant difference between the two experimental conditions (as an example the enrichment obtained with F-NRAS was 5.5 ± 0.8 and 7.6 ± 0.7-fold with and without copper, respectively, *vide infra*). We also found that avoiding the Na_2S treatment provided more reproducible results, the precipitation and centrifugation steps leading to a loss of

MultiTASQ material. We therefore decided to skip the precipitation step, taking into account that the final amount of copper in the G4-containing solution is below 0.5% w/w.

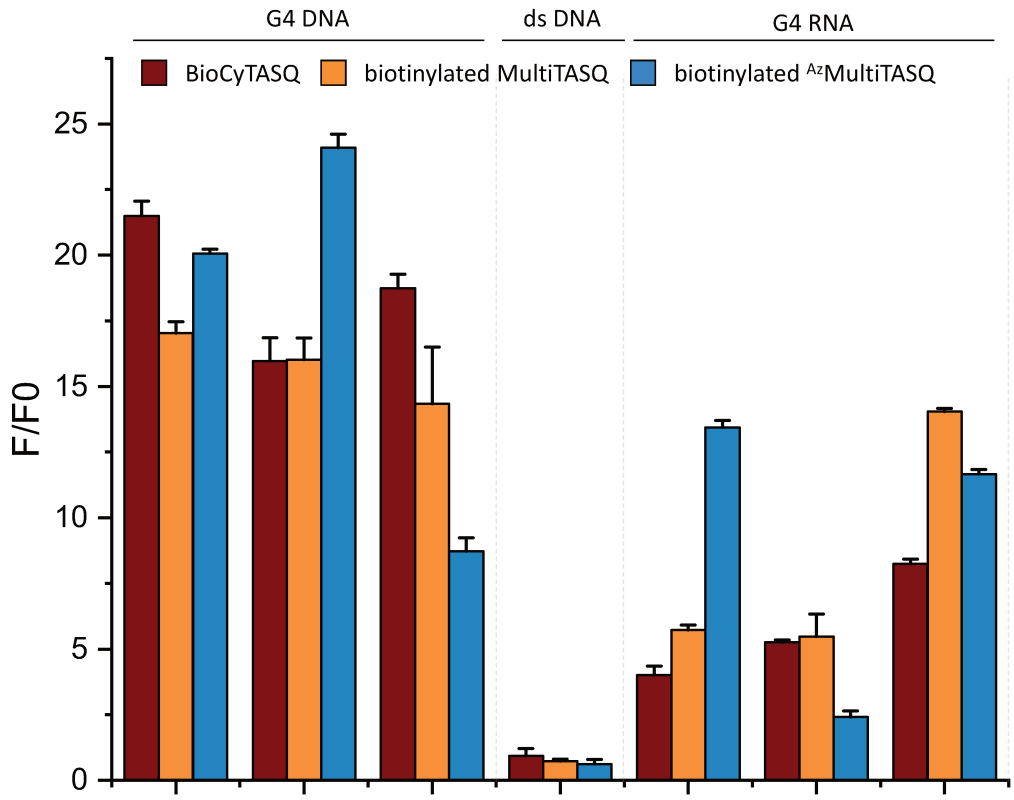
To functionalize **^{Az}MultiTASQ** with a biotin handle, the TASQ was mixed with a slight excess of **dibenzocyclooctyne-PEG$_4$-biotin conjugate** (1.1 mol. equiv.) and stirred for 1 h at 37°C (with HPLC-MS monitoring to stop the reaction after completion). The crude mixture was then incubated with FAM-labeled oligonucleotides and streptavidin magnetic beads following the classical fluorescence G4-pull-down protocol[128]. As a control, the same oligonucleotides were also magnetically precipitated with **BioCyTASQ** for a direct comparison of the performance of the three probes.



*Scheme 9: CuAAC reaction between **MultiTASQ** and azide-PEG$_3$-biotin; SPAAC reaction between **^{Az}MultiTASQ** and DBCO-PEG$_4$-biotin. Both probes are used without further purification for fluorescence G4 pull-down experiments (see below).*

As seen in *Figure 36*, the ability of the three TASQs (**BioCyTASQ**, biotinylated **MultiTASQ** and biotinylated **^{Az}MultiTASQ)** to capture G4s is very similar. All of them show better performances with G4-DNA (between 21.4 ± 0.6 and 17.0 ± 0.4-fold enrichment for F-Myc, 24.1 ± 0.5 and 15.9 ± 0.8 for F-22AG, and 18.8 ± 0.5 and 8.7 ± 0.5 with F-SRC) over G4-RNA (between 4.0 ± 0.3 and 13.4 ± 0.3-fold enrichment with F-TERRA, 5.5 ± 0.8 and 2.4 ± 0.3 for F-NRAS, and 8.2 ± 0.2 and 14.0 ± 0.2 for F-VEGF), along with an excellent G4-selectivity, as no enrichment can be seen with the control dsDNA F-duplex. Therefore, the biotinylated

MultiTASQ and **^{Az}MultiTASQ** have been found to behave as **BioCyTASQ**, with the notable exception of F-TERRA which is strongly enriched by the biotinylated **^{Az}MultiTASQ** (4.0 ± 0.3 *vs* 13.4 ± 0.3-fold). These results clearly demonstrate that CuAAC and SPAAC work with TASQs and provide highly efficient molecular baits. This data prompted us to investigate their G4-capture properties in more relevant conditions, that is, in the qPCR pull-down protocol described above.



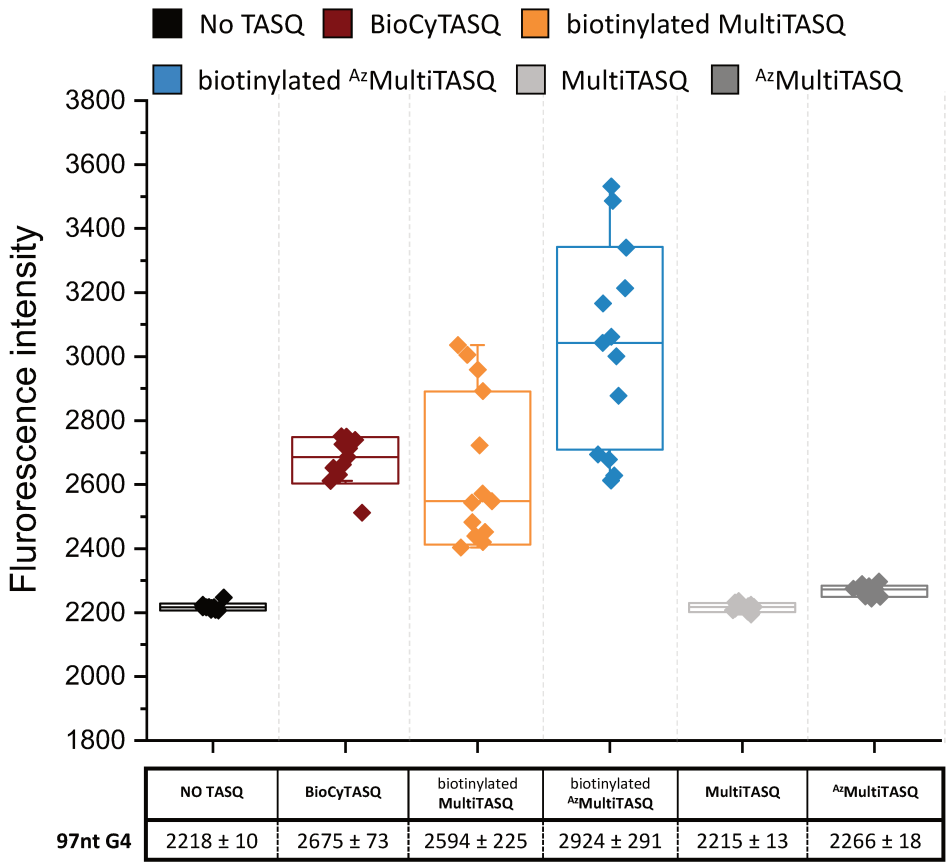
	F-Myc	F-22AG	F-SRC	F-duplex	F-TERRA	F-NRAS	F-VEGF
BioCyTASQ	21.5 ± 0.6	15.9 ± 0.9	18.8 ± 0.5	0.9 ± 0.3	4.0 ± 0.3	5.3 ± 0.9	8.2 ± 0.2
biotinylated **MultiTASQ**	17.0 ± 0.4	16.0 ± 0.8	14.3 ± 2.1	0.7 ± 0.1	5.7 ± 0.8	5.5 ± 0.8	14.0 ± 0.2
biotinylated **^{Az}MultiTASQ**	20.0 ± 0.2	24.1 ± 0.5	8.7 ± 0.5	0.6 ± 0.2	13.4 ± 0.3	2.4 ± 0.3	11.6 ± 0.2

*Figure 36: Fluorescence G4-pull-down results (expressed as fold-change) for experiments performed with **BioCyTASQ** (wine), biotinylated **MultiTASQ** (yellow) and biotinylated **^{Az}MultiTASQ** (blue) (10 µM, 10 mol. equiv.) and the G4-DNA F-Myc, F-22AG and F-SRC; the G4-RNA F-TERRA, F-NRAS and F-VEGF, and the control F-duplex (1 µM) (n=6).*

IV.3.d. G4 chemo-precipitation monitored by qPCR

The protocol developed to assess the G4-capturing abilities of **BioTriazoTASQ** and **BioCyTASQ** (cf. §III.4.d) was adapted to investigate the properties of biotinylated **MultiTASQ** and **^{Az}MultiTASQ**. Biotin handles were clicked to each TASQ prior to their use (CuAAC for **MultiTASQ**; SPAAC for **^{Az}MultiTASQ**, see IV.3.c.) and the properties of the two biotinylated

TASQs were compared to those of **BioCyTASQ**. We included two additional controls, *i.e.*, the unclicked **MultiTASQ** and **^{Az}MultiTASQ**. The five probes were thus incubated with a 97-nt long G4-forming sequence (cf. §III.4.d) for 2 h at room temperature; after magnetic precipitation, removal of the supernatant and resuspension in a fresh buffer, the residual DNA was amplified by qPCR, whose efficiency was monitored here by the SYBR green fluorescence level reached at the end of the amplification (33 cycles).



	NO TASQ	BioCyTASQ	biotinylated MultiTASQ	biotinylated ^{Az}MultiTASQ	MultiTASQ	^{Az}MultiTASQ
97nt G4	2218 ± 10	2675 ± 73	2594 ± 225	2924 ± 291	2215 ± 13	2266 ± 18

*Figure 37: qPCR pull-down results (SYBR green fluorescence intensity) for experiments performed without molecule (control) or with **BioCyTASQ** (wine), **biotinylated MultiTASQ** (yellow), **biotinylated ^{Az}MultiTASQ** (blue), **MultiTASQ** (grey) and **^{Az}MultiTASQ** (dark grey) (40 µM, 10 mol. equiv.) and the G4-strand (4 µM) (n=9).*

As seen in *Figure 37*, no fluorescence increase was observed for the control wells without TASQ ('no TASQ') and with the unclicked **MultiTASQ** and **^{Az}MultiTASQ**, thus validating the need of a biotin moiety to magnetically isolate G4 using streptavidin beads. We observed that the biotinylated **MultiTASQ** and **^{Az}MultiTASQ** can efficiently isolate G4s from solution, in a manner that is reminiscent to what is observed with **BioCyTASQ**. The final fluorescence of the three probes stands in the same range, with that of the biotinylated **^{Az}MultiTASQ** (2924 ± 291)

a bit higher than that of **BioCyTASQ** (2675 ± 73), which itself is a bit higher than that of the biotinylated **MultiTASQ** (2594 ± 225). Altogether, these experiments confirm the results collected *via* the fluorescence G4-pull-down assay, confirming the suitability of both CuAAC and SPAAC with TASQs that provide efficient molecular baits.

*Table 2 : Summary of **BioCyTASQ**, **MultiTASQ** and **^{Az}MultiTASQ** biophysical properties*

FRET-melting assay ($\Delta T_{1/2}$, °C)						
	G4-DNA			**G4-RNA**		
	F-Myc-T	F21T	F-kit2-T	F-TERRA-T	F- NRAS -T	F-VEGF-T
MultiTASQ	5.3 ± 0.5	2.5 ± 0.2	5.8 ± 0.2	11.0 ± 0.9	11.8 ± 0.7	11.9 ± 0.4
^{Az}MultiTASQ	6.4 ± 0.2	2.2 ± 0.3	7.7 ± 1.1	12.4 ± 0.6	14.6 ± 1.1	11.3 ± 0.4
Fluorescence pull-down assay (Normalized change in FAM fluorescence intensity)						
	G4-DNA			**duplex-DNA**	**G4-RNA**	
	F-Myc	F22AG	F-SRC	F-duplex	F-TERRA	F-NRAS	F- VEGF
BioCyTASQ	21.5 ± 0.6	15.9 ± 0.9	18.8 ± 0.5	0.9 ± 0.3	4.0 ± 0.3	5.3 ± 0.9	8.2 ± 0.2
MultiTASQ	17.0 ± 0.4	16.0 ± 0.8	14.3 ± 2.1	0.7 ± 0.1	5.7 ± 0.8	5.5 ± 0.8	14.0 ± 0.2
^{Az}MultiTASQ	20.0 ± 0.2	24.1 ± 0.5	8.7 ± 0.5	0.6 ± 0.2	13.4 ± 0.3	2.4 ± 0.3	11.6 ± 0.2
$^{APP}K_D$ (Conc nM)						
PhenDC3 (1 :1)	**BioCyTASQ**			**MultiTASQ**		**^{Az}MultiTASQ**
58 ± 0.2	1013 ± 110			1440 ± 140		900 ± 90
qPCR pull-down assay (SYBRGreen fluorescence intensity)						
No TASQ	**BioCyTASQ**	**Biotin-MultiTASQ**		**MultiTASQ**	**Biotin-^{Az}MultiTASQ**	**^{Az}MultiTASQ**
2218 ± 10	2675 ± 73	2594 ± 225		2215 ± 13	2924 ± 291	2266 ± 18

To further exploit their versatility, the clickable **MultiTASQ** and **^{Az}MultiTASQ** were used for *in situ* click imaging, that is, for imaging G4s in cells conjugating TASQs once in their genomic sites with fluorescent partners. The design of both **MultiTASQ** and **^{Az}MultiTASQ** ensures that the alkyne and azido appendages, due their small size and non-HB nature, do not impact interaction with the target in cells, making them suited to be live-incubated.

IV.3.e. *In situ* G4 click imaging

As above, we verified the reactivity of **MultiTASQ** for click chemistry performed with **AlexaFluor488-azide**, using copper sulfate as Cu(II) source, THPTA as ligand and sodium ascorbate as reducing agent. The reaction completion was monitored by HPLC-MS (*Figure 38*), which showed complete conversion of the starting material to the **AlexaFluor488-MultiTASQ** conjugate in 1 h at room temperature.

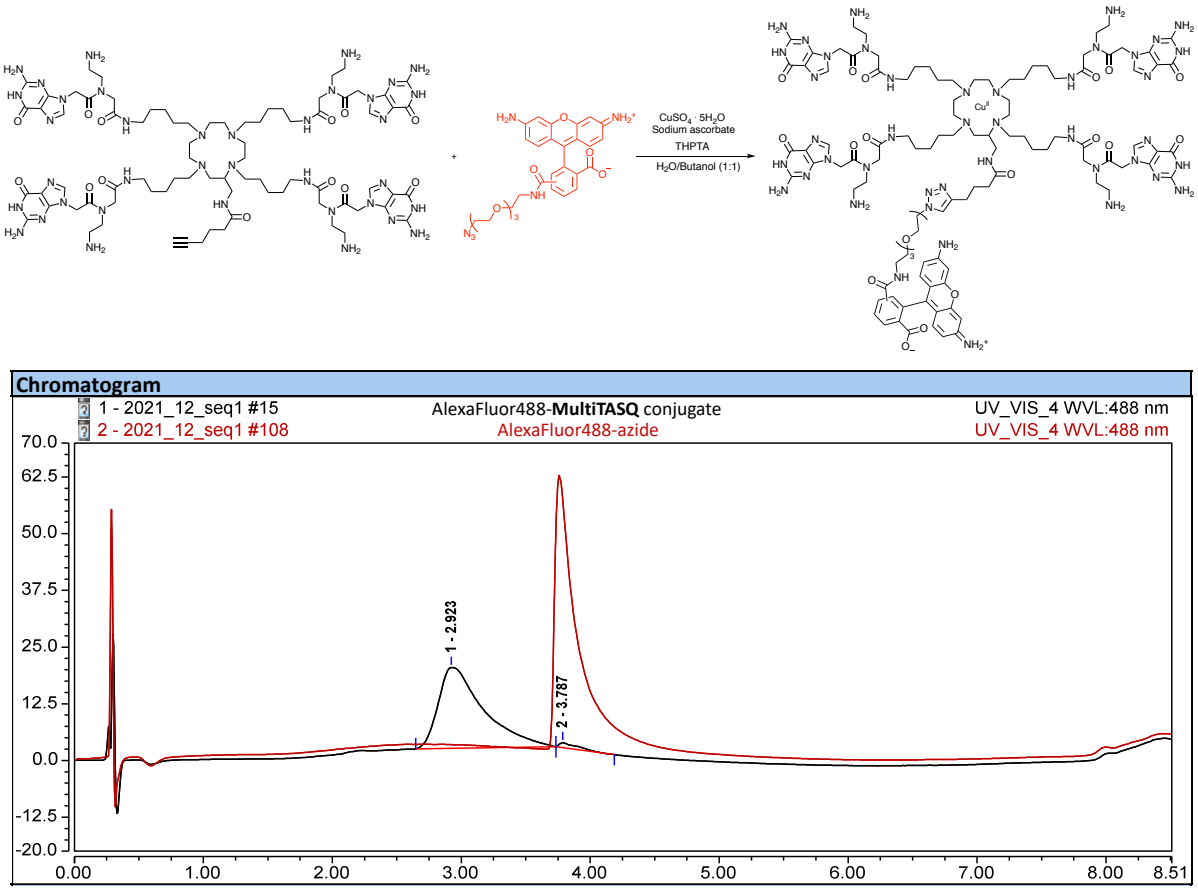


*Figure 38: CuAAC reaction between AlexaFluor488-azide (red trace) and **MultiTASQ**; the black trace corresponds to the complete conversion of the starting material to the target molecule.*

We thus decided to perform the same reaction directly in cells. MCF7 human breast cancer cells were incubated either live or post-fixation with **MultiTASQ** (5 µM for 24 hours in live cells and 10 µM for 1 hour in fixed cells). In both instances, the click reaction was performed in cells using an azido-functionalized fluorophore, either AlexaFluor488-azide (λ_{ex} = 493 nm, λ_{em} = 514- 524 nm, green) or AlexaFluor594-azide (λ_{ex} = 591 nm, λ_{em} = 604- 624 nm red), to show the modularity of the *in situ* click imaging protocols (*Figure 39*).

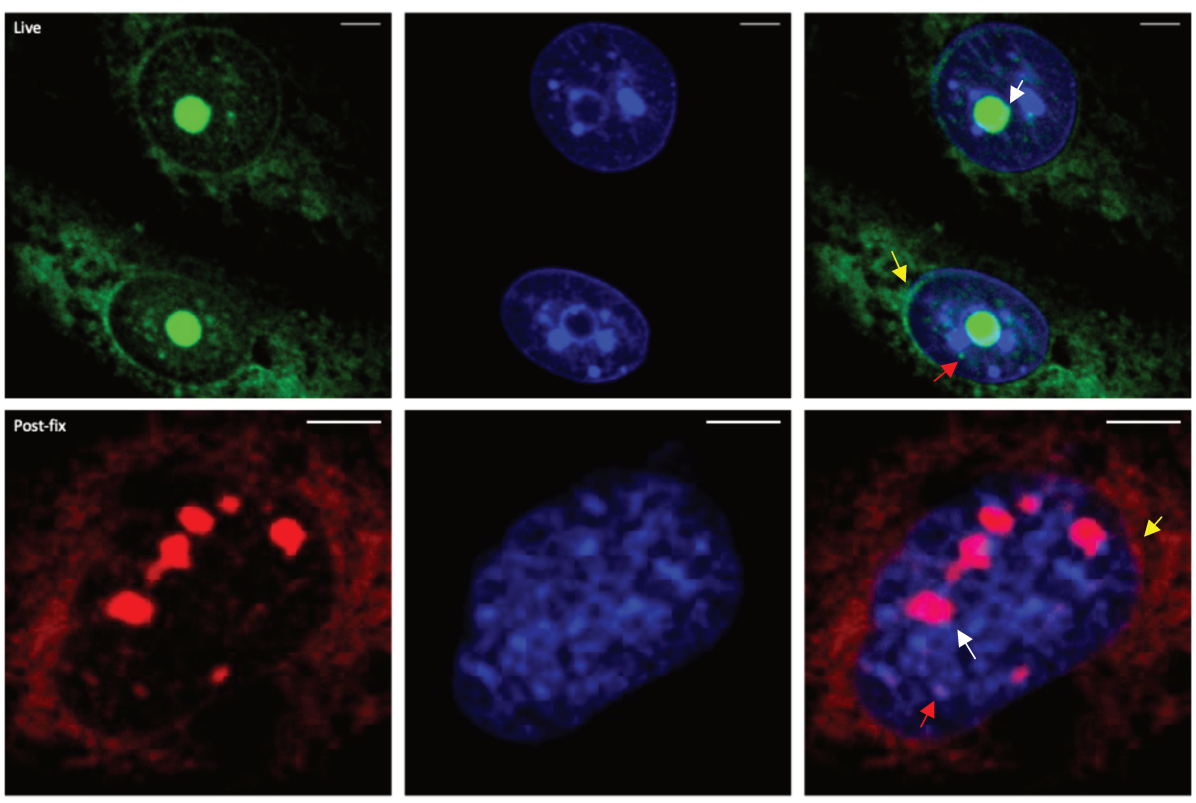


*Figure 39: High-resolution optical images of MCF7 cells treated with **MultiTASQ** (1 µM, 24 h), clicked with AlexaFluor488-azide (green) in live cell (scale bar = 10 µm); cells clicked with AlexaFluor594-azide (red) after cell fixation (scale bar = 10 µm) and chromatin staining by DAPI (blue). Both fixed and live cell incubation shows perinuclear region labelling (yellow arrow), nucleoli (white arrows) and non-nucleoli nucleoplasmic site (red arrow).*

The obtained staining pattern showed an important nucleolar staining (white arrow) and smaller nucleoplasmic *foci* (red arrow), reminiscent of what was obtained for **BioCyTASQ** (cf §II.5.e) and **N-TASQ** (further discussed below). Importantly, in both conditions (live-cell treatment or post fixation treatment), the signal-to-noise ratio is largely improved as compared to the pre-targeted G4-imaging performed with **BioCyTASQ**, which of course originates in the complete bioorthogonality of the CuAAC protocol. This resulted in better quality images; however, we did not further exploit them yet (for instance *via* colocalization with BG4, with an anti-γH2AX antibody or *via* treatment with DNase or RNase, etc.) because our aim was here limited to assess the efficiency of the *in situ* CuAAC with TASQs in the context of cell imaging.

The experiment was repeated using **^{Az}MultiTASQ**. As above, we first validated the reactivity of this TASQ for SPAAC reaction with a home-made BCN-rhodamine conjugate (kindly provided by Thibaud Bailly). As for biotin-DBCO, **^{Az}MultiTASQ** was incubated for 24 hours in water at 37 °C, unfortunately, even with longer reaction times (up to 48 h) did not lead to a complete conversion of the starting material, for a reason that remains unexplained to date (*Figure 40*).

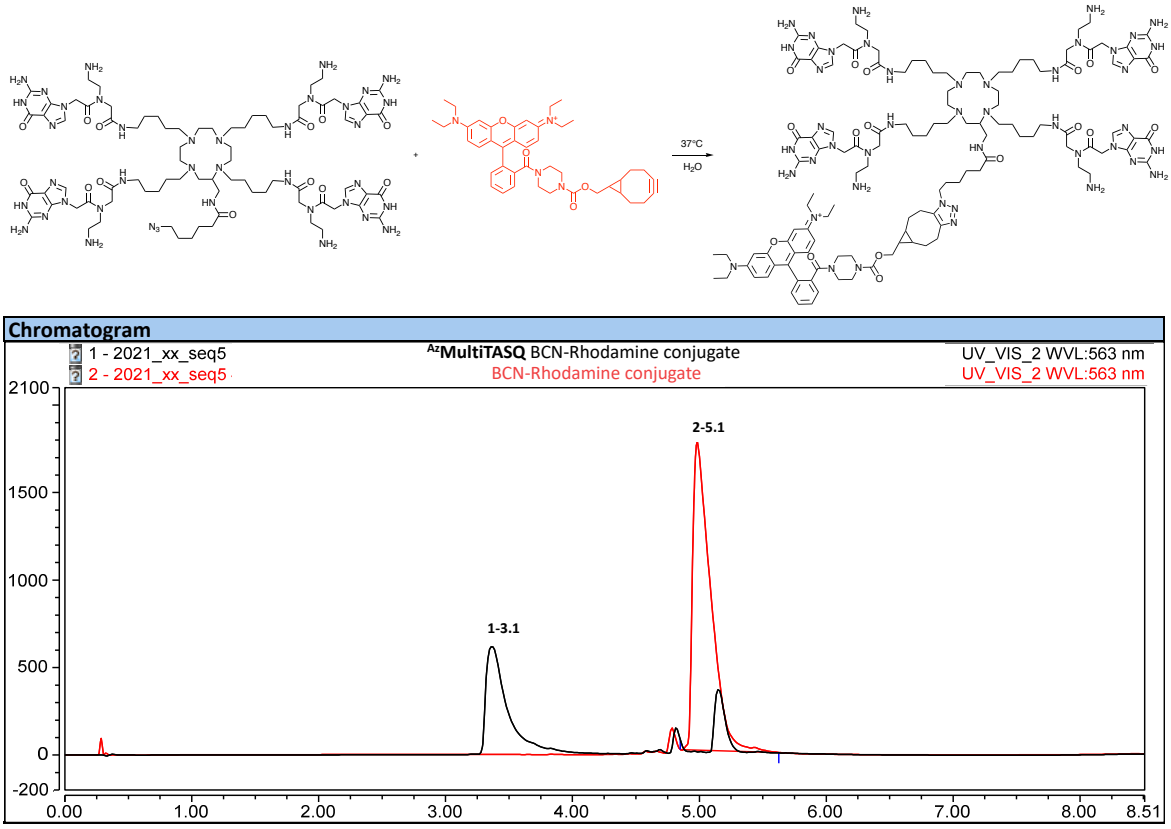


*Figure 40: SPAAC reaction between bicyclo[6.1.0]-nonyne (BCN)-conjugated with a hodamine (red trace) and the **^{Az}MultiTASQ**. the black trace corresponds to the complete conversion of the starting material to the target molecule.*

Nonetheless, we decided to perform cell-based investigations to assess whether this TASQ can be conveniently used for visualizing G4s. Fixed MCF7 cells (ice cold methanol) were incubated for 1 h with **^{Az}MultiTASQ** (20 µM) and, after washings, reacted with BCN-rhodamine (10 µM) at 37°C for 30 min. After washings with a Triton-enriched PBS solution, cells were imaged (λ_{ex} = 552 nm, λ_{em} = 571- 582 nm). As seen in the *Figure 41*, a strong signal can be observed in the perinuclear region (yellow arrow) for cells exposed to Rhodamine-BCN conjugate, with and without **^{Az}MultiTASQ.** This high background noise can be explained by the ability of Rhodamine to interact with diverse cellular components, including

mitochondria[162]. In literature, the use of a scavenger such as an Azido-PEG conjugate (11-Azido-3,6,9-trioxaundecan-1-amine) to react with the rhodamine in excess has already been described to facilitate the removal of the probe and thus, to increase the picture quality[163]. Besides the perinuclear region, a classical TASQ accumulation was observed notably in the nucleoli region (white arrow). However, these images cannot be further exploited owing to their very low resolution.

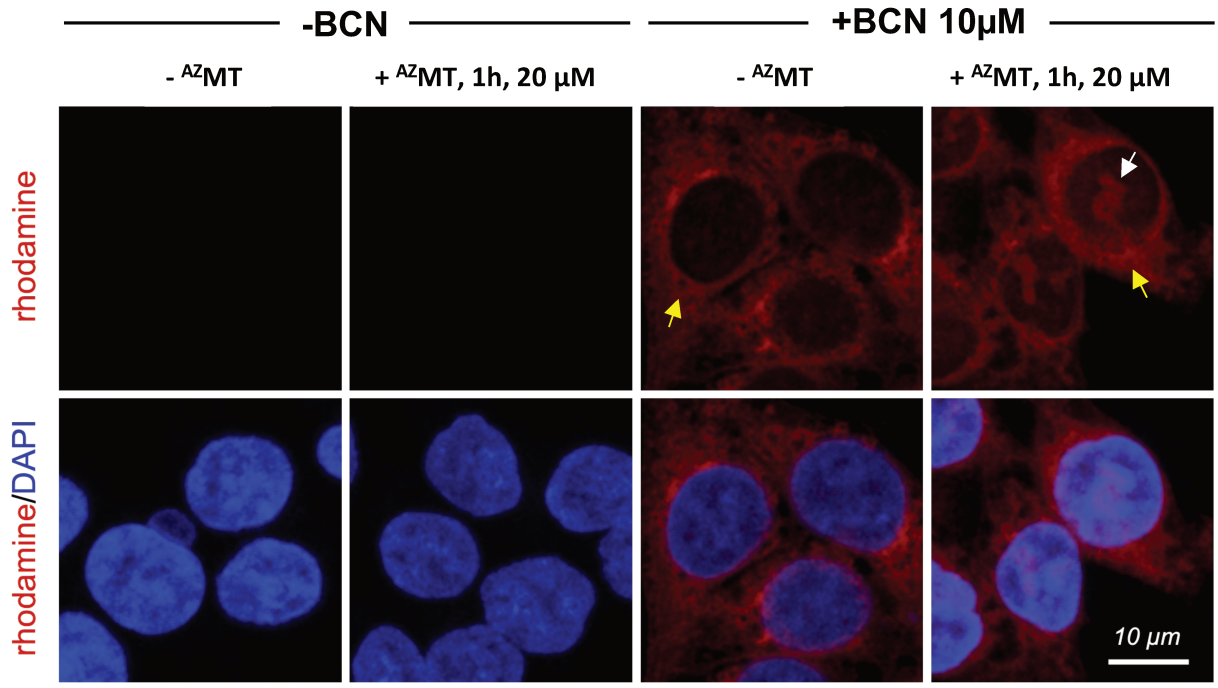


*Figure 41: High-resolution optical images of MCF7 cells treated with ^{Az}**MultiTASQ** (20 µM, 1 h), clicked with and without Rhodamine BCN (10 µM, red) in fixed cell; chromatin staining by DAPI (blue). Yellow arrows indicate the perinuclear region stained by Rhodamine BCN with and without the probe, white arrows indicate the nucleoli region stained only when ^{Az}**MultiTASQ** is present.*

These preliminary investigations indicated the need to develop either an efficient washing protocol or to change the nature of the fluorophore partner. Once images with a more acceptable quality will be obtained, we will investigate the use of this TASQ in live cells. This will correspond to the ideal use of ^{Az}**MultiTASQ**, as it uniquely offers the possibility of performing cell imaging without fixation step. Cellular studies in a native environment, coupled with real-time tracking of G4-formation in live cells, could thus exploit the full potential of this TASQ.

IV.4. Conclusion:

By using two bioorthogonal appendages (alkyne and azide), we have increased the TASQs versatility and opened ways towards their conjugation with a higher number of partners including commercially available biotin derivatives and fluorophores conjugated with the appropriate reactive partners (azides or BCN). Even if the synthesis of **MultiTASQ** and **^{Az}MultiTASQ** (6 and 3% overall chemical yield, respectively) were not as efficient as that of **BioCyTASQ**, the two probes showed excellent G4-interacting properties. Their affinities for G4s, evaluated through the determination of $\Delta T_{1/2}$ or $^{APP}K_D$ values, were found to be similar to that of **BioCyTASQ**, thus showing that the nature of the appendage on the AMC template (biotin, alkyne, azide) does not impact the G4-binding of this family of molecules. Their excellent selectivity for G4s, evaluated by competitive FRET-melting experiments, is also found to be similar to that of **BioCyTASQ**. All these features make **MultiTASQ** and **^{Az}MultiTASQ** highly valuable G4-ligands. The chemical reactivity of the alkyne and azide appendage was first assessed by performing both CuAAC and SPAAC with their corresponding clickable biotin partners. This approach allowed for using the resulting biotinylated TASQs in G4 pull-down experiments, performed with either fluorescently-labeled G4s or G4s included in a long DNA strand. As above, the G4-capture properties of the biotinylated **MultiTASQ** and **^{Az}MultiTASQ** proved to be as good as that of **BioCyTASQ**, indicating that changing a biotin handle for a clicked biotin results in TASQs displaying improved versatility without losing the initial enticing G4-binding ability. We further demonstrated this versatility clicking the two TASQs to different fluorophores, both *in vitro* and in cells. The images obtained through this pre-targeting and post-labeling protocol, referred to as *in situ* click imaging, show a better signal-to-noise ratio compared to the **BioCyTASQ** ones, obtained using Cy3-SA labeling, which thus validates the hypothesis according to which both **MultiTASQ** and **^{Az}MultiTASQ** do not suffer from unspecific interactions with cellular components.

The full potential of this new generation TASQs still deserves to be fully assessed (*e.g.*, for sequencing applications); however, we decided to keep on increasing the versatility of TASQ *via* the introduction of a photoactivable moiety on the **MultiTASQ** scaffold, which results in a new G4-probe ideally suited to perform both genomics and proteomics.

photoMultiTASQ:

The successful introduction of the alkynyl moiety into the TASQ scaffold paved the way to include another function, using an appendage comprising both an alkynyl group and a diazirine function (*i.e.*, 3-[3-(but-3-yn-1-yl)-3H-diazirin-3-yl]propanoic acid, *Figure 42*). In the literature, the use of G4-ligands that can covalently trap G4s and their surrounding partners is a rapidly growing field[164-166], aiming at identifying new G4-interacting molecules (chiefly proteins) and determine their biological role[167-168]. The photoactivatable diazirine, being UV-light sensitive and small-sized, is indeed an ideal candidate to be included in the TASQ structure as it was recently used in a photo-capture protocol that identified new G4-binding proteins (G4BP) by proteomics[169], and it is small and inert enough not to disturb the target engagement of the resulting TASQ. This new TASQ, referred to as **photoMultiTASQ** (this molecules was patented as well)[159], is activable by irradiation at 365 nm, which triggers the formation of a highly reactive carbene that creates covalent links with surrounding molecules. This allows for a great control over the activation stimuli (light) and at wavelengths that are minimally absorbed by proteins or nucleic acids. We envisioned that once the TASQ/G4 complex formed, it would be possible to covalently trap either nucleic acids (to confirm the identity of their targets in cells *via* sequencing) or proteins (to identify G4BP partners *via* proteomics) by UV-irradiation. Indeed, CuAAC will be implemented for coupling the alkynyl group with a **biotin-PEG-azide** conjugate, which leads to biotinylated TASQ/G4 and/or TASQ/G4/G4BP complexes readily isolated by magnetic precipitation. The analysis of the isolated complex (after dissociation by thermal denaturation and lysis) will be finally conducted by either sequencing or proteomic analyses.

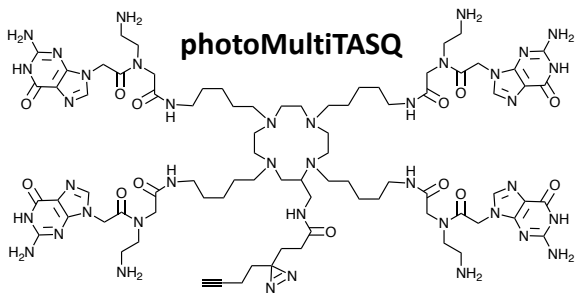
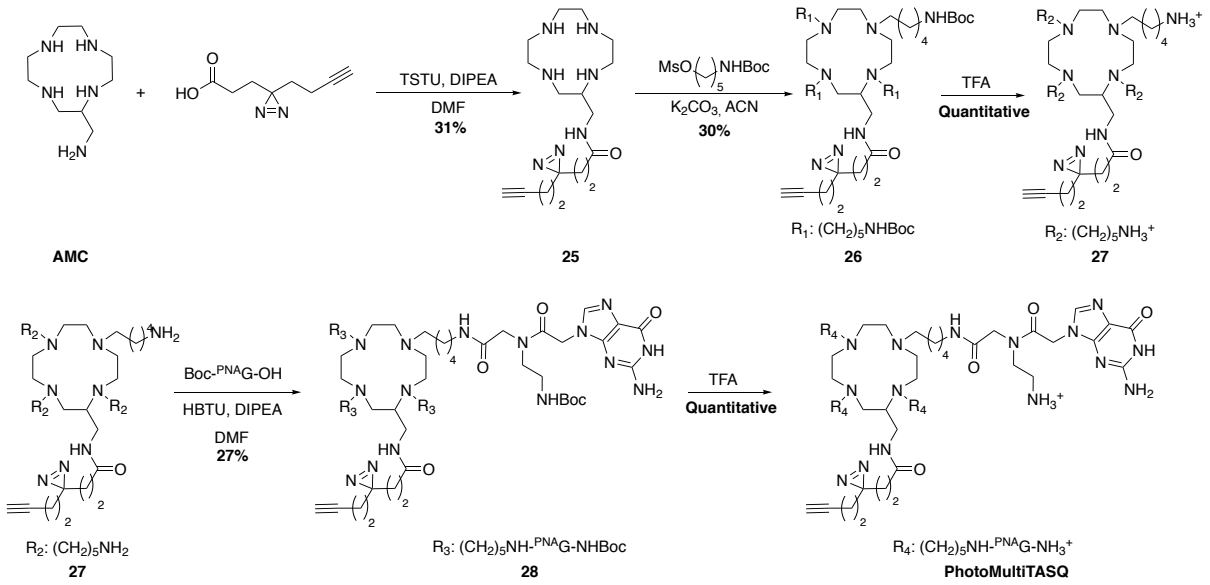


*Figure 42: **photoMultiTASQ** structure*

IV.5. PhotoMultiTASQ synthetic pathway:

The synthesis of **photoMultiTASQ** (*Scheme 10*) is directly inspired by that of **MultiTASQ**: **AMC** was first coupled with 3-[3-(but-3-yn-1-yl)-3H-diazirin-3-yl]propanoic acid in presence of TSTU and DIPEA, to provide the functionalized macrocycle **25** (31% chemical yield). This conjugate was then reacted with the previously used aliphatic arms to yield to compound **26** in 30% chemical yield. After Boc-deprotection of the terminal amines of **26**, the Boc-^{PNA}G monomer was coupled in presence of HBTU and DIPEA to obtain the Boc-protected **photoMultiTASQ** with a 3% overall chemical yield over 4 steps. The **photoMultiTASQ** synthesis still needs to be optimized, but we faced supply issues (the only European provider of the diazirine appendage, Enamine Ltd, is located in Ukraine), that is why we performed a series of *in vitro* evaluations to characterize the G4-binding abilities of **photoMultiTASQ**.



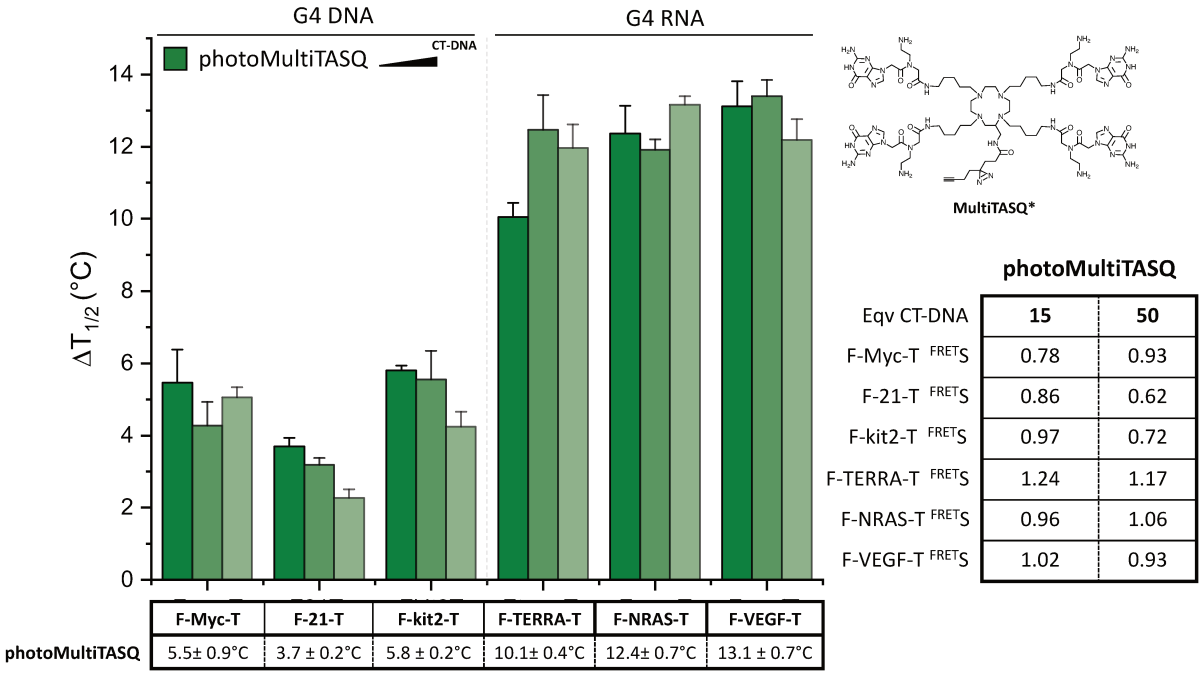
*Scheme 10: **photoMultiTASQ** synthetic pathway*

IV.6. The G4-interacting properties of photoMultiTASQ:

IV.6. a. FRET Melting:

As above, the G4-binding abilities of **photoMultiTASQ** were assessed by FRET-melting assay. The apparent affinity was investigated against a panel of doubly labeled oligonucleotides comprising 3 DNA (F-21-T, F-Myc-T and F-Kit-T) and 3 RNA QFSs (F-TERRA-T, F-NRAS-T and F-VEGF-T). As expected, **photoMultiTASQ** triggers a weak but significant stabilization of G4-DNA ($\Delta T_{1/2}$ values between 3.7 ± 0.2 and 5.8 ± 0.2 °C), with a better stabilization of G4-RNA ($\Delta T_{1/2}$

between 10.1 ± 0.4 and 13.1 ± 0.7 °C). We also performed competitive FRET-melting experiments in the presence of an excess (15 or 50 mol. equiv.) of CT-DNA: the selectivity values (^{FRET}S, described above) demonstrate the good (DNA) to excellent (RNA) selectivity of **photoMultiTASQ** for G4s over dsDNA (with ^{FRET}S > 0.62 for DNAs, ^{FRET}S > 0.93 for RNAs, *Figure 43*).



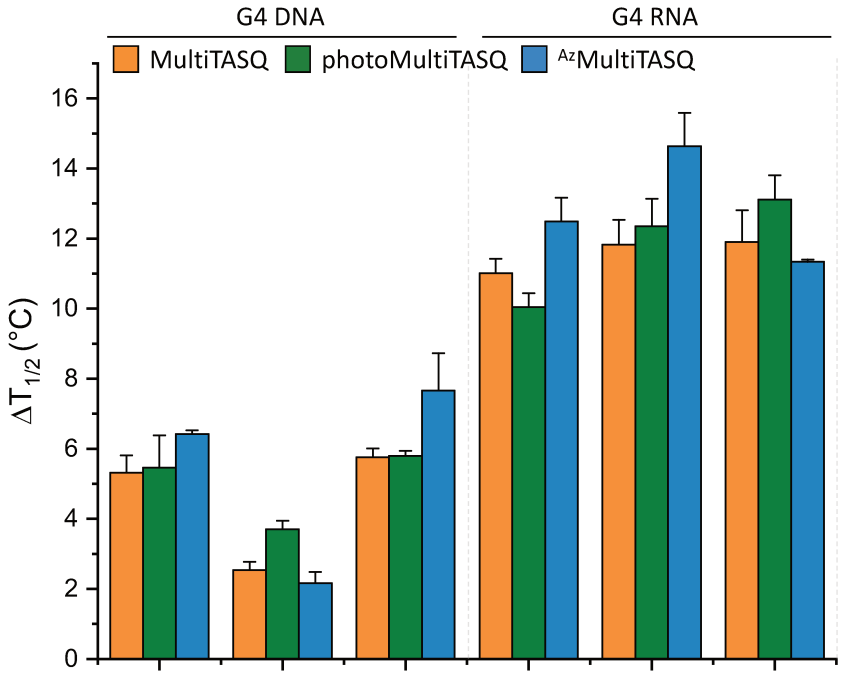
photoMultiTASQ	F-Myc-T	F-21-T	F-kit2-T	F-TERRA-T	F-NRAS-T	F-VEGF-T
	5.5± 0.9°C	3.7 ± 0.2°C	5.8 ± 0.2°C	10.1± 0.4°C	12.4± 0.7°C	13.1 ± 0.7°C

photoMultiTASQ

Eqv CT-DNA	15	50
F-Myc-T ^{FRET}S	0.78	0.93
F-21-T ^{FRET}S	0.86	0.62
F-kit2-T ^{FRET}S	0.97	0.72
F-TERRA-T ^{FRET}S	1.24	1.17
F-NRAS-T ^{FRET}S	0.96	1.06
F-VEGF-T ^{FRET}S	1.02	0.93

*Figure 43: FRET-melting results ($\Delta T_{1/2}$, °C) for experiments performed with **photoMultiTASQ** (1 µM, 5 mol. equiv.) and the G4-DNA F-Myc-T, F21T and F-kit2-T, the G4-RNA F-TERRA-T, F-NRAS-T and F-VEGF-T with increasing concentration of CT-DNA (15 and 50 equivalents). Selectivity values (^{FRET}S) calculated for **MultiTASQ** and ^{Az}**MultiTASQ** (n=3).*

We also compared the apparent affinity (FRET-melting results) of **photoMultiTASQ**, **MultiTASQ** and ^{Az}**MultiTASQ**: as seen in *Figure 44*, the trends are identical for the 3 probes against both DNA and RNA G4s, confirming that they interact with G4s according to a similar binding mode.



	F-Myc-T	F-21-T	F-kit2-T	F-TERRA-T	F-NRAS-T	F-VEGF-T
MultiTASQ	5.3±0.5°C	2.5±0.2°C	5.8±0.2°C	11.0±0.4°C	11.8±0.7°C	11.9±0.9°C
photoMultiTASQ	5.5±0.9°C	3.7±0.2°C	5.8±0.2°C	10.1±0.4°C	12.4±0.7°C	13.1±0.7°C
^{Az}MultiTASQ	6.4±0.2°C	2.2±0.3°C	7.7±1.1°C	12.5±0.4°C	14.6±1.0°C	11.3±0.6°C

*Figure 44: FRET-melting results ($\Delta T_{1/2}$, °C) for experiments performed with **MultiTASQ**, **photoMultiTASQ** and **^{Az}MultiTASQ** (1 µM, 5 mol. equiv.) and the G4-DNA F-Myc-T, F21T and F-kit2-T, the G4-RNA F-TERRA-T, F-NRAS-T and F-VEGF-T (n=3).*

IV.6.b. Diazirine activation:

To date, the biophysical properties of **photoMultiTASQ** have only been partially investigated. As the reactivity of the alkynyl appendage of the closely related **MultiTASQ** was validated, we first focused on the study of the photoactivation of the diazirine moiety, which turned out to be challenging. In the literature, several activation protocols have been reported, using different, quite often expensive lamps. We tried to use a home-made lamp (kindly provided by Richard Decréau) and monitor the covalent linkage of **photoMultiTASQ** with its G4-partner by HPLC. Indeed, the oligonucleotides being extremely polar compounds, they can be analyzed by RP-HPLC using ion-pairing reagents: the mobile phases are composed by a buffer solution (pH 7.2) of triethylammonium acetate (TEAA) and acetonitrile. The positively charged ammonium salts act as a counter-ions for the negatively charged phosphate backbone of DNA, which drastically increases the lipophilicity of the oligonucleotide (*Figure 45*). This allows for its interaction with the lipophilic stationary phase, and thus, a better retention in a

conventional C18 column. This technique can be carried over denaturing or non-denaturing conditions, in order to investigate the presence of secondary structures[170].

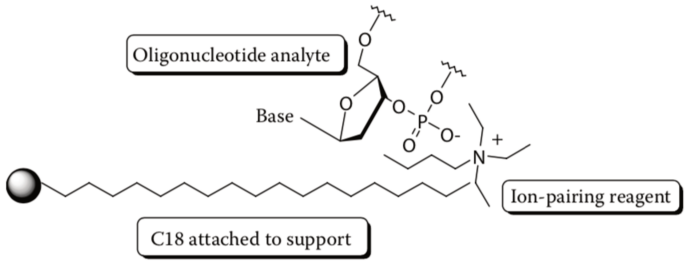


Figure 45: Association of an oligonucleotide with an ion-pairing reagent at the surface of a C18 support.

As TASQs have a high molecular weight (*ca*. 2 kDa) and display four positive charges, we hypothesized that the covalent linkage of **photoMultiTASQ** with a DNA G4 would result in a modification of the retention time of the G4, on the basis of a global change in mass and polarity of the final adduct[171]. This technique can be easily coupled with high resolution mass spectrometry, which could be used to confirm the formation of the covalent adduct.

To this end, we opted for analyses performed at 70°C (denaturing conditions), which might confirm the covalent nature of this complex. As a model G4, we choose a well-known DNA QFS SRC, as it was available in the lab. SRC was pre-folded and then incubated with **photoMultiTASQ** (100 µM, 5 mol. equiv. of TASQ) for 20 min. The solution was then placed into an ice bath and irradiated for 5, 15 and 30 min using a lamp at a maximum wavelength of 365 nm (10 MW). Different distances between the light source and the samples were investigated, ranging from 5 to 30 cm. After irradiation, the mixture was thermally denaturated (90 °C for 5 min) and diluted (5-fold) with the TEAA mobile phase.
Unfortunately, despite massive efforts, we failed in obtaining RP-HPLC traces showing difference in retention time between the starting material and the reaction mixture (*Figure 46*). This can be explained either by a lack of separation of the two compounds (**photoMultiTASQ**/SRC adduct and the starting materials) or to a lack of activation of diazirine moiety. We believe that the irradiation source was the actual problem (the lamp being quite old, home-made and without validation protocol). These experiments must thus be repeated with a more suited experimental setup.

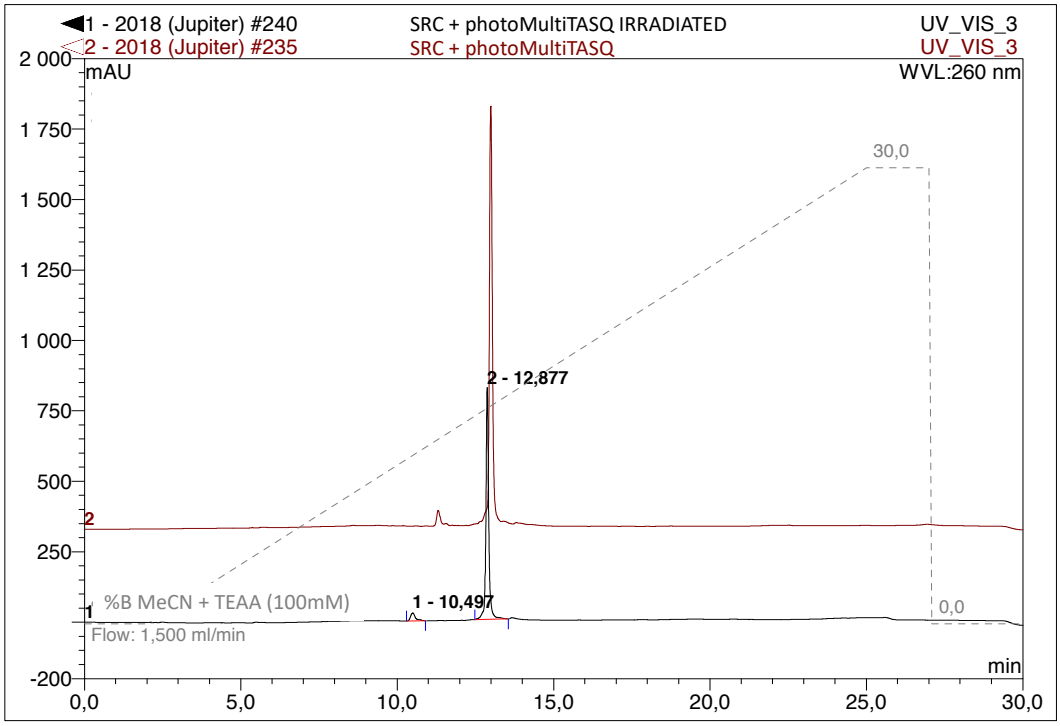


*Figure 46 : HPLC traces of SRC and **PhotoMultiTASQ** (5 mol. equiv.) before and after irradiation for 1 hour at 5 cm. Gradient from 0% to 30% in 30 minutes; Buffer A: TEAA 100mM, Buffer B: 90% MeCN +10% Buffer A.*

IV.7. Conclusion:

The design and synthesis of **photoMultiTASQ** was concomitant with that of the photoactivable probe **photoPDS**, synthetized by the S. Balasubramanian's group [168]. The latter shows excellent G4-binding skills, good G4-affinity capture properties once biotinylated and excellent G4BP trapping upon irradiation[167-168]. The covalent binding ability of **photoPDS** was validated by trapping nucleolin and hnRNP A1, two well-known G4BPs, and led to the identification of new G4BPs such as UHRF1, transcription termination factor TTF2, ATP-dependent RNA helicases DDX1 and DDX24 and pre-mRNA-splicing factor RBM22.

While our prototypal synthesis was efficient enough to isolate pure **photoMultiTASQ** in decent quantities, we failed in activating the diazirine moiety, which remains to be done to validate the relevance of our strategy. We thus plan to buy soon a dedicated photochemistry lamp (*e. g.*, UVP CL-1000 Ultraviolet Crosslinker, Fisher Scientific) that could solve this issue in order to evaluate the performance of our probe in co-binding-mediated protein profiling (CMPP) experiments.

Chapter 5:

NaphthoTASQ (N-TASQ) and ^{Tz}N-TASQ

In the previous chapters, we have shown that biotinylated TASQs (**BioTASQ**, **BioCyTASQ** and **BioTriazoTASQ**) and clickable TASQs (**MultiTASQ** and ^{Az}**MultiTASQ**) can be used as probes for pre-targeting G4 imaging (the former) and *in situ* click imaging (the latter). This approach, though efficient, requires additional manipulations (*i.e.*, labeling with Streptavidin-Cy3 conjugate (the former), click chemistry with *ad hoc* functionalized fluorophores (the latter)) for G4s to be detected in cells. An alternative is to use probes that allow for a direct G4-visualization without further manipulation. In this chapter, we will discuss how the imaging of G4s was achieved using **N-TASQ**, and how we have been able to both improve its chemical access and synthesize a new derivative referred to as ^{Tz}**N-TASQ**.

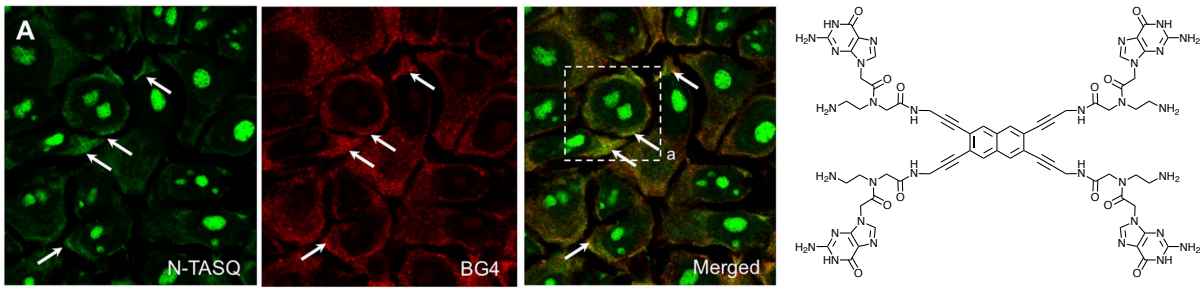


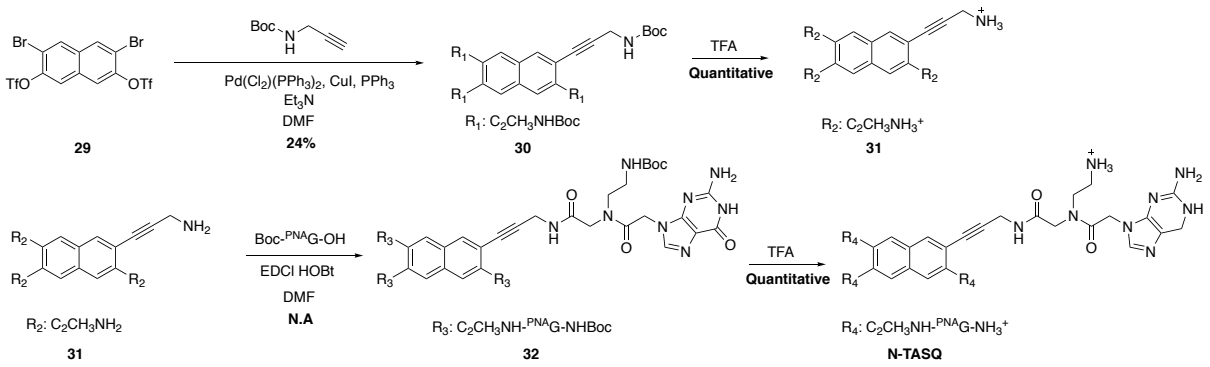
*Figure 47: Images of fixed MCF7 cells labelled with BG4 (10µg/mL, red channel) and **N-TASQ** (100µM, green channel). White arrows indicate some of the most representative **N-TASQ/BG4** co-straining sites. Left: the chemical structure of **N-TASQ**.*

In 2015, **N-TASQ** was published as a twice-as-smart G4-ligand, that is, a ligand that is both a smart ligand (which adopts its G4-binding conformation in the presence of G4s) and a smart fluorescent probe (whose fluorescence is turned-on upon interaction with G4s)[118-119]. Indeed, the central naphthalene template was shown to fluoresce when the intramolecular G-quartet is assembled (the 'closed' conformation, *i.e.*, upon the interaction with G4s) while its fluorescence was mostly quenched when free in solution (*i.e.*, the G arms are free). This results from a combination of G-mediated quench effect and a high degree of freedom of the TASQ in its 'open' conformation (energy dissipation *via* vibrational motion)[172]. The interaction with G4s forces the synthetic G-quartet to assemble, which 'freezes' the Gs in a position that hampers energy loss by vibration. Therefore, **N-TASQ** fluorescence is turned on upon

interaction with G4s, which was firstly validated *in vitro* and then proven in cells, allowing the very first direct visualization of G4-RNA in living cells[118] (*Figure 47*).

N-TASQ is a valuable G4-probe (fluorescence *turn-on* properties along with a good cell-permeability and no toxicity (IC_{50} > 100 µM)), but its synthesis was found to be particularly troublesome. The synthetic scheme comprises only four reactions steps but the key step, the quadruple Sonogashira-Heck-Cassar coupling (*see below*), turned out to be the bottleneck of this synthesis.

V.1. N-TASQ synthetic pathway:

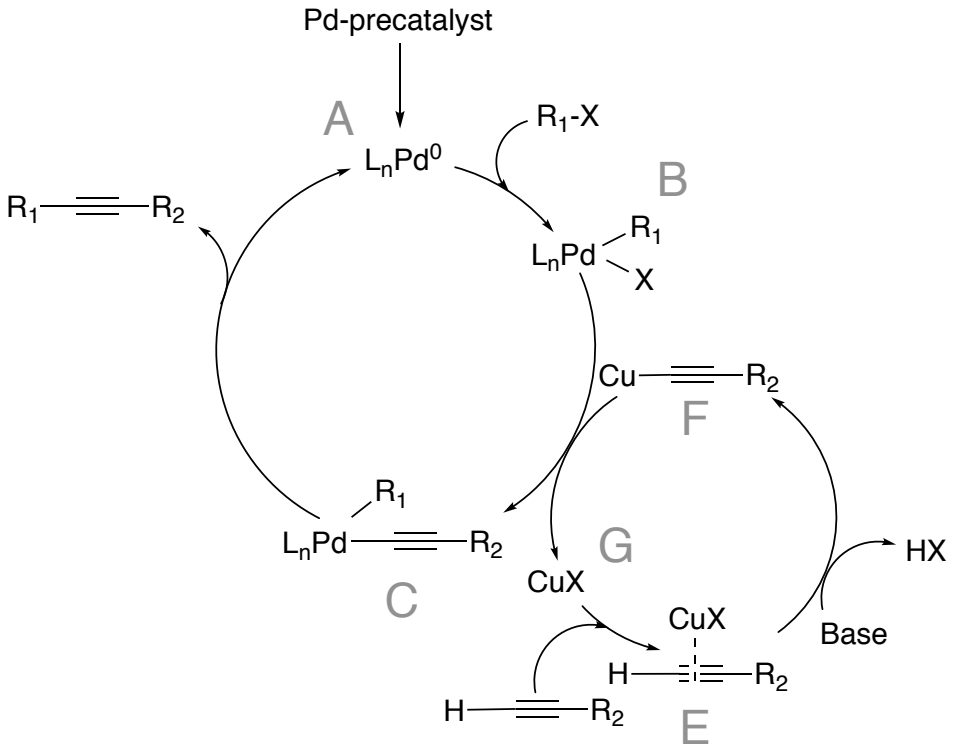


*Scheme 11: **N-TASQ** initial synthetic pathway.*

As seen in *Scheme 11*, the starting compound **29** (prepared according to Shinamura procedure[173]) reacted with the Boc-protected propargylamine under the conditions of the Sonogashira-Heck-Cassar[174-176] coupling ($Pd(PPh_3)_2Cl_2$, CuI, PPh_3 and NEt_3) in DMF under inert atmosphere at 100 °C. The reaction was stirred overnight and, after the reaction was stopped, compound **30** was isolated thanks to a column chromatography on silica gel followed by recrystallization (24% chemical yield). This compound was then deprotected (TFA, quantitative) and the resulting compound **31** was coupled with Boc-^{PNA}G using classical peptide coupling conditions (EDCI/HOBt) to obtain the protected **N-TASQ**. The target molecule was obtained after a TFA treatment (storage form) and deprotected just before use (TFA, quantitative)[118].

We identified the key step as the isolation of compound **30**. While the global yield of this step reported in literature is not particularly low (24%, which correspond to a yield of 70% for each coupling reaction), the reaction was found to be successful in a small scale only (<100 mg) and, unfortunately, poorly reproducible. With the help from our collaborators from IOCB Prague (Petra Menova's group), in the context of Jiri Ledvinka's joint Master 2 internship, we focused on the optimization of this step with the aim of scaling it up and improving its reproducibility.

V.1.a. Sonogashira-Heck-Cassar coupling :



Scheme 12: Sonogashira-Heck-Cassar coupling proposed mechanism.

The cross-coupling mechanism[177] is based on the formation of Pd(0) source (A, *Scheme 12*) that undergoes oxidative addition by reaction with an halogenated compound (R_1-X) to form compound *B* that reacts with copper acetylide Cu-C≡C-R_2 by transmetallation to lead to compound *C*. It is worth noting that in the particular case of substrate **29**, each successful cross-coupling enriches the naphthalene core and leads to the formation of a less reactive substrate for the next coupling. Thus, it seems reasonable to say that performing the quadruple cross-coupling in *ca*. 30% yield is already satisfying. Compound *C* then undergoes

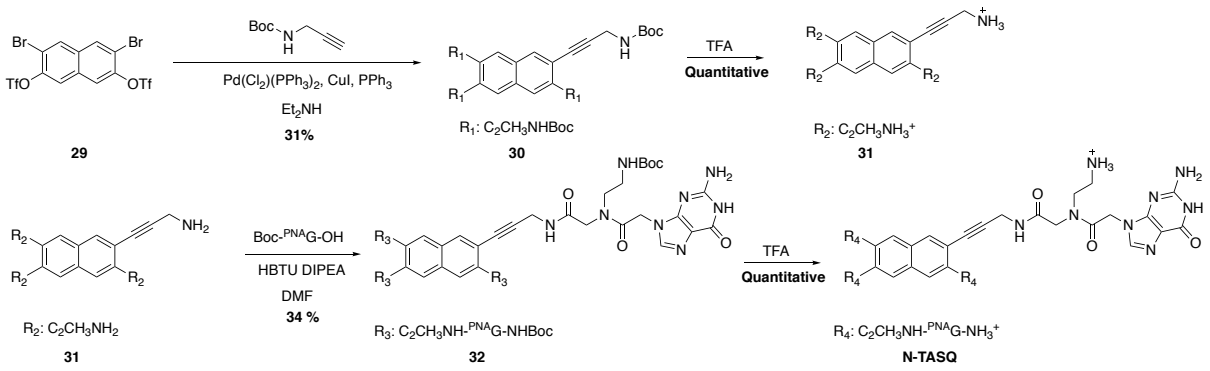
reductive elimination to restore the Pd(0) source and the target molecule R_1-C≡C-R_2. The copper acetylide necessary to create the compound *C* was generated thank to the copper catalytic cycle: the terminal alkyne H-C≡C-R_2 is first complexed by the copper; then the addition of a base or a proton acceptor generates the aforementioned copper acetylide that will undergo trans-metalation and restore the copper-halogen species. The addition of triphenylphosphine might be necessary to complete the coordination sphere of Pd(0) in compound *A*. To improve the reaction conversion and the scale of this quadruple coupling in our case, we focused on two parameters, the palladium source and the base.

In the original **N-TASQ** synthesis *(Scheme 11*), the chosen Pd(II) source was $Pd(PPh_3)_2Cl_2$ that was preactivated for 1 h at 100°C in DMF in the presence of triphenylphosphine to generate $Pd(PPh_3)_4$. The Pd(0) was then reacted directly *via* a drop-wise addition of *N*-Boc-propargylamine under inert atmosphere. We first performed this reaction as described in literature[118], and obtained compound **30** with a yield of 32% after purification by column chromatography and recrystallization. We tried to scale up the process, going from 125 to 500 mg of starting material **29**, which resulted in a decreased yield (25%). We also decided to investigate the role of the starting oxidation state of the catalyst, performing the reaction in the same condition but directly using $Pd(PPh_3)_4$ as a Pd(0) source. Unfortunately, it was not possible to observe any conversion or isolate the desired compound for unknown reasons, indicating that the *in situ* generation of Pd(0) was the better option. In the literature, gram-scale Sonogashira-Heck-Cassar couplings were described using the base in large excess, that is, directly used as a solvent. We thus adapted these conditions: the palladium catalyst and the copper iodide were mixed directly in anhydrous diethylamine with compound **29** and the mixture was heated at 50°C. Propargylamine was then added and the reaction was stirred overnight at 50°C. We firstly performed the reaction on small scale (125 mg) and then over 1 g and obtained compound **30** in 21 and 31% yield respectively, thus demonstrating the suitability of these conditions to larger quantities. Palladium loading was also decreased from 40% in the conditions described in literature to 2.4% mol of $Pd(PPh_3)_2Cl_2$ in the optimized conditions, which thus represented an important gain in Pd consumption.

Table 3: ***Sonogashira-Heck-Cassar coupling conditions explored for the synthesis of 30***

Catalyst	Pd [mol%]	Base	Solvent	Mass of 29 [mg]	Yield [%]
$Pd(PPh_3)_2Cl_2$	40	Et_3N	DMF	125	32
$Pd(PPh_3)_2Cl_2$	40	Et_3N	DMF	500	25
$Pd(PPh_3)_4$	40	Et_3N	DMF	125	0
$Pd(PPh_3)_2Cl_2$	2.4	Et_2NH	-	125	20
$Pd(PPh_3)_2Cl_2$	2.4	Et_2NH	-	250	22
$Pd(PPh_3)_2Cl_2$	2.4	Et_2NH	-	1000	31

While giving satisfying results, these conditions (*i. e*., $Pd(PPh_3)_2Cl_2$, diethylamine as a solvent), still have several issues, notably concerning the monitoring of the reaction (monitoring by TLC and HPLC-MS was unsuccessful). Also, the purification of compound **30** by column chromatography provided a mixture of compounds that required additional precipitation steps, which consequently drastically decreased the yield of the reaction. To improve the isolation of compound **30**, we envisioned that the precipitation of triphenylphosphine oxide prior to column chromatography could potentially improve the isolation process[178]. Another optimization was to change the amide coupling reagents from EDCL/HOBt to HBTU. This modification, which proved effective for the synthesis of **BioCyTASQ**, afforded compound **32** in 34% yield (*Scheme 13*). Unfortunately, a comparison of the efficiency of this step with the original **N-TASQ** synthesis was not possible since the yield for the isolation of compound **32** was not reported. Globally, the synthesis of **N-TASQ** was performed with a chemical yield of *ca*. 11% over 4 steps.

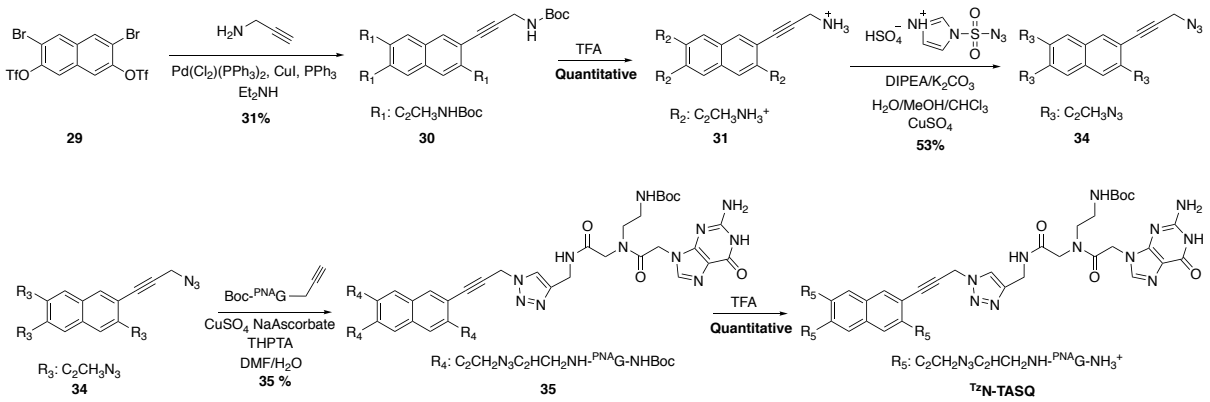


Scheme 13: ***N-TASQ*** *improved synthetic pathway.*

N-TASQ is among the most interesting probes for G4-imaging but its synthesis is still challenging. As a consequence, during the course of this study, a significant effort was invested to improve its synthesis. We believe that, thanks to the aforementioned optimizations, we are on the verge of making **N-TASQ** commercially available or, at least, providing this compound to other laboratories in decent quantities and purity. The improved accessibility of **30** allowed to further experiment with the nature of the spacer, which allowed us for designing a new probe: **^{Tz}N-TASQ.**

V.2. ^{Tz}N-TASQ synthetic pathway:

As shown above, the chemical access to compound **30** was optimized; we then adapted the amide-to-triazole switch made for the design of **BioTriazoTASQ** to design a new turn-on fluorescent probe in which the linkage between the G arms and the naphthalene core is achieved by CuAAC. The resulting, probe was referred to as **^{Tz}N-TASQ**.



*Scheme 14: **^{Tz}N-TASQ** synthetic pathway.*

Compound **34** was prepared *via* a diazotransfer reaction[179] (*Scheme 14*). Imidazole-1-sulfonyl azide was used to convert the primary amine of compound **31** for an azido group (compound **34**). This reaction was performed in biphasic conditions: chloroform was used to solubilize the deprotonated compound **31**, while MeOH to dissolve the diazotransfer agent. This biphasic reaction mixture was stirred vigorously at room temperature for 2 h to lead to compound **34**. This compound was then reacted with the alkynylated ^{PNA}G used for the synthesis of **BioTriazoTASQ** in the presence of copper sulfate, sodium ascorbate and the copper ligand (THPTA). The CuAAC reaction was monitored with HPLC-MS and the crude mixture was

purified by RP-HPLC, which led to the desired compound **35** (global chemical yield: 5.8 %). **^{Tz}N-TASQ** was quantitatively obtained after a final TFA treatment.

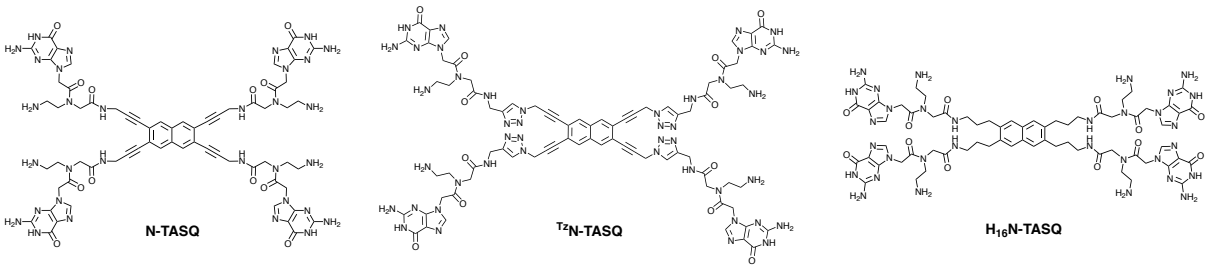


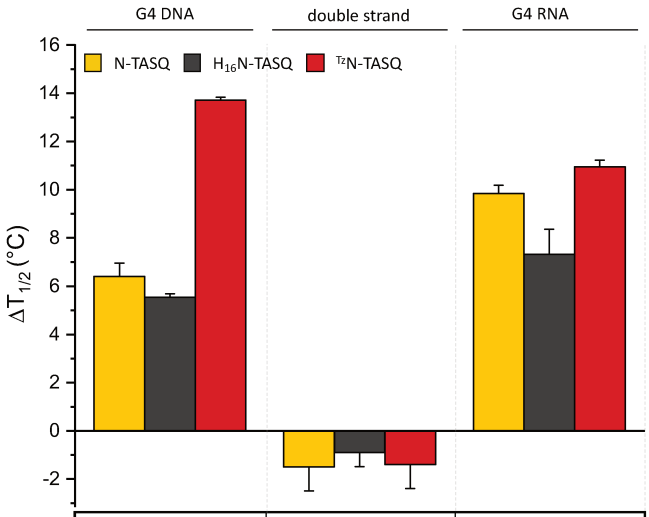
*Figure 48: Structures of **N-TASQ**, **^{Tz}N-TASQ** and **H_{16}N-TASQ**.*

V.3. The G4-interacting properties of N-TASQ, ^{Tz}N-TASQ and H_{16}N-TASQ:

We thus investigated the G4-interacting properties of **^{Tz}N-TASQ** *in vitro*, with a systematic comparison with the parent compound **N-TASQ** along with another probe called **H_{16}N-TASQ** (*Figure 48*) in which the four alkyne connections were hydrogenated. This probe was developed to investigate the role of the triple bonds in the fluorescence properties of the TASQs. The synthesis of **H_{16}N-TASQ** was achieved by Jiri Ledvinka, in the framework of his M1 internship in the Menova's group at IOCB, Prague, CZ. Briefly, the hydrogenation of compound **31** was performed using H_2 and Pd/C (73% yield). the substrate was then reacted with Boc-^{PNA}G under classical peptide coupling conditions (EDCI/HOBt), prior to a final TFA treatment.

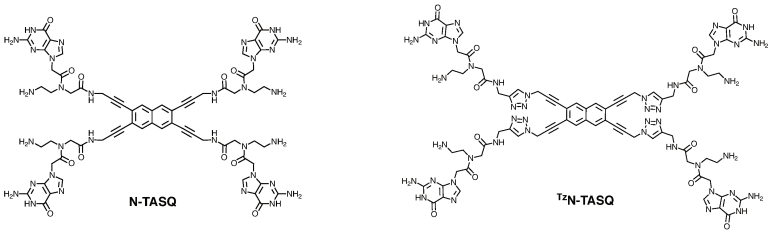
V.3.a FRET Melting:

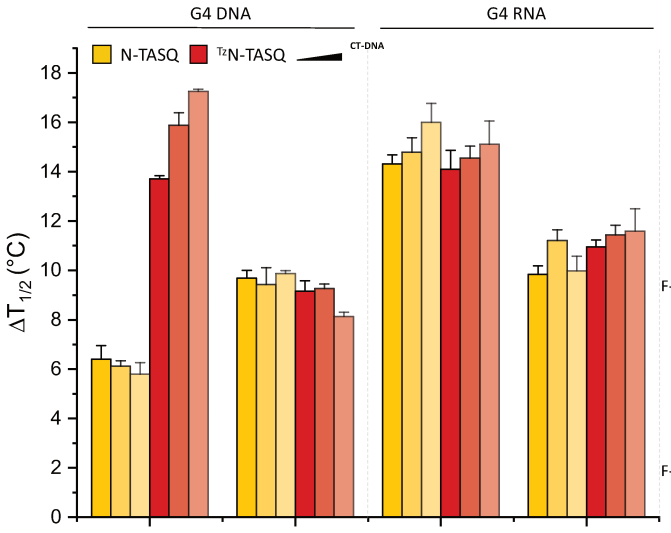
To determine the G4-binding abilities of **N-TASQ**, **^{Tz}N-TASQ** and **H_{16}N-TASQ**, the three TASQs were submitted to competitive FRET-melting experiments against both DNA and RNA G4s in the presence of increasing amounts of CT-DNA (as competitor). As seen in *Figure 49*, these evaluations showed that **H_{16}N-TASQ, ^{Tz}N-TASQ,** and **N-TASQ** efficiently interact with G4s (the G4-DNA F21T, the G4-RNA F-VEGF-T), while they do not stabilize dsDNA (F-duplex-T). It is interesting to note that **^{Tz}N-TASQ** was found 2-fold more efficient than both **N-TASQ** and **H_{16}N-TASQ** for the stabilization of F-21-T ($\Delta T_{1/2}$ = 6.4 ± 0.6, 5.4 ± 0.2 and 13.2 ± 0.2 °C, respectively), likely as a result of the presence of the triazole rings (*vide supra*), while the stabilization of F-VEGF-T by the three ligands stood in the same range (from 7.3 ± 1.1 °C for **H_{16}N-TASQ** to 10.9 ± 0.3 °C for **^{Tz}N-TASQ**).



	F-21-T	F-duplex-T	F-VEGF-T
N-TASQ	6.4 ± 0.6°C	-1.5 ± 0.1°C	9.8 ± 0.4°C
H_{16}N-TASQ	5.4 ± 0.2°C	-0.9 ± 0.6°C	7.3 ± 1.1°C
^{Tz}N-TASQ	13.2 ± 0.2°C	-1.4 ± 1.0°C	10.9 ± 0.3°C

*Figure 49: Melting results ($\Delta T_{1/2}$ °C) for experiments performed with **N-TASQ** (yellow), **H_{16}N-TASQ** dark grey) and **^{Tz}N-TASQ** (red) (1 µM, 5 mol. equiv.) and the G4-DNA F21T, the double strand DNA F-duplex-T and the G4-RNA F-VEGF-T (n=3).*





	F-21-T	F-Myc-T	F-TERRA-T	F-VEGF-T
N-TASQ	6.4 ± 0.6°C	9.3 ± 0.3°C	14.3 ± 0.4°C	9.9± 0.4°C
^{Tz}N-TASQ	13.7 ± 0.1°C	9.2 ± 0.4°C	14.1 ± 0.8°C	10.9 ± 0.3°C

N-TASQ

Eqv CT-DNA	15	50
F-21-T ^{FRET}S	0.96	0.90
F-Myc-T^{FRET}S	0.94	1.06
F-TERRA-T ^{FRET}S	1.03	1.11
F-VEGF-T ^{FRET}S	1.14	1.01

^{Tz}N-TASQ

Eqv CT-DNA	15	50
F-21-T ^{FRET}S	1.12	1.26
F-Myc-T ^{FRET}S	1.01	0.89
F-TERRA-T ^{FRET}S	1.02	1.07
F-VEGF-T ^{FRET}S	1.07	1.06

*Figure 50: FRET-melting results ($\Delta T_{1/2}$, °C) for experiments performed with **N-TASQ** and **^{Tz}N-TASQ** (1 µM, 5 mol. equiv.) with the G4-DNA F-21-T and F-Myc-T, the G4-RNA F-TERRA-T and F-VEGF-T with increasing concentration of CT-DNA (15 and 50 equivalents). Selectivity values (^{FRET}S) calculated for **N-TASQ** and **^{Tz}N-TASQ** (n=3).*

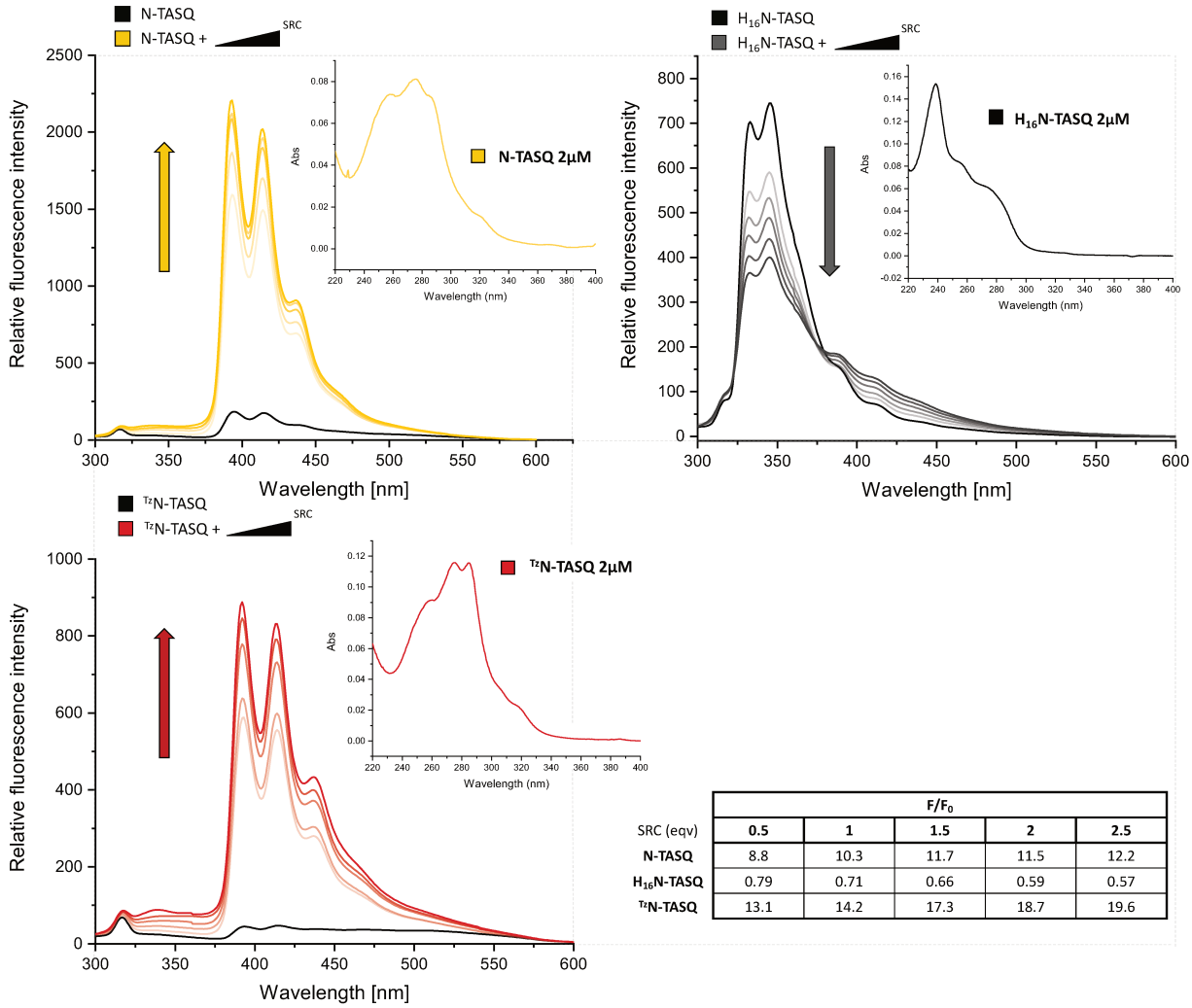
To further investigate the influence of the triazole rings, the performances of both **^{Tz}N-TASQ** and **N-TASQ** were evaluated against other labeled G4s, two DNA (F21T and F-Myc-T) and two RNA (F-TERRA-T and F-VEGF-T), in the presence of increasing amounts of CT-DNA (15 and 50 mol. equiv.). Again, collected results (*Figure 50*) showed that the two TASQs have a good G4-affinity and above all, an excellent selectivity for G4s over dsDNA (with ^{FRET}S >0.94 and 1.01 for both **N-TASQ** and **^{Tz}N-TASQ**, respectively, in the presence of 15 mol. equiv. of CT-DNA).

In conclusion the three probes did effectively bind to G4s, despite a difference in the nature of the G arms; we thus decided to investigate their *in vitro* turn-on fluorescence properties when interacting with G4s.

V.3.b Fluorescence titration:

To this end, each TASQs was dissolved in a K^+-rich buffer and the fluorescence was measured upon excitation at 286 nm (the wavelength used with **N-TASQ**, which actually corresponds to a DNA-mediated excitation of bound TASQs[180]). Then, increasing amounts of the G4-DNA SRC were added (from 0.5 to 2.5 mol. equiv., 0.5 mol. equiv. step) and the fluorescence emission was measured (between 300 and 650 nm). The fold increase of the fluorescence intensity at the maximum of emission (392nm for **N-TASQ**, **^{Tz}N-TASQ**, and 350 nm **H_{16}N-TASQ** respectively) was calculated with respect to the fluorescence of the TASQ alone (F/F_0).

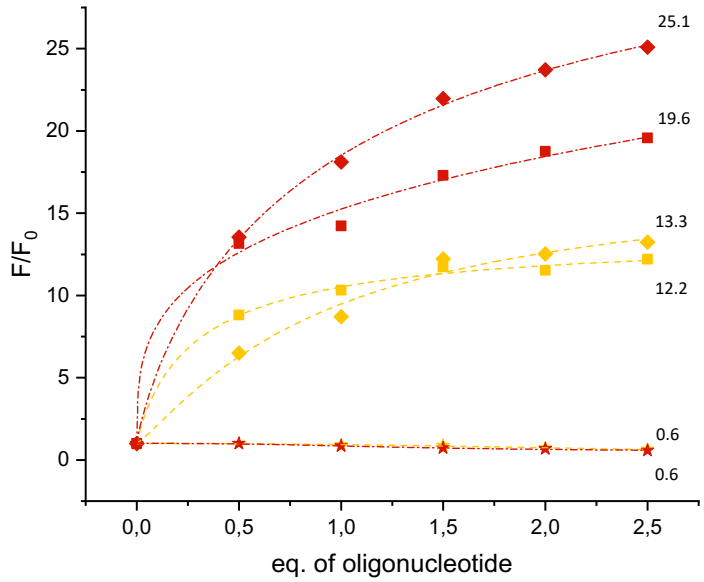
The low residual fluorescence of **N-TASQ** and **^{Tz}N-TASQ** indicated that in both cases the quenching is effective, especially in the case of **^{Tz}N-TASQ** (37 *vs.* 198 a.u. at 392 nm for **N-TASQ** in the same experimental conditions). Upon addition of 0.5 mol. equiv. of G4 only, the fluorescence is turned on for both molecules (*Figure 51*, *ca*. 9- and 13-fold increase in the fluorescence signal for **N-TASQ** and **^{Tz}N-TASQ**, respectively). Of note, the fluorescence signature of both probes presents two equally intense emission peaks at 392 and 414 nm, with a small shoulder at around 437 nm, which is in line with the fluorescence of a **N-TASQ** derivative; furthermore, both probes present an important Stokes shift (116 nm, $\lambda_{max\ abs}$: 286 nm, $\lambda_{max\ em}$: 392 nm).



SRC (eqv)	F/F$_0$				
	0.5	**1**	**1.5**	**2**	**2.5**
N-TASQ	8.8	10.3	11.7	11.5	12.2
H$_{16}$N-TASQ	0.79	0.71	0.66	0.59	0.57
^{Tz}N-TASQ	13.1	14.2	17.3	18.7	19.6

*Figure 51: Fluoresce titration of **N-TASQ** and **^{Tz}N-TASQ** and **H$_{16}$N-TASQ** upon addition of 0, 0.5, 1, 1.5, 2, 2.5 eq of a quadruplex DNA (SRC). Uv-Vis of the three probes (**N-TASQ**, **^{Tz}N-TASQ** and **H$_{16}$N-TASQ**) at 2µM.*

Upon addition of 2.5 mol. equiv. of SRC, the fluorescence of both molecules is further increased (*ca*. 12- and 20-fold), with an overall better performance of **^{Tz}N-TASQ** as compared to **N-TASQ** (*Figure 52*). Interestingly, **^{Tz}N-TASQ** showed both a lower fluorescence in the absence of G4-DNA and a higher relative fluorescence than **N-TASQ** at all G4 concentrations, which is really promising for imaging applications as it could result in a better signal-to-background ratio than **N-TASQ**. The control compound **H$_{16}$N-TASQ** displayed no fluorescence enhancement upon addition of G4s (a decrease was observed that could be attributed to a dilution effect), despite a good G4-affinity. It is worth noting that the level of fluorescence of **H$_{16}$N-TASQ** prior to G4 addition is sensibly higher compared to **N-TASQ** and **^{Tz}N-TASQ** (703 *vs*. 198 and 32 a.u., respectively).



N-TASQ

SRC	Eqv	F/F$_0$	ds17	Eqv	F/F$_0$	VEGF	Eqv	F/F$_0$
	0.5	8.8		0.5	0.99		0.5	6.5
	1	10.3		1	0.91		1	8.7
	1.5	11.7		1.5	0.87		1.5	12.2
	2	11.5		2	0.72		2	12.5
	2.5	12.2		2.5	0.64		2.5	13.2

^{Tz}N-TASQ

SRC	Eqv	F/F$_0$	ds17	Eqv	F/F$_0$	VEGF	Eqv	F/F$_0$
	0.5	13.1		0.5	0.99		0.5	13.6
	1	14.2		1	0.83		1	18.1
	1.5	17.2		1.5	0.71		1.5	22.0
	2	18.7		2	0.67		2	23.7
	2.5	19.6		2.5	0.58		2.5	25.1

*Figure 52: Fluorescence titrations (fold change) of **N-TASQ** (yellow) and **^{Tz}N-TASQ** (red) upon addition of increasing amounts of SRC (square), VEGF (rhombus) and dsDNA (star).*

Preliminary MD studies (performed by Dr. G. Paragi, Univ. of Pecs, Hungary) showed that, in absence of G4, the more flexible alkyl linkers of **H$_{16}$N-TASQ** fail to position the G moieties above the template, that is, in an adequate position to quench the template fluorescence. On the contrary, **N-TASQ**, which comprises more constrained alkynyl linkers, readily adopts a conformation in which G arms perfectly stack above the template, thus enabling an optimal quenching. This result showed how crucial are the alkynyl linkers for the fluorescence properties of the **N-TASQs**, as their replacement by sp^3 carbons greatly improves the flexibility of the molecule. In the presence of a G4, the fluorescence of **N-TASQ** is turned on thanks to a complete redistribution of the G electrons over the intramolecular folded synthetic G-quartet, which both contributes to the stability of the quartet and relieve the naphthalene template from its electronic restraint, making it fluorescent again[172, 181]. We observed a similar behavior with **^{Tz}N-TASQ**, where the introduction of G4s results in an increased fluorescence emission. Altogether, these data highlight the crucial role of the linkers as they participate to the template's quenching in absence of G4s, and allow for the G-quartet to be intramolecularly formed, which relieve the template's fluorescence. It is also worth noting that the two maxima of emission **H$_{16}$N-TASQ** are at lower wavelengths (325 and 350 nm) as compared to **N-TASQ**, indicating that the triple bonds also contribute to extend the conjugation of the π system; we also observed a smaller Stokes shift in comparison with **N-TASQ** and **^{Tz}N-TASQ** (Stokes shift = 111 nm, $\lambda_{max\ abs}$: 239 nm, $\lambda_{max\ em}$: 350 nm).

Similar fluorescence measurements were performed with a G4-RNA (VEGF) and a double-stranded DNA (ds17). Results seen in *Figure 52* indicate that both **N-TASQ** and **^{Tz}N-TASQ** responded to both G4-DNA (up to 12- and 20-fold enhancement) and G4-RNA (up to 13- and 25-fold enhancement) while their fluorescence is not affected by the addition of ds17, thus confirming their excellent selectivity. **N-TASQ** does not discriminate G4-DNA and RNA (12- to 13-fold improvement) while **^{Tz}N-TASQ** exhibited a slightly better response for G4-RNA (25-fold enhancement with VEGF) than for G4-DNA (20-fold enhancement with SRC). The 'preference' of triazole containing TASQ towards RNA is in line with what observed with **BioCyTASQ** and **BioTriazoTASQ.** We hypothesize that the increased accessibility of the upper tetrads of G4-RNAs can be more suited to welcome more sterically hindered TASQs, thus accommodating better **^{Tz}N-TASQ** or **BioTriazoTASQ** compared to **N-TASQ** and **BioCyTASQ** respectively. It is important to remember that the triazole containing TASQs are not selective towards RNA G4s, but interact more efficiently (fluorescence pull-down, fluorescence detection) compared their alkynylated counterparts.

V.4. Cellular imaging N-TASQ *vs* ^{Tz}N-TASQ:

We thus incubated MCF7 cells with **N-TASQ** and **^{Tz}N-TASQ** to directly compare the two probes in both live and fixed cells. By chemical design, **N-TASQ** and **^{Tz}N-TASQ** absorbs light below 320nm, this restriction precludes the use of confocal laser scanning microscope (CLSM) as they are commonly equipped with lasers adjusted at longer wavelengths (usually at 408, 488 and 555nm, the blue, green and red channels respectively). Nonetheless, it was possible to use CLSM thanks to the fact that **N-TASQ** behaves as a red edge probe[119, 182]. The red edge effect (REE), known since 1970[183-184], primarily concerns material sciences since highly organized media (vitrified glasses, dense polymeric matrices, etc.) are particularly prone to this optical feature and its caused by a conformation 'freezing' of the probe in highly crowded environment. The distinctive signature of REE is characterized by a dependence of the wavelength of the emission on the wavelength of excitation, $\lambda_{em\ max} = f(\lambda_{ex})$[185] (*Figure 53*). This phenomenon, which has been thoroughly investigated with **N-TASQ** by a former PhD student Aurélien Laguerre[119], allows its use at higher wavelength (λ_{ex}: 488 nm, λ_{em}: 535- 613 nm). The REE, while advantageous in our hands, can also be problematic for co-incubation studies:

since virtually every irradiation window can be used to excite **N-TASQ**, there is no channels dedicated to these probes only, which makes co-localization studies difficult and requiring a careful calibration of the experimental conditions (*Figure 47*).

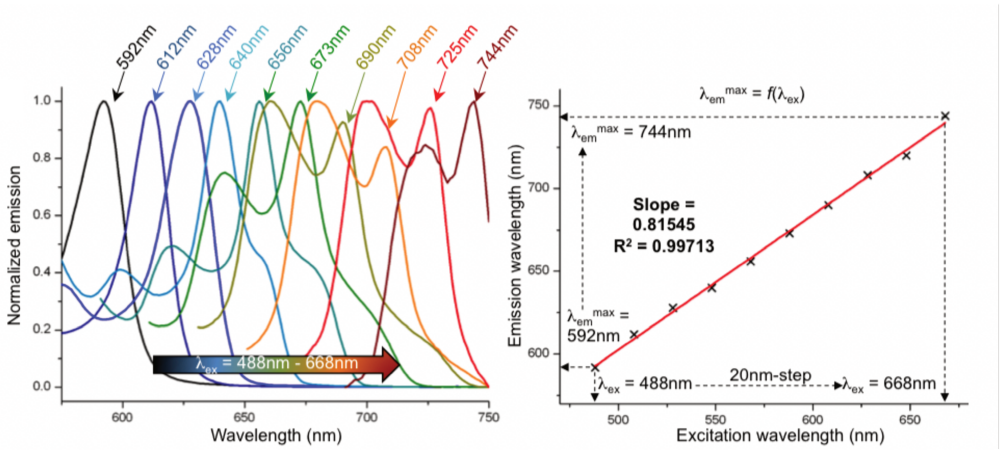


*Figure 53: Dependence of the emission maximum ($\lambda_{em\ max}$) on the excitation wavelength (λ_{ex}): fluorescence spectra recorded for an experiment carried out with **N-TASQ** (10µM) and TERRA (5µM) with λ_{ex} between 488 and 668nm (every 20nm)[119].*

Both probes were thus incubated for 3 h in fixed and permeabilized MCF7 cells (ice cold methanol) at a concentration of 50 μM for **N-TASQ** and 100 μM for **^{Tz}N-TASQ** (*Figure 54*). The overall quality of these images was quite low; however, the distribution of both probes was found to be similar with an accumulation of fluorescence in the nucleoli (white arrows) and in the perinuclear region (red arrows) (λ_{ex} = 405 nm, λ_{em} = 450- 530 nm). Both **N-TASQ** and **^{Tz}N-TASQ** proved to be effective, with **N-TASQ** showing a less diffuse staining.

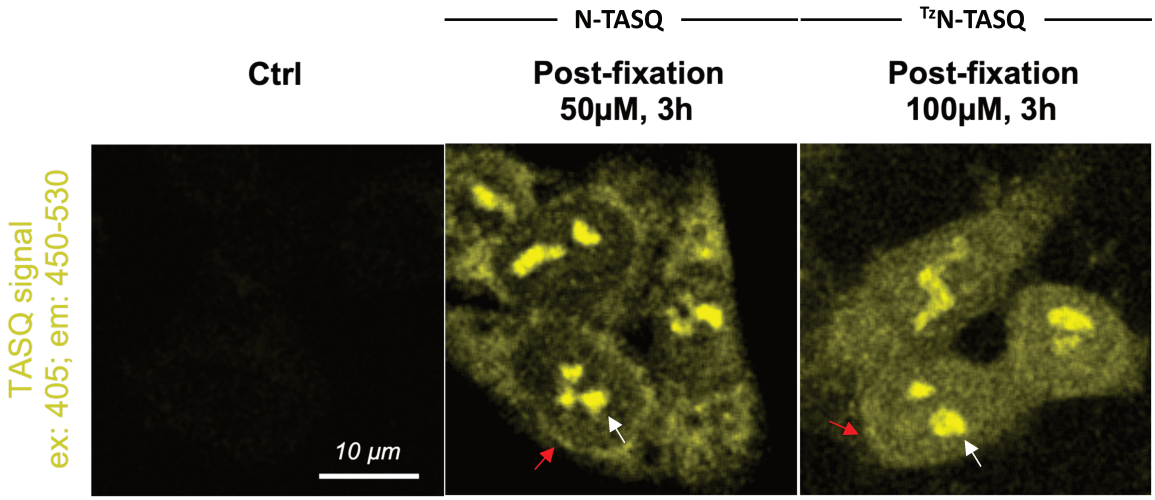


*Figure 54: MCF7 cells are fixed either with ice cold methanol and then incubated with **N-TASQ** (50µM) **^{Tz}N-TASQ** (100µM) for 3 hours before mounting steps with Fluoromount. Confocal analyses carried out with lasers. White arrows indicate some of the most representative nucleoli staining sites, in red the perinuclear region.*

From these preliminary images, we can only conclude that **^{Tz}N-TASQ** behaves similarly to **N-TASQ**. Live-cell incubation will also be soon investigated to highlight if any difference between the two probes can be found. Finally, the *in vitro* results that show a higher fluorescence enhancement with G4-RNA will also be investigated in cells, evaluating how RNase and DNase treatments impact the emission of the two probes.

V.5. Conclusions:

While the fluorescence turn-on properties of **N-TASQ** both *in vitro* and *in vivo* have been thoroughly exploited, its chemical accessibility was particularly complicated, which has strongly limited its exploitation. Here, we have presented optimized experimental conditions that improved the global chemical yield of the key step of the synthesis, which has made larger scale production possible. We also designed and synthesized a new version of **N-TASQ** referred to as **^{Tz}N-TASQ**, in which the G arms were connected to the central naphthalene template by CuAAC. **^{Tz}N-TASQ** showed better turn-on fluorescence properties than **N-TASQ** upon G4 addition, with a slight preference for G4-RNA. Both molecules were effectively used as post-fixation cell imaging probes, offering a direct visualization of G4s in cells. Unfortunately, both **N-TASQ** and **^{Tz}N-TASQ** cannot be conveniently used for co-incubation studies (*e.g.,* with concomitant immunodetection) as both of them are red-edge probes, which have intrinsic limitations. To tackle this issue, we devised the synthesis of new G4-probes (**SquareTASQ** series), which will be discussed in *Chapter 6*.

Chapter 6:

SquareTASQs

While **N-TASQ** and **^{Tz}N-TASQ** exhibit very desirable selectivity and fluorescence turn-on properties, they both suffer from two major drawbacks: *i)* it is necessary to irradiate in the UV region (286 nm) to observe their fluorescence *in vitro*, which limits their optical imaging application (due to high cellular autofluorescence at this wavelength); and *ii)* they display a red edge effect[182], which results in a very broad emission window (in the blue, green and red microscope channels, although with different intensity), which makes colocalization studies with other markers difficult. We therefore envisioned the development of a more classical G4 probe with a fluorescence emission closer to the near infra-red (NIR) region[186], which could be used to perform colocalization studies. Excitation close to the NIR window should also greatly decrease the background signal due to cellular autofluorescence, thus improving the signal-to-noise ratio. Ideally an excellent fluorescent probe should also have a big Stokes shift that indicate a low tendency to self-quenching trough energy transfer[187] and high quantum yield that ensures a good ratio between the number of photons absorbed and emitted[188]. To achieve this, we designed a new probe in which the central template *i*) is easy to synthesize, *ii)* absorbs and emits light close to the NIR region, and *iii)* can be tetra-functionalized with a pseudo-C_4 symmetry. Our new design was based on a squaraine core (λ_{em} around 640 nm)[189]. Two main types of fluorescent squaraines are described in literature, which are either aniline-derived or heterocyclic derivatives (*Figure 55*)[190]. While the second one is often considered as the best choice owing to its improved stability and high fluorescence efficiency[191], its heterocyclic chemical core can be hardly derivatized. On the other hand, the aniline moiety of the first category can be easily alkylated with the moiety of our choice, which makes aniline-based squaraines the best candidates for our new family of TASQs.

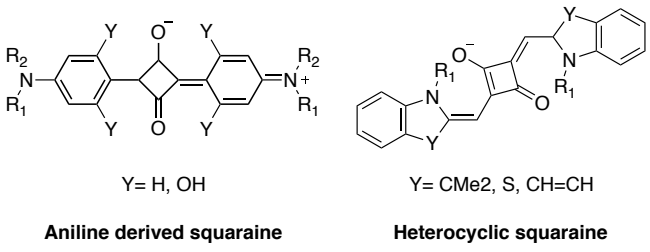
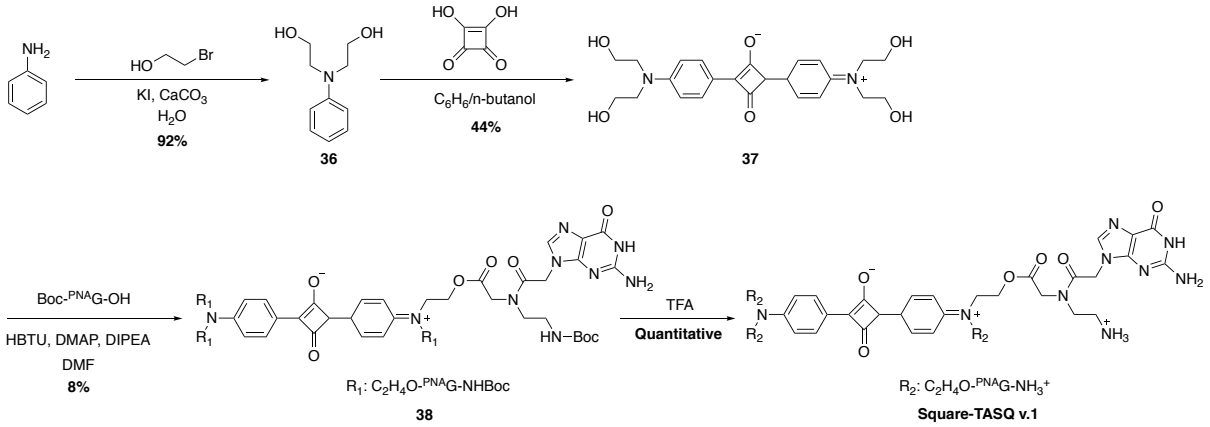


Figure 55: Aniline-derived and heterocyclic squaraines structures

VI.1. Square-TASQ v.1 synthetic pathway:

The first **Square-TASQ (v.1)** was developed thanks to a straightforward approach (*Scheme 15*) in which the aniline core was bis-functionalized with 2-bromoethanol to obtain the compound **36,** with a satisfactory 92% chemical yield by heating the reaction at reflux overnight[192]. This aniline was then reacted with squaric acid using a Dean-Stark apparatus to obtain the compound **37**[193], which conveniently precipitated out of the reaction mixture upon cooling to room temperature (44% yield). After filtration, compound **37** was coupled with ^{PNA}G through a Steglich esterification, this reaction mimics the classical amide coupling using HBTU as coupling reagent in basic conditions (DIPEA) with the addition of 4-dimethylaminopyridine (DMAP) as a more reactive leaving group (8% yield)[194]. After isolation of the desired compound **38** by RP-HPLC, the desired **Square-TASQ v.1** was obtained with a TFA treatment, with a global yield of 3.3% over four steps (not optimized).



*Scheme 15: **SquareTASQ v.1** synthetic pathway.*

VI.2. The G4 interacting properties of Square-TASQ v.1:

To determine the G4-interacting properties of this probe, we investigated its ability to stabilize G4s by FRET-melting experiments. To our surprise, this probe failed in stabilizing G4s: we noted a strong discoloration after every cooling/heating cycle necessary to defrost the probes, that we ascribed to a high chemical and thermal instability. In the literature, the use of similarly constructed squaraine is reported as colorimetric indicator for thiols[189], which attack the electron-poor squaraine core, resulting in a probe discoloration. We thus attributed the lack of G4-thermal stabilization to the chemical instability of the probe during the experiments and set out to rethink the SquareTASQ design.

VII.3. Square-TASQ v.2 synthetic pathway:

We identified two possible weak points in the design of **Square-TASQ v.1**: *i)* the electron poor central template (which could be subjected to nucleophilic attack), and *ii)* the ester linkage between the G arms and the functionalized aniline, which could be hydrolyzed. We thus decided to tackle the two issues concomitantly, introducing four electron-donor group on the aromatic core (hydroxyl groups) and changing the Steglich esterification for a CuAAC reaction to obtain more stable triazole linkages. The introduction of four hydroxyl group aims at shielding the central template against nucleophilic attacks by increasing both its electronic density and steric hindrance, as already demonstrated in literature with a similar squaraine derivative that remains stable even in the presence of an excess of thiols[195]. The triazole linkage will be here used as bioisosters of amide bonds (as above), allowing for exploiting both the efficiency and orthogonality of click chemistry (*Figure 56*).

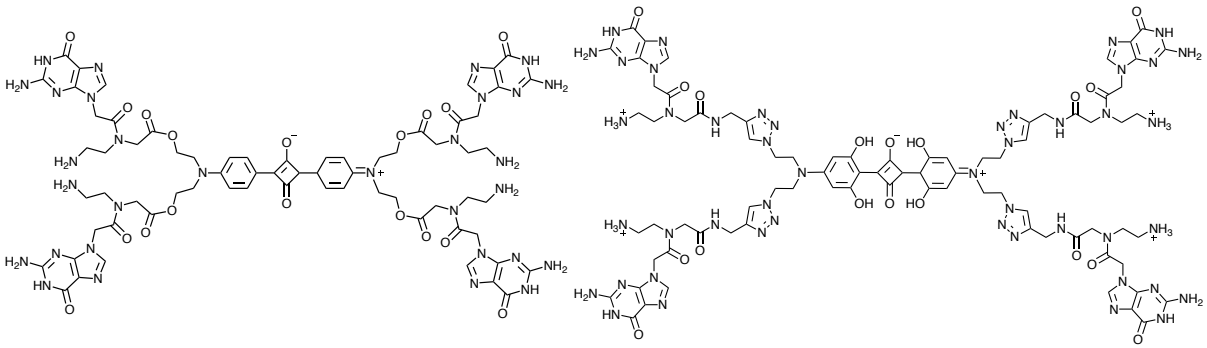
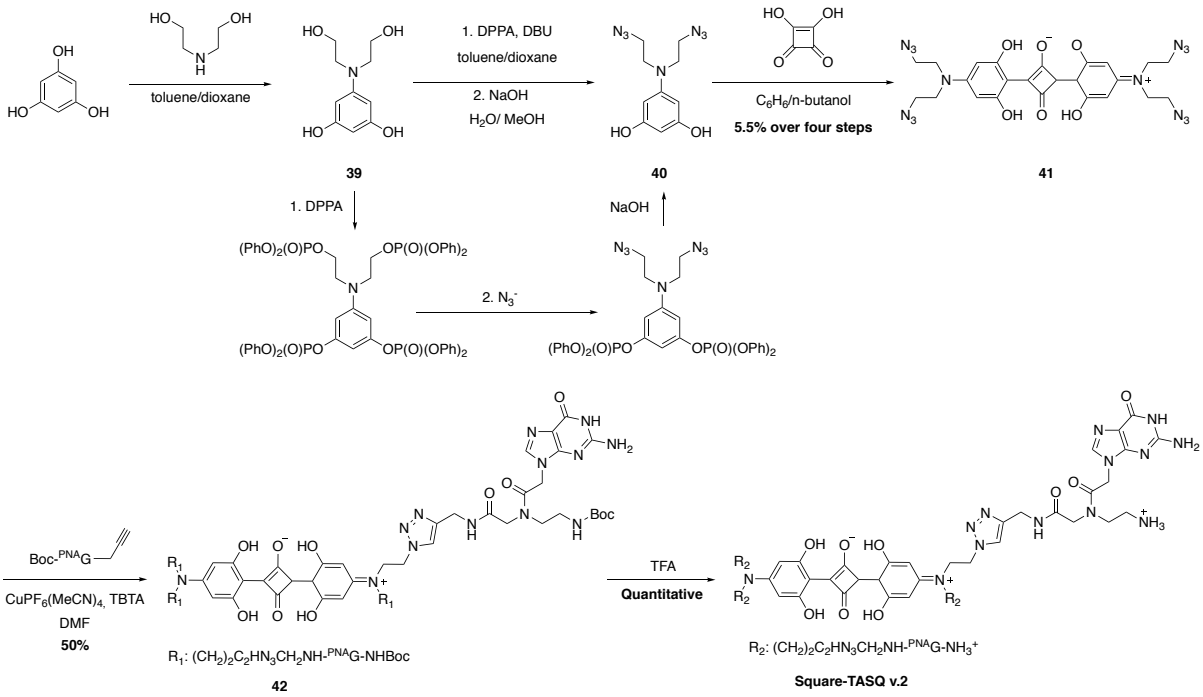


*Figure 56: **Square-TASQ v.1** and **Square-TASQ v.2** structure.*

The synthesis of this new Square-TASQ, referred to as **Square-TASQ v.2**, relied on the use of phloroglucinol as starting material (*Scheme 16*). The aniline derivative **39** was prepared by heating at reflux the mixture of toluene and dioxane overnight[196]; then, the crude aniline was successfully converted (on a test scale) to the desired azide-functionalized aniline **40** using diphenylphosphoryl azide (DPPA) and DBU as a base[197]. This procure relies on the reaction of the four hydroxyl groups with diphenylphosphoryl azide and the regiospecific nucleophilic attack of the released azides on the aliphatic phosphoryl esters. Finally, a basic treatment with sodium hydroxide hydrolyses the phenol esters to provide **40**. The crude aniline **40** was then reacted with squaric acid under conditions classically used for squaraine formation (Dean-Stark distillation). The target compound **41** precipitated out from the reaction mixture upon cooling to room temperature (5.5 % yield over 3 steps). The precipitate was then reacted

with the alkynylated ^{PNA}G arms used previously (cf. Chapters 3 and 6): the conditions used here were directly inspired by those implemented for the synthesis of both **TriazoTASQ** and **BioTriazoTASQ**, *i.e.*, with $(CH_3CN)_4CuPF_6$ as a copper source and TBTA as a copper ligand. After RP-HPLC purification, compound **42** was obtained with a 50% chemical yield. After a final TFA treatment, the desired probe **Square-TASQ v.2** was obtained with a global chemical yield of 2.8% over five steps (not optimized).



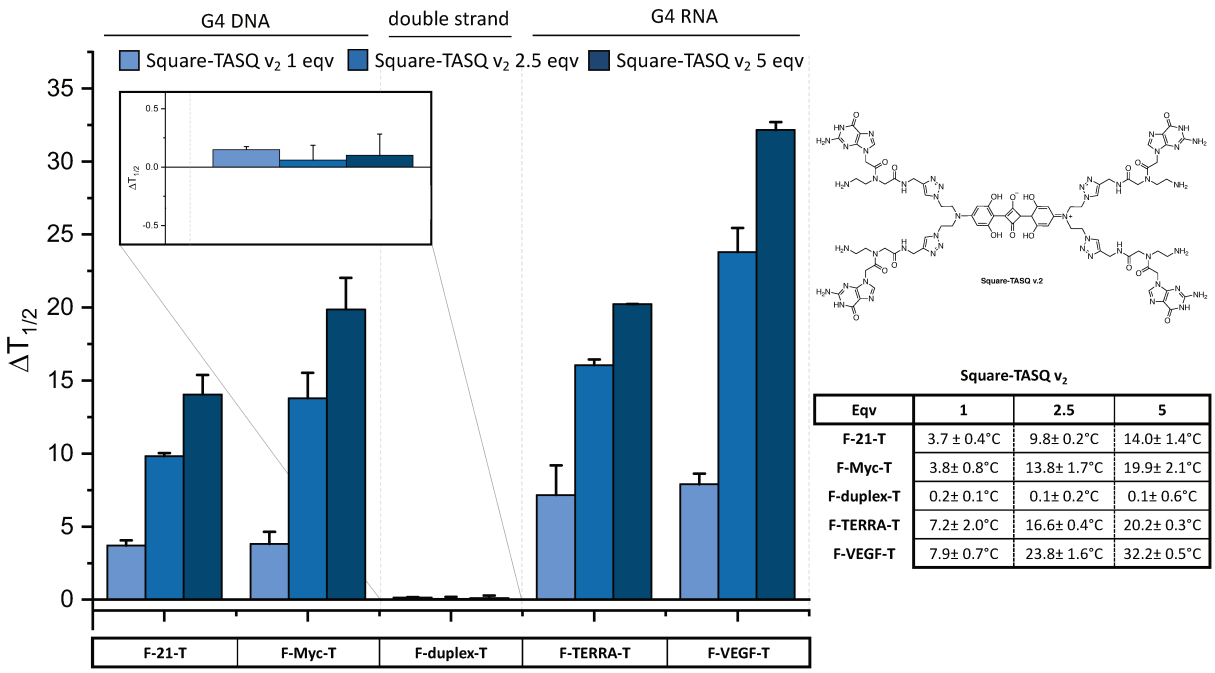
*Scheme 16: **SquareTASQ v.2** synthetic pathway.*

VI.4. The G4-interacting properties of Square-TASQ v.2:

VI.4.a FRET-Melting experiments:

To investigate whether **Square-TASQ v.2** is a valuable G4-ligand, we performed a series of FRET-melting experiments with both DNA (F-21-T, F-Myc-T) and RNA QFS (F-TERRA-T, F-VEGF-T), as well as dsDNA (F-duplex-T) as control. To our great surprise, **Square-TASQ v.2** showed very high $\Delta T_{1/2}$ (up to >30 °C), making it the TASQ with the highest $\Delta T_{1/2}$ values observed to date. These results also indicated that the new design clearly tackled the issue faced with **Square-TASQ v.1**. In more details, **Square-TASQ v.2** (5 mol. equiv.) raised the melting temperature by 14.0 ± 1.4 and 19.9 ± 2.1 °C with the G4-DNA F-21-T and F-Myc-T, respectively, clearly outperforming both **BioCyTASQ** and **^{Tz}N-TASQ** (by more than 10°C). With the G4-RNA

F-TERRA-T and F-VEGF-T, **Square-TASQ v.2** reached the highest melting temperatures of all TASQs ($\Delta T_{1/2}$ = 20.2 ± 0.3 and 32.2 ± 0.5 °C, respectively), which indicated a very strong G4 apparent affinity. With the double strand control F-duplex-T, **Square-TASQ v.2** showed no stabilization, indicating a very high level of selectivity towards G4s. These values, while extremely satisfactory, are difficult to rationalize and further experiments are now required to fully understand the origins of such an efficient G4-interaction.

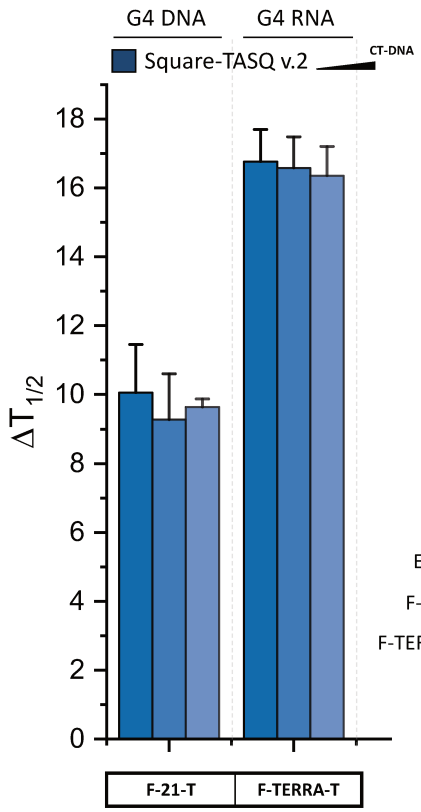


Square-TASQ v_2

Eqv	1	2.5	5
F-21-T	3.7 ± 0.4°C	9.8± 0.2°C	14.0± 1.4°C
F-Myc-T	3.8± 0.8°C	13.8± 1.7°C	19.9± 2.1°C
F-duplex-T	0.2± 0.1°C	0.1± 0.2°C	0.1± 0.6°C
F-TERRA-T	7.2± 2.0°C	16.6± 0.4°C	20.2± 0.3°C
F-VEGF-T	7.9± 0.7°C	23.8± 1.6°C	32.2± 0.5°C

*Figure 57: FRET-melting results ($\Delta T_{1/2}$ °C) for experiments performed with **Square-TASQ v.2** with increasing amount of probe: 0.2 µM, 1 mo. Equiv., 0.5 µM, 2.5 mol. equiv and 1 µM, 5 mol. equiv. and the G4-DNA F21T, F-Myc-T, the double strand DNA F-duplex-T, the G4-RNA F-TERRA-T and F-VEGF-T (n=6).*

To further investigate the G4-stabilization ability of **Square-TASQ v.2**, we also performed FRET-melting assay with different concentrations of the probe, the results seen in *Figure 57* show a good a dose-response pattern. This excellent affinity was in line with an excellent selectivity; indeed, the G4-stabilization of **Square-TASQ v.2** was unaffected by the presence of an excess of unlabeled CT-DNA (15 and 50 mol. equiv., *Figure 58*), with ^{FRET}S values ranging from 0.92 to 0.98. These excellent results thus led us to investigate the fluorescence properties **Square-TASQ v.2** upon G4-binding.



Square-TASQ v.2		
Eqv CT-DNA	15	50
F-21-T ^{FRET}S	0.92	0.96
F-TERRA-T ^{FRET}S	0.98	0.97

*Figure 58: Competitive FRET-melting results ($\Delta T_{1/2}$, °C) for experiments performed with **Square-TASQ v.2** (0.5 µM, 2.5 mol. equiv.) with the G4-DNA F-21-T, the G4-RNA F-TERRA-T with increasing concentration of CT-DNA (15 and 50 equivalents). Selectivity values (^{FRET}S) calculated for **Square-TASQ v.2** (n=3).*

VI.4.b Fluoresce titration:

Similarly to **N-TASQ** and **^{Tz}N-TASQ, Square-TASQ v.2** emission spectra was measured for the probe alone in K^+ rich buffered solution (cacodylate buffer 100 mM K^+, pH: 7.2) and then, upon addition of increasing amount of oligonucleotides (G4-DNA and RNA, and dsDNA). The probe was excited at 620 nm and the emission spectra measured from 630 to 750 nm. We decided to irradiate at this specific wavelength since, as it will be discussed later on, this wavelength corresponds to a clear isosbestic point, which allows for ruling out different absorption contributions to the modifications of the fluorescence emission observed. The emission of **Square-TASQ v.2** alone in solution (2µM) was rather low (*ca*. 50 a.u. at 647 nm), which is an ideal feature for a turn-on probe: we interpreted this by the tendency of squaraine to self-aggregate, which quenches its fluorescence, but also the presence of the G arms that could also quench the squaraine fluorescence by electron or energy transfer. The exact nature of the non-radiative deactivation is here not fully elucidated, and further studies are already ongoing to better comprehend it.

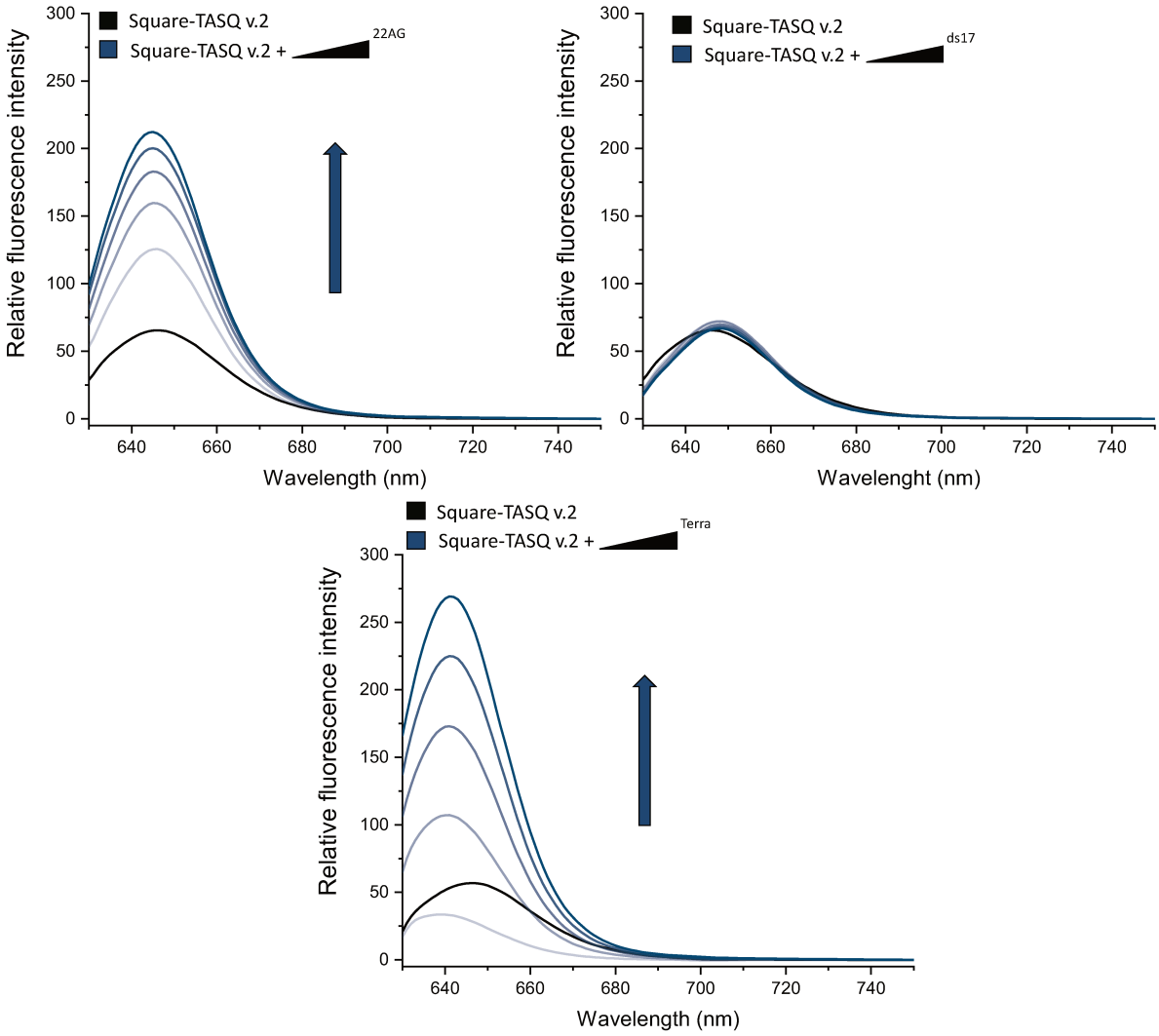
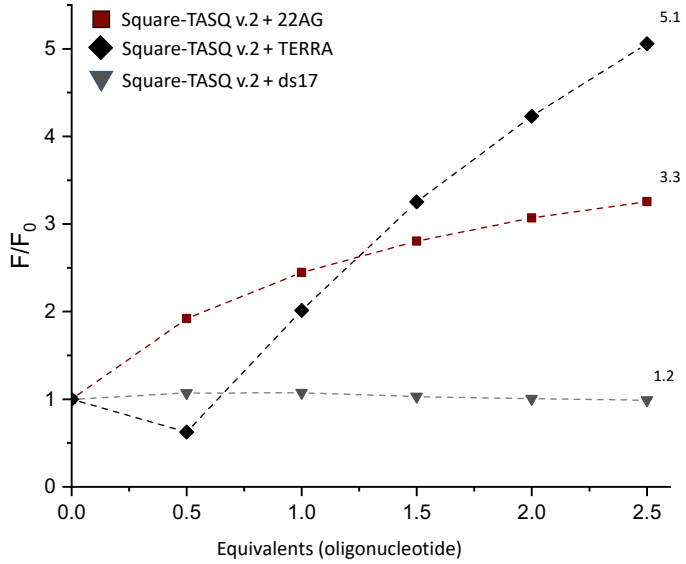


*Figure 59: Fluorescence (λ_{ex}: 620 nm) titration of **Square-TASQ v.2** (2 µM) upon addition of 0, 0.5, 1, 1.5, 2, 2.5 eq of a quadruplex DNA (22AG), double helix hairpin (ds17) and quadruplex RNA (TERRA).*

To investigate possible aggregation formation, the impact of the solvent of experiments (DMSO *vs*. H_2O) on the probe absorption spectra will be discussed hereafter (§VI.4.c). When both DNA and RNA G4s were added, an enhancement of the fluorescence emission was observed: the addition of increasing amounts of 22AG triggered an increase of **Square-TASQ v.2**'s fluorescence in a dose-response manner, reaching *ca*. 200 a.u. in the presence of 2.5 mol. equiv. The shape of the fluorescence band is marginally affected by the addition of the G4, with a small (2 nm) blue shift of its maximun (λ_{em} = 645 nm). Even if very small, this shift might indicate a reorganization of the probe in the presence of the G4-target, likely a G4-mediated disaggregation-based fluorescence. The addition of TERRA resulted in a similar observation, with a maximum intensity of fluorescence of 269 a.u. at 641 nm. The small hypsochromic shift (6 nm) between the emission of the probe alone and the **Square-TASQ**

v.2/RNA complex might indicate again a reorganization of the probe upon G4-addition. In contrast, the addition of dsDNA did not trigger any significant modification in the **Square-TASQ v.2'**s fluorescence, highlighting that the enhancement of the fluorescence emission is G4-specific.

We cannot now rationalize how and why the fluorescence of this new probe increases upon addition of G4s and further photophysical investigations (along with MD studies) are needed to completely understand the properties of **Square-TASQ v.2**. Two possibilities can be hypothesized: on the one hand, if we consider the presence of non-fluorescent squaraine aggregates, the formation of TASQ/squaraine complex upon addition of G4 might result in the probe disaggregation, which ultimately results in an increase of its fluorescence level. On the other hand, the fact that the fluorescence increases upon formation of the conformationally constrained **Square-TASQ v.2**/G4 complex might indicate that free guanines could be the source of fluorescence quenching. The two hypotheses are not mutually exclusive but further investigations are needed to decipher their actual contributions.



Triazo-TASQ V_2

22AG	Eqv	F/F_0	ds17	Eqv	F/F_0	TERRA	Eqv	F/F_0
	0.5	1.92		0.5	1.26		0.5	0.57
	1	2.44		1	1.27		1	2.01
	1.5	2.80		1.5	1.23		1.5	3.25
	2	3.07		2	1.20		2	4.23
	2.5	3.26		2.5	1.18		2.5	5.06

*Figure 60: Fluorescence titrations (fold change) of **Square-TASQ v.2** (λ_{ex}: 620 nm, λ_{em}: 640 nm) upon addition of increasing amounts of 22AG (wine square), TERRA (black rhombus) and dsDNA (grey triangle).*

Even though the fluorescent variations were relatively low as compared to **N-TASQ** and **^{Tz}N-TASQ**, the F/F_0 values obtained upon addition of G4/dsDNA (*Figure 60*) indicate its exquisite specificity for G4s over dsDNA that, along with a fluorescence emission in the NIR (*ca*. 645

nm), make **Square-TASQ v.2** a promising probe for the direct visualization of G4s in cells, furthermore compatible with colocalization studies.

VI.4.c Aggregation studies:

To investigate the possible aggregation of this probe in solution, we recorded the absorption spectra of **Square-TASQ v.2** in an aqueous buffer (with and without G4) and in the highly dissociating solvent DMSO[198-199]. Our goal was to evaluate if a/ the UV-Vis signature of the TASQ is different in these two conditions, and b/ the addition of a G4 triggers a similar UV-Vis spectrum redistribution.

As seen in *Figure 61*, **Square-TASQ v.2** alone in a K^+-rich buffer displays a broad absorption spectrum with a peak centred at 641 nm and a shoulder around 590 nm; in DMSO, the shape changes drastically with a sharper peak centred at 641 nm. The intensity of the signals is also drastically different: 0.08 *vs*. 0.34 a.u. at 641 nm in water and organic solvent, respectively.

Broad spectra with shoulders are often associated with the presence of aggregates[200]. Upon addition of the G4 SRC in the aqueous buffer, the absorption spectrum drastically changes, with the decrease of the shoulder at 590 nm: this is particularly observable with normalized spectra, which show that both DMSO and G4-containing conditions provide superimposable spectra. This corroborates the hypothesis according to which a/ the probe is at least partially aggregated with free in solution, and b/ the addition of G4 triggers the dye disaggregation.

When recorded as a function of the time, the UV-Vis signature of **Square-TASQ v.2** (2 µM) in aqueous buffer indicated that the peak that likely corresponds to the monomeric form (at 641 nm) is preserved after 1h, while that that likely corresponds to the aggregate (at 590 nm) increases during 20 min. This rudimentary kinetics study seems to corroborate the hypothesis according to which **Square-TASQ v.2 slowly** aggregates over time when free in solution, which has an impact on its spectroscopic properties.

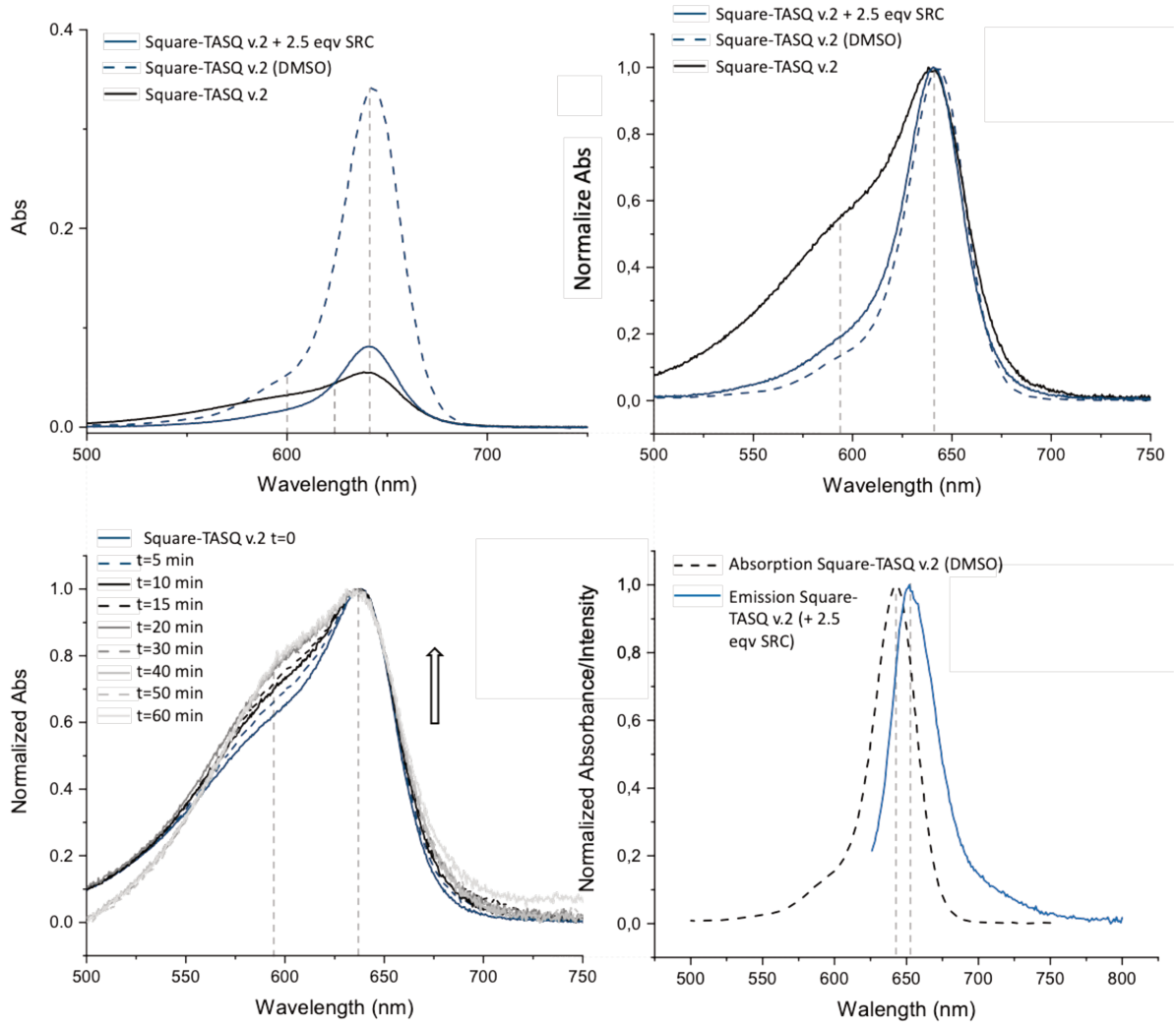


*Figure 61: i) Absorption spectra of **Square-TASQ v.2** (2µM) in cacodylate buffer 100 mM K⁺, pH: 7.2 (black), with the addition of 2.5 equiv. of SRC (blue) and in DMSO (blue dashed). ii) Normalized emission for **Square-TASQ v.2 (2µM)** in the three aforementioned conditions: cacodylate buffer without G4 (black) cacodylate buffer with G4 (blue) and DMSO (dashed blue). iii) Normalized absorption spectra as function of time (0, 5, 10,15,20,30,40,50 and 60 minutes) of **Square-TASQ v.2** (2µM) in cacodylate buffer 100 mM K⁺, pH: 7.2. iv) Normalized absorption spectra in DMSO (dashed black line) and normalized emission upon G4 addition (2.5 equiv. SRC, $\lambda_{ex:}$ 580 nm) (Blue).*

VI.4.d Quantum yield calculation:

Next, we decided to calculate the quantum yield of **Square-TASQ v.2** in the presence of G4s. The fluorescence of the probe (2 µM) in cacodylate buffer (100 mM K^+, pH: 7.2) was measured upon irradiation at 580 nm, selected for a direct comparison between the **Square-TASQ v.2**'s fluorescence and the reference compound **SulfoCyanine5** dye one (*Figure 63*)[201]. The spectra were recorded before and after the addition of the G4 SRC (2.5 mol. equiv.), which triggered a 7.7-fold increase in fluorescence emission. The UV-vis spectra of **Square-TASQ v.2** is modified upon addition of the G4, which again suggests a G4-mediated disaggregation.

Absorption and emission spectra were then measured at different concentrations ranging from 2 to 1 μM to calculate quantum yield (*Figure 62)*.

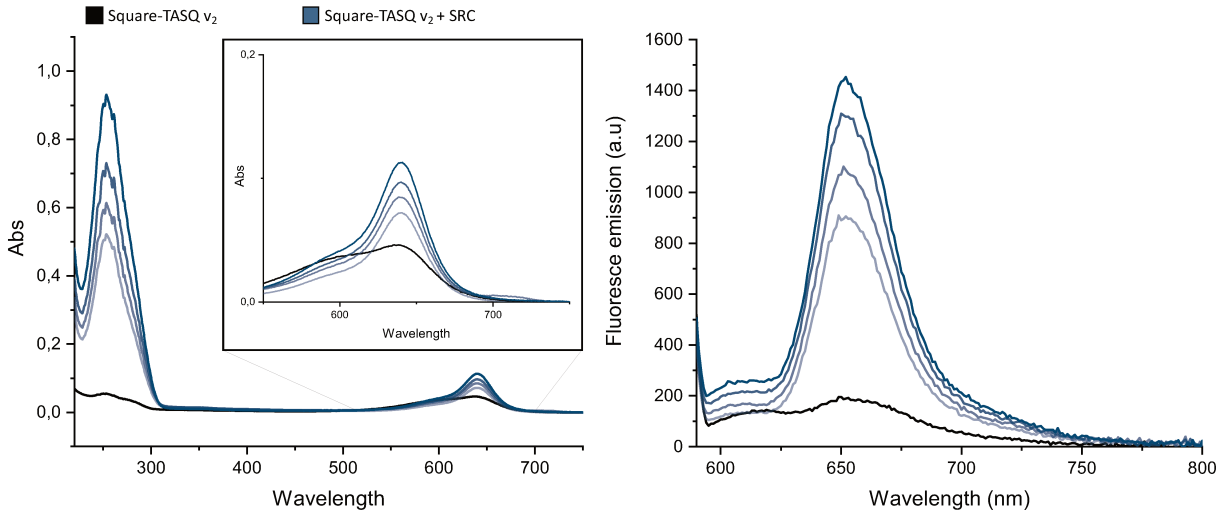


*Figure 62:Absorption and emission spectra (λ_{ex}: 580 nm) of **Square-TASQ v.2** alone and of a dilution of the **Square-TASQ v.2/G4** assembly with 2.5 equivalents of SRC (from 2µM to 1µM).*

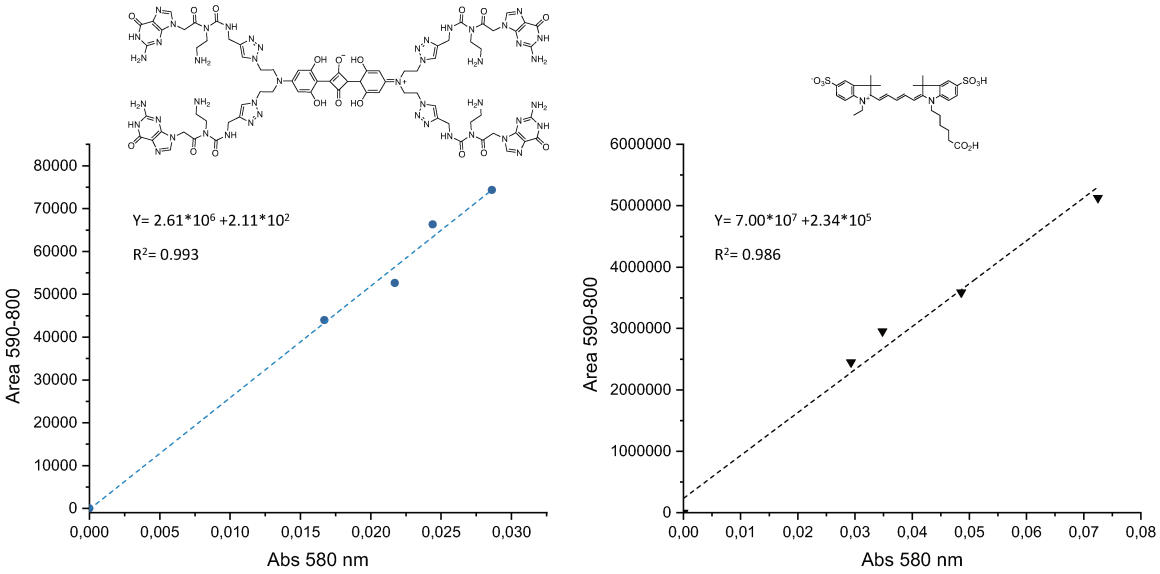


*Figure 63:Absorption at 580 nm plotted against the area emission curve from 590 to 800nm for **Square-TASQ v.2**/G4 assembly and **SulfoCy5** dye.*

The same titration (absorption and fluorescence) was performed with **SulfoCy5** (kindly provided by Pr. Antony Romieu): a 1 mg/mL stock solution of the probe was diluted in PBS and absorption and emission spectra were recorded (irradiation at 580 nm). To calculate the quantum yield, the integrals of the emission spectra of each probe were plotted against their

respective absorption values at the 580 nm. The slopes of the two different curves (*Figure 63*) were compared to the photoefficency of the cyanine dye (20%) and led to a quantum yield for the **Square-TASQ v.2**/G4 of around 1%. Despite having a very modest efficiency when compared to other fluorophores[188], we decided to investigate anyway whether **Square-TASQ v.2** could be used in cells.

VI.5. Cell imaging:

Preliminary cell-based studies were performed with MCF7 cells: **Square-TASQ v.2** (20 µM) was incubated for 1 h in fixed cells and the cells were then imaged by irradiation at 638 nm (λ_{em} from 643 to 722 nm). DAPI staining was also performed to label the nuclei.

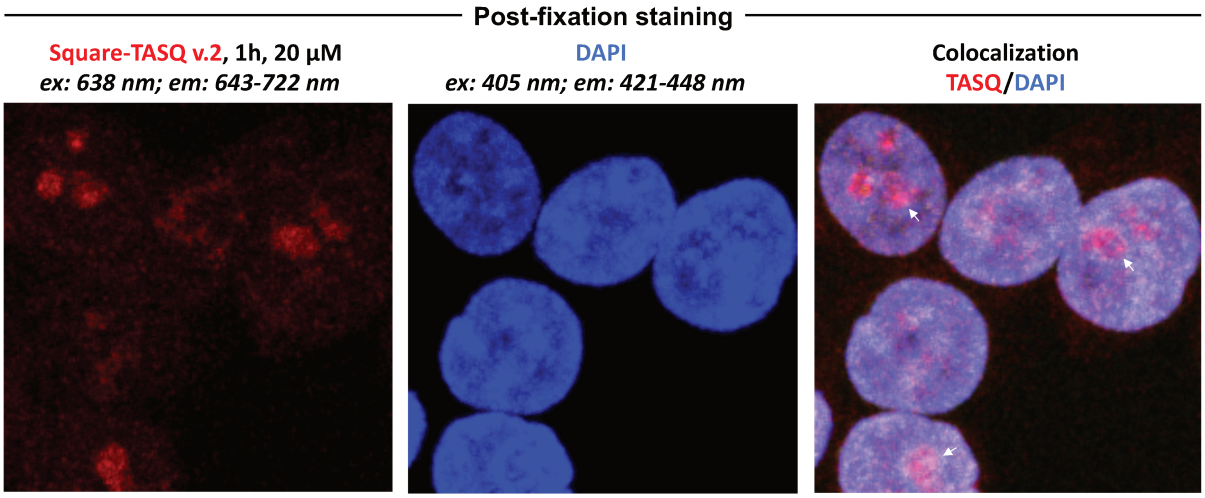


*Figure 64: MCF7 cells are fixed with cold methanol and incubated with **Square-TASQ v.2** (20 µM) for 1h before mounting steps in presence of DAPI (Fluoromount-GTM) and confocal analyses carried out with lasers.*

Contrarily to **N-TASQ** and **^{Tz}N-TASQ**, **Square-TASQ v.2** was designed to emit in the red channel. Although the staining was rather weak (due to a very low quantum yield), fluorescence *foci* were observed in the nucleoli region (white arrows), in a manner that is reminiscent to the other TASQs. The co-localization with DAPI did not pose any problem, indicating that **Square-TASQ v.2** did not suffer from the REE limitations. Further studies will be necessary to investigate the full potential of **Square-TASQ v.2** as direct G4-probe and to this end, live-cell experiments performed with **Square-TASQ v.2** (IC_{50} > 100 µM over 72 h) are currently ongoing.

VI.6. Conclusions:

We have focused here on the modulation of the fluorescence properties of TASQs by changing the central naphthalene core of **N-TASQ** and **^{Tz}N-TASQ** for a squaraine template to create the **Square-TASQs**. These TASQs emit close to the NIR region, which is particularly relevant for *in cella* applications in order to benefit from the low level of cell autofluorescence in this region. Two versions of **Square-TASQ** were developed: *i)* **Square-TASQ v.1** was unfortunately found to be thermally and chemically instable, and *ii)* **Square-TASQ v.2** that tackles these issues thanks to the presence of four phenol groups on the squarate core that provide both steric hindrance and electron donation. The linkage of the ^{PNA}G arms on the fluorescent core was done by CuAAC, which simplified the synthesis and provided a more robust TASQ. We then investigated the G4-binding properties of **Square-TASQ v.2** by FRET-melting, which was found to be an excellent G4-stabilizer (with the highest $\Delta T_{1/2}$ values ever observed with a TASQ). We believe that its charge, shape and high conformational freedom can explain these results. A preliminary fluorescent titration (λ_{ex}: 620 nm) showed that the addition of G4s triggers a small change in the emission spectra (blue shift) with a small increased in intensity, *via* a photophysical mechanism that is not yet fully understood.

The optimization of the synthesis of **Square-TASQ v.2** and a more detailed investigation of its photophysical properties will be the subject of a joined master's degree internship between Dijon and Prague. Additional MD investigations of the G4/probe interaction will also be required to rationalize the high level of G4-stabilisation (and thus, affinity) of the probe. However, preliminary optical imaging investigations have shown that **Square-TASQ v.2** does not suffer from REE limitations, which, when combined with its high G4-affinity and -selectivity and its NIR nature, make this probe a really promising tools to investigate G4 landscapes in cells.

Chapter 7

Conclusion and Perspective:

The goal of this PhD project was to develop efficient molecular tools that could be exploited to investigate the existence, prevalence and functional relevance of G4s in living cells. To this end, cell-permeable small molecules are invaluable as they can interrogate G4 biology *in situ*, with the aim of gaining precious insights into their actual biological roles. We focused here on the development of a unique class of biomimetic ligands referred to as Template-Assembled Synthetic G-Quartets, or TASQs, whose interaction with G4s relies on a 'like-likes-like' interaction between two G-quartets, one from the G4 and one from the TASQ. The main structural components of the TASQs are *i*) a central template, *ii*) a molecular appendage that can be used for introducing additional functionalities, and *iii*) four G arms organized around the central template that can fold into an intramolecular G-quartet.

We first focused on our template of choice, the aminomethylcyclen, or AMC, by producing several grams (*ca*. 20 g) of AMC improving both the synthesis and purification steps, which could lead us to envision a potential commercialization.

Next, we focused on the role that the linker between the template and the G arms plays by developing an alkylated version of **BioTASQ**, in which the amide bond-containing spacer of the **BioTASQ** was changed for a pentyl spacer. This modification resulted in an improved chemical access of the newly synthesized probe, the **BioCyTASQ**. We found that **BioCyTASQ** outperforms **BioTASQ** in every *in vitro* assay implemented, which makes it a better ligand and molecular bait for G4s. Indeed, both **BioCyTASQ** and **BioTASQ** were used to isolate G4s *in vitro* (fluorescence and qPCR pull-down assays) and from cell lysates, notably in plants where their performance in terms of G4-chemoprecipitation was directly compared with that of the well-established antibody BG4 (used for G4-immunoprecipitation). Both **BioCyTASQ** and **BioTASQ** proved to be excellent tools to fish G4s out of genomic extracts, thus being usable for detecting both G4-DNA (**G4DP-seq**) and G4-RNA (**G4RP-seq**), in both *in vitro* (*i.e*., after isolation of nucleic acids and proper G4 folding) and *in vivo* conditions (*i.e*., from fixed cells). To date, the full potential of **BioCyTASQ** still needs to be further investigated, notably in G4RP-

seq protocol which has been implemented with **BioTASQ** only. Also, the relevance of G4DP-seq must be assessed in mammalian cells. Furthermore, **BioTASQ** and **BioCyTASQ** were also used (thanks to a post-labeling step performed with a Cy3-labeled streptavidin) as imaging agents identifying G4 formation in cancerous cells *via* the so-called pre-targeted G4-imaging protocol. Incubation of the two TASQs in fixed and live cells highlighted their accumulation in perinuclear regions, nucleoli and nucleoplasmic sites, whose precise nature remains to be further investigated and validated.

We next developed another biotinylated TASQ in which triazole linkers were used to connect the AMC template with the ^{PNA}G arms. The synthesis of this new probe, named **BioTriazoTASQ**, was achieved thanks to the highly efficient click chemistry. While its chemical accessibility was not improved in comparison with **BioCyTASQ** in terms of overall chemical yield, this new ligand showed a unique *in vitro* property, that is, a strong bias towards the isolation of G4-RNA when compared to the properties of **BioCyTASQ**. The nature of this "preference" towards G4-RNA was investigated by a series of preliminary MD studies, highlighting that the more sterically demanding **BioTriazoTASQ** better fits the G4-RNA geometry. Unfortunately, **BioTriazoTASQ** has not yet been tested for its ability to isolate G4s from cell lysates (**G4RP-seq** and **G4DP-Seq**), but a direct comparison between the G4-DNA *vs.* G4-RNA isolating properties of **BioCyTASQ** and **BioTriazoTASQ** will be interesting in order to determine whether each probe could preferentially enrich samples of genomic *vs.* transcriptomic G4s. **BioTriazoTASQ** was also used for pre-targeted G4-imaging, which allowed for colocalizing G4 and DNA damage sites, thanks to the use of both Cy3-labeled streptavidin (for labeling TASQ) and an antibody raised against γH2AX (DNA damage marker). We were also able to provide a quantitative analysis of this colocalization, with *ca*. 5 % of common G4 and DNA damage sites.

We next focused on the nature of the TASQ appendage, changing the biotin of **BioTASQ**/**BioCyTASQ** for two handles capable of reacting bioothoghonally, *i. e.*, an alkyne and an azide appendage. The corresponding TASQ, named **MultiTASQ** and **^{Az}MultiTASQ**, respectively, are extremely selective G4-ligands that can be used in living cells as the nature of their appendages minimizes unspecific interactions with various cellular components. We validated two different click chemistry protocols with these new TASQs, *i.e*., CuAAC and

SPAAC, to couple **MultiTASQ** and **^{Az}MultiTASQ**, respectively, with a biotin handle: this model reaction allowed for validating the use of these news tools for pulling G4s down, with the aim of developing a novel sequencing protocol (so-called "click-seq") on the basis of what was done with the **G4RP-seq** and **G4DP-seq** protocols developed with **BioTASQ** and **BioCyTASQ**. Also, we performed a series of preliminary *in situ* click imaging experiments, which showed an improved signal-to-noise ratio as compared to that pre-targeted G4-imaging developed with **BioTASQ**, **BioCyTASQ** and **BioTriazoTASQ**. This validates the hypothesis according to which these new G4-probes minimize the undesired interactions with surrounding biomolecules in the cellular context. The synthesis and validation of **MultiTASQs** allowed for the design of a novel generation probe, in which a photoactivable handle was introduced. The G4-interacting properties of the corresponding **photoMultiTASQ** were validated and directly compared to those of the parent **MultiTASQ** compounds. The collected results demonstrated that the introduction of the diazirine moiety does not modify the G4-recognition of TASQs. We tried to covalently trap surrounding biomolecules by photoirradiation *in vitro* but our attempts were not successful, likely because of an unoptimized irradiation device. The potential of this probe is particularly relevant for G4-directed chemical biology as it can provide invaluable information about both G4-genomic/transcriptomic landscapes and G4-interacting proteins in cells (*Figure 65*).

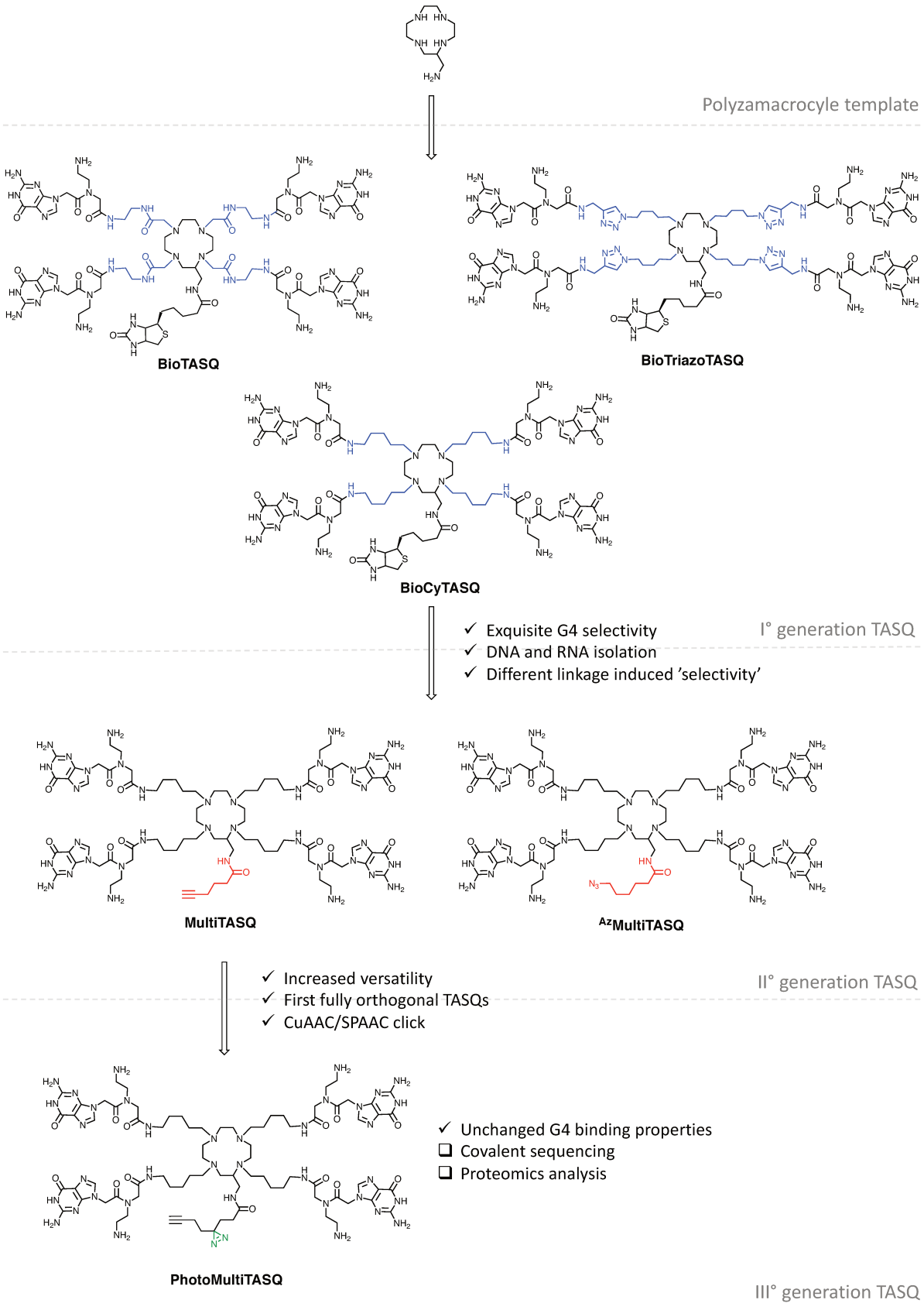


Figure 65: polyazamacrocycle based TASQ and their respective features developed this project.

Finally, while AMC-based TASQs were mostly used to isolate G4s, TASQs can also be used to visualize them in cells. It was shown with pre-targeted G4-imaging using **BioTASQ**, **BioCyTASQ** and **BioTriazoTASQ** with a fluorescently labeled streptavidin, with *in situ* click chemistry using **MultiTASQ** and **^{Az}MultiTASQ** with a clickable fluorophore. These approaches are interesting but they require multiple steps and provide images of perfectible quality. We thus decided to optimize an existing fluorescent G4-probe, **N-TASQ**, and to develop new probes allowing for the direct visualization of G4s in cells. To this end, we firstly revisited the synthesis of **N-TASQ** by improving the key step of its synthesis, a quadruple Sonogashira-Heck-Cassar reaction, in order to increase its chemical access. Then, inspired by the fruitful introduction of triazole linkers within the **BioTriazoTASQ** scaffold, we developed a new probe referred to as **^{Tz}N-TASQ**, which showed good fluorescence turn-on properties, even better to that of **N-TASQ**. Finally, we identified the main drawback of both **N-TASQ** and **^{Tz}N-TASQ**, which relies in the need to use excitation wavelengths in the UV-region, which is responsible for an elevated background fluorescence (cell autofluorescence). We thus changed the central naphthalene template of **N-TASQ** and **^{Tz}N-TASQ** for a squaraine core. The choice of this central template was limited due to two major requirements, *i.e.*, an emission in the near infrared region (NIR) and a pseudo C_4-symmetry. Two squaraine-based TASQ were developed: while the first version, **Square-TASQ v.1,** was found to be poorly stable, the second version, **Square-TASQ v.2**, was found to be an excellent G4 ligand. Its G4-binding abilities outperformed all the other TASQs reported to date, and its excitation at 620 nm potentially makes **Square-TASQ v.2** an excellent NIR G4-probe. Contrarily to **N-TASQ** and **^{Tz}N-TASQ**, **Square-TASQ v.2** is not a turn-on probe and its fluorescence is modestly modified upon interaction with its G4 targets. Preliminary cell-based investigations show that **Square-TASQ v.2** allows for visualizing G4s in the perinuclear and nucleoli area, in a manner that is similar to **N-TASQ** and **^{Tz}N-TASQ**. Further evaluations of **Square-TASQ v.2** fluorescence both *in vitro* and *in cella* are currently ongoing to better assess the potential of this new TASQ (*Figure 66*).

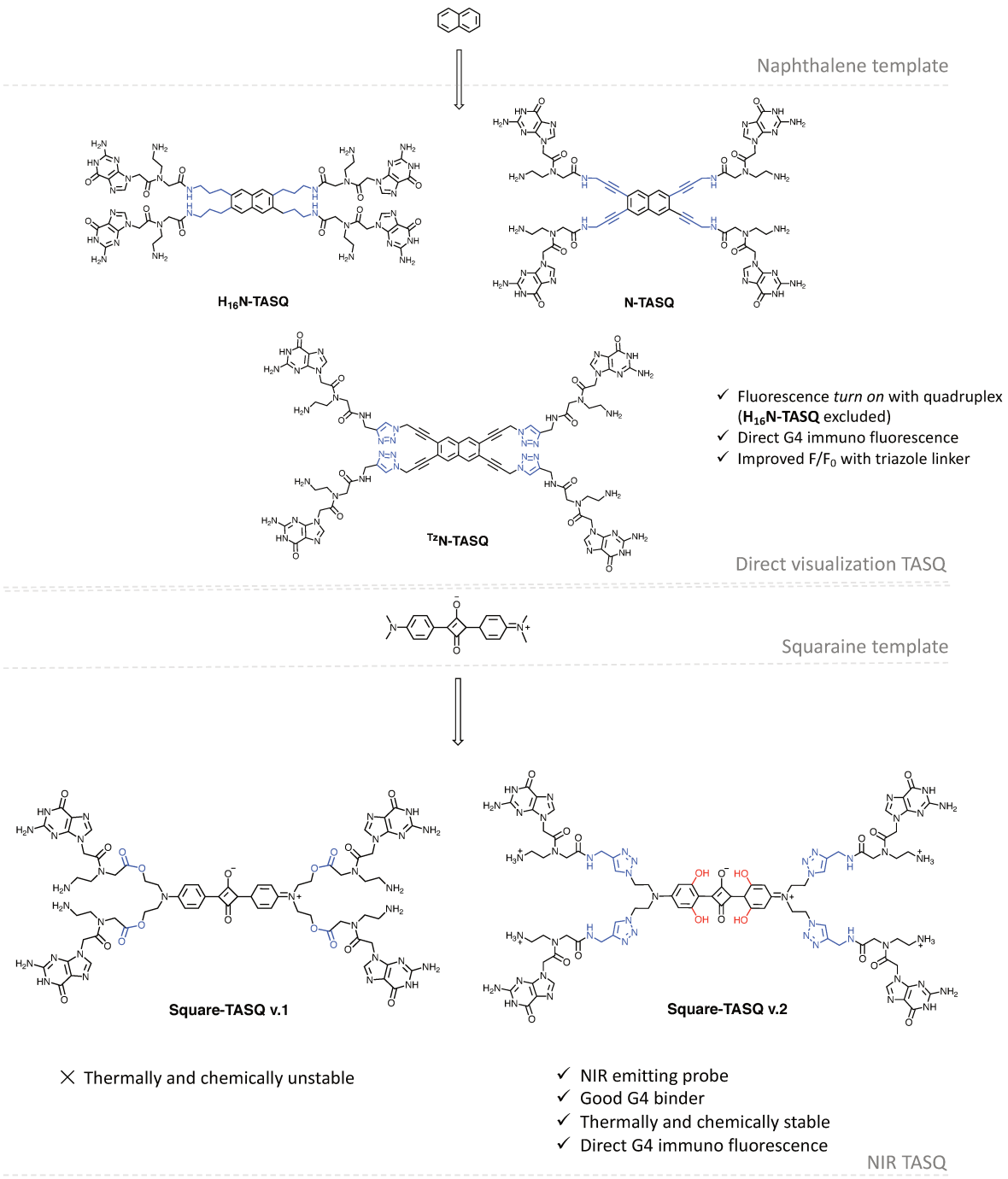


Figure 66: Naphthalene and Squaraine based TASQ and their respective features developed this project.

Bibliography

1. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S., The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020,** *21*, 459-474.
2. Rhodes, D.; Lipps, H. J., G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015,** *43* (18), 8627-8637.
3. Kwok, C. K.; Merrick, C. J., G-quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.* **2017,** *35* (10), 997-1013.
4. Steudel, H.; Brigl, P., Über die Guanylsäure aus der Pankreasdrüse. II. Mitteilung. **1910,** *68* (1), 40-51.
5. Lagnado, J., The story of quadruplex DNA–it started with a Bang! *The Biochemist* **2013,** *35* (2), 44-46.
6. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids. *Nature* **1953,** *171* (4356), 737-738.
7. Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R., Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953,** *171* (4356), 738-740.
8. Franklin, R. E.; Gosling, R. G., Molecular configuration in sodium thymonucleate. *Nature* **1953,** *171* (4356), 740-741.
9. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix formation by guanylic acid. *Proc Natl Acad Sci U S A* **1962,** *48* (12), 2013-8.
10. Hoogsteen, K., The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1959,** *12* (10), 822-823.
11. Hoogsteen, K. R., The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallogr.* **1963,** *16* (9), 907-916.
12. Largy, E.; Mergny, J. L.; Gabelica, V., Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. *Met Ions Life Sci* **2016,** *16*, 203-58.
13. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S., Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in Chemistry* **2016,** *4*.
14. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34* (19), 5402-5415.
15. Esposito, V.; Galeone, A.; Mayol, L.; Oliviero, G.; Virgilio, A.; Randazzo, L., A Topological Classification of G-Quadruplex Structures. *Nucleosides, Nucleotides & Nucleic Acids* **2007,** *26* (8-9), 1155-1159.
16. Yang, D.; Okamoto, K., Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med. Chem.* **2010,** *2* (4), 619-646.
17. Phan, A. T., Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **2010,** *277* (5), 1107-1117.
18. Dvorkin, S. A.; Karsisiotis, A. I.; Webba da Silva, M., Encoding canonical DNA quadruplex structure. *Science Adv.* **2018,** *4* (8), eaat3007.

19. Collie, G. W.; Parkinson, G. N., The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.* **2011,** *40* (12), 5867-5892.
20. Parkinson, G. N.; Lee, M. P.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-80.
21. Zhou, J.; Amrane, S.; Rosu, F.; Salgado, G. F.; Bian, Y.; Tateishi-Karimata, H.; Largy, E.; Korkut, D. N.; Bourdoncle, A.; Miyoshi, D.; Zhang, J.; Ju, H.; Wang, W.; Sugimoto, N.; Gabelica, V.; Mergny, J.-L., Unexpected Position-Dependent Effects of Ribose G-Quartets in G-Quadruplexes. *J. Am. Chem. Soc.* **2017,** *139* (23), 7768-7779.
22. Sen, D.; Gilbert, W., Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **1988,** *334* (6180), 364-366.
23. Williamson, J. R.; Raghuraman, M.; Cech, T. R., Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **1989,** *59* (5), 871-880.
24. Sundquist, W. I.; Klug, A., Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989,** *342* (6251), 825.
25. Todd, A. K.; Johnston, M.; Neidle, S., Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *33* (9), 2901-2907.
26. Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33* (9), 2908-2916.
27. Puig Lombardi, E.; Londoño-Vallejo, A., A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Res.* **2019,** *48* (1), 1-15.
28. Cagirici, H. B.; Budak, H.; Sen, T. Z., G4Boost: a machine learning-based tool for quadruplex identification and stability prediction. *BMC Bioinformatics* **2022,** *23* (1), 240.
29. Spiegel, J.; Adhikari, S.; Balasubramanian, S., The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry* **2020,** *2* (2), 123-136.
30. Lyu, K.; Chow, E. Y.-C.; Mou, X.; Chan, T.-F.; Kwok, Chun K., RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.* **2021,** *49* (10), 5426-5450.
31. Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S., High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **2015,** *33* (8), 877-881.
32. Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S. V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M., G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **2016,** *48* (10), 1267-1272.
33. Kwok, C. K.; Marsico, G.; Sahakyan, A. B.; Chambers, V. S.; Balasubramanian, S., rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Meth.* **2016,** *13* (10), 841-844.
34. Yang, S. Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A. G.; Wong, J. M.; Monchaud, D., Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **2018,** *9* (1), 4730.
35. Xu, Y., Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.* **2011,** *40* (5), 2719-2740.
36. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A., In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with

Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98* (15), 8572-8577.

37. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S., Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013,** *5* (3), 182-6.
38. Di Antonio, M.; Ponjavic, A.; Radzevičius, A.; Ranasinghe, R. T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L.-M.; Lee, S. F.; Klenerman, D.; Balasubramanian, S., Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **2020,** *12* (9), 832-837.
39. Umar, M. I.; Ji, D.; Chan, C.-Y.; Kwok, C. K., G-Quadruplex-Based Fluorescent Turn-On Ligands and Aptamers: From Development to Applications. *Molecules* **2019,** *24* (13), 2416.
40. Jafri, M. A.; Ansari, S. A.; Alqahtani, M. H.; Shay, J. W., Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med.* **2016,** *8* (1), 69-69.
41. De Lange, T., T-loops and the origin of telomeres. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (4), 323-329.
42. Harley, C. B., Telomere loss: mitotic clock or genetic time bomb? *Mut. Res.* **1991,** *256* (2), 271-282.
43. Watson, J. D., Origin of concatemeric T7DNA. *Nature New Biol.* **1972,** *239* (94), 197-201.
44. Olovnikov, A. M., A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *J. Theor. Biol.* **1973,** *41* (1), 181-190.
45. Olovnikov, A. M., Telomeres, telomerase, and aging: origin of the theory. *Exp. Gerontol.* **1996,** *31* (4), 443-448.
46. Hayflick, L., The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res.* **1965,** *37* (3), 614-636.
47. Blackburn, E. H.; Greider, C. W.; Szostak, J. W., Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat. Med.* **2006,** *12* (10), 1133-1138.
48. Bodnar, A. G.; Ouellette, M.; Frolkis, M.; Holt, S. E.; Chiu, C. P.; Morin, G. B.; Harley, C. B.; Shay, J. W.; Lichtsteiner, S.; Wright, W. E., Extension of life-span by introduction of telomerase into normal human cells. *Science* **1998,** *279* (5349), 349-52.
49. Djojosubroto, M. W.; Chin, A. C.; Go, N.; Schaetzlein, S.; Manns, M. P.; Gryaznov, S.; Harley, C. B.; Rudolph, K. L., Telomerase antagonists GRN163 and GRN163L inhibit tumor growth and increase chemosensitivity of human hepatoma. *Hepatology* **2005,** *42* (5), 1127-36.
50. Zahler, A. M.; Williamson, J. R., Inhibition of telomerase by G-quartet DNA structure. *Nature* **1991,** *350* (6320), 718.
51. Sun, D. Y.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H., Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.* **1997,** *40* (14), 2113-2116.
52. Yang, S. Y.; Chang, E. Y. C.; Lim, J.; Kwan, H. H.; Monchaud, D.; Yip, S.; Stirling, Peter C.; Wong, J. M. Y., G-quadruplexes mark alternative lengthening of telomeres. *NAR Cancer* **2021,** *3* (3).

53. Bryan, T. M.; Englezou, A.; Gupta, J.; Bacchetti, S.; Reddel, R., Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* **1995,** *14* (17), 4240-4248.
54. Bryan, T. M.; Englezou, A.; DallaPozza, L.; Dunham, M. A.; Reddel, R. R., Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* **1997,** *3* (11), 1271-1274.
55. Opresko, P. L.; Mason, P. A.; Podell, E. R.; Lei, M.; Hickson, I. D.; Cech, T. R.; Bohr, V. A., POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J Biol. Chem.* **2005,** *280* (37), 32069-80.
56. Xu, Y.; Suzuki, Y.; Ito, K.; Komiyama, M., Telomeric repeat-containing RNA structure in living cells. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (33), 14579-14584.
57. Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J., Telomeric repeat–containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **2007,** *318* (5851), 798-801.
58. Mei, Y.; Deng, Z.; Vladimirova, O.; Gulve, N.; Johnson, F. B.; Drosopoulos, W. C.; Schildkraut, C. L.; Lieberman, P. M., TERRA G-quadruplex RNA interaction with TRF2 GAR domain is required for telomere integrity. *Sci. Rep.* **2021,** *11* (1), 3509.
59. Kosiol, N.; Juranek, S.; Brossart, P.; Heine, A.; Paeschke, K., G-quadruplexes: A promising target for cancer therapy. *Mol. Cancer* **2021,** *20* (1), 1-18.
60. Moye, A. L.; Porter, K. C.; Cohen, S. B.; Phan, T.; Zyner, K. G.; Sasaki, N.; Lovrecz, G. O.; Beck, J. L.; Bryan, T. M., Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **2015,** *6* (1), 7643.
61. Prorok, P.; Artufel, M.; Aze, A.; Coulombe, P.; Peiffer, I.; Lacroix, L.; Guédin, A.; Mergny, J.-L.; Damaschke, J.; Schepers, A.; Cayrou, C.; Teulade-Fichou, M.-P.; Ballester, B.; Méchali, M., Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat. Commun.* **2019,** *10* (1), 3274.
62. Bryan, T. M., Mechanisms of DNA Replication and Repair: Insights from the Study of G-Quadruplexes. *Molecules* **2019,** *24* (19), 3439.
63. Bay, D. H.; Busch, A.; Lisdat, F.; Iida, K.; Ikebukuro, K.; Nagasawa, K.; Karube, I.; Yoshida, W., Identification of G-quadruplex structures that possess transcriptional regulating functions in the Dele and Cdc6 CpG islands. *BMC Mol. Biol.* **2017,** *18* (1), 17.
64. Jodoin, R.; Carrier, J. C.; Rivard, N.; Bisaillon, M.; Perreault, J.-P., G-quadruplex located in the 5'UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Res.* **2019,** *47* (19), 10247-10266.
65. Schiavone, D.; Guilbaud, G.; Murat, P.; Papadopoulou, C.; Sarkies, P.; Prioleau, M. N.; Balasubramanian, S.; Sale, J. E., Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J.* **2014,** *33* (21), 2507-20.
66. Zell, J.; Rota Sperti, F.; Britton, S.; Monchaud, D., DNA folds threaten genetic stability and can be leveraged for chemotherapy. *RSC Chem. Biol.* **2021,** *2*, 47-76.
67. Valton, A. L.; Hassan-Zadeh, V.; Lema, I.; Boggetto, N.; Alberti, P.; Saintomé, C.; Riou, J. F.; Prioleau, M. N., G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* **2014,** *33* (7), 732-46.

68. Lago, S.; Nadai, M.; Cernilogar, F. M.; Kazerani, M.; Domíniguez Moreno, H.; Schotta, G.; Richter, S. N., Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **2021,** *12* (1), 3885.
69. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H., Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA* **2002,** *99* (18), 11593-8.
70. Balasubramanian, S.; Hurley, L. H.; Neidle, S., Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **2011,** *10* (4), 261-275.
71. Shay, J. W.; Wright, W. E., Telomeres and telomerase: three decades of progress. *Nat. Rev. Genet.* **2019,** *20* (5), 299-309.
72. Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S., Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J. Med. Chem.* **2008,** *51* (18), 5641-5649.
73. Brosh, R. M.; Matson, S. W., History of DNA Helicases. *Genes* **2020,** *11* (3), 255.
74. Hou, X.-M.; Wu, W.-Q.; Duan, X.-L.; Liu, N.-N.; Li, H.-H.; Fu, J.; Dou, S.-X.; Li, M.; Xi, X.-G., Molecular mechanism of G-quadruplex unwinding helicase: sequential and repetitive unfolding of G-quadruplex by Pif1 helicase. *Biochem. J.* **2015,** *466* (1), 189-199.
75. Lejault, P.; Mitteaux, J.; Rota Sperti, F.; Monchaud, D., How to untie G-quadruplex knots and why? *Cell Chem. Biol.* **2021,** *28*, 436-455.
76. Moruno-Manchon, J. F.; Lejault, P.; Wang, Y.; McCauley, B.; Honarpisheh, P.; Scheihing, D. A. M.; Singh, S.; Dang, W.; Kim, N.; Urayama, A.; Zhu, L.; Monchaud, D.; McCullough, L. D.; Tsvetkov, A. S., Small-molecule G-quadruplex stabilizers reveal a novel pathway of autophagy regulation in neurons. *eLife* **2020,** *9*, e52283.
77. Kharel, P.; Becker, G.; Tsvetkov, V.; Ivanov, P., Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. *Nucleic Acids Res.* **2020,** *48* (22), 12534-12555.
78. Fay, M. M.; Lyons, S. M.; Ivanov, P., RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **2017,** *429* (14), 2127-2147.
79. Beaudoin, J.-D.; Perreault, J.-P., Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* **2013,** *41* (11), 5898-5911.
80. Beaudoin, J.-D.; Perreault, J.-P., 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* **2010,** *38* (20), 7022-7036.
81. Weldon, C.; Behm-Ansmant, I.; Hurley, L. H.; Burley, G. A.; Branlant, C.; Eperon, I. C.; Dominguez, C., Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA. *Nat. Chem. Biol.* **2017,** *13* (1), 18-20.
82. Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S., An RNA G-quadruplex in the 5 ' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **2007,** *3* (4), 218-221.
83. Bugaut, A.; Balasubramanian, S., 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **2012,** *40* (11), 4727-4741.
84. Cammas, A.; Millevoi, S., RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* **2017,** *45* (4), 1584-1595.

85. Robinson, J.; Raguseo, F.; Nuccio, S. P.; Liano, D.; Di Antonio, M., DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* **2021,** *49* (15), 8419-8431.
86. Neidle, S., Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016,** *59* (13), 5987-6011.
87. McLuckie, K. I. E.; Di Antonio, M.; Zecchini, H.; Xian, J.; Caldas, C.; Krippendorff, B. F.; Tannahill, D.; Lowe, C.; Balasubramanian, S., G-Quadruplex DNA as a Molecular Target for Induced Synthetic Lethality in Cancer Cells. *J. Am. Chem. Soc.* **2013,** *135* (26), 9640-9643.
88. Jackson, S. P.; Bartek, J., The DNA-damage response in human biology and disease. *Nature* **2009,** *461* (7267), 1071-1078.
89. Raguseo, F.; Chowdhury, S.; Minard, A.; Di Antonio, M., Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun.* **2020,** *56*, 1317-1324.
90. Monchaud, D., Quadruplex detection in human cells. In *Annu. Rep. Med. Chem.*, Neidle, S., Ed. Academic Press: 2020; Vol. 54, pp 133-160.
91. Haider, S. M.; Parkinson, G. N.; Neidle, S., Structure of a G-quadruplex-ligand complex. *J. Mol. Biol.* **2003,** *326* (1), 117-25.
92. Campbell, N. H.; Patel, M.; Tofa, A. B.; Ghosh, R.; Parkinson, G. N.; Neidle, S., Selectivity in ligand recognition of G-quadruplex loops. *Biochem.* **2009,** *48* (8), 1675-80.
93. Heald, R. A.; Modi, C.; Cookson, J. C.; Hutchinson, I.; Laughton, C. A.; Gowan, S. M.; Kelland, L. R.; Stevens, M. F. G., Antitumor polycyclic acridines. 8. Synthesis and telomerase-inhibitory activity of methylated pentacyclic acridinium salts. *J. Med. Chem.* **2002,** *45* (3), 590-597.
94. Shi, D. F.; Wheelhouse, R. T.; Sun, D.; Hurley, L. H., Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *J. Med. Chem.* **2001,** *44* (26), 4509-23.
95. 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. *Proc. Natl. Acad. Sci. USA* **2001,** *98* (9), 4844-4849.
96. Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S., The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **2005,** *65* (4), 1489-1496.
97. Fedoroff, O. Y.; Salazar, M.; Han, H.; Chemeris, V. V.; Kerwin, S. M.; Hurley, L. H., NMR-Based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochem.* **1998,** *37* (36), 12367-74.
98. 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. *J. Mol. Biol.* **2008,** *381* (5), 1145-1156.
99. Zuffo, M.; Guédin, A.; Leriche, E.-D.; Doria, F.; Pirota, V.; Gabelica, V.; Mergny, J.-L.; Freccero, M., More is not always better: finding the right trade-off between affinity and selectivity of a G-quadruplex ligand. *Nucleic Acids Res.* **2018,** *46* (19), e115-e115.

100. Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H., Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. *J. Am. Chem. Soc.* **2001,** *123* (6), 1262-3.
101. Lemarteleur, T.; Gomez, D.; Paterski, R.; Mandine, E.; Mailliet, P.; Riou, J.-F., Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem. Biophys. Res. Commun.* **2004,** *323* (3), 802-808.
102. De Cian, A.; DeLemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D., Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J. Am. Chem. Soc.* **2007,** *129* (7), 1856-1857.
103. Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S., A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **2008,** *130* (47), 15758-15759.
104. Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K. C.; Von Hoff, D.; Anderes, K.; Rice, W. G., Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res.* **2009,** *69* (19), 7653-7661.
105. Drygin, D.; Lin, A.; Bliesath, J.; Ho, C. B.; O'Brien, S. E.; Proffitt, C.; Omori, M.; Haddach, M.; Schwaebe, M. K.; Siddiqui-Jain, A.; Streiner, N.; Quin, J. E.; Sanij, E.; Bywater, M. J.; Hannan, R. D.; Ryckman, D.; Anderes, K.; Rice, W. G., Targeting RNA Polymerase I with an Oral Small Molecule CX-5461 Inhibits Ribosomal RNA Synthesis and Solid Tumor Growth. *Cancer Res.* **2011,** *71* (4), 1418-1430.
106. Hilton, J.; Gelmon, K.; Bedard, P. L.; Tu, D.; Xu, H.; Tinker, A. V.; Goodwin, R.; Laurie, S. A.; Jonker, D.; Hansen, A. R.; Veitch, Z. W.; Renouf, D. J.; Hagerman, L.; Lui, H.; Chen, B.; Kellar, D.; Li, I.; Lee, S.-E.; Kono, T.; Cheng, B. Y. C.; Yap, D.; Lai, D.; Beatty, S.; Soong, J.; Pritchard, K. I.; Soria-Bretones, I.; Chen, E.; Feilotter, H.; Rushton, M.; Seymour, L.; Aparicio, S.; Cescon, D. W., Results of the phase I CCTG IND.231 trial of CX-5461 in patients with advanced solid tumors enriched for DNA-repair deficiencies. *Nat. Commun.* **2022,** *13* (1), 3607.
107. Olivieri, M.; Cho, T.; Álvarez-Quilón, A.; Li, K.; Schellenberg, M. J.; Zimmermann, M.; Hustedt, N.; Rossi, S. E.; Adam, S.; Melo, H.; Heijink, A. M.; Sastre-Moreno, G.; Moatti, N.; Szilard, R. K.; McEwan, A.; Ling, A. K.; Serrano-Benitez, A.; Ubhi, T.; Feng, S.; Pawling, J.; Delgado-Sainz, I.; Ferguson, M. W.; Dennis, J. W.; Brown, G. W.; Cortés-Ledesma, F.; Williams, R. S.; Martin, A.; Xu, D.; Durocher, D., A Genetic Map of the Response to DNA Damage in Human Cells. *Cell* **2020,** *182* (2), 481-496.e21.
108. Masson, T.; Landras Guetta, C.; Laigre, E.; Cucchiarini, A.; Duchambon, P.; Teulade-Fichou, M.-P.; Verga, D., BrdU immuno-tagged G-quadruplex ligands: a new ligand-guided immunofluorescence approach for tracking G-quadruplexes in cells. *Nucleic Acids Res.* **2021,** *49* (22), 12644-12660.
109. Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P., Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **2012,** *8* (3), 301-310.
110. Grande, V.; Doria, F.; Freccero, M.; Würthner, F., An Aggregating Amphiphilic Squaraine: A Light-up Probe That Discriminates Parallel G-Quadruplexes. *Angew. Chem. Int. Ed.* **2017,** *56* (26), 7520-7524.

111. Stefan, L.; Monchaud, D., Applications of guanine quartets in nanotechnology and chemical biology. *Nat. Rev. Chem.* **2019,** *3* (11), 650-668.
112. Nikan, M.; Sherman, J. C., Template-Assembled Synthetic G-Quartets (TASQs). *Angew. Chem. Int. Ed.* **2008,** *47* (26), 4900-4902.
113. Bare, G. A. L.; Liu, B.; Sherman, J. C., Synthesis of a Single G-Quartet Platform in Water. *J. Am. Chem. Soc.* **2013,** *135* (32), 11985-11989.
114. Stefan, L.; Guedin, A.; Amrane, S.; Smith, N.; Denat, F.; Mergny, J.-L.; Monchaud, D., DOTASQ as a prototype of nature-inspired G-quadruplex ligand. *Chem. Commun.* **2011,** *47* (17), 4992-4994.
115. Murat, P.; Gennaro, B.; Garcia, J.; Spinelli, N.; Dumy, P.; Defrancq, E., The Use of a Peptidic Scaffold for the Formation of Stable Guanine Tetrads: Control of a H-bonded Pattern in Water. *Chem. Eur. J.* **2011,** *17* (21), 5791-5795.
116. Haudecoeur, R.; Stefan, L.; Denat, F.; Monchaud, D., A Model of Smart G-Quadruplex Ligand. *J. Am. Chem. Soc.* **2013,** *135* (2), 550-553.
117. Laguerre, A.; Stefan, L.; Larrouy, M.; Genest, D.; Novotna, J.; Pirrotta, M.; Monchaud, D., A Twice-As-Smart Synthetic G-Quartet: PyroTASQ Is Both a Smart Quadruplex Ligand and a Smart Fluorescent Probe. *J. Am. Chem. Soc.* **2014,** *136* (35), 12406-12414.
118. Laguerre, A.; Hukezalie, K.; Winckler, P.; Katranji, F.; Chanteloup, G.; Pirrotta, M.; Perrier-Cornet, J.-M.; Wong, J. M. Y.; Monchaud, D., Visualization of RNA-Quadruplexes in Live Cells. *J. Am. Chem. Soc.* **2015,** *137* (26), 8521-8525.
119. Laguerre, A.; Wong, J. M. Y.; Monchaud, D., Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci. Rep.* **2016,** *6* (1), 32141.
120. Yang, S. Y.; Lejault, P.; Chevrier, S.; Boidot, R.; Robertson, A. G.; Wong, J. M. Y.; Monchaud, D., Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat. Commun.* **2018,** *9* (1), 4730.
121. Rousselin, Y.; Sok, N.; Boschetti, F.; Guilard, R.; Denat, F., Efficient Synthesis of New C-Functionalized Macrocyclic Polyamines. *Eur. J. Org. Chem.* **2010,** *2010* (9), 1688-1693.
122. Wilchek, M.; Bayer, E. A., The avidin-biotin complex in bioanalytical applications. *Anal. Biochem.* **1988,** *171* (1), 1-32.
123. Renaud de la Faverie, A.; Hamon, F.; Di Primo, C.; Largy, E.; Dausse, E.; Delaurière, L.; Landras-Guetta, C.; Toulmé, J.-J.; Teulade-Fichou, M.-P.; Mergny, J.-L., Nucleic acids targeted to drugs: SELEX against a quadruplex ligand. *Biochimie* **2011,** *93* (8), 1357-1367.
124. Busto, N.; Calvo, P.; Santolaya, J.; Leal, J. M.; Guédin, A.; Barone, G.; Torroba, T.; Mergny, J.-L.; García, B., Fishing for G-Quadruplexes in Solution with a Perylene Diimide Derivative Labeled with Biotins. *Chem. Eur. J.* **2018,** *24* (44), 11292-11296.
125. Renard, I.; Grandmougin, M.; Roux, A.; Yang, S. Y.; Lejault, P.; Pirrotta, M.; Wong, J. M. Y.; Monchaud, D., Small-molecule affinity capture of DNA/RNA quadruplexes and their identification in vitro and in vivo through the G4RP protocol. *Nucleic Acids Res.* **2019,** *47* (11), 5502-5510.
126. Yang, S. Y.; Monchaud, D.; Wong, J. M. Y., Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq. *Nat. Protoc.* **2022,** *17* (3), 870-889.
127. El-Faham, A.; Albericio, F., Peptide Coupling Reagents, More than a Letter Soup. *Chem. Rev.* **2011,** *111* (11), 6557-6602.

128. Sperti, F. R.; Charbonnier, T.; Lejault, P.; Zell, J.; Bernhard, C.; Valverde, I. E.; Monchaud, D., Biomimetic, Smart, and Multivalent Ligands for G-Quadruplex Isolation and Bioorthogonal Imaging. *ACS Chem. Biol.* **2021,** *16* (5), 905-914.
129. De Rache, A.; Mergny, J.-L., Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* **2015,** *115*, 194-202.
130. Feng, Y.; Tao, S.; Zhang, P.; Sperti, F. R.; Liu, G.; Cheng, X.; Zhang, T.; Yu, H.; Wang, X.-e.; Chen, C.; Monchaud, D.; Zhang, W., Epigenomic features of DNA G-quadruplexes and their roles in regulating rice gene transcription. *Plant Physiol.* **2021,** *188* (3), 1632-1648.
131. Feng, Y.; Luo, Z.; Sperti, F. R.; Valverde, I. E.; Zhang, W.; Monchaud, D., Chemo-versus immuno-precipitation of G-quadruplex-DNA (G4DNA): a direct comparison of the efficiency of the antibody BG4 versus the small-molecule ligands TASQs for G4 affinity capture. **2022**. DOI: 10.21203/rs.3.rs-1717207/v1
132. Saha, P.; Panda, D.; Dash, J., The application of click chemistry for targeting quadruplex nucleic acids. *Chem. Commun.* **2019,** *55* (6), 731-750.
133. Rota Sperti, F.; Dupouy, B.; Mitteaux, J.; Pipier, A.; Pirrotta, M.; Chéron, N.; Valverde, I. E.; Monchaud, D., Click-Chemistry-Based Biomimetic Ligands Efficiently Capture G-Quadruplexes In Vitro and Help Localize Them at DNA Damage Sites in Human Cells. *JACS Au* **2022,** *2* (7), 1588-1595.
134. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002,** *41* (14), 2596-2599.
135. Tornøe, C. W.; Christensen, C.; Meldal, M., Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002,** *67* (9), 3057-3064.
136. Huisgen, R., 1,3-Dipolar Cycloadditions. Past and Future. *Angew. Chem. Int. Ed.* **1963,** *2* (10), 565-598.
137. Tron, G. C.; Pirali, T.; Billington, R. A.; Canonico, P. L.; Sorba, G.; Genazzani, A. A., Click chemistry reactions in medicinal chemistry: Applications of the 1,3-dipolar cycloaddition between azides and alkynes. *Medicinal Res. Rev.* **2008,** *28* (2), 278-308.
138. Hein, C. D.; Liu, X.-M.; Wang, D., Click chemistry, a powerful tool for pharmaceutical sciences. *Pharm. Res.* **2008,** *25* (10), 2216-2230.
139. Nwe, K.; Brechbiel, M. W., Growing applications of "click chemistry" for bioconjugation in contemporary biomedical research. *Cancer Biother. Radiopharm.* **2009,** *24* (3), 289-302.
140. Cañeque, T.; Müller, S.; Rodriguez, R., Visualizing biologically active small molecules in cells using click chemistry. *Nat. Rev. Chem.* **2018,** *2*, 202-215.
141. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004,** *126* (46), 15046-15047.
142. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009,** *48* (38), 6974-6998.
143. Rečnik, L.-M.; Kandioller, W.; Mindt, T. L., 1,4-Disubstituted 1,2,3-Triazoles as Amide Bond Surrogates for the Stabilisation of Linear Peptides with Biological Activity. *Molecules* **2020,** *25* (16), 3576.

144. Valverde, I. E.; Vomstein, S.; Fischer, C. A.; Mascarin, A.; Mindt, T. L., Probing the backbone function of tumor targeting peptides by an amide-to-triazole substitution strategy. *J. Med. Chem.* **2015,** *58* (18), 7475-7484.
145. Aucagne, V.; Berná, J.; Crowley, J. D.; Goldup, S. M.; Hänni, K. D.; Leigh, D. A.; Lusby, P. J.; Ronaldson, V. E.; Slawin, A. M. Z.; Viterisi, A.; Walker, D. B., Catalytic "Active-Metal" Template Synthesis of [2]Rotaxanes, [3]Rotaxanes, and Molecular Shuttles, and Some Observations on the Mechanism of the Cu(I)-Catalyzed Azide−Alkyne 1,3-Cycloaddition. *J. Am. Chem. Soc.* **2007,** *129* (39), 11950-11963.
146. Straub, B. F., μ-Acetylide and μ-alkenylidene ligands in "click" triazole syntheses. *Chem. Commun.* **2007,** (37), 3868-3870.
147. Rodionov, V. O.; Fokin, V. V.; Finn, M. G., Mechanism of the Ligand-Free CuI-Catalyzed Azide–Alkyne Cycloaddition Reaction. *Angew. Chem. Int. Ed.* **2005,** *44* (15), 2210-2215.
148. Ahlquist, M.; Fokin, V. V., Enhanced Reactivity of Dinuclear Copper(I) Acetylides in Dipolar Cycloadditions. *Organomet.* **2007,** *26* (18), 4389-4391.
149. Worrell, B. T.; Malik, J. A.; Fokin, V. V., Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. *Science* **2013,** *340* (6131), 457-460.
150. Rodionov, V. O.; Presolski, S. I.; Díaz Díaz, D.; Fokin, V. V.; Finn, M. G., Ligand-Accelerated Cu-Catalyzed Azide−Alkyne Cycloaddition: A Mechanistic Report. *J. Am. Chem. Soc.* **2007,** *129* (42), 12705-12712.
151. Knör, S.; Modlinger, A.; Poethko, T.; Schottelius, M.; Wester, H.-J.; Kessler, H., Synthesis of Novel 1,4,7,10-Tetraazacyclodecane-1,4,7,10-Tetraacetic Acid (DOTA) Derivatives for Chemoselective Attachment to Unprotected Polyfunctionalized Compounds. *Chem. Eur. J.* **2007,** *13* (21), 6082-6090.
152. Knör, S.; Modlinger, A.; Poethko, T.; Schottelius, M.; Wester, H. J.; Kessler, H., Synthesis of Novel 1, 4, 7, 10-Tetraazacyclodecane-1, 4, 7, 10-Tetraacetic Acid (DOTA) Derivatives for Chemoselective Attachment to Unprotected Polyfunctionalized Compounds. *Chem. Eur. J.* **2007,** *13* (21), 6082-6090.
153. Jamroskovic, J.; Obi, I.; Movahedi, A.; Chand, K.; Chorell, E.; Sabouri, N., Identification of putative G-quadruplex DNA structures in S. pombe genome by quantitative PCR stop assay. *DNA Repair* **2019,** *82*, 102678.
154. Marchetti, C.; Zyner, K. G.; Ohnmacht, S. A.; Robson, M.; Haider, S. M.; Morton, J. P.; Marsico, G.; Vo, T.; Laughlin-Toth, S.; Ahmed, A. A.; Di Vita, G.; Pazitna, I.; Gunaratnam, M.; Besser, R. J.; Andrade, A. C. G.; Diocou, S.; Pike, J. A.; Tannahill, D.; Pedley, R. B.; Evans, T. R. J.; Wilson, W. D.; Balasubramanian, S.; Neidle, S., Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. *J. Med. Chem.* **2018,** *61* (6), 2500-2517.
155. Rappold, I.; Iwabuchi, K.; Date, T.; Chen, J., Tumor Suppressor P53 Binding Protein 1 (53bp1) Is Involved in DNA Damage–Signaling Pathways. *J. Cell Biol.* **2001,** *153* (3), 613-620.
156. Salvati, E.; Doria, F.; Manoli, F.; D'Angelo, C.; Biroccio, A.; Freccero, M.; Manet, I., A bimodal fluorescent and photocytotoxic naphthalene diimide for theranostic applications. *Org. Biomol. Chem.* **2016,** *14* (30), 7238-7249.
157. Rossetto, D.; Avvakumov, N.; Côté, J., Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* **2012,** *7* (10), 1098-1108.

158. Bonner, W. M.; Redon, C. E.; Dickey, J. S.; Nakamura, A. J.; Sedelnikova, O. A.; Solier, S.; Pommier, Y., GammaH2AX and cancer. *Nat. Rev. Cancer* **2008,** *8* (12), 957-967.
159. Monchaud, D.; Valverde, I. E.; Lejault, P.; Rota Sperti, F. Biomimetic G-quartet compounds. WO2021198239, Oct 07, 2021, 2021.
160. Le, D. D.; Di Antonio, M.; Chan, L. K. M.; Balasubramanian, S., G-quadruplex ligands exhibit differential G-tetrad selectivity. *Chem. Commun.* **2015,** *51* (38), 8048-8050.
161. Marras, S. A. E.; Kramer, F. R.; Tyagi, S., Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res.* **2002,** *30* (21), e122-e122.
162. Chazotte, B., Labeling mitochondria with rhodamine 123. *Cold Spring Harb Protoc* **2011,** *2011* (7), 892-4.
163. Murrey, H. E.; Judkins, J. C.; am Ende, C. W.; Ballard, T. E.; Fang, Y.; Riccardi, K.; Di, L.; Guilmette, E. R.; Schwartz, J. W.; Fox, J. M.; Johnson, D. S., Systematic Evaluation of Bioorthogonal Reactions in Live Cells with Clickable HaloTag Ligands: Implications for Intracellular Imaging. *J. Am. Chem. Soc.* **2015,** *137* (35), 11461-11475.
164. Percivalle, C.; Doria, F.; Freccero, M., Quinone Methides as DNA Alkylating Agents: An Overview on Efficient Activation Protocols for Enhanced Target Selectivity. *Curr. Org. Chem.* **2014,** *18* (1), 19-43.
165. Cadoni, E.; Manicardi, A.; Fossépré, M.; Heirwegh, K.; Surin, M.; Madder, A., Teaching photosensitizers a new trick: red light-triggered G-quadruplex alkylation by ligand co-localization. *Chem. Commun.* **2021,** *57* (8), 1010-1013.
166. Verga, D.; Hamon, F.; Poyer, F.; Bombard, S.; Teulade-Fichou, M.-P., Photo-Cross-Linking Probes for Trapping G-Quadruplex DNA. *Angew. Chem. Int. Ed.* **2014,** *53* (4), 994-998.
167. Zhang, X.; Spiegel, J.; Martínez Cuesta, S.; Adhikari, S.; Balasubramanian, S., Chemical profiling of DNA G-quadruplex-interacting proteins in live cells. *Nat. Chem.* **2021,** *13* (7), 626-633.
168. Su, H.; Xu, J.; Chen, Y.; Wang, Q.; Lu, Z.; Chen, Y.; Chen, K.; Han, S.; Fang, Z.; Wang, P.; Yuan, B.-F.; Zhou, X., Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J. Am. Chem. Soc.* **2021,** *143* (4), 1917-1923.
169. Mackinnon, A. L.; Taunton, J., Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr. Protoc. Chem. Biol.* **2009,** *1*, 55-73.
170. McCarthy, S. M.; Gilar, M.; Gebler, J., Reversed-phase ion-pair liquid chromatography analysis and purification of small interfering RNA. *Anal. Biochem.* **2009,** *390* (2), 181-188.
171. Lena, A.; Benassi, A.; Stasi, M.; Saint-Pierre, C.; Freccero, M.; Gasparutto, D.; Bombard, S.; Doria, F.; Verga, D., Photoactivatable V-Shaped Bifunctional Quinone Methide Precursors as a New Class of Selective G-quadruplex Alkylating Agents. *Chem. Eur. J.* **2022,** *28* (35), e202200734.
172. Zhou , J.; Roembke, B. T.; Paragi , G.; Laguerre , A.; Sintim, H. O.; Fonseca Guerra , C.; Monchaud, D., Computational understanding and experimental characterization of twice-as-smart quadruplex ligands as chemical sensors of bacterial nucleotide second messengers. *Sci. Rep.* **2016,** *6*, 33888.

173. Shinamura, S.; Osaka, I.; Miyazaki, E.; Nakao, A.; Yamagishi, M.; Takeya, J.; Takimiya, K., Linear- and Angular-Shaped Naphthodithiophenes: Selective Synthesis, Properties, and Application to Organic Field-Effect Transistors. *J. Am. Chem. Soc.* **2011,** *133* (13), 5024-5035.
174. Dieck, a. H.; Heck, F., Palladium catalyzed synthesis of aryl, heterocyclic and vinylic acetylene derivatives. *J. Organomet. Chem.* **1975,** *93* (2), 259-263.
175. Cassar, L., Synthesis of aryl-and vinyl-substituted acetylene derivatives by the use of nickel and palladium complexes. *J. Organomet. Chem.* **1975,** *93* (2), 253-257.
176. Sonogashira, K.; Tohda, Y.; Hagihara, N., A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Lett.* **1975,** *16* (50), 4467-4470.
177. Doucet, H.; Hierso, J. C., Palladium-based catalytic systems for the synthesis of conjugated enynes by Sonogashira reactions and related alkynylations. *Angew. Chem. Int. Ed.* **2007,** *46* (6), 834-871.
178. Hergueta, A. R., Easy Removal of Triphenylphosphine Oxide from Reaction Mixtures by Precipitation with CaBr2. *Org. Proc. Res. Dev.* **2022,** *26* (6), 1845-1853.
179. Potter, G. T.; Jayson, G. C.; Miller, G. J.; Gardiner, J. M., An Updated Synthesis of the Diazo-Transfer Reagent Imidazole-1-sulfonyl Azide Hydrogen Sulfate. *J. Org. Chem.* **2016,** *81* (8), 3443-3446.
180. Anders, A., Selective laser excitation of bases in nucleic acids. *Applied Phys.* **1979,** *20* (3), 257-259.
181. Fonseca Guerra, C.; Zijlstra, H.; Paragi, G.; Bickelhaupt, F. M., Telomere Structure and Stability: Covalency in Hydrogen Bonds, Not Resonance Assistance, Causes Cooperativity in Guanine Quartets. *Chem. Eur. J.* **2011,** *17* (45), 12612-12622.
182. Yang, S. Y.; Amor, S.; Laguerre, A.; Wong, J. M. Y.; Monchaud, D., Real-time and quantitative fluorescent live-cell imaging with quadruplex-specific red-edge probe (G4-REP). *Biochim. Biophys. Acta* **2017,** *1861* (5, Part B), 1312-1320.
183. Eisinger, J.; Lamola, A. A.; Longworth, J. W.; Gratzer, W. B., Biological Molecules in their Excited States. *Nature* **1970,** *226* (5241), 113-118.
184. Weber, G.; Shinitzky, M., Failure of Energy Transfer between Identical Aromatic Molecules on Excitation at the Long Wave Edge of the Absorption Spectrum. *Proc. Natl. Acad. Sci. USA* **1970,** *65* (4), 823-30.
185. Cushing, S. K.; Li, M.; Huang, F.; Wu, N., Origin of Strong Excitation Wavelength Dependent Fluorescence of Graphene Oxide. *ACS Nano* **2014,** *8* (1), 1002-1013.
186. Suseela, Y. V.; Narayanaswamy, N.; Pratihar, S.; Govindaraju, T., Far-red fluorescent probes for canonical and non-canonical nucleic acid structures: current progress and future implications. *Chem. Soc. Rev.* **2018,** *47* (3), 1098-1131.
187. Lavis, L. D.; Raines, R. T., Bright Ideas for Chemical Biology. *ACS Chem. Biol.* **2008,** *3* (3), 142-155.
188. Lavis, L. D.; Raines, R. T., Bright Building Blocks for Chemical Biology. *ACS Chem. Biol.* **2014,** *9* (4), 855-866.
189. Liu, T.; Liu, X.; Valencia, M. A.; Sui, B.; Zhang, Y.; Belfield, K. D., Far-Red-Emitting TEG-Substituted Squaraine Dye: Synthesis, Optical Properties, and Selective Detection of Cyanide in Aqueous Solution. *Eur. J. Org. Chem.* **2017,** *2017* (27), 3957-3964.

190. Ilina, K.; MacCuaig, W. M.; Laramie, M.; Jeouty, J. N.; McNally, L. R.; Henary, M., Squaraine Dyes: Molecular Design for Different Applications and Remaining Challenges. *Bioconj. Chem.* **2020,** *31* (2), 194-213.
191. Grande, V.; Doria, F.; Freccero, M.; Würthner, F., An Aggregating Amphiphilic Squaraine: A Light-up Probe That Discriminates Parallel G-Quadruplexes. *Angew. Chem. Int. Ed.* **2017,** *56* (26), 7520-7524.
192. Lin, S. W.; Sun, Q.; Ge, Z. M.; Wang, X.; Ye, J.; Li, R. T., Synthesis and structure-analgesic activity relationships of a novel series of monospirocyclopiperazinium salts (MSPZ). *Bioorg. Med. Chem. Lett.* **2011,** *21* (3), 940-3.
193. Shin, I.-S.; Gwon, S.-Y.; Kim, S.-H., Chromogenic sensing of biological thiols using squarylium dye. *Spectrochim. Acta A* **2014,** *120*, 642-645.
194. Twibanire, J.-d. A. K.; Grindley, T. B., Efficient and Controllably Selective Preparation of Esters Using Uronium-Based Coupling Agents. *Org. Lett.* **2011,** *13* (12), 2988-2991.
195. Diehl, K. L.; Bachman, J. L.; Anslyn, E. V., Tuning thiol addition to squaraines by ortho-substitution and the use of serum albumin. *Dyes Pig.* **2017,** *141*, 316-324.
196. Wu, N.; Lan, J.; Yan, L.; You, J., A sensitive colorimetric and fluorescent sensor based on imidazolium-functionalized squaraines for the detection of GTP and alkaline phosphatase in aqueous solution. *Chem. Commun.* **2014,** *50* (34), 4438-4441.
197. Ishihara, K.; Shioiri, T.; Matsugi, M., SNAr azidation of phenolic functions utilizing diphenyl phosphorazidate. *Tet. Lett.* **2020,** *61* (7), 151493.
198. Kaczmarek-Kedziera, A.; Kędziera, D., Molecular aspects of squaraine dyes aggregation and its influence on spectroscopic properties. *Theor. Chem. Acc.* **2016,** *135* (9), 214.
199. McKerrow, A. J.; Buncel, E.; Kazmaier, P. M., Aggregation of squaraine dyes: Structure–property relationships and solvent effects. *Can. J. Chem.* **1995,** *73* (10), 1605-1615.
200. Mishra, A.; Behera, R. K.; Behera, P. K.; Mishra, B. K.; Behera, G. B., Cyanines during the 1990s: A Review. *Chem. Rev.* **2000,** *100* (6), 1973-2012.
201. Mujumdar, R. B.; Ernst, L. A.; Mujumdar, S. R.; Lewis, C. J.; Waggoner, A. S., Cyanine dye labeling reagents: Sulfoindocyanine succinimidyl esters. *Bioconj. Chem.* **1993,** *4* (2), 105-111.

Table of Contents Experimental Part:

Chemicals:

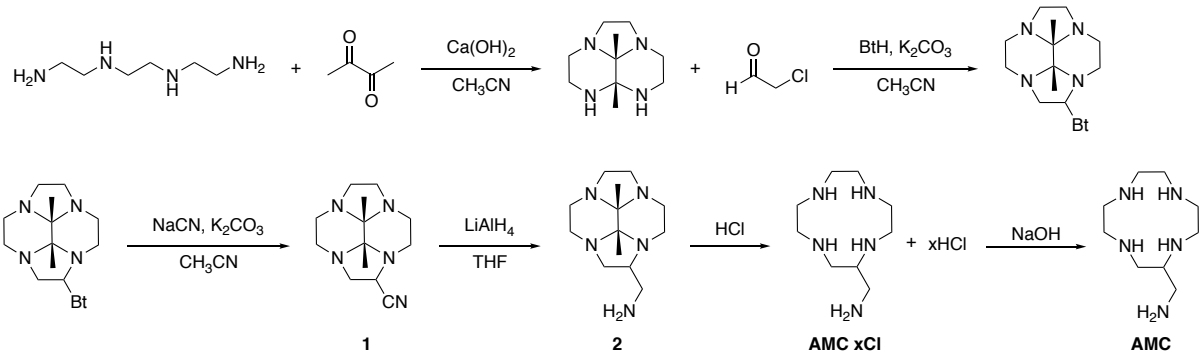
Chemicals were purchased from Alfa Aesar, Sigma Aldrich, or Iris Biotech GmbH. Solvents were purchased from VWR or Carlo Erba. Unless noted otherwise, all commercially available reagents and solvents were used without further purification. Dicalite was purchased from Carlo Erba. TLC were carried out on Merck DC Kieselgel 60 F-254 aluminum sheets. The spots were directly visualized by staining with ninhydrine or through illumination with a UV lamp (λ = 254 nm). Column chromatography purifications were performed manually on silica gel (40-63 µm) from Sigma-Aldrich (technical grade). Dry solvents (HPLC grade) were dried over alumina cartridges using a solvent purification system PureSolv PS-MD-5 model from Innovative Technology. Peptide synthesis-grade DIPEA and TFA were provided by Iris Biotech GmbH. HPLC-gradient grade MeCN used for HPLC-MS analyses was obtained from Carlo Erba. MeCN used in semi-preparative RP-HPLC purifications was obtained from Biosolve or VWR (technical, +99% but distilled prior to use). All aq. mobile-phases for HPLC were prepared using water purified with a PURELAB Ultra system from ELGA (purified to 18.2 MΩ.cm). Yields were calculated based on isolation of the compounds.

Instruments and methods:

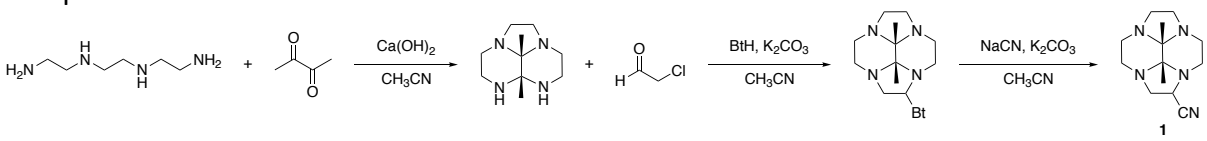
Lyophilization was performed with a Christ Alpha 2-4 LD plus. ^{1}H-, and ^{13}C- NMR spectra were recorded on a Bruker Avance 500 II HD (Bruker) at RT. Chemical shifts are expressed in parts per million (ppm) from the residual nondeuterated solvent signal. J values are expressed in Hz. Coupling constants (J) are reported in hertz (Hz). Standard abbreviations indicating multiplicity were used as follows: s = singlet, d = doublet, t =triplet, q = quadruplet, m = multiplet, br = broad. High resolution mass spectrometry analysis were recorded on LTQ Orbitrap XL or Orbitrap Exploris 240 mass spectrometers (Thermo Scientific) with an ESI ion source, or a Bruker MicroFlex LRF MALDI (Matrix-Assisted Laser Desorption/Ionisation) coupled with a time of flight detector (TOF) . HPLC-MS analyses were performed on a Thermo-Dionex Ultimate 3000 instrument (pump + autosampler at 20°C + column oven at 25°C) equipped with a diode array detector (Thermo-Dionex DAD 3000-RS) and a MSQ Plus single quadrupole mass spectrometer. The corresponding low-resolution mass spectra (LRMS) were recorded with this latter mass spectrometer, with an electrospray (ESI) source (HPLC-MS coupling mode). HPLC systems were equipped with a Phenomenex Kinetex C18 column, 2.6µm, 2.1 × 50 mm or a Jupiter Proteo 4 µm 90Å column, 250 x 4.6 mm. *Method*: from 5% t

o 100% H_2O/MeCN + 0.1% formic acid (FA) in 5 min. Particularly lipophilic compounds were analyzed using the same column with a method going from 50% to 100% H_2O/MeCN + 0.1% formic acid (FA) in 5 min. Final compounds (**BioTASQ, BioCyTASQ**, **TriazoTASQ**, **BiotriazoTASQ**, **MultiTASQ**, **^{Az}MultiTASQ**, **photoMultiTASQ**, **N-TASQ, ^{Tz}N-TASQ**, **Square-TASQ v.1 and Square-TASQ v.2**) were analyzed with a ACCUCORE column, method: from 0% to 20% MeCN/H2O+0.1% formic acid (FA) in 2.4 min and from 20% to 100% H_2O/MeCN + 0.1% formic acid (FA) in 2.6 min. Purifications by semi-preparative HPLC were performed on a Thermo-Dionex Ultimate 3000 instrument equipped with a RS Variable Detector (four distinct wavelengths). HPLC purification system was equipped with Jupiter Proteo 4 µm 90 column (250 x 21.2 mm, AXIA packed). Low-resolution mass spectra (LRMS) were recorded on a Thermo Scientific MSQ Plus single quadrupole equipped with an electrospray (ESI) source (LC-MS coupling) or a Brüker Amazon SL (direct injection in MeOH).

Aminomethylcyclen:



Compound **1** :



Technical grade triethylenetetramine (419 g, 60% purity as specified by the provider) was dissolved in toluene (800 mL) and cooled to 0°C with an ice bath. The solution was vigorously stirred and water (115mL) was slowly added monitoring the temperature to avoid excessive heating (< 36°C). A fine precipitate was formed and filtered. The solid was then dissolved in methanol, cooled with an ice bath and magnesium sulfate was added to get rid of the remaining water. After filtration and drying, the amine was obtained as a colorless oil (140 g, extraction yield 55%).

^{1}H NMR : (500 MHz, D_2O) δ 2.74 – 2.68 (m, 8H), 2.64 (t, *J* = 6.2 Hz, 4H).

The re-crystalized triethylenetetramine (140 g, 0.96 mol) was dissolved in acetonitrile (5 L) to which calcium hydroxide (142 g, 1.92 mol) was added and the solution was cooled to 0 °C and 2,3-butanedione (84 mL, 0.96 mol) was added dropwise to the suspension over 2 hours. After an additional 20 minutes stirring, the suspension was filtered and the reaction mixture was used in the next synthetic step without purification. The acetonitrile solution was cooled back to 0 °C and 1*H*-Benzotriazole (114 g, 0.96 mol) and K_2CO_3 (65 g, 1.92 mol) were added. Meanwhile, in an Erlenmeyer flask, chloroacetaldehyde (150 mL, 0.96 mol, 55% v/v in water) was cooled with an ice bath and magnesium sulfate was added to dry the solution. After filtration, the latter was added dropwise to the acetonitrile suspension of *bis*-aminal and the

reaction was allowed to stir for 2 hours at 0°C. Sodium cyanide (47 g, 0.96 mol) was carefully added and the reaction mixture was let to stir overnight at RT. The suspension was filtered on celite and the solvent was evaporated under reduced pressure. To the solid was first added chloroform (600 mL) and vigorously stirred, the suspension was then filtered on celite. After evaporation of the solvent under reduced pressure the same procedure was repeated with diehtylether (2L). Organic phases were pooled together and evaporated under vacuum. The resulting oil was dissolved in chloroform, added alumina (3 g of alumina per 1 g of crude) and the solvent was evaporated under vacuo. Two purification techniques were successfully carried out to obtain compound **1**, the mixture of alumina and reaction crude was used to perform a solid deposition for column chromatography (Method A), or **1** was extracted from the mixture of alumina and reaction crude by Soxhlet extraction (Method B).

Method A) Alumina column chromatography, (Pentane/DCM 0% to 30%) to obtain compound **1** (39 g, 0.25 mol) in 26% chemical yield.

^{1}H NMR: (500 MHz, $CDCl_3$) δ 4.22 (dd, *J* = 8.1, 6.0 Hz, 1H), 3.53 – 3.34 (m, 3H), 3.22 – 2.59 (m, 26H), 1.33 (s, 3H), 1.14 – 1.02 (m, 9H).

^{13}C NMR: (126 MHz, $CDCl_3$) δ 120.4, 119.7, 80.0, 79.5, 78.3, 78.0, 56.5, 55.7, 51.1, 50.8, 50.4, 50.2, 49.8, 49.7, 47.9, 47.4, 46.5, 45.8, 44.3, 43.8, 42.9, 42.6, 16.4, 14.4, 14.1, 13.9.

Method B) Soxhlet extraction with pentane for 42 hours to obtain the desired compound **1** in 38% chemical yield (6.8 g, 0.027 mol).

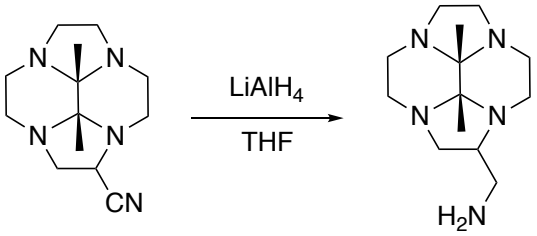
^{1}H NMR: (500 MHz, $CDCl_3$) δ 4.22 (dd, *J* = 8.1, 6.0 Hz, 1H), 3.56 – 3.37 (m, 3H), 3.23 – 2.49 (m, 26H), 1.33 (s, 3H), 1.18 – 1.00 (m, 9H).

Reported NMR[121]:

^{1}H NMR *(300 MHz, $CDCl_3$, 300 K): δ 4.21 (m, 1 H), 3.34–3.52 (m, 3 H), 2.48– 3.22 (m, 26 H), 1.31 (s, 3 H), 1.07 (m, 9 H).*

^{13}C NMR *(75 MHz, $CDCl_3$, 300 K): 120.4, 119.7 (CN), 80.0, 79.5, 78.3, 78.0 (N-C-N), 56.5, 55.7 (CH), 53.4, 51.1, 50.7, 50.3, 49.7, 47.9, 47.4, 46.5, 45.8, 44.2,43.8, 42.8, 42.6, (CH_2-α), 16.4, 14.4, 14.0, 13.8, (CH_3).*

Compound **2**:



1 **2**

Anhydrous tetrahydrofuran was added to a three-neck round-bottom flask and cooled at -78°C under inert atmosphere, to which a $LiAlH_4$ (72 mL, 0.173 mol, 2.4 M THF solution) was added under inert atmosphere. Compound **1** (21.4g, 0.086 mol) was dissolved in anhydrous THF (21.3 g, 0.086 mol) and added dropwise to the $LiAlH_4$ solution at -78°C. The cooling bath was removed and the reaction mixture was allowed to stir overnight at RT. The excess of $LiAlH_4$ was quenched by cooling the solution (-78°C) and gentle addition of water and the solvent was evaporated. The resulting solid was extracted three times with chloroform (3 x 500 mL) and filtered. After evaporation of the solvent under reduced pressure, the desired compound was isolated without further purification (19.9 g, 0.070 mol, 81% chemical yield).

^{1}H NMR (500 MHz, $CDCl_3$) δ 3.46 – 2.57 (m, 32H), 2.54 – 2.34 (m, 4H), 1.28 – 1.20 (m, 4H), 1.15 (dd, *J* = 6.5, 2.9 Hz, 4H), 1.12 – 1.08 (m, 4H), 1.08 – 1.04 (m, 4H).

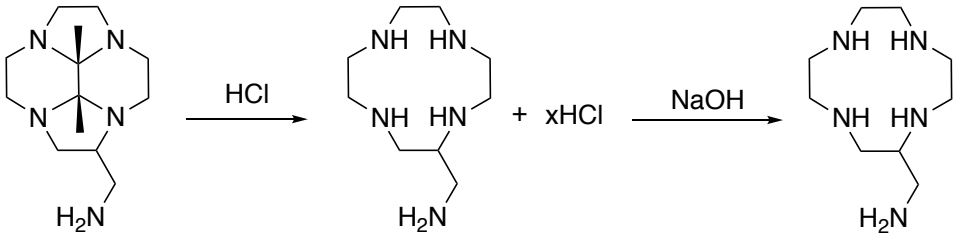
^{13}C NMR (126 MHz, $CDCl_3$) δ 80.3, 79.4, 78.4, 63.3, 60.4, 52.3, 51.4, 51.3, 50.5, 50.1, 46.4, 46.4, 46.1, 45.6, 45.6, 45.2, 44.80, 44.78, 44.4, 44.3, 43.9, 43.3, 23.0, 16.6, 13.98, 13.90.

Literature NMR:

^{1}H NMR (300 MHz, CDCl3, 300 K): 3.44-2.36 (m, 34 H), 1.55–1.05 (m, 16 H).

^{13}C NMR (75 MHz, CDCl3, 300 K): 80.3, 79.4, 78.4, 77.3 (N-C-N), 63.3, 60.5 (CH), 55.7, 52.3, 51.4, 51.3, 50.5, 50.1, 48.6, 46.1, 45.6, 45.2, 44.8 (CH_2- α), 44.4, 43.9, 43.3 (X2), 23.0, 16.6, 14.0, 13.9 (CH_3).

AMC:



2 **AMC xCl** **AMC**

Compound **2** (19.9 g, 0.07 mol) was dissolved in HCl (46 mL, 37% solution, 0.55 mol) and heated at reflux overnight. The precipitate was filtered and washed with cold ethanol and diethyl ether to obtain the desired compound as a hydrochloride salt (15.1 g, 0.011 mol, Y: 60%). Moreover, to obtain the basic form, the protonated AMC was resuspended in a 16 M aqueous solution of NaOH and extracted three times with chloroform. The organic phase was then dried with $MgSO_4$, filtered, and evaporated under vacuo to afford the desired cyclic polyamine AMC as an oil (quantitative).

Protonated form:

^{1}H NMR (500 MHz, Deuterium Oxide) δ 3.50 – 3.03 (m, 16H), 2.92 (m, 1H).

^{13}C NMR (126 MHz, D_2O) δ 51.4, 47.0, 45.6, 44.3, 44.0, 43.9, 42.3, 42.0, 39.3.

After extraction:

^{1}H NMR (500 MHz, *$CDCl_3$*) δ 2.75 – 2.41 (m, 17H), 2.01 (m, 6H).

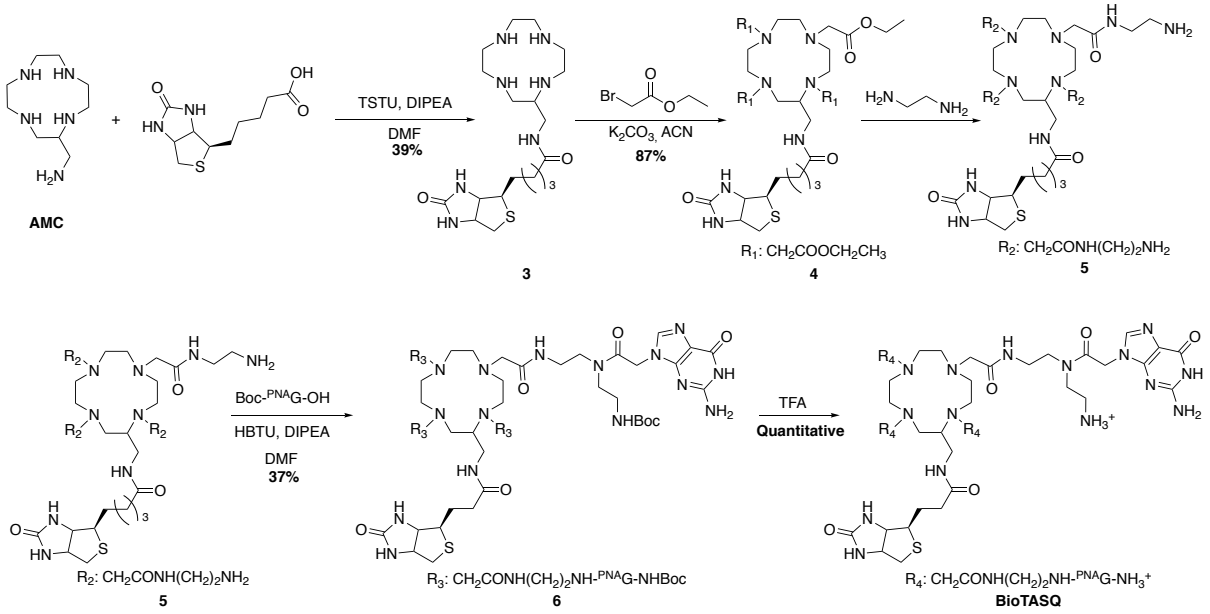
^{13}C NMR (126 MHz, Chloroform-*d*) δ 57.2, 49.3, 48.9, 47.4, 46.7, 46.3, 44.2, 44.1, 41.5.

Literature NMR:

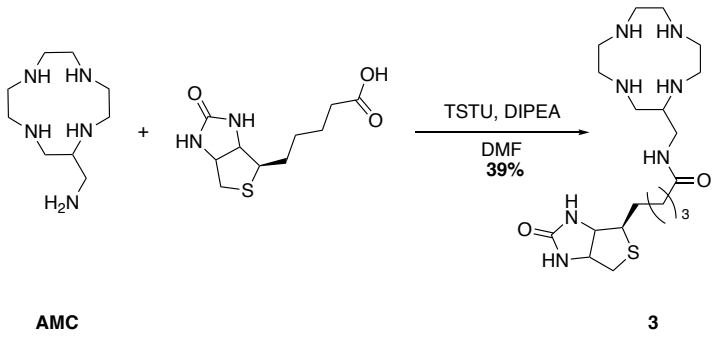
^{1}H NMR (300 MHz, $CDCl_3$, 300 K): δ 2.66–2.49 (m, 17 H), 1.75 (s, 6 H).

^{13}C NMR (75 MHz, $CDCl_3$, 300 K): 57.2 (CH), 48.9, 46.8, 46.7, 46.4, 46.3, 46.0, 44.6, 44.1 (CH2-α).

BioTASQ :



Compound **3** :



Biotin (652.3 mg, 2.67 mmol, 0.9 equiv.) was dissolved in anhydrous DMF (0.250 mL), to which TSTU (*N,N,N',N'*-Tetramethyl-*O*-(*N*-succinimidyl)uronium tetrafluoroborate) (803.7 mg, 2.67 mmol, 0.9 equiv.) and *N,N*-Diisopropylethylamine (DIPEA) (931 µL, 5.3 mmol, 1.8 equiv.) were added in order to obtain *N*-hydroxysuccinimide ester. The reaction was left to stir and the conversion was monitored by HPLC-MS until completion (usually 1h). The reaction mixture containing the activated ester was then added dropwise to a DMF (1 mL) solution containing AMC (594.9 mg, 2.96 mmol). The reaction was let to stir for one hour and, after removal of the solvent, the crude was dissolved in water and purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient from 2 to 50% over 50 minutes, r.t: 13 minutes).

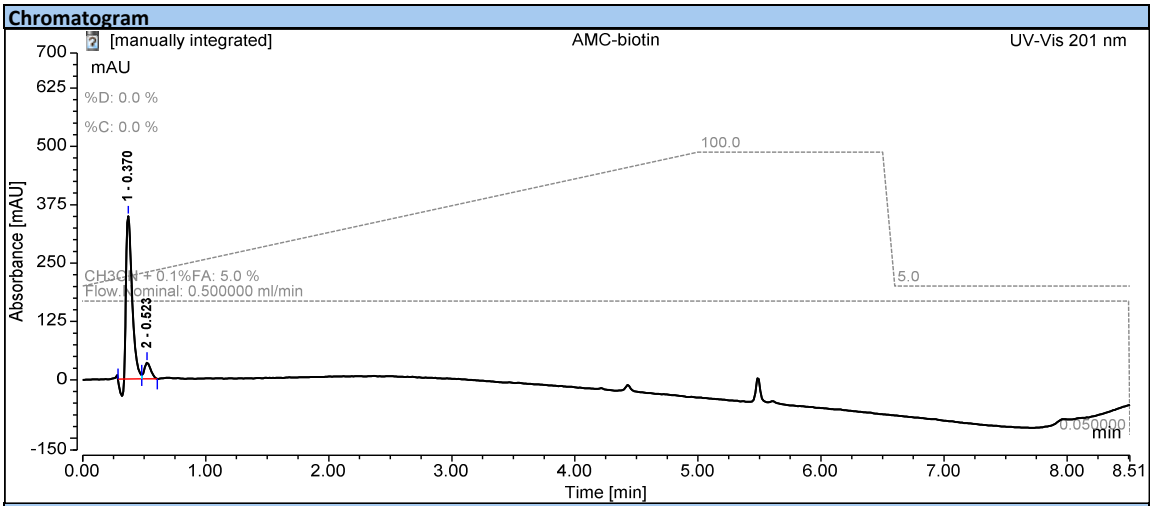
After evaporation of the solvents and lyophilisation, compound **3** was obtained in 39 % chemical yield (890.2 mg, 1.2 mmol).

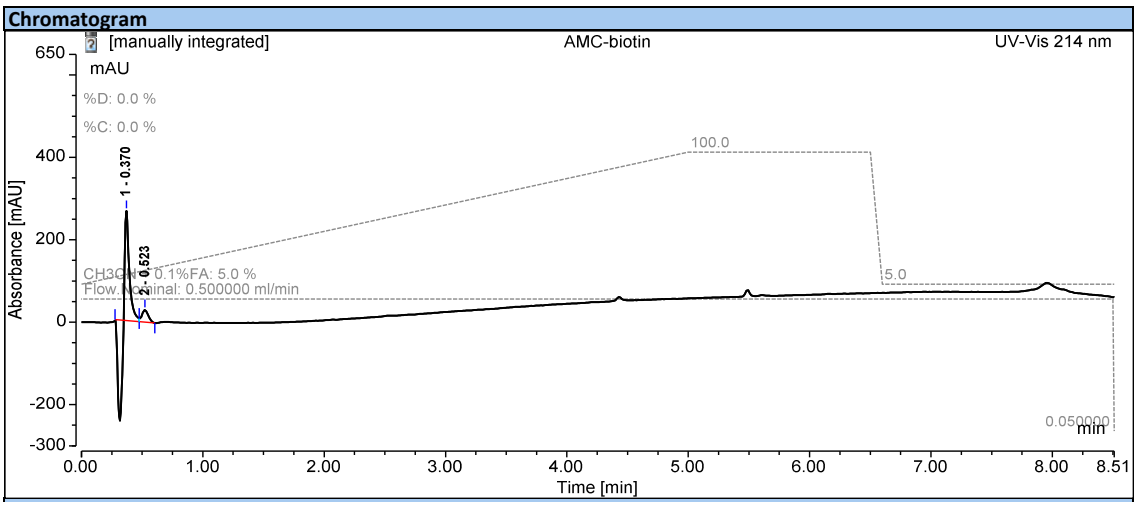
^{1}H NMR: (500 MHz, Deuterium Oxide) δ 4.60 (dd, *J* = 7.9, 4.9 Hz, 1H), 4.41 (dd, *J* = 8.0, 4.5 Hz, 1H), 3.51 – 2.73 (m, 20H), 2.30 (t, *J* = 7.3 Hz, 2H), 1.78 – 1.50 (m, 4H), 1.48 – 1.29 (m, 4H). (Signals of intracyclic amines and amide are missing due to exchange with the solvent)

^{13}C NMR: (126 MHz, Deuterium Oxide) δ 177.72, 165.35, 62.13, 60.27, 55.44, 51.95, 46.28, 44.23, 44.04, 42.62, 42.07, 39.65, 39.13, 39.09, 35.28, 28.01, 27.69, 24.88, 16.23.

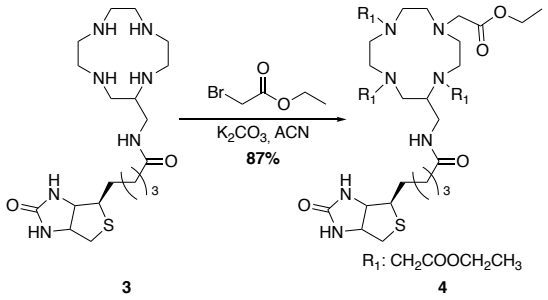
LR-MS: $[M+H]^+$ (m/z): 428.27 and 450.26 $[M+Na]^+$ (calcd. for $C_{19}H_{38}N_7O_2S$: 428.27)

HPLC:





Compound **4** :



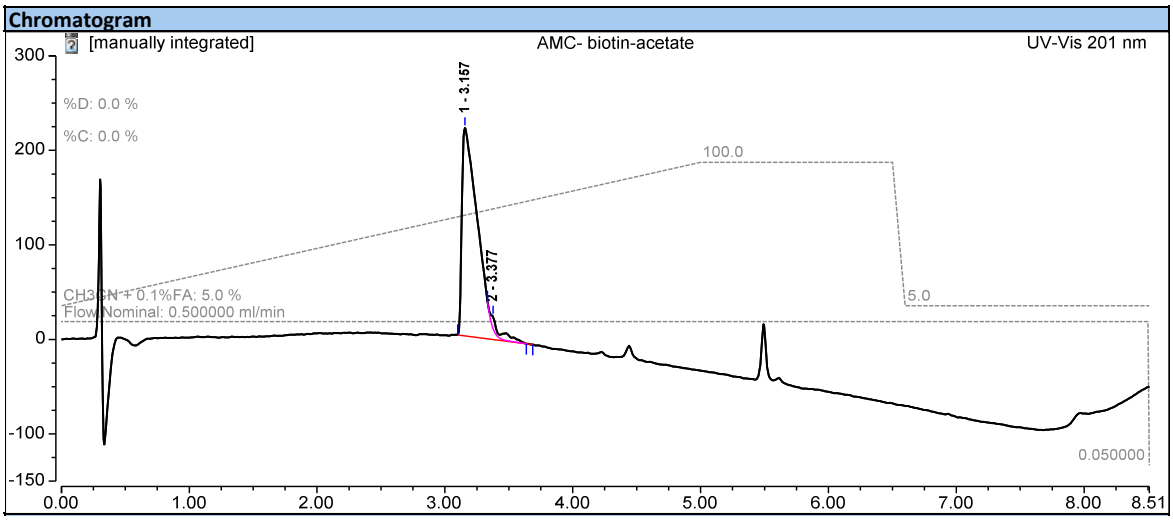
3 **4**

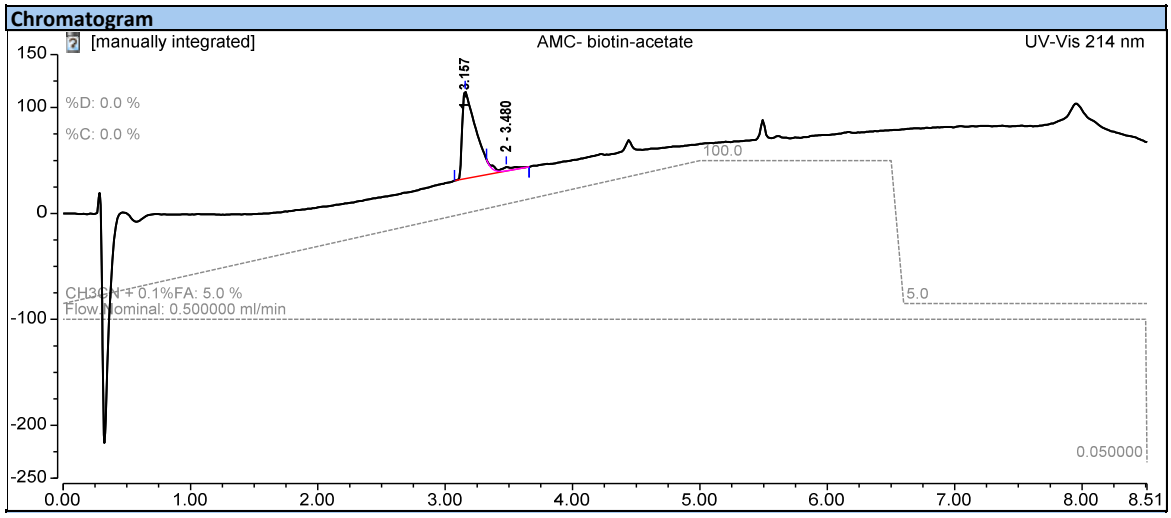
To a suspension of **3** (369.7 mg, 0.48 mmol, 1 equiv.) and K_2CO_3 (529.9 mg, 3.84 mmol, 8 equiv.) in acetonitrile (4 mL) was added ethylbromoacetate (212 µL, 1.92 mmol, 4 equiv.) and the mixture was left to stir at 40°C and the conversion was monitored by RP-HPLC-MS (full conversion 48 h). The reaction mixture was filtered and the solvent was evaporated under reduced pressure. The crude was purified by semi-preparative RP-HPLC (gradient from 30 to 100% over 50 minutes, r.t: 16 minutes) to afford **4** as a white solid (322.6 mg, 0.42 mmol, chemical yield 87 %).

^{1}H NMR (500 MHz, d^6-DMSO): δ 7.30 (s broad, 1H); 5.50- 5.09 (m, 2H); 4.26-4.07 (m, 8H); 3.46-2.04 (m, 27H), 1.65-1.42 (m, 2H); 1.29-1.18 (m, 18H); 0.88 (m, 3H).

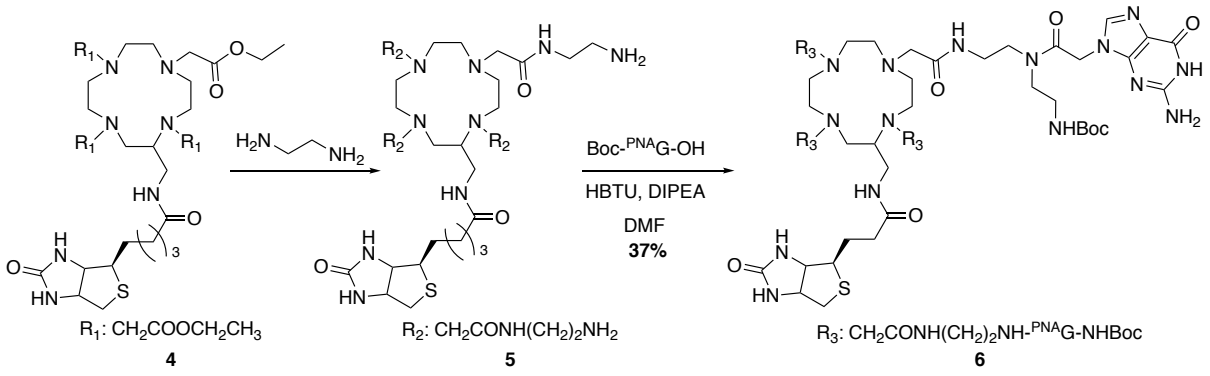
HR-MS (MALDI TOF): $[M+H]^+$ (m/z): 772.43 (calcd. for $C_{35}H_{62}N_7O_{10}S$: 772.42).

HPLC:





Compound **6** :

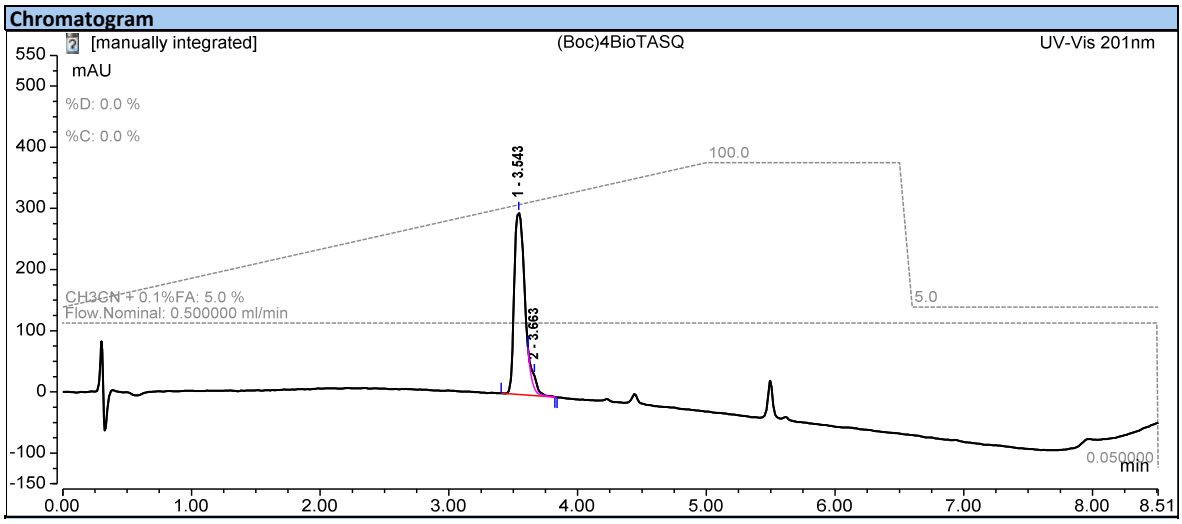


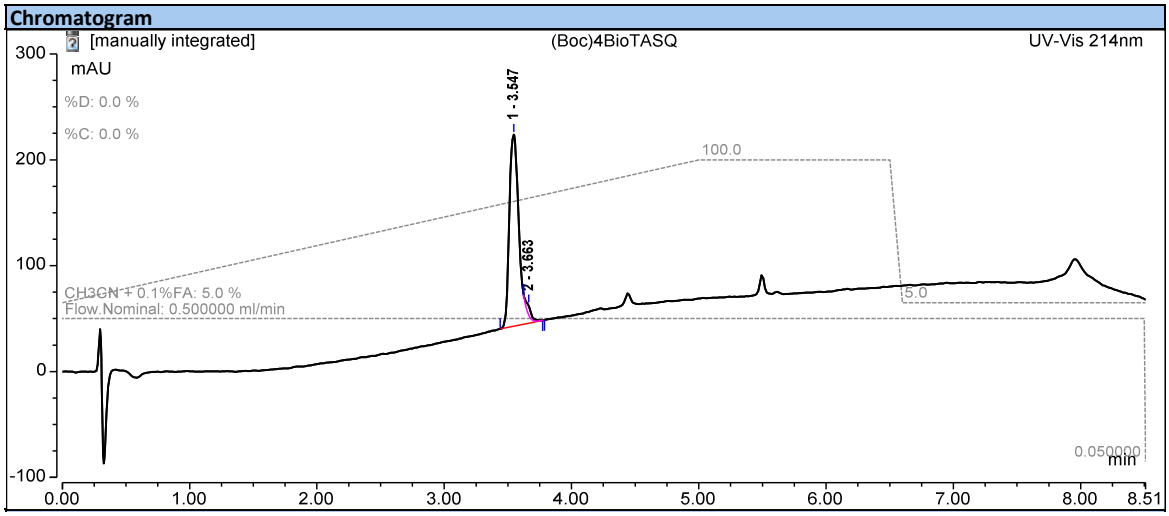
A solution of **4** (30.1 mg, 0.039 mmol, 1 equiv.) in ethylene diamine (1 mL) was stirred at room temperature for 8 days, after which ethylene diamine was thoroughly evaporated under reduced pressure. The residue was then directly used in the following step. Boc^{PNA}G (63.9 mg, 0.16 mmol, 4.1 equiv.), HBTU (60.6 mg, 0.16 mmol, 4.1 equiv.) and DIPEA (27 µL, 0.078 mmol,4 equiv.) were dissolved in DMF (1.6 mL). The mixture was stirred at room temperature for 5 min and a solution of **5** and DIPEA (54 µL, 0.31 mmol, 8 equiv.) in DMF (2 mL) was added dropwise. The reaction mixture was left to stir overnight at room temperature and the conversion was monitored by HPLC-MS. After reaction completion, the solvent was removed under reduced pressure. The residue was purified by semi-preparative HPLC to afford **6** as a white solid (34.5 mg, 0.014 mmol, 37% chemical yield).

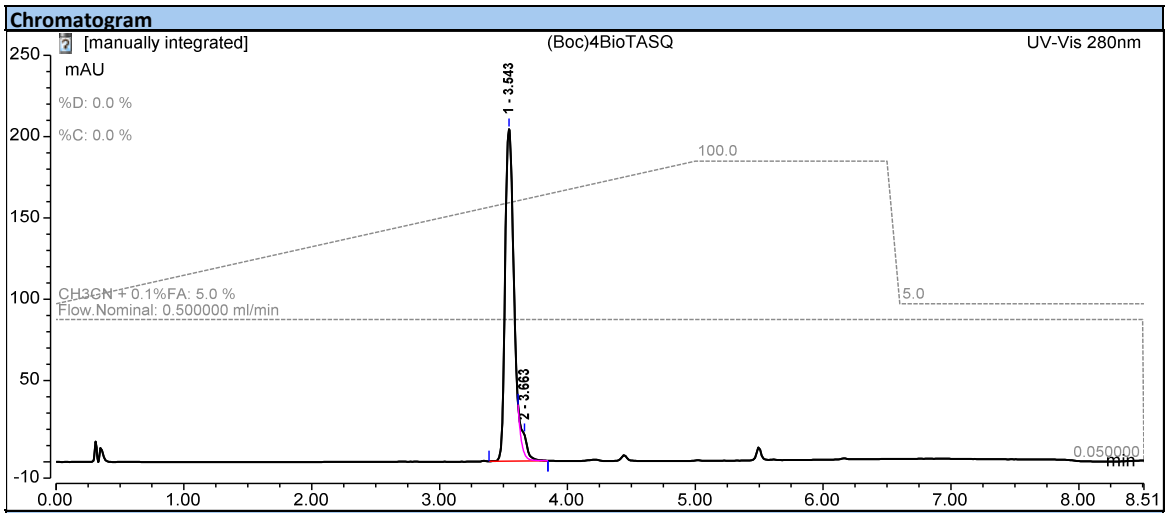
^{1}H NMR (500 MHz, d_6- DMSO): δ 10.77 (s broad, 2H); 7.9-7.42 (m, 14H); 7.05 (s broad, 2H); 6.90-6.62 (m, 5H); 6.17 (dd, 5H); 5.32- 4.84 (m, 9H); 4.26-4.03 (m, 10H); 3.52-3.07 (m, 43H), 1.82 (m, 5H); 1.39-0.99 (m, 58H).

HR-MS (MALDI-TOF, m/z): $[M+H]^+$ = 2393.22 (calcd for $C_{99}H_{153}N_{43}O_{26}S$ = 2393.17)

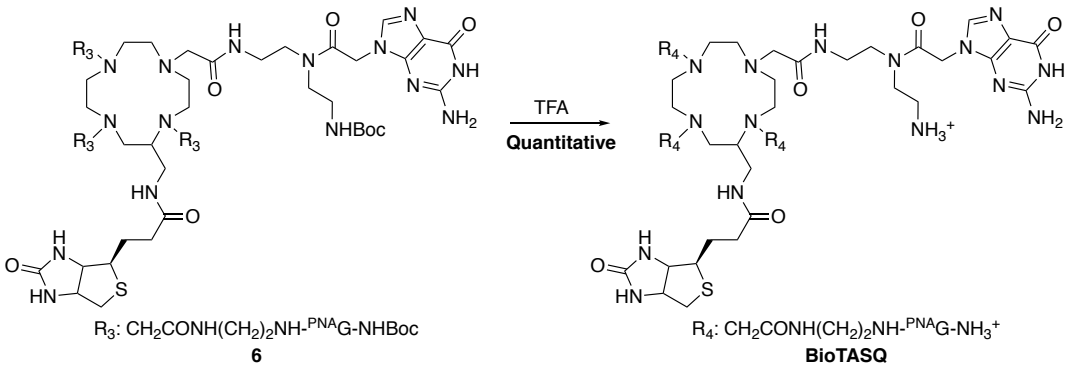
HPLC:







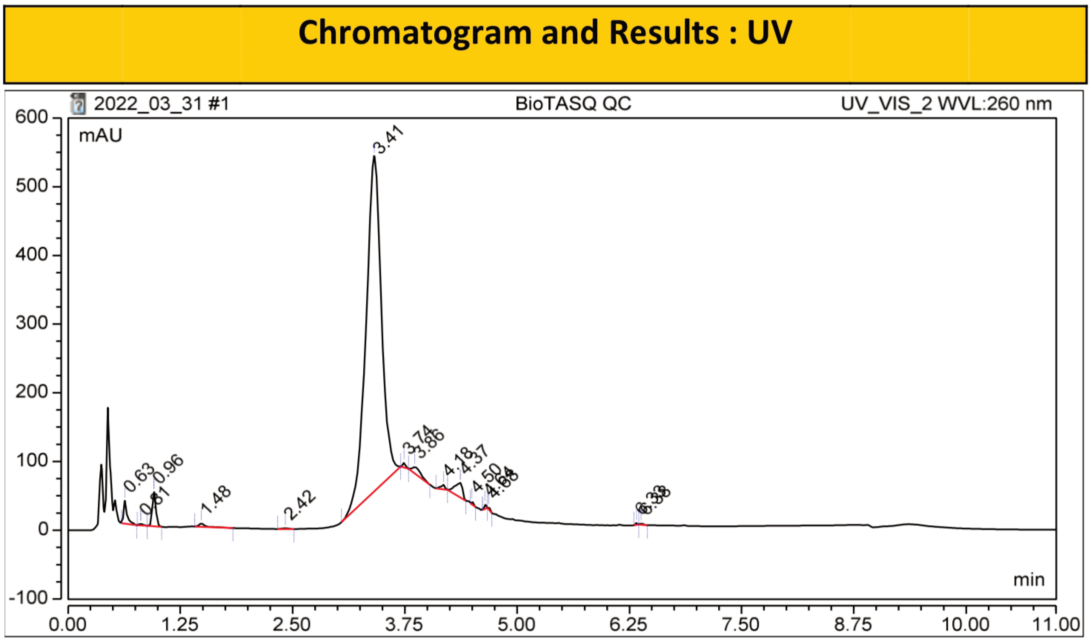
BioTASQ :



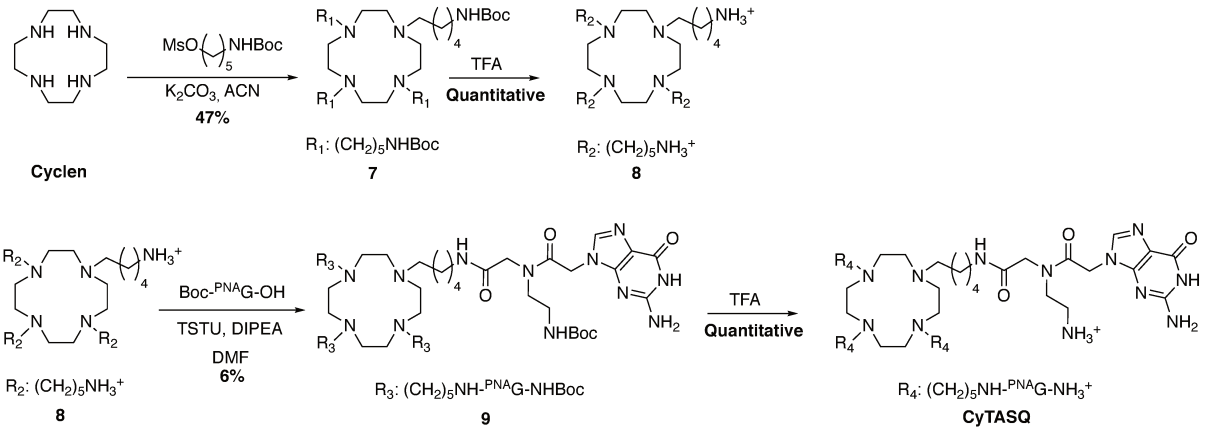
Compound **6** (1.95 mg, 0.81 µmol) was dissolved in TFA (100 µL) and the complete deprotection was assessed *via* RP-HPLC-MS, the mixture was diluted with water and the compound was freeze-dried to afford **BioTASQ.** (2 mg, 0.81 µmol, chemical yield 100%).

HR-MS :$[M+H]^{2+}$ *m/z* = 967.04040 (calc. for $C_{79}H_{121}N_{43}O_{18}S$: 967.03802)

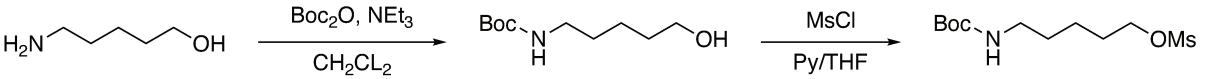
HPLC:



CyTASQ:



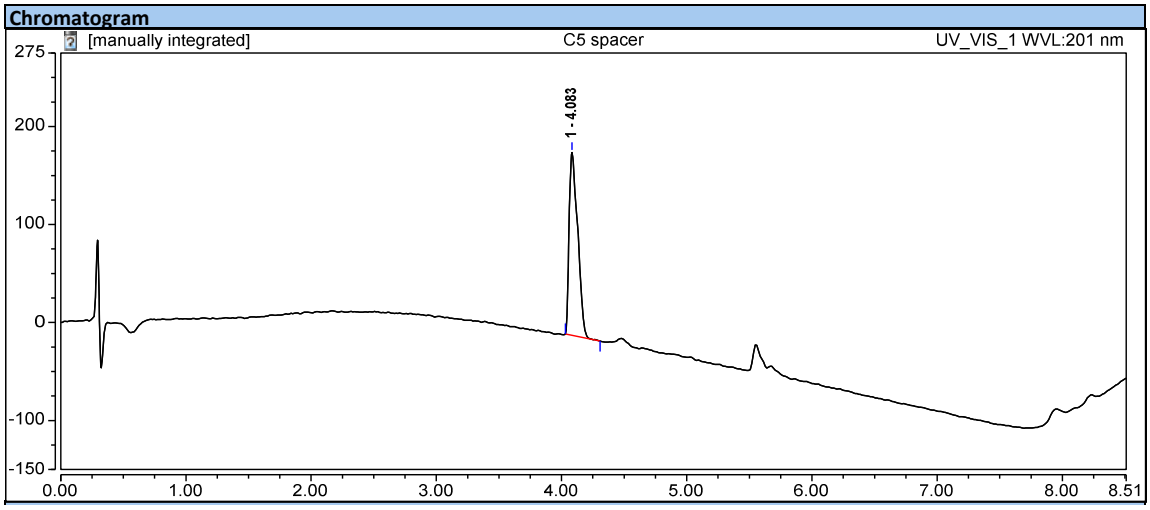
5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate :



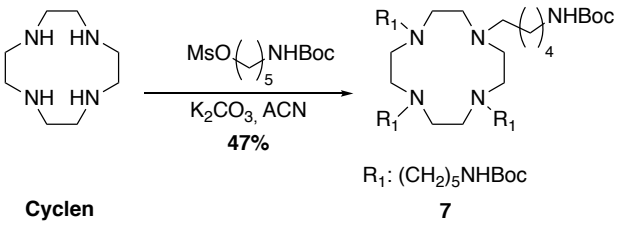
To a solution of 5-aminopentanol (1 g, 9.7 mmol) in CH_2Cl_2 (100 mL) was added triethylamine (1.5 mL, 10.7 mmol, 1.1 equiv.) and the solution was cooled down to 0°C with an ice bath. Di-*tert*-butyl dicarbonate (2.114 g, 9.7 mmol, 1 equiv.) was added portion-wise and the solution

was allowed to warm to RT and stir overnight. The reaction was monitored by TLC (CH_2Cl_2/MeOH 9:1) and the solvent was evaporated after completion. The residue was purified by silica gel chromatography (CH_2Cl_2/MeOH 4% to 6%) to afford 5-((*tert*-butoxycarbonyl)amino)pentanol. The latter (1 g, 4.5 mmol) was then dissolved in THF (40 mL) with pyridine (2mL, 45.2 mmol, 10 equiv.) and MsCl (2mL, 45.2 mmol, 10 equiv.) was added dropwise. The solution was allowed to stir overnight at RT. THF was evaporated under vacuum and to the mixture were then added 30 mL of ethyl acetate and 60 ml of acidified brine (30 mL of aq solution HCl 1M and 30 mL of brine). The aqueous phase was extracted 3 times with ethyl acetate (30 mL each), the organic phases were pulled together, dried over $MgSO_4$ filtered and dried under vacuum. The residue was purified by precipitation, dissolving the crude into a minimal amount of CH_2Cl_2 and adding an excess of cold heptane. After filtration the desired compound was successfully isolated as an amorphous solid (2.53 g, 9.0 mmol, 93% chemical yield).

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.16 (t, J = 6.4 Hz, 2H), 3.06 (t, J = 6.4 Hz, 2H), 2.94 (s, 3H), 1.80 – 1.60 (m, 2H), 1.49 – 1.42 (m, 2H), 1.38 (s, 9H).



Compound **7**:



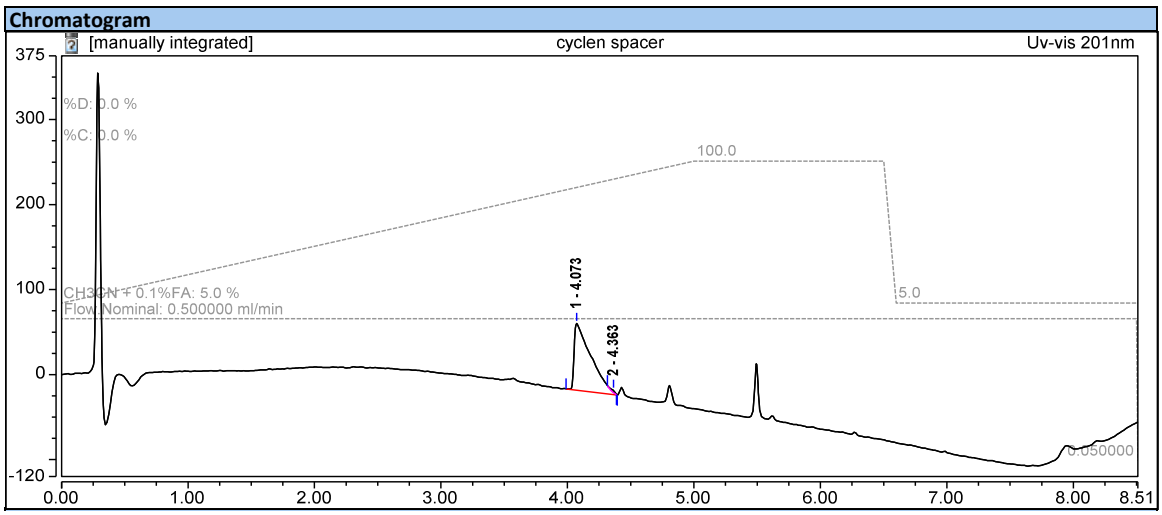
Cyclen **7**

To a solution of cyclen (29.2 mg, 0.17 mmol, 1 equiv.) in MeCN (1.5 mL) was added 5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate (382.2 mg, 1.36 mmol, 8 equiv.) and K_2CO_3 (187.9 mg, 1.36 mmol, 8 equiv.) and the solution was stirred for 48 hours at 70°C until complete conversion of the starting material. The crude mixture was filtered, and concentrated under vacuum. The residue was then purified by silica gel column chromatography (CH_2Cl_2/MeOH, 9:1). After evaporation of the solvents, the compound **7** was obtained (74.2 mg, 0.08 mmol, 47 % chemical yield).

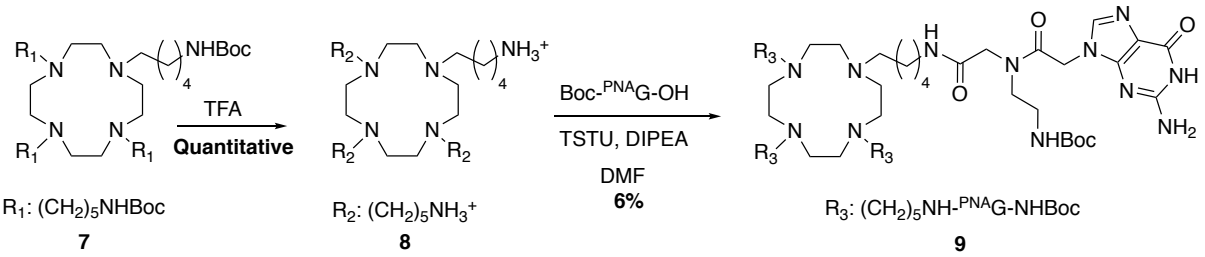
^{1}H NMR (500 MHz, $CDCl_3$): δ 3.10 (q, *J* = 6.7 Hz, 8H), 2.82 (d, *J* = 87.2 Hz, 24H), 2.07 (s, 4H), 1.52 (m, *J* = 15.1, 7.8, 6.8 Hz, 16H), 1.43 (s, 36H), 1.31 (h, *J* = 7.2, 6.6 Hz, 8H).

ESI-LRMS: $[M+H]^+$ *m/z* = 913.95 (calcd. for $C_{48}H_{96}N_8O_8$:913.74).

HPLC :



Compound **9**:

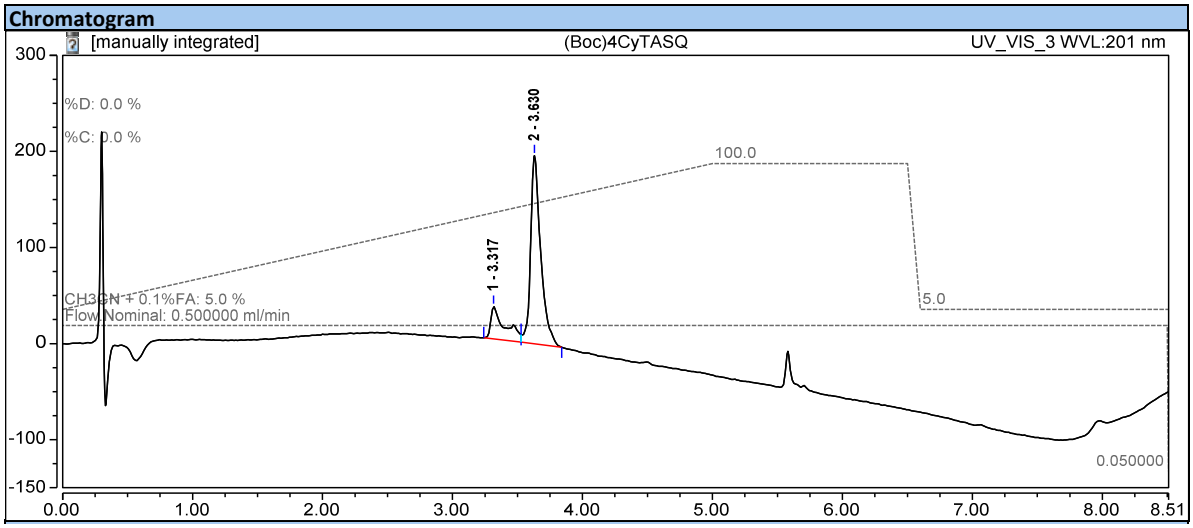


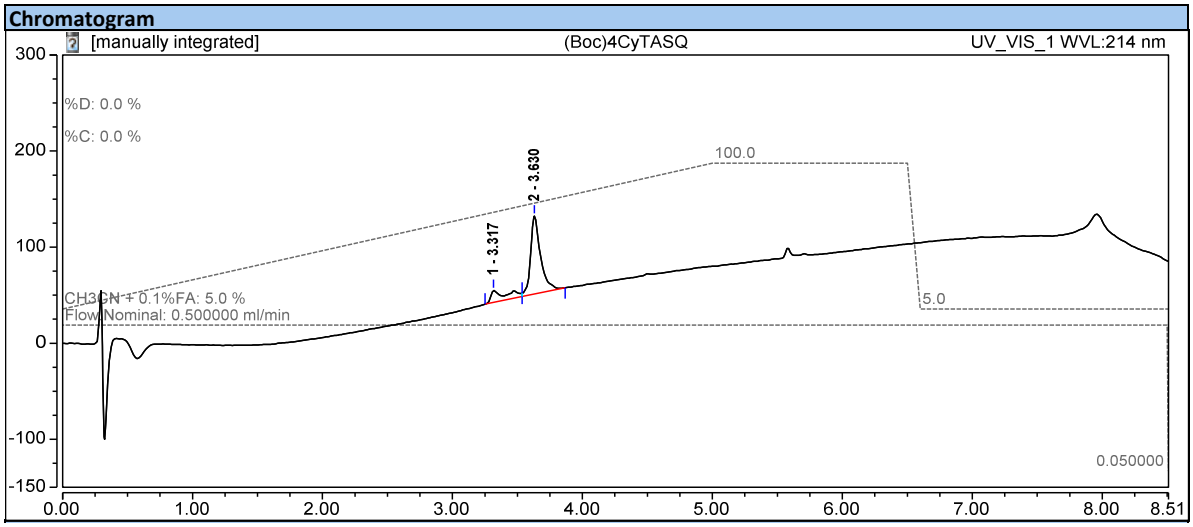
Compound **7** (90.1 mg, 0.098 mmol, 1 equiv.) was stirred with 2 mL of TFA during 1 hour. After evaporation of the TFA, the desired compound **8** was obtained. Boc-^{PNA}G (180.8 mg, 0.44 mmol, 4.5 equiv.), TSTU (132.4 mg, 0.44 mmol, 4.5 equiv.) were dissolved in DMF (1 mL), and DIPEA was added (68 µL, 0.39 mmol, 4 equiv.). After 1 h, complete activation of the carboxylic acid was assessed by HPLC and a solution of **8** and DIPEA (135 µL, 0.78 mmol, 8 equiv.) in DMF (1 mL) was added to the mixture. The mixture was stirred at RT for 48 hours. The solution was then concentrated under vacuum, added water (2mL), and purified by RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 40 % over 20 minutes). After evaporation of the solvents, compound **9** was obtained (12.3 mg, 5.92 µmol, 6 % chemical yield). As seen in HPLC-MS track, after isolation, partial deprotection of one of the Boc moieties was observed (rt: 3.32 min).

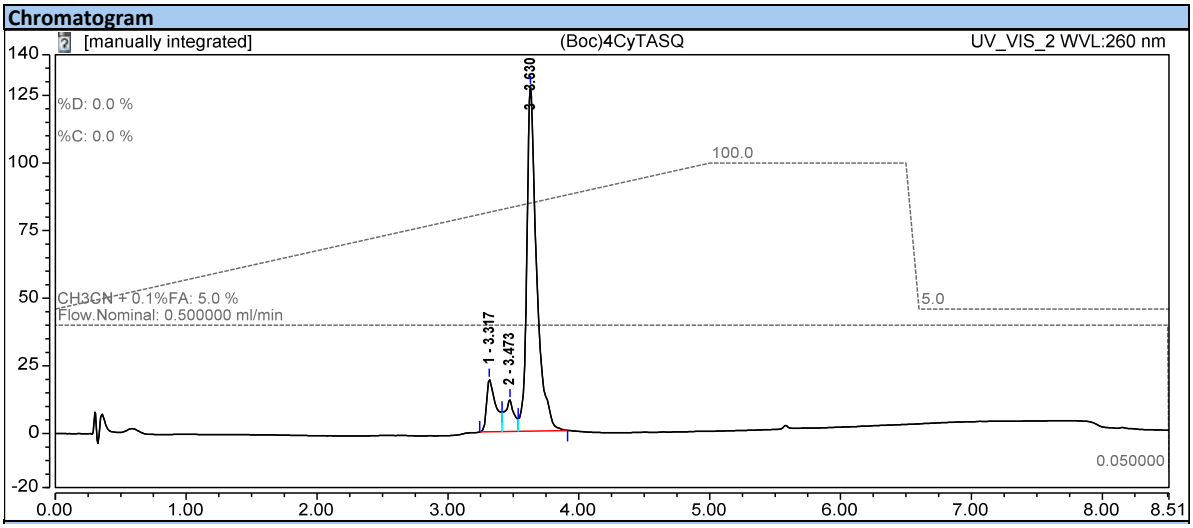
^{1}H NMR (500 MHz, DMSO-d_6) δ 10.94 (s, 4H), 7.89 (q, *J* = 21.0, 14.7 Hz, 8H), 6.64 (s, 8H), 5.00 (s, 4H), 4.86 (d, *J* = 19.5 Hz, 4H), 4.12 (s, 3H), 3.87 (s, 5H), 3.61 – 3.37 (m, 8H), 3.22 (d, *J* = 6.9 Hz, 8H), 3.19 – 3.08 (m, 8H), 3.07 – 2.96 (m, 20H), 2.78 (d, *J* = 19.4 Hz, 8H), 1.56 – 1.43 (m, 12H), 1.36 (d, *J* = 9.3 Hz, 36H), 1.31 – 1.11 (m, 12H).

ESI-HRMS: $[M+H]^+$ *m/z* = 2079.17259 (calc. for $C_{92}H_{148}N_{36}O_{20}$: 2079.17434).

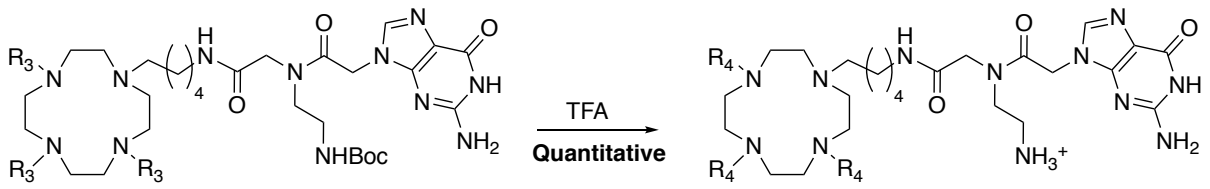
HPLC :


Chromatogram
[manually integrated]
(Boc)4CyTASQ
UV_VIS_3 WVL:201 nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.317
2 - 3.630
100.0
5.0
0.050000
300
200
100
0
-100
-150
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
(Boc)4CyTASQ
UV_VIS_1 WVL:214 nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.317
2 - 3.630
100.0
5.0
0.050000
300
200
100
0
-100
-150
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
(Boc)4CyTASQ
UV_VIS_2 WVL:260 nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.317
2 - 3.473
3 - 3.630
100.0
5.0
0.050000
140
125
100
75
50
25
0
-20
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51

CyTASQ :



R_3: $(CH_2)_5NH$-^{PNA}G-NHBoc

9

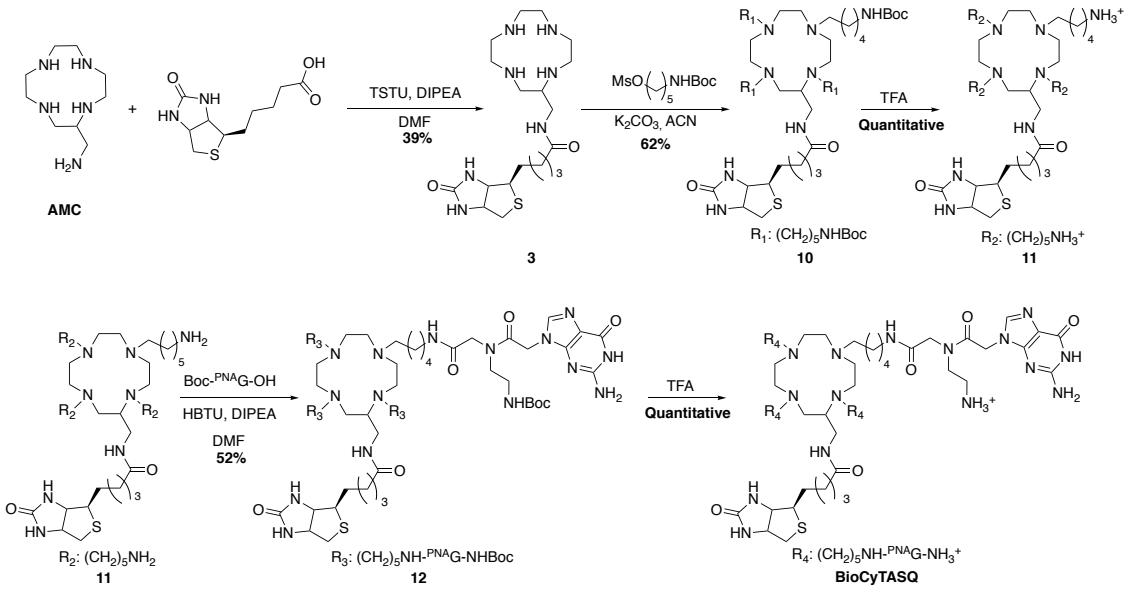
R_4: $(CH_2)_5NH$-^{PNA}G-NH_3^+

CyTASQ

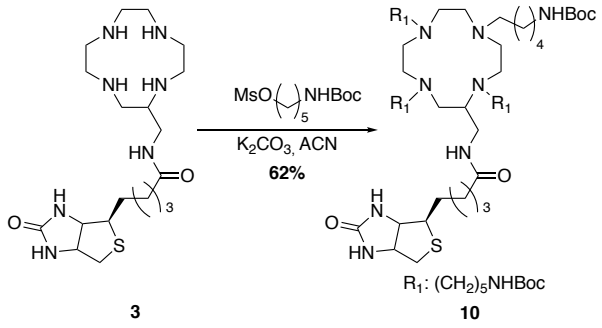
Compound **9** was dissolved in 1 mL of TFA and stirred for 1 hour. The complete deprotection was assessed by HPLC-MS. TFA was removed under reduced pressure, the residue lyophilized and used without further purification.

ESI-HRMS: $[M+H+Na]^{2+}$ m/z = 850.47707 (calcd. for $C_{72}H_{116}N_{36}O_{12}Na$: 850.47692)

BioCyTASQ :



Compound **10** :



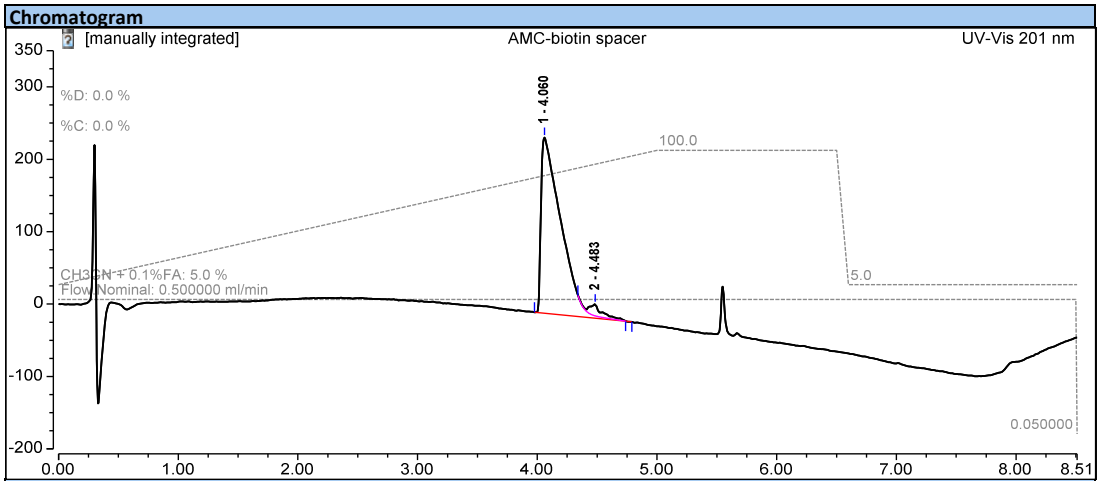
To a solution of **3** (246.8 mg, 0.32 mmol, 1 equiv.) in acetonitrile (4 mL) was added 5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate (720 mg, 2.6 mmol, 8 equiv.) and K_2CO_3 (472 mg, 5.9 mmol), the mixture was stirred at 70°C. After 48 h, an additional amount of 5-((tert-butoxycarbonyl)amino)pentyl methanesulfonate (392 mg, 1.40 mmol, 5 equiv.) was added to guarantee the complete substitution of the cyclen derivative and the mixture left to stir at reflux for an additional 9 hours. The crude mixture was then filtered, and concentrated under vacuum. The residue was purified by semi-preparative RP-HPLC with a H_2O/MeCN + 0.1% TFA mixture (gradient of 30 to 100 % over 40 minutes, r.t: 19 minutes) to obtain compound **10** (232.5 mg, 0.20 mmol, 62% chemical yield).

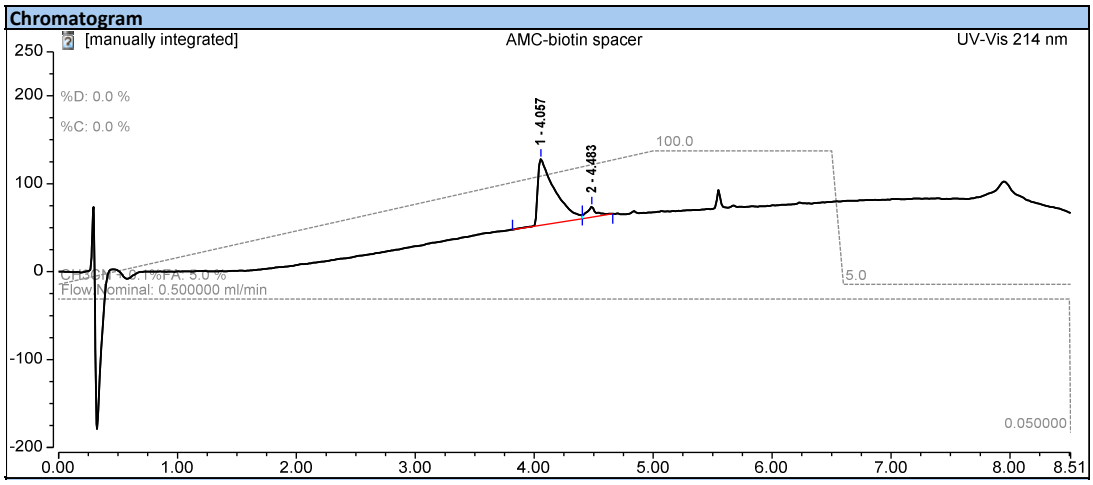
^{1}H NMR: (500 MHz, DMSO-d_6) δ 8.01 (s, 1H), 7.84 (s, 2H), 7.70 (s, 1H), 6.80 (d, *J* = 14.5 Hz, 3H), 6.40 (s, 2H), 4.32 (dd, *J* = 7.8, 4.9 Hz, 1H), 4.21 – 4.14 (m, 1H), 4.13 (dd, *J* = 7.8, 4.4 Hz, 1H), 3.40 – 3.34 (m, 5H), 3.16 (d, *J* = 7.5 Hz, 2H), 3.09 (dt, *J* = 10.0, 5.3 Hz, 1H), 3.05 (s, 13H), 2.92 (tt, *J* = 10.6, 5.6 Hz, 8H), 2.80 (tt, *J* = 13.1, 5.7 Hz, 3H), 2.71 (s, 4H), 2.13 – 2.06 (m, 2H), 1.66 (tt, *J* = 14.1, 7.0 Hz, 2H), 1.54 (s, 11H), 1.53 (t, *J* = 11.4 Hz, 1H), 1.37 (d, *J* = 2.2 Hz, 39H), 1.30 (s, 9H), 1.28 (d, *J* = 7.8 Hz, 2H).

^{13}C NMR: (126 MHz, DMSO-d_6) δ 162.88, 158.51, 158.24, 155.79, 117.50, 115.15, 77.54, 70.58, 70.34, 61.25, 59.40, 55.60, 36.69, 30.85, 29.05, 28.43, 28.34, 22.41.

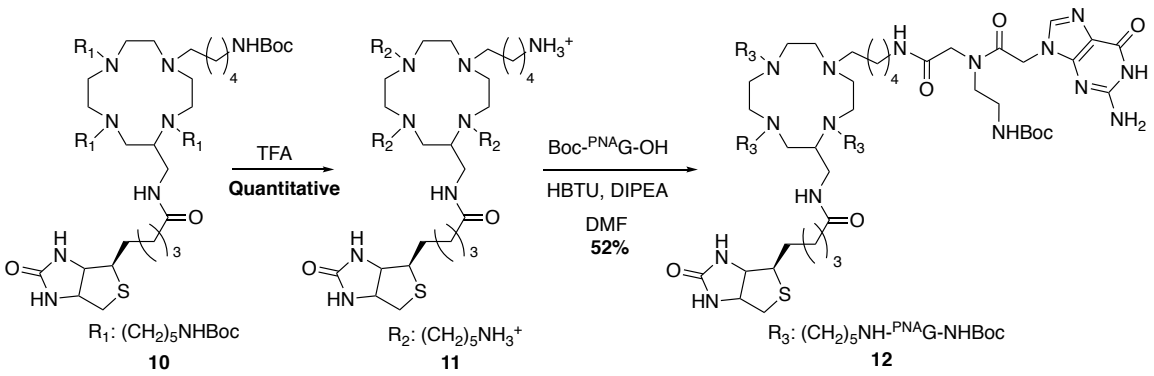
MALDI-TOF-MS : $[M+H]^+$ 1168.09 /z = (calcd. for $C_{59}0H_{113}N_{11}O_{10}S$: 1168.68).

HPLC:





Compound **12**:

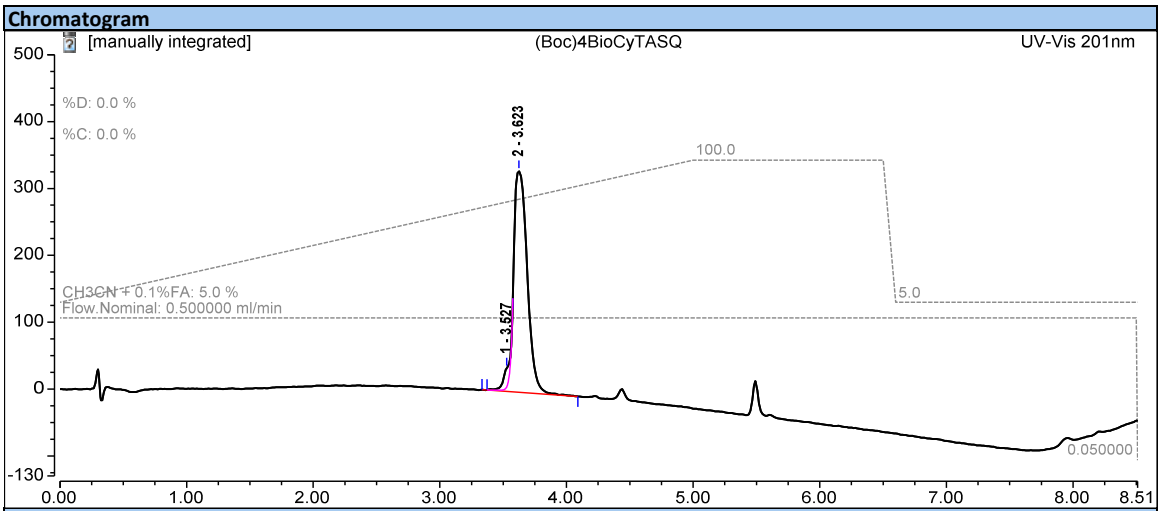


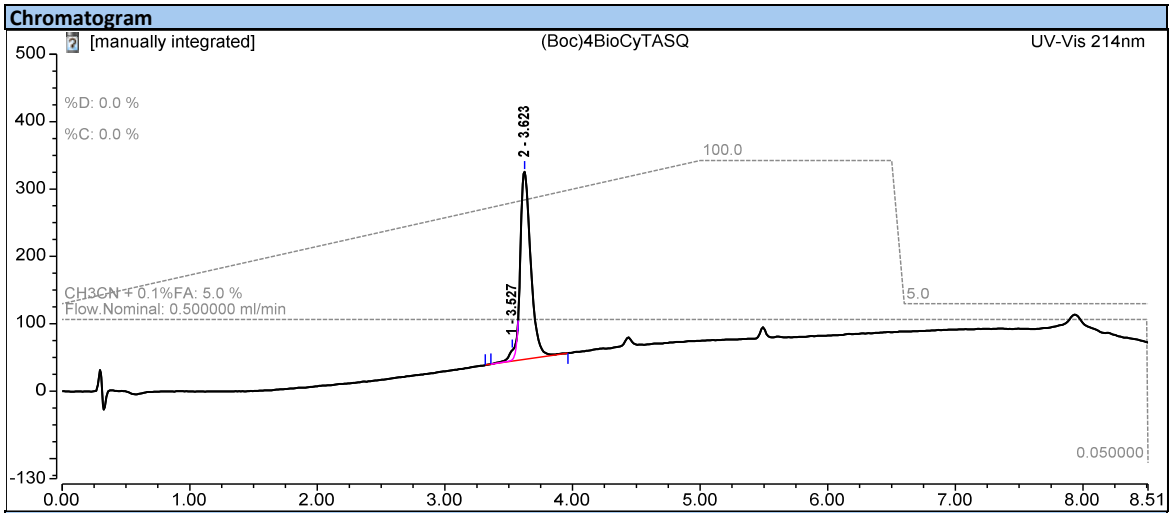
Compound **10** (17.9 mg, 0.015 mmol, 1 equiv.) was stirred with 250 µL of TFA during 1 hour to deprotect the amines. After evaporation of the TFA, complete deprotection of compound **10** was observed *via* HPLC-MS. TFA was removed under reduced pressure and after addition of DMF (1 mL), to the solution was added DIPEA (21 µL, 0.12 mmol). In another round bottom flask, Boc-^{PNA}G (27.5 mg, 0.067 mmol, 4.4 equiv.) was dissolved in DMF (1 mL), and HBTU (25.4 mg, 0.067 mmol, 4.4 equiv.) and DIPEA were added (11 µL, 0.06 mmol). The reaction mixture was added dropwise to the basic solution of **11**, the resulting mixture was stirred at room temperature for three hours, monitoring its completion by HPLC-MS. The solution was then concentrated under vacuum, added a mixture of MeCN and water (50/50, 1 mL), and purified by RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 15 to 65 % over 50 minutes, r.t: 21 minutes). After evaporation of the solvents, the compound **12** was obtained (18.3 mg, 7.84 µmol, 52% chemical yield)

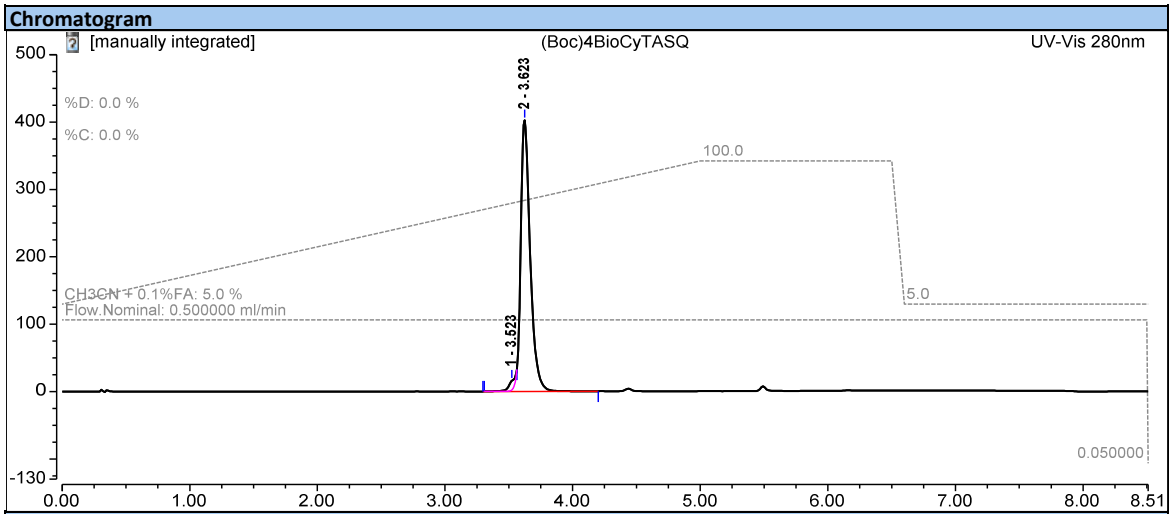
^{1}H NMR: (500 MHz, DMSO-d_6) δ 10.76 (s, 2H), 10.72 (s, 2H), 8.27 (s, 2H), 8.04 (s, 1H), 7.88 (s, 3H), 7.64 (s, 3H), 7.03 (s, 2H), 6.77 (s, 1H), 6.66 (s, 1H), 6.50 (s, 6H), 6.41 (s, 2H), 4.97 (s, 4H), 4.83 (s, 2H), 4.31 (d, *J* = 6.5 Hz, 1H), 4.11 (s, 2H), 3.87 (s, 3H), 3.47 (s, 4H), 3.28 (s, 2H), 3.24 – 3.19 (m, 4H), 3.15 (dd, *J* = 7.3, 4.2 Hz, 1H), 3.06 (s, 18H), 3.02 (s, 2H), 2.81 (d, *J* = 8.3 Hz, 1H), 2.76 (s, 1H), 2.59 (s, 1H), 2.57 (s, 1H), 2.31 (s, 1H), 1.47 (s, 22H), 1.37 (d, *J* = 8.4 Hz, 29H), 1.25 (q, *J* = 7.8, 7.3 Hz, 8H), 1.14 (s, 1H).

ESI-HRMS: $[M+H+Na]^{2+}$ m/z = 1178.63633 (calc. for $[C_{103}H_{166}N_{39}O_{22}SNa]^{2+}$: = 1178.63519).

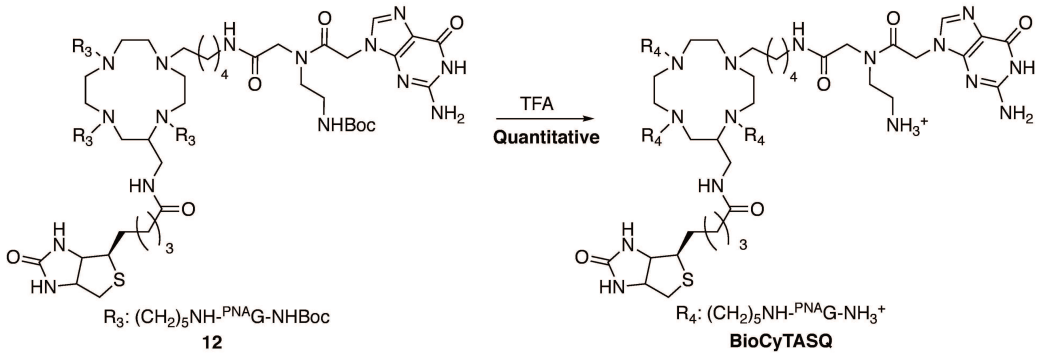
HPLC track:


Chromatogram
[manually integrated]
(Boc)4BioCyTASQ
UV-Vis 201nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
2 - 3.623
1 - 3.527
100.0
5.0
0.050000


Chromatogram
[manually integrated]
(Boc)4BioCyTASQ
UV-Vis 214nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
2 - 3.623
1 - 3.527
100.0
5.0
0.050000


Chromatogram
[manually integrated]
(Boc)4BioCyTASQ
UV-Vis 280nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
2 - 3.623
1 - 3.523
100.0
5.0
0.050000

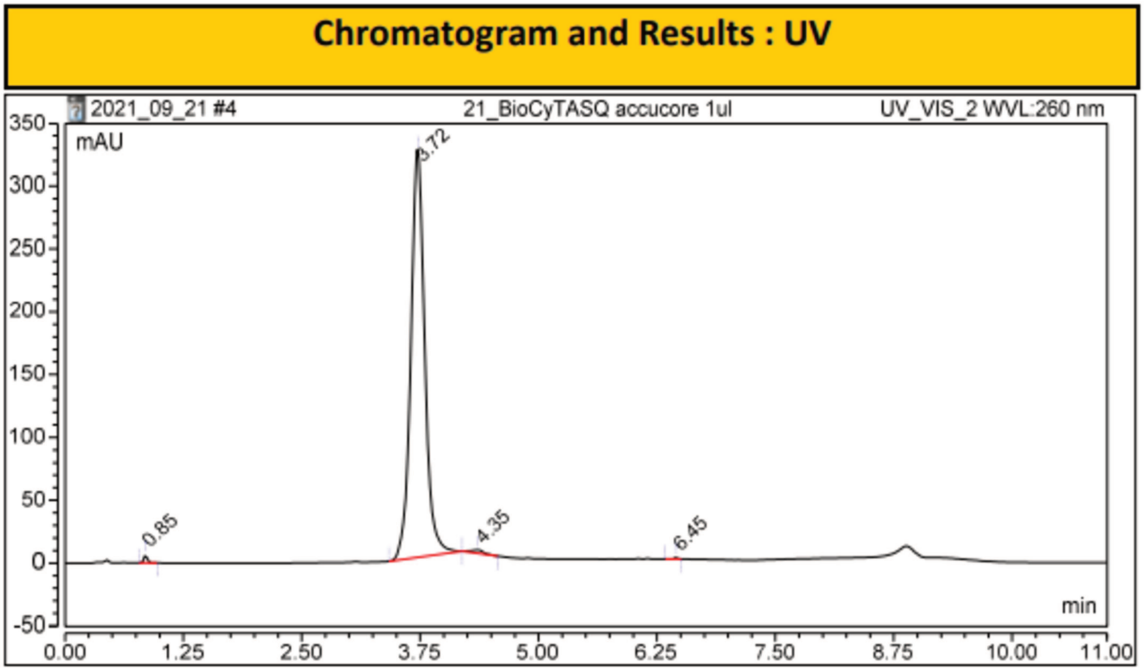
BioCyTASQ:



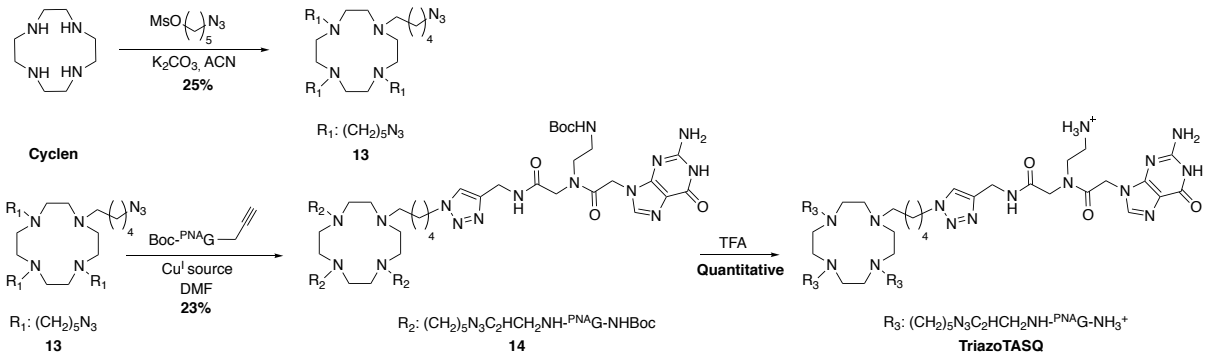
Compound **12** (1 mg, 0.43 µmol) was dissolved in TFA (100 µL) and the complete deprotection was assessed *via* HPLC. TFA was removed under reduced pressure and the target compound was obtained quantitatively.

ESI-HRMS: $[M+H]^{2+}$ m/z = 967.04040 (calc. for $[C_{83}H_{135}N_{39}O_{14}S]^{2+}$: =967.03802)

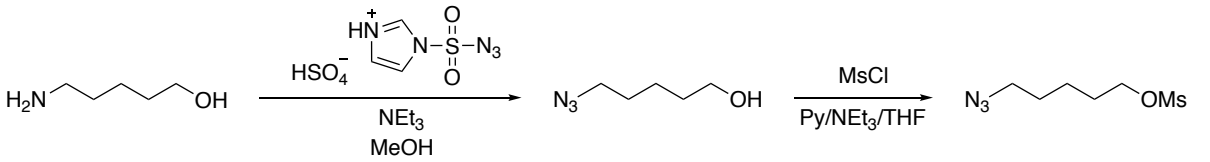
HPLC track:



TriazoTASQ:



5-azidopentyl methanesulfonate:



To a solution of 5-aminopentanol (1.0 g, 9.6 mmol, 1 equiv.) and 1*H*-imidazole-1-sulfonyl azide·H_2SO_4 (2.6 g, 9.6 mmol, 1 equiv.) in MeOH (100 mL) was added triethylamine (6.6 mL, 48.2 mmol, 5 equiv.). The reaction was stirred for 1 h, after which 2 additional equiv. of triethylamine were added in order to reach pH > 9-10. The reaction was stirred overnight and completion was assessed by TLC (CH_2Cl_2/MeOH 9:1). The solvent was evaporated and 50 mL of EtOAc and 100 mL of an aq. HCl 1M solution were then added to the mixture. The aqueous phase was extracted with EtOAc (3 x 30 mL). The organic phases were pooled together, dried over $MgSO_4$, filtered and dried under vacuum. The crude product was dissolved in a mixture of THF (100 mL) and triethylamine (6.6 mL, 48.2 mmol, 5 equiv.), and MsCl (1.5 mL, 19.3 mmol, 2 equiv.) was added slowly at 0 °C. After 2 hours, TLC analysis (9.5/0.5, CH_2Cl_2/MeOH) indicated a partial conversion of the starting material. In order to reach completion, an additional equiv. of triethylamine (3 mL) and pyridine (10 mL) were added to the mixture, stirred at RT, and the reaction followed by TLC. The solvent was evaporated, to the mixture was then added 50 mL of EtOAc and 100 mL of an aqueous solution of HCl 1M. The aqueous phase was extracted 3 times with EtOAc (30 mL each); the organic phases were pooled together, dried over $MgSO_4$, filtered and dried under vacuum.

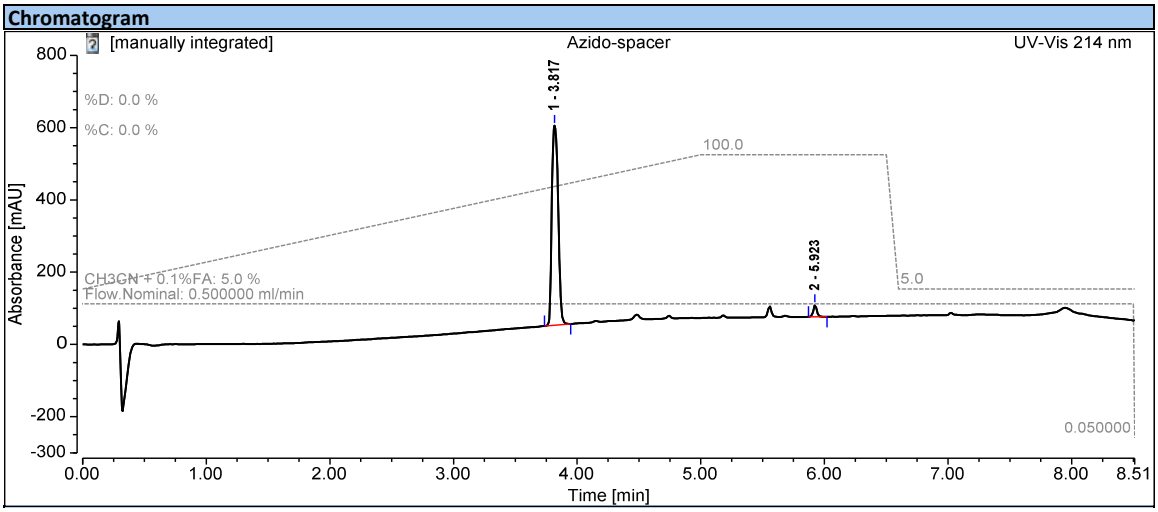
The residue was purified by silica gel column (CH_2Cl_2/MeOH 0% to 4%) to afford 5-azidopentyl methanesulfonate (836.1 mg, 4.1 mmol, 42 % chemical yield over two steps).

^{1}H NMR (500 MHz, $CDCl_3$) δ 4.19 (t, *J* = 6.5 Hz, 2H), 3.26 (t, *J* = 6.5 Hz, 2H), 2.97 (s, 3H), 1.75 (dt, *J* = 14.0, 6.5 Hz, 2H), 1.60 (dt, *J* = 14.0, 6.5 Hz, 2H), 1.47 (dt, *J* = 8.0, 6.5 Hz, 2H).

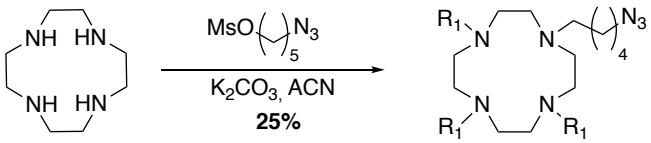
^{13}C NMR (126 MHz, $CDCl_3$) δ 69.7, 51.1, 37.3, 28.6, 28.2, 22.7.

IR : 2092.10 cm^{-1} (azide), 1168.46 cm^{-1} (sulfonyl)

HPLC :



Compound **13** :



Cyclen

R_1: $(CH_2)_5N_3$

13

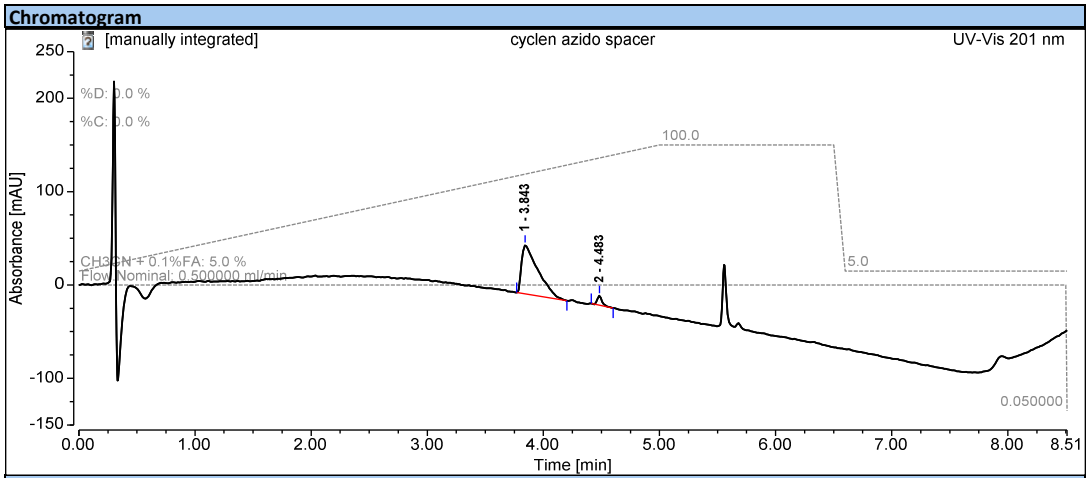
To a solution of cyclen (68.9 mg, 0.4 mmol, 1 equiv.) in acetonitrile (6 mL) were added 5-azidopentyl methanesulfonate (497.3 mg, 2.4 mmol, 6 equiv.) and K_2CO_3 (456.1 mg, 3.2 mmol, 8 equiv.). The reaction was heated at 50 °C overnight and adequate conversion was assessed by HPLC-MS. The mixture was filtered and the solvent evaporated under vacuum. The crude material was purified by RP-HPLC (H_2O/MeCN + 0.1% TFA, gradient from 5% to 60% over 45 minutes). The solvent was removed by freeze-drying to provide compound **13** (61.7 mg, 0.1 mmol, 25% chemical yield).

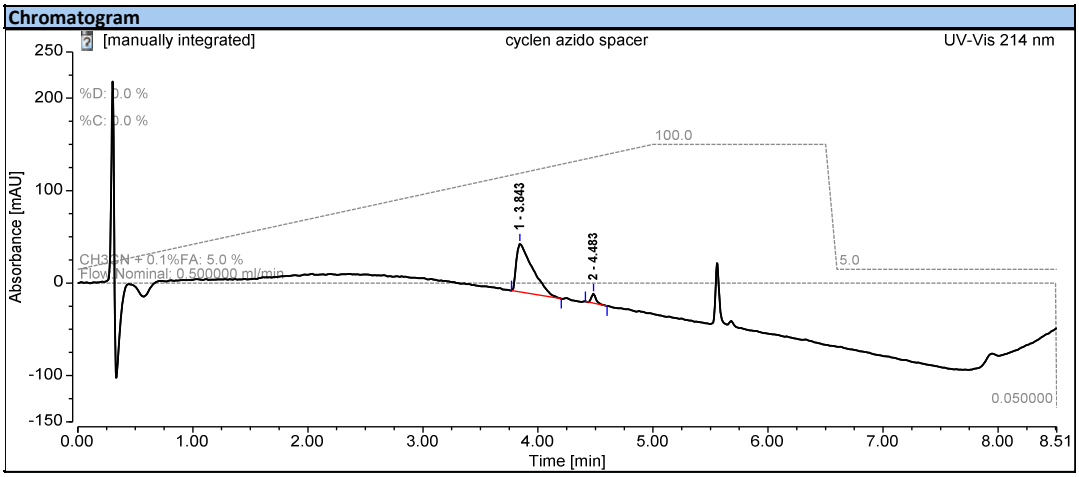
^{1}H NMR (500 MHz, $CDCl_3$) δ 3.24 (t, *J* = 6.5 Hz, 8H), 3.11 – 2.86 (m, 24H), 1.75 – 1.53 (m, 16H), 1.48 – 1.30 (m, 8H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 53.2, 51.1, 48.5, 28.4, 24.2, 21.6.

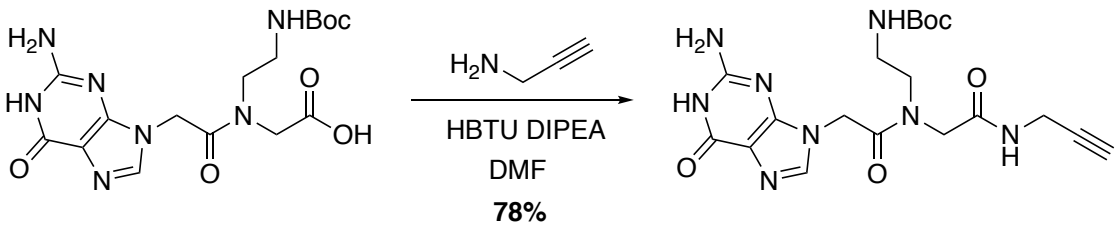
ESI-HRMS $[M+H]^+$ m/z = 617.49366 (calcd. for $C_{28}H_{57}N_{16}$: 617.49466).

HPLC :





Alkynilated Guanine :



Boc-^{PNA}G (199.3 mg, 049 mmol, 1 equiv.), propargylamine (67 µL, 1.0 mmol, 2 equiv.) and HBTU (379.2 mg, 1.0 mmol, 2 equiv.) were dissolved in anhydrous DMF (5 mL). DIPEA was added (250 µL, 1.5 mmol, 3 equiv.) and the solution stirred for 1 hour at RT (the complete conversion was assessed by HPLC). The solution was then concentrated under vacuum, and the target compound purified by RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient from 2% to 35% over 45 minutes), **Alkynilated Guanine** was obtained as a white solid after lyophilization (169.5 mg, 0.38 mmol, 78 % chemical yield).

Two isomers, probably corresponding to *cis/trans* isomers were identified by NMR

^{1}H NMR (500 MHz, Methanol-d_4):

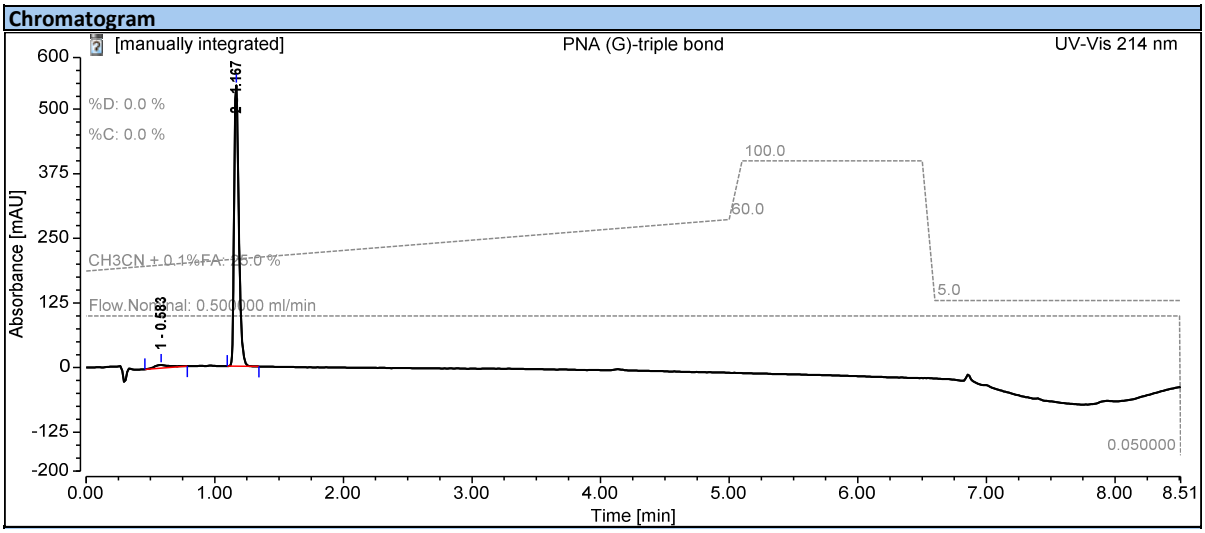
Major isomer: δ 8.43 (s, 1H), 5.24 (s, 2H), 4.09 (s, 2H), 3.98 (d, *J* = 2.5 Hz, 1H), 3.60 (t, *J* = 6.5 Hz, 1H), 3.38 (t, *J* = 6.5 Hz, 2H), 2.58 (t, *J* = 2.5 Hz, 1H), 1.42 (s, 9H).

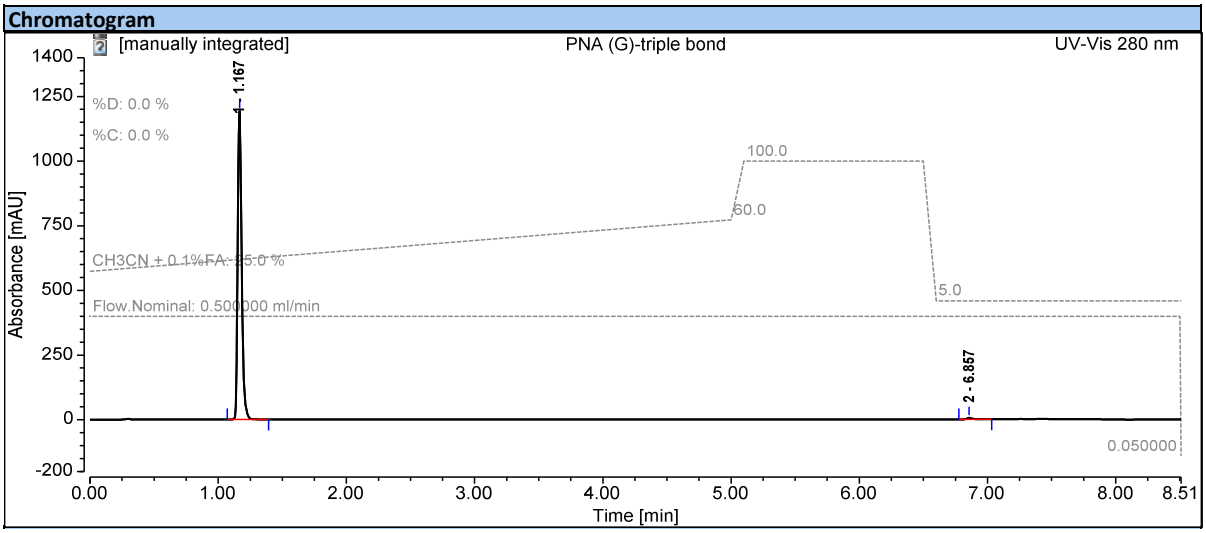
Minor isomer δ 8.47 (s, 1H), 5.10 (s, 2H), 4.27 (s, 2H), 4.07 (s, 2H), 3.46 (t, *J* = 6.4 Hz, 1H), 3.19 (t, *J* = 6.5 Hz, 2H), 2.67 (t, *J* = 2.5 Hz, 1H), 1.48 (s, 9H).

^{13}C NMR (126 MHz, Methanol-d_4) δ 167.6, 166.9, 79.1, 78.9, 71.2, 71.0, 49.6, 48.3, 48.1, 47.8, 47.6, 47.4, 47.2, 47.0, 44.6, 38.4, 28.2, 27.5, 27.4.

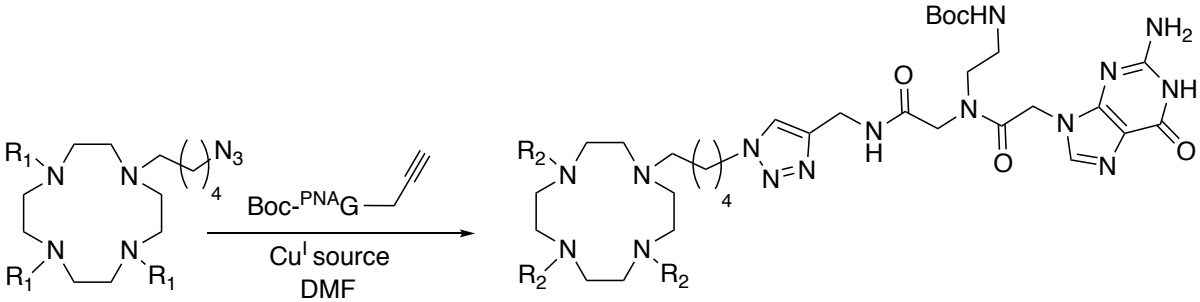
ESI-HRMS $[M+Na]^+$ m/z = 469.19193 (calcd. for $C_{19}H_{26}N_8NaO_5$: 469.19184).

HPLC : (gradient from 25% to 60% of MeCN in 5 minutes)





Compound **14** :



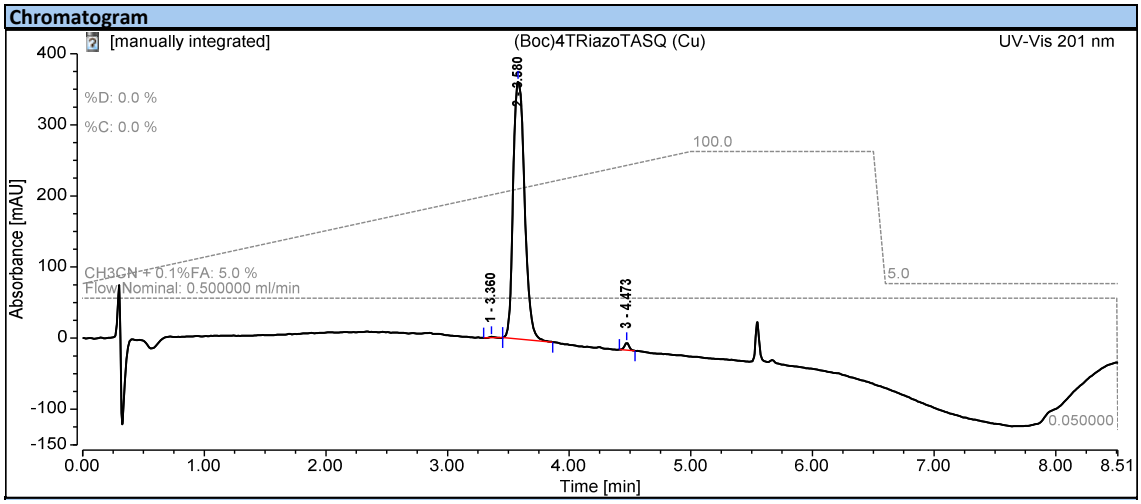
In anhydrous DMF (0.7 mL) were mixed the alkynylated guanine (13.8 mg, 31 µmol, 4 equiv.), compound **13** (4.7 mg, 7.7 µmol, 1 equiv.) and the copper ligand *tris*-(benzyltriazolylmethyl)amine (TBTA) (2.1 mg, 3.9 µmol, 0.5 equiv.). The catalyst $(MeCN)_4Cu{\cdot}PF_6$ (4.3 mg, 11.6 µmol, 1.5 equiv.) was added to the mixture which was stirred at RT for 3 hours (conversion assessed by HPLC-MS). The solvent was evaporated under vacuum and the crude product purified by RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 60% over 45 minutes). After lyophilization of the solvents, **$(Boc)_4$TriazoTASQ·Cu(II)** (**14·Cu(II)**) was obtained (4.34 mg, 1.76 µmol, 23 % chemical yield), presence of coper was assessed by ICP-MS and HPLC-MS.

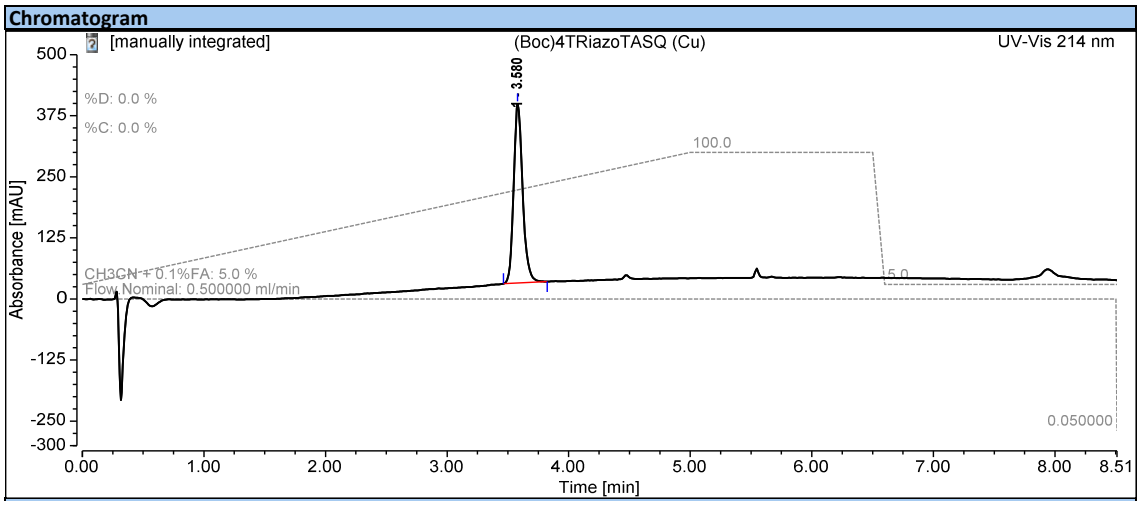
Compound 14·Cu(II) :

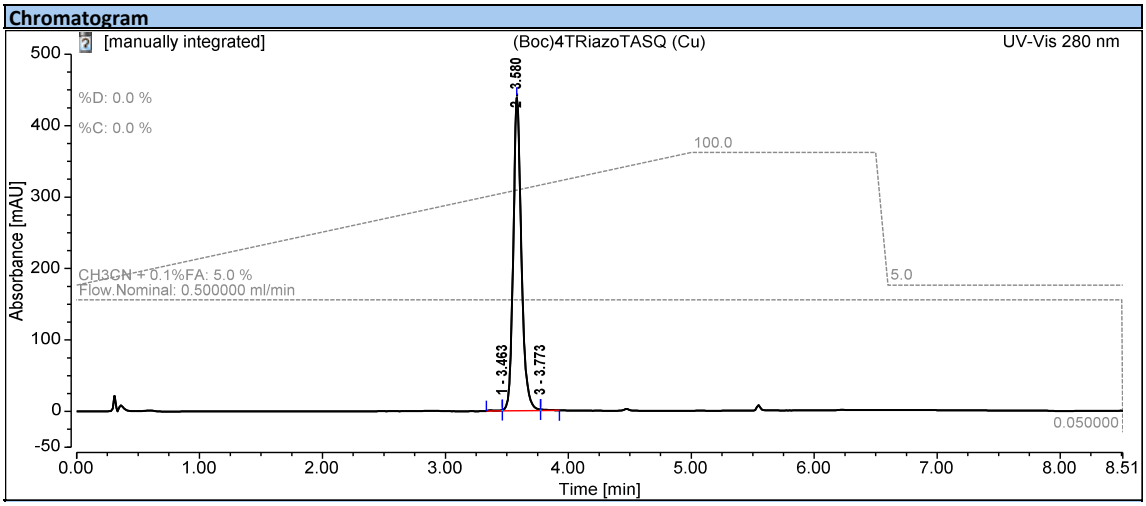
ICP-MS : 1.91 % of Copper

ESI-HRMS $[M+Cu+H]^{2+}$ m/z = 1233.11523 (calcd. for $C_{104}CuH_{160}N_{48}O_{20}$: 1232.11318)

HPLC :


Chromatogram
[manually integrated]
(Boc)4TRiazoTASQ (Cu)
UV-Vis 201 nm
%D: 0.0 %
%C: 0.0 %
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
2 - 3.580
1 - 3.360
3 - 4.473
Absorbance [mAU]
Time [min]


Chromatogram
[manually integrated]
(Boc)4TRiazoTASQ (Cu)
UV-Vis 214 nm
%D: 0.0 %
%C: 0.0 %
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
1 - 3.580
Absorbance [mAU]
Time [min]


Chromatogram
[manually integrated]
(Boc)4TRiazoTASQ (Cu)
UV-Vis 280 nm
%D: 0.0 %
%C: 0.0 %
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
2 - 3.580
1 - 3.463
3 - 3.773
Absorbance [mAU]
Time [min]

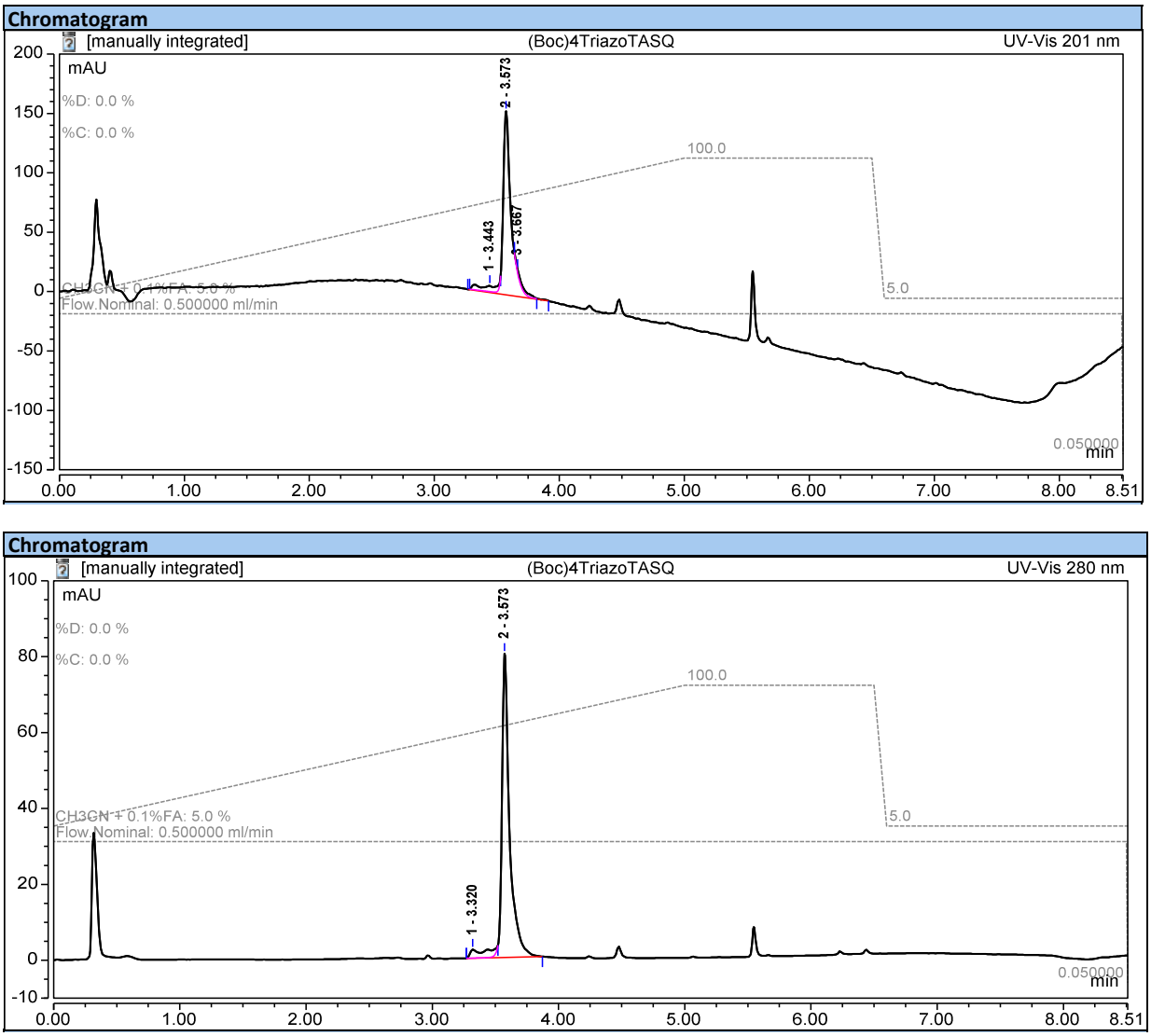
14·Cu(II) (4.34 mg, 1.76 µmol, 1 equiv.) was dissolved in MeCN (200 µL) and Na_2S (0.5 mg, 7 µmol, 4 equiv.) in H_2O (200 µL) was added. After 2 minutes, centrifugation was performed (4000 RPM, 15 min), the supernatant was removed, the pellet washed with MeCN (200 µL) and H_2O (200 µL) and the solvents were lyophilized to afford **14** (4.24 mg, 1.76 µmol). Demetallation was monitored by elemental analysis and HPLC-MS.

Compound 14:

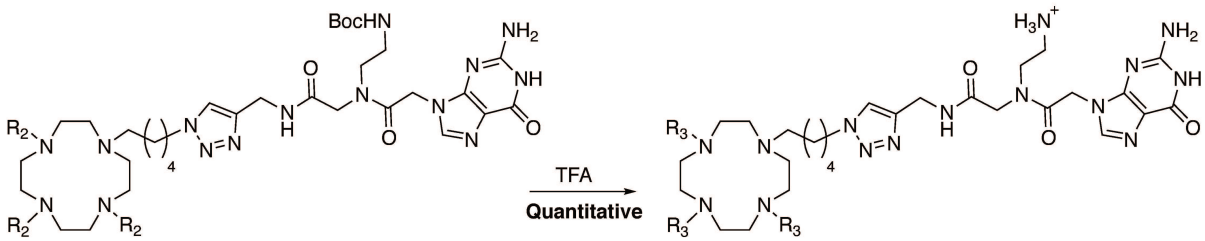
ICP-MS : 0.02 % of Copper

Maldi-TOF MS : $[M+H]^+$ m/z = 2403.38 (calcd. for $C_{104}H_{161}N_{48}O_{20}$: 2403.74)

HPLC :



TriazoTASQ :



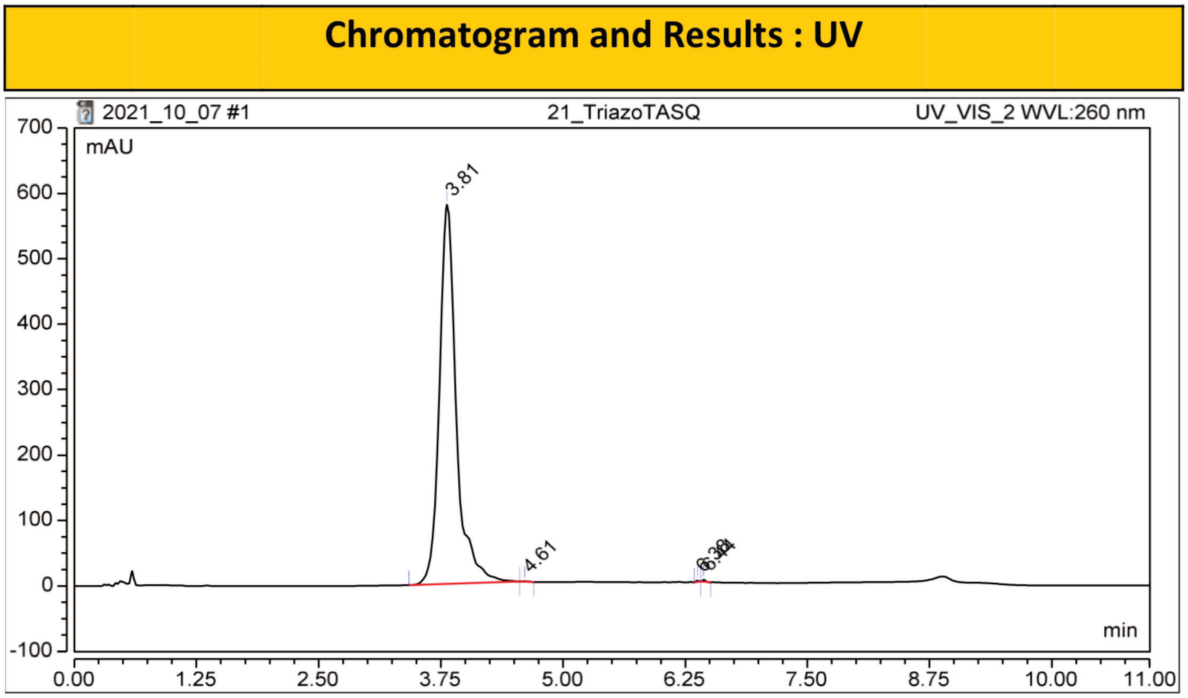
R_2: $(CH_2)_5N_3C_2HCH_2NH$-^{PNA}G-NHBoc
14

R_3: $(CH_2)_5N_3C_2HCH_2NH$-^{PNA}G-NH_3^+
TriazoTASQ

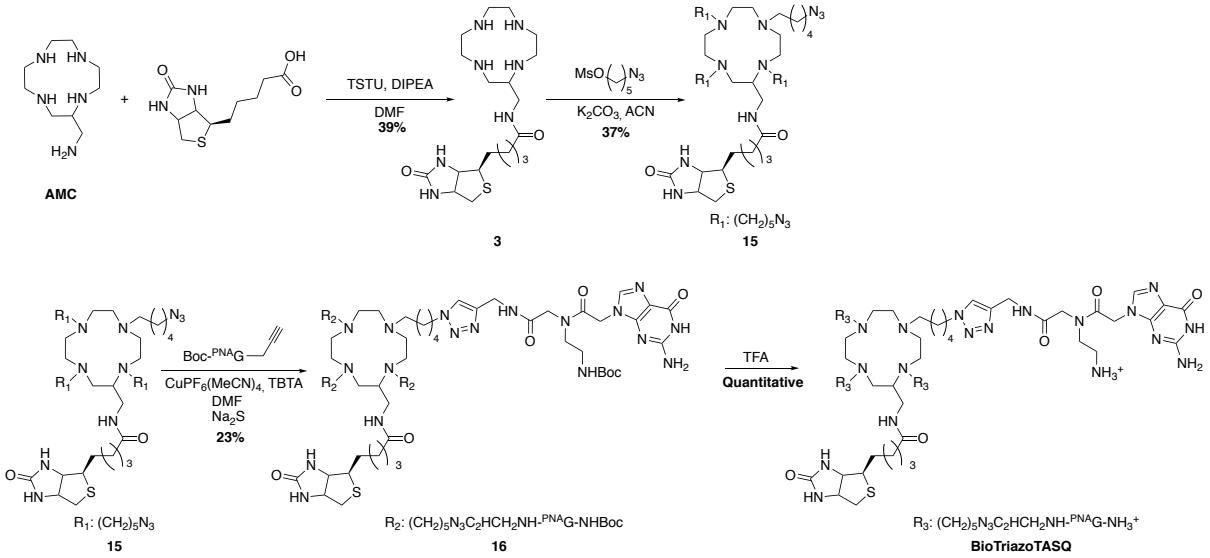
14 (1.0 mg, 0.4 µmol) was stirred in TFA (50 µL) at 25 °C for 1 hour. Complete conversion was assessed by HPLC and the mixture was diluted in water (100 µL). **TriazoTASQ** (0.98 mg, 0.4 µmol) was obtained after lyophilization.

ESI-HRMS $[M+2H]^{2+}$ m/z = 1001.54657 (calcd. for $C_{84}H_{130}N_{48}O_{12}$: 1001.55134).

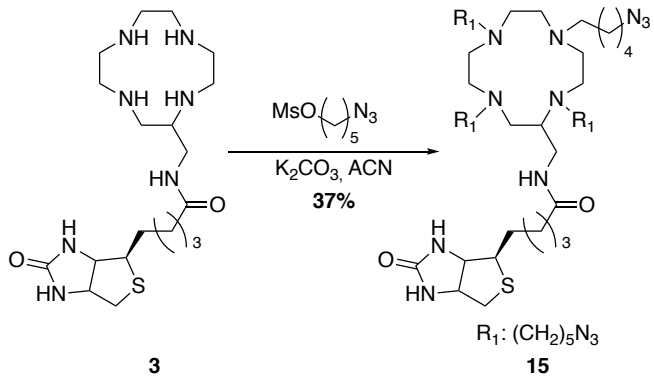
HPLC :



BioTriazoTASQ :



Compound **15** :



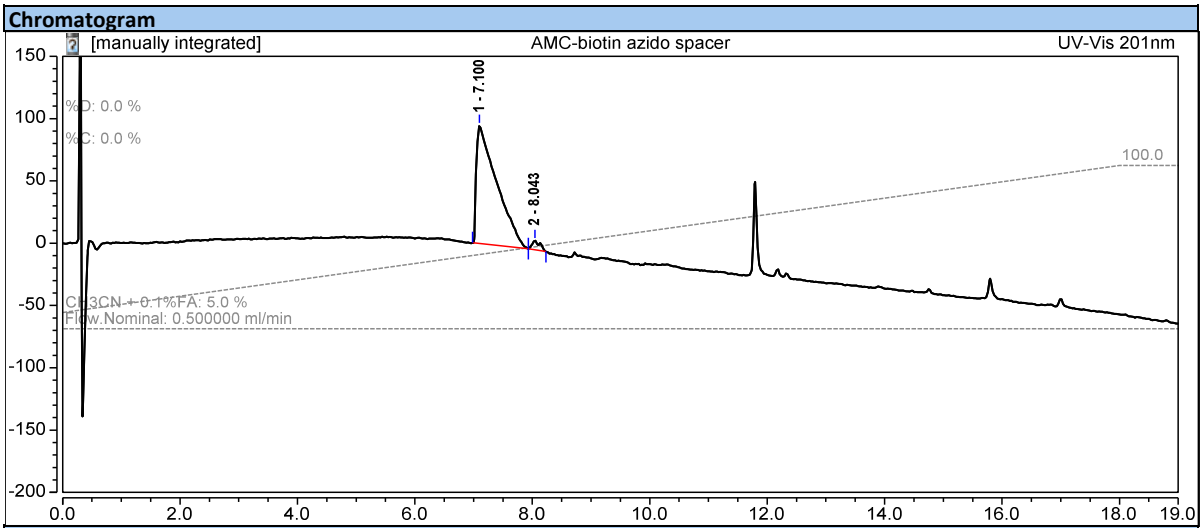
Biotin-AMC (**3**) (99.9 mg, 0.13 mmol, 1 equiv.) was mixed with 5-azidopentyl methanesulfonate (140.8 mg, 0.7 mmol, 6 equiv.) in MeCN (3 mL). To which K_2CO_3 (127.3 mg, 0.9 mmol, 8 equiv.) was added and the mixture was stirred at 50 °C for 72 hours (conversion assessed by HPLC-MS). After which the mixture was filtered, the solvent evaporated under vacuum and the crude purified by RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 2 to 50 % over 45 minutes). After evaporation of the solvents, compound **15** was obtained as a white solid (41.9 mg, 0.048 mmol, 37 % chemical yield).

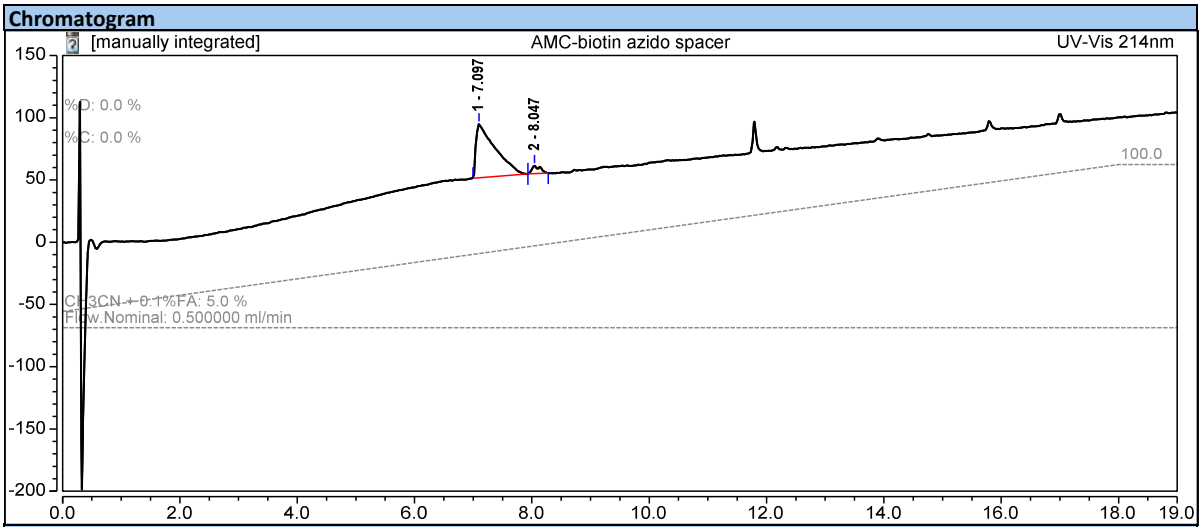
^{1}H NMR (500 MHz, D_2O) δ 4.61 (dd, *J* = 8.0, 4.5 Hz, 1H), 4.41 (dd, J = 8.0, 4.5 Hz, 1H), 3.36 – 2.60 (m, 37H), 2.30 (t, *J* = 10.5, 10.5 Hz, 2H), 1.75 – 1.54 (m, 19H), 1.50 – 1.40 (m, 6H), 1.38 – 1.27 (m, 4H). (Signals of intracyclic amines and amide are missing due to proton solvent exchange)

^{13}C NMR (126 MHz, D_2O) δ 161.1, 160.8, 160.5, 160.2, 77.2, 76.9, 76.6, 60.8, 51, 50.8, 50.8, 50.8, 40.2, 34.8, 28.4, 28.3, 28.3, 28.1, 28.1, 25, 24.2, 23.8, 23.7, 23.7, 21.8.

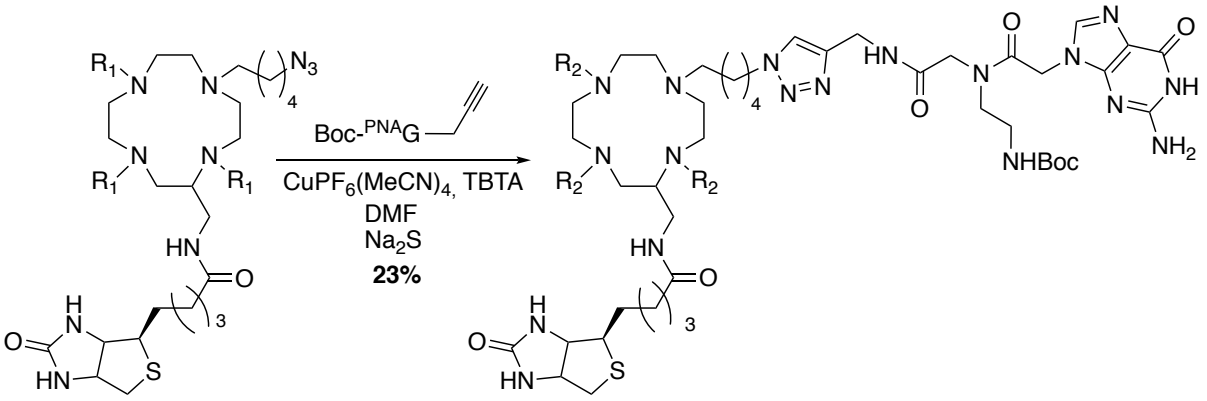
MALDI-TOF MS $[M+H]^+$ m/z = 872.51 (calcd. for $C_{39}H_{74}N_{19}O_2S$: 873.20).

HPLC: (from 5% to 100% MeCN in 18 min)





Compound **16** :

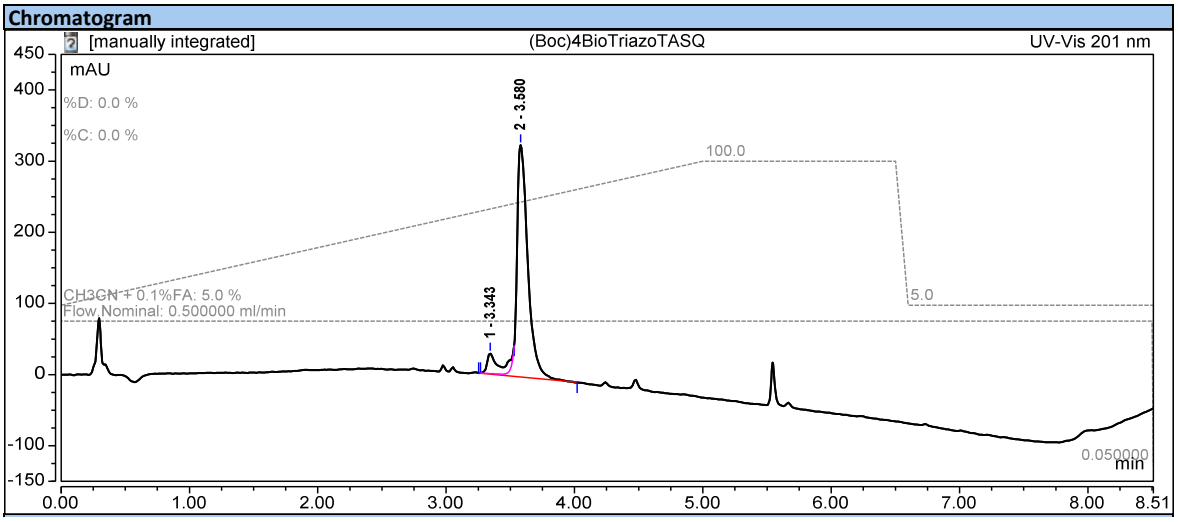


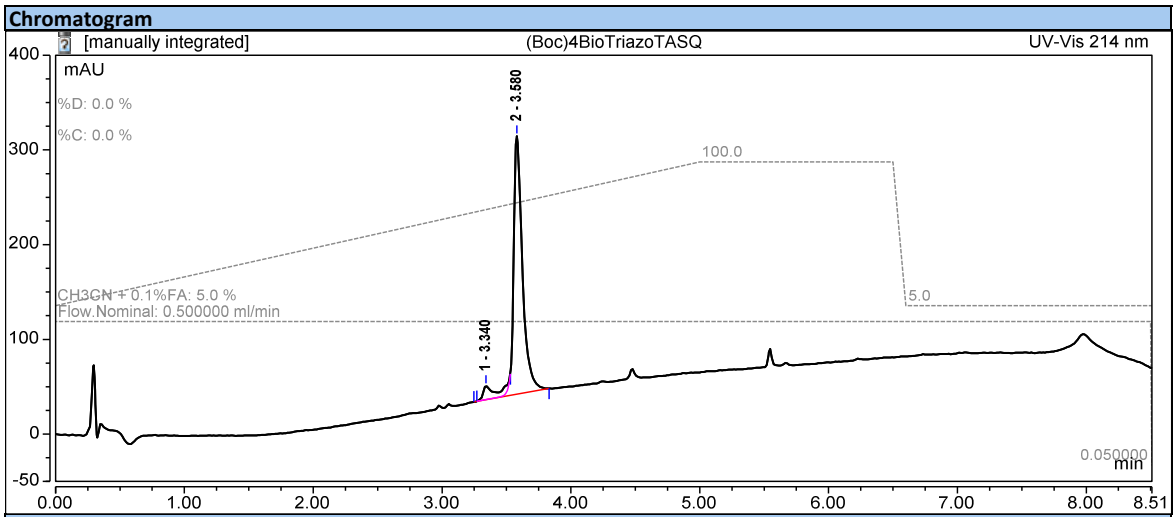
In anhydrous DMF (1.8 mL) were mixed the previously described alkynylated guanine (27.6 mg, 61.9 µmol, 4.2 equiv.), **15** (12.7 mg, 0.015 mmol, 1 equiv.) and TBTA (3.9 mg, 7.5 µmol, 0.5 equiv.). $(MeCN)_4Cu{\cdot}PF_6$ (8.2 mg, 22.5 µmol, 1.5 equiv.) was added to the mixture and stirred at RT till reaction completion (6 h, assessed by HPLC-MS). The solvent was evaporated under vacuum and the crude mixture purified by RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 60% over 45 minutes). After lyophilization of the solvents, **$(Boc)_4$BioTriazoTASQ·Cu(II)** was obtained (9.17 mg, 3.4 µmol, 23 % chemical yield) and subsequently demetalated after dissolution in MeCN (200 µL) and addition of Na_2S (1.1 mg, 13.6 µmol, 4 equiv.) in H_2O (200 µL). After 2 min, centrifugation was performed (4000 RPM, 15 min), the supernatant removed and **$(Boc)_4$BioTriazoTASQ 16** (8.96 mg, 3.4 µmol) isolated after lyophilization.

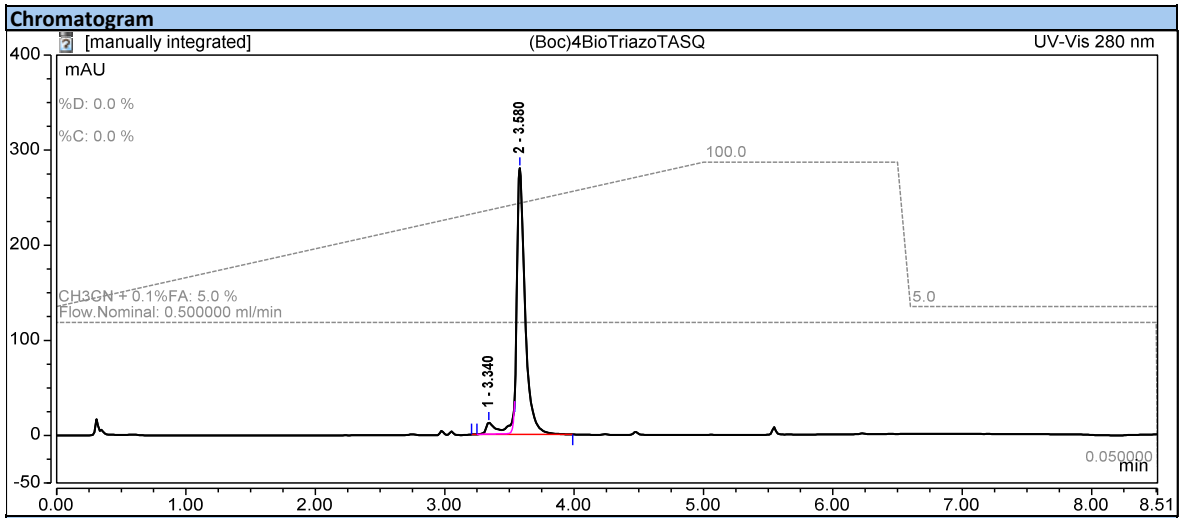
ESI-HRMS $[M+H+Na]^{2+}$ m/z = 1340.20428 $C_{115}H_{178}N_{51}NaO_{22}S$: 1340.19925).

ICP-MS : 0.19 % of Copper

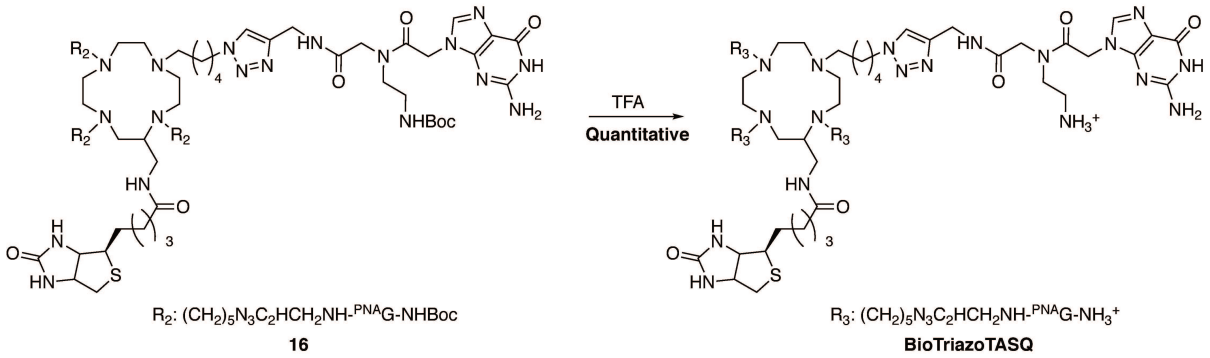
HPLC :


Chromatogram
[manually integrated]
(Boc)4BioTriazoTASQ
UV-Vis 201 nm
mAU
%D: 0.0 %
%C: 0.0 %
2 - 3.680
100.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.343
5.0
0.050000
min
450
400
300
200
100
0
-100
-150
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
(Boc)4BioTriazoTASQ
UV-Vis 214 nm
mAU
%D: 0.0 %
%C: 0.0 %
2 - 3.680
100.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.340
5.0
0.050000
min
400
300
200
100
0
-50
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
(Boc)4BioTriazoTASQ
UV-Vis 280 nm
mAU
%D: 0.0 %
%C: 0.0 %
2 - 3.680
100.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.340
5.0
0.050000
min
400
300
200
100
0
-50
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51

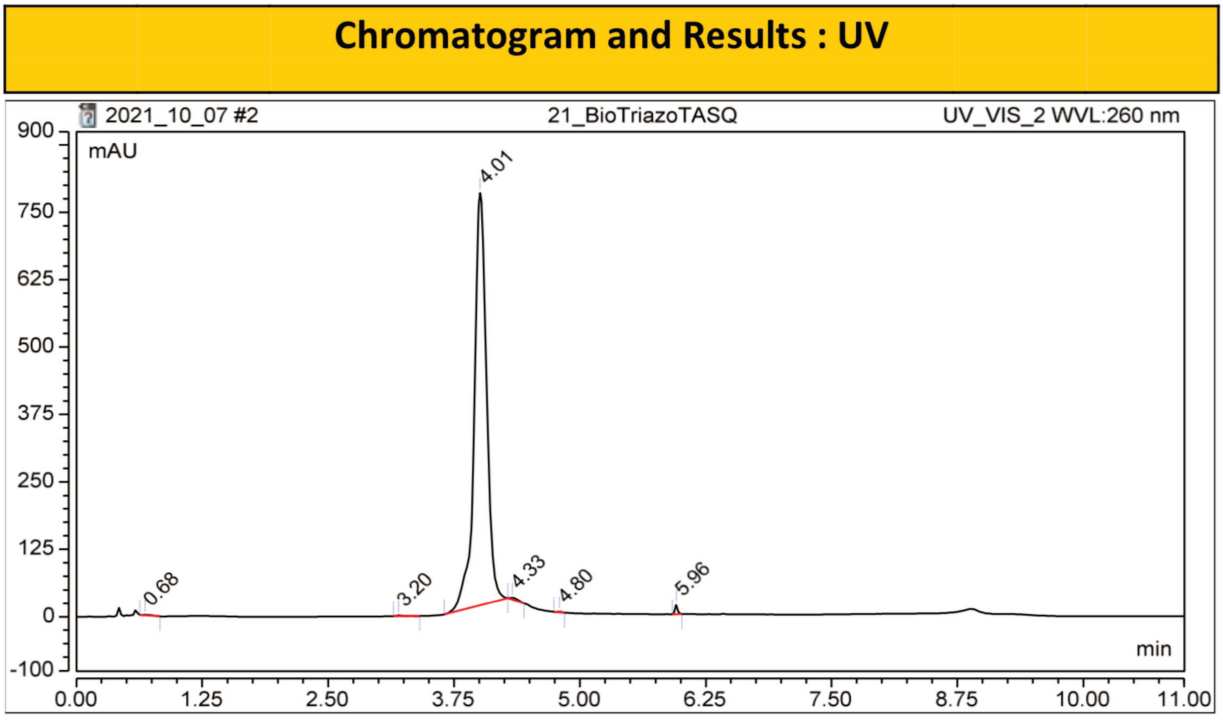
BioTriazoTASQ :



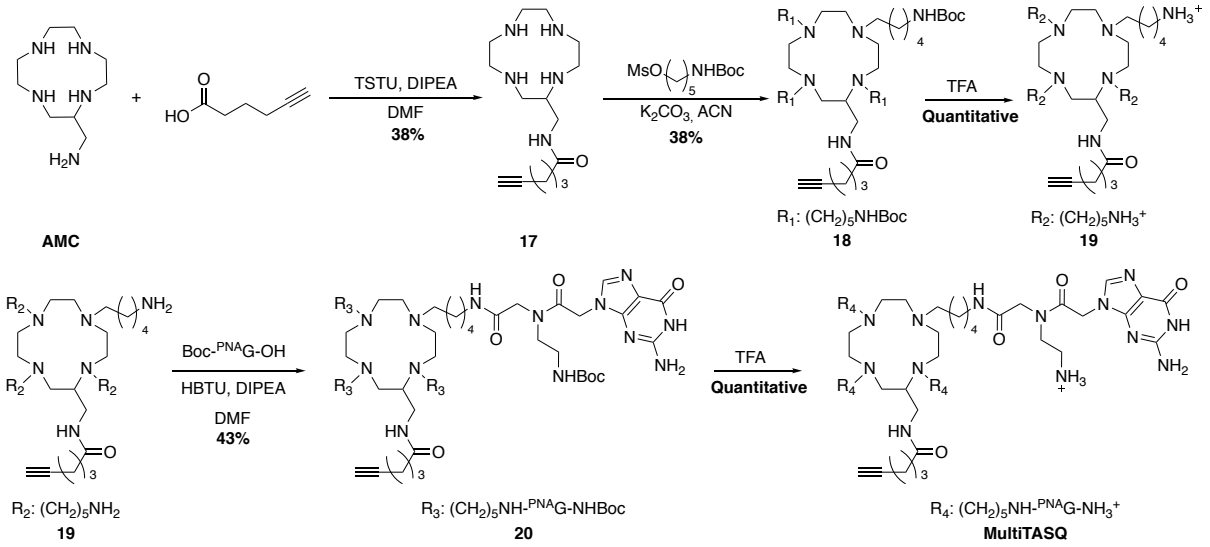
(Boc)$_4$BioTriazoTASQ (**16**) (2.8 mg, 1.1 µmol) was mixed in TFA (50 µL) at RT for 1 hour. After the completion of the reaction (assessed by HPLC) **BioTriazoTASQ** (2.9 mg, 1.1 µmol) was isolated by lyophilization.

ESI-HRMS $[M+2H]^{2+}$ m/z = 1129.10726 (calcd. for $C_{95}H_{147}N_{51}O_{14}S$: 1129.10342).

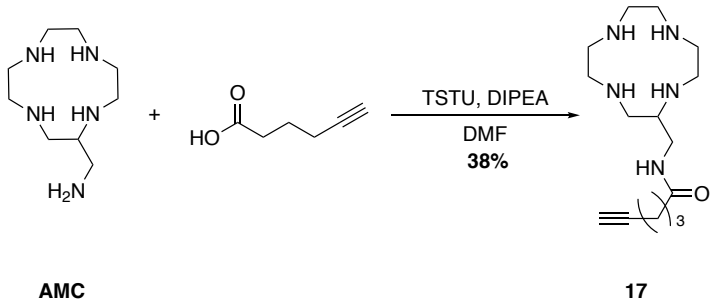
HPLC :



MultiTASQ:



Compound **17**:



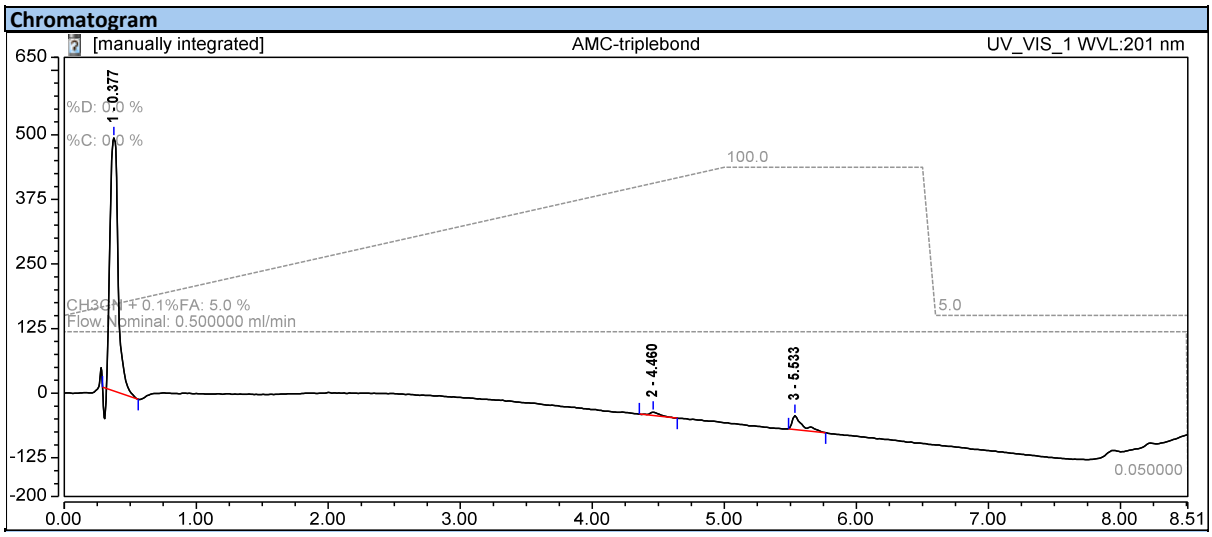
To a solution of hexynoic acid (167 µL, 0.61 mmol, 0.9 equiv.) and DIPEA (315 µL, 1.34 mmol, 2 equiv.) in DMF (2 mL) was added TSTU (228 mg, 0.74 mmol, 1.1 equiv.) and the solution was allowed to stir 1 hour until complete conversion of the starting material assessed through HPLC-MS. To a solution of **AMC** (134.7 mg, 0.67 mmol, 1 equiv.) in DMF (10 mL) at room temperature was added dropwise the latter reaction mixture containing *N*-hydroxysuccimide hex-5-ynoate (1 mL/3 hrs). The reaction was carefully monitored by HPLC-MS and the addition was stopped as soon as the *bis*-substituted compound was observed. The solution was then concentrated under vacuum, added water and purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 50% over 35 minutes, r.t: 12.5 minutes). After evaporation of the solvents and lyophilization, compound **17** was obtained (159.3 mg, 0.25 mmol, 38% chemical yield).

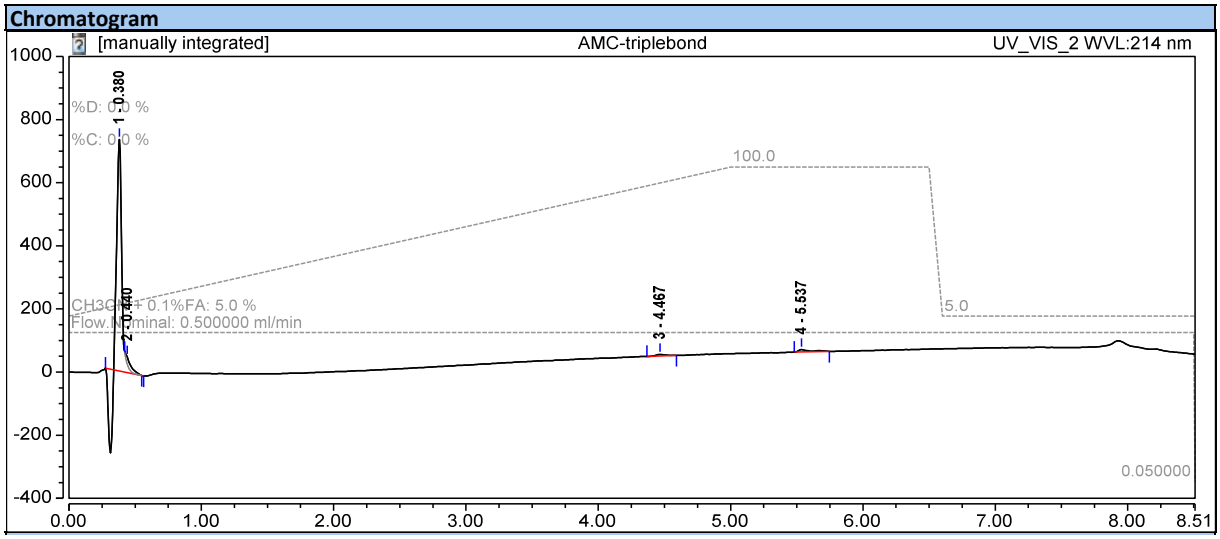
^{1}H NMR: (500 MHz, D_2O): δ. 3.70-2.77 (m, 14H), 2.37 (t, J = 7.4 Hz, 1H), 2.23 (m, 2H), 1.80 (m, 1H), 1.36 (m, 6H) (signals of intracyclic amines and amide are missing due to proton solvent exchange).

^{13}C NMR: (126 MHz, D_2O): δ. 177.04, 162.83, 119.09, 117.16, 115.23, 84.58, 69.91, 52.26, 46.42, 44.18, 43.89, 42.08, 40.51, 39.01, 23.75.

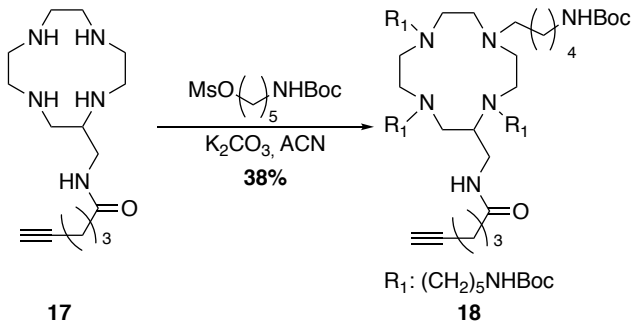
ESI-HRMS: $[M+H]^+$ m/z = 296.2434 (calcd. for $C_{15}H_{30}N_5O$: 296.2450)

HPLC:





Compound **18**:



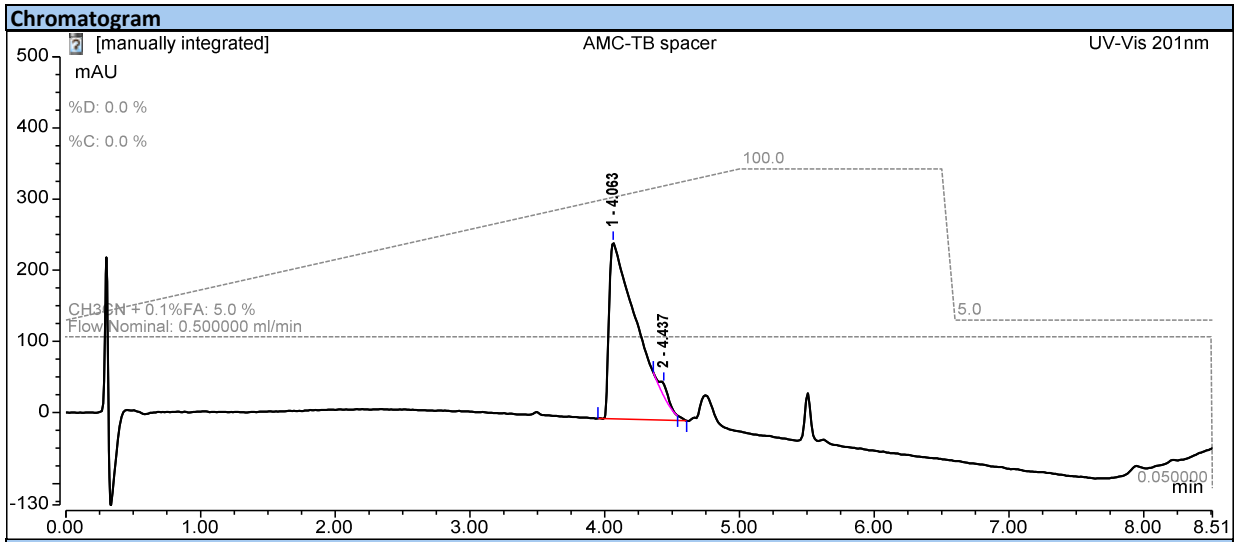
17 **18**

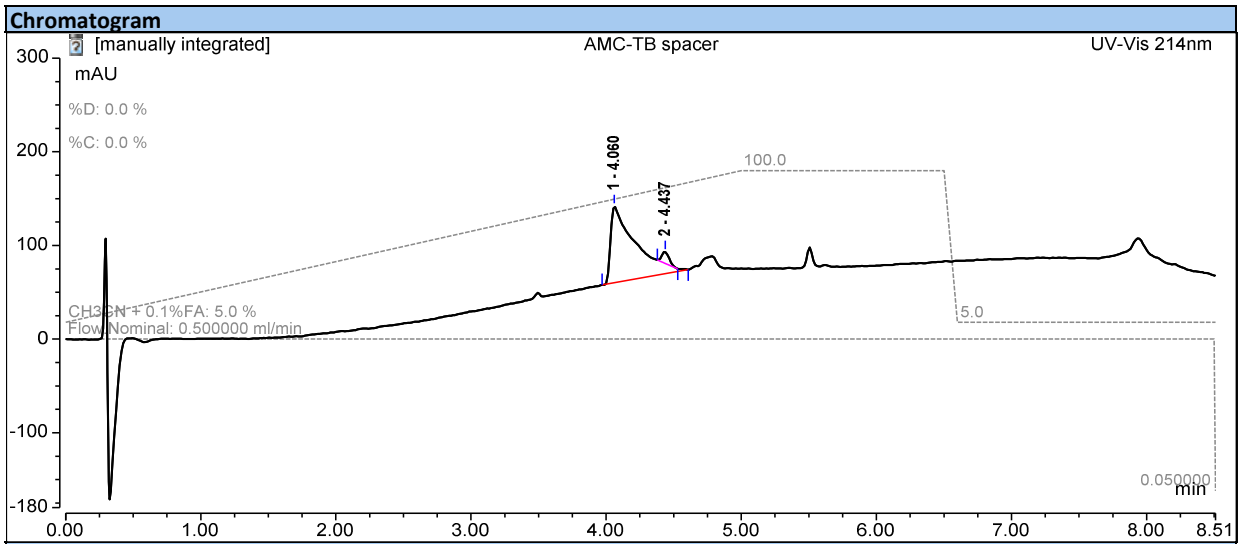
To a solution of compound **17** (536.4 mg, 0.84 mmol, 1 equiv.) in MeCN (7 mL) were added 5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate (1882.7 mg, 6.7 mmol, 8 equiv.) and K_2CO_3 (814.2 mg, 6.7 mmol, 8 equiv.). The solution was heated at 70°C and it was allowed to stir for 72 hours, the reaction was monitored through HPLC-MS. After adequate conversion, the crude was filtered, and concentrated under vacuum. The residue was then purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 60 % over 50 minutes, r.t: 22 minutes), obtaining compound **18** as a white solid after lyophilization. (341.5 mg, 0.32, mmol, 38 % chemical yield).

^{1}H NMR: (500 MHz, Methanol-d_4) δ 4.28 (t, *J* = 6.2 Hz, 1H), 3.67 (s, 1H), 3.17-3.03 (m, 15H), 2.98-2.77 (m, 10H), 2.74-2.69 (m, 3H), 2.35 (t, *J* = 7.5 Hz, 2H), 2.29 (s, 1H), 2.24 (m, 2H), 2.05 (s, 1H), 1.86-1.78 (m, 7H), 1.77-1.65 (m, 7H), 1.61-1.50 (m, 14H), 1.48-1.36 (m, 36H) (signals of intracyclic amines and amide are missing due to proton solvent exchange).

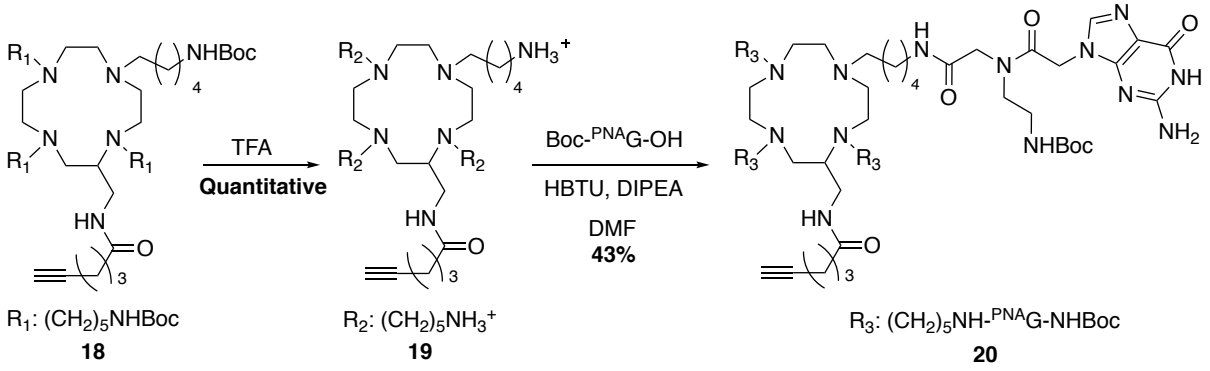
ESI-HRMS: $[M+H]^+$ m/z = 1036.81107 (calcd. for $C_{55}H_{105}N_9O_9$: 1036.81080).

HPLC:


Chromatogram
[manually integrated]
AMC-TB spacer
UV-Vis 201nm
mAU
%D: 0.0 %
%C: 0.0 %
100.0
5.0
1 - 4.063
2 - 4.437
CH3CN + 0.1%FA: 5.0 %
Flow Nominal: 0.500000 ml/min
0.050000
min
500
400
300
200
100
0
-130
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
AMC-TB spacer
UV-Vis 214nm
mAU
%D: 0.0 %
%C: 0.0 %
100.0
5.0
1 - 4.060
2 - 4.437
CH3CN + 0.1%FA: 5.0 %
Flow Nominal: 0.500000 ml/min
0.050000
min
300
200
100
0
-100
-180
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51

Compound **20**:

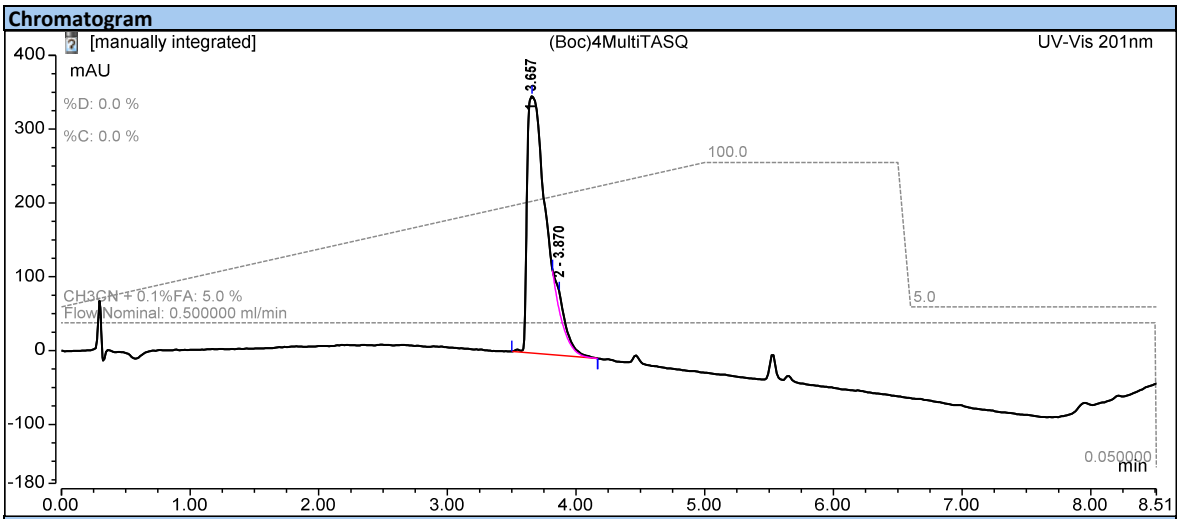


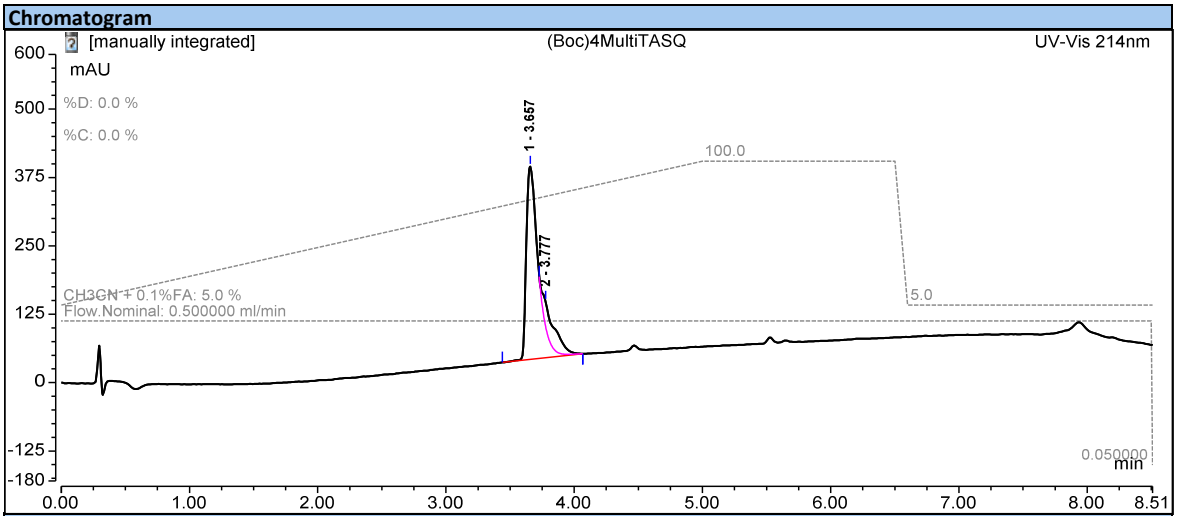
Compound **18** (14.8 mg, 0.014 mmol, 1 equiv.) was stirred with 500 µL of TFA during 1 hour to deprotect the amines. After TFA evaporation, the compound was directly engaged in following reaction without purifications. Boc-^{PNA}G (25.8 mg, 0.063 mmol, 4.5 equiv.) and HBTU (23.9 mg, 0.063 mmol, 4.5 equiv.) were dissolved in DMF (1 mL) and DIPEA was added (12 µL, 0.056 mmol 4 equiv.) The mixture was let to stir for five minutes and added to the solution containing compound **19** (14.8 mg, 0.014 mmol, 1 equiv.) and DIPEA (3 µL, 1 equiv.) in DMF (1 mL). The mixture was stirred at RT for 3 hours and the completion of the four couplings was assessed by HPLC-MS. The solution was then concentrated under vacuum, added a mixture of water and MeCN (50/50, 2mL), and purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 15% over 5 minutes then from 15 to 65% over 50 minutes, r.t: 37 minutes). After evaporation of the solvents by freeze drying, the compound **20** was obtained (13.5 mg, 6.14 µmol, 43% chemical yield).

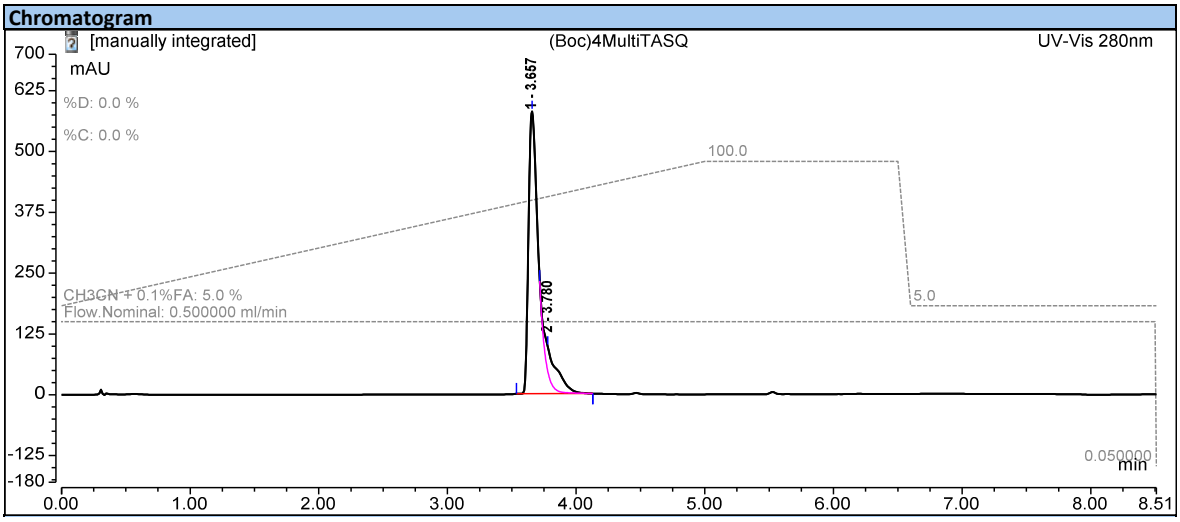
^{1}H NMR: (500 MHz, Methanol-d_4) δ 8.22 (m, 2H), 7.96 (m, 4H), 5.06 (m, 10H), 4.95 (m, 4H), 4.15 (s, 2H), 3.91 (s, 8H), 3.55 (s, 8H), 3.45-3.25 (m, 14H), 3.16-2.77 (m, 24H), 2.34 (s, 3H), 2.18 (s, 1H), 2.13 (m, 2H), 1.70 (m, 4H), 1.45-1.33 (m, 42H) 1.31 (s, 10H), 1.26-0.94 (m, 14H).

ESI-HRMS: $[M+H]^{2+}$ m/z = 1101.13051 (calcd. for $C_{99}H_{157}N_{37}O_{21}$: 1101.12501).

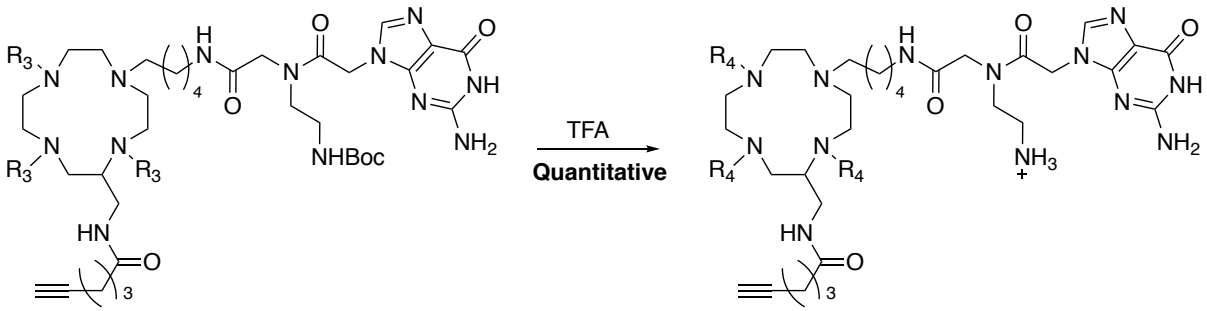
HPLC:


Chromatogram
[manually integrated]
(Boc)4MultiTASQ
UV-Vis 201nm
mAU
%D: 0.0 %
%C: 0.0 %
1 - 3.657
2 - 3.870
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
min


Chromatogram
[manually integrated]
(Boc)4MultiTASQ
UV-Vis 214nm
mAU
%D: 0.0 %
%C: 0.0 %
1 - 3.657
2 - 3.777
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
min


Chromatogram
[manually integrated]
(Boc)4MultiTASQ
UV-Vis 280nm
mAU
%D: 0.0 %
%C: 0.0 %
1 - 3.657
2 - 3.780
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
min

MultiTASQ:



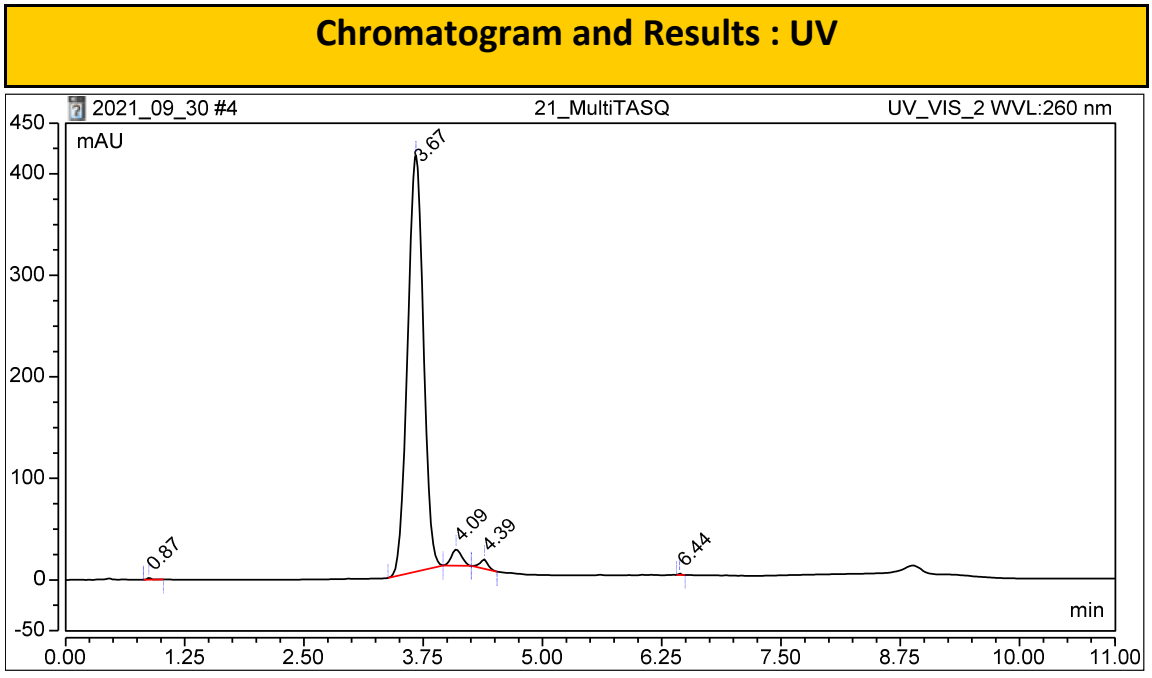
R_3: $(CH_2)_5NH$-^{PNA}G-NHBoc
20

R_4: $(CH_2)_5NH$-^{PNA}G-NH_3^+
MultiTASQ

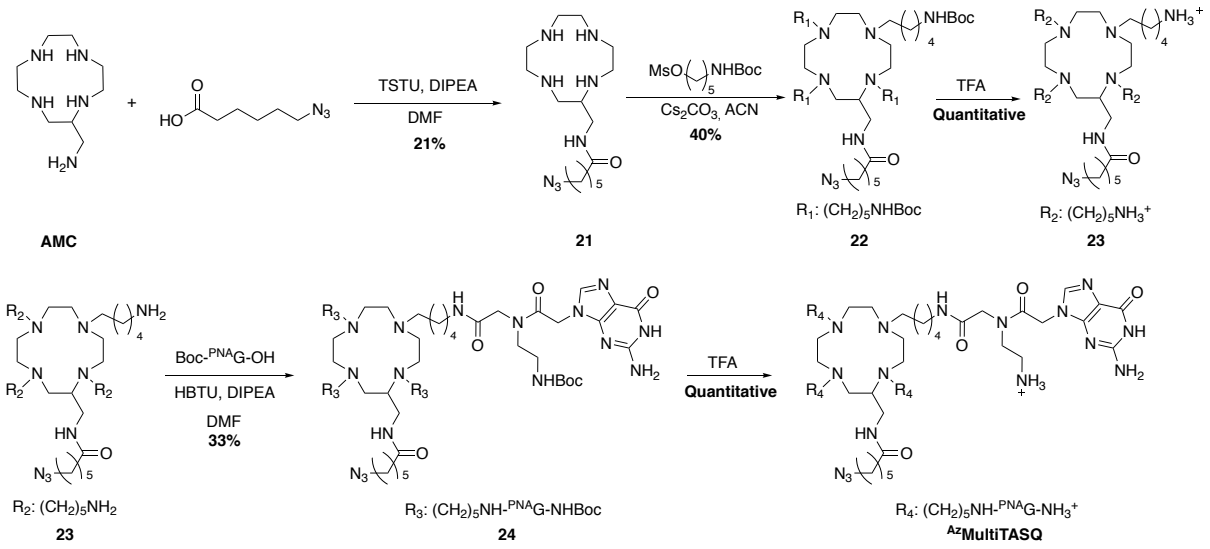
Compound **20** (1.95 mg, 0.9 µmol) was dissolved in TFA (100 µL) and the complete deprotection was assessed *via* HPLC, the mixture was diluted in water and the compound was freeze dried to afford **MultiTASQ.** (2 mg, 0.9 µmol, chemical yield 100%).

ESI-HRMS: $[M+H]^{2+}$ m/z = 901.01988 (calcd. for $C_{79}H_{125}N_{37}O_{13}$: 901.02015)

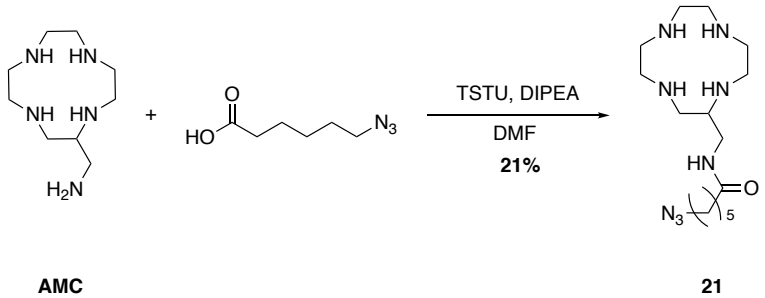
HPLC:



^{Az}MultiTASQ:



Compound **21**:



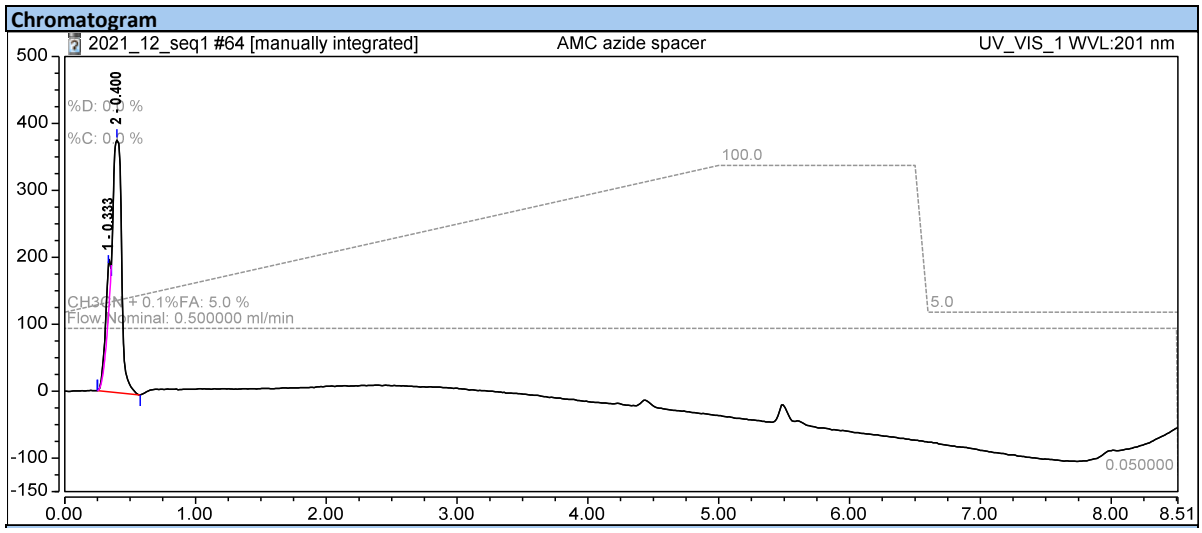
To a solution of 6-Azido-hexanoic acid (414 mg, 2.64 mmol, 0.9 equiv.) and DIPEA (533 µL, 3.21 mmol, 1.1 equiv.) in DMF (2 mL) was added TSTU (995.7 mg, 3.21 mmol, 1.1 equiv.) and the solution was allowed to stir 1 hour until complete reaction of the starting material (assessed through HPLC-MS). To a solution of **AMC** (588.9 mg, 2.92 mmol, 1 equiv.) in DMF (250 mL) at room temperature was added dropwise the latter reaction mixture containing N-hydroxylsuccimide-6-azido-hexanate (1 mL/3 hrs). The reaction was carefully monitored by HPLC-MS and the addition was stopped as soon as the *bis*-substituted compound was observed. The solution was then concentrated under vacuum, added water and purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 50% over 35 minutes, r.t: 13 minutes). After evaporation of the solvents, compound **21** was obtained (416.0 mg, 0.61 mmol, 21% chemical yield).

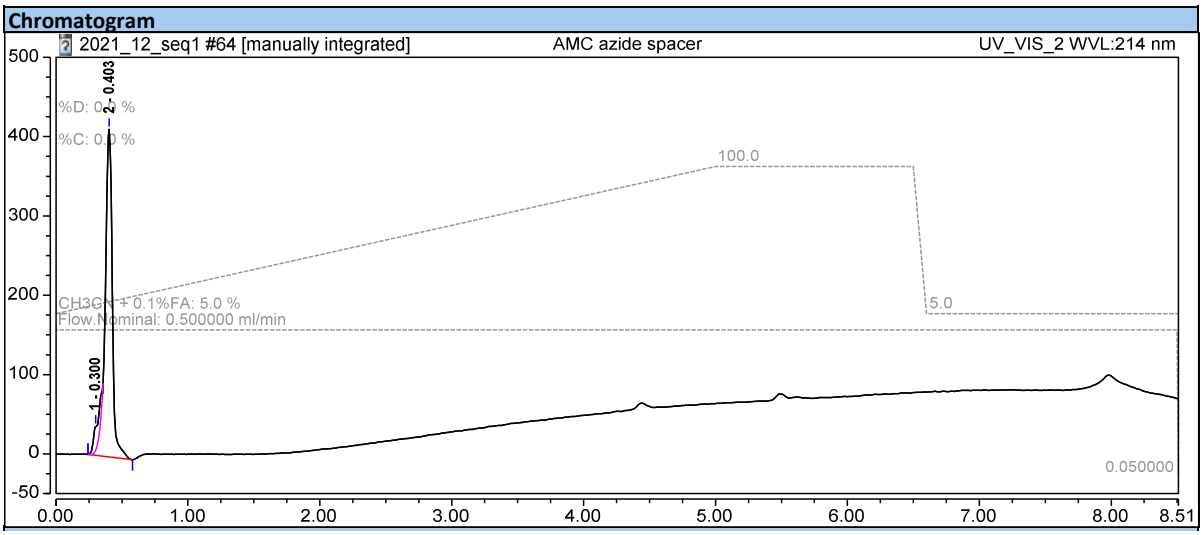
^{1}H NMR: (500 MHz, D_2O): δ 3.31 (d, *J* = 5.4 Hz, 2H), 3.21 (t, *J* = 6.8 Hz, 2H), 3.17 – 2.81 (m, 15H), 2.19 (t, *J* = 7.4 Hz, 2H), 1.51 (m, 4H), 1.27 (m, 4.1 Hz, 2H) (signals of intracyclic amines and amide are missing due to proton solvent exchange).

^{13}C NMR: (126 MHz, D_2O): δ 177.8, 119.7, 117.4, 115.1, 112.8, 51.0, 44.3, 43.9, 42.7, 42.1, 39.0, 35.4, 27.6, 25.5, 24.7.

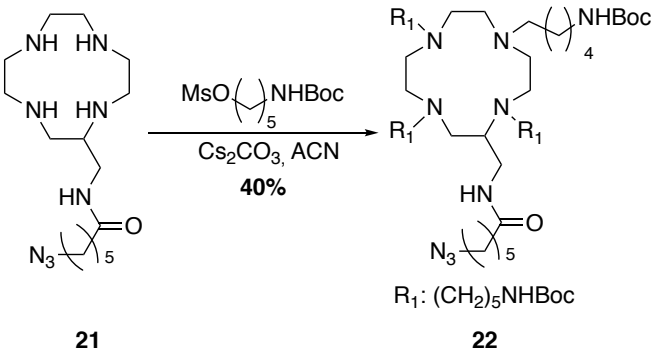
MALDI-TOF : $[M+H]^+$ m/z = 341.33 (calcd. for $C_{15}H_{33}N_8O$: 341.27)

HPLC:





Compound **22**:



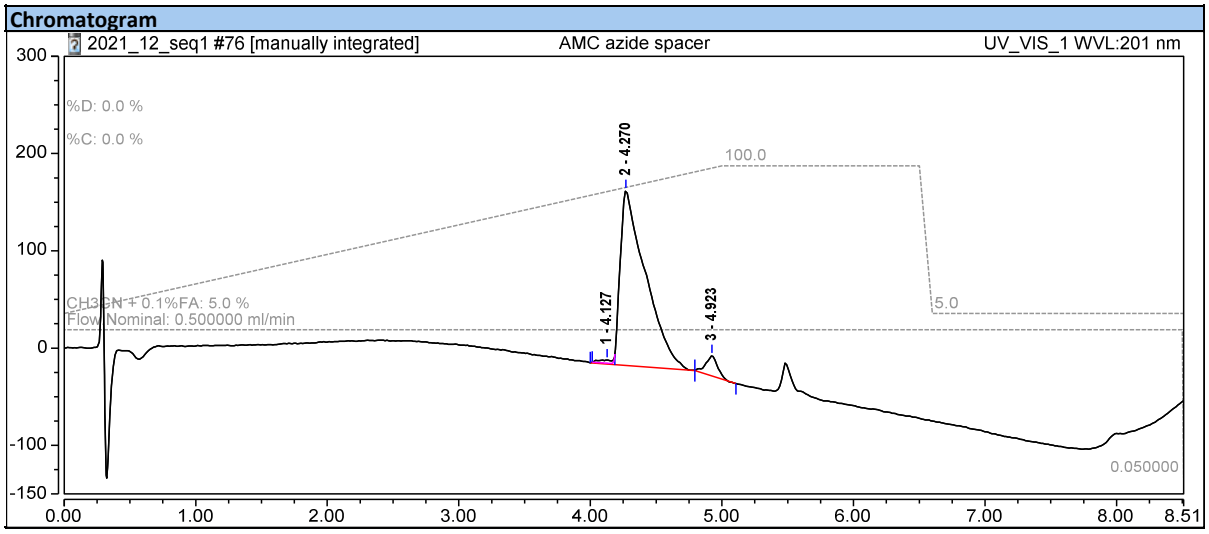
21 **22**

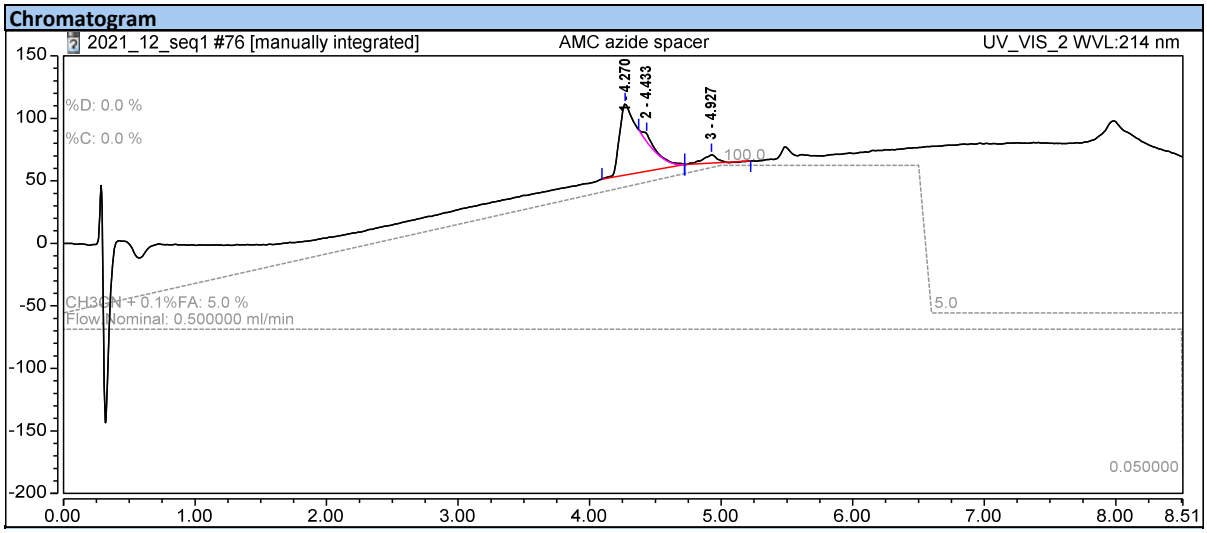
To a solution of compound **21** (145.7 mg, 0.21 mmol, 1 equiv.) in MeCN (2.5 mL) was added 5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate (576.6 mg, 2.1 mmol, 10 equiv.) and Cs_2CO_3 (668.3 mg, 2.1 mmol, 10 equiv.). The solution was heated at 50 °C and it was allowed to stir for 72 hours, the reaction was monitored through HPLC-MS. After complete conversion, the crude was filtered, and concentrated under vacuum. The residue was then purified by semi-preparative RP-HPLC with a H_2O/MeCN + 0.1% TFA mixture as mobile phase (gradient of 5 to 60 % over 50 minutes, r.t: 22 minutes). Compound **22** was isolated as a white solid (92.4 mg, 0.085, mmol, 40 % chemical yield).

^{1}H NMR: (500 MHz, Methanol-d_4) δ 3.73 – 3.25 (m, 6H), 3.22 (s, 4H), 3.19 – 3.02 (m, 9H), 2.96 (m, 10H), 2.75 (t, *J* = 36.3 Hz, 7H), 2.15 (t, *J* = 7.5 Hz, 2H), 1.84 – 1.49 (m, 10H), 1.48 – 1.38 (m, 9H), 1.34 (s, 42H), 1.26 – 1.17 (m, 4H) (signals of intracyclic amines and amide are missing due to proton solvent exchange).

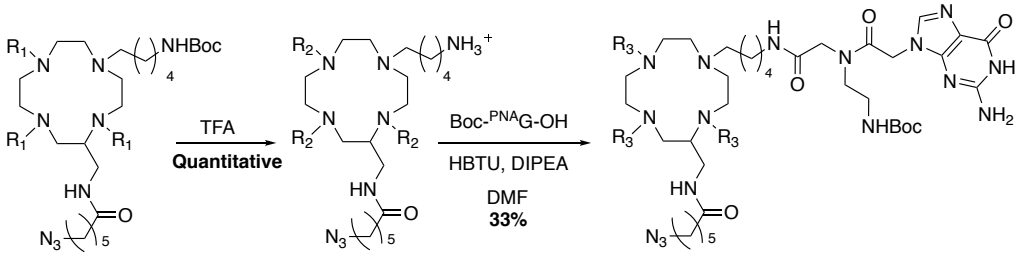
MALDI-TOF : $[M+H]^+$ m/z = 1081.99 (calcd. for $C_{55}H_{109}N_{12}O_9$: 1081.84).

HPLC:





Compound **24**:



R_1: $(CH_2)_5NHBoc$ **22**

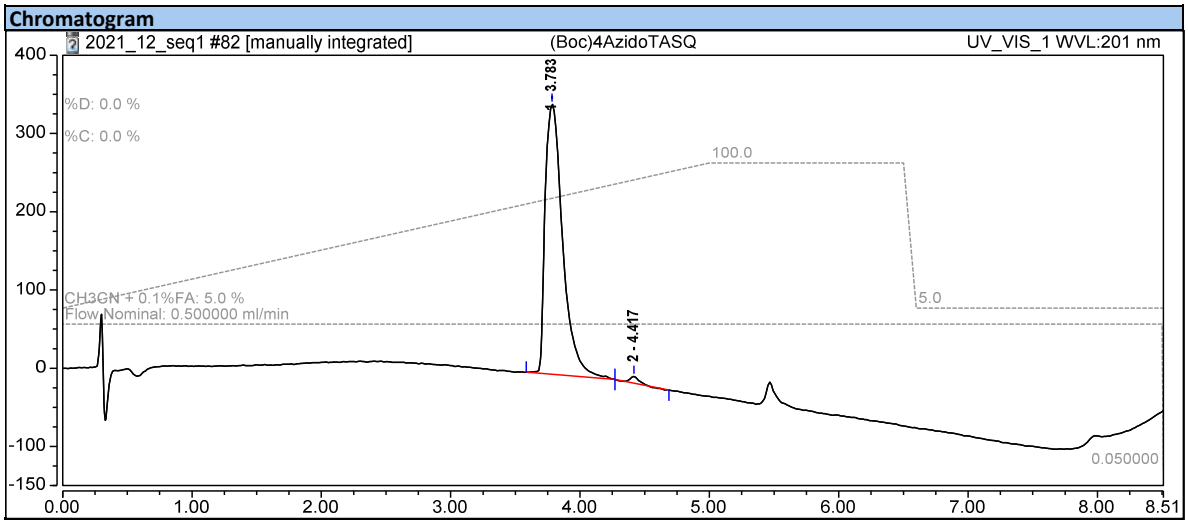
R_2: $(CH_2)_5NH_3^+$ **23**

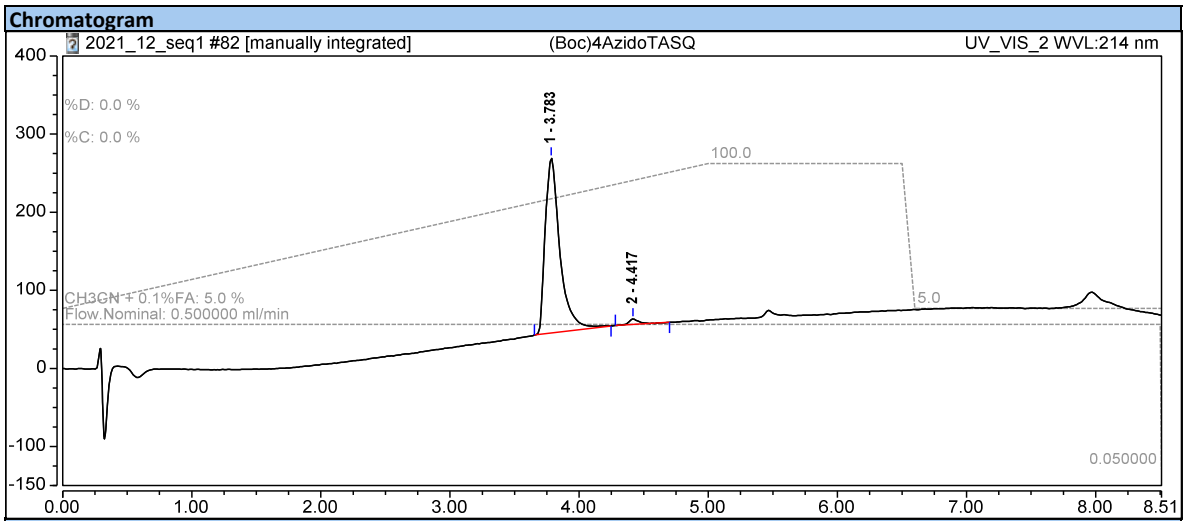
R_3: $(CH_2)_5NH$-PNAG-NHBoc **24**

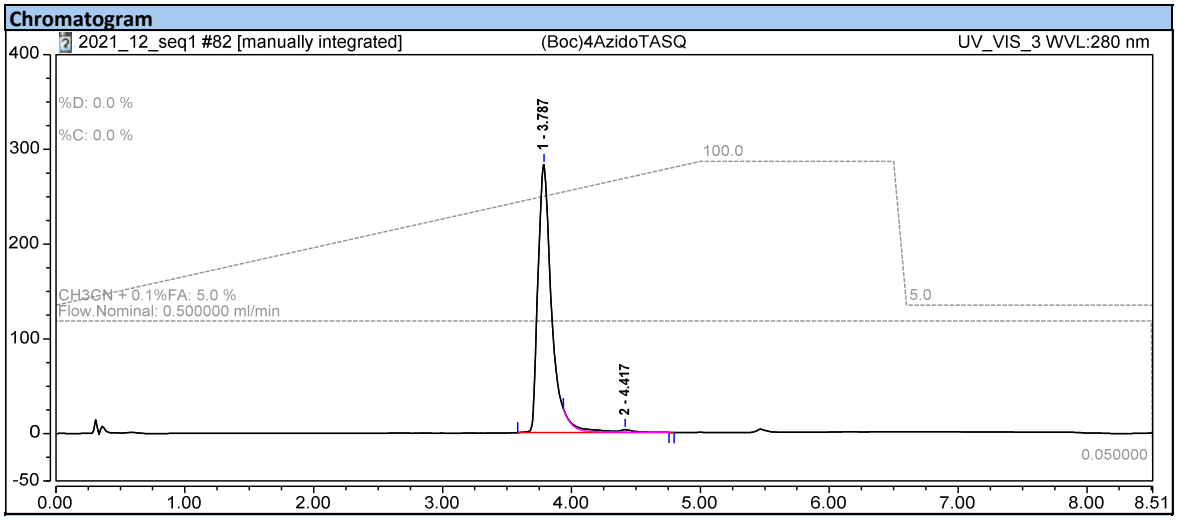
Compound **22** (50.6 mg, 0.042 mmol, 1 equiv.) was stirred with 500 µL of TFA for 1 hour to deprotect the amines. After evaporation of the TFA (completion of the deprotection assessed by HPLC-MS), compound **23** was used without purification. In a second-round bottom flask, Boc-^{PNA}G (76.1 mg, 0.186 mmol, 4.4 equiv.), HBTU (70.5 mg, 0.19 mmol, 4.4 equiv.) were dissolved in DMF (1 mL), and DIPEA was added (43 µL, 6 equiv.) The mixture was added to the solution containing previously described compound **23** (50.6 mg, 0.042 mmol, 1 equiv.) and DIPEA (3 µL, 1 equiv.) in DMF (1 mL). The mixture was stirred at RT for 3 hours and the completion of the four couplings was assessed by HPLC-MS. The solution was then concentrated under vacuum, added a mixture of water and MeCN (50/50, 2mL), and purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 5 to 15% over 5 minutes then from 15 to 65% over 50 minutes, column *A*, r.t: 37 minutes). After evaporation of the solvents, compound **24** was obtained (33.6 mg, 0.015 mmol, 33% chemical yield).

ESI-HRMS: $[M+H]^{2+}$ m/z = 1123.64146 (calcd. for $C_{99}H_{161}N_{40}O_{21}$: 1123.64136).

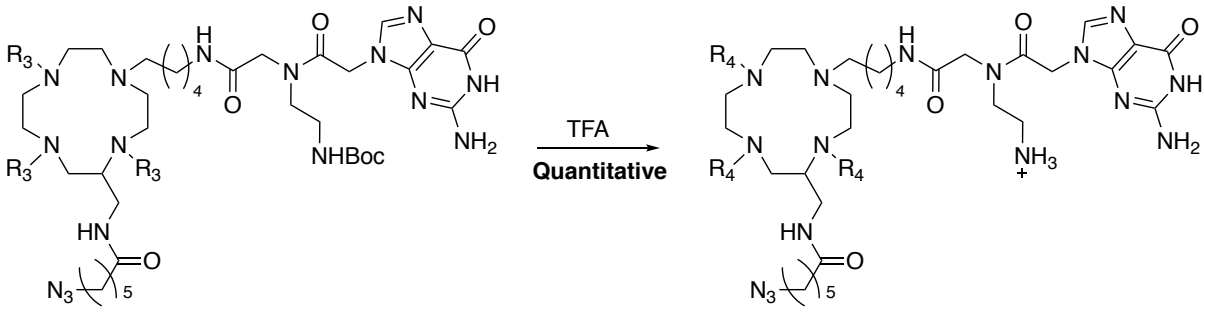
HPLC:







^{Az}MultiTASQ:



R_3: $(CH_2)_5NH$-^{PNA}G-NHBoc

24

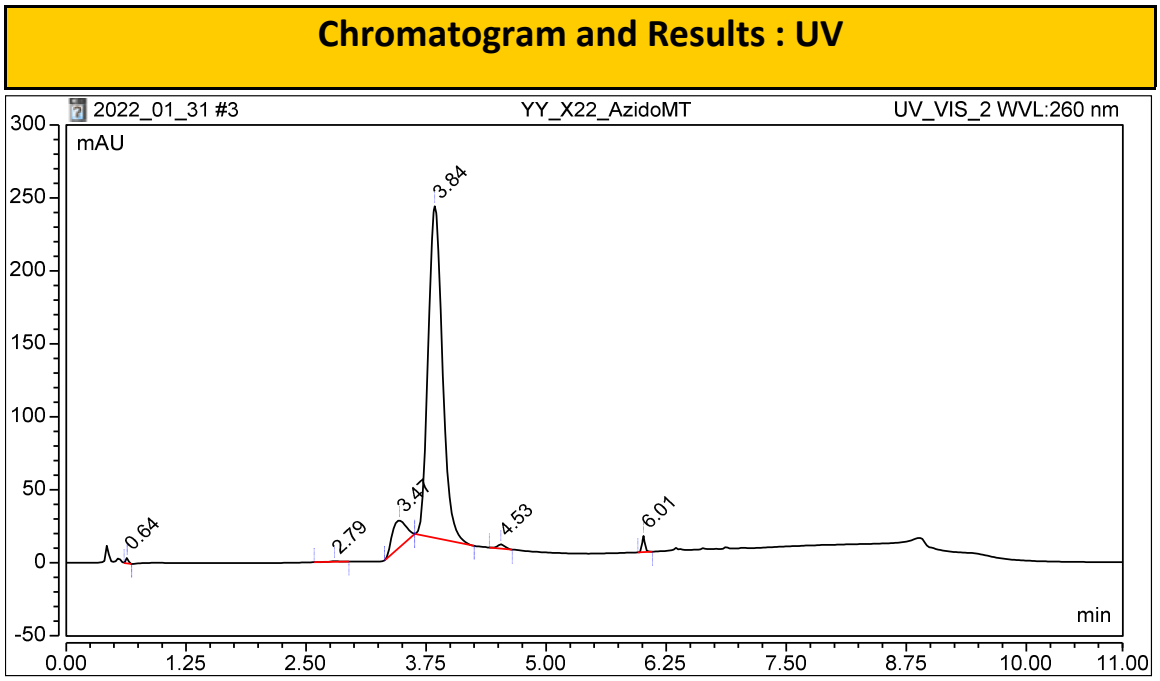
R_4: $(CH_2)_5NH$-^{PNA}G-NH_3^+

^{Az}MultiTASQ

Compound **24** (1.95 mg, 0.9 µmol) was dissolved in TFA (100 µL) and the complete deprotection was assessed by HPLC, the mixture was diluted in water and the compound was freeze-dried to afford **AzidoMultiTASQ.** (2 mg, 0.9 µmol, chemical yield 100%).

ESI-HRMS: $[M+H]^{3+}$ m/z = 616.026740 (calcd. for $C_{79}H_{129}N_{40}O_{13}$: 616.02676).

HPLC:

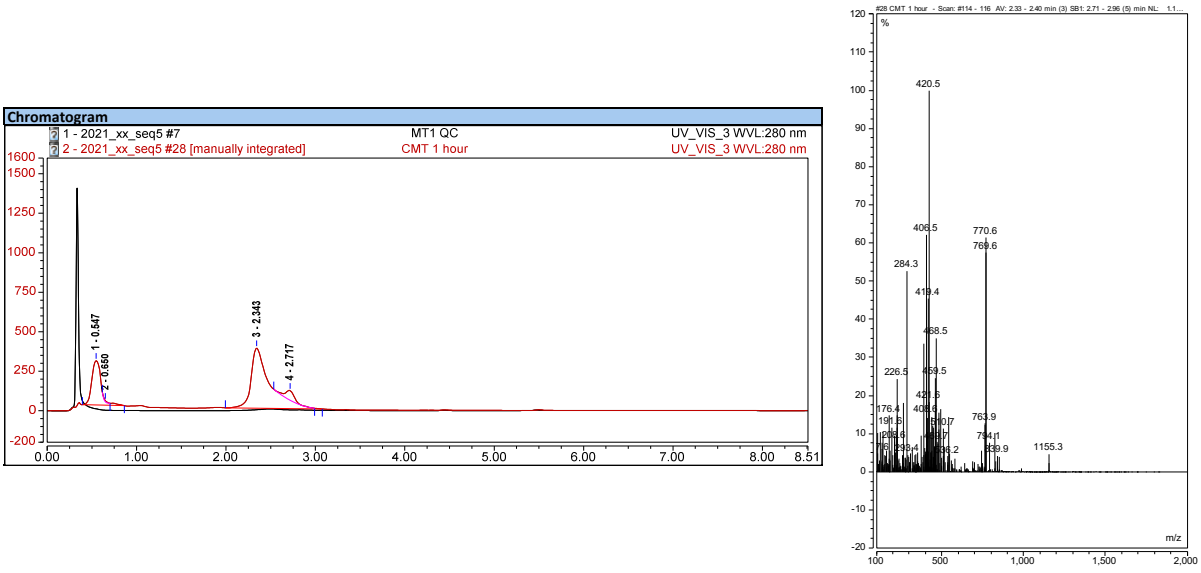


Click Protocol:

CuAAC:

To perform CuAAC a soluble source of Cu(I) is necessary, so in micro-tube 5 µL of a 1M solution of $(MeCN)_4Cu{\cdot}PF_6$ in DMSO are mixed with 7 µL of an aqueous solution of THPTA 1M (*tris*(3-hydroxypropyltriazolylmethyl)amine), the mixture was stirred to provide a dark blue color solution. To which were added 10 µL of sodium ascorbate (1 M in water) to provide a colorless of Cu(I) (227 mM). In a second micro-tube, 20 µL of a 5 mM solution of **MultiTASQ** (water/1-Butanol 1:1) were mixed with 1.1 µL of Azido-PEG_4-Biotin conjugate (100 mM in water), to which 4µL of the aforementioned Cu(I) containing mixture were added. The reaction was left to stir for one hour after which the complete conversion was assessed by HPLC-MS. The solution was let to decant for further 15 minutes and a blue precipitated formed. After centrifugation the supernatant was removed and diluted to 1 mM (presence of target compound attested by HPLC-MS).

LR-MS: $[M + Cu]^{2+}$ m/z = 1153.3 (calcd. for $C_{97}H_{157}N_{43}O_{18}SCu$: 1153.6).

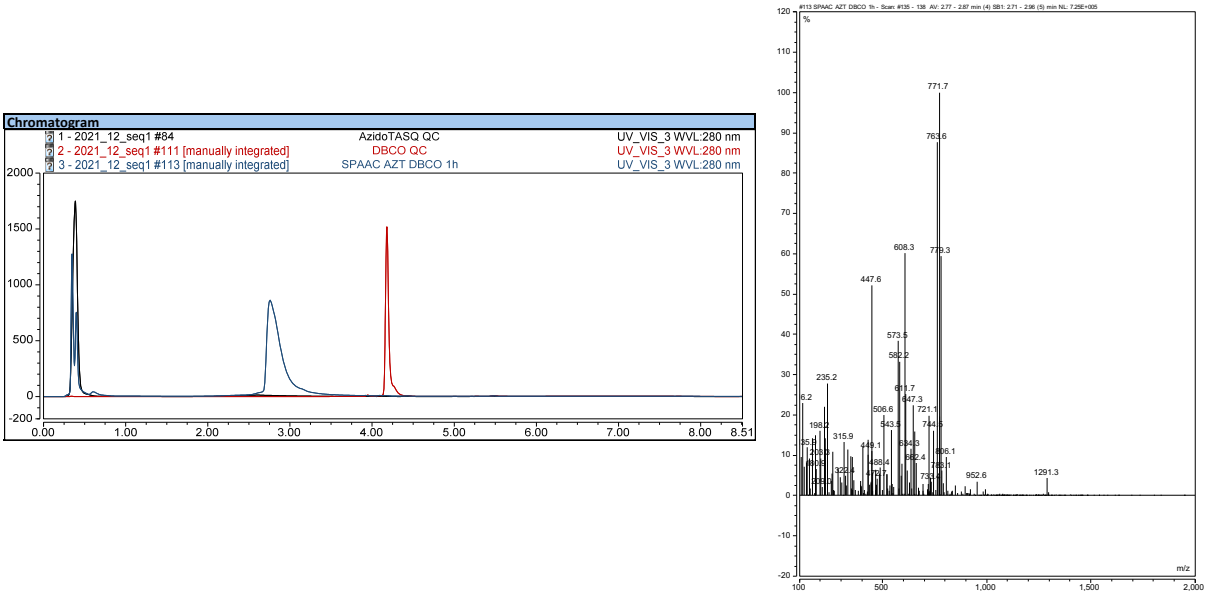


In black MultiTASQ prior to the reaction, in red the biotinylated MultiTASQ after 1 hour of reaction.

SPAAC:

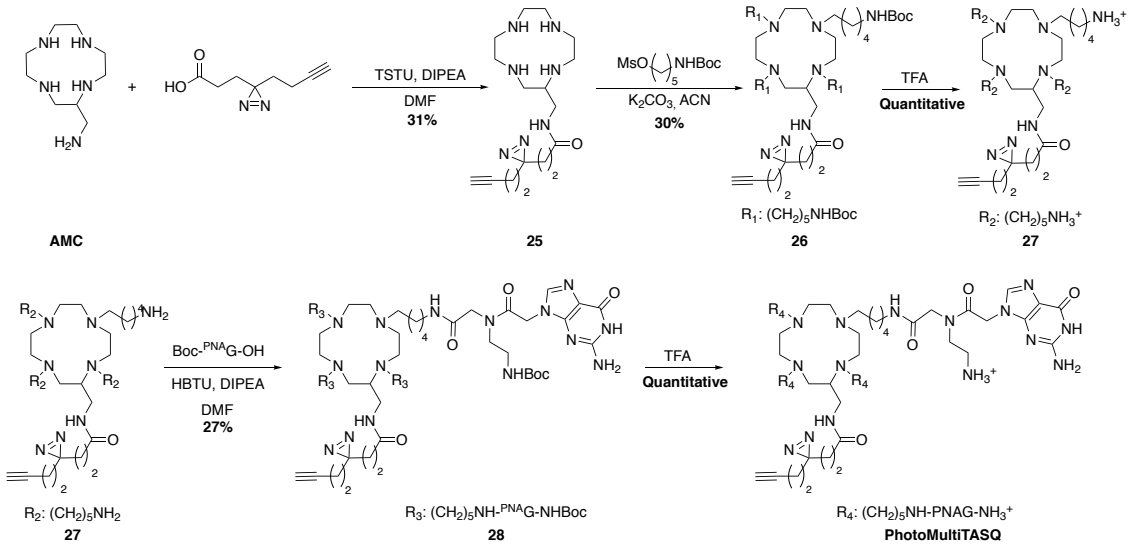
In a micro-tube, 20 µL of a 1mM solution of **^{Az}MultiTASQ** were mixed with 2.2 µL of 10mM solution of commercial dibenzocyclooctyne-PEG$_4$-biotin conjugate (DBCO-PEG$_4$-Biotin), the reaction was stirred at 37°C for 1 hour until complete conversion of the starting material was observed by HPLC-MS. The TASQ solution was used without further purification.

LR-MS: $[M+H]^{2+}$ m/z = 1291.3 (calcd. for $C_{118}H_{179}N_{45}O_{21}S$: 1291.9).

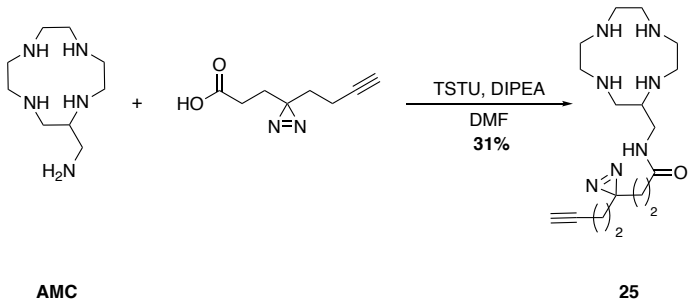


In black AzidoTASQ prior to the reaction, in red the Dibenzocyclooctyne-PEG$_4$-biotin conjugate *before the reaction, in blue the biotinylated AzidoTASQ conjugated after 1hour at 37°C.*

photoMultiTASQ:



Compound **25**:



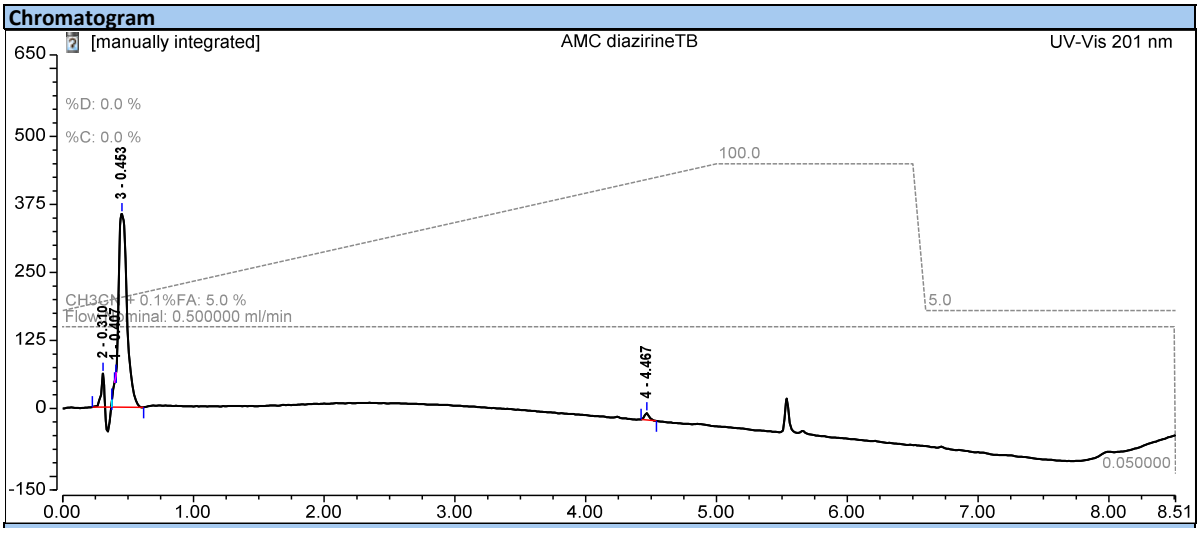
To a solution of 3-(3-But-3-yn-1yl)-3H-Diazirin-3yl) propanoic acid (18.0 mg, 0.11 mmol, 1 equiv.) and DIPEA (40 µL, 0.22 mmol, 2.1 equiv.) in DMF (3 mL) was added TSTU (42.1 mg, 0.14 mmol, 1.3 equiv.) and the solution was allowed to stir 1h until complete conversion of the starting material (monitored by HPLC). To a solution of **AMC** (29.3 mg, 0.14 mmol, 1.3 equiv.) in DMF (3 mL) at room temperature was added dropwise the latter reaction mixture containing the activated ester (1 mL/3 hrs). The reaction conversion was carefully monitored by HPLC-MS. The solution was then concentrated under vacuum, added water and purified by semi-preparative-HPLC in a H_2O/MeCN + 0.1% TFA mixture (gradient of 2 to 35 % over 35 minutes, r.t: 12.5 minutes). After evaporation of the solvents, the compound **25** was obtained (29.7 mg, 0.043 mmol, 31% chemical yield) (exact amount of TFA not quantified).

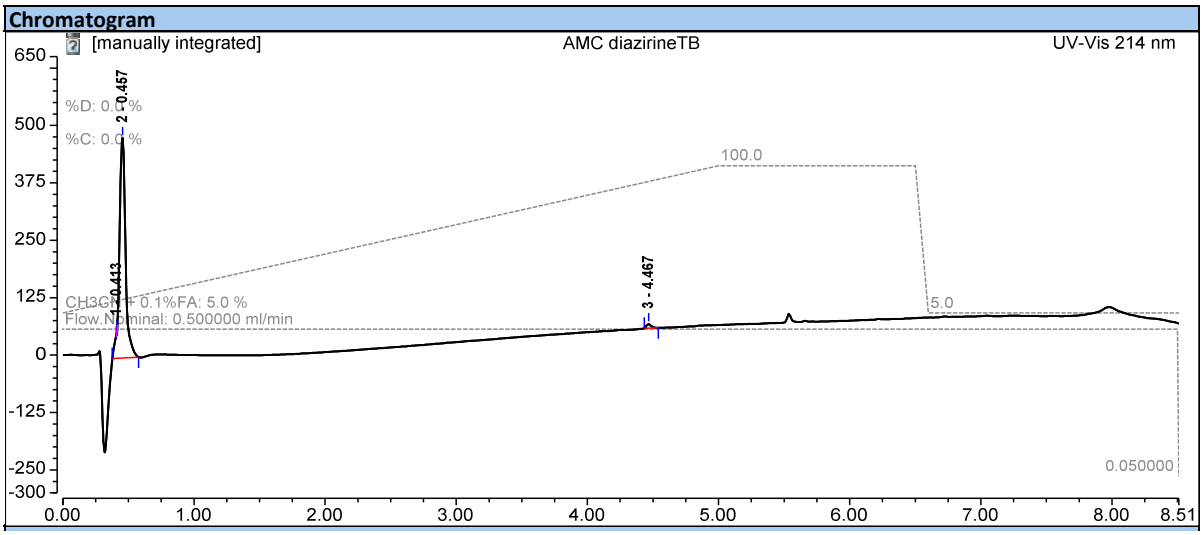
^{1}H NMR: (500 MHz, Deuterium Oxide) δ 3.40 (d, *J* = 5.4 Hz, 2H), 3.13 (m, 15H), 2.37 (s, 1H), 2.12 (t, *J* = 7.1 Hz, 2H), 2.03 (m, 2H), 1.83 (t, *J* = 7.1 Hz, 2H), 1.66 (t, *J* = 7.1 Hz, 2H) (signals of intracyclic amines and amide are missing due to proton solvent exchange).

^{13}C NMR: (126 MHz, Deuterium Oxide) δ 175.8, 119.8, 117.4, 115.1, 112.8, 84.3, 69.9, 52.1, 46.4, 44.3, 42.6, 39.0, 30.9, 29.8, 28.8, 27.6, 12.4.

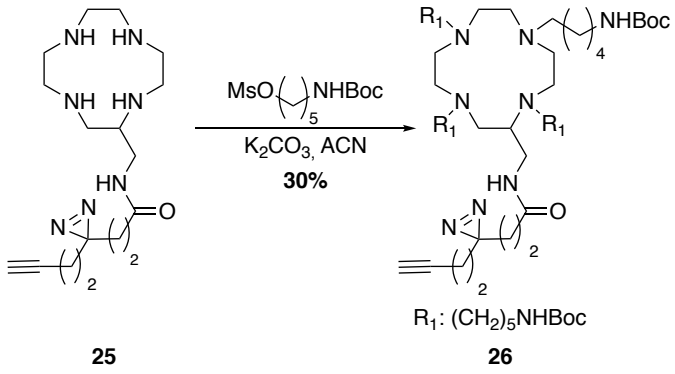
MALDI-TOF: $[M+H]^+$ m/z = 350.27 (calcd. for $C_{17}H_{31}N_7O$: 350.26)

HPLC:





Compound **26**:



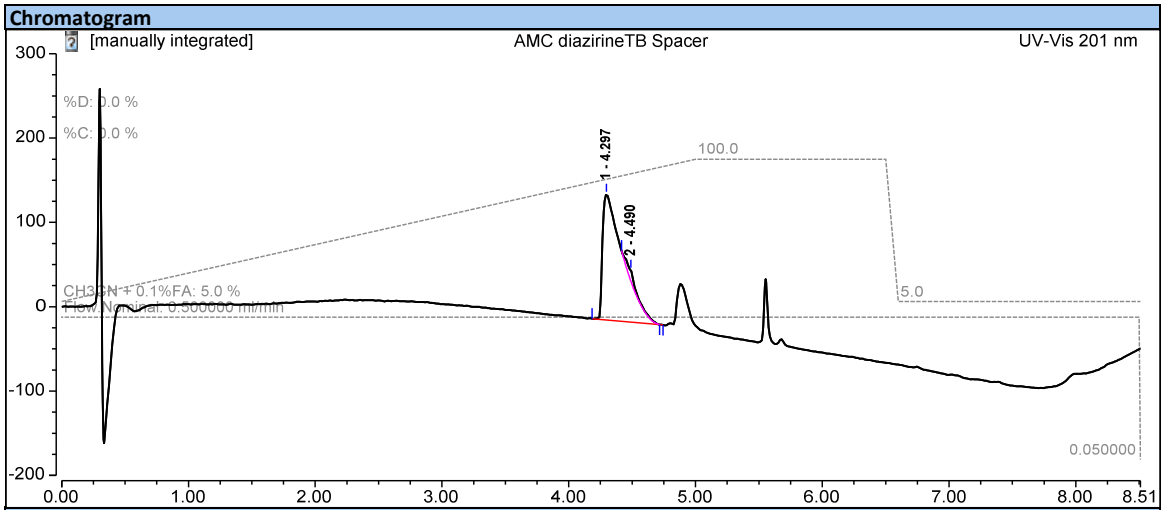
25 **26**

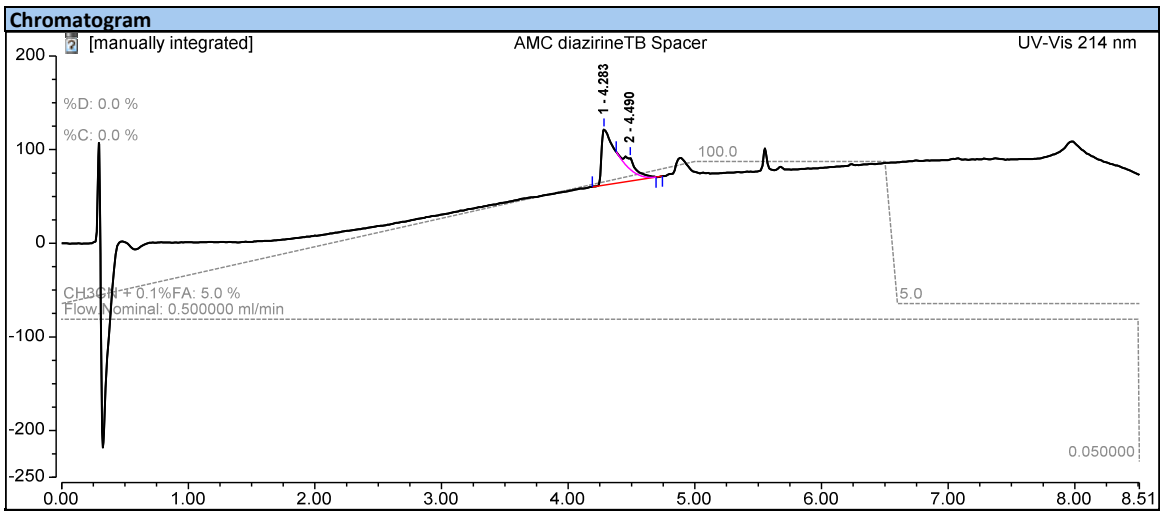
To a solution of **25** (34.2 mg, 0.049 mmol, 1 equiv.) in MeCN (1 mL) was added 5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate (82.6 mg, 0.29 mmol, 6 equiv.) and K_2CO_3 (40.1 mg, 0.29 mmol, 6 equiv.). The solution was heated at 60°C and it was allowed to stir for 48 hours. Two additional equivalents of 5-((*tert*-butoxycarbonyl)amino)pentyl methanesulfonate were added in order to complete the conversion of **25** and the mixture was allowed to stir for another 24 h. An adequate conversion of the reaction was monitored through HPLC-MS, after which the crude was filtered and concentrated under vacuum. The residue was then purified by semi-preparative-HPLC in a H_2O/MeCN + 0.1 % TFA mixture (gradient of 5 to 60 % over 50 minutes, r.t: 37 minutes), obtaining the compound **26** after solvent removal by lyophilization (16.0 mg, 14.5 µmol, 30% chemical yield, 83% purity).

^{1}H NMR: (500 MHz, Methanol-d_4) δ 3.79 – 3.34 (m, 8H), 3.27 – 2.99 (m, 20H), 2.83 (d, *J* = 41.3 Hz, 8H), 2.30 (t, *J* = 2.6 Hz, 1H), 2.04 (m, 4H), 1.85 (m, 4H), 1.63 (m, 4H), 1.54 (m, 10H), 1.44 (s, 42H), 1.37 – 1.21 (m, 5H) (signals of intracyclic amines and amide are missing due to proton solvent exchange).

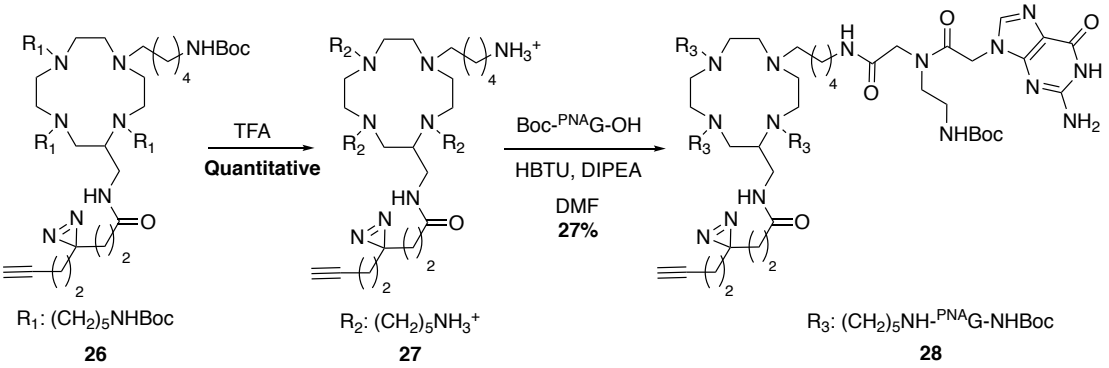
ESI-HRMS: $[M+H]^+$ m/z = 1090.83405 (calcd. for $C_{57}H_{107}N_{11}O_9$: 1090.83260).

HPLC:


Chromatogram
[manually integrated]
AMC diazirineTB Spacer
UV-Vis 201 nm
%D: 0.0 %
%C: 0.0 %
1 - 4.297
2 - 4.490
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow Nominal: 0.500000 ml/min
0.050000
300
200
100
0
-100
-200
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
AMC diazirineTB Spacer
UV-Vis 214 nm
%D: 0.0 %
%C: 0.0 %
1 - 4.283
2 - 4.490
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow Nominal: 0.500000 ml/min
0.050000
200
100
0
-100
-200
-250
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51

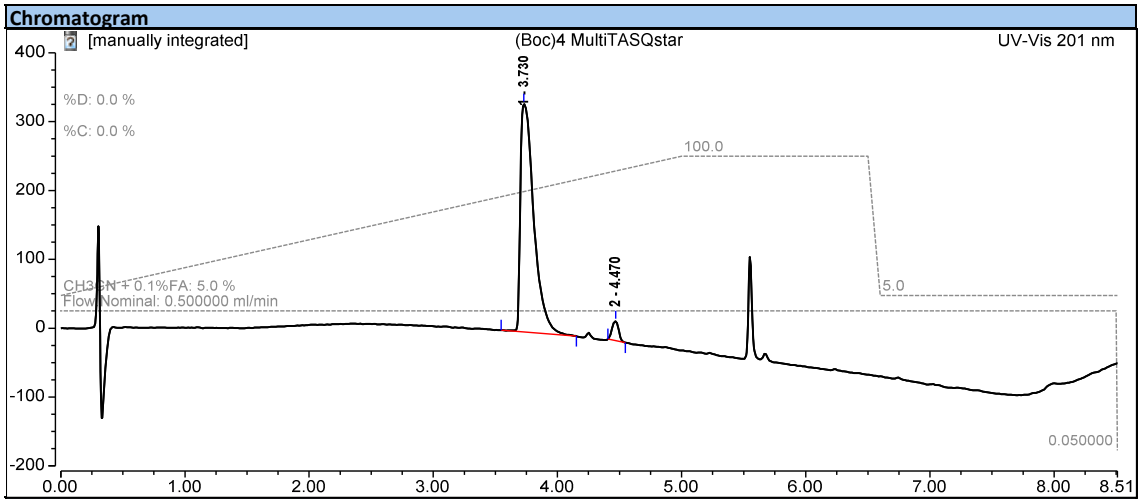
Compound **28**:

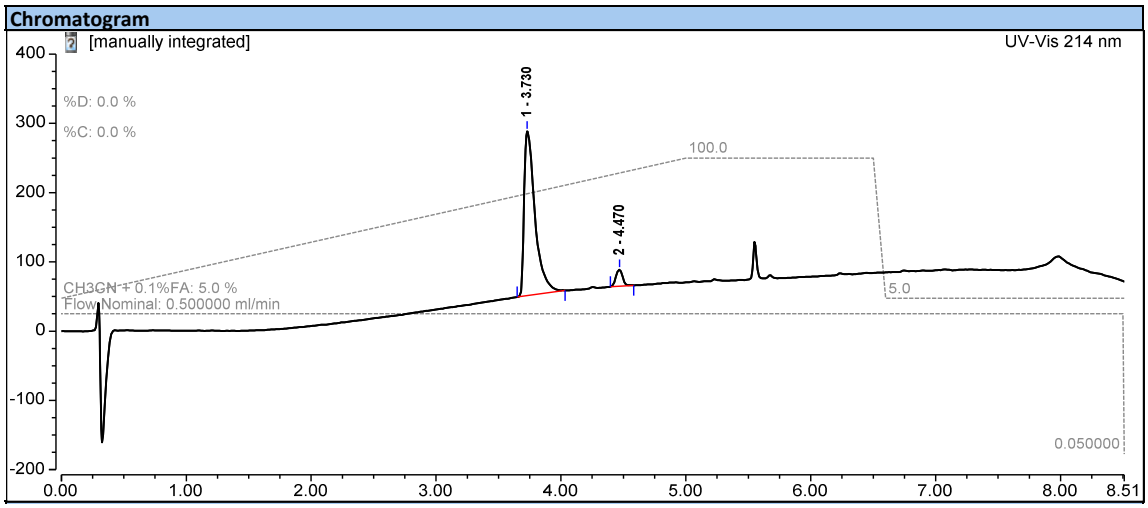


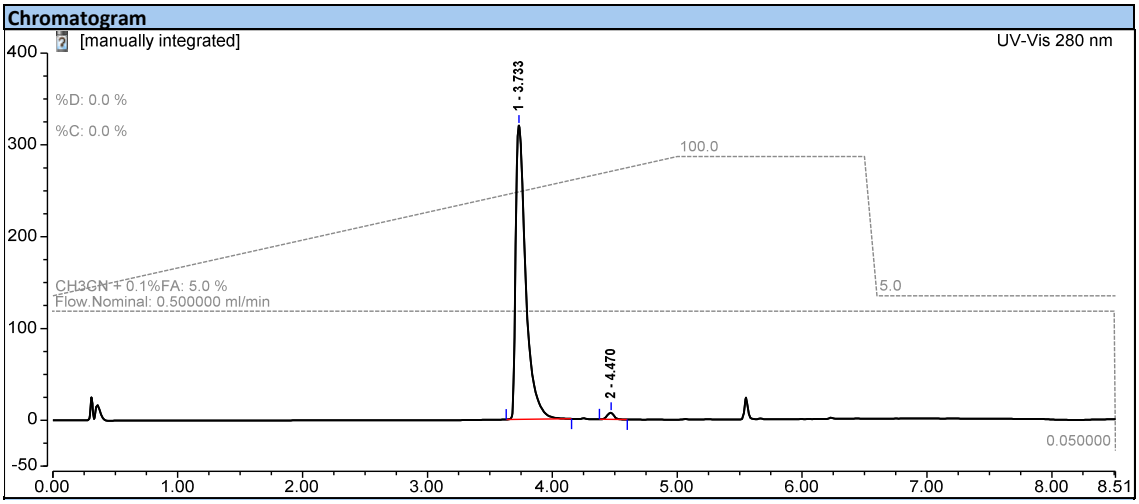
Compound **26** (15.1 mg, 0.014 mmol, 1 equiv.) was stirred with 500 µL of TFA during 1 hour. After evaporation of the TFA, the complete deprotection of **26** was observed with HPLC-MS and compound **27** was obtained (9.7 mg, 0.014 mmol, 100 % chemical yield). Boc-^{PNA}G (22.5 mg, 0.55 mmol, 4 equiv.) and HBTU (23.5 mg, 0.062 mmol, 4.4 equiv.) were dissolved in DMF (1 mL) to which DIPEA was added (10 µL, 4 equiv.). The activated ester containing mixture was added to a DMF (1 mL). solution of **27** (9.66 mg, 0.014 mmol, 1 equiv.) and DIPEA (10 µL, 4 equiv.). The mixture was stirred at RT for 3 hours. After observing its full conversion by HPLC-MS, the solution was concentrated under vacuum, added a mixture of water and MeCN (50/50, 2mL), and purified by semi-preparative RP-HPLC in a H_2O/MeCN + 0.1 % TFA mixture (gradient of 5 to 15 % over 5 minutes then from 15 to 65% over 50 minutes, column *A*, r.t: 29 minutes). After lyophilization of the solvents, the compound **28** was obtained (8.43 mg, 3.74 µmol, 27% chemical yield) (exact amount of TFA not quantified).

ESI-HRMS: $[M+H+Na]^{2+}$ m/z = 1139.12403 (calcd. for $C_{10}H_{159}N_{39}O_{21}Na$: 1139.12689).

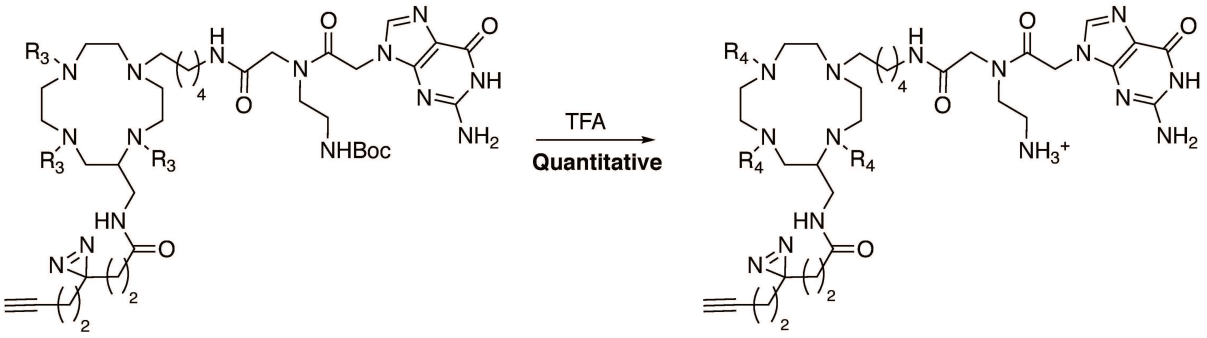
HPLC:


Chromatogram
[manually integrated]
(Boc)4 MultiTASQstar
UV-Vis 201 nm
%D: 0.0 %
%C: 0.0 %
3.730
100.0
2 - 4.470
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
400
300
200
100
0
-100
-200
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
UV-Vis 214 nm
%D: 0.0 %
%C: 0.0 %
1 - 3.730
100.0
2 - 4.470
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
400
300
200
100
0
-100
-200
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51


Chromatogram
[manually integrated]
UV-Vis 280 nm
%D: 0.0 %
%C: 0.0 %
1 - 3.733
100.0
2 - 4.470
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
0.050000
400
300
200
100
0
-50
0.00
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
8.51

photoMultiTASQ:



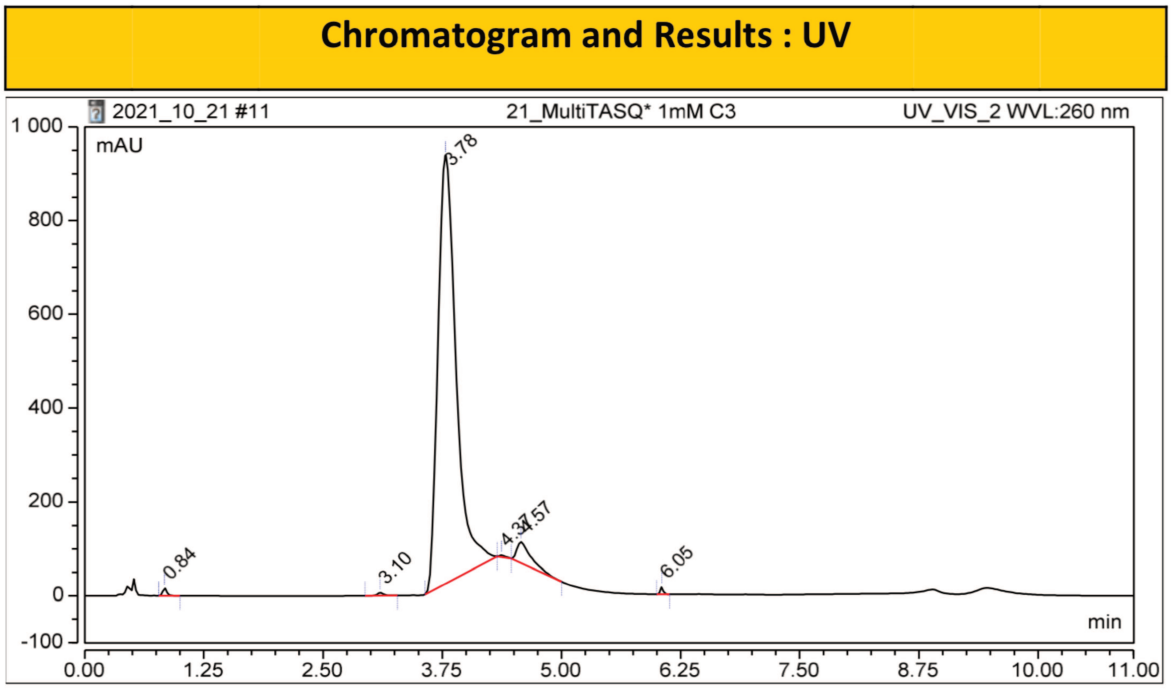
R_3: $(CH_2)_5NH$-^{PNA}G-NHBoc
28

R_4: $(CH_2)_5NH$-PNAG-NH_3^+
PhotoMultiTASQ

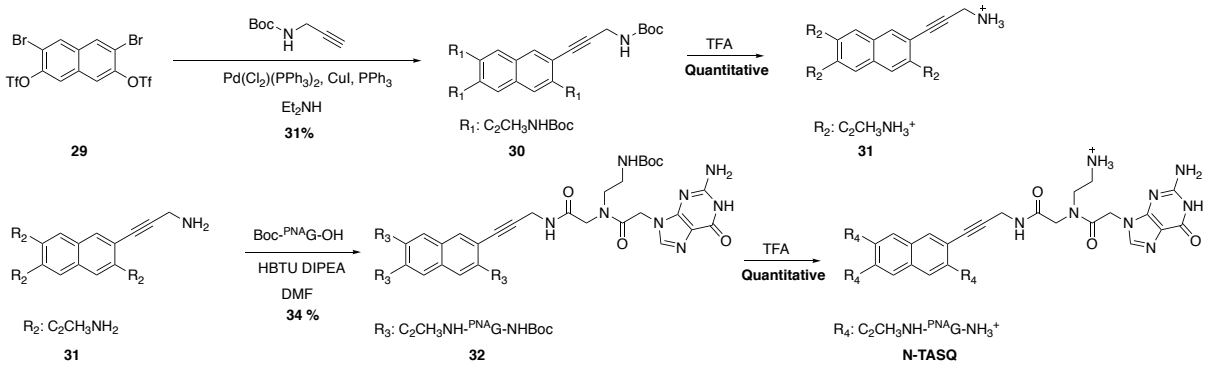
28 (1.95 mg, 0.9 µmol) was dissolved in TFA (200 µL) and the complete deprotection was assessed *via* HPLC. The mixture was diluted with water and the compound was freeze-dried to achieve **photoMultiTASQ.** (2 mg, 0.9 µmol, 100%).

ESI-HRMS: $[M+2H]^{2+}$ m/z = 928.03214 (calcd. for $C_{81}H_{127}N_{39}O_{13}$: 928.03105).

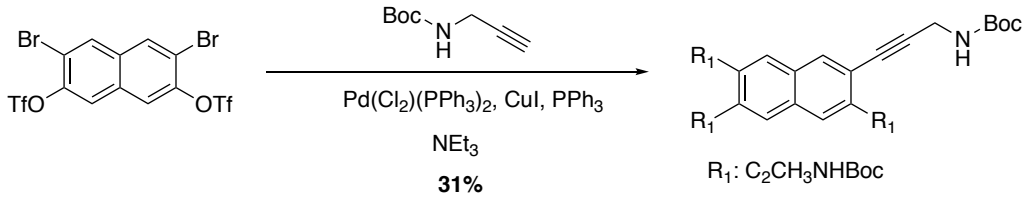
HPLC:



N-TASQ:



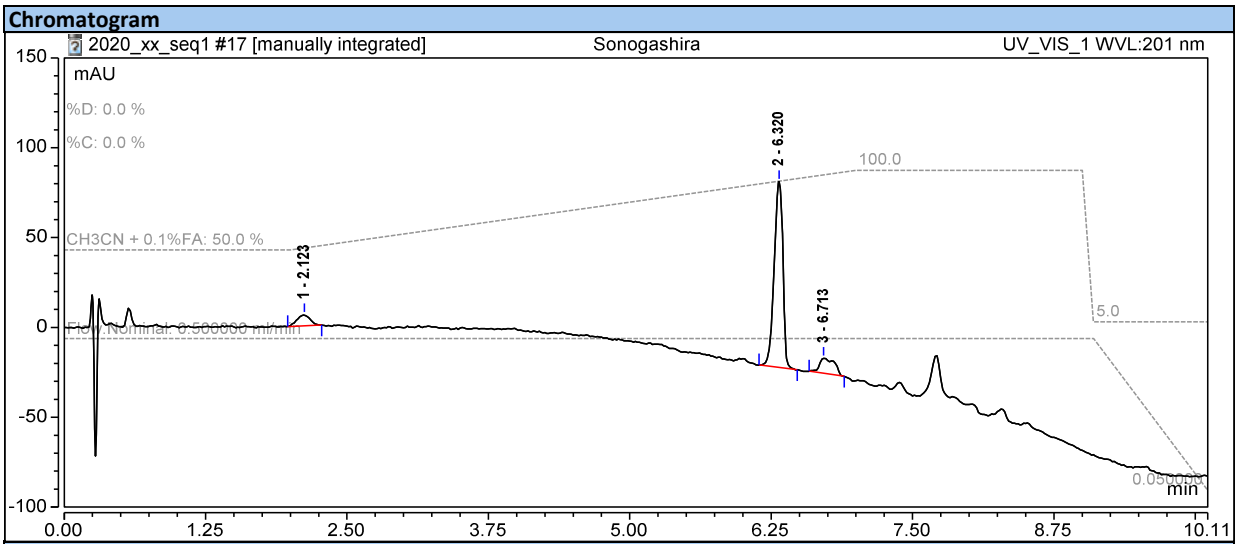
Compound **30**:

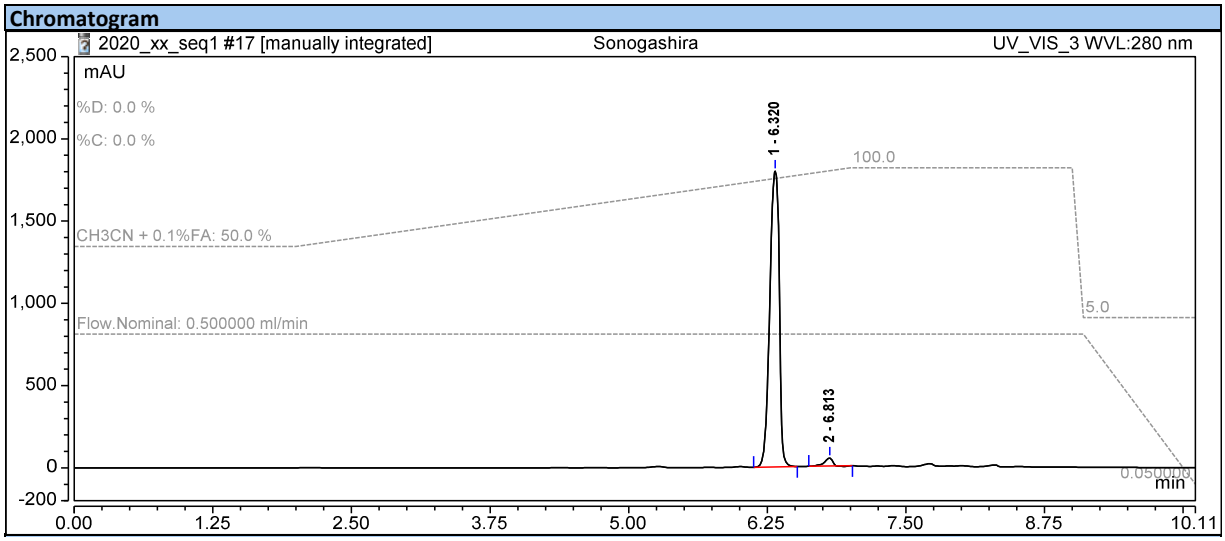


3,6-Dibromonaphthalene-2,7-diyl *bis*(trifluoromethanesulfonate) (**29**) (1.0 g, 1.72 mmol, 1 equiv.), CuI (3.9 mg, 20.6 μmol, 1.2 mol%), $Pd(PPh_3)_2Cl_2$ (28.9 mg, 41.2 μmol, 2.4 mol%) and *N*-Boc-propargylamine (1.280 g, 8.246 mmol, 4.8 equiv.) were introduced in a dry Schlenk tube under argon. Dry and degassed diethylamine (8 mL) was added with a syringe and the solution was stirred at 50 °C overnight. After which, the diethylamine was removed under vacuum and the residue was adsorbed on Dicalite (dry-load, 24 g). The crude was purified by automated flash chromatography using an 80 g FlashPure irregular silica column (Büchi) (from 10% to 50% of EtOAc in heptane in 50 min and flowrate of 45 mL/min). The desired compound crystallized out of the collection tubes after an overnight. The supernatants containing the pure desired compound and the crystals were collected and pooled together and the solvent was removed under vacuo. The residue was then dissolved in hot EtOAc/heptane (150 mL) and **30** was recrystallized by leaving the mixture at -19 °C overnight. The obtained crystals of product were collected and the filtrate was concentrated until more crystals were formed. In total, 390 mg (0.53 mmol, 31% chemical yield) of white solid product were obtained. ^{1}H NMR data were in agreement with literature[118].

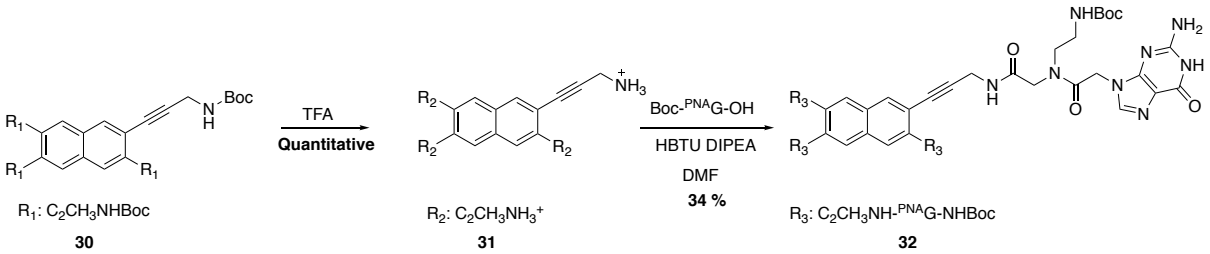
^{1}H NMR (500 MHz, Chloroform-d) δ 7.74 (s, 4H, Ar-H), 5.28 (br s, 4H, N-H), 4.23 (s, 8H, CH_2), 1.47 (s, 36H, *t*-Bu).

HPLC: (from 50% to 100% CH_3CN in 5 minutes)





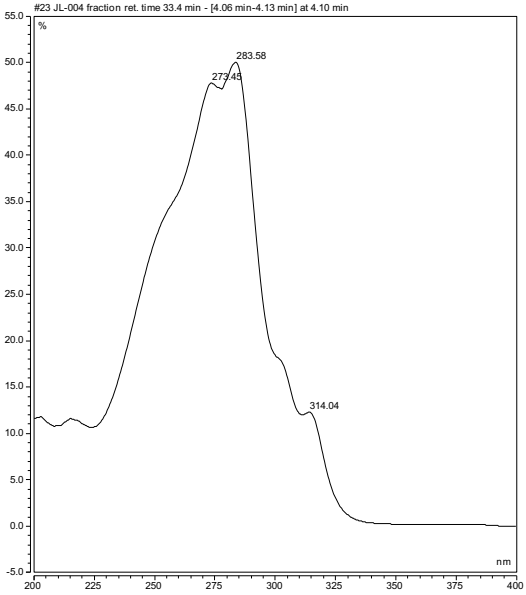
Compound **32**:



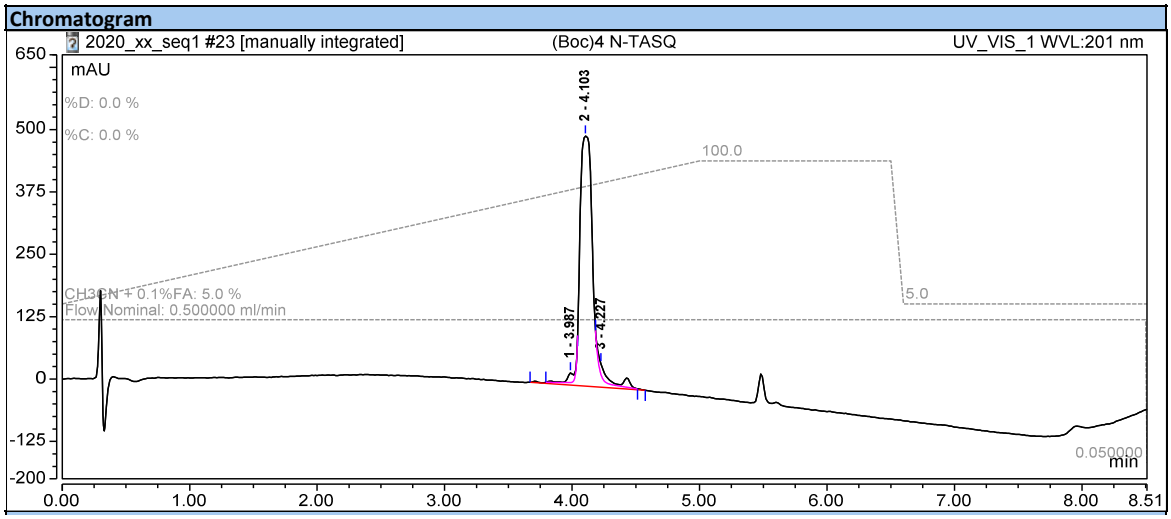
Compound **30** (51.1 mg, 0.07 mmol, 1 equiv) was dissolved in TFA (200 μL) and complete deprotection was assessed by HPLC-MS. Boc-^{PNA}G (124.3 mg, 0.31 mmol, 4.4 equiv) was mixed with HBTU (115.1 mg, 0.31 mmol, 4.4 equiv) and DIPEA (318 μL, 1.82 mmol) in DMF (1 mL) in separated round bottom flask and let to stir for 5 minutes. After TFA evaporation, **31** was dissolved in DMF and DIPEA was also added until neutralization, this solution was added to the latter and the reaction was left to stir for four hours at RT, until HPLC-MS showed complete conversion. The solvent was evaporated and the residue was purified by semi-preparative RP-HPLC (gradient of 5 to 15% over 5 minutes then from 15 to 65% over 50 minutes, r.t: 35 minutes). After lyophilization, **32** was obtained as a solid (44.5 mg, 0.024 mmol, 34% chemical yield).

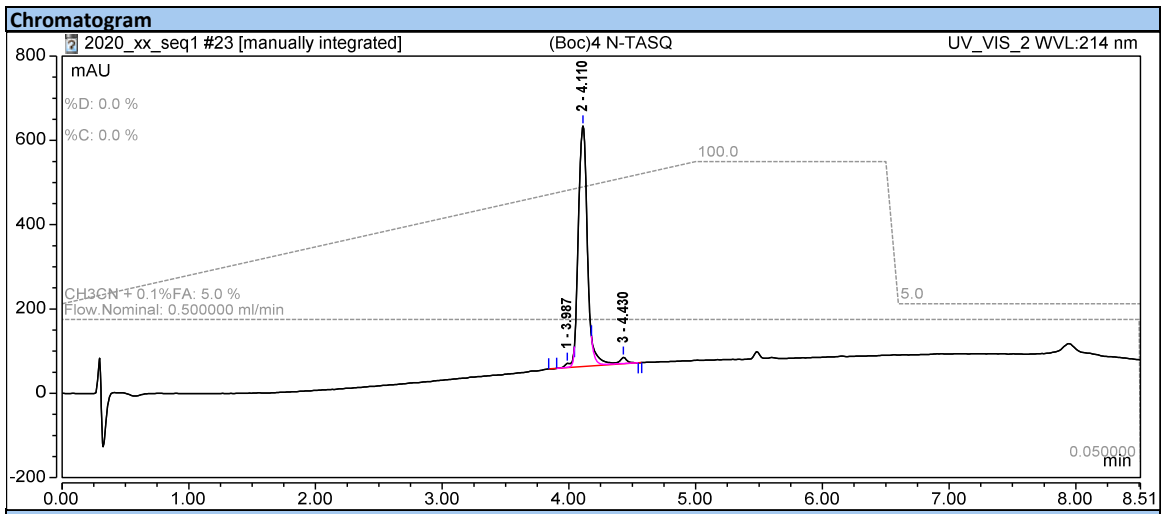
ESI-HRMS $[M+3H]^{3+}$ *m/z* = 635.94393 (calculated for $C_{86}H_{107}N_{32}O_{20}^{3+}$: 635.94410).

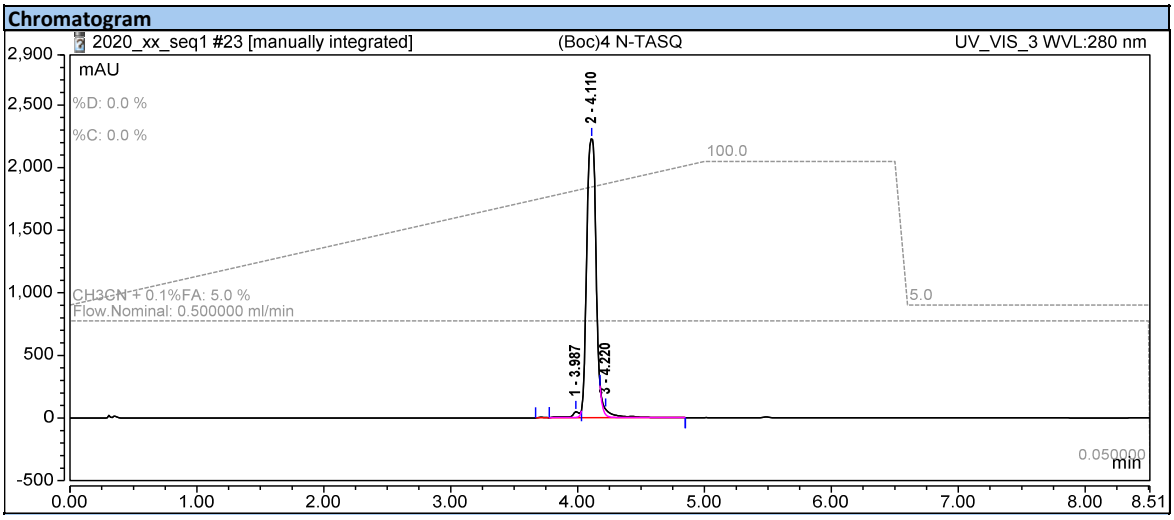
UV-Vis:



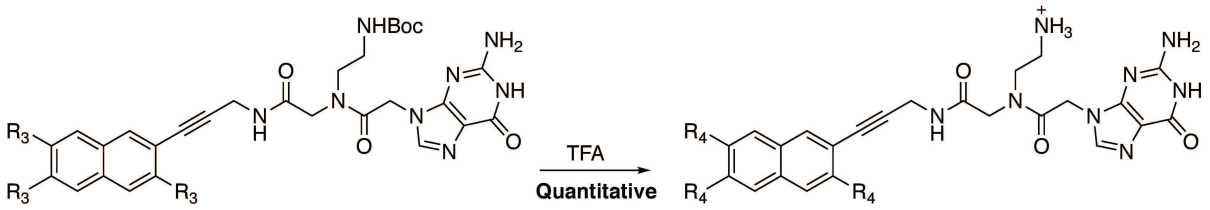
HPLC:


Chromatogram
2020_xx_seq1 #23 [manually integrated]
(Boc)4 N-TASQ
UV_VIS_1 WVL:201 nm
mAU
%D: 0.0 %
%C: 0.0 %
2 - 4.103
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.987
3 - 4.227
0.050000
min


Chromatogram
2020_xx_seq1 #23 [manually integrated]
(Boc)4 N-TASQ
UV_VIS_2 WVL:214 nm
mAU
%D: 0.0 %
%C: 0.0 %
2 - 4.110
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.987
3 - 4.430
0.050000
min


Chromatogram
2020_xx_seq1 #23 [manually integrated]
(Boc)4 N-TASQ
UV_VIS_3 WVL:280 nm
mAU
%D: 0.0 %
%C: 0.0 %
2 - 4.110
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 3.987
3 - 4.220
0.050000
min

N-TASQ:



R_3: C_2CH_3NH-^{PNA}G-NHBoc

32

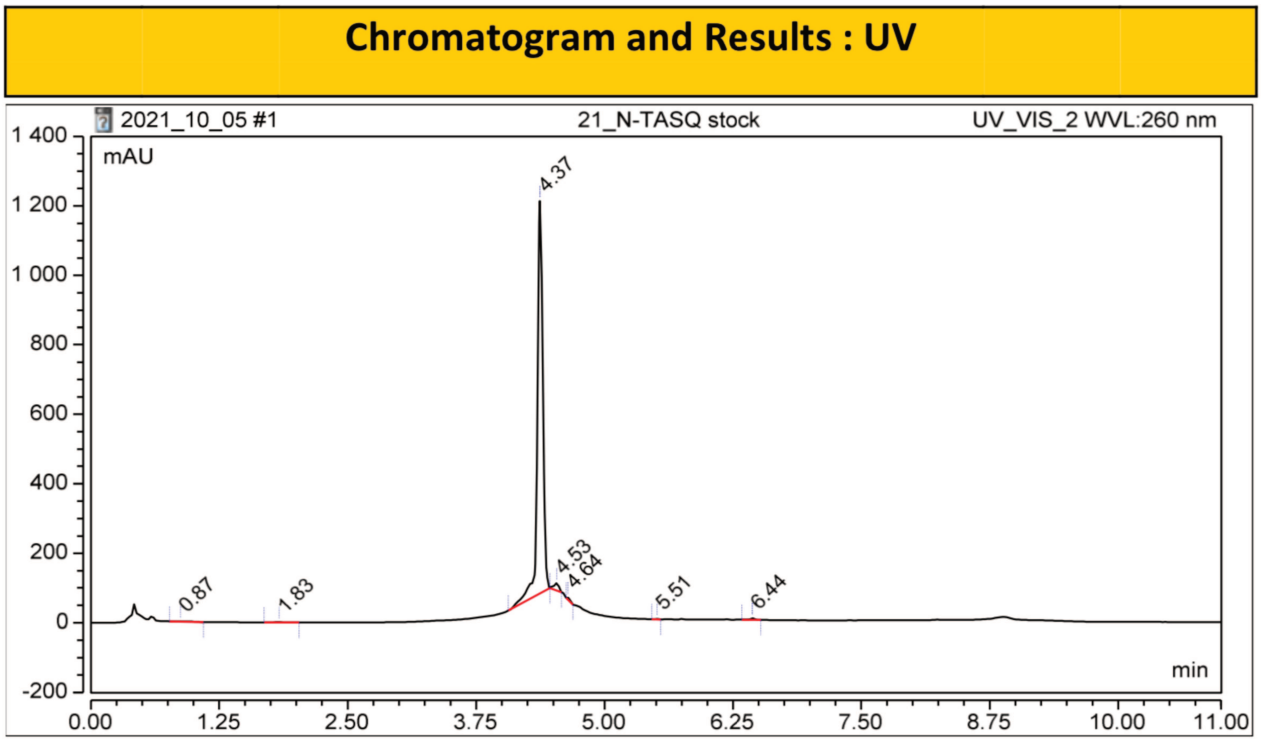
R_4: C_2CH_3NH-^{PNA}G-NH_3^+

N-TASQ

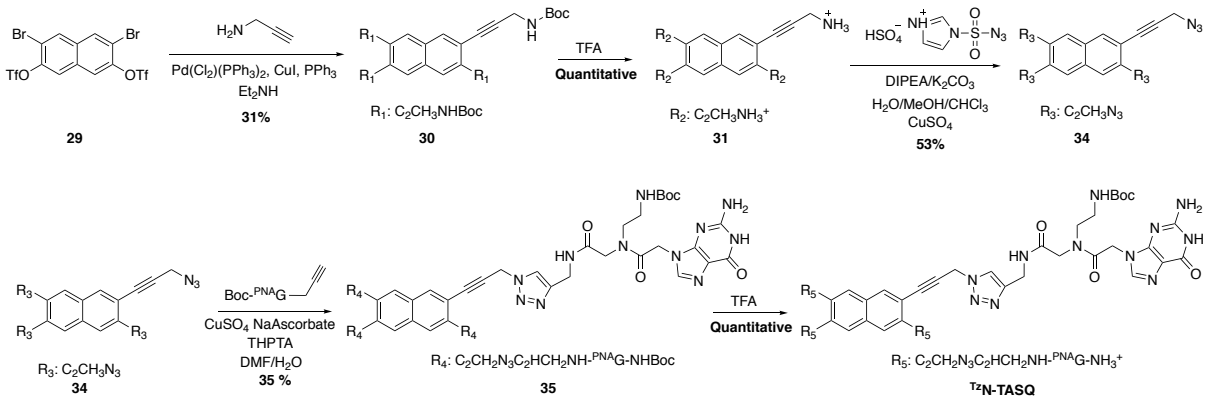
32 (1.95 mg, 1.0 µmol) was dissolved in TFA (200 µL) and the complete deprotection was assessed *via* HPLC, the mixture was diluted in water and the compound was freeze-dried to afford **N-TASQ**. (2 mg, 1.0 µmol, quantitative).

ESI-LRMS: $[M+2H]^{2+}$: *m/z* = 753.4 (calculated for $C_{66}H_{74}N_{32}O_{12}^{2+}$: 753.3).

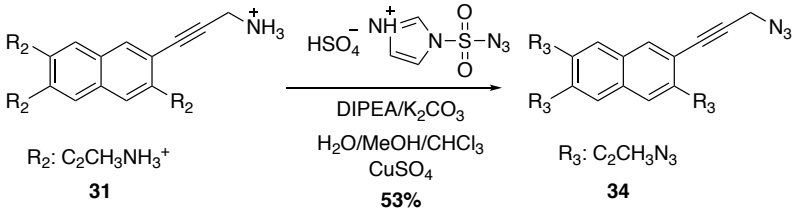
HPLC:



^{Tz}N-TASQ:



Compound **34**:



Compound **31** (50.4 mg, 0.067 mmol, 1 equiv.) was dissolved in deionized water (2 mL) and the pH was adjusted to 10 by the addition of DIPEA. Imidazolesulfonylazide (sulfate salt, 110 mg, 0.41 mmol, 6 equiv) was dissolved in a separate vessel in MeOH (1 mL) and DIPEA was slowly added with stirring and cooling in ice until the pH reached 10. This solution was then added to the solution of **31**, followed by the addition of $CuSO_4 \cdot 5H_2O$ (0.67 mg, 2.7 µmol, 4 mol%) dissolved in deionized water (30 µL). To the water/ methanol mixture, $CHCl_3$ (4 mL) was added to the mixture and the pH was further basified (pH 11 to 12) by addition of 10% aqueous K_2CO_3. The biphasic mixture was then stirred at maximum RPM for 2 hours, resulting in orange coloration of the chloroform layer. The layers were then separated, the aqueous layer was extracted with $CHCl_3$ (3 × 15 mL) and the combined organic phases were dried with sodium sulfate. After filtration of the drying agent, the solution was diluted with a minimal amount of pentane (10 mL) and passed through a a silica gel short-pass column chromatography. The silica was then washed with a 1/1 mixture of DCM and pentane (100 mL) to eluted the desired compound. The collected filtrate was concentrated under

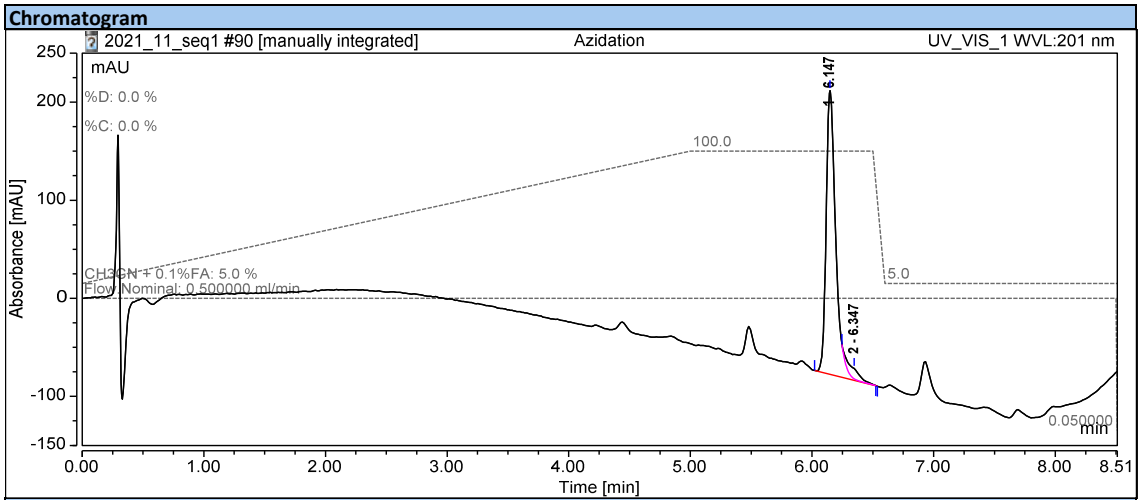
vacuum and dried *in vacuo*, affording the product (16.0 mg, 0.036 mmol, 53% chemical yield) as a white crystalline solid.

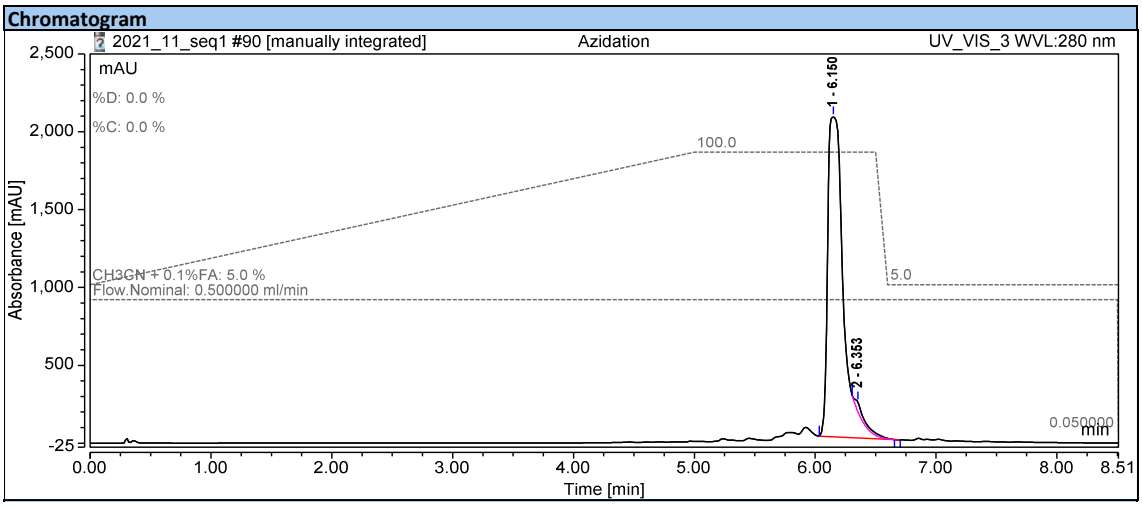
^{1}H NMR: (500 MHz, Chloroform-*d*) δ 7.93 (s, 4H), 4.24 (s, 8H).

^{13}C NMR: (126 MHz, Chloroform-*d*) δ 132.2, 131.4, 122.9, 86.3, 85.3, 40.8.

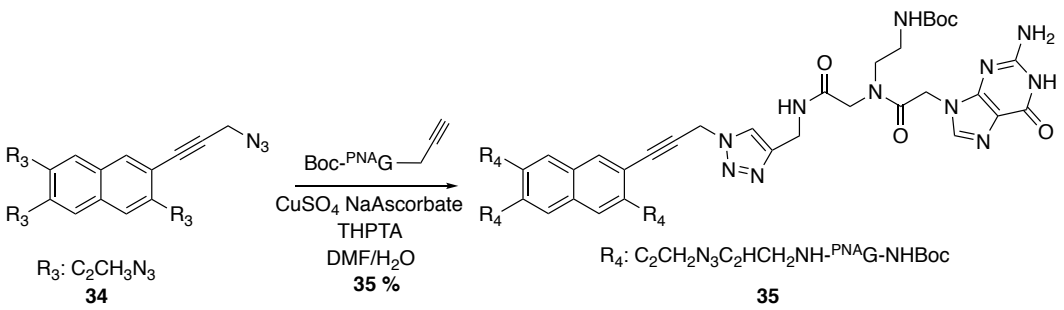
ESI-HRMS: impossible to obtain due to the low ionization of the compound **34**.

HPLC:





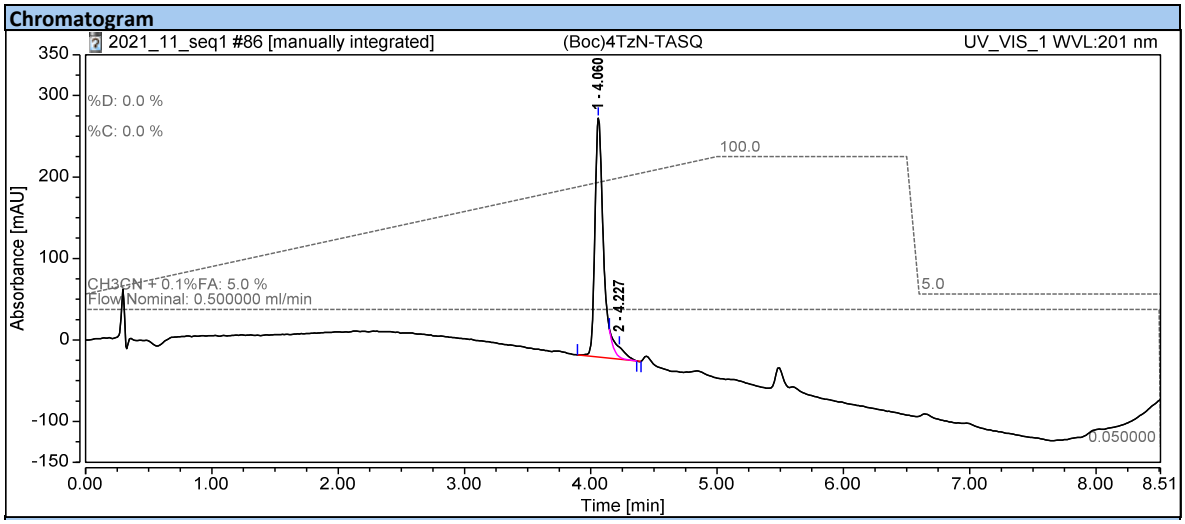
Compound **35**:

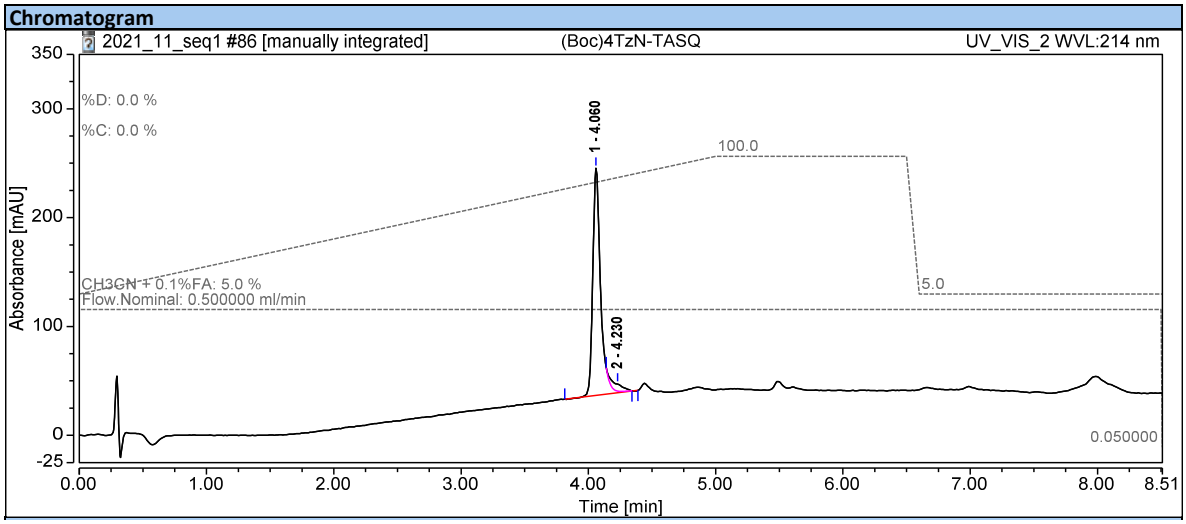


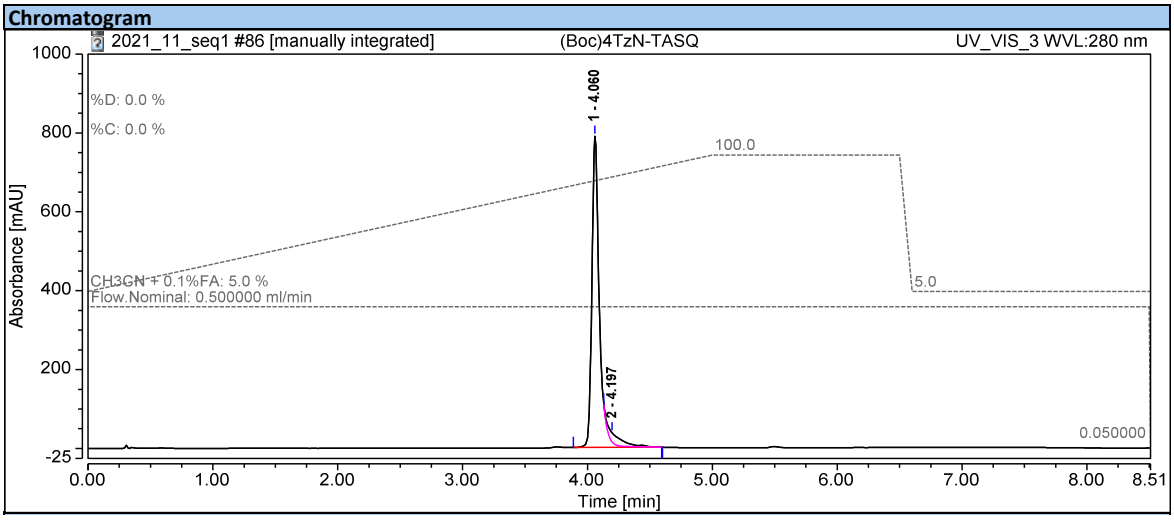
Compound **34** (11.7 mg, 0.026 mmol, 1 equiv.) was dissolved in anhydrous DMF (0.9 mL) and the alkynylated guanine (51.7 mg, 0.12 mmol, 4.6 equiv.) was added. Three aqueous solutions were prepared separately: $CuSO_4 \cdot 5H_2O$ (1.31 mg, 5.26 µmol, 0.2 equiv., 33 µL), THPTA (4.56 mg, 10.5 µmol, 0.4 equiv., 33 µL), and sodium ascorbate (2.08 mg, 10.5 µmol, 0.4 equiv., 33 µL). To the Cu(II) solution was added the solution containing the copper ligand (THPTA), resulting in deep blue complex. To this mixture was then added the solution of sodium ascorbate, producing a colourless Cu(I)·THPTA aqueous solution. The prepared solution was added to the stirred DMF solution of **34,** the reaction was monitored every 30 minutes and 0.2 equiv. of fresh catalyst-complex mixture (99 µL) were added if the HPLC analysis showed incomplete conversion. After two additions of catalyst, an additional 1 equiv. of alkyne (11.8 mg, 26.3 µmol, 1.0 equiv.) in DMF (100 µL) was added in to further improve the conversion of **34** in 15 min. After 20 minutes the conversion was stil partial and therefore another portion of catalyst was added as above, together with another aliquot of the alkyne (11.8 mg, 26.3 µmol, 1.0 equiv.) in DMF (100 µL). The conversion of starting compound **34** was complete after 15 minutes (monitored by HPLC). The reaction mixture was then purified by preparative HPLC (gradient of 5 to 15% over 5 minutes then from 15 to 65% over 50 minutes, r.t: 41minutes). Compound **35** (20.2 mg, 9 µmol, 35%) was obtained as a white solid after lyophilisation of the solvents.

ESI-HRMS: $[M+2H]^{2+}$ *m/z* = 1115.47820 (calculated for $C_{98}H_{118}N_{44}O_{20}^{2+}$: 1115.47790).

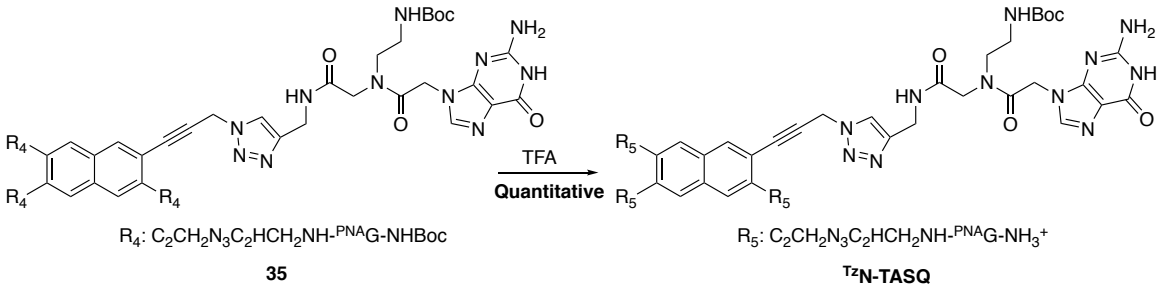
HPLC:


Chromatogram
2021_11_seq1 #86 [manually integrated]
(Boc)4TzN-TASQ
UV_VIS_1 WVL:201 nm
%D: 0.0 %
%C: 0.0 %
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 4.060
2 - 4.227
0.050000
Absorbance [mAU]
Time [min]


Chromatogram
2021_11_seq1 #86 [manually integrated]
(Boc)4TzN-TASQ
UV_VIS_2 WVL:214 nm
%D: 0.0 %
%C: 0.0 %
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 4.060
2 - 4.230
0.050000
Absorbance [mAU]
Time [min]


Chromatogram
2021_11_seq1 #86 [manually integrated]
(Boc)4TzN-TASQ
UV_VIS_3 WVL:280 nm
%D: 0.0 %
%C: 0.0 %
100.0
5.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 4.060
2 - 4.197
0.050000
Absorbance [mAU]
Time [min]

^{Tz}N-TASQ:

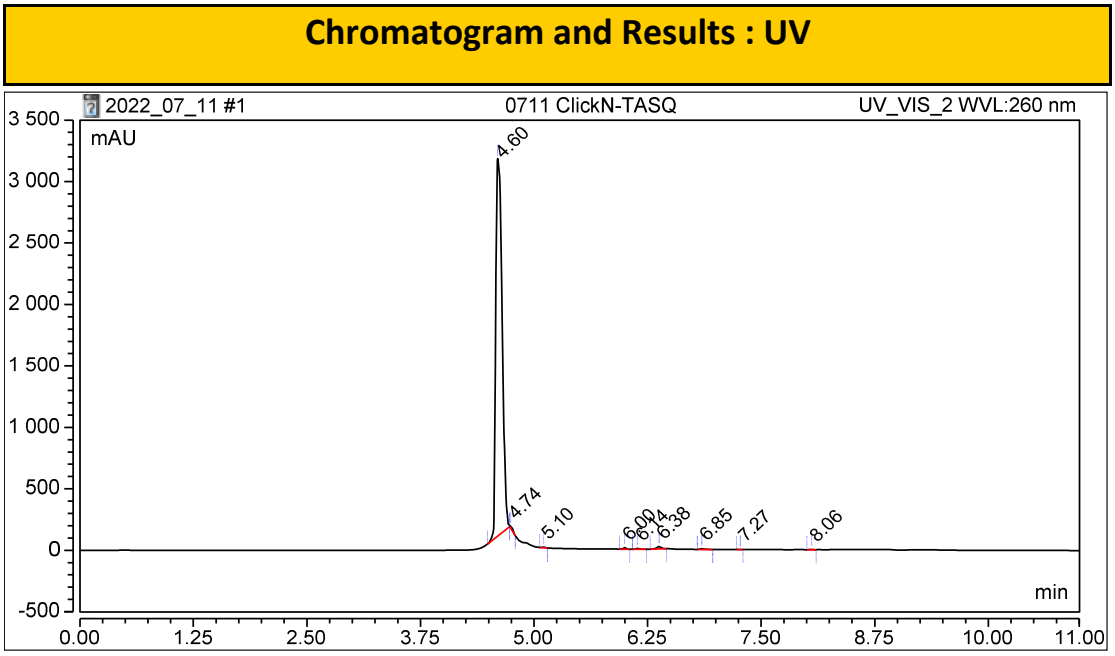


35 **^{Tz}N-TASQ**

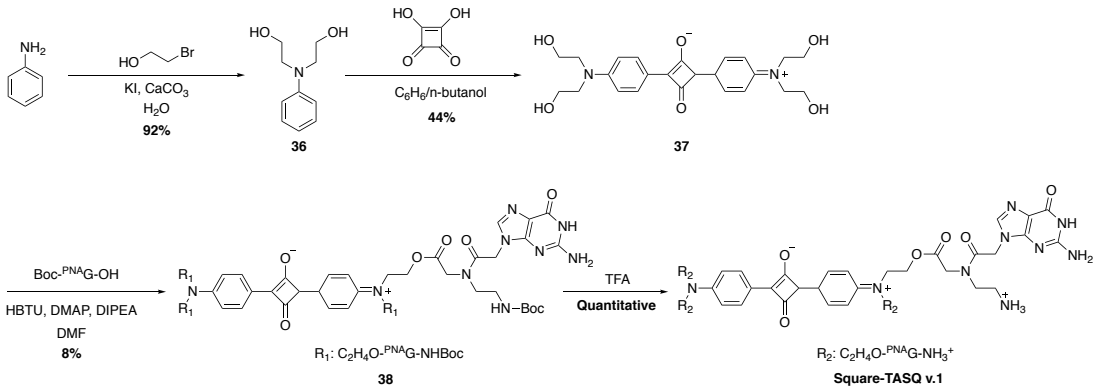
Compound **35** (1.95 mg, 0.9 µmol) was dissolved in TFA (200uL) and the complete deprotection was assessed *via* HPLC, the mixture was diluted in water and the compound was freeze dried to achieve **^{Tz}N-TASQ**. (2 mg, 0.9 µmol, 100%).

ESI-HRMS: $[M+4H]^{4+}$ m/z = 458.19015 (calculated for $C_{78}H_{88}N_{44}O_{12}^{4+}$: 458.19016).

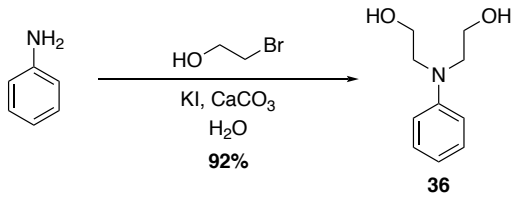
HPLC:



Square-TASQ V_1:



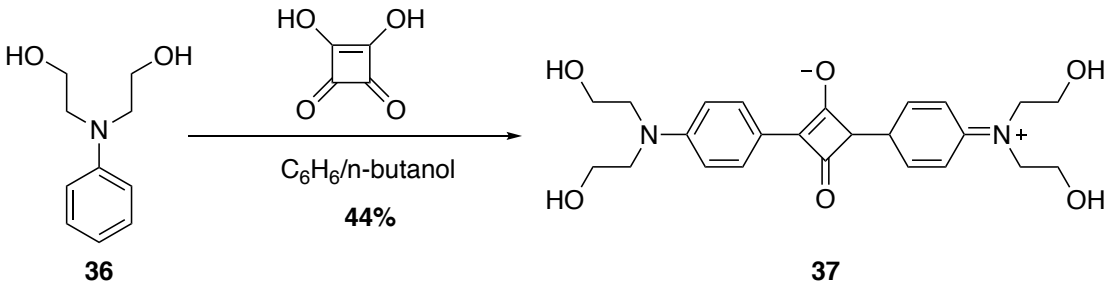
Compound **36**:



Aniline (2.7 mL, 30 mmol, 1 equiv.), 2-bromoethanol (6.4 mL, 90 mmol, 3 equiv.), calcium carbonate (6.0 g, 60 mmol, 2.0 equiv.) and potassium iodide (0.50 g, 3 mmol, 0.1 equiv.) in water (42 mL) were stirred at reflux overnight (reaction completion was checked by LC-MS and TLC). The reaction mixture was filtered, and the precipitate was washed with small amount of ethyl acetate. The two layers were separated in a separatory funnel. The aqueous layer was extracted with ethyl acetate (3 × 50 mL) and the combined organic layers were washed with brine (50 mL), dried with sodium sulfate, and filtered. The solvent was removed under vacuum, the residue was deposited on Dicalite (50 g) and the product was purified by flash chromatography (330 g FlashPure irregular SiO_2 column eluted with mobile phase 2 to 6 % MeOH/ DCM at a flow of 100 mL/min). Compound **36** (5.02 g, 92%) was obtained as a brownish oil. ^{1}H NMR data were in agreement with literature[192].

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.26 – 7.21 (m, 2H), 6.76 – 6.72 (m, 1H), 6.72 – 6.68 (m, 2H), 3.84 (t, *J* = 4.9 Hz, 4H), 3.57 (t, *J* = 4.9 Hz, 4H), 3.41 (br s, 2H).

Compound **37**:

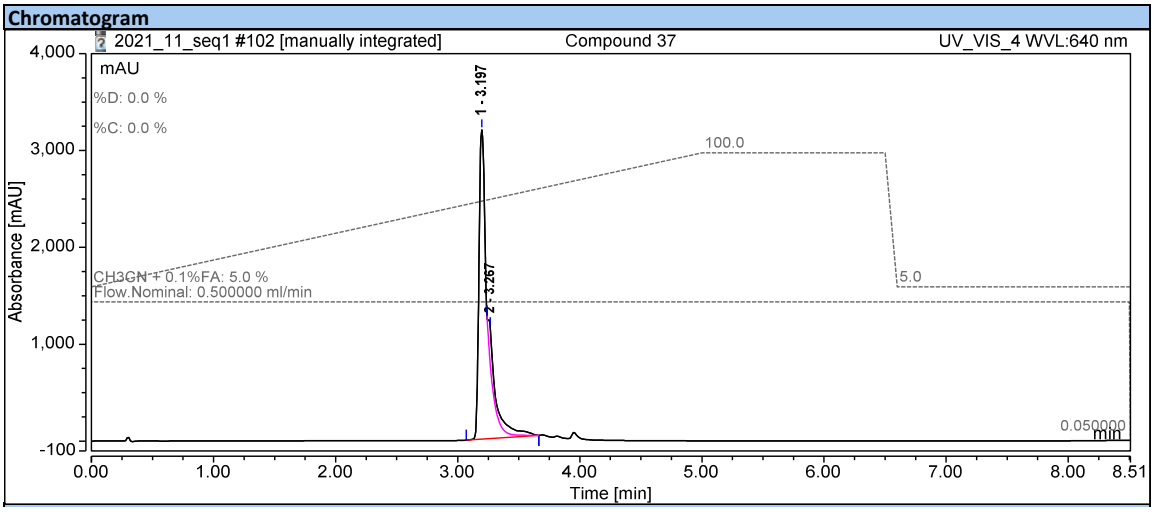


36 **37**

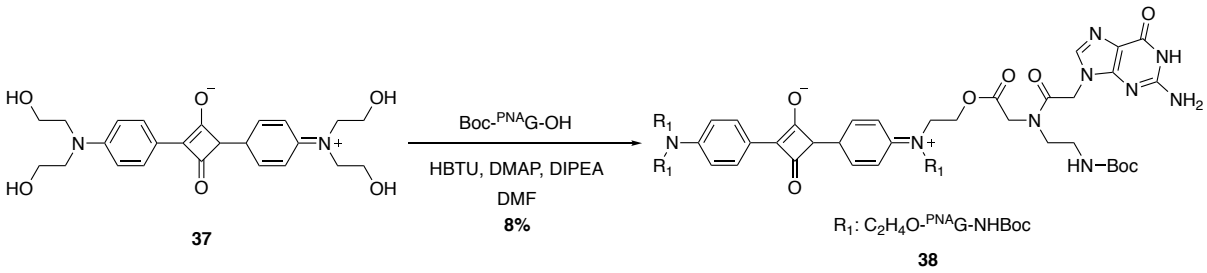
N,*N*-Bis(2-hydroxyethyl)aniline (**36**) (1.81 g, 10 mmol, 2 equiv.) was dissolved in a mixture of benzene and *n*-butanol (*V*/*V* = 1/1, 40 mL) and squaric acid (0.57 g, 5 mmol, 1 equiv.) was added. The mixture was stirred under reflux in Dean-Stark apparatus for azeotropic removal of water for 14 hours. The mixture was left to cool to room temperature and the solid dark green compound **37** (968 mg, 2.2 mmol, 44%) was filtered off and washed with pentane. ^{1}H NMR data were in agreement with literature [193].

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.10 (d, *J* = 9.2 Hz, 4H, Ar-H), 7.01 (d, *J* = 9.4 Hz, 4H, Ar-H), 4.88 (br s, 4H, OH), 3.71 – 3.60 (m, 16H, CH_2).

HPLC:



Compound **38**:

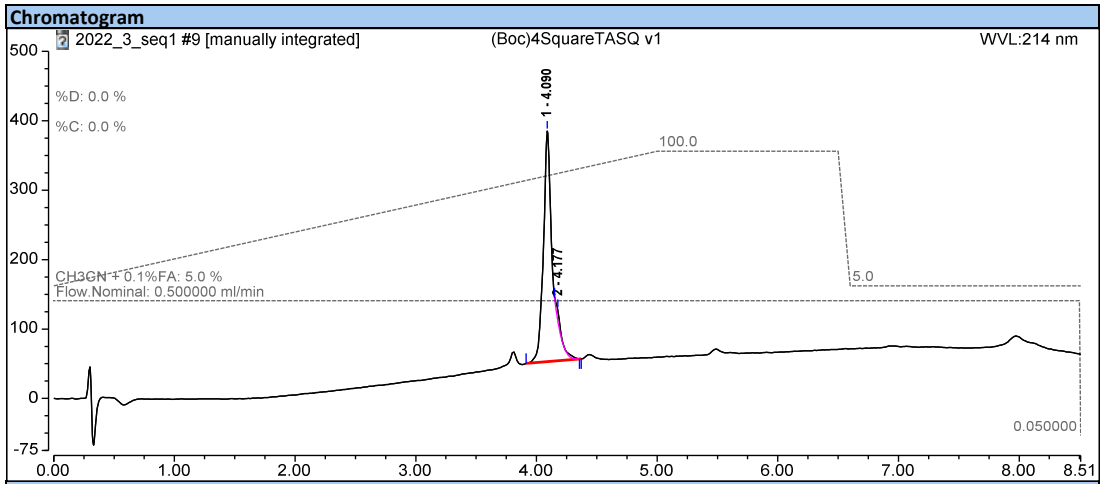


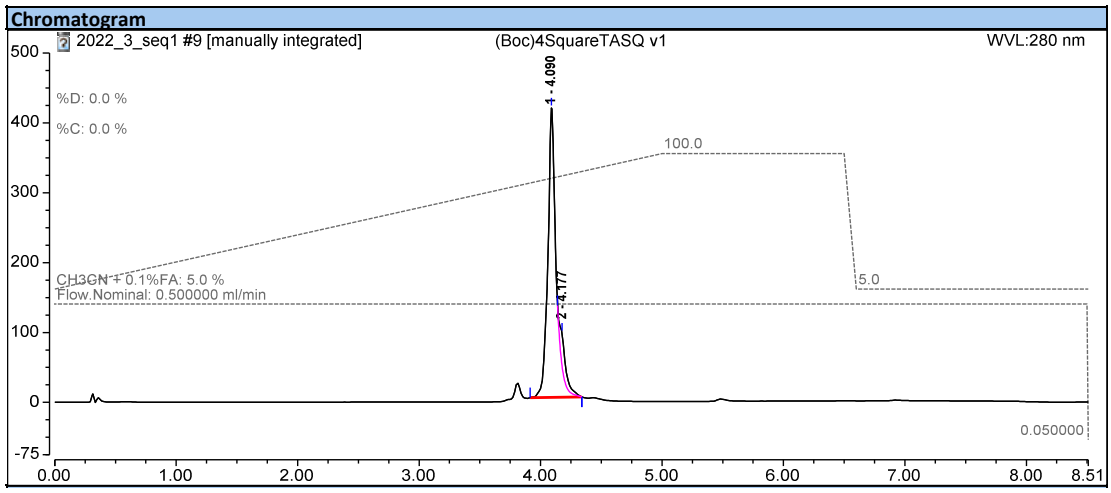
37 **38**

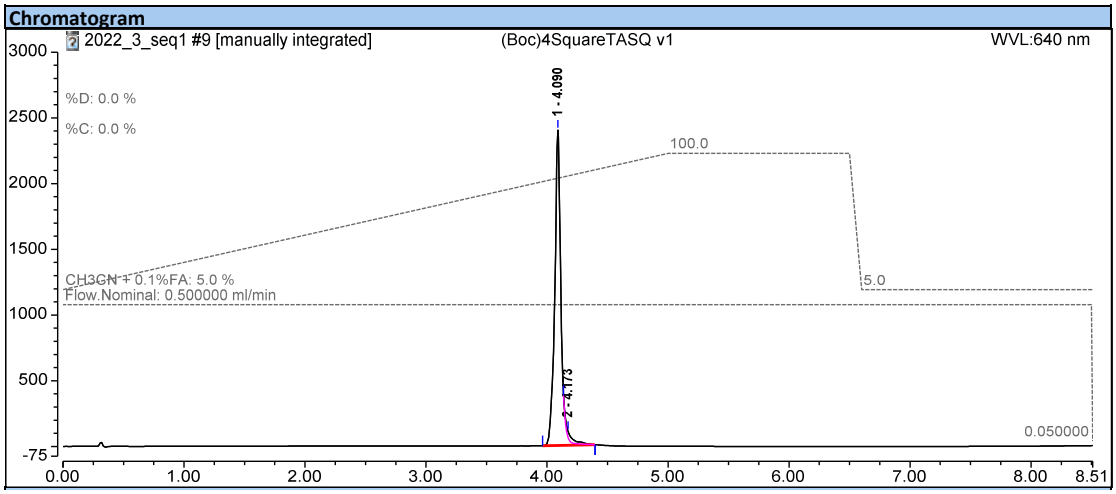
Boc-^{PNA}G (186 mg, 0.45 mmol, 4 equiv.) and HBTU (172 mg, 0.45 mmol, 4 equiv.) were dissolved in dry DMF (1 mL) in a dry flask under argon atmosphere and DIPEA (159 µL, 0.91 mmol, 8 equiv.) was added. This mixture was stirred for 30 minutes at room temperature. Then, the squaraine **37** (50 mg, 0.12 mmol, 1 equiv.) was dissolved in dry DMF (1.2 mL) and added dropwise. A further aliquot of HBTU (172 mg, 0.45 mmol, 4 equiv.) and DIPEA (80 µL, 0.45 mmol, 4 equiv.) was then added. DMAP (13.9 mg, 0.15 mmol, 1 equiv.) was added to improve the conversion. After 30 minutes, another equivalent of Boc-^{PNA}G (46.5 mg, 0.11mmol, 1 equiv.) was added together with HBTU (43.0 mg, 0.11 mmol, 1 equiv.) and DIPEA (40 µL, 0.23 mmol, 2 equiv.). This resulted in complete conversion of the starting squaraine dye according to HPLC (3h). The reaction mixture was purified by semi-preparative HPLC (gradient of 5 to 15% over 5 minutes then from 15 to 65% over 50 minutes, r.t: 40 minutes) and the fractions containing compound **38** were lyophilised, affording product (18.4 mg, 9.6 µmol, 8%) as a blue-green solid.

ESI-HRMS $[M+3H]^{3+}$ m/z = 669.28601 (calculated for $C_{88}H_{115}N_{30}O_{26}^{3+}$: 669.28608).

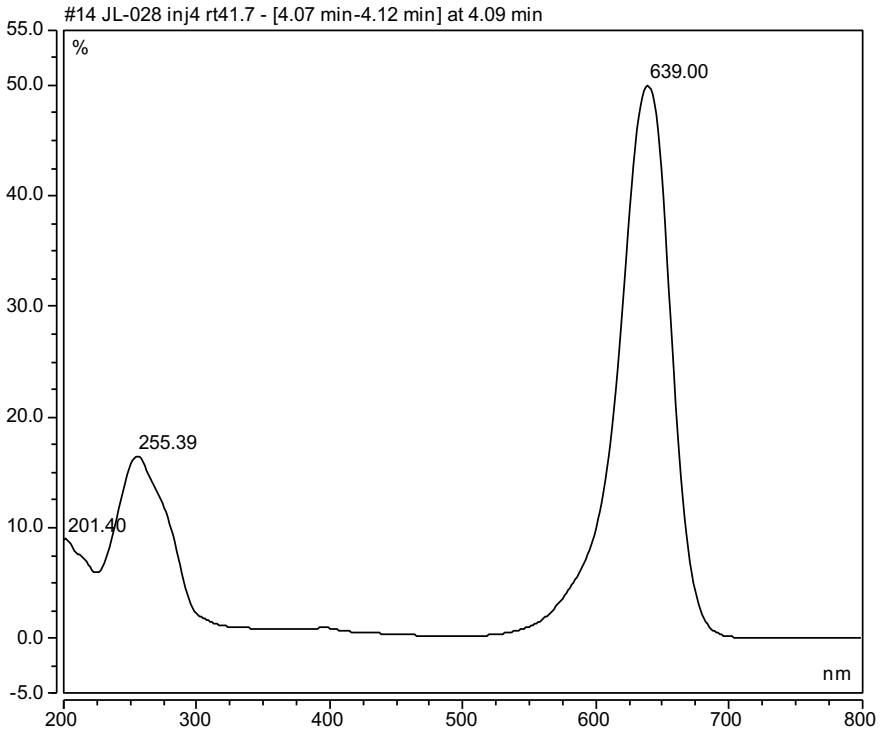
HPLC:


Chromatogram
2022_3_seq1 #9 [manually integrated]
(Boc)4SquareTASQ v1
WVL:214 nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 4.090
2 - 4.177
100.0
5.0
0.050000

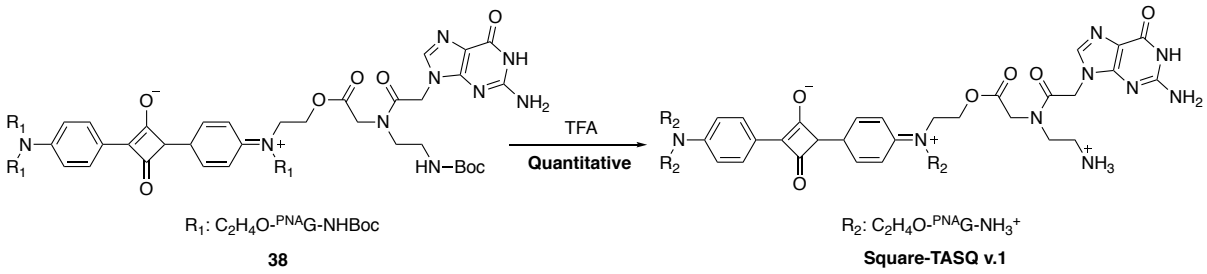

Chromatogram
2022_3_seq1 #9 [manually integrated]
(Boc)4SquareTASQ v1
WVL:280 nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 4.090
2 - 4.177
100.0
5.0
0.050000


Chromatogram
2022_3_seq1 #9 [manually integrated]
(Boc)4SquareTASQ v1
WVL:640 nm
%D: 0.0 %
%C: 0.0 %
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
1 - 4.090
2 - 4.173
100.0
5.0
0.050000

Uv-Vis Spectra:



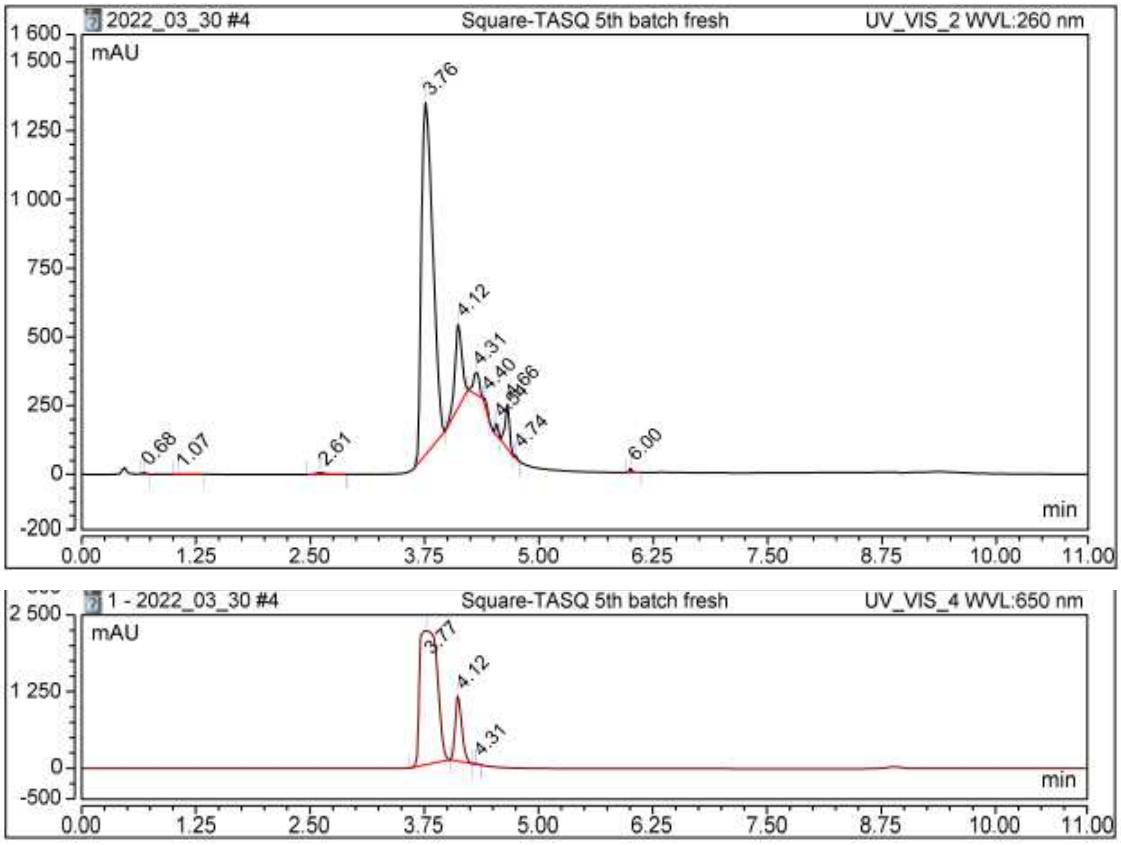
Square-TASQ v.1:



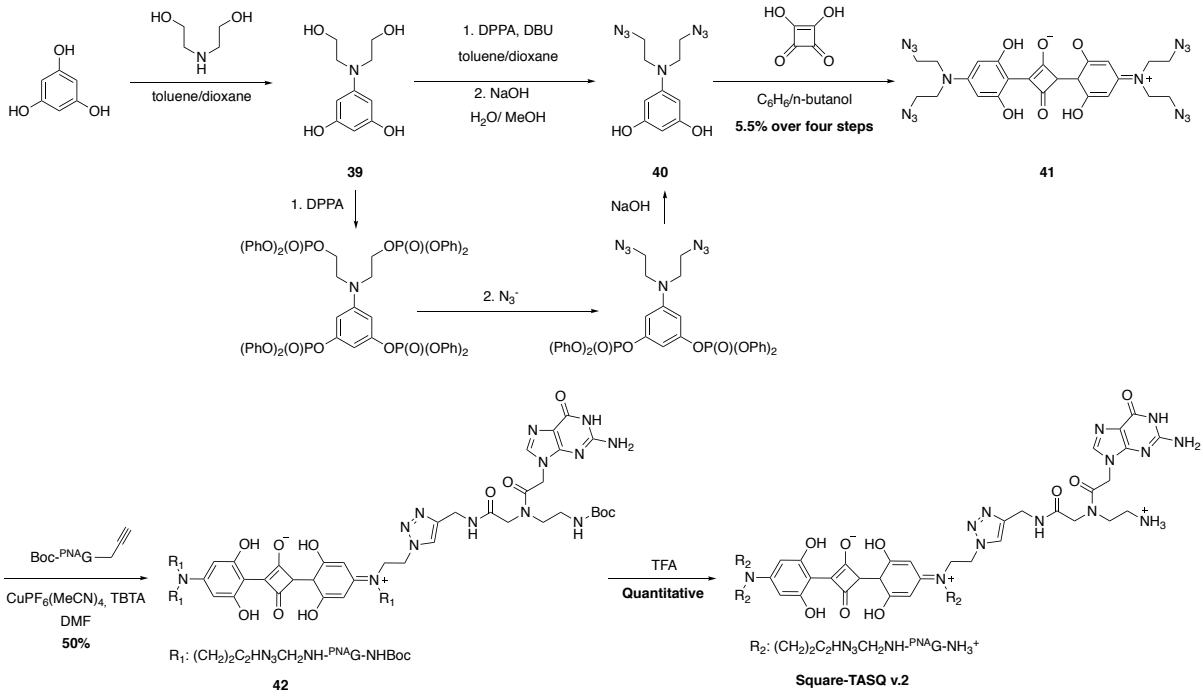
38 (1.95 mg, 0.97 µmol) was dissolved in TFA (200 µL) and the complete deprotection was monitored by HPLC. After 20 minutes, the mixture was diluted in water and the compound was freeze dried to achieve **Square-TASQ v_1** (2 mg, 0.97 µmol, 100%).

ESI-HRMS $[M+2H]^{2+}$ *m/z* = 803.32076 (calculated for $C_{68}H_{82}N_{30}O_{18}^{2+}$: 803.32062).

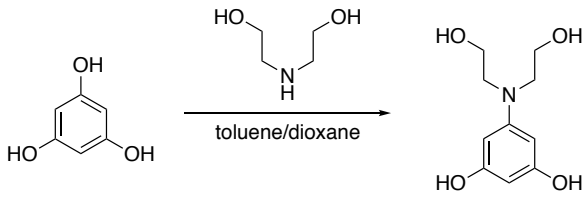
HPLC:



Square-TASQ V_2:



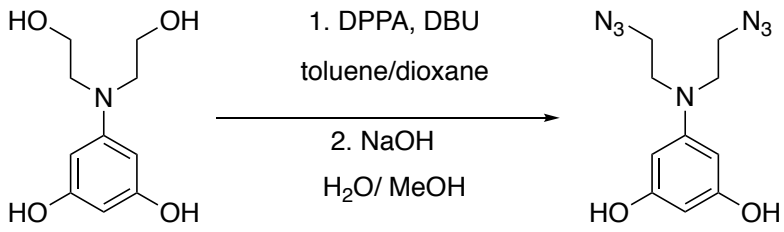
Compound **39**:



39

Diethanolamine (4.20 g, 40.0 mmol, 1.3 equiv.) and phloroglucinol (5.04 g, 31.1 mmol, 1 equiv.) were heated under reflux in a mixture of toluene and *n*-butanol (*V*/*V* = 1/1, 200 mL) overnight. The reaction mixture was then concentrated under vacuum and the resulting red oil partially solidified after drying *in vacuo*. The crude product (10.64 g) was used in the next step without further purification.

Compound **40**:

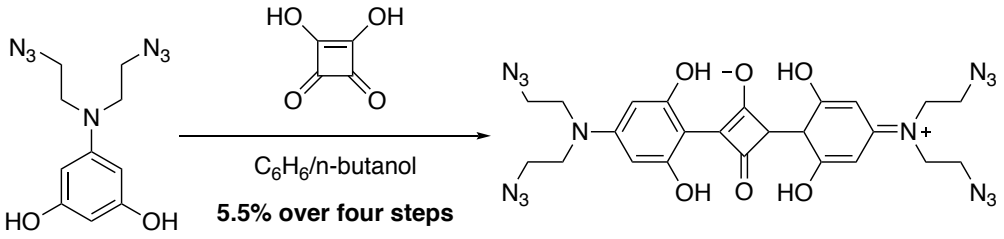


39 **40**

Crude 3,5-dihydroxy-*N*,*N*-bis(2-hydroxyethyl)aniline (**39**) (500 mg, *ca.* 2.4 mmol.) was suspended in toluene (10 mL) and dioxane (20 mL). DBU (1473 µL, 9.9 mmol, 4.1 equiv) was added, followed by dropwise addition of diphenylphosphoryl azide (DPPA) (2.2 mL, 10.3 mmol, 4.3 equiv). The reaction mixture was then heated under reflux for 3.5 hours under vigorous stirring, HPLC-MS indicated conversion of the starting material to bis(diphenyl phosphate) of the product. The reaction mixture was concentrated under vacuum and the residue was dissolved in DCM (50 mL) and Brine (50 mL). The organic layer was separated and the aqueous layer was extracted with DCM (3 × 50 mL). The combined organic extracts were dried with K_2CO_3, the drying agent was filtered off and the solvent removed under vacuum.

The residue was purified by column chromatography (neat DCM to 6 vol% MeOH in DCM as a mobile phase) to remove most of the polar impurities. The fractions containing the bis(diphenyl phosphate) ester of the product were concentrated under vacuum and the residue was dissolved in methanol (45 mL) and water (5 mL). The phosphate was hydrolysed by the addition of 10% aqueous sodium hydroxide (2.5 mL) and the solution was stirred until complete conversion (HPLC monitor, approx. 1 hour). Methanol was removed by evaporation and the remaining aqueous solution was extracted with DCM (2 × 50 mL). The aqueous layer was then neutralized with hydrochloric acid (1 M, aqueous) and extracted a second time with DCM (4 × 50 mL). The combined organic phases were dried with sodium sulfate, the drying agent was filtered off and the solvent was removed under vacuum. The desired product (155 mg) was obtained as a brown oil and used without further purification in the following step.

Compound **41**:



40 **41**

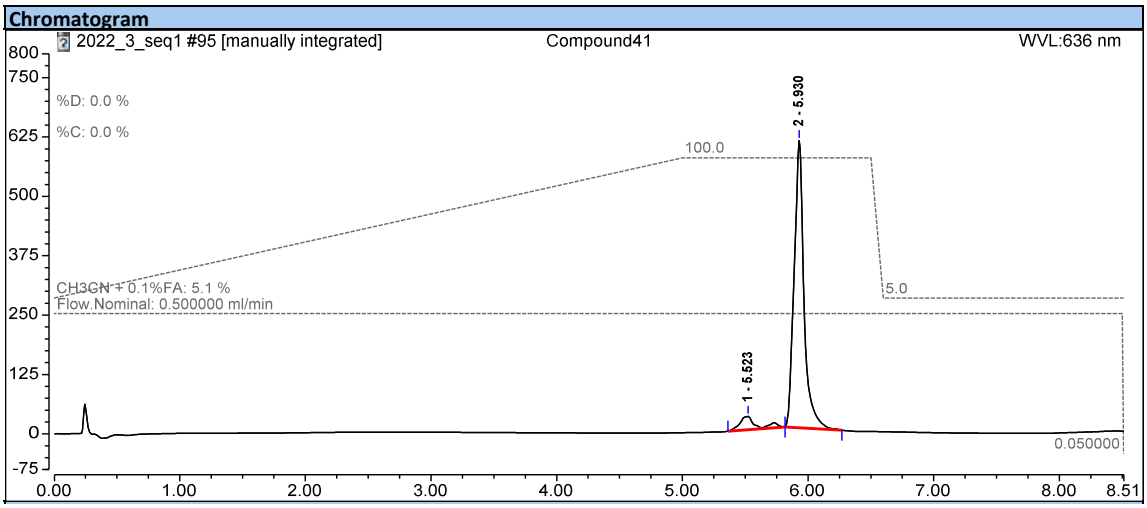
The crude containing compound **40** (155 mg, ca. 0.59 mmol, 2 equiv.) was dissolved in a mixture of toluene and *n*-butanol (*V*/*V* = 1/1, 40 mL) and heated under reflux with squaric acid (33.6 mg, 0.29 mmol, 1 equiv.) in a Dean-Stark apparatus for 6.5 hours. The solution was allowed to cool down to room temperature and the product was allowed to crystallize by leaving the crude mixture at 4 °C overnight. The solid was filtered off and washed with toluene, pentane and diethyl ether. The product was obtained as a solid (48.8 mg, 0.014 mmol, 5.5% from phloroglucinol).

^{1}H NMR (500 MHz, DMSO-d_6) δ 11.12 (s, 4H, OH), 6.01 (s, 4H, Ar-H), 3.71 (t, *J* = 6.1 Hz, 8H, CH_2), 3.59 (t, *J* = 5.9 Hz, 8H, CH_2).

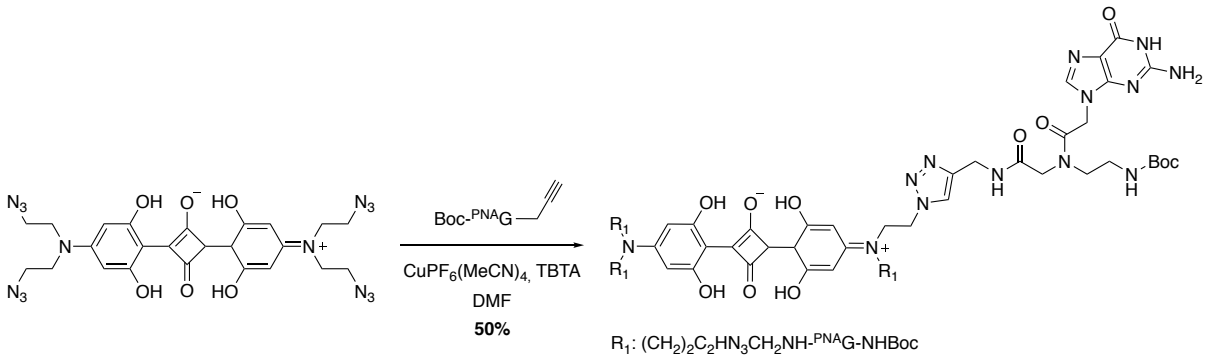
^{13}C NMR (126 MHz, DMSO-d_6) δ 180.7, 163.6, 162.3, 158.1, 102.4, 93.7, 49.4, 48.5.

ESI-HRMS $[M+H]^+$ m/z = 605.20768 (calculated for $C_{24}H_{25}N_{14}O_6^+$: 605.20760).

HPLC:



Compound **42**:



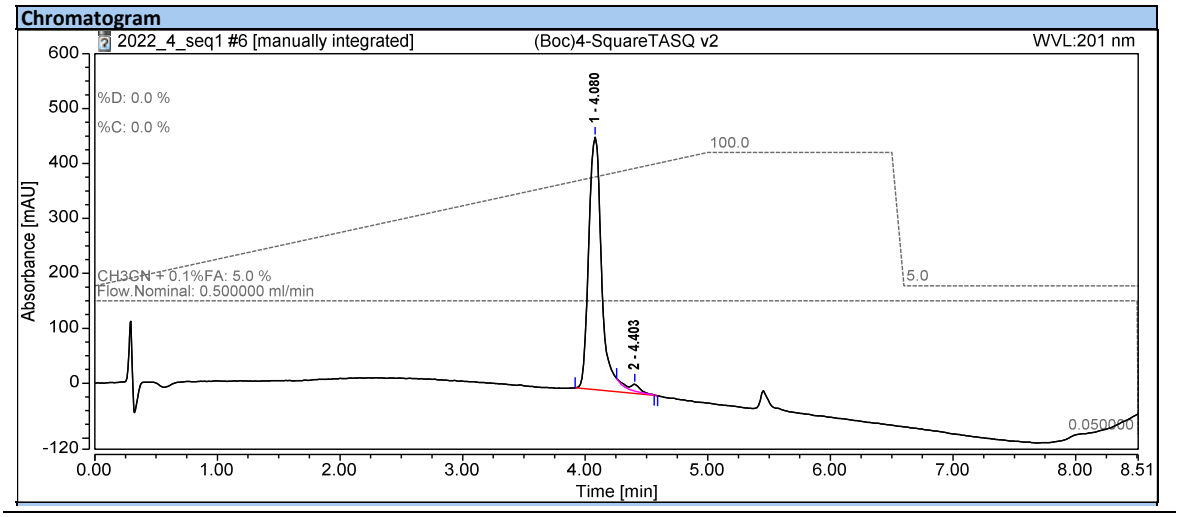
41 **42**

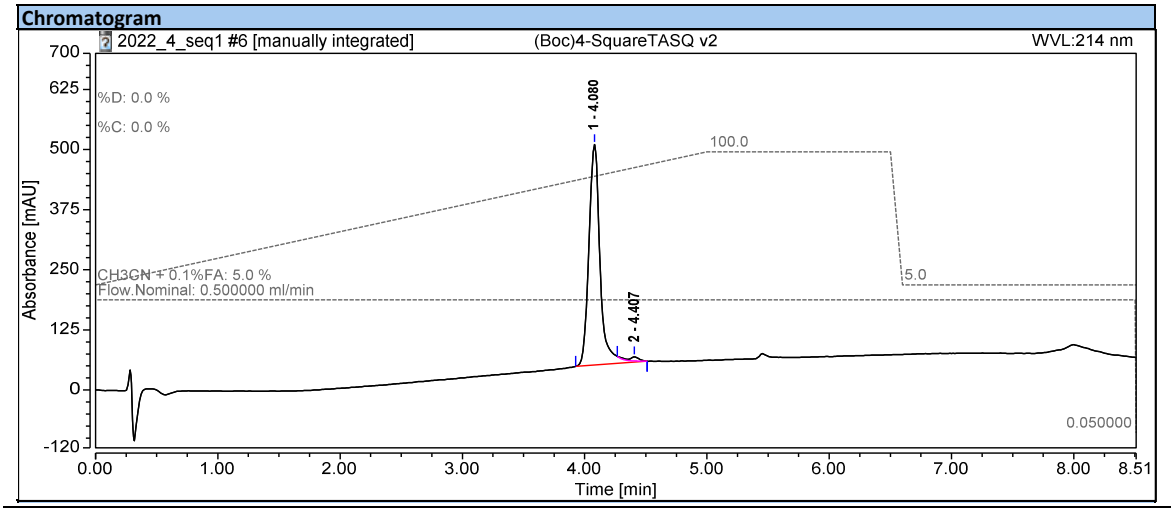
Compound **41** (26.0 mg, 43.0 µmol.) and the propargyl-PNA derivative previously described (84.5 mg, 189.2 µmol, 4.4 equiv.) were stirred in DMF (3 mL). As copper source in a separate vessel an aqueous solution of $CuSO_4 \cdot 5H_2O$ (2.1 mg, 8.6 µmol, 33 µL, 0.2 equiv.), THPTA (7.5 mg, 17.2 µmol, 33 µL, 0.4 equiv.) and sodium ascorbate (8.5 mg, 43.0 µmol, 33 µL, 1 equiv.) were added in that order. The Cu(I) containing solution was then added to the mixture containing compound **41** and the propargyl-PNA derivative and let to stir for 30 minutes. As the conversion was still low after 30 minutes according to HPLC, a solution of copper iodide

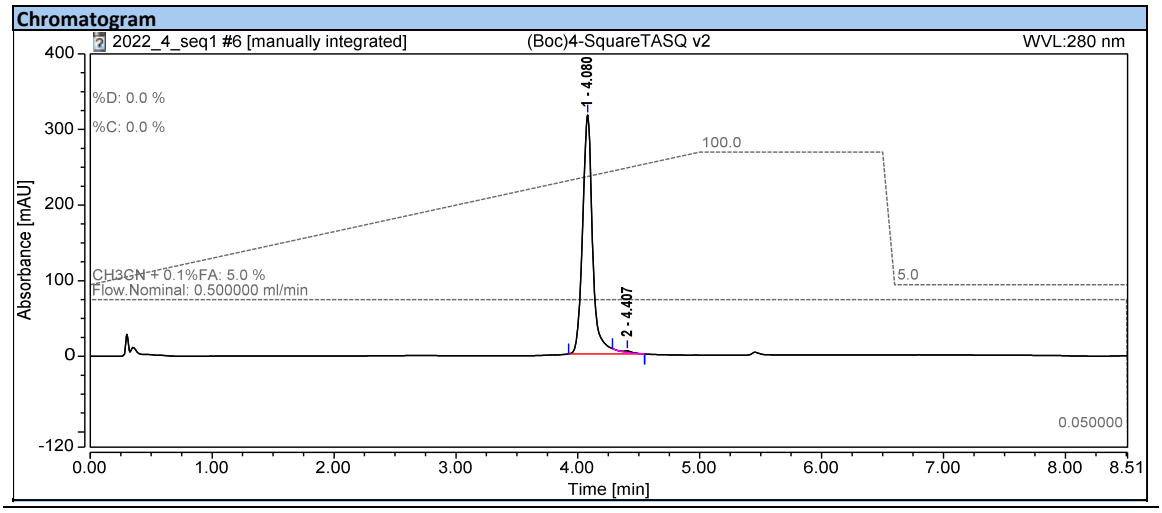
(3.3 mg, 17.3 µmol, 0.4 equiv.) and TBTA (18.3 mg, 34.5 µmol, 0.8 equiv.) in DMSO (0.5 mL) was added to improve the conversion of the starting material. After no improvement was observed, $Cu(MeCN)_4PF_6$ (6.4 mg, 17.2 µmol, 0.4 equiv.) was chosen as copper source, showing progression in the reaction conversion. After an additional 30 minutes a second addition of $Cu(MeCN)_4PF_6$ (6.4 mg, 17.2 µmol, 0.4 equiv.) resulted in complete conversion to the target compound. The reaction mixture was purified by semi-preparative HPLC (gradient of 5 to 15% over 5 minutes then from 15 to 65% over 50 minutes, r.t: 40 minutes) and the fractions containing product were lyophilised affording the product **42** (51.6 mg, 21.5 µmol, chemical yield 50%) as a blue solid.

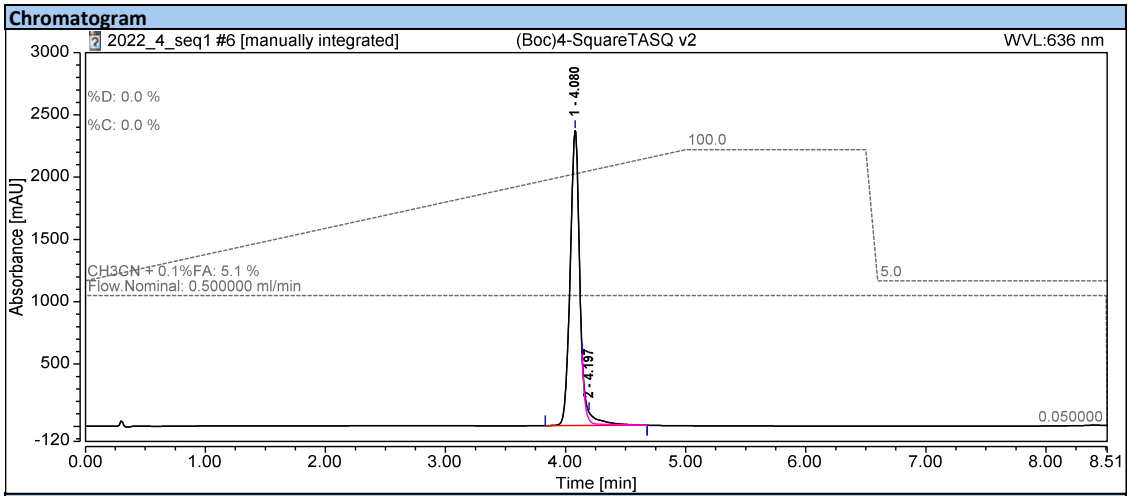
ESI-HRMS $[M+2H]^{2+}$ *m/z* = 1195.51294 (calculated for $C_{100}H_{130}N_{46}O_{26}^{2+}$: 1195.51267).

HPLC:

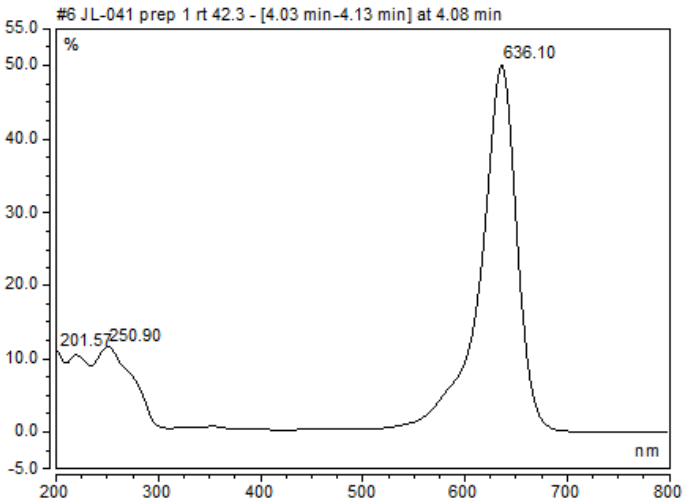




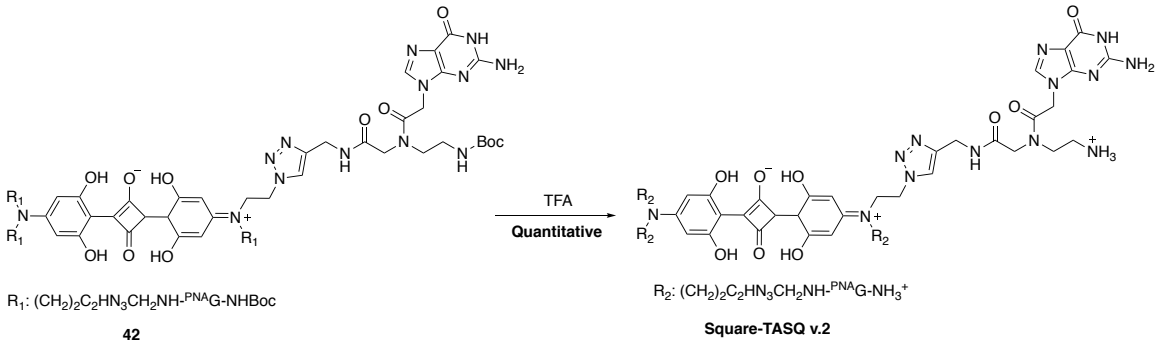

Chromatogram
2022_4_seq1 #6 [manually integrated]
(Boc)4-SquareTASQ v2
WVL:280 nm
%D: 0.0 %
%C: 0.0 %
1 - 4.080
100.0
CH3CN + 0.1%FA: 5.0 %
Flow.Nominal: 0.500000 ml/min
5.0
2 - 4.407
0.050000
Absorbance [mAU]
Time [min]


Chromatogram
2022_4_seq1 #6 [manually integrated]
(Boc)4-SquareTASQ v2
WVL:636 nm
%D: 0.0 %
%C: 0.0 %
1 - 4.080
100.0
CH3CN + 0.1%FA: 5.1 %
Flow.Nominal: 0.500000 ml/min
5.0
2 - 4.197
0.050000
Absorbance [mAU]
Time [min]

Uv-Vis:


#6 JL-041 prep 1 rt 42.3 - [4.03 min-4.13 min] at 4.08 min
%
636.10
201.57
250.90
nm

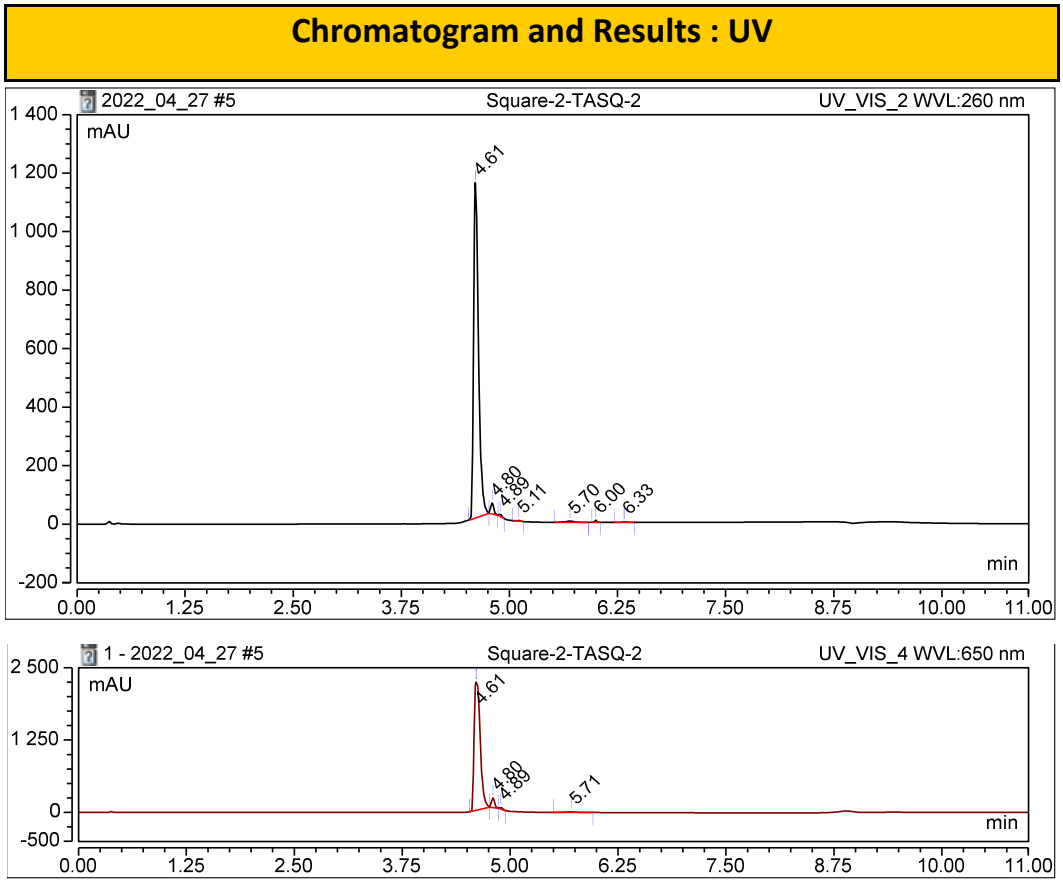
Square-TASQ v.2:



42 (1.95 mg, 0.8 µmol) was dissolved in TFA (200 µL) and the complete deprotection was assessed *via* HPLC, the mixture was diluted in water and the compound was freeze dried to achieve **Square-TASQ v.2** (1.96 mg, 0.8 µmol, 100%).

ESI-HRMS $[M+3H]^{3+}$ *m/z* = 663.94098 (calculated for $C_{80}H_{99}N_{46}O_{18}$: 663.94097).

HPLC:



Biophysical assays:

Oligonucleotides preparation:

Preparation of stock solutions. The lyophilized DNA/RNA strands (purchased from Eurogentec) were firstly diluted in deionized water (18.2 MΩ.cm resistivity) at 500 µM (for monomolecular (*i.e.,* F21T, F-Myc-T, F-TERRA, F-VEGF and F-duplex-T for FRET-melting experiments; F-Myc, F-22AG, F-SRC and F-duplex for pull-down assay; Cy5-Myc for $^{APP}K_D$ measurements) and bimolecular structures (*i.e.,* ds17 and CT-DNA). All DNA structures were prepared in a Caco.K buffer, comprising 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl. Monomolecular structures were prepared by mixing 40 µL of the constitutive strand (500 µM) with 8 µL of a lithium cacodylate buffer solution (100 mM, pH 7.2), plus 8 µL of a KCl/LiCl solution (100 mM/900 mM) and 24 µL of water. Bimolecular structures were prepared by mixing 40 µL of each constitutive strand (500 µM) with 16 µL of a lithium cacodylate buffer solution (100 mM, pH 7.2), plus 16 µL of a KCl/LiCl solution (100 mM/900 mM) and 48 µL of water. The actual concentration of each DNA sample was determined *via* dilution to 1 µM (expressed in motif concentration) for monomolecular structures (*i.e.,* 4 µL in 996 µL water) and bimolecular structures (*i.e.,* 8 µL in 992 µL water) through UV spectral analysis at 260 nm (after 5 min at 90°C) with the following molar extinction coefficient values provided by the manufacturer. The higher-order DNA/RNA structures were folded according to two procedures: (a) for the monomolecular architectures, solutions were heated (90°C, 5 min), cooled on ice (10 minutes) and then stored at least overnight (4°C); (b) for the folding of bimolecular structures, the solutions were heated (90°C, 5min), gradually cooled (65, 60, 55, 50, 40 and 30°C (60mn/step), 25°C (2hr)) and then stored at least overnight (4°C).

FRET-Melting:

FRET-melting experiments were performed in a 96-well format using a Mx3005P qPCR machine (Agilent) equipped with FAM filters (λ_{ex} = 492 nm; λ_{em} = 516 nm) in 100 µL (final volume) of 10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl (F-duplex-T) or plus 1 mM KCl/99 mM LiCl (F-Myc-T and F-Terra-T) with 0.2 µM of labelled oligonucleotide and 1 µM of TASQ. Competitive experiments were carried out with labelled oligonucleotide (0.2 µM), 1 µM TASQ and 15 and 50 mol. equiv. of the unlabelled competitor ds17 or Ct-DNA. After an initial equilibration step (25°C, 30 s), a stepwise increase of 1°C every 30s for 65 cycles to reach 90°C was performed, and measurements were made after each cycle. Final data were analyzed with Excel (Microsoft Corp.) and OriginPro®9.1 (OriginLab Corp.). The emission of FAM was normalized (0 to 1), and $T_{1/2}$ was defined as the temperature for which the normalized emission is 0.5.

Fluorescence pull-down:

The streptavidin MagneSphere® beads (Promega) were washed 3 times with TrisHCl buffer containing 1 mM KCl, 99 mM LiCl and 10 mM $MgCl_2$. Either BioCyTASQ or BioTriazoTASQ (10 µM) was mixed with 5'-labelled oligonucleotides (F-ON, 1 µM: F- 22AG, F-Myc, F-duplex, F-NRAS and F-TRF2), MagneSphere® beads (32 µg) in the same TrisHCl buffer (320 µL final volume) and stirred for 1 h at 25 °C. The beads were immobilized (fast centrifugation (< 2 s), magnet) and the supernatant removed. The solid residue was resuspended in 320 µL of TBS 1X buffer, heated for 8 min at 90 °C (gentle stirring 800 r.p.m.) and then centrifuged for 2 min at 8900 rpm. The supernatant was taken up for analysis (magnet immobilization), after being distributed in 3 wells (100 µL each) of a 96-well plate, using a ClarioStar® machine (BMG Labtech) equipped with FAM filters (λ_{ex} = 492 nm; λ_{em} = 516 nm). Competitive experiments were carried out with labelled oligonucleotide (1 µM), 10 µM TASQ and in the presence of the double strand DNA (ds17) competitor (25 mol. equiv.). Data were analyzed with Excel (Microsoft Corp.) and OriginPro®9.1 (OriginLab Corp.); each analysis comprises: a/ 3 control wells with F-ON and beads only, in order to quantify the non-specific F-ON/bead binding, the FAM emission of the solution was normalized to 1; and b/ 3 wells comprising solutions that resulted from experiments performed with F-ON, TASQs and beads, in order to quantify the actual TASQs capture capability when compared to the control experiments.

G4-chemo-precipitation monitored by qPCR

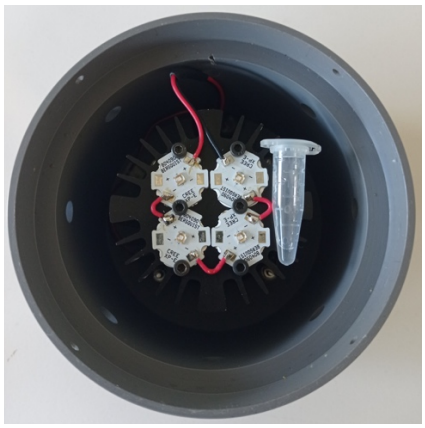
The pull-down experiments were performed as above, with the following modifications: a/ the buffer was replaced here by the so-called G4RP buffer (25 mM Tris pH 7.4 plus 150 mM KCl, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40); b/ the incubation time was changed for 2 h at 25 °C; c/ the output was changed for qPCR analyses: polymerase reactions were carried out in triplicate in 96-well format using a Mx3005P qPCR machine (Agilent) equipped with FAM filters (λ_{ex} = 492 nm; λ_{em} = 516 nm) in 20 µL (final volume) of G4-1R primer (1 µL, 300 nM), TASQ/ODN mixture (3.7 µL) in 10 µL iTaqTM Universal SYBR® Green Supermix (Bio-Rad) + KCl (5.3 µL, 100 mM). After a first denaturation step (95 °C, 5 min), a two-step qPCR comprising a hybridization step (85 °C, 10 s) and an elongation step (60 °C, 15 s) for 33 cycles was performed, and measurements were made after each cycle. Final data were analyzed with OriginPro®9.1 (OriginLab Corp.). The starting emission (first qPCR cycle) of SYBR Green (FI) was set to 2200 fluorescence intensity and the FI at the 33th cycle was used for calculation. Three biological triplicates (n = 3) were used.

Apparent K_D measurements:

2 µL of a concentrated solution (50 µM) containing Cy5-myc previously folded as described above were dilute in the assay buffer (50mM TrisHCl, pH 7.2, 150mM KCl, 0.05 v/v % Triton X-100) in order to achieve a concentration of 20nM. A dilution series were prepared with concentrations ranging from 1mM to 1 pM. 10 µL of each concentration of ligand were added to 90 µL of the aforementioned oligo solution (final oligo concentration: 18 nM) and let to incubate for 1 hour at 25°C. The solutions were transferred into a 96-well plate and after centrifugation the end-point fluorescence was measured using a ClarioStar® machine (BMG Labtech) equipped with Cyanine5 filters (λ_{ex} = 610 nm; λ_{em} = 675 nm). Data were analyzed with Excel (Microsoft Corp.) and OriginPro®9.1 (OriginLab Corp.); the values were normalized ranging from 0 to 100, and the percentage of ligand bound was calculated subtracting the normalized values to 100. Three biological triplicates (n = 3) were used to calculate an average and adequate fitting curves were generated.

Diazirine activation IP-HPLC:

SRC (500 µM) was folded as described above (10 mM lithium cacodylate buffer (pH 7.2) plus 10 mM KCl/90 mM LiCl) and diluted to 20 µM with a cacodylate buffer (pH 7.2) 1mM. **photoMultiTASQ** (5 equiv.) was added and let to reciprocate for 20 minutes (sample was split in two: non irradiated control and irradiate sample). The microtube containing half of the sample was directly placed under the light source (λ_{em}: 365 nm, 10 MW) with an ice bath and irradiated for 1 hour. Both samples were diluted (4 µM) with concentrated running buffer (TEAA 200mM) and heated at 90°C for 10 minutes. The two samples were analyzed with a Jupiter C18 (4µm, Proteo 90 Å, 250 x 4.6 mm), Buffer A: 100 mM triethylammonium acetate (TEAA, pH: 7.2), Buffer B: 10% Buffer A, 90% acetonitrile, gradient from 0% to 30% of Buffer B in 30 minutes.



Led lamp used for irradiation with a microtube used ad reference.

Fluorescent titration:

Spectra were recorded on a JASCO FP8500 spectrofluorimeter in a 10mm path-length quartz semi-micro cuvette (Starna). Fluorescence experiments were carried out in 500 µL (final volume) of 100mM lithium cacodylate buffer (pH 7.2) + 90mM LiCl/10mM KCl, with **N-TASQ**, **^{Tz}N-TASQ** and **H_{16}N-TASQ** (2µM) alone or in presence of DNA/RNA oligonucleotides (from 0 to 2.5 equiv) being either duplexes (ds17) or quadruplexes (SRC, 22AG and VEGF). Spectra (λ_{ex}= 286nm, λ_{em}= 300- 600nm, Ex and Em slits = 5nm, 1nm pitch, 2s response, scan speed = 500nm.mn^{-1}) were recorded 5 minutes (at 25°C) after the addition of the oligonucleotides. The fluorescence fold increase was calculated normalizing the fluorescence of the probe alone to one, calculating the increase accordingly.

As for **N-TASQ**, **^{Tz}N-TASQ** and **H_{16}N-TASQ**, fluorescence experiments were carried out in 500 µL (final volume) of 100mM lithium cacodylate buffer (pH 7.2) + 90mM LiCl/10mM KCl with **Square-TASQ v.2** (2µM) alone in presence of DNA/RNA oligonucleotide (from 0 to 2.5 equiv) being either duplexes (ds17) or quadruplexes (22AG and TERRA). Spectra (λ_{ex} = 620nm, λ_{em} = 630- 750 nm, Ex and Em slits = 10nm, 1nm pitch, 2s response, scan speed = 200nm.mn^{-1}) were recorded 5 minutes (at 25°C) after the addition of the oligonucleotides.

Absorbance studies:

Spectra were recorded on a Varian Cary 50 Scan UV-VIS Photometer in 10mm path-length quartz cuvette. Absorbance experiments were carried out in 500 µL (final volume) of DMSO or 100mM lithium cacodylate buffer (pH 7.2) + 90mM LiCl/10mM KCl, with **Square-TASQ v.2** (2µM) alone or in presence of DNA oligonucleotide (2.5 equiv of SRC). Absorption spectra as a function of time were recorded on a Varian Cary 50 Scan UV-VIS Photometer in 10mm path-length quartz cuvette. 500 µL final volume of 100mM lithium cacodylate buffer (pH 7.2) + 90mM LiCl/10mM KCl with **Square-TASQ v.2** (2µM); spectra measured every 5 minutes for twenty minutes and then every ten minutes, data absorption spectra values were normalized from 0 to 1 with OriginPro®9.1 (OriginLab Corp.).

Quantum yield evaluation:

UV-visible spectra were obtained on a Varian Cary 50 Scan (single-beam, software Cary WinUV) spectrophotometer by using a rectangular quartz cell, pathlength: 10 mm, chamber volume: 0.5 mL), at 25 °C (using a temperature control system combined with water circulation). The absorption spectra of **Square-TASQ v.2** (before and after incubation with G4 oligonucleotide) and sulfoindocyanine dye Cy 5.0 were recorded in cacodylate or PBS buffer within the concentration range 1 to 2 µM. Fluorescence spectra were recorded on an HORIBA Jobin Yvon Fluorolog spectrofluorometer (software FluorEssence) at 25 °C (using a temperature control system combined with water circulation), with the same rectangular quartz cell used for absorption measurements. The following set of parameters was used: shutter: Auto Open, Ex/Em slits = 7 nm for recording emission spectra and Ex/Em slits = 7 nm for recording excitation spectra, integration time = 0.1 s, 1 nm step, HV(S1) = 950 V. All fluorescence spectra were corrected. Relative fluorescence quantum yield of SqarTASQ (after

incubation with G4 oligonucleotide) was measured in cacodylate buffer at 25 °C by a relative method using the suitable standard (sulfoindocyanine dye Cy 5.0: Φ_F = 20% in PBS (10 mM phosphate + 2.7 mM KCl + 137 mM NaCl, pH 7.3), excitation at 580 nm; no dilution between absorption and fluorescence measurements[201]). The following equation was used to determine the relative fluorescence quantum yield:

$$\Phi_F(x) = (A_S/A_X)(F_X/F_S)(n_X/n_S)^2\Phi_F(s)$$

where A is the absorbance (in the range of 0.01-0.1 A.U.), F is the area under the emission curve, n is the refractive index of the solvents (at 25 °C) used in measurements, and the subscripts s and x represent standard and unknown, respectively. The following refractive indices was used for PBS and cacodylate buffer: 1.337.

Molecular dynamics:

The 3D structures of the ligands were manually constructed with UCSF Chimera, placing the four guanines in a plan with an average distance between the heavy atoms forming the hydrogen bonds of 3.5 Å. After a geometry optimization procedure, the RESP charges were computed and the GAFF parameters obtained using a combination of antechamber and acype programs. Water was described with TIP3P. Bonds containing a hydrogen were constrained with LINCS. A value of 8 Å was used for the non-bonding interactions cutoffs (standard value for the Amber force fields). The ligand structures were submitted to NPT molecular dynamics simulations in cubic boxes of 43 Å length. Velocities were generated at 200K and the systems were heated up to 350 K in 100 ps and were then kept at that temperature for 100 ps (velocity-rescale thermostat, 1 fs timestep). We then ran production simulations with a 2 fs timestep (velocity-rescale thermostat); for each ligand, 3 independent simulations of 500 ns were performed with Gromacs2021, during which the positions of the guanine atoms were restrained with a smooth potential of 1000 kJ/mol/nm2 to their initial position in order to force the guanine tetrade to be formed. Post-analysis (including SASA calculations) were done using Gromacs tools and the Plumed library.

Culture Cells assays :

Cellular pull down conditions :

Plant material :

Seeds of rice (Oryza sativa) cultivar Nipponbare (Japonica) were pregerminated at 25 °C for 3 days under dark condition. Uniformly germinated seeds were transferred to pots containing the nutrient soil and grown in a greenhouse with an automatically controlled condition: 28 to 30 °C and a 14 h/10 h light/dark cycle. For **G4DP-seq** in vivo: two-week-old rice seedlings were cut into 1-1.5 cm in length and merged into 1 % of formaldehyde (v/v) in HEPES buffer pH = 8.0 (20 mM HEPES, 1 mM EDTA, 100 mM NaCl and 1 mM PMSF) for cross-linking at 25 °C for 10 min in vaccuo. After quenching the excess of formaldehyde by adding 0.125 M final concentration of glycine followed by vacuuming for additional 5 min, the cross-linked leaves were rinsed with autoclaved H_2O and air-dried. For **G4DP-seq** *in vitro* and *in vivo*: cross-linked or native leaves were ground into fine powder using liquid nitrogen; after several washing steps using NIB (nuclear isolation buffer, containing spermine, spermidine and mercaptoethanol), NWB (nuclear washing buffer, containing triton X-100) and NDB (nuclear digestion buffer), chromatin (for cross-linked leaves) and genomic DNA (for native leaves) were isolated and used directly for downstream experiments (or stored at -80 °C for later use).

G4DP-seq protocol in vitro :

The purified rice genomic DNA was diluted with 1X sonication buffer (50 mM Tris-HCl, 10 mM EDTA and 1 % SDS w/v, pH = 8.0), then fragmented into 100 - 500 bp in size using the water-based Biorupter (Diagnode). A total of 5 µg purified fragmented genomic DNA was diluted in a G4-stabilizing buffer (150 mM KCl and 10 mM Tris-HCl, pH = 7.5), denatured at 95 °C for 10 min, then refolded thanks to a slow return to 25 °C. The refolded DNA was diluted with G4DP buffer (10 mM Tris-HCl, 5 mM EDTA, 150 mM KCl, 0.5 mM DTT, 0.5 % NP-40, pH = 7.4), then incubated with 100 µM of **BioTASQ**/biotin/**BioCyTASQ** with a constant rotation at 4 °C overnight. 30 µl of prewashed streptavidin-coupled Dynabeads were added the reaction above for another 4 h-incubation at 4 °C. The washed beads with **BioCyTASQ**/**BioTASQ** bound G4 DNA were eluted twice with 200 µl elution buffer (0.1 M NaHCO3 and 1 % SDS w/v) at 65

°C for 15 min each. Precipitated DNA and input DNA was purified using phenol/chloroform extraction followed by cold ethanol precipitation. All libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB, E7645S) for paired-end mode sequencing on Illumina NovaSeq platform.

G4DP-seq protocol in vivo :

The purified cross-linked nuclei were fragmented into sizes ranging from 100 - 500 bp in size in sonication buffer (10 mM Tris-HCl, 5 mM EDTA, 150 mM KCl, 0.5 mM DTT, 0.5 % NP-40 and 0.5 % SDS, pH = 7.4) using the water-based Biorupter. After being centrifuged at 12,000 rpm at 4 °C for 10min, the supernatant containing fragmented chromatin was carefully transferred to a new 1.5 mL tube and kept on ice. The fragmented chromatin was diluted with incubation buffer (10 mM Tris-HCl, 5 mM EDTA, 150 mM KCl, 0.5 mM DTT and 0.5 % NP-40, pH = 7.4) until the final concentration of SDS was below 0.1%. After keeping 1/10 volume of diluted supernatant as input, the remaining volume was incubated with 30 µl of BioTASQ/biotin at 4 °C overnight, then incubated with 30 µL of washed Dynabeads for another 4 h at 4 °C. The washed **BioTASQ**-bound DNA was eluted with 200 µl elution buffer twice at 65 °C for 15 min each, then reverse cross-linked at 65 °C overnight. The de-crosslinked DNA was purified using phenol/chloroform extraction and cold ethanol precipitation. Precipitated DNA and input DNA was used for library preparation as described above.

Motif prediction :

G4 motifs within G4 peaks were predicted using MEME-ChIP (http://memesuite.org/tools/meme-chip) with the following parameters: minimum width 5 bp, maximum width 25 bp. Only the top significantly enriched motifs (*i.e.,* with the highest E-values) are shown in *Figures 16 and 17*.

PFQSs identification and fold-enrichment analyses *:*

Putative G4-forming sequences (PG4FSs) were identified by screening the whole genome sequences using QuadParser. The fold-enrichment of PG4FSs (G_2+$L_{1\text{-}12}$) was calculated relative to random controls across the genome (bedtools shuffle command, observed values divided by average of 100 randomizations values).

Cellular imaging conditions

BioTASQ/BioCyTASQ:

MCF7 cells were routinely cultured in 75 cm2 tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% Penicillin-Streptomycin (10,000 units/mL penicillin, 10,000µg/mL streptomycin, Gibco) mixture. Cells were subcultured twice a week using standard protocols. Round coverslips (12 mm) were sterilized with 70% ethanol before cell seeding. MCF7 cells were seeded at a density of 6 × 104 cells per coverslip on chambered coverslips (24 well-plate) and allowed to recover overnight. In the case of live-cell labeling, cells were incubated with either **BioTASQ** or **BioCyTASQ** (1 µM, 24h) at 37 °C, washed once with PBS 1X, fixed and permeabilized with ice cold MeOH for 10 min at RT, washed with PBS 1X, then labeled with Streptavidin-Cy3 (1µg/mL) as below. In the case of postfixation labeling, seeded cells were fixed and washed with PBS 1X (3x), then incubated with 1 µM either **BioTASQ** or **BioCyTASQ** for 1 h at 25 °C, washed with PBS 1X (3 × 5 min), then incubated for 1 h at 25 °C in a light-tight box with Streptavidin-Cy3 (1 µg/mL) and washed with PBS 1X (3 × 5 min) and once with H_2O. Cells were mounted onto glass microscope slides with FluoromountG (Southern Biotech) containing DAPI (2.5 µg/mL). The cells were imaged with a Leica TCS SP8 confocal laser-scanning microscope with a 63× oil objective, collected through the following channels: DAPI (excitation: 340−380 nm; emission: < 425 nm), GFP (excitation: 450−490 nm; emission: 500−550 nm) and Cy3 (excitation: 518−562 nm; emission: > 580 nm). Images were processed with ImageJ (https://fiji.sc).

BioTriazoTASQ:

MCF7 cells were routinely cultured in 75 cm2 tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM, Dutscher L0104) supplemented with 10% fetal bovine serum (FBS, Dutscher S1810) and 1% Penicillin-Streptomycin (Pen-Strep: 5.0 U.mL-1 Pen/5.0 µg.mL-1 Strep, Gibco 15140122) mixture. Cells were subcultured twice a week using standard protocols.

Round coverslips (12 mm) were sterilised with 70% ethanol before cell seeding. MCF7 cells were seeded at a density of 6.104 cells per coverslip on chambered coverslips (24 well-plate) and allowed to recover overnight. Cells were incubated with either BioCyTASQ or

BioTriazoTASQ (1 µM, 24h) at 37 °C, washed once with PBS 1X, fixed and permeabilized with cold MeOH for 10 min at 4 °C, washed with PBS 1X (L0615 Dutscher, 3 x 5 min), then incubated with: 1/ the blocking buffer (PBS 1X/BSA 1%/Triton X-100 0.1%, 10 min at RT); 2/ Streptavidin-Cy3 (ThermoFisher Scientific 434315, 1 µg/mL, 1 h at RT); 3/ PBS 1X/Triton X-100 0.1% (3 x 5 min); 4/ anti-gH2AX antibody (Merck Millipore 05-636, mouse monoclonal Ab, 1/2000, 2 h in dark chamber); 5/ PBS 1X/Triton X-100 0.1% (3 x 5 min); 6/ anti-mouse Ig-AF647 antibody (donkey, 1/500, 45 min in dark chamber); 7/ PBS 1X/Triton X-100 0.1% (3 x 5 min); 8/ DAPI (1 µg/mL, 5 min in dark chamber); and 9/ PBS 1X (3 x 5 min). Cells were mounted (Fluoromount-G) and imaged with a Leica TCS SP8 confocal laser-scanning microscope with a 63X oil objective, collected through the following channels: DAPI (emission 448-475 nm), AF647 (emission 644-696 nm) and Cy3 (emission: 558-575 nm). One biological triplicate (n= 3) were used. For statistical hypothesis tests Student's *t*- test and Welch's unequal variances *t*-test were used depending on variances equality.

PDS treatment: Round coverslips (12 mm) were sterilised with 70% ethanol before cell seeding. MCF7 cells were seeded at a density of 6.104 cells per coverslip on chambered coverslips (4 well-plate) and allowed to recover overnight. Cells were incubated with either PDS (5 µM, 6 h, 0.5% DMSO) or supplemented DMEM-DMSO solution (6 h, 0.5% DMSO; untreated) at 37 °C, washed once with PBS 1X (L0615 Dutscher), fixed and permeabilized with cold MeOH for 10 min at 4 °C, washed with PBS 1X (3 x 5 min), then incubated with: 1/ the blocking buffer (PBS 1X/BSA 1%/Triton X-100 0.1%), 10 min at RT; 2/ BioTriazoTASQ (10 µM) or no TASQ (control) with anti-gH2AX antibody (Merck Millipore 05-636, mouse monoclonal Ab, 1/1000), overnight in dark chamber, at 4 °C; 3/ PBS 1X/Triton X-100 0.1% (3 x 5 min); 4/ Streptavidin-Cy3 (ThermoFisher Scientific 434315, 1 µg/mL) with anti-mouse Ig-AF647 antibody (donkey, 1/500), 45 min in dark chamber, at RT; 5/ PBS 1X/Triton X-100 0.1% (3 x 5 min); 6/ DAPI (1 µg/mL), 5 min in dark chamber; and 7/ PBS 1X (3 x 5 min). Cells were mounted (Fluoromount-G) and imaged with a Leica TCS SP8 confocal laser- scanning microscope with a 63X oil objective, collected through the following channels: DAPI (emission 448-475 nm), AF647 (emission 644-696 nm) and Cy3 (emission: 558-575 nm). Images were processed with ImageJ (https://fiji.sc) and *foci* quantification made according to home-made FIJI add-ons. One biological triplicate (n= 3) were used. For statistical hypothesis tests Student's *t*- test and Welch's unequal variances *t*-test were used depending on variances equality.

MultiTASQ and ^{Az}MultiTASQ :

MCF7 or MDA-MB-231 cells were routinely cultured in 75 or 175 cm^2 tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% Penicillin-Streptomycin (10,000 units/mL penicillin, 10,000µg/mL streptomycin, Gibco) mixture. Cells were subcultured twice a week using standard protocols. For microscopy experiments, MCF7 cells were seeded on glass coverslips in a 24 well-plate for 24 h at 37 °C. Cells were either treated live with **MultiTASQ** (5 µM in DMEM, 24h) then fixed (ice cold methanol, 10 min), or fixed and treated with **MultiTASQ** (10 µM in PBS, 1h). Coverslips were washed with PBS (3*5 min), and click staining is carried out with AlexaFluor488- or AlexaFluor594-azide (1 µM) in PBS enriched with 0.05% IGEPAL CA-630, 1 mM $CuSO_4$ and 10 mM sodium ascorbate for 30 min, washed with PBS+0.1% Triton (3*5 min), incubated with DAPI (10 min, 1 µg/mL in PBS) and mounted with Fluoromount. Confocal imaging was performed using a confocal laser-scanning microscope (Leica TCS SP8) with a 63× objective lens and LASX software (Leica Microsystems CMS GmbH). The samples were excited at 405 nm (DAPI) and 552 nm (Alexa594) and the fluorescence collected at 409 − 499 nm (DAPI) and 560 – 640 nm (Alexa594). Image processing was carried out using ImageJ.

Cells were cultivated as above, for microscopy experiments MCF7 cells were seeded on glass coverslips in a 24 well-plate for 24 h at 37 °C, fixed with ice cold methanol for 10 min and incubated with ^{Az}**MultiTASQ** (20 µM) for one hour. Cells were then incubated for 30 minutes with a PBS solution of BCN-Rhodamine (10 µM). After extensive washing with a triton containing PBS solution (0.1%, 3 times), cover slips were incubated with DAPI and mounted using Fluoromount. Samples were excited at 405 nm (DAPI) and 552 nm (Rhodamine). Fluorescence collected at 440 − 468 nm (DAPI) and 571 – 582 nm (Rhodamine). Image processing was carried out using ImageJ.

N-TASQ and ^{Tz}N-TASQ :

MCF7 or MDA-MB-231 cells were routinely cultured in 75 or 175 cm^2 tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% Penicillin-Streptomycin (10,000 units/mL penicillin, 10,000µg/mL streptomycin, Gibco) mixture. Cells were subcultured twice a week using standard protocols. For microscopy experiments, MCF7 cells were seeded on glass coverslips in a 24 well-plate for 24 h at 37 °C. For microscopy experiments, MCF7 cells were seeded on glass coverslips in a 24 well-plate for 24 h at 37 °C. Cells were fixed and permeabilized with ice cold methanol and incubated with **N-TASQ** (50 µM) and ^{Tz}**N-TASQ** (100 µM), cover slips were mounted using Fluoromount and samples were excited at 405 nm. Fluorescence collected at 450 − 530 nm, image processing was carried out using ImageJ.

Square-TASQ v.2:

MCF7 cells were routinely cultured in 75 or 175 cm^2 tissue culture flasks (Nunc) at 37 °C in a humidified, 5% CO_2 atmosphere in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% Penicillin-Streptomycin (10,000 units/mL penicillin, 10,000µg/mL streptomycin, Gibco) mixture. Cells were subcultured twice a week using standard protocols. For microscopy experiments, MCF7 cells were seeded on glass coverslips in a 24 well-plate for 24 h at 37 °C. Cells were fixed (ice cold methanol, 10 min) and incubated with **Square-TASQ v.2** (20 µM in PBS, 24h). Coverslips were washed with PBS (3*5 min), and incubated with DAPI (10 min, 1 µg/mL in PBS) and mounted with Fluoromount. Confocal imaging was performed using a confocal laser-scanning microscope (Leica TCS SP8) with a 63× objective lens and LASX software (Leica Microsystems CMS GmbH). The samples were excited at 405 nm (DAPI) and 638 nm (**Square-TASQ v.2**) and the fluorescence collected at 409 − 499 nm (DAPI) and 560 – 640 nm (**Square-TASQ v.2**). Image processing was carried out using ImageJ.