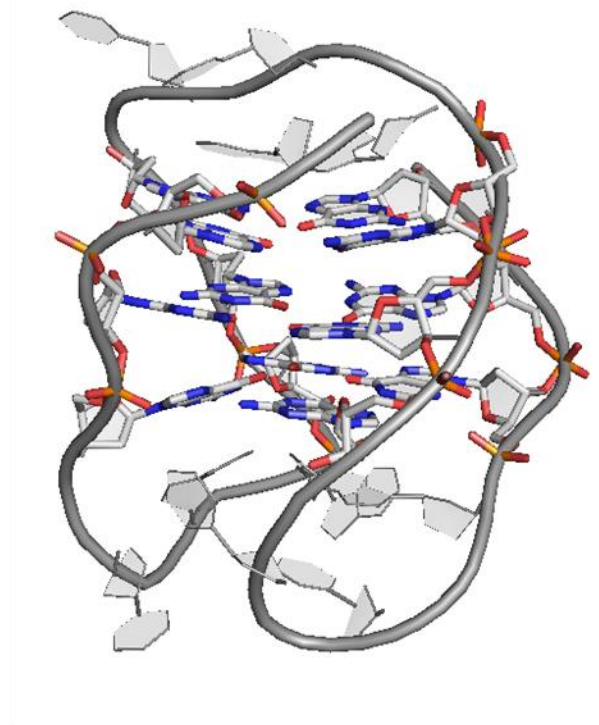**Universidad de Granada**
Programa de Doctorado en Química

DISEÑO Y SÍNTESIS DE LIGANDOS DE G-CUADRUPLEXES QUE INCORPORAN CARBOHIDRATOS COMO POTENCIALES FÁRMACOS ANTICANCERÍGENOS

TESIS DOCTORAL

MATILDE ARÉVALO RUIZ

GRANADA, MAYO 2017

Universidad de Granada
Programa de Doctorado en Química

Consejo Superior de Investigaciones Científicas

Instituto de Parasitología y Biomedicina López Neyra

Memoria presentada por la Licenciada
Matilde Arévalo Ruiz para optar al grado
de Doctora en Química

Director: Dr. Juan Carlos Morales Sánchez

La doctoranda **Matilde Arévalo Ruiz** y el director de la tesis **Juan Carlos Morales Sánchez**.

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 22 de Mayo de 2017

Director de la Tesis   Doctorando

Fdo.: Juan Carlos Morales Sánchez   Fdo.: Matilde Arévalo Ruiz

AGRADECIMIENTOS/AKNOWLEDGEMENTS

En primer lugar, me gustaría agradecer a mi director de tesis, el Dr. Juan Carlos Morales Sánchez por haberme dado la oportunidad de realizar la tesis en su grupo. No me considero una persona especialmente acertada en la toma de decisiones, pero la de dejar Granada para incorporarme a su grupo fue la mejor decisión que he tomado. Gracias por tu verdadera vocación científica (que logras contagiar), tu confianza en mí y por poner a mi disposición todos los medios que has podido para que aprendiera.

Gracias al Ministerio de Educación por concederme la beca FPU (marzo 2013-marzo 2017) con la que he podido llevar cabo todo este trabajo, así como al Consejo Superior de Investigaciones Científicas (CSIC), como institución pública que ha gestionado todo el proceso. También agradecer al Catedrático de la Universidad de Granada, Francisco Santoyo, por aceptar la tutela de la presente Tesis Doctoral.

Gracias a los técnicos de ambos centros. De Sevilla, a Gloria por el servicio de masas, a todo el personal de RMN, y especialmente a Fran, una persona imprescindible para todos los que pasamos por ese instituto. De Granada, a Samuel por el servicio de masas y a Alí por el de RMN. A Alberto Zafra, por permitirme usar su HPLC siempre que lo he necesitado. También al grupo de orgánica de la Facultad de Ciencias, por ayudarme simpre que lo he necesitado tanto a nivel de laboratorio como de docencia. Así mismo, agradecer al departamento de Ingeniería Química, especialmente a Antonio, por ayudarme con la docencia en el último año.

I would like to thank to Dr. Jean Louis Mergny for letting me stay in his laboratory in Bordeaux. It was a pleasure to work with a team constituted by the best researchers on the G-quadruplex field. Thanks to Prof Sara Richter as well for my short stage in Padua and for allowing me to experience the work with the HIV in a P3.

A nivel personal, de mi etapa en Sevilla dejé atrás a muchos compañeros y amigos que hicieron de mi estancia en el CIC Cartuja mucho más que un lugar de trabajo. Para empezar, como no podía ser de otra forma, gracias a Empar, a la que considero mucho más que una compañera, una amiga. Gracias por tu apoyo y por hacer que me riera hasta de mí misma. Gracias a Mónica, mi compañera de campana y veladora de mi seguridad, a quien le tengo un enorme cariño; y a mi otra compañera de gimnasio, Cristina. A Ricardo, gracias por ayudarme con tu sabiduría sintética. Del laboratorio de glicosistemas tengo muy buenos recuerdos gracias a todas las personas que formaban parte de él: María José, Juan Carlos, Ainhoa, Pepe, Carmen, Javi, José Juan, José Luis, Abel y Rocío. Gracias a Cynzia, por ser un encanto y acogerme en Milán como la mejor anfitriona.

Más allá del laboratorio de glicosistemas, los organometálicos también contribuyeron a que tenga muy buenos recuerdos de esta etapa: Astrid, Carmen, John Jairo, Laura, María…Haré una mención especial a Natalia, porque no existe persona más agradable con la que compartir piso. También a mi otra compañera, Sofía, la portuguesa con acento sevillano más entrañable. Por último, también mencionaré la otra ala del instituto que albergaba a mis chicos de materiales:

Antonio, Nico y Jaime; y a la felicísima Lilian. Todas estas personas hicieron de mis dos años en Sevilla una experiencia inolvidable, y espero conservar su amistad siempre.

De mi etapa en Granada, destaco sobre todo al señor Efres. No puedo expresar el agradecimiento que siento por todo el apoyo, tanto técnico como personal, que me has dado. Técnico por tu ayuda informática, en el laboratorio y tus clases de inglés. Personal por ser capaz de aguantarme hasta cuando ni yo misma era capaz de hacerlo. Por todo ello, pasaste de caerme tremendamente mal a ser alguien realmente especial. Tampoco me olvido de las clases de historia y geografía, porque Efres educa y divierte. También le dio las gracias a Pablo, por su inestimable ayuda, su eternísima paciencia (soy consciente de lo realmente cansina que puedo llegar a ser a veces), y por soportar como un campeón mis momentos de "achuchamiento". Gracias a los chicos de prácticas y demás personas que pasaron fugazmente por nuestro laboratorio: Ismael, María, Marisa y Sonia, porque fue un placer teneros en el laboratorio. El López Neyra nos acogió de la mejor manera posible, y aunque más tarde, conocí a gente estupenda dentro de él, a los que estoy segura de que no les perderé la pista: Norma, Raquel, Irene, Clara y muchos mas.

From Dr. Jean Louis Mergny's lab, thanks to Samir and Carmelo for his patience, I know it was not easy…and thanks to Aurore, the first one to help me in the lab. In addition, Nassima, Aurore, Oscar, Amina and Lylia contributed to the nice ambient there. I do not forget Laura, I wanted to mention her separately because she made my last stage in Bordeaux the best. And last but not least, the family who I felt mine there, Lallie and her daughters: merci beaucoup par tout, vous êtes les meilleures! From Padua, Martina was very nice with me inside and outside the Prof. Sara Richter's lab.

Le quiero dar las gracias a mi gente de Granada, porque siempre encontraban palabras de ánimo y apoyo. Cuando se conservan amistades así después de tantos años es porque verdaderamente merecen la pena. Por ello, gracias a Lucía, Maria de Mar, Cristina, Ana y Carmen.

Pasando a la familia, la sección que considero más importante, primero debo agradecer a mis padres, José Luis y Enriqueta, que me han apoyado en todas mis decisiones, aunque no fuera fácil para ellos. Por las mudanzas, el continuo ánimo, la fe en mí que siempre me demostráis, por montaros en un avión para ir a verme y por muchas muchas otras cosas más, GRACIAS. Sabéis que os quiero y no podría pedir más a unos padres como vosotros. A mi hermana Ángeles que, aunque estuviéramos en la distancia sabía que podía contar con ella para todo lo que necesitara. Y a mi cuñado Juanma, porque los Skype con él siempre eran más divertidos. Gracias a los dos, por hacerme sentir tan bien en vuestra casa y vuestro apoyo incondicional.

A mi tita Angelita, gracias por creer tanto en mí, verme como alguien al que yo no veo ni de lejos. Por tus riquísimas comidas y tu apoyo, muchas gracias. A mis tíos Matilde y Miguel, a quienes veo como un ejemplo de fortaleza, gracias por verme siempre con esos ojos de cariño, los mismos con los que les veo yo. Y gracias a mi prima Pilar, a la que dedico todo este trabajo que ahora me resulta insuficiente.

A mi prima

Pilar Sánchez

"Triunfa victorioso el emperador,
y no venciera si no hubiese peleado; y
cuanto mayor fue el peligro en la batalla,
tanto es mayor en el triunfo la alegría"

San Agustín (398 d.C)

Abbreviations

5-FU: 5- fluorouracil

A: adenine

A549: adenocarcinomic human alveolar basal epithelial cells

Ac: acetyl

ACN: acetonitrile

AIDS: acquired immune deficiency syndrome

ATP: Adenosine triphosphate

BLM: Bloom Syndrome Protein

C: cytosine

Carb-NDI: carbohydrate NDI conjugate

carb-PTDZ: carbohydrate-modified phenyl ditriazole

CC50: concentration with 50% of cytotoxic growth effect

CD: circular dichroism

CDCl3: chloroform

COSY: correlation spectroscopy

CuAAC: Cu-catalyzed alkyne azide cycloaddition

d: doublet

D2O: deuterium oxide

DBU: 2,3,4,6,7,8,9,10-Octahydropyrimidol[1,2-a]azepine

DCQAs: dicaffeoylquinic acids

dd: double doublet

DMAP: 4-dimethylaminopyridine

DMEM: Dulbecco's Modified Eagle Medium

DMF: dimethylformamide

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

dsDNA: double-stranded DNA

DU145: prostate cancer cell line derived from metastatic brain

EGFP: enhanced green fluorescent protein

EGFR: epidermal growth factor receptor

ESI: electrospray ionization

ESI-MS: Electrospray Ionization Mass Spectrometry

EtOAc: ethyl acetate

FAD: flavin adenine dinucleotide

FAM: fluorescein

FRET: Förster resonance energy transfer

G: guanine

G4: G-quadruplex

Gal: galactose

Glc: glucose

Glc-Pts: glucose-platinum conjugates

GLUT: glucose transporter

GTP: guanosine triphosphate

Hex: hexane

HFF: human foreskin fibroblasts

HIF-1: hypoxia inducible factor

HIV: human immunodeficiency virus

HPLC: high performance liquid chromatography

HRMS (FAB): high-resolution mass spectrometry

HT-29: human colorectal adenocarcinoma

hTel: human telomeric sequence

IC$_{50}$: concentration with 50% of inhibitory growth effect.

iFBS: inactivated fetal bovine serum

Abbreviations

IPBLN: Instituto de Parasitología y Biomedicina "López Neyra"

ITAG: ionic-liquid as soluble functional supports

Kd: dissociation equilibrium constant

LC50: median lethal dose

LOEC: lowest observed effect concentration

LRMS (ES)

LTR: long terminal repeat

m: multiplet

MDA-MB-435: melanoma breast cancer xenograft model

Me: methyl

MeOH: methanol

mer: a synonym for repeat unit in chemistry

MIA-PacA-2: pancreatic cancer cell line

mRNA: messenger RNA

MW: microwave

Mw: molecular weight

NADH: reduced form of nicotinamide adenine dinucleotide

NCI: National Cancer Institute

NCp: nucleocapsid protein

NDI: naphthalene diimide

NMR: nuclear magnetic resonance

NOEC: no observed effect concentration

NOESY: nuclear Overhauser effect spectroscopy

Nts: nucleotides

P3: Biosafety level 3 laboratory

PCNA: proliferating cell nuclear antigen clamp protein

PDB: protein data bank

PDH: pyruvate dehydrogenase

PI3K: phosphoinositide 3-kinase

PKM2: pyruvate kinase isozyme M2

ppm: one part per million

PQS: potential quadruplex sequences

PQS: putative quadruplex sequences

PTDZ: phenyl ditriazole

r.t.: room temperature

Rap1: the transcriptional repressor/activator protein 1

RNA: ribonucleic acid

ROS: reactive oxygen species

RRE: rev response element

RTEL1: Regulator of Telomere elongation helicase 1

RU: response unit

SDS: sodium dodecyl sulfate

sFv: intracellular single-chain antibody

SGLT: sodium-dependent glucose transporter

SI: selectivity index

siRNA: small interfering RNA

SPR: surface Plasmon resonance

SRB: sulforhodamine B

ssDNA: single-stranded DNA

T: thymine

t: triplet

TAMRA: tetramethylrhodamine

TAR: transactivation responsive element region

TBAF: Tetra-n-butylammonium fluoride

TBTA: Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine

t-BuONO: tert-Butyl nitrite

Abbreviations

TDS: thermal difference spectra

TEA: triethylamine

TERRA: RNA transcript from DNA human telomeric sequence

TFA: trifluoroacetic acid

THF: tetrahydrofuran

TIN2: TRF1 interacting protein 2

T_m: melting temperature

TMAA: trimethylammonium acetate

$TMSN_3$: trimethylsilyl azide

TNP: transdominant negative protein

TRAP: telomeric repeat amplification protocol

TRF1: telomeric repeat binding factor 1

TRF2: telomeric repeat binding factor 2

TSS: transcription start site

U: uracil

UPLC: Ultra Performance Liquid Chromatography

UV: ultraviolet

WHO: World Health Organization

WI38: lung normal human cell line

WSP: Werner Syndrome Protein

ΔT_m: change in melting temperature

RESUMEN

La presente tesis doctoral está centrada en el desarrollo de nuevos ligandos de G-quadruplexes y su potencial uso como agentes anticancerígenos. Los G-quadruplex son estructuras secundarias que se forman en secuencias del ADN y ARN ricas en guaninas, vinculadas a la regulación de importantes procesos biológicos como la replicación y la transcripción de determinados genes. Se han sintetizado y evaluado dos familias de compuestos: fenil ditriazoles y naftalen diimidas. El punto común en ambas familias es la incorporación de carbohidratos en la estructura con el fin de conferir selectividad a la estructura global de los ligandos, así como mejorar sus características farmacológicas.

La **familia de feniltriazoles** (PTDZ) fue sintetizada mediante la optimización de la cicloadicción 1,4- de Huisgen (click chemistry) para conjugar carbohidratos en forma de azida al núcleo aromático de 1,3-dietinil benceno. Con este método se sintetizaron compuestos simétricos, con azúcares neutros como cadenas laterales; y compuestos asimétricos, con un azúcar neutro como cadena lateral y una cadena cargada positivamente como segunda cadena lateral (ver esquema).
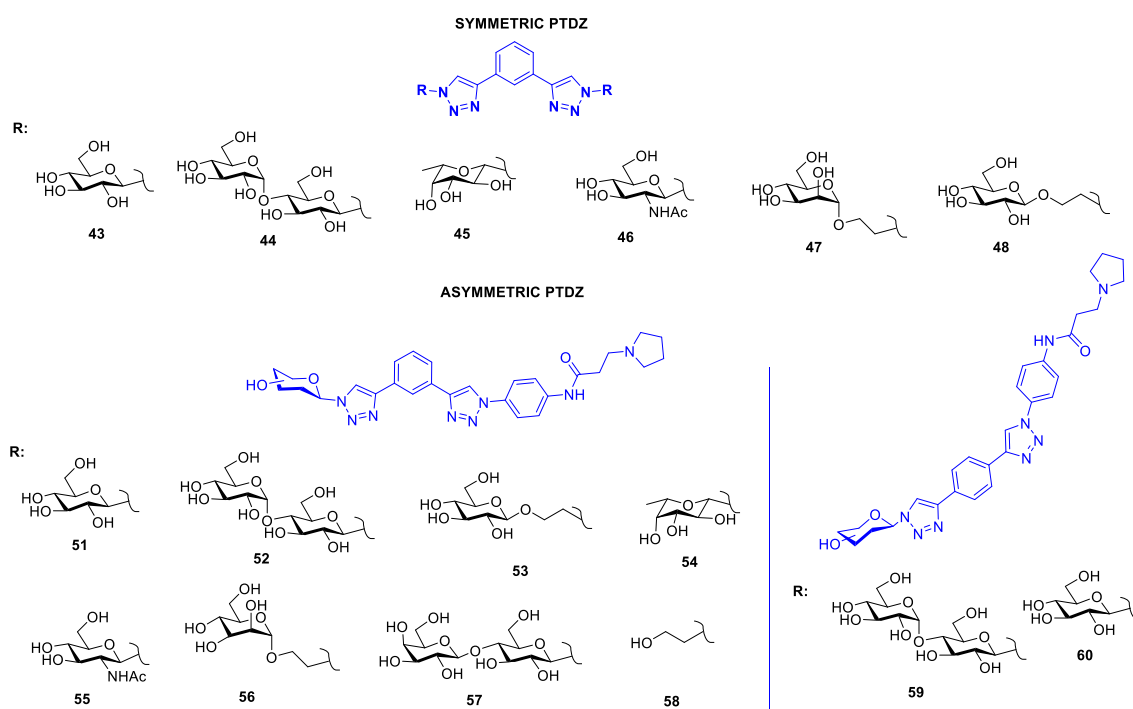
Se midió la capacidad de dichos compuestos para estabilizar los quadruplex mediante estudios de FRET-melting. Los simétricos no mostraron capacidad alguna de estabilización mientras que los asimétricos incrementaban la T_m del quadruplex del telómero humano hasta 13°C a una concentración de ligando 30 μM. Además, no mostraron ninguna habilidad para estabilizar secuencias de ADN en forma de dúplex, por lo que son bastante selectivos para quadruplex. Los resultados fueron muy similares para todos los compuestos por lo que cabe pensar que dicho efecto se debe a la cadena cargada, común en todos ellos. Los experimentos adicionales de CD, SPR y masas confirmaron la selectividad de esta familia, aunque la afinidad no resultó ser tan potente

como en otros ligandos de quadruplex previamente publicados. Los estudios de RMN con varias secuencias formadoras de quadruplex mostraron resultados bastantes interesantes. Con la secuencia del telómero humano se pudo localizar el área del quadruplex que más se ve afectada por la interacción de unos de los ligandos asimétricos.

La actividad anticancerígena observada en esta familia fue muy limitada cuando se evaluó su citotoxicidad sobre varias líneas cancerígenas humanas (MCF7, HT-29, LN-229 o HeLa). No se observó descomposición en agua o en medio de cultivo de los compuestos, pero parece que su escasa actividad se debe a la baja entrada celular.

La **familia de napftalen-diimidas** (NDI) fue sintetizada en colaboración con el Prof. Mauro Freccero (Universidad de Pavia) siguiendo la misma estrategia sintética que con la familia anterior: incorporación de carbohidratos a un núcleo aromático constituido por una naftalen-diimida bisubstituída a través de click chemistry. Se incluyó además como control un ligando sin azúcar (ver esquema).

61: β-Glc-NDI
62: β-GlcNAc-NDI
63: β-6-dGlc-NDI
64: β-Malt-NDI

65: α-manC2-NDI
66: β-glcC2-NDI

67: prop-NDI
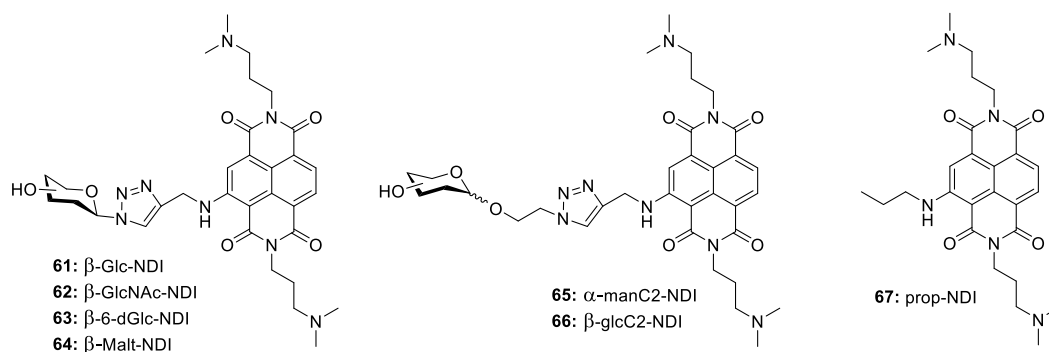
Los resultados de los estudios biofísicos y biológicos han sido publicados recientemente en Chemistry, DOI: 10.1002/chem.201604886.

A tan solo 2 μM, esta familia incrementó la T_m del telómero humano en potasio con valores cercanos a los 20°C, por lo que son mucho más potentes que los feniltriazoles. El ensayo de competitividad demostró que son bastante selectivos a las estructuras de quadruplex.

La citotoxicidad observada fue también muy superior a la de la familia anterior, con valores de IC_{50} que permanecen en el rango nanomolar o bajo micromolar. Sí que se observaron diferencias según el carbohidrato presente en la estructura. El derivado con glucosa unida directamente a la NDI presentó la menor toxicidad, mientras que el compuesto sin azúcar **67** y los que presentan una cadena de dos carbonos entre el azúcar (glucosa o manosa, **65** y **66**) y el núcleo aromático presentaron la mayor toxicidad.

Se hicieron estudios de entrada celular mediante citometría de flujo y espectroscopía de fluorescencia. Se observaron diferencias importantes entre el compuesto control **67** y los conjugados de carbohidrato-NDI. El compuesto control penetra en la célula de manera casi instantánea y sin depender de la temperatura, mientras que los conjugados carb-NDIs dependían de la temperatura, tiempo de incubación y de la disponibilidad del carbohidrato. Estos resultados pudieron relacionarse con los diferentes datos de citotoxicidad obtenidos: aquellos que entran en

mayor proporción y con más rapidez resultaron ser los más citotóxicos. Además, cuando las células fueron previamente incubadas con inhibidores de GLUT (transportadores de glucosa), los resultados de entrada celular parecen indicar que los derivados con linker **65** y **66** son transportados, al menos parcialmente, por GLUT4.

La fluorescencia inherente en los compuestos permitió localizarlos dentro de la célula mediante microscopía confocal, viéndose en mayor proporción en el núcleo de la célula.

Se realizaron ensayos de actividad antiviral frente a VIH con ambas familias en el laboratorio de Sara Richter (Universidad de Padua), pero ninguno de los compuestos desarrolló ninguna actividad interesante.

CHAPTER I:
GENERAL INTRODUCTION

1. QUADRUPLEX STRUCTURES AS NEW THERAPEUTIC TARGETS.
1.1. Quadruplex as non-canonical DNA structures.

Nucleic acids are molecules which carry all organism's genetic information. Due to their potential in explaining life itself; the constitution, replication and expression of these acids have been extensively studied since their discovery. It is now known that long chains of sugars are linked together via a phosphodiester bond between C3' and C5' atom of the subsequent sugar. The carbohydrate, which in RNA is D-ribose and in DNA is 2-deoxy-D-ribose, is also bonded through C1' to a purine or pyrimidine base. The common bases are adenine, guanine and cytosine whilst thymine is present only in DNA and uracil in RNA (figure **1.1**).
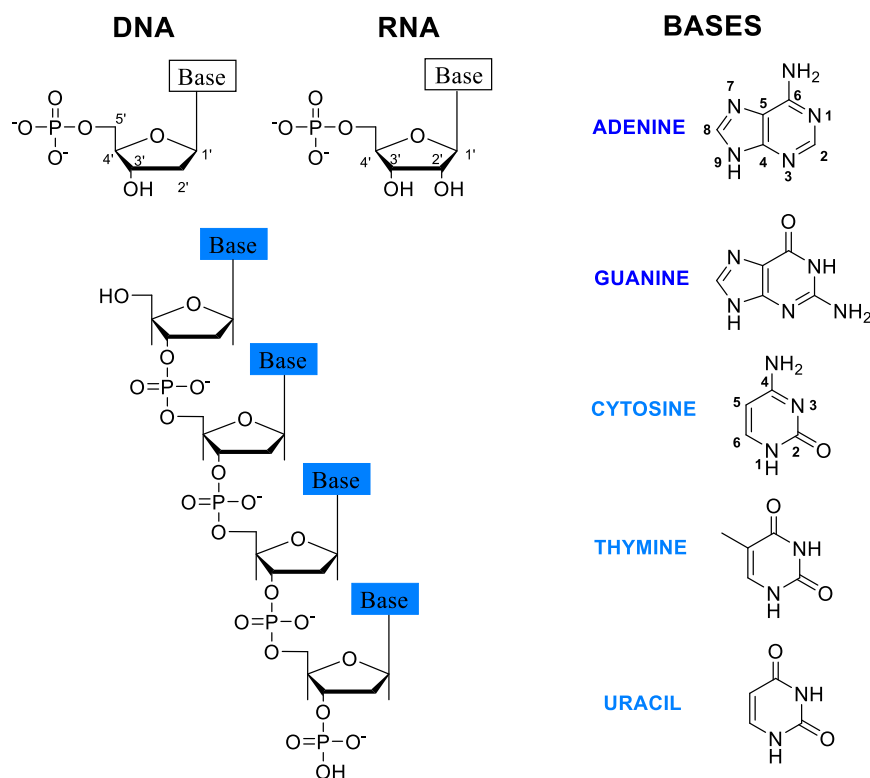
Figure 1.1. Left: DNA composition and simple backbone chain. Right side: nucleobases in DNA and RNA, purines (dark blue) and pyrimidines (light blue).

DNA and RNA oligonucleotides give rise to complex structural arrangements. The most well-known DNA structure was described by Watson-Crick in 1953,[1] where a duplex formed by two DNA antiparallel strands (5'→3' and 3'→5') results in a right-handed helix. This conformation -which is the most common in living beings and called **form B**- is stabilized by base pairing through hydrogen bonds and π-π stacking forces. The sugar adopts C2'-endo and C3'-exo conformation and the glycosidic-base bond an *anti*-conformation, enabling 10.2-10.4 base pairs per turn of helix in a diameter of around 2 nm. Right-handed helix **form A**, with 11 base pair per turn in a diameter of 2.3 nm, and left-handed helix **form Z**, which is only feasible synthetically at high salt concentration,[2] are other possible DNA conformations (figure **1.2**).
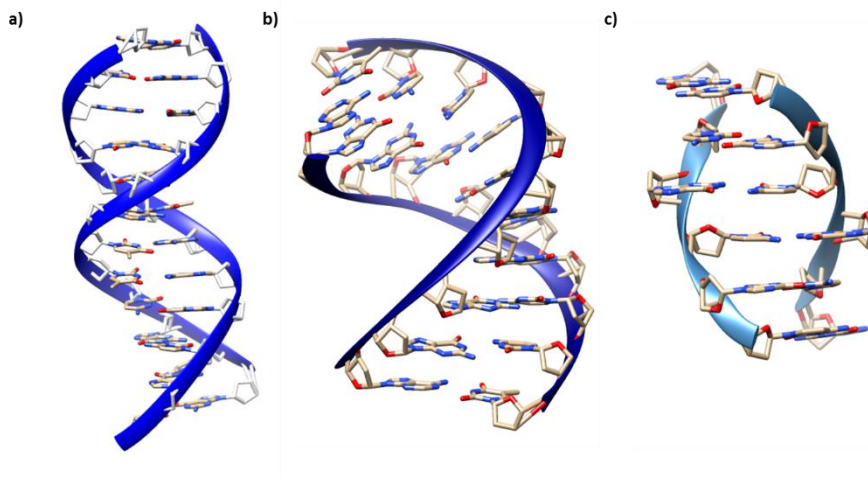
Figure 1.2. a) Duplex DNA structures. Form B (left side) described by Watson and Crick in 1953 (PDB: 1BNA). b) Form A (middle) from PDB 440D and c) form Z (right side) from PDB 1DCG.

Nevertheless, duplex DNA based on Watson-Crick hydrogen bonds are not the exclusive conformation of nucleic acid sequences. Hoogsteen hydrogen bonds lead to alternative non-canonical DNA structures, such as triplex or quadruplex DNA conformations. This new bonding system forms novel hydrogen interactions between adenine's N8, thymine's O8 and the protonated cytosine's N3 (figure **1.3**) allowing insertion of strands to DNA helix conformations.[3]
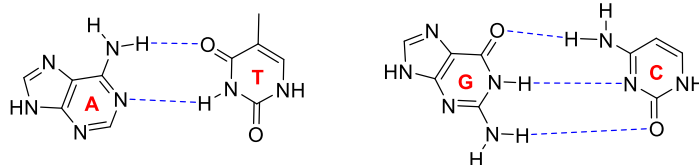
Figure 1.3. Interaction between nucleic acid bases. A. Classical Watson-Crick base pair. B. Alternative Hoogsteen base pair.

Two pyrimidine and one purine strand joined by Hoogsteen interactions along the major groove of the purine strand constitute **DNA triplex**. Their biological role is not clear yet but it appears to be formed by introns of a variety of crucial human genes related to development and signaling.[4]

G-quadruplex (**G4**) are interesting non-canonical DNA structures which have been receiving the attention of the scientific community. This interest is focused towards their structure and function as nanosensors and their potential role in cells seem to be linked to the regulation in replication and expression of genetic information. Basically, G4 are guanine-rich sequences in DNA or RNA which acquire a distinctive four stranded morphology. Bang *et al.*[5] initially reported that guanylic acid concentrated solutions form a dense gel, dilucidating its potential to form superior structures. This effect was also measured by Gellert *et al.*[6] through fibre-diffraction studies. The present work focuses in this kind of structures. We will describe in detail their composition, structure and importance in biology.

1.1.1. GENERAL G-QUADRUPLEX FEATURES

The **quartet** or **tetrad** is the G-quadruplex basic unit which is constituted by four associated guanines in a planar structure and stabilized by two series of Hoogsteen hydrogen bonds (figure **1.4a**). These bonds are formed between the N1-imino hydrogen and the O6-carbonyl oxygen, and between the N7-nitrogen and the hydrogen from an N2-amino group. Hence, each guanine is an acceptor and donor of two hydrogen bonds. Then, two or more quartets stack up through π-π interactions giving rise to the quadruplex arrangement (figure **1.4b**).
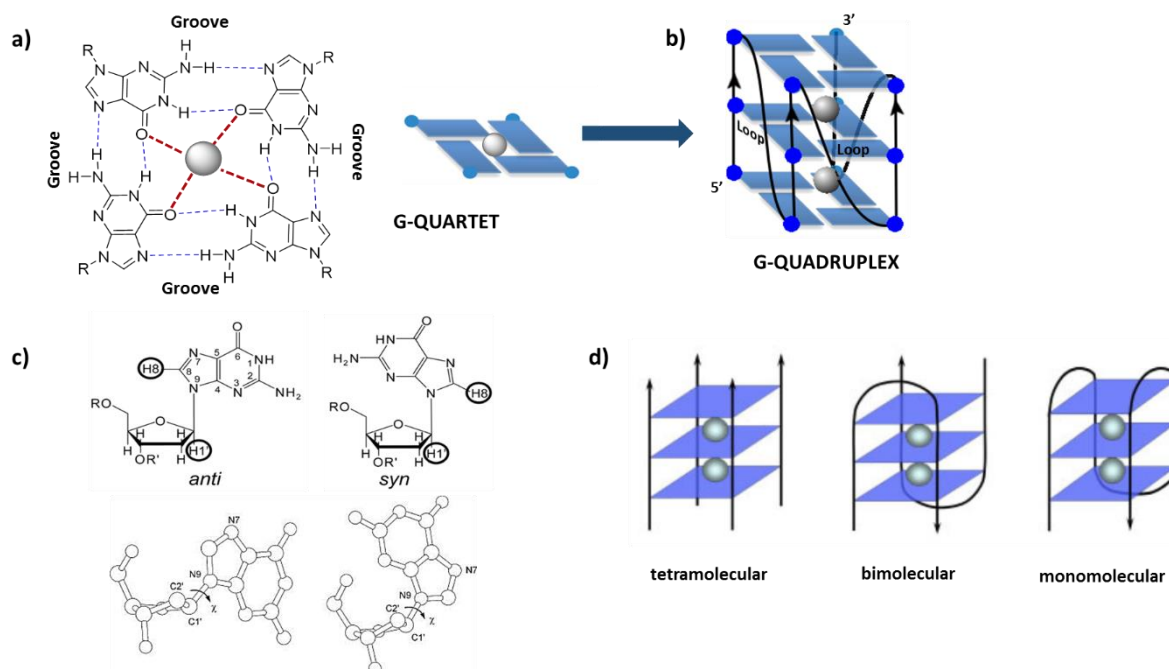
Figure 1.4. a) G-quartet and the interaction network with the gray ball representing the cation and blue dashed lines representing the hydrogen bonds. b) Resultant quadruplex from three G-quartets stacking. c) *anti* and *syn* guanine glycosidic bond conformation. d) different quadruplex arrangements according to the number of strands.

The quadruplex integrity is dependent on a monovalent metal cation which is hosted in the central cavity of the structure and is able to coordinate the quartets through the negative electrostatic potential generated by the guanines carbonyl groups.[7,8] The strength of stabilization of the monovalent cations goes in the order $K^+>Na^+>NH_4^+>Li^+$. The most common stabilizing cation is potassium, but sodium and ammonium are also feasible. The cation-hosting cavity

location can also vary and depends on the quadruplex nature and the cation size. For example, Na^+ is placed in the plane of the G-tetrad whilst K^+ and NH_4^+ are located in between two successive quartets.

All quadruplexes have as common features the presence of at least two G-quartets stacked together with a right-handed rotation and connected by loops. Between the four strands, we can find four spaces called **grooves** whose dimensions depend on the size and nature of the **loops**, and the pattern of glycosidic angles. The guanine-glycosidic bonds are able to adopt *anti* or *syn* conformations (figure **1.4c**) which define the guanine grooves and also affect the overall quadruplex topology. The G4-groove dimensions confer important structural features.

According to the number of strands, a quadruplex can be formed by: one (unimolecular or intramolecular G4), two (bimolecular G4), three (trimolecular) or four strands (tetramolecular G4) of DNA or RNA (figure **1.4d**).

1.1.2. QUADRUPLEX STRUCTURE

1.1.2.1. Quadruplex polymorfism.

Quadruplexes present many 3D conformations. This topological diversity is the result of: the oligonucleotide composition (length and sequence), the orientation it adopts, the conformation of the loops and the nature of the stabilizing cation.

- Strand sequence and length: DNA or RNA which are able to form G4 have a general sequence of $G_{\geq 3}$ X_n $G_{\geq 3}$ X_o $G_{\geq 3}$ X_p $G_{\geq 3}$. Guanines (G) can be 3 or more and X denote any base (A,T, G or C) between 2 and 20, which configure the loop's length and disposition.[9]
- Strand orientation: DNA and RNA phosphate backbone direction runs from either 5'→3' or from 3'→5'. Inside said structures, G4 motifs are not restricted to antiparallel strand orientation as DNA duplex structures are and additionally have the *anti* or *syn* glycosidic bond contribution to the orientation. When all these factors are taken into account, four G-tetrad structure cores can arise (figure **1.5a**):
    - Parallel: with the same direction of all strands and the same conformation of the glycosidic bond (*anti* or *syn*).
    - Antiparallel: two guanosines in *syn* conformation and two in *anti* (*syn·syn·anti·anti* and *syn·anti·syn·anti*).
    - Hybrid (3+1): a mix of parallel and antiparallel conformations with three guanosines with the same conformation and one with the opposite orientation (*syn·anti·anti·anti* and *anti·syn·syn·syn*).
- Configuration of the loops: loops has the strongest influence on G4 variability. When the loop connects two quartets with parallel adjacent strands they are propellers. When they run across a side between two adjacent anti-parallel strands they are lateral or edge-wise. And finally, they are diagonal when joining diagonally across the top of antiparallel strands (figure **1.5b**).

- Cation nature: G4 conformation can also be dependent on the cation present in the media. An example is the human telomeric G4-forming sequence, which adopts a different conformation in presence of sodium or potassium.
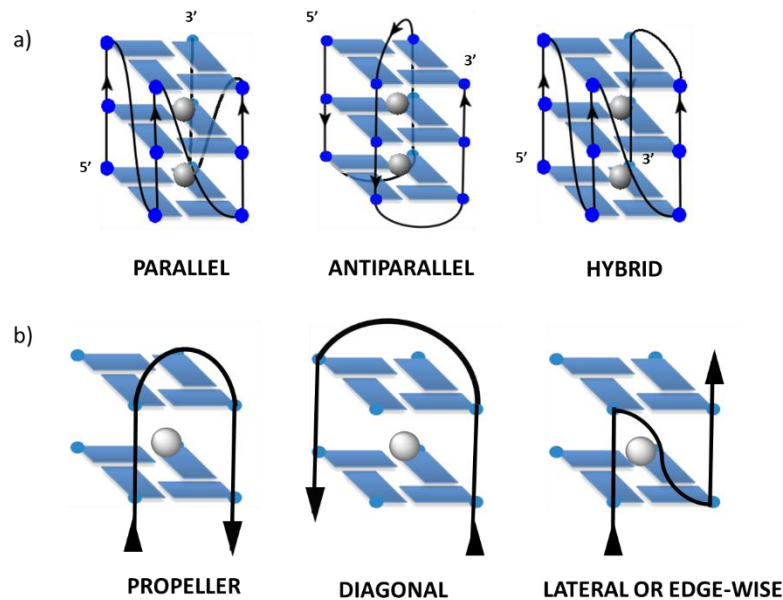
Figure 1.5. a) Different G4 topologies. b) Loop disposition types.

Hence, quadruplexes can be classified by:

➤ Strand number:
  ○ Intramolecular: one strand
  ○ Intermolecular: two, three or four strands.
➤ Strand orientation:
  ○ Parallel
  ○ Antiparallel
  ○ Hybrid
➤ Loop conformation:
  ○ Lateral
  ○ Diagonal
  ○ Parallel

Therefore, there is a vast G4-topologic variability described *in vitro*. Many PDB structures have been reported to date, some examples are shown in figure **1.6**.

G4 structures in **G-rich RNA** have greater polymorphism restrains than DNA because of the presence of the 2' hydroxyl groups which forces an *anti*-conformation orientation of the glycosidic bond. This is why most of the described G4-RNA forming sequences present a parallel conformation. Additionally, they have higher G4-folding propensity and increased stability respect to DNA due to the absence of the complementary RNA strand.[10]
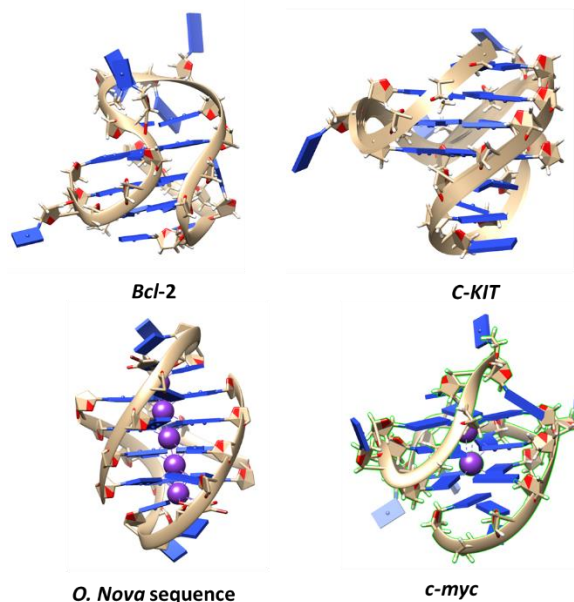
Figure 1.6. Different G4 quadruplex topologies examples: Bcl-2 promoter (PDB: 2F8U) Hybrid G4 formed K^+ in containing two lateral loops and one side loop; c-kit promoter (PDB:2O3M) Monomeric G4 formed in K^+ containing two single-residue double-chain-reversal loops, a two-residue loop, and a five-residue stem-loop; Crystal structure of native d[$G_4T_4G_4$] from *O.nova* in K^+ showing a bimolecular anti-parallel (PDB code: 1JPQ); c-myc promoter. Parallel G4 formed in K^+(PDB:1XAV).

1.2.   Quadruplexes in biology: prevalence and relevance.

Quadruplex studies are currently immersed in a positive vicious circle, where knowledge in the biological function of G4 increases the research effort and funding and vice versa. This circle is sustained by the great potential G4 exhibits towards many unsolved pathologies we humans are exposed to. Still, many unanswered questions about the role of G-quadruplexes within the organisms remain. Here we will review the most relevant discoveries and hypothesis, focusing firstly on the search of potential G-quadruplex forming sequences and continuing with their biological functions.

1.2.1.   PREVALENCE IN THE HUMAN GENOME.

Nowadays we are able to access and work with the complete genome of many organisms, including the human genome. This makes possible the identification of putative quadruplex sequences (**PQS**) through algorithms. The two first approximations were developed by Tood *et al.*[11] and Huppert *et al.*,[12] using $G_{3-5}X_nG_{3-5}X_oG_{3-5}X_pG_{3-5}$ as a basic search sequence. In these studies, they considered at least four repetitions of 3-5 guanines sequences separated by loops with 1-7 bases ($X_nX_oX_p$). This algorithm -called *Quadparser*-identified in the human genome more than 375,000 sequences as PQS, many clustered in long repetitive sections which they denominated "quadruplex islands".

Mergny *et al.*[13] recently demonstrated that *Quadparser* is too restrictive with the loop length which consequently resulted in big amounts of false negatives (missing quadruplex forming sequences) and false positives (detecting non-forming quadruplex as forming sequences) results. To overcome this situation, they developed a new algorithm called *G4Hunter,* able to acknowledge

sequential G-richness and skewness in calculations and provide the quadruplex-forming probability for all detected sequences. Moreover, Amrane *et al.*[14] recently described in the human mini-satellite CEB25 a quadruplex with the following sequence: d(GGTGGG**TGTAAGTGT**GGGTGGG) as a PQS. By having a 9 central nucleotide loop it was previously discarded by bioinformatics methods which evidence their limitations.

These G-quadruplex genome searches have located the PQSs mainly in two places:

1. In telomeric ends with a general repeating motif (**TTAGGG**)$_n$, suggesting their telomere length-regulation role.
2. Near some gene transcription start site regions (TSS) which pin-points to their importance in gene expression regulation role.

Less frequently, PQS have been found near most of the known 250,000 human replication origins,[15] as well as 5'UTR regions in mRNA[16] and TERRA (corresponding RNA-telomeric sequence).[17] They have also been found in G-rich micro- and mini-satellites within the ribosomal DNA near transcription binding sites and mitotic and meiotic double-strand break sites.

Several scientific G4-quadruplex searches with algorithms has demonstrated that these structures are not limited to human DNA and RNA, but are also present in other mammalian and non-mammalian species including lower organisms, such as parasites,[18] bacteria[19] and viruses.[20] Conservation of several PQS sequences along different families of the animal kingdom and different kingdoms suggests that G4-structures may be biologically important. Additionally, it has been confirmed that more evolved species -such as mammals- have better conserved quadruplexes in their genetic code.[21]

Biophysical confirmation using different techniques follow PQS genomic identification in order to separate false positives from real quadruplex forming sequences *in vitro*. Usually, up to five biophysical techniques are used to confirm quadruplex stability, including: CD, NMR, UV-melting or mass spectrometry (see section **1.3.3**.).

1.2.2. EVIDENCE OF G-QUADRUPLEX PRESENCE *in vivo*

After *in vitro* confirmation of quadruplex forming, a legitimate question to make is: can G4 structures actually fold *in vivo* or are these just artifacts of forced conditions (pH, buffer, cation nature and concentration…) to which DNA is exposed to? To answer this question, selective antibodies and small molecules have been developed to bind to G4 structures and help *in vivo* visualization. Schaffitzel *et al.*[22] were the first to observe the G-quadruplexes formation during ciliate replication in 2001 using high-affinity single chain specific antibodies in the telomeres of *Stylonychia lemnae*.

Later Biffi *et al.*[23] developed a high affinity and specificity monoclonal antibody (BG4) to detect intramolecular quadruplexes which were then applied to a range of human cell lines. They were able to locate them both in the telomeres and -in higher proportion- non-telomeric regions (figure **1.7**). They concluded that quadruplex formation is dependent on the cycle of the cell; whereas some are detected in all phases, most are seen during the S phase where the G4-structure numbers spike.
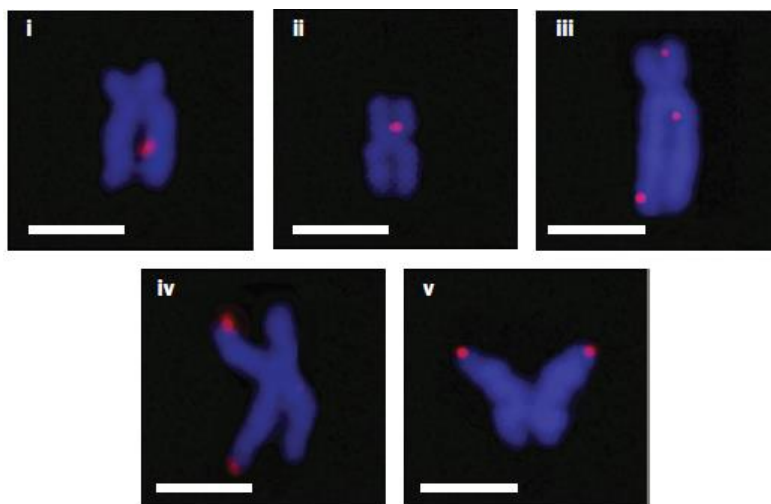
Figure 1.7. Immunofluorescence for BG4 on metaphase chromosomes isolated from Hela cervical cancer cells. I-III non-telomeric quadruplexes. IV-V telomeric G4.[23]

They continued their research by using the BG4 antibody on cancer and normal human tissues. Non-neoplastic tissues presented a diversified G4 staining pattern not affected by cell type within the tissue. However, when they compared liver cancer and normal liver tissues from the same patient, substantial differences were observed including higher G4-positive proportion in cancer cell samples (60.3% vs 18.3% of those found in the non-tumor sample). This quadruplex proportion was maintained in stomach cancer samples they also analyzed (~3.1 fold more in cancer than in non-cancer).[24] They concluded that G4 has a strong association with cancer formation and/or progression possibly due to unresolved G4 during DNA replication. This promoted the tumor trademark of genome instability and cancer.

Another approach to resolve this questions has been the use of small G4 fluorescent binders. *Biffi et al.*[23] treated simultaneously cells with the G4-binder Pyridostatin and GB4 antibody to obtain higher G-quadruplex detection values than with the GB4 alone (figure **1.8a**). Granotier *et al.*[25] radioactively labeled the compound 360A (^{3}H-360A) with tritium to visualize the telomeres in normal and tumoral human cell lines chromosomes (figure **1.8b**).
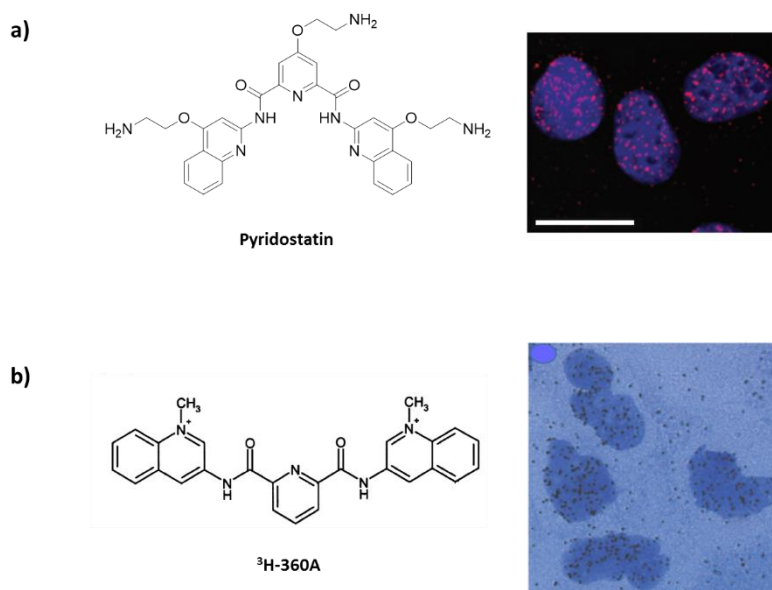
Figure 1.8. a) Structure of the G-quadruplex binding ligand Pyridostatin and increase in BG4 foci number in U2OS cells after treatment with pyridostatin. Nuclei are counterstained with DAPI (blue).[23] b) 360A structure and autoradiographs of T98G cultured with (B) 3H-360A for 24 h, showing accumulation of the radioactive compound in nuclei. Nuclei were stained with Mayer's hemalum solution.[25]

1.2.3. G-QUADRUPLEX AT TELOMERES.

Telomeres are a region of repetitive nucleotide sequences located at linear chromosomes endings to protect eukaryotic chromosome extremities from degradation and fusion. They play a key role in cell life and are involved in cellular aging and cancer. A DNA telomere generally consists of 100-200 nucleotides tracks with short repetitive sequences of $(TTAGGG)_n$, highly conserved in vertebrates. Due to its guanine-rich sequence it was suggested -and then confirmed- that it folds into a quadruplex structure. Sundkist and Wesley[26] reported the first telomeric-G4 in *Tetrahymena thermophile* with a d(GGGGTT) repeating telomeric tandem, which evidenced telomeric-G4 formation in invertebrates as well.

During replication, DNA-polymerase is unable to finish the telomere synthesis by itself. In somatic cells, telomeres are shortened by each cycle until the minimum Hayflick limit is reached, activating cell senescence and apoptosis. Therefore, this process acts as a natural mechanism of cellular division control (figure **1.9**).[27] Cells with high replication levels produce **Telomerase** enzymes to prevent telomere shortening through its reverse transcriptase activity adding nucleotides to the chromosome 3'ending (figure **1.9a**).[28] The catalytic human telomerase subunit (hTERT protein) uses the RNA subunit (hTR) as template to add 5'-GGTTAG-3' repeats to the ends. The hTERT activity is inhibited by intramolecular G4 formed in the single strand 3'-ending (figure **1.9b**). This avoids template hybridization. For this reason, quadruplexes are suspected to preserve chromosome integrity through telomerase inhibition.[29] Hence, we can hypothesize that quadruplex formation and stabilization through G4-ligands can prevent recognition of telomerase-RNA and G-strand overhang. Actually, telomerase is overexpressed in 80-85% of cancer types, making them an interesting target for tumor treatments.[30]
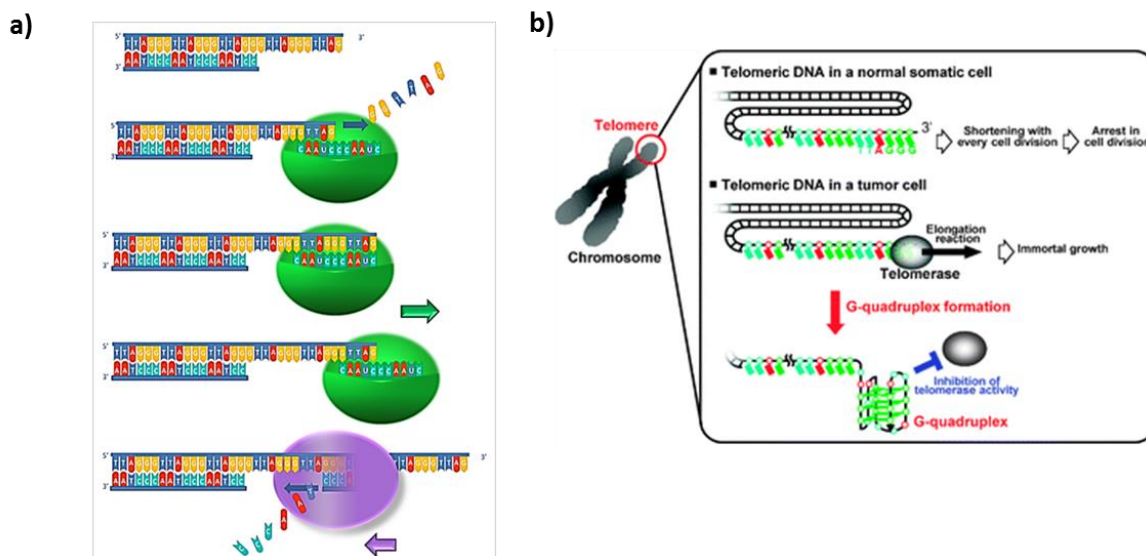
Figure 1.9. a) Mechanism by telomerase elongates the telomeres: telomerase binds to 3' flanking end of telomere that is complementary to telomerase RNA, then bases are added using RNA as template. Telomerase relocates, second step is repeated and DNA polymerase complements the lagging strand. Image modified from "Working principle of telomerase," by Fatma Uzbas (CC BY-SA 3.0). b) mechanism of telomerase inhibition by G4 formation at telomeric regions.

Human telomeric DNA is constituted by 2-10 kb double stranded sequence with 50-100 deoxynucleotides G-rich sequence on the 3' overhand.[31] It is the most studied G4-forming sequence due to its abundance inside the human genome. To date, the folding and unfolding regulation mechanism of G4-telomeric structures has not been completely resolved. It is known that telomere duplex-DNA is unwound during the replication process, aided by the telomeric specific binding proteins complex -Shelterin- whose tasked with telomere length regulation.[32] On the other hand, telomere binding protein POT1 (hPOT1 in humans) destabilizes quadruplex formation at telomeric DNA ends protecting them.[33]Therefore, there is a complex mechanism of regulation of G4-telomeric folding and unfolding with many protein and regulation factors involved. Actually, G-quadruplex on telomeric regions seem to have both beneficial and negative effects. On one side, they can cap the telomeres by inhibiting telomerase activity and consequently displaying a potential anticancer effect. On the other side, quadruplex on the replication fork-both in telomeric and non-telomeric regions- can stop the replication progress which could lead to genome instability. For that reason, the G4-regulation is essential for the correct cell genome stability.[34]

1.2.3.1. DNA-human telomeric G4-structure.

Depending on the supplemented cation, in addition to the G-tracks number and length of the telomeric sequence, different topologies have been observed according to NMR and X-ray crystallography studies. Thus, the human telomeric sequence fold into hybrid or antiparallel conformation with potassium or sodium buffer conditions, respectively.

The first structure was resolved by Wang *et al.*,[35] employing d(AG$_3$(T$_2$AG$_3$)$_3$) (**22AG**, 22 mer) as intramolecular G4-forming sequence, with four guanine rich repeats. They carried out the structural determination in sodium containing solution forming an antiparallel structure with two lateral loops and one diagonal loop. Consequently, two backbone orientations are in one direction

and the other two are antiparallel, obtaining a basket-like appearance (figure **1.10a**). The crystal structure of the same sequence in potassium conditions described by Parkinson *et al.*[36] revealed a parallel structure with loop nucleotides positioned on the exterior of the G4-core in a propeller like arrangement (figure **1.10b**).
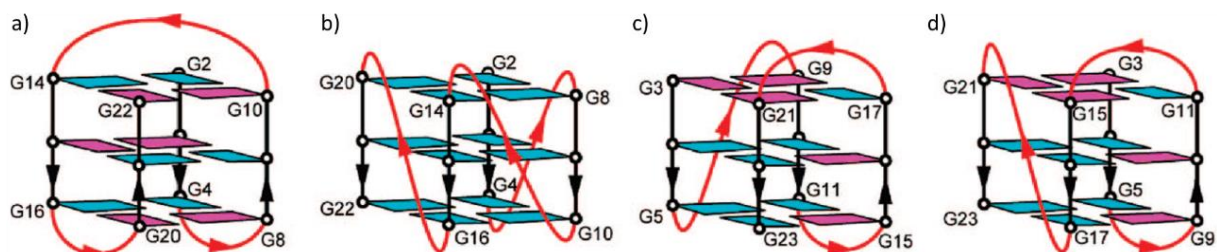
Figure 1.10. Schematic structures of human telomeric sequence. (a) basket-type form observed for d[A(G$_3$T$_2$A)$_3$G$_3$] in Na$^+$ solution (b) propeller-type form observed for d[A(G$_3$T$_2$A)$_3$G$_3$] in a K$^+$-containing crystal, (c) (3+1) Form 1 observed for d[TTG$_3$(G$_3$T$_2$A)$_3$A] in K$^+$ solution, and (d) (3+1) Form 2 observed for d[TA(G$_3$T$_2$A)$_3$G$_3$T$_2$] in K$^+$ solution. Loops are colored red; *anti* and *syn* guanines are colored cyan and magenta, respectively.[37]

Phan *et al.*[38] reported an alternative loop organization in potassium solution conditions, forming two intramolecular (3+1) quadruplex (figures **1.10c and d**). The 22 mer sequence can adopt multiple conformations under potassium conditions which complicates the NMR resolution. So, **24TTA**, a new 24 mer d(T$_2$G$_3$(T$_2$AG$_3$)$_3$A) sequence was characterized to favor a major conformation in potassium, improving the NMR spectrum quality.[37] The glycosidic conformation of the first tetrad is *anti·syn·syn·syn*, and *syn·anti·anti·anti* in the rest, according to the (3+1) G-tetrad core, with three tracks in one direction and the fourth ones in the opposite. In this structure, one loop is reversal and two edgewise (figure **1.10c**).

The main question is: which topology is closer to the real one in cells? The answer is still controversial. Several studies[(39–42)] suggest that the intramolecular G-quadruplex structure observed in potassium containing crystals[36] are the major form. Taking into account the greater potassium than sodium abundance in biological systems, structural resolution studies in potassium gain relevance.

1.2.3.2. RNA-human telomeric sequence.

Telomeric DNA can be transcribed into variable lengths (0.1-9.0 kb) non-coding RNA transcripts, commonly named TERRA, by a DNA-dependent RNA polymerase (Pol-II).[17,43,44] It has been demonstrated that (UUAGGG) repeats form stable parallel G-quadruplex structures in potassium conditions,[45] both in solution[46] and *in silico* (figure **1.11**).[47] These RNA higher order structures are involved in key cellular processes such as telomere length regulation (via telomerase activity), TERRA-telomeric-DNA hybrids formation, and DNA-damage response.[48–50] A wide range of proteins bind TERRA to regulate their localization and local concentration. For instance, Shelterin complex proteins TRF1 and TRF2 bind TERRA, and TRF2 binds TERRA-RNA quadruplex and DNA telomeric quadruplex at the same time.[51,52] However, it is not still clear how these structures display their roles and the consequences of their stabilization within the medicinal field.
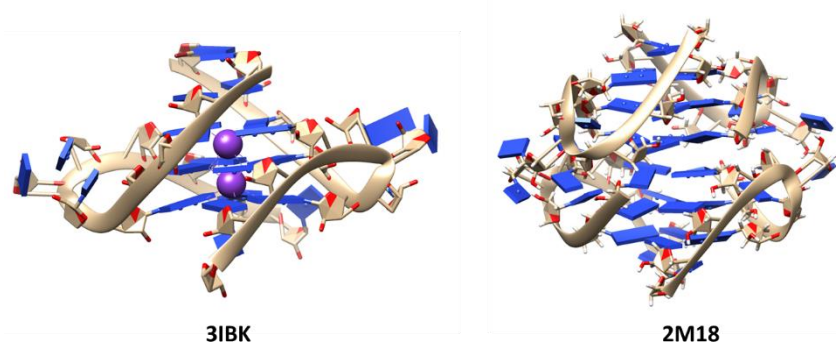
Figure 1.11. PDB structures published for human TERRA sequence.3IBK: crystal structure. 2M18: stacking structure of two G-quadrupex TERRA sequence blocks in K$^+$ solution.[53]

1.2.4. G-QUADRUPLEXES AT PROMOTER REGIONS. REGULATION OF TRANSLATION AND REPLICATION.

The G4 bioinformatic searches showed that in addition to chromosome endings, about 50% of human genes have PQS near promoter regions in oncogenes and regulatory genes.[54,55] It suggests G4 may interfere in expression regulation. The possible DNA G-quadruplex intervention in transcription is summarized in figure **1.12**.
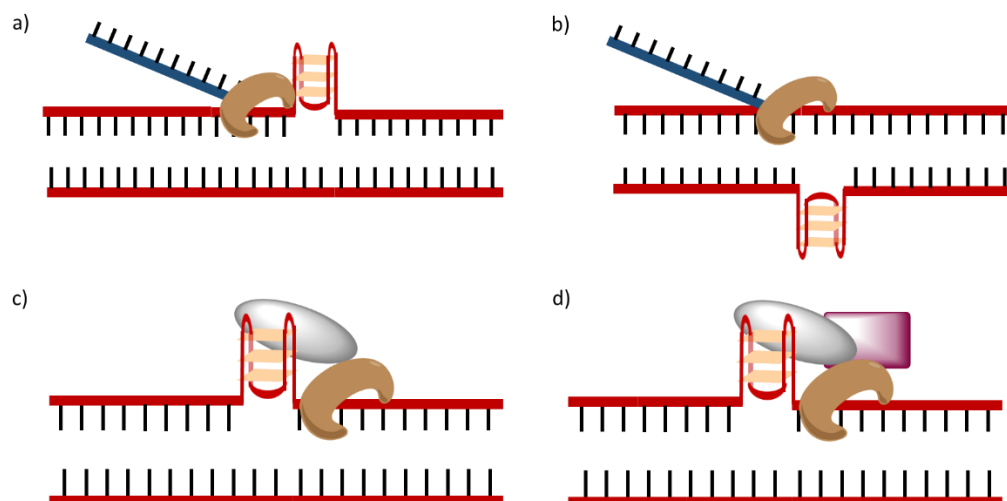
Figure 1.12. Possible roles of DNA G-quadruplex in transcription by RNA-polymerase (brown). a) Blocking transcription: transcription is inhibited due to the formation of a stable G-quadruplex. b) Facilitating transcription: formation of a stable G4 in the complementary strand enhances transcription. c) Stimulating transcription: G-quadruplex binds proteins (gray) that stimulate the transcription process. d) Repressing transcription: G-quadruplex binds protein that repress transcription directly (gray) or via other proteins (purple).

Two main hypothesis are discussed on the field about the role of G-quadruplex sequences on the TSS (transcription start site):[56]
a) Quadruplex formation just downstream of the TSS activates transcription.
b) Further downstream G4 may impede RNA polymerase II action and therefore DNA transcription.

Other model suggests G4s hinder normal nucleosome formation and duplex-DNA binding proteins interaction, which block transcription.[57,58]

G4 requires single stranded DNA to be able to form. Logically, this means that at least one double strand opening mechanism and regulation must exist to achieve G4 formation, and protein **helicases** are thought to be the main responsible. Helicases break the double-stranded structure as a precursor for cellular machinery access to:

o DNA for its replication, transcription, reparation and recombination,
o Chromosomes for its segregation,
o Telomeres for its maintenance.

Hence, helicases play a key role in genome homeostasis. Many helicases have been described to unwind quadruplexes on *in vitro* studies, but *in vivo* confirmation is still pending. To date, helicases regulate gene expression in a G4-dependent manner through a not clarified mechanism. However, differential expression levels in particular genes suggest their relation.[59,60]

Mendoza *et al.*[61] recently reviewed the helicases-G4 relation, and described two main Helicase superfamilies (SF):
- **SF1:** which include Pif1 (*Saccharomyces cerevisiae* helicase) and DNA2 helicases.
  o Pif1 is region-specific helicase in telomeres and mitochondria.
  o DNA2 is a nuclease-helicase which seems to resolve G4-telomeric structures according to several *in vivo* assays.[62]
- **SF2:** which include RecQ, FANCJ, Bloom Syndrome Protein (BLM), Werner Syndrome Protein (WSP) and RTEL1.
  o <u>FANCJ</u> acts over stable G4-structures because it displays unwinding preference towards G4 over DNA duplex, alleviating the replication damage generated.[63]
  o <u>RTEL1</u> (Regulator of Telomere elongation helicase 1) unwinds G4 in an ATPase-dependent manner. Its lack of action drives to unresolved T-loops during DNA replication, losing the telomere circle.[64]
  o <u>RecQ helicases</u> are an enzymatic family that include BLM, WSP and RECQ4. They are responsible of preserving DNA integrity during replication through C-terminal protein domain (RQC) which can bind to a DNA structure and unfold a G4. They are also related to disorders associated with cancer and ageing when their functionality fails. BLM helicase unwinds intermolecular G4 in a 3'-5' direction manner, with preference for G4 structures over duplex.[65] However, Tan *et al.*[66] reported that BLM unfolds intramolecular G4 less efficiently than duplex DNA, making the selectivity of BLM towards intermolecular G4 patent. In both TSS downstream locations reported with PQS, BLM and WRN helicases would denature the structures attenuating the G4 effects.

Outside these SF, **RHAU** is another important helicase associated to AU-rich sequences which can unwind duplex RNA, RNA-DNA hybrids, triplexes and G4 of both DNA and RNA. Furthermore, it has been reported that RHAU has great affinity for DNA and RNA quadruplexes.[67]

Most of non-telomeric PQS described are located in oncogenes or cancer-related genes, whose G4 folding near promoter region have been confirmed in some cases *in vitro*. Among them

we find c-*MYC*,[68] c-*KIT*,[69,70] *HIF*,[71,72] *BCL-2*,[73,74] *RET*,[75] k-*RAS*,[76] h-*RAS*,[77] *HSP90*,[78] b-*RAF*,[79] the androgen receptor,[80] VEGF[74] and MET.[81] The sequence and consequent topology is different for each one, in contrast to the telomeric one. The first non-telomeric G4-sequence identified was the corresponding nuclease hypersensitive element in the *c-MYC* promoter region gene,[68,82,83] an oncogene whose expression is reduced by G4-stabilizing ligand addition. It has been hypothesized about the role of these non-telomeric G4s.

In c-*KIT* gene, the G4 have four G-tracks repeats with a total length of 21-22-mer,[68,70] in c-*MYC* is a 27 mer with five G-tracks[84–86] and the longest G4-forming sequence c-*BCL*-2 has 39 mer and six G-tracks, increasing the topology possibilities.[87] The structural details knowledge is essential for the selective targeting by ligands. Some of these genes are overexpressed in certain cancers, thus researchers have focused their efforts in finding the way to break their expression through G4-stabilizating ligands. In c-*MYC* or *RAS* genes, the impossibility of using their corresponding proteins as therapeutic targets makes the G4-strategy the most feasible.

1.2.5. QUADRUPLEXES IN NON-HUMAN GENOME

PQS have also been located in lower organisms such as yeasts, parasites, bacteria or viruses, which has amplified the interest on G4 as therapeutic targets. Examples of them are: *Saccharomyces cerevisiae*,[88] *Escherichia coli*,[89,90] *Herpes Simplex 1*,[91] *papilloma*,[92] *Epstein-Bar*[93] or *SAR*[94] viruses.

Perrone *et al.*[95] searched for PQS in the HIV virus, finding G4 forming sequences in the NEF gene coding region and within the long terminal repeat (LTR) promoter.[20] Amrane *et al.*[96] identified, characterized and studied another very conserved HIV G4-LTR promoter sequence by CD and NMR techniques (HIV PRO1: 5′-TGGCCTGGGCGGGACTGGG-3′) responsible for provirus transcription. This quadruplex structure -composed by two G-tetrads- adopts an antiparallel topology with an additional Watson-Crick CG base pair (figure **1.13a**). G2-G18-G12-G9 and G3-G8-G13-G17 constitute the first and the second tetrads, respectively, which have opposite hydrogen bond directionalities around them. The CD spectrum shows the typical antiparallel quadruplex signatures pattern: a positive peak at 295 nm and a negative one at 265 nm (figure **1.13b**).
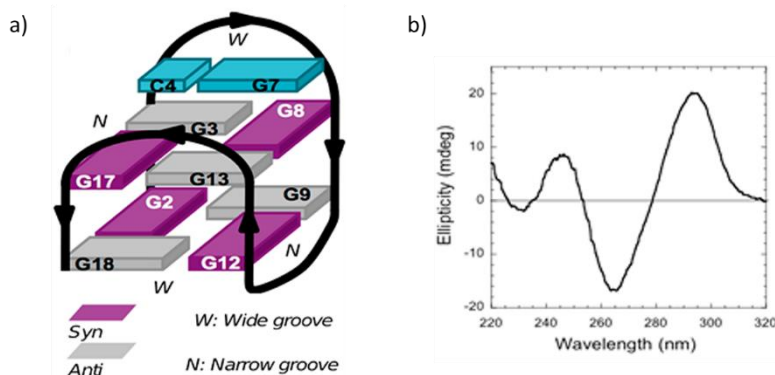
Figure 1.13. HIV-PRO1 biophysics characterization. a) Topology of the hybrid G-quadruplex structure. b) Circular dichroism spectrum recorded at 25 °C at a concentration of 5 µM in a buffer composed of 20 mM potassium phosphate pH 6.9 supplemented with 70 mM KCl.

1.3.   G-quadruplex ligands.

The role of G4 in biological systems makes them potential therapeutic targets. Small molecules have been designed and synthesized to bind G4. These G4-ligands could act by two different mechanisms:

✓ G4-stabilization among a gen promoter region which leads to transcription inhibition[55,97] and consequently lower protein expression which is especially useful in oncogenes (c-*KIT*, k-*RAS*, c-*myc*, B-*RAF*, *BCL*-2...) and G4-associated to cancer genes (VEGF growth factor, androgen receptor, HIF, hTERT[98] and HSP90[99]).

✓ Telomeric-G4 stabilization inhibits telomerase and/or telomere maintenance, which results in induced DNA shortening, cellular senescence and apoptosis,[100,101] making overexpressed telomerase cancer cell lines particularly susceptible.

With this goal in mind, many researchers have focused on developing G4-stabilizing ligands as therapeutical tools. As we have described previously, the ample variety of G4-topologies and prevalence makes selectivity the key feature to take into account. Hence, G4-ligands must have:

➢ Quadruplex *versus* duplex selectivity: the duplex form is the most abundant DNA arrangement and the ligands must be selective only and exclusively towards G-quadruplex structures to avoid side effects.

➢ Selectivity among different quadruplexes: DNA-quadruplex *versus* RNA-quadruplex and between different sequences and topologies (telomeres *versus* oncogene promoter regions, non-human quadruplexes….) would be highly desirable. In fact, the success on developing G4-selective binders would possible avoid side effects. To achieve so, G4-grooves dimensions and loop sizes which differentiate terminal quartet surface area and ligand-accessibility can be exploited.[102] Still, this aim has been found to be very challenging and to date, no ligand has been found selective for a certain G4 structure, even with partial topological selectivity success in human telomeric quadruplex.[103]

1.3.1.   G-QUADRUPLEX LIGANDS GENERAL FEATURES.

We can observe several very common structural features on the reported G4-ligands:[104]

o A large **aromatic central core**: polycyclic heteroaromatic systems which stack on top of the terminal G-quartet by π-π interactions. Acridine derivatives, naphthalene diimides or anthraquinones are some examples (see below).

o At least one flexible **cationic side chain** able to interact with the backbone phosphates and fit into the quadruplex grooves. They increase the global compound hydrophilicity and contribute to selectivity due to their grooves-interaction. Aliphatic chains with final primary, secondary or tertiary amine motifs are the most frequent chains, including-NMe_2, -NEt_2 and small rings as pyrrolidine, piperidine or piperazine. They are charged at physiological pH and enhance G4-ligand affinity.

In general, small molecules containing these previous features are strong G4-binders able to stabilize the structure *in vitro* at low concentrations. In contrast, they involve huge structures with high molecular weights (normally >600) and liphophilicity that do not meet the Lipinski Rule

of Five,[105] by which the appropriate properties for a drug-like compound are defined as follow:

- No more than 5 hydrogen bond donors (total number of N-H and O-H bonds).
- No more than 10 hydrogen bond acceptors (total number of N or O).
- Molecular weight less than 500 daltons.
- An octanol-water partition coefficient (log P) lower than 5.

In addition, they normally display not desirable ADME (absorption, distribution, metabolism and excretion) properties to become a drug. Actually, Quarfloxine (see below) is the only G4-ligand which has progressed to clinical evaluation.

1.3.2. G-QUADRUPLEX LIGAND FAMILIES.

a. The **first ligands** were developed as telomerase inhibitors. Among them we find anthraquinones,[106] porphyrins,[107] triazines[108] and acridines (figure **1.14**).[109,110] The most representative compound from this group is BRACO-19 (a 3,6,9-trisubstituted acridine derivative), developed by Neidle's group[111] with activity against prostate cancer[112] and uterus carcinoma[113] cell lines at sub-micromolar concentration range. The X-Ray crystallographic structure of the human telomeric sequence and BRACO-19 complex was resolved which let us observe the binding mode of the G4-ligand with G4 hTel: the core stacks on top of the quadruplex by π-π aromatic interactions and the side chains reach the G4-loops.[114] More recently, Sara Richter's group used this compound to stabilize the HIV-LTR sequence fragment and found antiviral activity.[115]
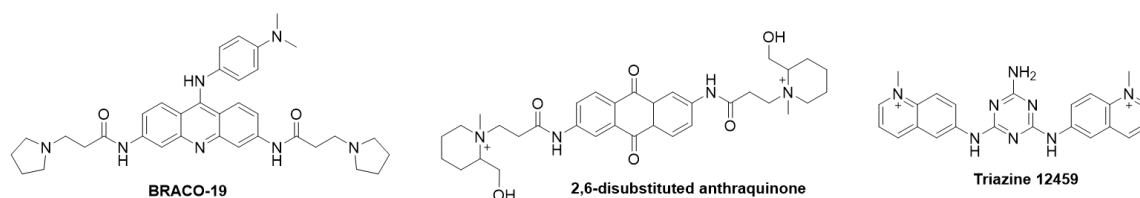
Figure 1.14. Some examples from the first G4-ligands family.

b. Aromatic natural compounds have been investigated as quadruplex ligands as well, like the alkaloid Berberine[116] and Telomestatine (figure **1.15**).[117] Berberine is an antibiotic with anticancer activity associated with telomerase inhibition[118] whose crystal structure with the human telomeric sequence has been described.[119] Telomestatin is a macrocyclic product with atypical G4-ligand structural features, which have five unsubstituted oxazole rings, two methyloxazoles and one thiazoline ring constituting a cyclic product with neutral charge. Despite of its configuration, it is slightly non-planar and overlaps on top of the terminal G-quartet by π-π aromatic interactions with high affinity and selectivity.[120] It acts as a very efficient telomerase inhibitor[121] and is active against several human tumor cell lines such as multiple myeloma,[122] childhood neuroblastoma[123] and acute leukemia.[124] On the other hand, this product has solubility problems which makes it not suitable for oral treatments. To solve this problem, more water-soluble derivatives have been prepared. For instance, macrocyclic pyridyl polyoxazoles (*e.g.* HXDV[125]) are selective through G4-DNA and RNA, and show antiproliferative activity against breast cancer cell lines in the nanomolar range. Even in MDA-MB-435 breast tumour xenograft model they could reduce the tumor size,[126] although the relation between *in vivo* results and quadruplex stabilization has not been clarified.
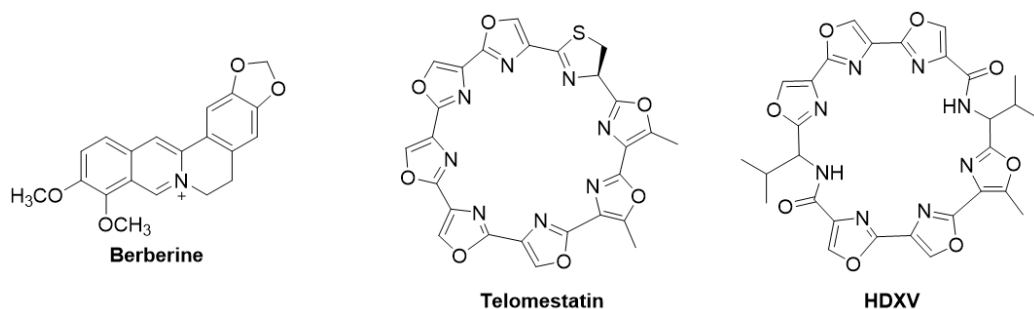
Figure 1.15. examples from the natural derivatives as G4-ligands.

c. Synthetic macrocycles based on quinacridine,[127] oxazole[128] and quinolone[129] have been proposed as good G4-specific binders and potent telomerase-inhibitory activity in both cell-free and cell-based systems assays (figure **1.16**).[130] Acyclic compounds tend to be flexible but rapidly adopt a hallmark horseshoe-shape which mimics the macrocyclic structure of Telomestatin, and binds the terminal G-quartet by π-π interactions. For instance, this occurs with PhenDC$_3$ and c-myc quadruplex according to NMR studies.[131] PhenDC$_3$ structure has a central phenantroline flanked symmetrically by two quinolinium groups (cationic), and has a slight propeller twist at equilibrium, complementing the guanines quartet. The *meta-* substitution on the phenyl rings assures the selectivity to G4 due to the consequent change to a horseshoe-like conformation, in contrast to *para-* derivatives, which are non-selective DNA minor groove binders.[132] With the same acyclic ligand configuration approach, five membered oxazole or triazole rings have been described, such as TOxaPy which discerns between different G4.[133] Pyridostatin is a selective G4-binder based on the N,N'-bis(quinolinyl)pyridine-2,6-dicarboxamide scaffold and produces DNA damage in high PQS-concentration sites.[134] It adopts a flat but flexible conformation provided by an internal hydrogen bond network which assistance the settlement into the quadruplex structure. It poses an optimal electronic density on the aromatic surface to interact with the G-quartet. In addition, the nitrogen lone pairs coordinate with a water molecule to sequester potassium in the center,[135] facilitating even more the interaction between both entities. All these features make it a strong and selective G4-binder. The same group developed pyridostatin analogs with better G4-stabilization properties and selective growth arrest in cancer cell lines versus normal cell lines.[136]
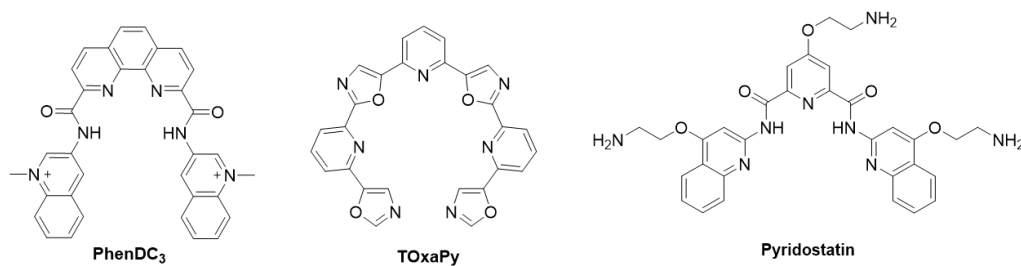
Figure 1.16. Some examples of synthetic macrocycles.

d. Perylene derivatives such as perylenedicarboximides do have the characteristic G4-ligand presentation: big aromatic core (tri-, tetra- or hepta-cyclic) and different length basic side chains (figure **1.17**). They can induce G4 structures and also inhibit the telomerase enzyme.[137–139] The most representative family member are PIPER (*N,N'*-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide) derivatives, pointing the importance between the charged nitrogen atoms in the side chain and the aromatic core:[140] too short side chains and/or weak base amine motifs drive to worse inhibitory ability. Other researchers developed more hydrosoluble

perylenes[141] and coronenes (CORON)[142] derivatives with three or four side chains and several basic motifs, with better G4-stabilization and telomerase inhibition properties. Substituted **naphthalene diimides** family shows very high affinity to quadruplex,[143,144] and crystal structures of the hTel with five NDI derivatives are available.[182,183] A chapter of this thesis is focused on this G4-ligands family, thus they will be described later in more detail.
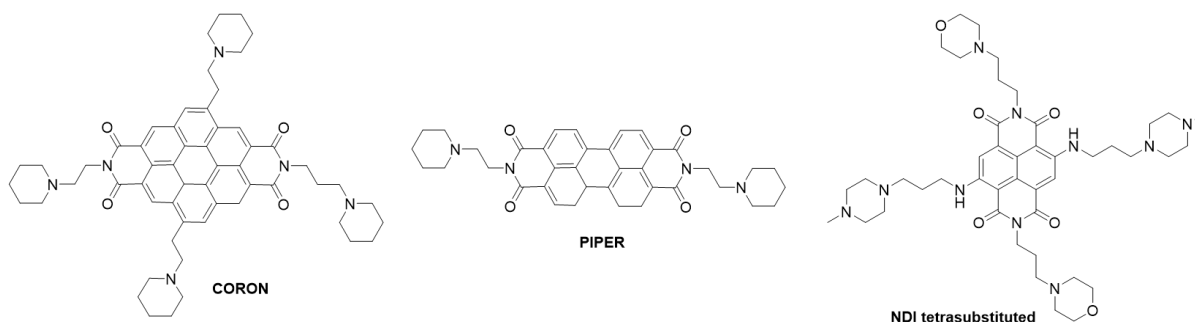
Figure 1.17. Perylene derivatives

e. Another family include **metal complexes**, extended polycyclic aromatic compounds with a coordinated metal cation (commonly Pt^{II}, Mn^{II}, Cu^{II} or Zn^{II}). The metal organizes the polyaromatic in an optimal planar geometry in order to stack it on top of the terminal quartet.[147,148] To highlight, metallophorphyrines derivatives with TMPyP (meso-methylpyridinium-substituted porphyrin) as representation (figure **1.18**) shows high affinity but poor selectivity towards G4, binding to duplex DNA at the same time.[149] The addition of the G4-ligand TMPyP4 affects the expression of numerous genes,[150,151] possibly due to the RNA polymerase progress hampering by G4-ligand complex.

f. The synthetic compound **Quarfloxin** (CX-3543)[152] is a pentacyclic benzo[b]pyrido[3,2,1-kl] phenoxazine (figure **1.18**), originally designed to bind c-myc, but latter studies demonstrated it acts on ribosomal DNA (rDNA) G4-forming sequences. The authors suggested Quarfloxin inhibits rDNA-nucleolin interactions and therefore RNA polymerase I transcription. It reduced cell growth in MIA-PaCa-2 pancreatic cancer and minimized the tumor volume in MDA-MB-231 breast cancer xenograft models.[153] Its importance resides in being the first G4-ligand to progress to clinical phases. In phase I it was well-tolerated and in phase II was tested over carcinoid/neuroendocrine tumors. For further clinical phases, Tetragene Inc. has in-licensed it.

g. Some of the reported **groove binders** compounds have the particularity to bind through the G4-grooves instead of the typical π-π aromatic stacking. For instance, the peptidic oligoamide Distamycin (figure **1.18**) shows groove-binding according to the NMR results,[154] and similar binding has been assumed for the antibiotic Neomycin with telomeric G4,[155] although this type of interaction is not well-known.

24
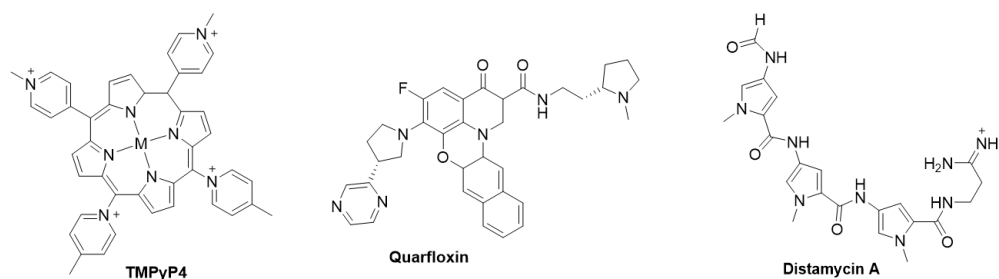
Figure 1.18. A metal complex, Quarfloxin and a groove binder

It exists the necessity to find new G4-binders with better drug-like properties. Several works have tried to find smaller molecules among libraries containing structurally varied compounds, but meeting the Lipinski Rule of Five. According to this idea, Rahman *et al.*[156] published the novel telomeric G4-ligands search using the NCI (National Cancer Institute) library, where they found some interesting compounds able to stabilize the human telomeric sequence (figure **1.19a**), screened by FRET melting assay. The same approached was used by Pagano *et al.*[157] with the evaluation of six drug-like molecules as selective G4-ligands, where they found two interesting compounds (figure **1.19b**) with G4-stabilizing properties and the capacity to induce telomere damage.
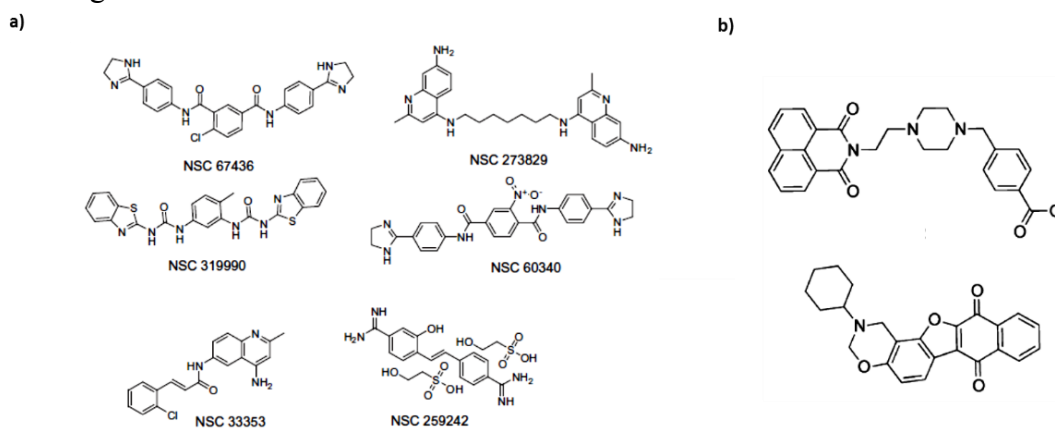
Figure 1.19. Some examples of drug-like compound evaluated as G4-binders. a) examples from Rahman report.[156] b) examples from Pagano report.[157]

1.3.3. SCREENING METHODS FOR G-QUADRUPLEX LIGANDS.

The first step within G4-drug development is studying the affinity and selectivity of the compounds towards quadruplex. When we have a large library of compounds to screen, we need fast and sensitive techniques such as FRET melting assay, CD, or SPR to evaluate them. Since each one has its limitations, it is recommended to use more than one. Once we choose the best candidates based on the previous techniques, the next step is to study the interactions with the corresponding G4 structures more deeply, and to examine their possible applications.

1.3.3.1. FRET (Fluorescence Resonance Energy Transfer) melting assay.

It is the first choice technique for simultaneous screening with many compounds and to minimize the amount of oligonucleotide to use at the same time. The FRET methodology, designed by Dr. Mergny's group,[158] is based on fluorophore modified G4s. The G4 probe consists on attaching FAM (fluorescein) as fluorescence donor and TAMRA (tetramethylrhodamine) as acceptor to the 5' and 3' sequence endings, respectively, through covalent linkage. When the quadruplex is folded both fluorophores are close together and TAMRA quenches the fluorescence transmitted by FAM. When it unfolds fluorophores separate increasing the fluorescence signal (figure **1.20a**). Oligonucleotide denaturation takes place by gradual temperature application, and the sample melting temperature (T_m) is obtained from the fluorescence melting curve (figure **1.20b**).

The experimental conditions were reviewed by De Cian *et al.*[159] such as buffer conditions, experimental settings and T_m calculation. Experiments are carried out on multiple well plates (commonly 96 or 384 well plate) with small sample volume per well (20-25 µL) and oligonucleotide amount (0.2 µM), which let us screen many compounds and sequences at the same time.
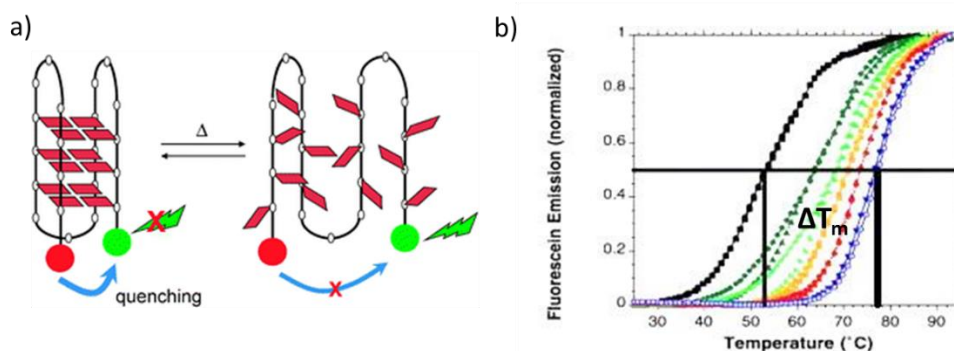
Figure 1.20. a) Schematic representation of the process. b) Normalized melting curves of F21T (human telomeric sequence) alone (black) or in presence of a neomycin-capped acridine at different concentrations (green, 0.5 µM; light green, 1 µM; orange, 2 µM; red, 4 µM; blue, 8 µM).

Compounds are added to the G4-sequence sample and the fluorescence at different temperatures is measured. Obtained T_m results are compared with the compound-free control curve and then subtracted to get ΔT_m values for each ligand-G4 interaction. So, T_m is obtained for each sample (T_mS) and the ΔT_m is calculated by applying the formula: $\Delta T_m = T_mS - T_mC$; where T_mC is the compound-free control T_m.

Despite of its previously described advantages, FRET melting assay has also some problems:

- When the screened compound is highly fluorescent it could interfere in the final results. Thus, we should choose another technique without fluorescence measurement, such as CD or SPR.
- When the compound presents more affinity to the fluorophores than to the quadruplex, false positives could appear, which increase the T_m without quadruplex stabilization. This is why it is convenient to check the obtained ΔT_m using another technique.

- We have to take into account the possibility by which fluorophores can affect quadruplex folding. Normally, the sequences are separated from the fluorophores by a few nucleotides in order to avoid this phenomenon, but it is advisable to compare the final G4 structure with and without the corresponding fluorophores.[160]

1.3.3.2. CD (Circular Dichroism) and CD melting assay.

This technique is based on the different left and right circularly polarized light absorption by chiral molecules. It gives secondary structure information regarding helices or sheet conformations of proteins. The CD characterization of quadruplexes was published by Karsisiotis *et al.*[161] They reported each structural group presents a typical CD signals pattern (figure **1.21**). The characteristic signals in the 210 to 340 nm wavelength range of the CD spectrum result from the π-π stacking interaction between the quartets and the glycosidic angle torsion (*syn* or *anti*). This distinctive signals are:[162]

- For parallel quadruplexes: a maximum at ~260 nm and a minimum about 240 nm due to the *syn·syn* or *anti·anti* glycosidic bond conformation (figure **1.21a**).
- For antiparallel quadruplexes: a maximum at ~290 and a minimum at ~260 nm due to the *syn·anti* or *anti·syn* conformations (figure **1.21b**).
- Parallel and antiparallel hybrid conformation give rise to the sum of both specific signals in the spectrum, including positive bands at 290 and 265 nm and a minimum at 240 nm (figure **1.21c**). An example is the human telomeric sequence in potassium buffer conditions.
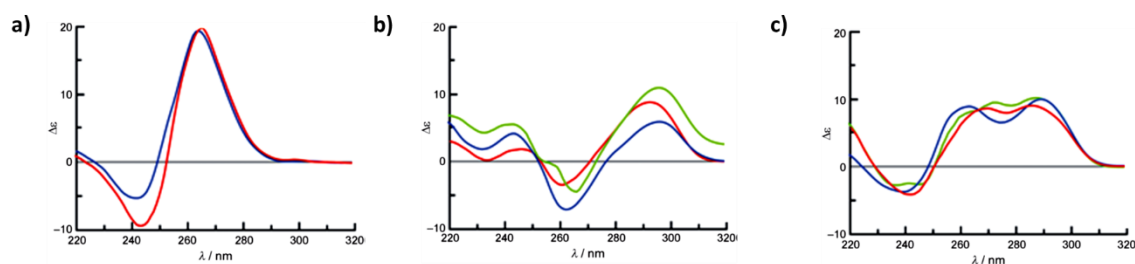
Figure 1.21. CD spectra for different quadruplex topologies. a) parallel strand quadruplex I-1XAV (red) and I-2O3M (blue). b) Hybrid quadruplex II-JSL (green), II-2GKU (red), II-11a (blue). c)antiparallel quadruplex III-201D (green), III-143D (blue), III-2KM3 (red).

For quadruplexes, circular dichroism is the simplest technique to determine the overall fold of a potential G4-forming sequence. It needs a small amount of DNA or RNA sample (3-10 µM) in the appropriate buffer. Its simplicity makes it one of the first techniques to apply when showing potential G4 structures, measuring the effect of compound addition, buffer modifications or cation changes on the G4-structure.

This technique allows to study the influence of the ligand on the global G4-structure following the evolution of the CD signals pattern. Larger sample amounts (about 3-10 µM sequence concentration) are needed than on FRET studies and simultaneous samples measurement is not possible. However, it provides additional information about interaction features, depending on the assay:

- **Titration experiment**: starting from the G4-folded CD spectrum, the ligand solution is added stepwise and the corresponding spectrum is recorded on each step. When the G4 specific signals intensity increases after ligand addition it denotes stabilization effect. By contrast, signal intensity decreasing reflects distortion on the G4-structure (figure **1.22a**).

- **Thermodynamic titration:** a large amount of ligand is added before oligonucleotide annealing, and then the CD spectrum is recorded. It allows to study the ligand effect over the G4-folding. When the G4-forming sequence is able to fold into more than one topology, such as the human telomeric sequence, it is interesting to see if the compound favors one of them.

- **CD melting assay:** based on the same FRET melting assay principle, we calculate the ΔT_m but monitoring a characteristic CD-signal disappearance *versus* temperature, from which the T_m value is extracted (figure **1.22b**).
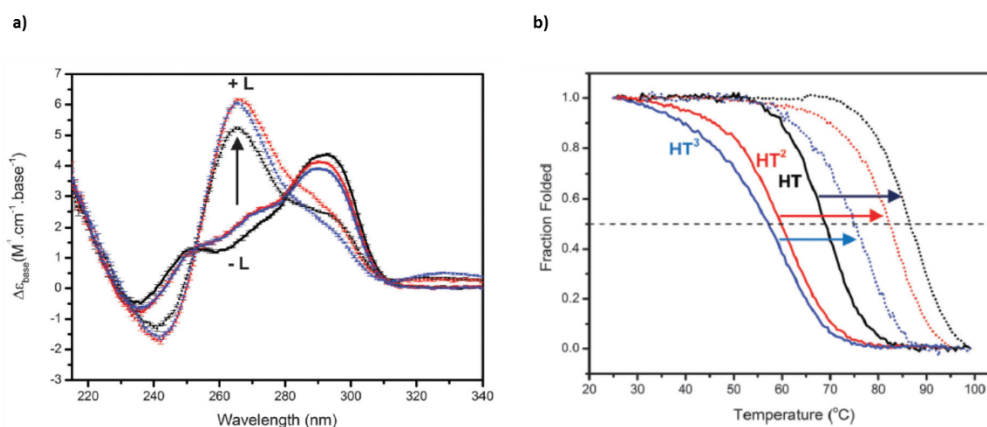
Figure 1.22. Examples of CD spectra. a) Normalized CD spectra of three quadruplex unbound (solid lines color coded) in 100 mM KCl, 10 mM K_2HPO_4/KH_2PO_4, pH 7 and in the presence of the ligand DR4-47 (dashed lines) at 1:5 ratios, showing a large ligand induced increase in ellipticity at 265 nm (arrow). b) CD thermal unfolding profiles (at 290 nm) for HT, HT2 and HT3 quadruplex (solid lines color coded) and shifts (shown by arrows) when bound to ligand RHPS4 (dotted lines) in 100 mM KCl, 10 mM K_2HPO_4/KH_2PO_4, pH 7 with two equivalents of ligand per quadruplex (HT=3.83 mM, HT2 = 1.53 mM, HT3 = 0.61 mM).[163]

With this technique we avoid the false positives found on the FRET melting assay, but some compounds can present CD signal at the measured DNA range. A control spectrum with the ligand in the same buffer conditions is recommended before carrying out the experiment to assure there is no ligand influence. Another limitation in the CD melting assay is the presence of G4-conformations mixture which complicates the T_m calculation.

1.3.3.3. UV-spectroscopy.

Nucleic acid stability in DNA and RNA sequences has been long studied by UV-spectrometry melting curve profiling. This technique has been successfully applied to quadruplexes which have given rise to distinctive melting curves. Specifically, the G4-melting profile is inverse respect to the DNA duplex one. Commonly for DNA or RNA duplexes, melting assays are carried out at 260 nm because a significant absorption difference between the folded and unfolded structures exists (25% difference) (figure **1.23a**). At this wavelength, G4 structural differences are less pronounced (4%), so the melting experiments are recorded at 290 nm where the difference in absorption of both arrangements is around 50% (figure **1.23b**).[164] Additionally,

melting curves are reversible where the reverse process of temperature gradient will provide the same curve.
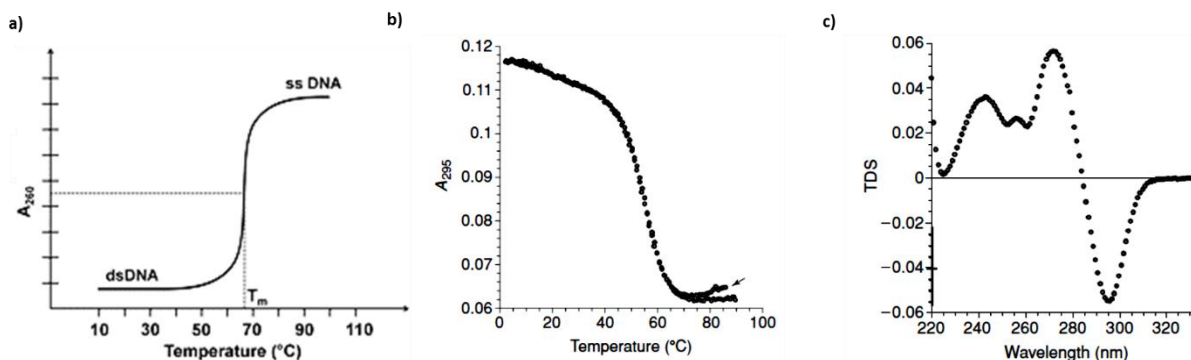
Figure 1.23. a). A typical DNA melting curve; the temperature at which half of the DNA is denatured is termed as the melting temperature (T_m) measured at 260 nm. b) A reversible thermal denaturation curve Cooling-heating profiles at 295 nm of a 3-μM sample of the modified vertebrate telomeric G-strand oligonucleotide, d(AG$_3$(T$_2$AG$_3$)$_3$) in 20 mM lithium cacodylate (pH 7.2) in the presence of 100 mM NaCl. The heating and cooling profiles are superimposable. The arrow indicates the reverse process. c) The thermal difference spectrum of the modified human telomeric sequence.[165]

This technique allows us to characterize the structure stability and kinetics by analyzing the effect of different buffers, cations and strand concentrations. Usually, cacodylate buffer[165] (pK$_a$= 6.14) is the recommended one for G4-melting studies (including FRET and CD melting assays as well) because:

-It does not absorb light in the studied UV-region.
-Its pK$_a$ is temperature-independent.
-It does not interact with other constituents in the sample.

Likewise FRET melting assay, UV-melting curves with and without the ligand provide the ligand ΔT_m. Within UV-melting curves, <u>TDS</u> (Thermal Difference Spectra) affords a specific signature related to the nucleic acid arrangement which complements other information acquired from other techniques. It is obtained by the subtraction of a spectrum recorded at 90 °C, a temperature above the T_m (unfolded structure), and a spectrum at 20 °C, a temperature below the T_m (folded structure). The typical tetrameric TDS spectrum shows a negative band at 295 nm (a consequence of G-tetrad stacking) and positive bands around 240 and 270 nm whilst parallel folding shows a less pronounced negative band than antiparallel ones (figure **1.23c**).[166]

1.3.3.4. <u>NMR spectroscopy.</u>

NMR spectroscopy has been commonly employed for small molecule characterization in organic synthesis, peptide and protein structure resolution and since 1971 also for DNA structural studies.[167]

The G-tetrad association displays characteristic signals in the 10-12 ppm chemical shift range generated by the imino protons,[168] in contrast to the 13-14 ppm signals of Watson-Crick hydrogen bonds.[169] Additionally, the proton exchange occurs more slowly in G4 than in the duplex structures. The number of imino protons matches the implicated guanines in the quartets. For instance, twelve sharp signals spawn in the sequence shown in figure **1.24** evidence a three tetrads

G4. In doing so, the number of imino peaks may not correspond to the theoretical expectations. This usually means that several conformations co-exist in the solution sample and this scenario complicates the structure resolution. To resolve such situation, it is possible to modify the sequence or sample conditions to favor a specific topology and eliminate the conformational mixture, simplifying structure resolution.[170] Additional bi-dimensional experiments, such as NOESY, HMQC, etc…, are required to assign the structure's proton residues, whose analysis allows the resolution of: structure, G-tetrad alignments, strand directionalities, loop connectivity and groove dimensions.
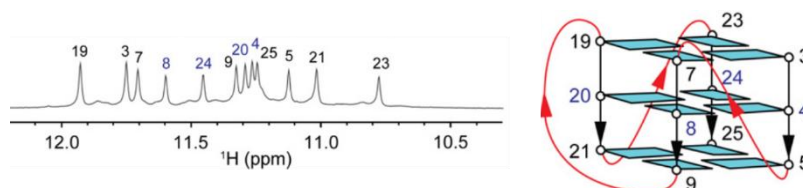
Figure 1.24. Left side: NMR imino region of the 26 mer CEB25 minisatellite G4 forming sequence. Right side: schematic presentation of the corresponding topology.[14] Numbers correspond to the base number in the oligonucleotide sequence.

NMR can give structural information on the ligand-G4 binding, pointing to the molecular interactions involved. The main disadvantage of NMR studies are the amount of sample and the high purity needed; DNA must be pure enough to allow NMR peaks assignment without impurities distortion. Nevertheless, its sensitivity let us detect even weak interactions and compounds with low affinity constant.

After NMR characterization of the targeted G4, including every assigned proton, we can run different types of experiments:

- **Titration experiments:** stepwise addition of the ligand allows to follow the changes on the chemical shift of the different ligands from the original G4-spectrum. We may identify the ligand binding site in some cases, as well as dissociation constant (K_d) for millimolar and micromolar ligand quantities. Simple $1D^1H$-NMR spectra are enough to appreciate chemical shift changes especially on the imino and the aromatic protons (the easiest to follow due to their well separated disposition), whilst the methyl group of the thymine bases changes in lower proportion and are more complicated to appreciate. As mentioned, imino protons are the most common signals affected, with a more intense shift when the ligand stacks on the G4-edge. In these experiments we can find two situations:
    - High affinity ligands: tend to provide a different set of signals which increases with ligand addition in detriment of the original ones. This happens progressively at target saturation states in a slow exchange on the chemical shift time scale. Dystamicin is an example of this situation (figure **1.25a**).[171]
    - Low affinity ligands: DNA signals shift with ligand addition which suggests the formation of short-lived complex on the NMR scale. Example in figure **1.25b**.
- **Full structural resolution:** More complex NMR-experiments (2D-NOESY, HMQC, COSY, etc…) are carried out. Higher concentration of the sample is normally needed (~1 mM DNA). Partial or total resonance assignment identifies the involved ligand-G4 interacting residues. G4 size and complexity is directly proportional to the difficulty on its resolution. However, if we

are able to identify most of the atom-scale structural changes, a 3D approximation of the interactions can be reached. To do so, it is possible to use the 2D-spectra positional restrictions found in the NOESY experiments to perform the computational calculations of the complex structure.
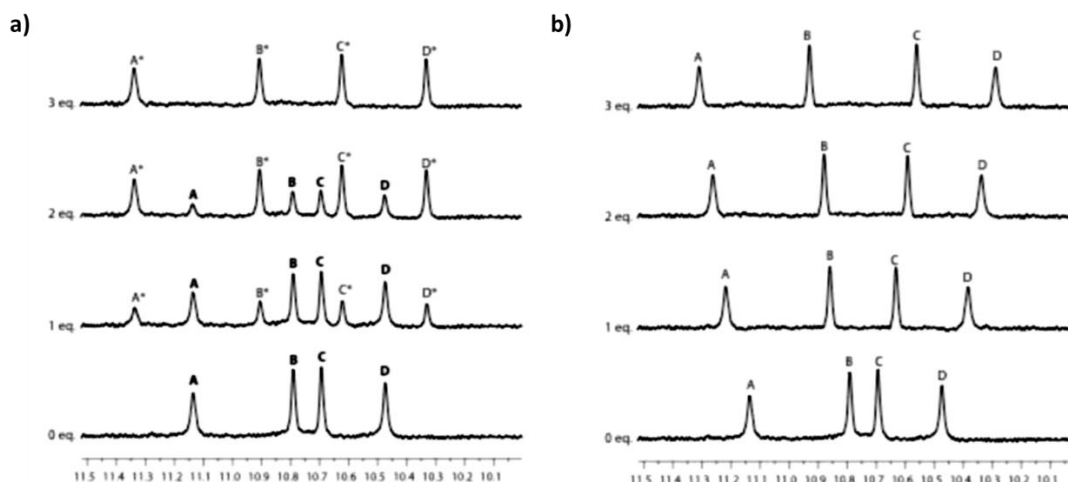
Figure 1.25. Simulation of two different NMR titrations. a) example of binding with high affinity, where original signals of the target (bold letters) are gradually replaced by new ones (letter marked with asterisk). b) example of binding with low affinity where the NMR signals of the target drift upon addition of the ligand.

Only a few structures of G4-ligand complexes have been solved[172–174] due to the technical problems such as difficulty in obtaining well-defined complexes, molecular aggregation and the lack of intermolecular NOEs.

1.3.3.5. Mass spectrometry.

Electrospray ionization-mass spectrometry is a sensitive method to detect non-covalent interactions, since it allows the formation of the complex in the gas phase. This technique allows the determination of structures and the stoichiometry of complexes even when constituted by supramolecular components. G-quadruplexes are an example of structures that can be analyzed using this methodology. This is achieved by gasification and ionization of samples which form charged fractions that can later be detected and analyzed to obtain their mass-to-charge ratios (m/z). To do so, the negative environment generated by the G4 phosphate backbones in solution is accommodated by negative ion generation and detection, allowing the elucidation of the number of quadruplex strands and the cation number involved in stabilizing its interactions (example in figure **1.26**).
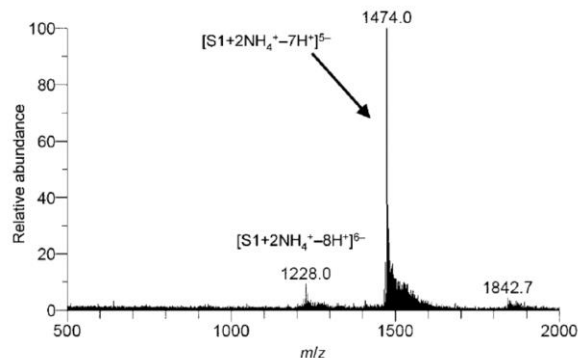
Figure 1.26. ESI mass spectrum of the bcl-2 G-rich sequence d(G$_3$CGCG$_3$AG$_2$A$_2$G$_5$CG$_3$) in 20 mM NH$_4$OAc buffer (pH 7.0).[175]

Electrospray spectrometry is combined with ion-mobility chromatography to improve G4 studies. Experimental conditions are crucial to keep the complex stability during gassing phase where ammonium cations are used instead of the common sodium and potassium ions. The reason resides in the method's low tolerance of salt concentration. Due to adducts generation by counter-ion condensation during nucleic acids droplet evaporation [DNA/RNA+nNa$^+$/K$^+$] experimental sensibility decreases. Hence, ammonium and trimethylammonium acetate (TMAA) are used as electrolytes to avoid the adducts. NH$_4^+$ and PO$_3^-$ proton transfection after solvent evaporation makes ions non coordinated with G-tetrads evaporate as NH$_3$. They influence the G4-topology, for instance, ammonium buffers give rise to an intermediate topology between that observed in Na$^+$ and K$^+$ in the human telomeric sequence. The ammonium cation intercalates between G-quartets, so the number of G4-quartet present can be calculated with the formula: N° G4-Quartet = N° cation +1. At 150 mM ammonium acetate, just a small fraction of the phosphate groups remain negatively charged.[176]

Electrospray ionization-mass spectrometry provides the determination for the mass, the interaction energy, as well as the stoichiometry of the complex. It has been applied to the study of both intra- and intermolecular quadruplex and their interaction with a variety of ligands. The sample is prepared by mixing the annealed DNA solution with the ligand, which is injected in the mass spectrometer via the electrospray source. Ions are analyzed according to their mass-to-charge ratio using the time of flight analyzer and each *m/z* number of ions is counted on the detector. In figure **1.27**, the typical electrospray mass spectrum of a DNA-ligand mixture is exposed, which shows three different mass species corresponding to free DNA (D), 1:1 and 2:1 ligand-DNA complexes (DL and DL2), each at three different charge states (z =6, 5 and 4).[177]
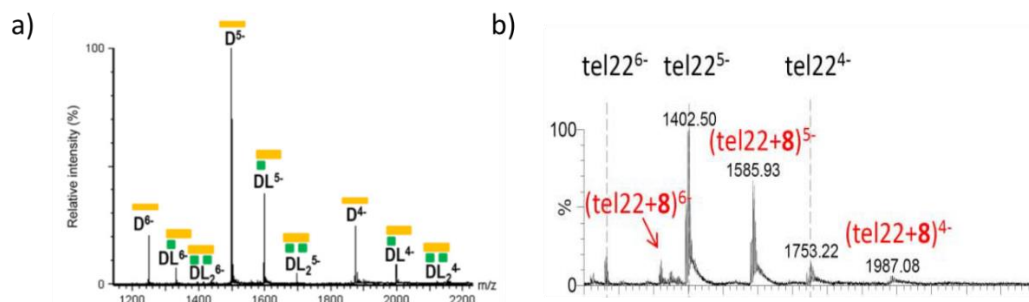
Figure 1.27.a)typical mass spectrum. b) Example from a BRACO derivative.[178]Representative ESI-MS spectra of mixtures between 5 μM telomeric Gquadruplex dAGGG(TTAGGG)₃ (tel22) and 5 μM the ligand 8.

The advantages of the mass spectrometry are its high sensitivity, accuracy, low sample consumption and rapidity. It gives one signal for each different mass species, so the stoichiometry of each existing complex is detected. In addition, from each mass peak, it is possible to determine the number of DNA strands, cations and ligand number involved in the complex. This method is not only qualitative, it also provides quantitative information about the complexes, such as concentration, from the relative intensities of mass spectral peaks, and equilibrium binding constants determination. Competition assays with different G4 or DNA samples are available to determine the ligand selectivity as well.

Many G4-ligands have been assayed through this technique, for instance the planar ligands PIPER, TMPyP4 and quinacridines,[179] porphyrins,[180] Telomestatin,[181] or BRACO derivatives.[178]

1.3.3.6.   SPR (Surface Plasmon Resonance).

We extend the binding study to bimolecular complexes kinetic-behavior with SPR, which provides the complex dissociation equilibrium constant (K_d). This technique is based on the detection of changes of the reflecting index at the sensor surface. The equipment detects a mass-change on the surface due to the binding of the ligand to the attached G-quadruplex. The ligand solution is injected within the appropriate running buffer over the 5'-biotinylated sequence, which is attached to the surface through a Streptavidin linkage. That change is transformed into a photoelectric response (RU: response unit) *versus* time. Binding leads to signal increment from the background, whilst complex dissociation causes RU descent. The curve shape gives information about the binding features:

❖ Fast dissociation gives rise to sharp curves, with short equilibrium states, characteristic of low-affinity binders.
❖ Strong binders produce slow dissociations with long equilibrium states.

In figure **1.28** some examples are represented.
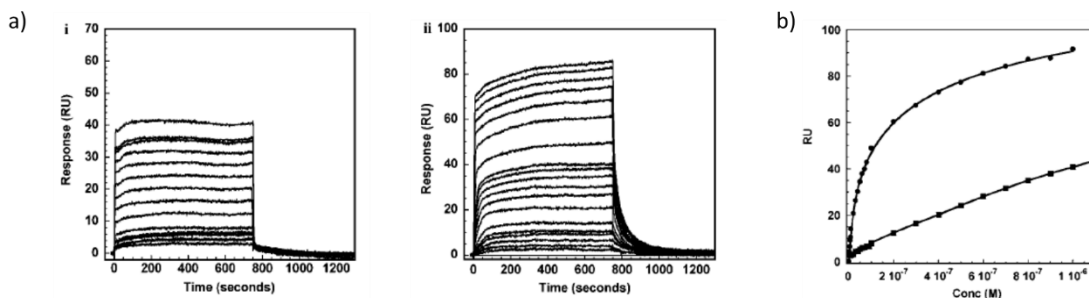
Figure 1.28. SPR experiments with a G4-ligand. a) Representative sensorgrams for compound binding to (i) duplex (CG eight base pair hairpin d[(CG)$_8$]) and (ii) quadruplex (human telomeric d[AG$_3$(T$_2$AG$_3$)$_3$]) DNA sequences. The lines are best fits to the steady-state RU (response units) values, which are directly proportional to the amount of bound compound. (b) Direct binding plot for the compound of RU versus concentration of the unbound: (●) quadruplex; (▪) duplex. The lines are obtained by nonlinear least-squares fits of the data.[182] The duplex line shape presents a non-specific binding behavior while the quadruplex is specific.

Titration assays let us determine the K$_d$ value. The RU in equilibrium obtained from each ligand concentration provides the titration curve. The curves are fitted through the application of different equations, which rely on the ligand behavior. The curve shape is fitted in a different way whether binding is specific or non-specific (see figure **1.28b**).

Similarly to the FRET melting assay, SPR is a fast, accurate and high-throughput method. It has been thoroughly applied in protein-antibody and protein-ligand complexes. Nevertheless, the application to this field is not completely optimized, despite several works have reported its use for quadruplex ligand screening.

2.   CARBOHYDRATES IN BIOLOGY.

2.1.   Composition and biological functions.

Chemically, carbohydrates are polyhydroxy aldehydes or ketones with five (pentoses), six (hexoses), seven (heptoses) or eight (octoses) carbon atoms. They can also be classified depending on the number of units, where monosaccharides have one unit, disaccharides two, oligosaccharides several units and polysaccharides have many units forming complex systems. Additionally to this complexity, it is possible to find monosaccharides with identical chemical formula but different spatial configuration and therefore different global properties. Hence, configurational and conformational isomerism play an important role in carbohydrates.

The most abundant pentoses in biological systems are L-arabinose, D-ribose, 2-deoxy-D-ribose, and D-xylose. For hexoses, D-glucose, D-fructose, D-galactose, and D-mannose are the most common; and maltose (α-glc (1→4)-glc), lactose (β-gal (1→4)-glc) and cellobiose (β-glc (1→4)-glc) among disaccharides (figure **1.29**).
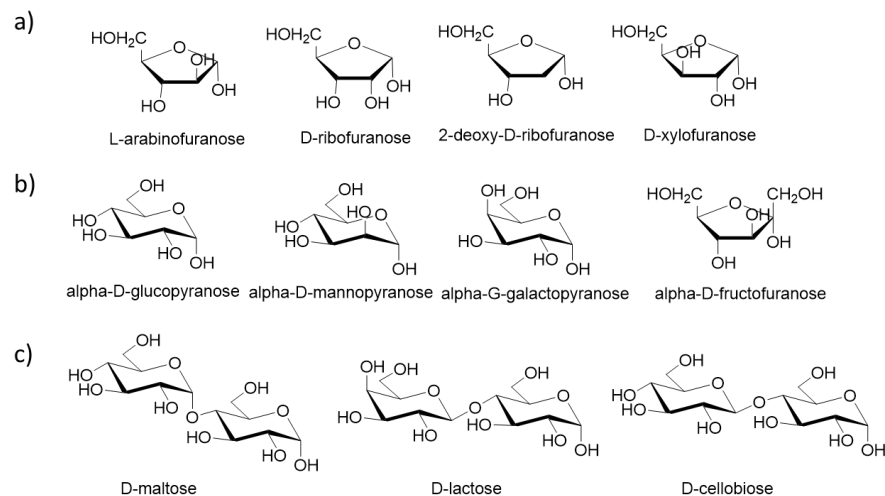
Figure 1.29. Most common oligosaccharides: a) pentoses b) hexoses c) disaccharides

Together with proteins, lipids and nucleic acids, carbohydrates are essential biomolecules. They perform important roles by themselves or conjugated to the other components such as glycoproteins and glycolipids. Actually, glycosylation is one of the most ubiquitous posttranslational modifications in cells to produce the glycoconjugates. By themselves, they constitute the main energy source for living organisms and also display structural functions such as polysaccharides building blocks cellulose and glycogen. In the case of glycoproteins and glycolipids, they are involved in molecular recognition in the cell surface polyproteins and in a large number of signaling processes, and tasks involved in immune reactions. It would be hard to summarize in this work all the extensive functions they undertake, so as examples we will mention carbohydrate based matrix cell interactions, cell-cell adhesions and viral invasions.

2.2. Carbohydrate cell-uptake

Glucose hydrophilicity obstructs its translocation through the lipid membrane, so transport needs to be facilitated by proteins. Two are the responsible transport proteins for glucose cell uptake: the sodium-dependent glucose transporter (SGLT) family and the GLUTs family, summarized in table **1.1**.[183,184]

- <u>SGLT (sodium-glucose co-transporter) family</u>: these are proteins tasked with fast glucose absorption through active transport using the electrochemical potential difference gain as energy source. When sodium is transported in favor of its concentration gradient, it can modify SGLT conformation and enable glucose translocation (via *symport*). Between the six described isoforms, **SGLT-1**, **SGLT-2** and **SGLT-3** are the most known. SGLT-1 is the glucose and galactose essential absorption transporter in the intestine and imports 2 Na/glucose. The glucose specific transporter SGLT-2 is used to reabsorb around 90 % of the urine-dissolved glucose at renal tubules epithelia and imports 1 Na/glucose, leaving the rest of the reabsorption to SGLT-1.

- <u>Superfamily GLUT</u>: these are proteins which uptake glucose through facilitated diffusion in favor of its concentration gradient. 14 isoforms exist classified in three different families:

  o **Class I family**: constituted by GLUTs 1,2,3 and 4 isoforms. **GLUT-1** is a high glucose affinity transporter, being mainly located at high glucose consumption tissues and glucose distribution such as erythrocytes and brain, although it is thoroughly distributed in the body. **GLUT-2** shows low glucose affinity and it is found in liver, kidney, small intestine and pancreas. It is very sensitive to glucose concentration in the media, involved in blood glucose concentration regulation. GLUT-2 recognizes also fructose in lower proportion. **GLUT-3** is the transporter with the largest affinity for glucose and commonly co-located with GLUT-1. **GLUT-4** shows high glucose affinity as well, and is located mostly in muscular cardiac and muscular tissues, and adipocytes. They response at insulin signal, regulating glucose metabolism in muscle.

  o **Class II family**: GLUTs 5,7,9 and 11 isoforms are included. **GLUT-5** is a specific fructose transporter, mainly expressed in small intestine. Low levels are found in erythrocytes, kidney, spermatozoon, skeletal muscle and adipocyte tissue. **GLUT-7** has been located in liver. **GLUT-9** is expressed overall in kidney and liver, and less in small intestine, lung and leucocytes. **GLUT-11** is a fructose transporter with low glucose affinity, and sited in skeletal and cardiac muscle.

  o **Class III family**: GLUTs 6,8,10,12 and 13 isoforms complete the family. **GLUT-6** has low glucose affinity. **GLUT-8** has been located in testicle, blastocysts and brain, and in lower amount in spleen, prostate, small intestine, heart and skeletal muscle. It shows high affinity to 2-deoxy-glucose, as well as **GLUT-10**. GLUT-10 is distributed in liver, pancreas, cardiac muscle, lung, brain, skeletal muscle and kidney. **GLUT-12** was found in MCF-7 cell line (breast cancer), in addition to skeletal muscle, adipose tissue and small intestine. **GLUT-13** is a H/inositol transporter mainly expressed in glia cells and some neurons.

TRANSPORTER	SUBSTRATES	MAJOR SITES OF EXPRESSION
GLUT1	Glucose (K_m 3 mmol/L) Galactose (K_m 17 mmol/L) Mannose (K_m 20 mmol/L) Glucosamine (K_m 2.5 mmol/L) DHA (K_m 1.1 mmol/L)	Ubiquitous distribution
GLUT2	Glucose (K_m 17 mmol/L) Galactose (K_m 92 mmol/L) Mannose (K_m 125 mmol/L) Glucosamine (K_m 0.8 mmol/L) Fructose (K_m 76 mmol/L)	Kidney, small intestine, liver, pancreas
GLUT3	Glucose (K_m 1.4 mmol/L) Galactose (K_m 8.5 mmol/L) Mannose Xylose DHA	Brain, testis
GLUT4	Glucose (K_m 5 mmol/L) DHA (K_m 0.98 mmol/L) Glucosamine (K_m 3.9 mmol/L)	Muscle, fat, heart
GLUT5	Fructose (K_m 6 mmol/L)	Intestine, kidney, testis
GLUT6	Glucose	Spleen, brain, leukocytes
GLUT7	Glucose (K_m 0.3 mmol/L) Fructose (K_m 0.2 mmol/L)	Small intestine, colon, testis, prostate
GLUT8	Glucose (K_m 2 mmol/L)	Testis, brain, adrenal gland, liver, spleen, fat, lung
GLUT9	Glucose (K_m 0.61 mmol/L) Fructose (K_m 0.42 mmol/L) Urate (K_m 0.9 mmol/L)	Liver, kidney, placenta, pancreas
GLUT10	2-deoxy-D-glucose (K_m 0.3 mmol/L) Galactose	Heart, lung, brain, liver, muscle, pancreas, placenta, kidney
GLUT11	Glucose (K_m 0.16 mmol/L) Fructose (K_m 0.16 mmol/L)	GLUT11-A- heart, muscle, kidney GLUT11-B- placenta, fat, kidney GLUT11-C-heart, muscle, fat, pancreas
GLUT12	Glucose	Muscle, heart, fat, small intestine, prostate, placenta
HMIT	Myoinositol (K_m 0.1 mmol/L)	Brain
GLUT14	N/A	Testis
SGLT1	Glucose and galactose	Intestine, kidney
SGLT2	Glucose	Kidney, lower in liver, thyroid, muscle and heart.
SGLT3	Glucose	Muscle, lung, uterus, testis, intestine.

Table 1.1. Monosaccharide membrane transporter proteins with the main substrates and locations.

2.3. Warburg effect in cancer cells.

In tumorigenic tissues, oncogenic mutations and epigenetic changes on the cell cycle control mechanisms provoke exorbitant capacity for proliferation, one of the hallmark of cancer. Such proliferation rate requires vast amounts of energy consumption which induce cells to alter their machinery to increase sugar provision, the main source of energy.

In <u>aerobic conditions</u>, normal tissues metabolize the glucose by a highly regulated process. Starting from glucose hydrolysis, pyruvate is obtained to be incorporated into the Krebs Cycle to produce NADH, the necessary fuel for the later oxidative phosphorylation to maximize ATP production. The regulation takes place on each enzyme involved in the process by different mechanisms including external signals of glucose availability or oxygen conditions. 30-32 ATP units, NADH, $FADH_2$, 2 GTP, ROS (reactive oxygen species), CO_2 and H_2O are generated per glucose by this pathway. On the contrary, under <u>anaerobic conditions</u> such pyruvate is simply metabolized to lactate by lactate dehydrogenase producing 2 ATP/glucose, by a less energy-efficient process (figure **1.30**).
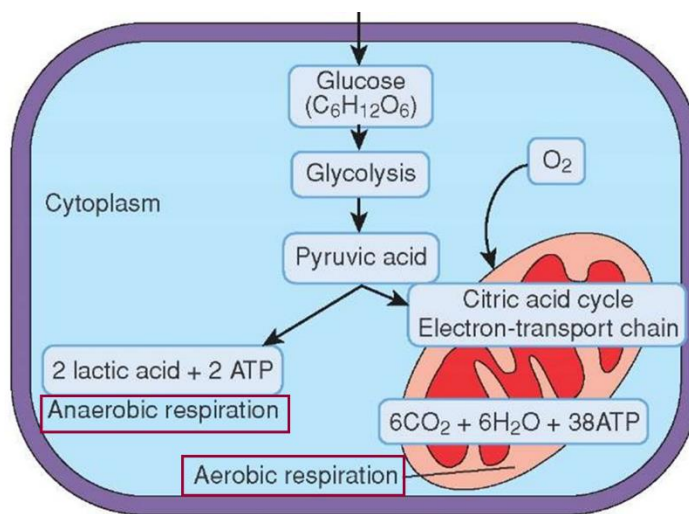

Figure 1.30. Glycolysis at aerobic and anaerobic condition in normal cells.

Otto Warburg observed that cancer cells metabolize glucose in a different pathway, since they tend to "ferment" glucose into lactate despite of the oxygen accessibility to support oxidative phosphorylation (**aerobic glycolysis**).[185] It could seem contradictory cause they use a less efficient metabolism even though they require more energy to support the high division rates, but some explanations have been proposed for this paradox:

- One possible explanation is that inefficient ATP production would be a problem only when sources are limited, which it is not the cancer cells case. The nutrients supply increase in circulating blood and modification of all glucose utilization machinery increases its uptake.
- Another possible explanation is that tumor cells need biomass production to divide. If all glucose is metabolized upon CO_2 production in oxidative phosphorylation they do not have material available to build macromolecules. From glucose and glucosamine

aerobic glycolysis they obtain lactate and alanine to synthesize the rest of essential constituents.

- In addition, excess metabolized glucose by mitochondrial phosphorylation would lead to abundant ROS production, highly harmful for cell survival.

Cancer cells change their glucose metabolism through different mechanisms including oncogenes activation and regulation of transcription factors (see reviews[186,187]) (figure **1.31**):

- Via **AKT oncogene**, the phosphoinositide 3-kinase (**PI3K**) signaling pathway is involved in both glucose metabolism and growth control, since it regulates glucose uptake and its utilization through transporter expression regulation, glucose capture enhancement by hexokinase and stimulate phosphofructokinase activity.[188] GLUT transporters and hexokinase are the main two proteins associated with increase glucose uptake in cancer cells.

- **C-MYC oncogene** activates glycolytic enzymes such as lactate dehydrogenase[189] and interacts with HIF-1 (hypoxia inducible factor) which indirectly activates p21^{cip1}, a key cyclin-dependent kinase inhibitor. When myc is overexpressed, HIF induces Hexokinase-2 and pyruvate dehydrogenase kinase transcription resulting in increased aerobic glycolysis.[190,191]

- The **P53 tumor suppressor** controls metabolic genes and alters glucose metabolism by induction of TIGAR, which inhibits phosphofructokinase redirecting the glucose to the pentose phosphate pathway and NADPH production.[192] This mechanism is an intracellular defense against ROS, hence NADPH reduces glutathione, the ROS reducing agent.

- **Mutated Krebs cycle enzymes** such as succinate dehydrogenase and fumarate dehydratase activate HIF1α which regulates a pleiotropic response involved in the altered expression of 450 genes including the glycolytic ones.[193]

- Many oncogenes are **tyrosine kinases** signaling associated with cell proliferation, characterized by glucose metabolism regulation. PK-M2 pyruvate kinase stimulate glycolytic pathway and the glucose utilization into biomass production.[194]
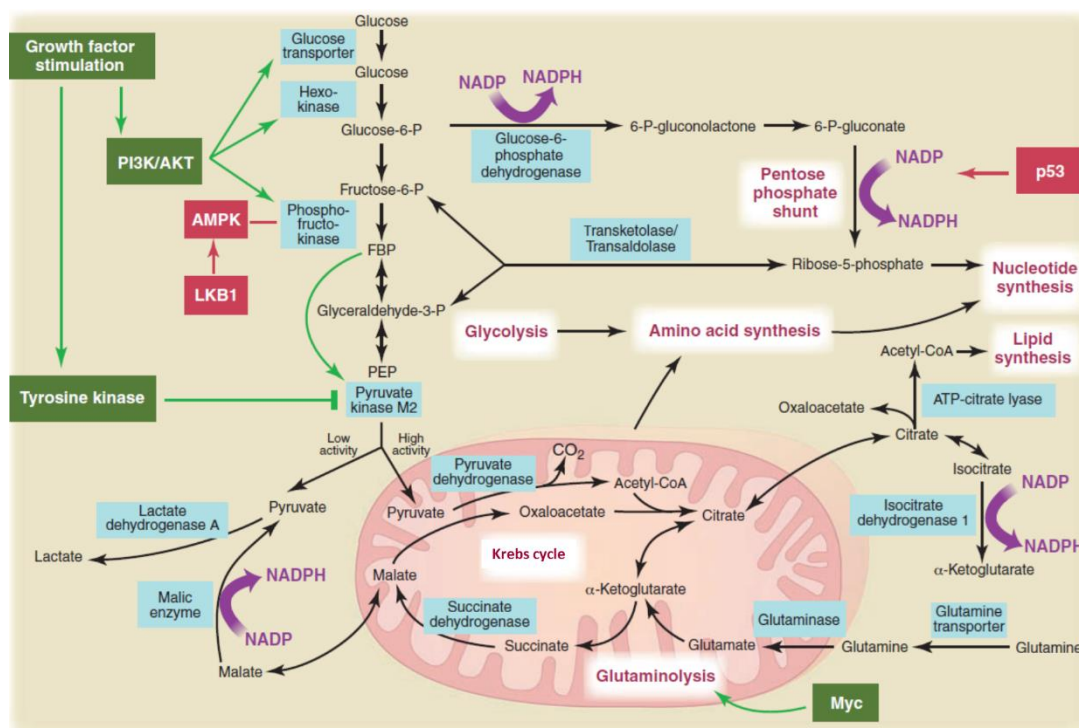
Figure 1.31.Metabolic pathway activated in cancer cell lines.

The above pathways normally decrease the cell response mechanism to external regulation signals, such as growth factors that regulates the proliferation behavior in normal cells. Therefore, the organism is not able to regulate such cells and they grow uncontrollably. As result, glucose uptake and consumption increase by 200-fold supporting the fast cell proliferation rates in malignant cells and inducing cancer progression.

Targeting cancer is always a challenge. The searching for specific treatments still continues since current therapies lead in many cases to severe side effects or are not effective at all. Regarding particular cancer cells behavior such as altered glucose metabolism may be a potential strategy to achieve selective tumor cells death.

2.4. Overexpressed glucose transporter in cancer.

We mentioned that cancer cells enhance glucose source to support their high proliferation rate and glucose transporters overexpression is one of the developed modification to increase such sustenance. Glucose transporter overexpression has been reported for some cancer types, including SGLT and GLUT transporters (table **1.2**):[195]

SGLT-1 has been found overexpressed in colorectal, head and neck, prostate tumors, primary tumors, and metastatic lesions of lung, and pancreatic adenocarcinomas. EGFR (epidermal growth factor receptor) associates with SGLT-1 and stabilize it. **SGLT2** is expressed in colorectal, gastrointestinal, head and neck, and kidney tumors as well as in carcinomas and leukaemia. It was found in primary tumors and metastatic lesions of lung, pancreatic adenocarcinomas, and head and neck cancers.[196,197]

GLUT1 location is ubiquitous in normal tissues, then it is not surprising it is overexpressed in many tumors: hepatic, pancreatic, breast, esophageal, brain, renal, lung, cutaneous, colorectal, endometrial, ovarian, and cervical cancers.[198] Increased expression of GLUT1 has been shown to be correlated with a poor prognosis in a variety of tumors.[199] **GLUT2** is highly expressed in gastric tumors, breast, colon cancers and liver carcinoma.[200,201] In pancreatic cancers, GLUT 2 was detectable in 75% of human PanINs (pancreatic intraepithelial neoplasia).[202] **GLUT3** is induced in many cancer lines as well as under hypoxia and has been found expressed in all the melanocytic lesions from melanocytic nevi, Spitz nevi, and primary cutaneous malignant melanomas, both benign and malignant.[203] Its expression was observed in non-Hodgkin lymphoma samples also, where GLUT1 was not found.[204]Furthermore, it has been highly located in brain, breast, gastric, head and neck, lung, meningiomas, oral squamous cell carcinoma, and ovarian. **GLUT4** is associated to breast[205] and gastric cancers.[200]

With **GLUT-5** contradictory results are observed, some reports indicate their low expression in breast cancer[206] whilst another defend a high expression in comparison with normal breast tissue, which may indicate that human breast cancer cells have a specialized capacity to transport fructose.[207] According to some studies, **GLUT12** was overexpressed in prostate adenocarcinomas, with low GLUT1 expression[208] and breast cancer cell line MCF-7.

Transporter	Tissue overexpression
GLUT1	Various tumors
GLUT2	Liver, breast, pancreatic, colon and gastric carcinoma
GLUT3	Lung, brain, breast, bladder, laryngeal, prostate, gastric, head and neck, ovarian and oral squamous carcinoma
GLUT4	Colon, lymphoid, breast, thyroid, pancreatic and gastric carcinoma
GLUT5	Breast, renal, colon, liver, testicular and lymphoid carcinoma
GLUT6	Breast, pancreatic and endometrial carcinoma, uterine leiomyoma
GLUT8	Endometrial and lymphoid carcinoma, multiple myeloma
GLUT9	Liver, lung, skin, thyroid, kidney, adrenal, testicular and prostate carcinoma
GLUT10	No info available
GLUT11	Multiple myeloma, prostate carcinoma
GLUT12	Breast, prostate, lung and colorectal carcinoma, rhabdomyosarcoma, oligodendroglioma, oligoastrocytoma, astrocytoma
HMIT	No info available
GLUT14	No info available

Table 1.2. Summarized GLUT transporters and tissues where they were found overexpressed.[184]

2.5. Carbohydrate-drug conjugates

Carbohydrate-containing drugs have been designed as therapeutic agents against different pathologies trying to take advantage of a facilitated cell entrance through GLUT proteins and to improve drug solubility or toxicity. This approach has been used with several anticancer drugs (figure **1.32**) since cancer cells overexpress GLUT transporters to favor glucose uptake.
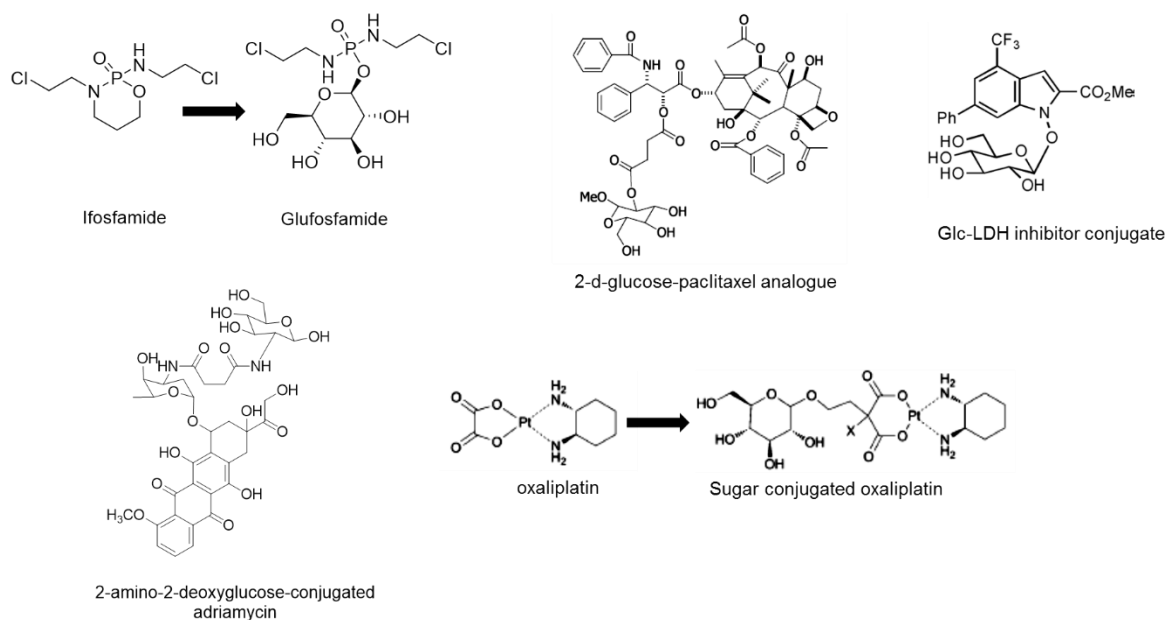
Figure 1.32. Examples of glucose-drug conjugates.

This concept was firstly applied by Pohl *et al.*[209] with the development of **glufosfamide**, a glucose derivative of ifosfamide mustard. Glufosfamide showed lower toxicity and better tolerance than its aglycone predecessor on tumor-bearing animals with drug accumulation increase in cancer tissues. The authors proposed glufosfamide facilitated cell entrance by GLUT1 as explanation of its cancer selectivity, since renal reabsorption of filtered drug could be blocked by pre- and cotreatment with phlorizin, a glucose transporter inhibitor. It has recently advanced to metastatic pancreatic cancer phase III clinical trials.

The same idea has been applied to some taxols such as Paclitaxel, a potent anticancer drug used for the treatment of breast, ovarian, and lung carcinomas, but with serial solubility problems. **2'-Paclitaxel-methyl-2-glucopyranose** conjugate improved pharmaceutical properties (by increased water solubility) and selectivity to MCF-7 breast cancer cell line without cytotoxic effects on non-tumor cell lines,[210] with the same GLUT favored uptake as potential selectivity mechanism.

Adriamycin is a potent antitumor agent in many types of solid tumors but its high cardiotoxicity and multidrug resistance problems limit its use. On the contrary, **Adriamycin** conjugated with **2-amino-2-deoxy-D-glucose** and succinic acid was reported to target tumor cells through GLUT1 transporters with better inhibition rate in tumor cells and lower toxicity than Adriamycin itself, even on animal experiments.[211]

More examples among **glucose-platinum conjugates** have been recently reported, created to avoid the typical neurotoxic effects of this kind of drugs. Oxaliplatin is especially potent in many of the cell lines and tumors that are known to be resistant to other platinum derivatives such as cisplatin or carboplatin. However, in addition to the neurotoxic side effects, oxaliplatin treatment is also accompanied by thrombocytopenia. The sugar conjugated (trans-R,R-cyclohexane-1,2-diamine)-malonatoplatinum(II) exhibited high water solubility, GLUTs

dependency and higher potent cytotoxicity in human cancer cells, with 9-30 times better therapeutic index and more favorable maximum tolerated dose than oxaliplatin.[212]

Patra *et al.*[213] developed a similar strategy with glucose-platinum conjugates going into detail about the cell entrance mechanism via glucose transporters through *in vitro* experiments. The same group studied how the position of substitution on the D-glucose affects the activity and GLUT1 selectivity in Glucose-Platinum conjugates (figure **1.33**).[214] The C1α- and C2-substituted Glc-Pts accumulated in cancer cells most efficiently and were more potent on DU145 cancer cells compared to the others, whereas the C3-Glc-Pt was the worst uptaken derivative. They also observed differences between the α- and β- isomers on cellular uptake and activity profiles. Nevertheless, the C2 derivative had the highest GLUT1 specific internalization, which was related to the best cancer targeting ability.
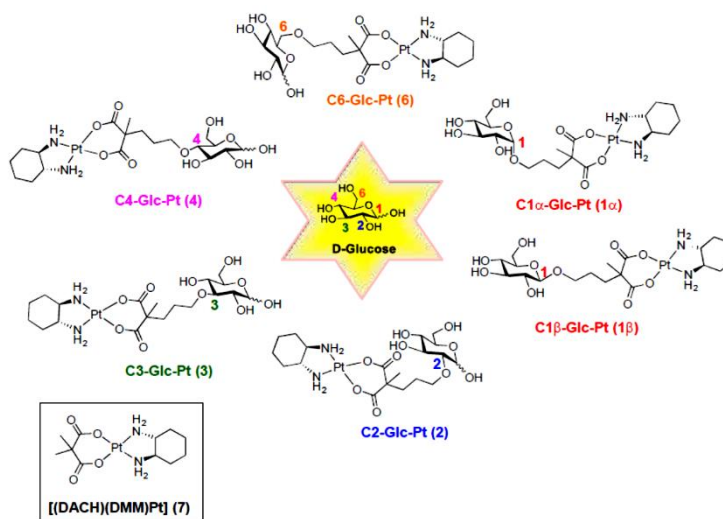
Figure 1.33. Structures of C1-C6-substituted positional isomers (1-6) of a glucose-platinum conjugate and aglycone 7.

All above strategies were based on the cancer selectivity by overexpressed GLUT transporters targeting but glucose metabolism is altered through more mechanisms in cancer cells. Another approach was explored by Calvaresi *et al.*,[215] who combined a **lactate dehydrogenase inhibitor** with a glucose to dually target the Warburg effect in tumor cells: enhancement of glucose uptake and increased glycolysis. The conjugate improved potency against cancer cells and showed increased cell permeability compared to its aglycone.

These examples suggest carbohydrate-conjugated drugs improve the original pharmacokinetics compounds properties as well as their selectivity diminishing harmful side effects.

2.6. G-quadruplex ligands incorporating carbohydrates.

Our research group have previously studied carbohydrate–DNA interactions in double helices and in G4 by using carbohydrate oligonucleotide conjugates. Highly polar mono- and disaccharides were capable of stacking onto the terminal DNA base pair of a duplex by CH/π

interactions as shown by NMR spectroscopy (figure **1.34**).[216] These results revealed carbohydrate-DNA base stacking as a potential recognition motif to be used in drug design. We also found that carbohydrate–thrombin-binding aptamer conjugates showed stacking interactions between the sugar and a DNA guanine tetrad together with hydrogen bonding and hydrophobic contacts.[217]These results pointed out to the potential use of mono and disaccharides as part of a drug interacting with G4-quadruplexes trying to improve binding affinity and selectivity.
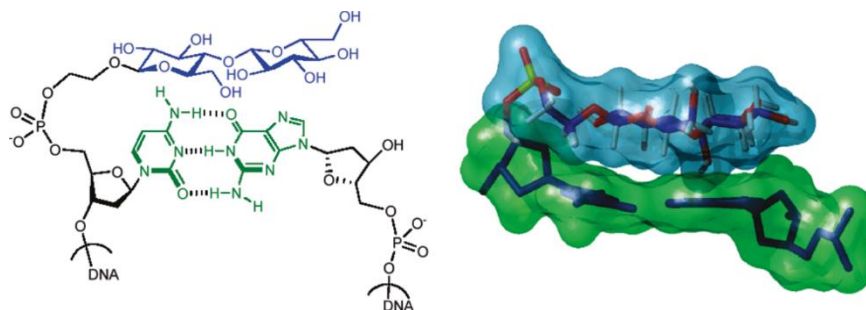
Figure 1.34. Disaccharide stablishing CH/π stacking onto the terminal DNA base pair of a duplex

G4-ligands modified with carbohydrates have been previously reported. Bugaut *et al.* [218] were the first one incorporating sugars on a G4-binder by preparation of a series of carbohydrates-based **oxazole–peptide macrocycle** containing α- and β- substituted derivatives of lyxose and xylose (figure **1.35a**). They showed how subtle changes, as single stereogenic center, can result in differences in affinity towards c-*kit* and c-*myc* G-quadruplex forming sequences, introducing the hypothesis of selectivity due to the carbohydrates motifs.
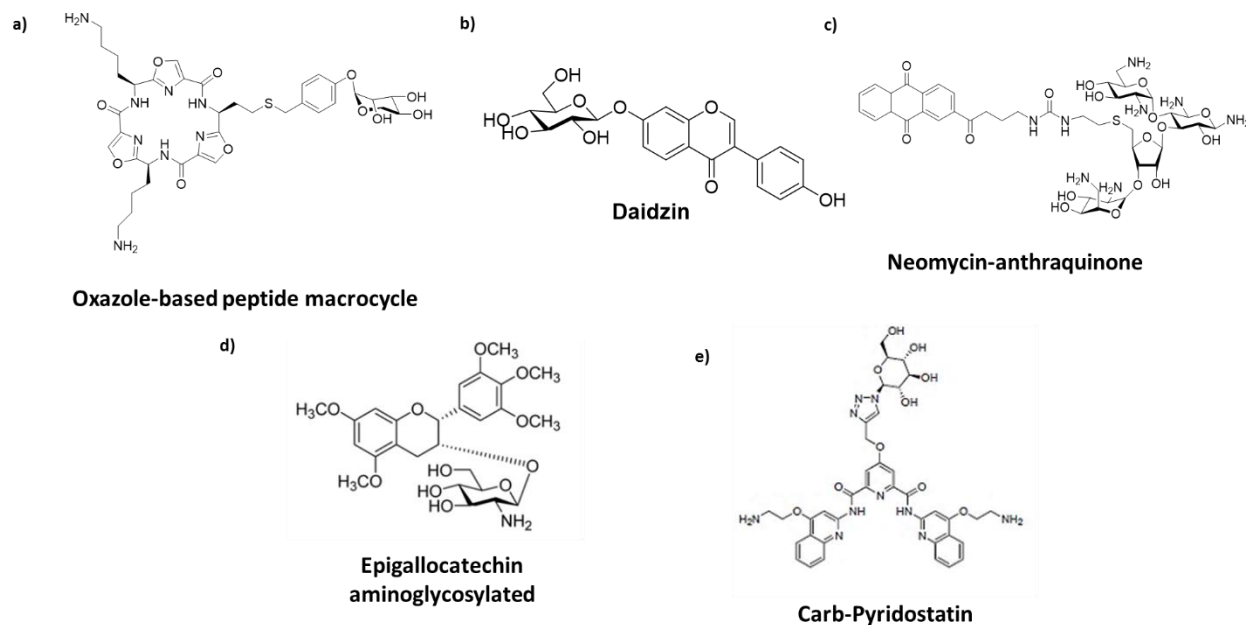

Figure 1.35. Several examples of carbohydrates-G4-ligands conjugates.

As natural product, isoflavones such as glycosidic **daidzin** (figure **1.35b**) or **genistin** were reported as G4-ligands capable of inducing a structural change of the polymorphic human telomeric (hTel) G4 into its antiparallel conformation while they destabilize duplex structure.[219]

The amino sugar **neomycin** was reported as G4-groove binder candidate due to its flexible nature when *Oxytricha nova* telomeric DNA was used as target.[220] Later, the same group created a perylene-neomycine conjugate with two neomycine units linked by a long chain to the perylene aromatic core. UV-thermal melting and absorption binding studies demonstrated a human telomeric binding via dual recognition mode with the perylene core stacking on top of the quadruplex.[221] Likewise, a neomycin-anthraquinone conjugate (figure **1.35c**) displayed double recognition of the hTel with improved binding features with respect to the individual components.[222] Penta-methylated **epigallocatechin** glycosylation with aminoglycosides (figure **1.35d**) was also described as stronger G4-binder than their precursor Epigallocatechin, with more affinity towards the human telomeric RNA-quadruplex versus the corresponding DNA-sequence. Molecular docking simulations located that glucosamine moiety resided in the G4-groove.[223]

The side chains of the potent G4-ligand **Pyridostatin** were modified by incorporating different moieties linked via click chemistry, and among them we find carbohydrates.[136] These carbohydrates-pyridostatin conjugates (figure **1.35e**) improved the human telomeric stabilization properties in comparison to the unsubstituted analogues. However, the cytotoxic results against tumor cancer cell lines were moderate, which authors explained by the differences in compounds cell permeability and uptake. The same ligands were synthesized via *in situ* click-chemistry, using a DNA-target mixture as catalyst in order to create more selective compounds.[224]

Recently, ***N*-phenanthroline glycosylamine copper(II) complexes**, both from D- and L-carbohydrate series, have been reported to bind hTel (figure **1.36**).[225] Here, the G4-ligands interaction mode involved the recognition of the DNA grooves and/or loops instead of the classical π–π stacking interactions with the external tetrads of the G-quadruplex. The conjugates discriminated among hTel conformations, since binding affinity was higher in potassium buffer conditions than in sodium.
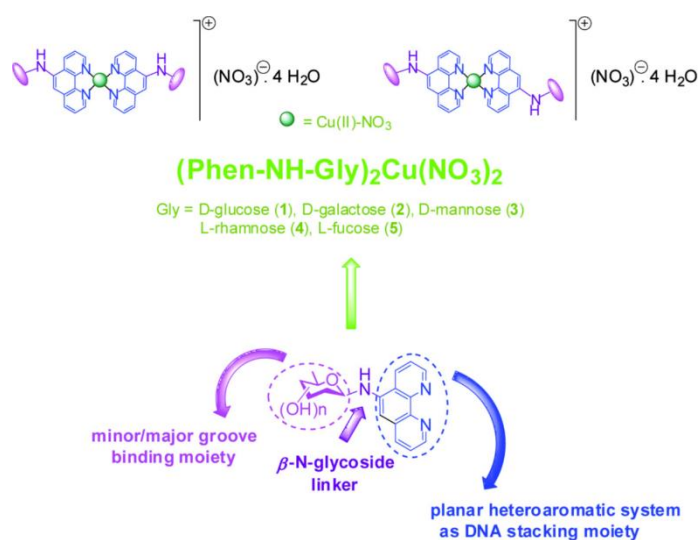
Figure 1.36. Structural features of N-(1,10-phenanthrolin-5-yl)-b-glycopyranosylamines and their copper(II) complexes.

These examples prove the advantages of incorporating carbohydrates to the G4-ligands structure, such as the possibility to introduce a new G4-interaction mode to the grooves/loops, the

most specific structural feature in a quadruplex. We plan to take advantage of this property by the conjugation of carbohydrates to G4 ligands. At the same time, we also expect to increase solubility and bioavailability, and also to promote selective entrance into tumor cells through overexpressed GLUTs.

3. AIMS.

This thesis is focused on the preparation of new carbohydrates-modified G4-ligands as mainly potential antiproliferative drugs. The aims are summarized in four points:

❖ To design and synthesize a new series of G4-ligands incorporating carbohydrates in their structures: phenyl ditriazole and naphthalene diimide derivatives.

❖ To evaluate the interaction of the new carbohydrate modified derivatives with different G4-quadruplexes, using biophysical techniques (FRET melting assay, CD, NMR, mass spectroscopy and SPR) to obtain the affinity and selectivity profile.

❖ To study their antiproliferative activity over different human cancer cell lines, and the cytotoxicity on non-cancer cell lines.

❖ To investigate the potential role of the carbohydrate motif on the facilitated cell uptake by GLUT transporters.

CHAPTER II:
## CARBOHYDRATE-BASED PHENYL DITRIAZOLE DERIVATIVES AS NEW G4-LIGANDS

1. INTRODUCTION.

1.1. Phenyl ditriazole derivatives as G-quadruplex ligands and their applications.

Moorhouse *et al.*[226,227] developed a new G4-ligand series which presents 1,3-di(1,2,3-triazole-4-yl) benzene as the aromatic core system and different tertiary amines as cationic side chains. Their goal was to create quadruplex-selective molecules capable of reaching clinical application. The potential G4-affinity would be displayed by the heteroaromatic 1,3-ditriazole system, which interacts by π-π stacking with the top G-quartet.

The synthesis was based on the "click chemistry" approach connecting the central core with the lateral chains through the Cu(I)-catalyzed Huisgen cycloaddition reaction.[228] Fragments containing tertiary amines-azide were linked to the 1,3-diethynylbenzene core to build up the five contiguous but not-fused aromatic system, including two triazoles (scheme **2.1**).
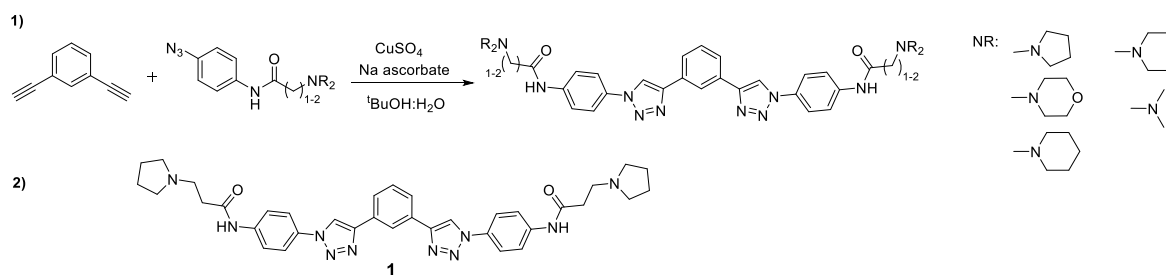
Scheme 2.1. 1) General click chemistry reaction used in the reference. 2) The pyrrolidinyl derivative **1**, the best candidate chosen by the authors.

Binding to G4 for these compounds was examined by FRET melting assay over the human telomeric quadruplex and a duplex sequence studying their affinity and selectivity. The classical G4-binders BRACO-19 and Telomestatin were used as controls. The nature of the tertiary amine affected the ΔT_m value as follows: pyrrolidinyl>dimethylamine>piperidinyl. They determined the best candidate was the one with two carbon length chain and the pyrrolidinyl motif as tertiary amine (**1**) since, according to FRET melting assay results, it showed a ΔT_m of 17.8°C on the human telomeric sequence whilst did not affect duplex stability. To confirm its selectivity, they carried out a competition assay with increasing amount of calf thymus DNA. Even at high competitor concentration, ΔT_m mainly remained the same value in comparison with the decrease on ΔT_m for BRACO-19 control.

Molecular modelling of the ligands with a telomeric-parallel quadruplex model exposed the same tendency than FRET results, with larger calculated free energies of binding and binding constants for compound **1**. The core system constituted by the central phenyl ring and the triazole linkers stacked over the tetrads (figure **2.1a**). Both the amide-nitrogen and the tertiary amine contributed to hydrogen-bond formation with quadruplex phosphate groups, which enhanced structure stabilization (figure **2.1b**). Additionally, tertiary amine groups protonation at physiological pH favored also electrostatic interactions with the negatively charged backbone atoms in the loops. Finally, telomerase inhibition activity was evaluated through TRAP-assay and the best candidate exhibited $^{tel}EC_{50}$ of 13.2 μM.

a)

b)
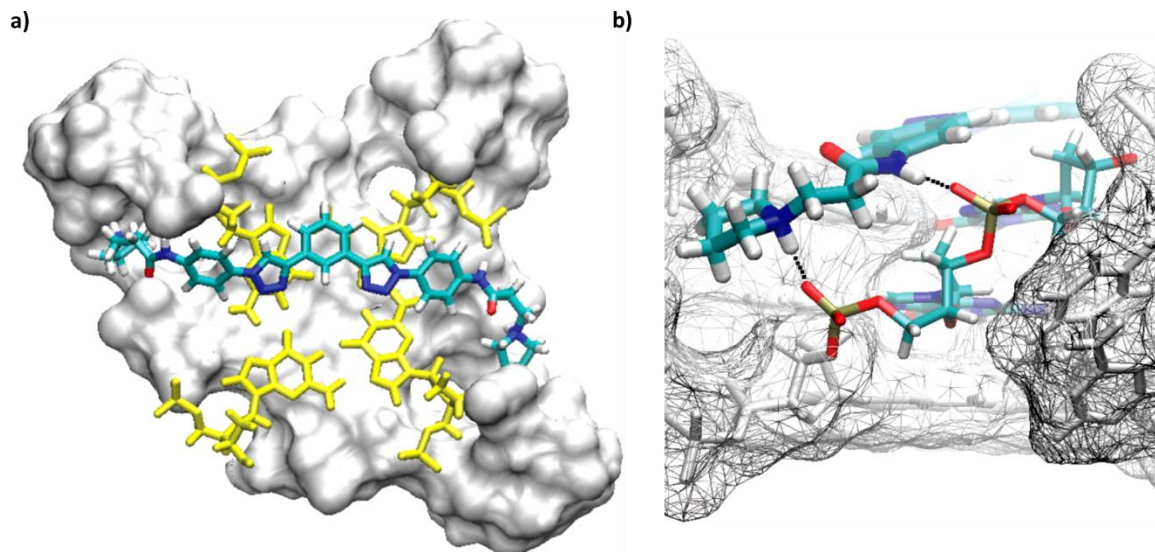
Figure 2.1.a) Top view of ligand 43 making π-π stacking interactions with the guanine tetrad (yellow). The ligand side chains are shown lying in the grooves created by the TTA loops. b) Interactions involving the amide nitrogen atom and the positively-charged tertiary amine with the negatively-charged phosphate groups in the loops. Hydrogen bonds are shown as dashed lines. The ligand is shown in a thick stick representation.

The pyrrolidinyl-candidate (compound **1**) was later used to estimate its cytotoxicity over MCF-7 and A-549 cancer cell lines obtaining an IC_{50} of 3.6 µM and 3.4 µM, respectively;[229] while over a somatic cell line (WI38) was 11.9 µM (three times higher). In the same report, the authors combined this molecule with the antitumor agent Cis-Platin at sub-cytotoxic concentrations and at different stoichiometric relation to determine the possibility of synergetic effect from this combination. They observed a synergetic effect at ratios of 1:1, 1:2, 1:3 and 2:1 (Cis-platin: pyrrolidinyl-derivative), whilst at 3:1 ratio the effect was antagonistic. Results evidenced the candidate produced long-term growth arrest in combination with Cis-Platin for both cancer cell lines. The authors associated this effect with the telomerase inhibition activity of the compound.

1.2. Click chemistry and application in the Huisgen 1,3-dipolar cycloaddition.

The term "click chemistry" was introduced by Sharpless[230] to describe a new strategy to easily conjugate small molecules mimicking the linkages observed in nature, such as nucleic acids, proteins and polysaccharides which are polymers of subunits linked through carbon–heteroatom bonds. According to Sharpless, this reaction must be "modular, wide in scope, give very high yields, generate only inoffensive byproducts, and be stereospecific". In addition, the process must be characterized by "simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation."

They applied this concept to the following chemical transformations:

➢ Cycloadditions of unsaturated species, especially 1,3-dipolar cycloaddition reactions, but also the Diels Alder transformations.

➢ Nucleophilic substitution chemistry, particularly ring-opening reactions of strained heterocyclic electrophiles such as epoxides, aziridines, aziridiniumions, and episulfoniumions.

➢ Carbonyl chemistry of the "non-aldol type", such as formation of ureas, thioureas, aromatic heterocycles, oxime ethers, hydrazones, and amides.

➢ Additions to carbon-carbon multiple bonds, especially oxidative cases such as epoxidation, dihydroxylation, aziridination, and sulfenyl halide addition, but also Michael additions of Nu-H reactants.

Among the first chemical transformation we find the **Huisgen 1,3-dipolar cycloaddition** of alkynes and azides to yield 1,2,3-triazoles.[228] Azides were avoided for a long time due to their low safety, however they are highly useful in click chemistry because they are stable at water, oxygen and many chemical reaction conditions. Triazole is chemically inert to reactive conditions such as oxidation, reduction and hydrolysis. When the triazole cycloaddition is carried out at high temperature, a mixture of the 1,4- and 1,5- regioisomers is obtained (scheme **2.2a**), but when copper (I) is introduced as catalyst, the reaction progresses to give regiospecifically the 1,4-disubstituted 1,2,3-triazole specie with high yields (scheme **2.2b**). The catalyst is prepared *in situ* from Cu^{II} salts, cheaper and cleaner than the corresponding Cu^{I} salts. Commonly, CuSO$_4$·5H$_2$O is used with this goal, and ascorbic acid or sodium ascorbate as the reductant agent. Reaction is completed at these conditions in a range of 6-36 hours at room temperature in variety of solvents (always including water). Organic solvent can be avoided considering that starting material total solution is dispensable.
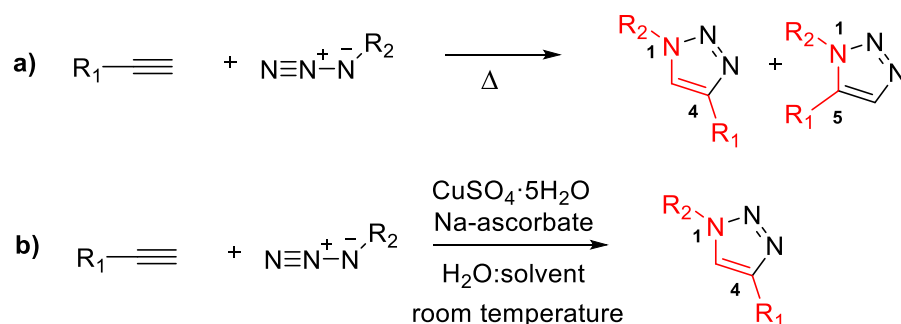
Scheme 2.2. General "click chemistry" schemes. a) Reaction without Cu catalyst and heat. b) Incorporation of Cu as catalyst in the reaction to give 1,4-triazole derivatives.

Sharpless[228] proposed a reaction mechanism with Cu^{I}-acetylide derivative. The complete cycle is shown below in scheme **2.3**.
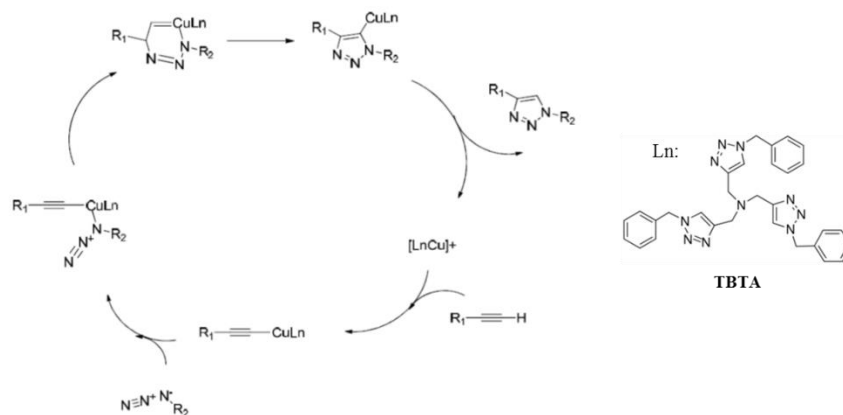
Scheme 2.3. Mechanistic cycle of Huisgen 1,3-dipolar cycloaddition with copper as catalyst and TBTA as ligand.

Afterwards, Sharpless' group[231] synthesized and incorporated TBTA (Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine) to the reaction mixture to act as a ligand which stabilizes the Cu^I from disproportionation, oxidation and increases its catalytic activity (scheme **2.3**). The next step was the reduction of the reaction-time by microwaves, which has significantly shortened the reaction times to minutes, keeping regioselectivity, purity and good yields.[232]

Generally, the triazole formation is unaffected by bulk and electronic properties of both azide and alkyne but due to differential rates, reaction conditions must be optimized depending on both azide and alkyne nature.

2.   AIMS.

+ To synthetize new carbohydrate-phenyl ditriazole derivatives (**carb-PTDZ**), using a "click chemistry" approach:
    o Synthesis of different azido-sugars motifs.
    o Click chemistry optimization to obtain symmetric phenyl ditriazole derivatives with the same carbohydrate motif in both sides of the aromatic core.
    o Click chemistry optimization to achieve asymmetric phenyl ditriazole derivatives, with a carbohydrate motif on one side and the pyrrolidinyl motif on the other side.
+ To characterize the interaction between the new compounds and different G4-forming sequences by biophysical assays:
    o FRET melting assay; to measure the stabilization over both DNA and RNA G4s, as well as their selectivity versus duplex DNA.
    o CD assays; to study the G4-topological distortion by the interaction with new ligands.
    o SPR experiments; to study the kinetic behavior of the binding.
    o NMR studies; to investigate the ligand-G4 complexes, with two human telomeric sequences and an HIV-G4 structure.
+ To investigate the potential anticancer application of the new compounds:
    o To measure cytotoxicity on several tumoral and non-tumoral cell lines.

3.   RESULTS AND DISCUSSION.

Our first goal was to develop new G4-ligands with a smaller aromatic core than the classical binders and to incorporate carbohydrates in the structure to confer them better properties as drug-like molecules. The strategy employed by Neidle's team to create their pyrrolidinyl-phenyl ditriazole (**PTDZ**) derivative inspired us to design new carbohydrate-modified phenyl ditriazole (**carb-PTDZ**) compounds as selective G-quadruplex-ligands. The innovation resides in incorporating carbohydrate residues into the structure to improve the compound solubility, G4-selectivity and antiproliferative activity taking advantage of the Warburg effect displayed by cancer cell lines.

3.1.   Synthesis of new phenyl ditriazole derivatives.

Taking the synthesis of the compound **1** as inspiration, we developed two kind of compounds: symmetric structures, with the same sugar motif as side chains; and asymmetric molecules, with a sugar on one side and the same pyrrolidinyl- residue employed by Moorhouse *et al.*[226] on the other side. To do so, diethynylbenzene would be reacted with different azide derivatives. Symmetric compounds were built in one "click chemistry" step and the asymmetric ones required two "click chemistry" steps. General reaction scheme is shown below (scheme **2.4**).
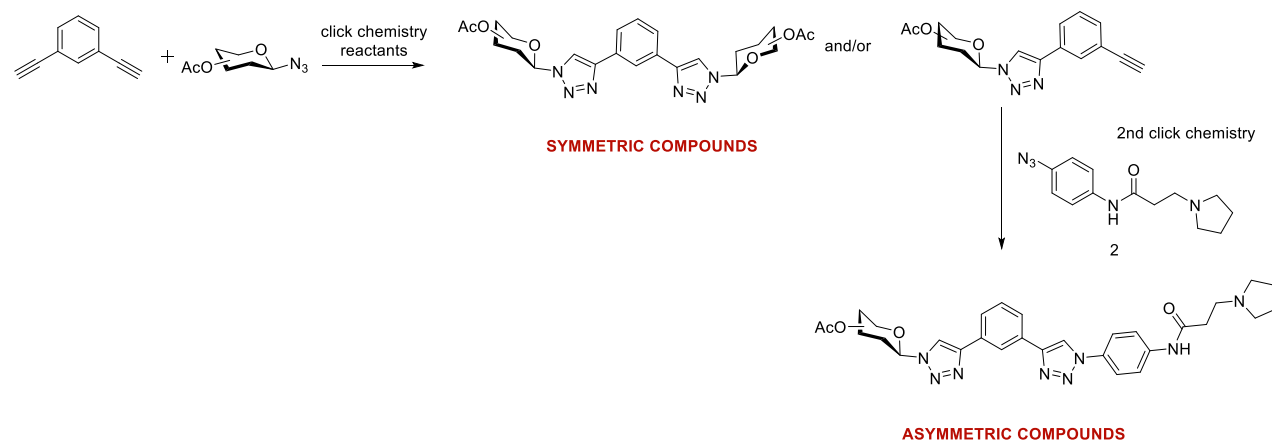
Scheme 2.4. General scheme synthesis of the new ligand by "click chemistry" approach.

3.1.1. SYNTHESIS OF AZIDO-SUGAR DERIVATIVES.

We selected a series of mono- and disaccharides with varied stereochemistry and two possible presentations: a) directly attachment to sugar C1 or b) attachment to C1 using an ethylene spacer. Hence, we prepared 1-azidoglycosides (compounds **3-6**) and 2-azidoethyl glycosides (compounds **10** and **14**).

We followed reported procedure to prepare 1-azido- glucose (**3**), fucose (**4**), lactose (**5**) and maltose (**6**) derivatives (scheme **2.5**). We started from the peracetylated carbohydrate, which was purchased or prepared by reaction of the natural sugar with acetic anhydride in pyridine. Then, the glycosyl bromide was easily prepared by reacting the acetylated derivative with hydrogen bromide

in acetic acid.[233] Without further purification (due to bromide product instability), we reacted the bromide with NaN$_3$ in a water/acetone solution mixture. The β-azido derivative was obtained with moderate to good yields (43-80%).[234]

Scheme 2.5. a) General scheme synthesis of the azido-sugar derivatives. b) Azido sugars afforded by the previous method: 1-azido glucose, fucose, lactose and maltose.

We also synthesized peracetylated 1-azide-*N*-acetylglucosamine (**8**), 2-azidoethyl glucopyranoside (**14**) and 2-azidoethyl mannopyranoside (**10**) for further exploring carbohydrate diversity and presentation. The corresponding azido derivatives were prepared by different synthetic routes (scheme **2.6**).

Scheme 2.6. a) Synthesis of 1-azido *N*-acetyl glucosamine derivative **8**. b) synthetic route for preparation of mannose with C2 linker azido derivative **10**. c) Synthetic route for preparation of the azido C2 linker glucose derivative **14**. d) Synthesis of ethanolazide **15**.

The peracetylated 1-azido-*N*-acetylglucosamine derivative **8** was synthetized in two-steps as reported in the literature (scheme **2.6a**).[235–237] Firstly, *N*-acetylglucosamine was converted into 2-*N*-(acetyl)-2-deoxy-3,4,6-triacetate α-D-glucosamine chloride **7** by reaction with acetyl chloride. Secondly, the corresponding chloride derivative was reacted with TMSN$_3$ and TBAF as catalyst to obtain the azido sugar **8** with good yield (67%).

Synthesis of mannose C2-linker derivative **10** was carried out according to the reported methodology (scheme **2.6b**).[238,239] The linker was incorporated by glycosylation of peracetylated mannose with 2-bromoethanol using boron trifluoride diethyl etherate Lewis acid as the promotor. We reacted the bromide with NaN_3 after 48 h reaction at 80 °C with good yield (84%). Ideally, the same steps could be applied to the corresponding glucopyranoside to obtain the derivative **14**, but the substituents in C2 at equatorial position deactivates the anomeric acetyl group. For this reason, an alternative glycosylation via trichloroacetimidate formation was applied (scheme **2.6c**).[240] To do so, the acetyl anomeric position was deprotected with hydrazine monohydrate[241] before trichloroacetimidate formation using trichloroacetonitrile and DBU as catalyst. Then, the glycosylation was carried out using the same procedure as for mannose. Finally, the bromide was replaced by NaN_3 as described above resulting in pure β-anomeric product **14** (53% of yield).

On the other hand, we prepared a "control" compound without any sugar on its structure (scheme **2.6d**). This compound will help us to figure out the real contribution of the carbohydrate to the quadruplex affinity, selectivity or biological properties of the ligand. For this goal, we prepared a C2-azido linker **15**. It was easily synthesized from 2-bromo-ethanol by reaction with NaN_3 and $^nBu_4NBr$.[242]

The last step in all synthetic route to synthesize the azidosugar was the substitution of the halide with the azido group. The main handicap was that both starting material and product shared similar Rf (retardation factor), which complicated reaction monitoring by TLC. It was important to keep the reaction enough time to assure total conversion and check by NMR the shift of the proton and carbon at the anomeric position. The bromide derivative presented lower chemical shift than the azido product.

3.1.1.1. Charged azido sugars

Regarding the general G4-ligand structural motif (central aromatic core surrounded by side chains containing charged groups), we designed molecules with positively charged sugars as side chains to be attached to the aromatic core. In collaboration with Prof. M. Carmen Galán (University of Bristol), we planned to connect 2-methylimidazole to the C6- azido glucose derivative. Her group has reported the ITAG-carbohydrate synthesis to employ them on the ionic-liquid field, useful in oligosaccharide organic synthesis.[243–245] They incorporated the ITAG motif at the anomeric position through S_N2 displacement of a leaving group (bromide, tosyl-,mesyl-,…) with potassium tetrafluoroborate and heat. We applied these conditions to add the motif at C6 glucose position.

First, we tried to incorporate pyrrolidine via tosyl- substitution (scheme **2.7a**). 1-azido-β-glucopyranoside **16** was reacted with *p*-toluene sulfonic chloride, followed by acetylation with acetic anhydride. The desired product **18** was obtained with good yield (64%). The displacement of the tosyl group was tried by direct reaction with pyrrolidine in methanol. The TLC assay showed a new product but it could not be isolated. Second, we tried the reaction of the tosyl with n-methylimidazole using high temperature and two different solvents (scheme **2.7b**) but no reaction occurred.[246] Third, we tried using mesylate as leaving group (scheme **2.7c**).[247] Despite of the successful mesylation of compound **16**, the subsequent S_N2 lead to a reaction mixture difficult to

purify due to the product's high polarity and retention over silica. We decided to acetylate the resulting imidazole derivative **22** to facilitate its purification, but two purification steps were needed to obtain the corresponding pure compound **20** (flash chromatography and SephadexG25) at 10% yield.
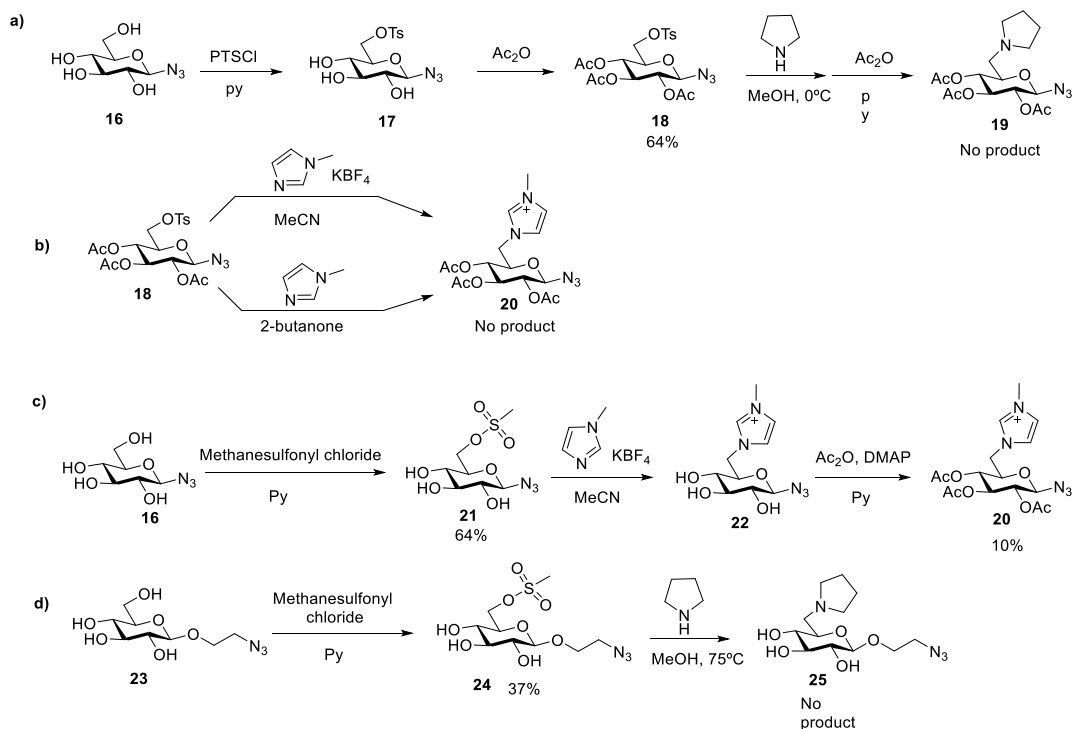
Scheme 2.7. Synthetic routes applied to afford positively charged glucose.

The same approach was used with C2-linker azido sugar **23** and pyrrolidine as potential cationic motif. We introduced mesylate to the azido-linker derivative following the previous reaction conditions to get the desired product **24** (scheme **2.7d**). Nevertheless, pyrrolidine addition was not successful using methanol as solvent and keeping the reaction at 75°C for 4 days.

The synthesis of charged carbohydrates was not an easy task, their polarity hindered reaction monitoring and purification steps, and common flash chromatography was not enough to reach the desired compounds with decent yields. After all efforts, we only were able to obtain compound **20** with low yield.

3.1.2. CLICK CHEMISTRY OPTIMIZATION FOR MONO- AND DITRIAZOLE COMPOUNDS.

The next step was attaching the obtained azido sugars to the aromatic core via click chemistry with triazole ring as connector element. The Huisgen 1,3-dipolar cycloaddition of alkynes and azides to yield 1,2,3-triazoles via click chemistry has been highly exploited in organic synthesis. Apparently, reactants are already stablished.[248] However, the chemical variety and properties of both components (azido- and alkynyl- motifs) force to optimize the reaction

conditions for each individual case. Fortunately, sugars have been conjugated to other organic molecules both as the azido- and propargyl- reaction component to create glycodendrimers or carbohydrate-clusters in the past.[249–252]

Our first strategy consisted on building **symmetric compounds**, with the neutral carbohydrate-motifs as lateral chains. The alkynyl motif was 1,3-diethynylbenzene **A**, with two alkynyl groups available for two cycloadditions. As second strategy, we planned to prepare **asymmetric compounds**, which forced us to control the reaction progress until reaching the compound with only one triazole, leaving free one alkynyl residue for further reaction. To do so, we optimized the amounts of the common Huisgen reactants -$CuSO_4 \cdot 5H_2O$ as copper salt, sodium ascorbate as reductant agent and TBTA as ligand -and reaction conditions such as temperature, reaction time, the use of microwave, and the use of the protected or unprotected azido-sugar **B** derivatives. The optimization was carried out with the azido-glucopyranoside derivative due to its easy and fast preparation. The table **2.1** summarizes the different conditions examined.
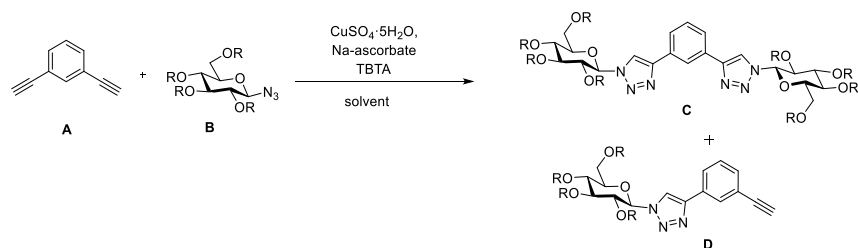
Entry	R	Azide eq	CuSO$_4$·5H$_2$O eq	Na-ascorbate eq	TBTA	Solvent	Temp	Time	% Product
1	H	3	0.1	0.5	0.2	H$_2$O:^{t}BuOH (1:1)	rt	24h	36% **D**(yield)
2	H	4	0.1	0.4	0.2	H$_2$O:THF (1:1)	60°C mw	1h 30 min	98% **C**(yield)
3	H	4	0.1	0.4	0.2	H$_2$O:THF (1:1)	60°C mw	25 min	Mainly **C** Low proportion **D**
4	Ac	4	0.1	0.4	0.2	H$_2$O:THF (1:1)	60°C mw	25 min	27% **C** 73% **D**
5	H	4	0.2	0.7	0.7	H$_2$O:THF (1:1)	60°C mw	30 min	Not purified
6	Ac	4	0.2	0.7	0.7	H$_2$O:THF (1:1)	60°C mw	1h	80%(yield) **C** 0% **D**
7	Ac	3	0.05	0.5	0	H$_2$O:THF (1:1)	130°C mw	30 min	100%(yield) **C** 0% **D**
8	Ac	3	0.1	0.4	0	H$_2$O:THF (1:1)	60°C mw	25 min	30% (yield)**D**
9	Ac	2	0.1	0.4	0	H$_2$O:THF (1:1)	60°C mw	65 min	14% (yield)**C** 55% (yield)**D**
10	Ac	2	0.1	0.4	0.2	H$_2$O:THF (1:1)	60°C mw	1h	76%(yield) **C** 18%(yield)**D**

Table 2.3. Click chemistry reaction conditions employed for optimization of symmetric and asymmetric compounds synthesis.

^{t}BuOH is frequently the organic co-solvent used in this kind of reaction, but in our case we noticed solubility problems with our azido sugars, therefore we decided to used THF instead (entries **1** and **2**). Furthermore, we realized reaction yield and purification was favored when peracetylated carbohydrates were employed in comparison to the unprotected ones (entries **3-6**). Finally, elimination of TBTA and increased reaction temperature (130 °C, MW) showed the best conditions to obtain the symmetric compounds (table entry **7**). Once the reaction was completed, QuadraSil^{TM} MP was added to scavenge the copper salt catalyst, then the mixture was filtered and the crude easily purified by flash chromatography to remove the azido sugar excess. Following

these conditions, we synthesized the desired symmetric compounds shown in figure **2.2** with good yields in most cases.
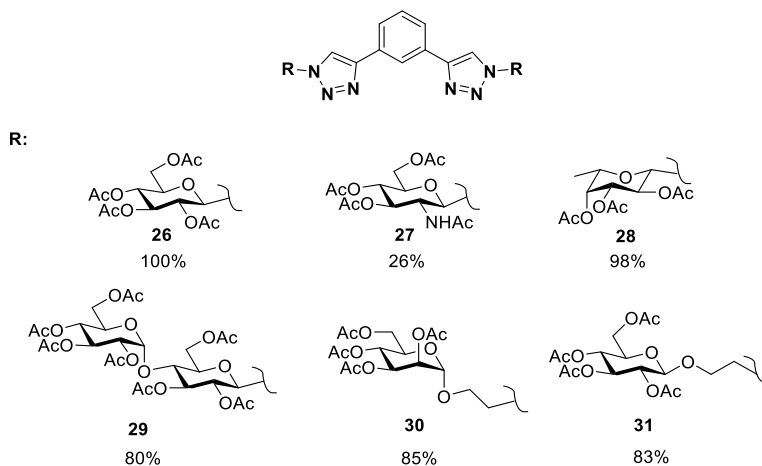
Figure 2.2. Final symmetric compounds achieved.

To achieve the mono-triazole derivative we used less amount of the azido sugar and lower reaction temperature (entries **9-10**). The best conditions were those in entry **9**. However, when these conditions were applied to other sugars, we found a range of reactivity. Reaction monitoring through TLC was required, considering that some azido carbohydrates reacted quicker than others. Additionally, we included the synthesis of the mono-triazole-1,4-diethynylbenzene **F**, instead of 1,3-diethynylbenzene **E** (see below), to investigate how a structural change at the aromatic core could influence binding of G-quadruplexes and their cytotoxicity. In that case, just glucose and maltose derivatives were prepared. We summarize the reaction time and obtained yields for each derivative in table **2.2.**
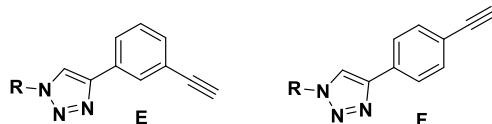
Aromatic core	R	Reference	Time (min)	Yield (%)
E		**32**	65 min	50
E		**33**	30 min	80
E		**34**	15 min	53
E		**35**	15 min	43
E		**36**	25 min	50
E		**37**	15 min	41
E		**38**	30 min	30
F		**39**	30 min	15
F		**40**	45 min	36

Table 2.4. Monotriazole reaction times and yields.

We previously mentioned the design of a non-sugar control PTDZ derivative with the objective to study the actual effect of the carbohydrate. To prepare it, 2-azidoethanol (**15**) was chosen to be linked to the aromatic core via click chemistry as well. Nevertheless, this azido derivative had to be treated more carefully considering the explosive risk in short chain azides. For this reason, we synthesized the mono triazole at special reaction conditions, avoiding neither microwave nor high temperature (scheme **2.8a**). Reaction with only 1 equivalent of 2-azidoethanol **15** was stirred at room temperature for 2 h. The azide was very reactive but still the monotriazole **41** was isolated with 26% of yield.
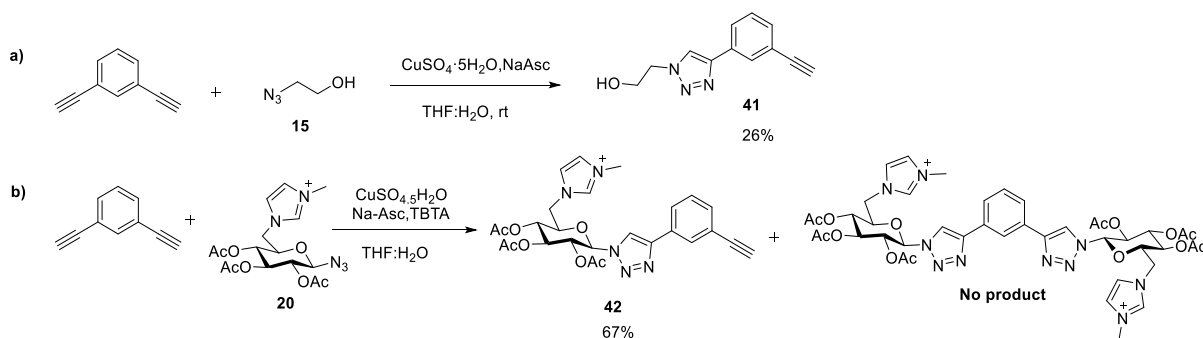
Scheme 2.8.a) Preparation of 2-azidoethanol and its monotriazole derivative. b) click chemistry with the methylimidazole derivative to obtain the di and c) monotriazoles.

We tried to incorporate the 6-methyl-imidazole azido glucose **20** with the previous procedures to prepare both symmetric and asymmetric derivatives, but we found several problems. When we applied the conditions to obtain the symmetric compound, we observed that only the mono-triazole **42** was formed, even forcing reaction conditions adding more amount of copper salt, sodium ascorbate and TBTA (scheme **2.8b**). We isolated the monotriazole **42** by reverse chromatography C18 getting a 67% of yield. In fact, it seems this particular motif may act as a copper (II) chelator.[253]

3.1.3. SYNTHESIS OF SYMMETRIC PHENYL DITRIAZOLE LIGANDS.

We chose the acetylated sugars to perform the click chemistry reaction, therefore the last step to achieve the final symmetric compounds was to deprotect them using basic conditions. To do so, the pure acetylated compounds were dissolved in methanol and sodium methoxide solution was added to provide the basic conditions (figure **2.3**).[254] Amberlite IR®120 ion-exchange resin was then incorporated to neutralize the mixture reaching the pure final **symmetric PTDZ** derivative without further purification step and good yields.
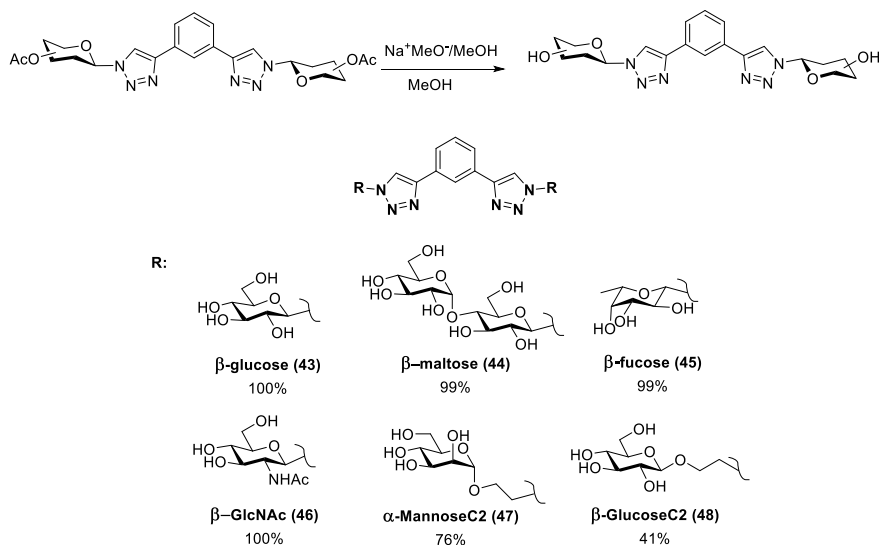
Figure 2.3. Final reaction step to unprotect the peracetylated carbohydrates and the final symmetric PTDZ derivatives.

3.1.4.   SYNTHESIS OF ASYMMETRIC PHENYL DITRIAZOLE LIGANDS.

Considering that an alkynyl- motif was still available in the mono-triazole derivatives, another cycloaddition with different azido derivative was used to obtain the asymmetric-PTDZ compounds with two different lateral chains. The second azide to be incorporated was Neidle's pyrrolidinyl side chain (**2**, see next section). Therefore, the final compounds had a sugar on one side, and a cationic side chain on the other side.

The second click chemistry was carried out using a modified cycloaddition based on Neidle's team procedure and using the previous mono-triazole fragments. We noticed solution problems with the azido sugar in the ᵗBuOH:H₂O mixtures, so we used THF instead of ᵗBuOH. Neidle's group applied 130 °C in the microwave reaction, but sometimes THF was evaporated at such temperature conditions, hence we reduced the temperature to 60 °C. We added TBTA at 0.06 equivalents to compensate the reduced temperature. Reaction was monitored by checking starting material disappearance, though reaction was normally completed in 30 min. After quenching with QuadraSil™ MP as previous cycloadditions, purification was more difficult. Resultant products presented high polarity and similar to the azide reactant, thus flash chromatography was not enough to separate the product and we decided to use semi-preparative HPLC for proper purification. Taking into account that the carbohydrate was still protected, we decided to carry out the deacetylation on the reaction mixture previously to the HPLC purification step. Final asymmetric compounds and yields are described below (scheme **2.9**). The same procedure was applied to the synthesis of the *p*-phenyl ditriazole derivatives (scheme **2.10**).
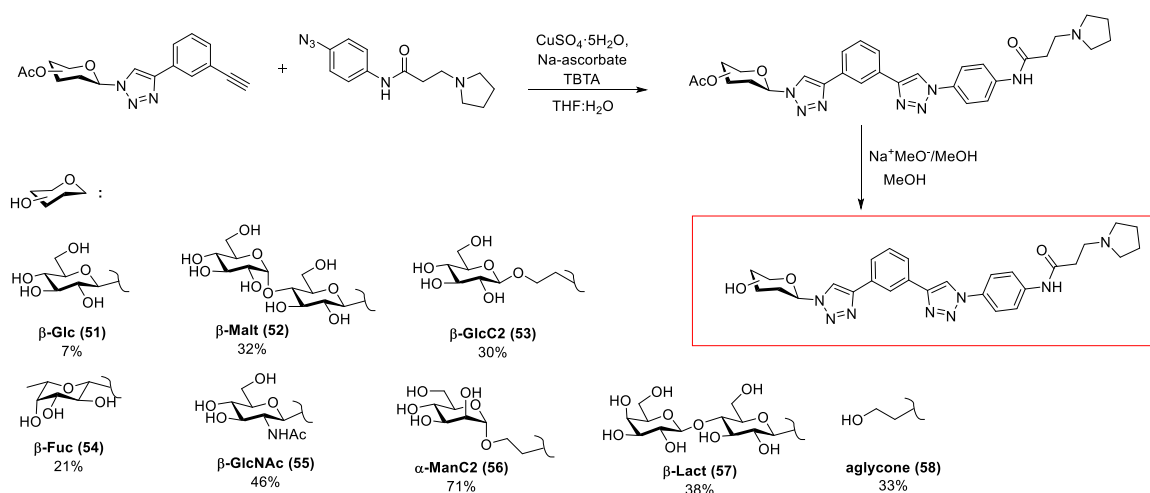
Scheme 2.9. Click chemistry reaction to achieve the final *m*-asymmetric PTDZ ligands, and the structures of the compounds synthetized.
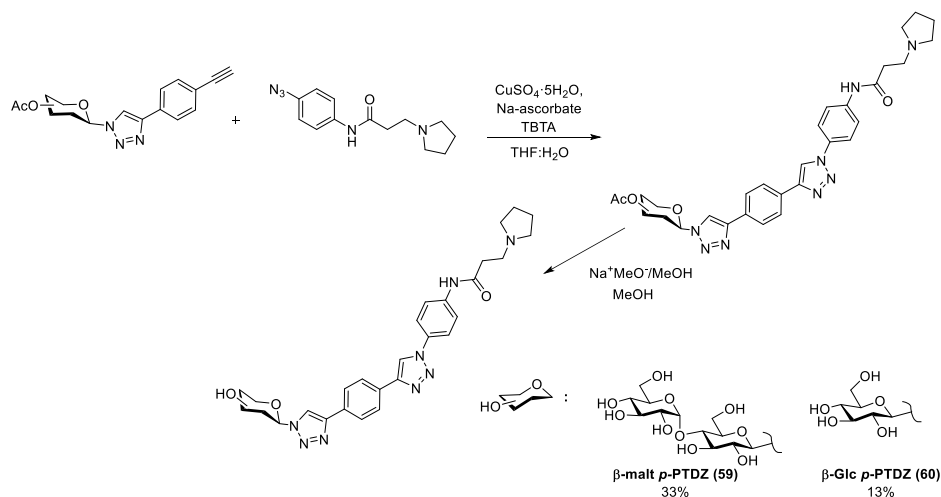
Scheme 2.10. Click chemistry applied to synthetize the *p*-asymmetric PTDZ ligands and the final structures.

We applied this cycloaddition to the monotriazole derivative with 6-methyl-imidazole glucopyranoside **42** with the purpose to reach an asymmetric compound with two different cationic side chains (scheme **2.11**). But again, the second cycloaddition did not progress, probably because the methyl-imidazole motif was scavenging the copper catalyst.
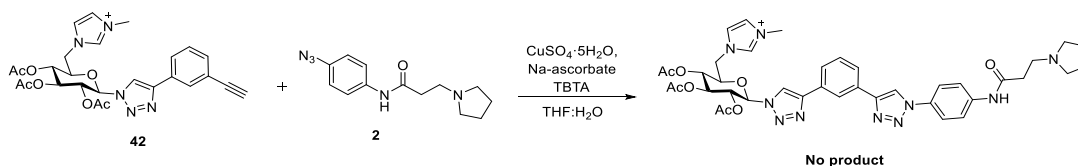
Scheme 2.11. Click chemistry with monotriazole derivative **42**.

Generally, when more than one alkynyl- motif exists, click chemistry takes place to obtain symmetric compounds with the same residues attached to the triazole connectors. Only a few references are reported using asymmetric click chemistry,[255–258] and most of them proceeded with an intermediate alkynyl protection step.[256–258] It is important to remark we were able to customize the molecule with the desire motif linked to each side by controlling the click chemistry reaction conditions.

3.1.5. <u>SYNTHESIS OF NEIDLE'S COMPOUND AS REFERENCE.</u>

We used as control Neidle's phenyl ditriazole G4-ligand with pyrrolidinyl motif as side chains (**1**).[226] We prepared it following the synthetic route reported by the authors and shown in scheme **2.12**. Firstly, the *N*-(4-azidophenyl)-(3-pyrrolidin-1-yl) propionamide **49** was synthesized in three reaction steps. Starting from *p*-nitro aniline, amide formation and S_N2 substitution with pyrrolidine were carried out in one step. The nitro group was then reduced to amine, which was converted into the azide by reaction with sodium azide at acidic conditions. The purification of the azide **2** was particularly complicated: azide isolation was achieved using alumina instead of silica gel and polar eluent mixtures to avoid the product retention over the solid phase. Finally, the click reaction was carried out as described, but reaction time was increased from 30 to 45 min. Three

purification steps were needed to obtain the pure compound: flash chromatography, Sephadex LH20 and finally reverse chromatography C18.
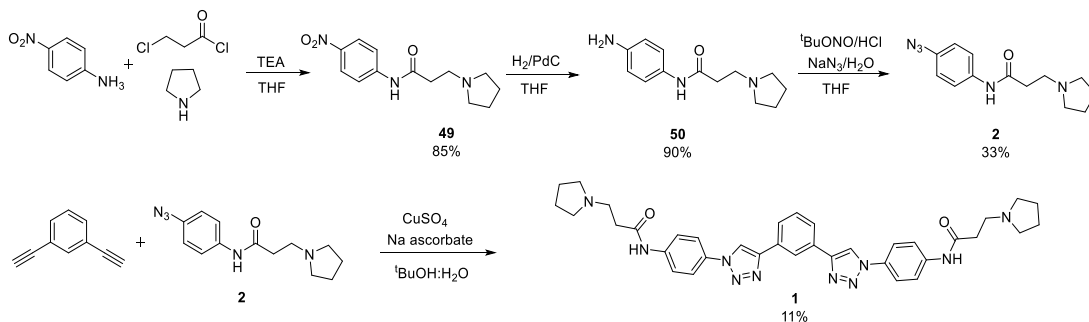
Scheme 2.12. Synthetic route employed in the reference compound.

3.2.   Biophysical studies with G-quadruplexes.

Next, we evaluated the carbohydrate modified phenyl ditriazole derivatives (carb-PTDZ) as G-quadruplex ligands using different biophysical techniques available in the field. This section was developed in Dr. Jean-Louis Mergny's laboratory (Institut Européen de Chimie et Biologie, IECB), placed in Bordeaux. His group is a reference in the G-quadruplex field: they study different aspects such as structure, helicases activity, new ligands development and biological evaluation. They have optimized biophysical techniques for the most complete binding study of G4-structures, and I had the opportunity to evaluate our own compounds and receive their advice during several stays there.

FRET melting assay was the first technique we used, which allowed us to discriminate the compounds between G4-binders and not binders. Once we selected the best candidates, we studied their binding with more detail using CD, SPR and NMR experiments. We could get preliminary mass spectrometry assays as well with the technical support of Dr. Frédéric Rosu.

3.2.1. FRET MELTING ASSAYS.

This technique allowed us to screen quickly and efficiently the G4-stabilization effect of the new carb-PTDZ derivatives. They have been evaluated over different DNA and RNA G4-forming sequences as well as over a DNA-duplex sequence to investigate their selectivity over the duplex and among different quadruplex structures.

First, we studied the ligand binding on the human telomeric G-quadruplex sequence (F21T: d (FAM-$G_3[T_2AG_3]_3$-TAMRA)) in potassium buffer conditions. Neidle's compound **1** was our reference ligand and PhenDC$_3$ was included as positive control. Results are shown in the table and graph of figure **2.4**.

a)

		1 µM	5 µM	10 µM
Symmetric PTDZ	43 (β-glc-PTDZ)	-0,33±0,60	-0,22±1,17	-0,88±1,94
	44 (β-malt-PTDZ)	0,01±0,30	0,15±0,96	0,46±0,64
	45 (β-fuc-PTDZ)	-0,53±0,97	-0,23±1,03	-0,97±1,71
	46 (β-glcNAc-PTDZ)	0,27±0,34	-0,13±0,77	0,26±1,53
	47 (α-manC2-PTDZ)	-0,22±0,51	-0,07±0,54	-0,52±0,32
	48 (β-glcC2-PTDZ)	0,15±0,50	-0,26±0,94	-0,44±0,98
Asymmetric m-PTDZ	51 (β-glc-m-PTDZ)	0,46±0,77	3,03±1,20	3,82±1,77
	52 (β-malt-m-PTDZ)	0,73±0,83	3,36±1,29	4,96±2,50
	53 (β-glcC2-m-PTDZ)	0,58±0,78	3,45±1,92	4,76±2,04
	54 (β-fuc-m-PTDZ)	0,64±0,62	2,78±1,17	4,84±0,49
	55 (β-glcNAc-m-PTDZ)	0,07±0,89	3,00±1,83	4,82±1,93
	56 (α-manC2-m-PTDZ)	1,24±0,45	5,61±0,11	8,03±0,38
	57 (β-lac-m-PTDZ)	1,49±0,34	5,59±0,35	7,94±0,53
	58 (aglycone-PTDZ)	1,46±0,04	8,03±0,07	9,05±0,64
Asymmetric p-PTDZ	59 (β-malt-p-PTDZ)	0,10±0,30	3,65±0,33	5,80±0,51
	60 (β-glc-p-PTDZ)	0,50±0,28	3,34±0,16	5,75±0,29
Controls	1	7,73±1,48	27,33±0,55	NA
	PhenDC3	17,63±0,85	NA	NA

b)

Figure 2.4.a)Table and b) bar graph with each ligand ΔT_m (°C) over the human telomeric sequence (F21T) at 1, 5 and 10 µM ligand concentration. Buffer with 10 mM lithium cacodylate pH 7.2 containing 10 mM KCl and 90 mM LiCl.

The general observation was the low ΔT_m exhibited by the new ligands when they were compared with the controls at the same concentration. It was not possible to determine the T_m with the controls at higher ligand concentrations than 1 or 5 µM because the stabilization was too high. **PhenDC₃** was clearly the strongest ligand since it displayed the highest ΔT_m (at just 1 µM concentration), and **1** showed the same value than the highest concentration (10 µM) for our compounds. Therefore, the replacement of a positively charged side chain with a carbohydrate decreased the affinity towards the G4-telomeric sequence.

When we compared the symmetric with the asymmetric PTDZ ligands, we observed that the symmetric ligands were not able to increase the G4-T_m at any concentration, while the asymmetric ones showed ΔT_m up to 9°C at 10 µM ligand concentration. Their ΔT_m varied from 4 (**51, asym-glc-PTDZ**) to 9°C (**58, asym-aglycone-PTDZ**) at 10 µM ligand concentration.

We did not perform a competition assay with added dsDNA due to the low binding affinity to the G4-telomeric sequence. Nevertheless, we increased the ligand concentration (20 and 30 µM) with some symmetric and asymmetric PTDZ derivatives in the following experiments to find a

concentration-dependence or saturation effect (figure **2.5**). The corresponding experiments confirmed the no-stabilization ability of the symmetric PTDZ ligands, even at 30 µM concentration. Symmetric neutral carbohydrates phenyl ditriazole ligands did not show any binding to the G4-telomeric sequence. It was not the case for the asymmetric G4-ligands, whose ΔT_m increased with the concentration, reaching 13.4 °C with the best compound (**53, asym-glcC2-PTDZ**) at 30 µM concentration.

		1 µM	5 µM	10 µM	20 µM	30 µM
Symmetric PTDZ	44 (β-malt-PTDZ)	0,4±0,42	0,9±0,46	1,2±0,60	-0,3±0,13	0,7±0,05
	46 (β-glcNAc-PTDZ)	0,7±0,28	0,7±0,55	1,0±1,78	-0,1±0,58	0,4±0,29
Asymmetric m-PTDZ	51 (β-glc-m-PTDZ)	0,9±0,66	3,8±1,02	4,5±1,48	7,5±0,17	10,5±0,33
	52 (β-malt-m-PTDZ)	1,1±0,83	4,2±1,16	5,6±2,29	7,3±2,65	11,7±0,02
	53 (β-glcC2-m-PTDZ)	1,0±0,67	4,2±1,50	5,5±1,70	10,5±0,20	13,4±0,18
	54 (β-fuc-m-PTDZ)	1,1±0,48	3,6±0,61	5,5±0,45	5,6±0,14	10,1±0,33
	55 (β-glcNAc-m-PTDZ)	0,5±0,53	3,8±1,36	5,5±1,63	8,3±0,03	12,6±0,32
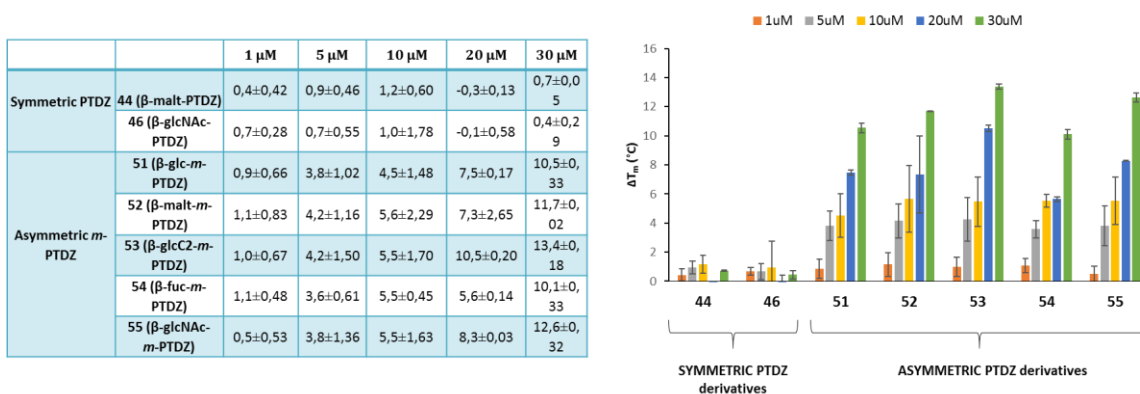
Figure 2.5. Table and bar graph with the ΔT_m (°C) over the human telomeric sequence (F21T) at 1, 5, 10, 20 and 30 µM ligand concentration. Buffer with 10 mM lithium cacodylate pH 7.2 containing 10 mM KCl and 90 mM LiCl.

The binding behavior of the asymmetric PTDZ ligands was very similar, not depending on the carbohydrate unit attached to the molecule. This result seems to point out that the central aromatic core and the charged side chain are mostly responsible for the G4-binding.

Once we analyzed the compounds stabilization ability over the G4-telomeric sequence, we studied their selectivity towards duplex DNA and over another G4-quadruplex topologies. Some of the asymmetric compounds, including control **1**, were assayed over a duplex forming sequence. While the new products did not show any stabilization effect, the control started to stabilize the duplex at 5 µM concentration and increased at 10 µM (figure **2.6a**). We observed better selectivity against dsDNA for our asymmetric PTDZ ligands than for Neidle's control. This is an important issue in the development of G4-ligands.

a)

FdxT

■ 1µM ■ 5µM ■ 10µM

b) **ΔT$_m$ at sodium buffer conditions**

	LIGANDS	F21T	FdxT
Symmetric PTDZ	43 (β-glc-PTDZ)	-0,2±0,20	-0,1±0,18
	44 (β-malt-PTDZ)	-0,5±0,18	-0,1±0,26
	45 (β-fuc-PTDZ)	0,6±0,51	0,2±0,10
	46 (β-glcNAc-PTDZ)	-0,2±0,23	-0,2±0,39
	47 (α-manC2-PTDZ)	0,5±0,31	0,3±0,00
	48 (β-glcC2-PTDZ)	0,2±0,26	-0,3±0,58
Asymmetric m-PTDZ	51 (β-glc-*m*-PTDZ)	6,9±0,25	0,9±0,29
	52 (β-malt-*m*-PTDZ)	6,4±0,00	1,1±0,04
	53 (β-glcC2-*m*-PTDZ)	5,4±0,98	0,7±0,16
	54 (β-fuc-*m*-PTDZ)	5,9±0,89	0,8±0,07
	55 (β-glcNAc-*m*-PTDZ)	5,8±0,15	0,6±0,37

Figure 2.6. a) Duplex (FdxT) stability assay at 1, 5 and 10 µM ligand concentrations. b) FRET melting assay with the human telomeric sequence (F21T) and duplex (FdxT) at sodium buffer conditions. Buffer with 10 mM lithium cacodylate pH 7.2 containing a) 10 mM KCl and 90 mM LiCl or b) 10 mM NaCl and 90 mM LiCl.

The previous experiments were carried out at potassium buffer conditions, at which human telomeric G4-sequence adopts a hybrid conformation. Sometimes, G4-ligands tend to be topology-selective, even on the same G4-forming sequence. The human telomeric sequence folds into a different topology at sodium buffer conditions. Thus, we examined ligand binding at such conditions using 30 µM ligand concentration and we also measured ligand binding to duplex DNA (FdxT sequence). We observed the ΔT_m decreased in all cases, with the best value of 6.9 °C at 30 µM, half of the value obtained at potassium buffer conditions (figure **2.6b**). Despite of the lower ΔT_m, the selectivity through quadruplex remained. Regarding these results, the new sugar-PTDZ ligands were more selective binding the hybrid conformation of the human telomeric sequence than the parallel topology or dsDNA.

Once we knew the new compounds were selective for quadruplexes, we decided to explore the stabilization effect over another quadruplex forming sequences. To do so, we chose a range of different G4-forming sequences, RNA and DNA, both from human or non-human genome, and with different topologies. The nature and topology features for each one are described in the table **2.3**. Ligand concentration assayed was 30 µM in all cases, taking into account the previous results. Only two symmetric compounds (**44** with maltose and **46** with glcNAc) were included in order to check any behavior change in that group. F21T was also included as internal control and compound **1** was incorporated although its results are not shown because the stabilization at such concentration was too high with every sequence. Actually, it was not possible to determine the T_m in any case, except for FdxT. The ΔT_m results are described in figure **2.7a**.

Oligonucleotide name	Sequence	Origin	Structure type
F21T	FAM-G$_3$[T$_2$AG$_3$]$_3$-TAMRA	Human Telomeric sequence	Polymorphic
F25CebT	FAM-AG$_3$TG$_3$TGTA$_2$GTGTG$_3$TG$_3$T-TAMRA	Human minisatellite	Parallel
FmycT	FAM-T$_2$GAG$_3$TG$_3$TAG$_3$TG$_3$TA$_2$-TAMRA	c-myc promoter	Parallel
FBom17T	FAM-[G$_2$T$_2$A]$_3$G$_2$-TAMRA	Bombyx telomere	Anti-parallel
F21CTAT	FAM- G$_3$[CTAG$_3$]$_3$-TAMRA	Human DNA telomere (variant)	Anti-parallel
FdxT	FAM-[TA]$_2$GC[TA]$_2$-HEG-[TA]$_2$GC[TA]$_2$-TAMRA	Intramolecular duplex	Duplex
FHIV321T	FAM-T$_2$G$_2$C$_2$TG$_3$CG$_3$ACTG$_3$A-TAMRA	HIV PRO1 gene	Antiparallel
FHIV32T	FAM-CAG$_3$AG$_2$CGTG$_2$C$_2$TG$_3$CG$_3$A-TAMRA	HIV PRO2 gene	Polimorphic
FHIV21T	FAM-TG$_3$CG$_3$ACTG$_4$AGTG$_2$C-TAMRA	(LTRIV)	Parallel
F21RT	FAM-G$_3$[U$_2$AG$_3$]$_3$-TAMRA	RNA human telomeric sequence (TERRA)	Parallel
FTRF2RT	FAM-CG$_3$AG$_3$CG$_4$AG$_3$C-TAMRA	UTR region in TRF2 gene	Parallel

Table 2.5. G4-forming sequences included in the FRET-melting assays.
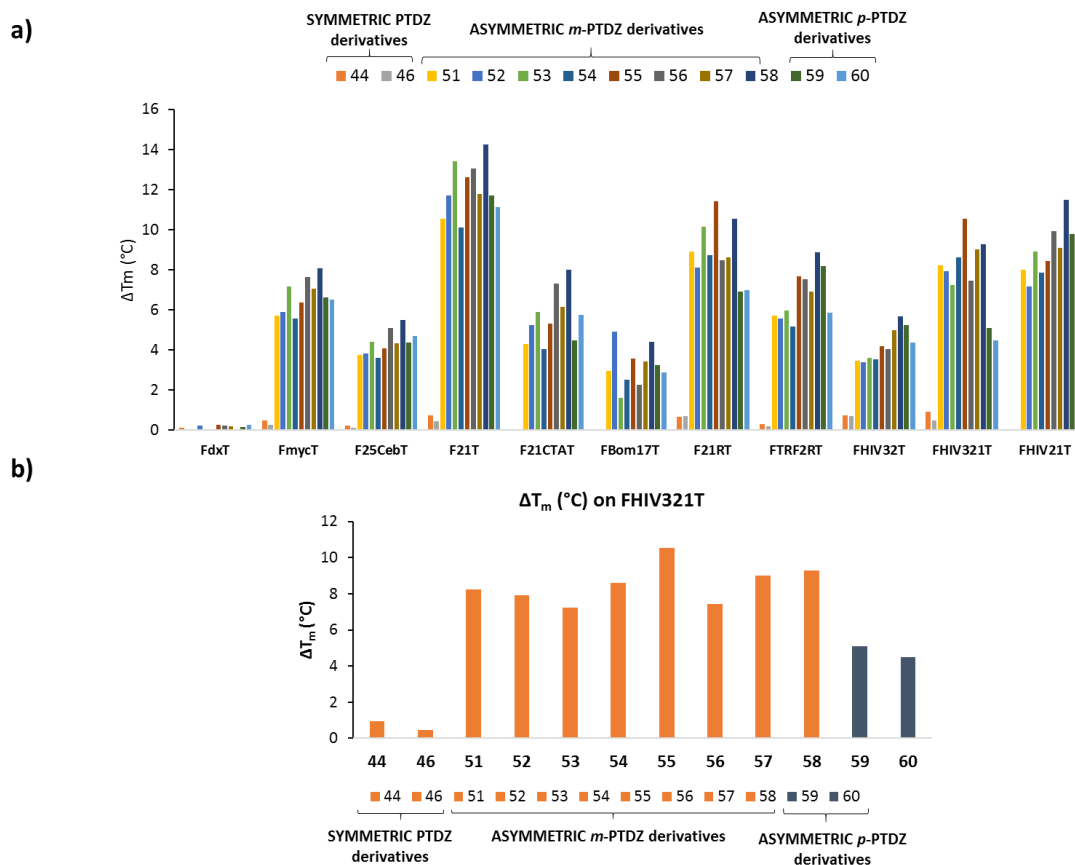
Figure 2.7. a) Bar graph showing the ΔT$_m$ (°C) obtained for each product and sequence. T$_m$ (°C) values for G4-sequences alone: F21T (50.1±0.43), F25CebT (68,5±0,22), Fmyc (66.8±2.9), FBom17T (43.8±0.38), F21CTAT (56.4±0.26) and FdxT (67.2±0.12), F21RT (51±4), FTRF2RT (62.8±1.1), FHIV32T (58.8±0.7), FHIV21T (55.9±0.9), and FHIV321T (55.1±0.6). b) Bar graph with the ΔT$_m$ (°C) for each product on FHIV321 sequence. *m*-PTDZ in orange bars and *p*-PTDZ in dark blue bars.

Once again, we observed no binding of the symmetric compounds to any G4-sequences. However, the asymmetric PTDZ ligands showed some selectivity towards the DNA-telomeric sequence (F21T), as well as its corresponding RNA-sequence; only compound **55 (asym-glcNAc-PTDZ)** showed a small preference for the RNA sequence. None of the ligands exhibited a good stabilization value with FBom17T, the *Bombyx* telomere sequence, which adopts an anti-parallel conformation. The asymmetric PTDZ compounds stabilized both parallel (F21RT) and antiparallel (HIV321T) quadruplex topologies and therefore we cannot conclude they were selective through a certain topology. Among the three HIV DNA-sequences, the asymmetric derivatives generally exhibited better results with the parallel FHIV21T G4-forming sequence than with FHIV321T (sequence corresponding to the promoter region of HIV-PRO1), but still very similar.

There were no many differences on G4-binding between the asymmetric ligands except for the FHIV321T and F21RT sequences, where the *p*-derivatives (**59** with maltose and **60** with glucose) showed lower ΔT_m than the corresponding *m*-derivatives (figure **2.7b**). It is possible that in these particular sequences the shape of the ligand may be important for the interaction. We studied in detail this phenomenon by NMR experiments in following sections for the interaction with FHIV321T.

The aglycone compound **58** displayed the same behavior than the carbohydrate modified PTDZ ligands, even with better results in most cases. Taking into account this fact, it seems the carbohydrate units are not playing a relevant role on the observed G-quadruplex selectivity.

Our reference compound (**1**) showed strong stabilization ability, hence we wanted to analyze its selectivity among the G4-sequences. To do so, we decreased the ligand concentration to 2 μM (figure **2.8**). Comparing with the results obtained for the carbohydrate-PTDZ ligands, its selectivity profile differed for some sequences: with Fmyc, F21RT, F21CTAT and HIV21T, the ligand displayed similar ΔT_m (about 20°C) as the second most stabilized sequences, while the new PTDZ derivatives showed lower selectivity towards Fmyc and F21CTAT sequences. Among the three HIV quadruplexes, compound **1** stabilized similarly FHIV32T and FHIV321T sequences, whilst the asymmetric PTDZ ligands displayed a clear preference for FHIV321T *versus* FHIV32T. It could seem that compound **1** was selective for quadruplex structures, however the small ΔT_m showed for the dsDNA FdxT may be a problem taking into account it is the most abundant DNA-topology inside the cell.
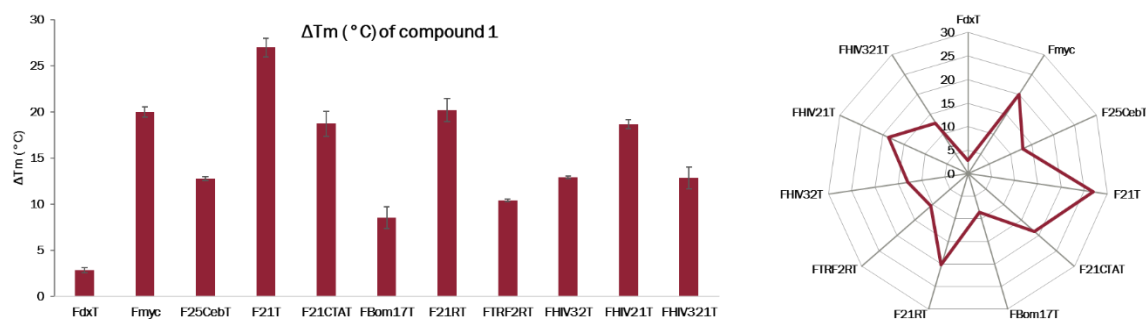
Figure 2.8. Selectivity profile for compound **1** in bars (left) and radar (right) graphs.

In conclusion:

- Symmetric PTDZ ligands were not able to stabilize any G4-sequences, indicating the necessity of unless one positive charge on the G4-ligand.
- The new asymmetric carbohydrate PTDZ ligands stabilized selectively quadruplex structures *versus* DNA duplexes. They were not strong G4-ligands in comparison to previously reported G4-binders such as PhenDC$_3$ or Neidle's control **1**, since their ΔT_m were lower. However, they were more selective towards quadruplex even at high concentration. They did not discriminate between different G4-topologies, although they showed certain degree of selectivity towards the DNA-human telomeric sequence (F21T).
- There was not a noticeable difference between the *m*- (**51-57**) and *p*-PTDZ derivatives (**59** and **60**), except for the FHIV321T sequence, where the *p*-derivatives presented ΔT_m 3-4 °C lower than the corresponding *m*- derivatives.
- The carbohydrate motif was not the responsible of the selectivity pattern, taking into account the results obtained for the aglycone PTDZ ligand (**58**).

3.2.2. <u>CIRCULAR DICHROISM (CD) EXPERIMENTS.</u>

As we described in the general introduction chapter, CD spectra provide information about secondary structures. Our goal was to investigate whether the new carb-PTDZ ligands modify the G-quadruplexes structure. With this purpose, titration experiments have been performed with some of the asymmetric carbohydrate PTDZ ligands.

The 22AG human telomeric sequence d(AG$_3$(T$_2$AG$_3$)$_3$) is commonly used in biophysical assays such as CD, and it differs only on one extra-adenine from the hTel sequence used in FRET studies. CD titration experiments have been performed with this sequence and 5 equivalents of several asymmetric carb-PTDZ ligands (glucose derivative **51** and glucose with linker C2 **53**), both in buffer containing either potassium or sodium salts. In potassium rich buffer, 22AG sequence presents a polymorphic topology: strong positive peaks at 290 and 250 nm, a positive shoulder around 260-270 nm and a negative signal at 235 nm. Under these conditions, the structure distortion produced by the ligands seems to be very weak, signals decreased slightly their intensity overall at 250 and 260 nm (figure **2.9a**). Compound **51** produced higher distortion on the G4-structure than the linker derivative **53**, although the change was small. On the contrary, 22AG adopts an antiparallel conformation at sodium buffer conditions, with positive signals at 295 and 245 nm, and a strong negative at 260 nm (figure **2.9b**). The same compounds with the 22AG in sodium buffer did not introduce any change in the signals pattern.
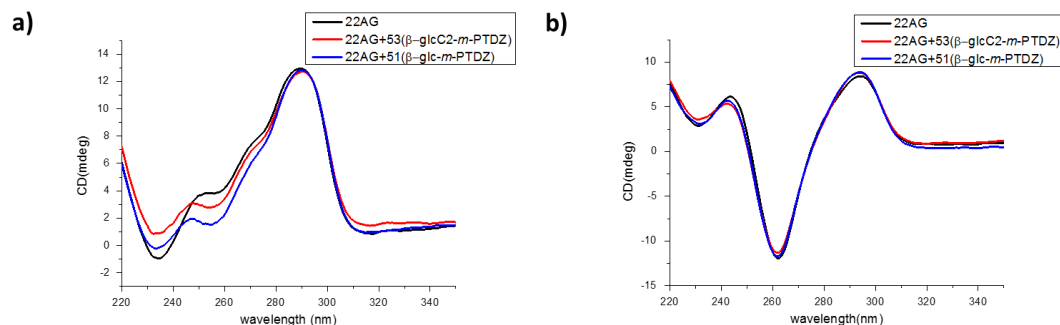
Figure 2.9. CD spectra with 22AG human telomeric sequence. a) 22AG at potassium buffer conditions alone (black) and with 5 equivalents of **53** (red) or **51** (blue). b) 22AG at sodium buffer conditions alone (black) and with 5 equivalents of **53** (red) or **51** (blue).

In conclusion, the antiparallel conformation was not affected by ligand titration while the hybrid one was slightly disturbed. This points out to different ligand interaction depending on the hTel topology.

We compared these results with the ones obtained for Neidle's compound **1**, analyzed by Garner *et al.*,[259] which added 4 equivalents to the hTel sample. The spectra pattern was clearly different from the previous ones, with a strong emerged band at 262 nm (figure **2.10**). The authors justified it by the increase of the parallel relative population, thus they proposed the ligand had preference for the parallel form of the G-quadruplex. These results disagree with our FRET melting results, where **1** did not displayed a clear preference for parallel G4-sequences.
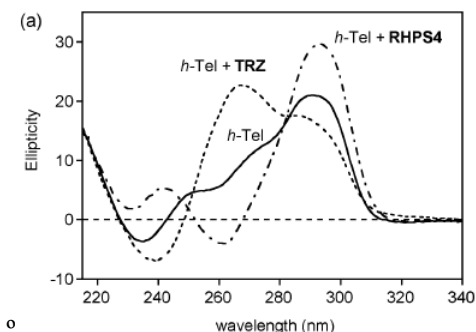
Figure 2.10. Image from the research developed by Garner et al.[259] CD spectra obtained from the titration experiment with 4 equivalents of TRZ (our compound **1**) and RHPS4.

Finally, we investigated the binding of several carbohydrate PTDZ ligands with two HIV sequences (HIV-ADR and HIV321), from the HIV-PRO2 and HIV-PRO1 sequences, respectively, found by Amrane *et al.*[96] 5 equivalents of each ligand were titrated and CD spectra were recorded (figure **2.11**).

The HIV-ADR sequence showed a hybrid conformation: two positive signals at 290 and 260 nm, and a negative one at 235 nm. The addition of the ligand **54** (fucose) did not show any modification in the spectrum (figure **2.11a**).
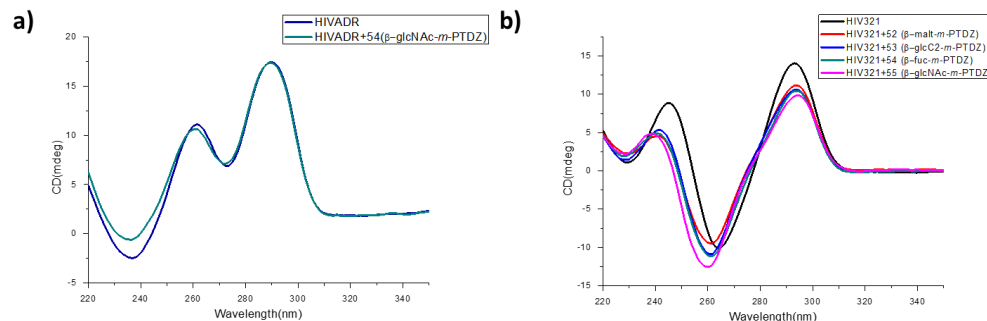
Figure 2.11. a) CD spectra with HIV-ADR alone (dark blue), and with 5 equivalents of **54** (asym fuc-*m*-PTDZ). b) CD spectra alone (black) and with 5 equivalents of **52** (asym malt-*m*-PTDZ), **53** (asym glcC2-*m*-PTDZ), **54** (asym fuc-*m*-PTDZ), or **55** (asym glcNAc-*m*-PTDZ). Buffer condition were 70 mM KCl and 20 mM potassium phosphate buffer pH 6.9.

The HIV321 sequence folds into an antiparallel G4: positive peak at 290 and 245 nm, and a strong negative at 260 nm, similar to the hTel sequence at sodium rich conditions. Again, we carried out titration experiments with different asymmetric carb-PTDZ products (**52**, malt-*m*-PTDZ, **53**, glcC2-*m*-PTDZ, **54,** fuc-*m*-PTDZ and **55,** glcNAc-*m*-PTDZ) and all of them behaved similarly showing a small distortion of the CD-spectrum (figure **2.11b**). The ligand glcC2-PTDZ **53** did not produce such distortion on the antiparallel form of the human telomeric sequence (22AG at sodium conditions), therefore the structure modification was not dependent on the parallel or antiparallel topology, it must be due to the interaction with another G4-structural element such as the grooves or loops (different for each G4-sequence).

CD spectroscopy allows the determination of the quadruplex T_m both alone or in presence of a ligand. Sometimes FRET melting assays provide false positive results when compounds interact with the fluorophores in the G4-probe. Furthermore, if the achieved ΔT_m in a FRET melting assay is very low, it is recommendable to confirm it by another technique to discard potential false positives. Taking into account the previous ΔT_m results, we performed **CD-melting** experiments with two human telomeric sequences (22AG and 24TTA) in order to check the ΔT_m at two different buffer conditions (sodium and potassium, respectively). Compound **53** (asym glcC2-*m*-PTDZ) was selected for this study.

We carried out melting experiments for each G4-sequence in the absence and in the presence of the PTDZ ligand **53**. The CD signal corresponding to 295 nm wavelength was followed at each recorded temperature. The T_m was obtained from the fitted resulting curve for each sample. The T_m for the 22AG sequence was 55.3 °C and with the ligand 59.6 °C, so the resultant ΔT_m was 4.3°C (figure **2.12a** and **b**). We repeated the experiment with the 24TTA human telomeric sequence at potassium rich buffer and the resulting T_m for the G4-sequence was 55.0 °C using 290 nm as reference wavelength (figure **2.12c** and **d**). However, when the compound was incorporated, the spectrum shape was unusual. The corresponding curve at 290 nm did not fit properly and the calculated T_m was exactly the same than without ligand. It is difficult to draw conclusions from this experiment; the ligand may destabilize this G4 structure or there was some manipulation mistake. It would be possible to demonstrate it by carrying out the reverse assay, where the sample with the G4-sequence and ligand is heated and the CD signals are recorded during the progressive cooling process.
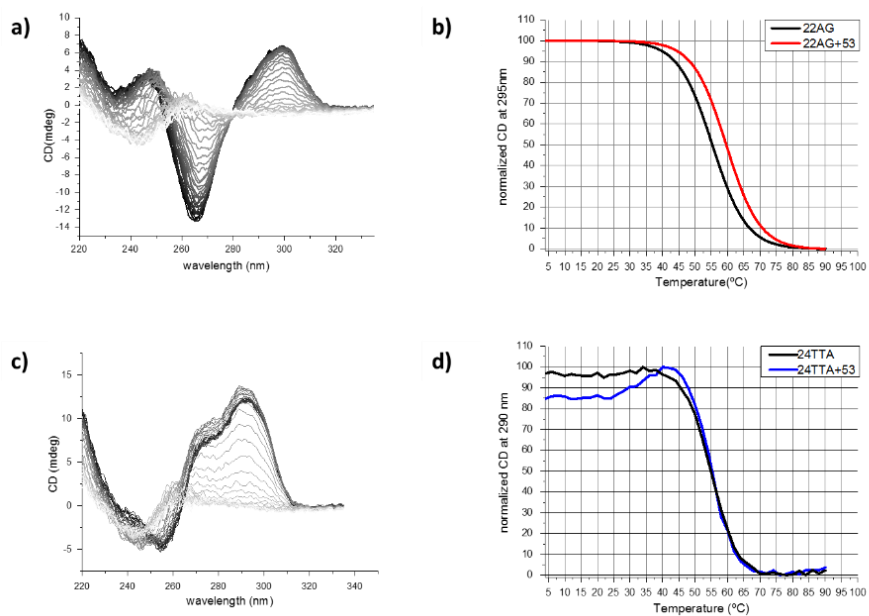
Figure 2.12.a) 22AG CD spectra at each recorded temperature. B) 22AG melting curve at 295 nm, alone (black) and with 3 equivalents of **53** (red). C) 24TTA CD spectra at each recorded temperature. D) 24TTA melting curve at 290 nm, alone (black) and with 3 equivalents of **53** (blue).

We can conclude from the CD experiments:

- All the asymmetric PTDZ ligands did not induce a severe distortion on the hTel-CD signal profile, neither the hybrid nor the antiparallel conformation. The signals were barely displaced on the hybrid form while the antiparallel remained intact. That could mean small changes in the mode of interaction with both topologies.

- The antiparallel conformation of the HIV321 sequence was not altered by any compound assayed. The changes on the CD-spectrum were not topology type-dependent.

- The CD melting assay confirmed the carbohydrate PTDZ ligands were capable of stabilizing the 22AG human telomeric-G4 sequence. Results were not conclusive with the 24TTA sequence at potassium rich buffer conditions.

3.2.3. SPR (SURFACE PLASMON RESONANCE) EXPERIMENTS.

Some of the asymmetric PTDZ ligands have been screened over different G-quadruplex forming sequences using Surface Plasmon Resonance. This technique provides the thermodynamic and kinetic parameters of DNA-ligand interactions, including K_d and the interaction profile as specific or non-specific ligand. Although the methodology has not been completely optimized for quadruplex, it provides additional information about binding.

The biotinylated DNA sequences assayed are described in the corresponding experimental section and were attached on a streptavidin surface. The hTel 22AG, two HIV sequences and c-myc were the selected G4s for this experiment. A duplex sequence and a mutated 22AG-not able to fold into a quadruplex structure- were also included as controls to discriminate between G4 and another DNA-structures. The selected PTDZ ligands were: **53** (glcC2-*m*-PTDZ), **52** (malt-*m*-

73

PTDZ), **54** (fuc-*m*-PTDZ), **59** (malt-*p*-PTDZ), **58** (aglycone-PTDZ) and **1** (our control) at varying concentrations.

Firstly, it was necessary to assure the ligands did not interact with the streptavidin surface. To do so, compounds solutions at different concentrations were flown into the potassium running buffer to find the most tolerable concentration. This value was 50 µM for all the new carbohydrate PTDZ products. However, Neidle's compound increased the signal at higher concentration than 20 µM. Considering these results, the new asymmetric ligands were assayed at 20, 16, 12, 8, 4, 2, and 0.5 µM; and the control **1** at 4, 2 ,1, 0.5, 0.25 and 0.125 µM concentrations.

The response signal depends on the molecular weight, hence comparing results between different weight compounds is not accurate. For this reason, we compared the results for each compound between the different sequences and the general selectivity profile among them. So, we carried out titration experiments for each ligand with four G4-sequences (22AG, c-myc, HIV32 and HIV321), the mutated 22AG (22AG mut) and the duplex control (Hairp) (figure **2.13**).
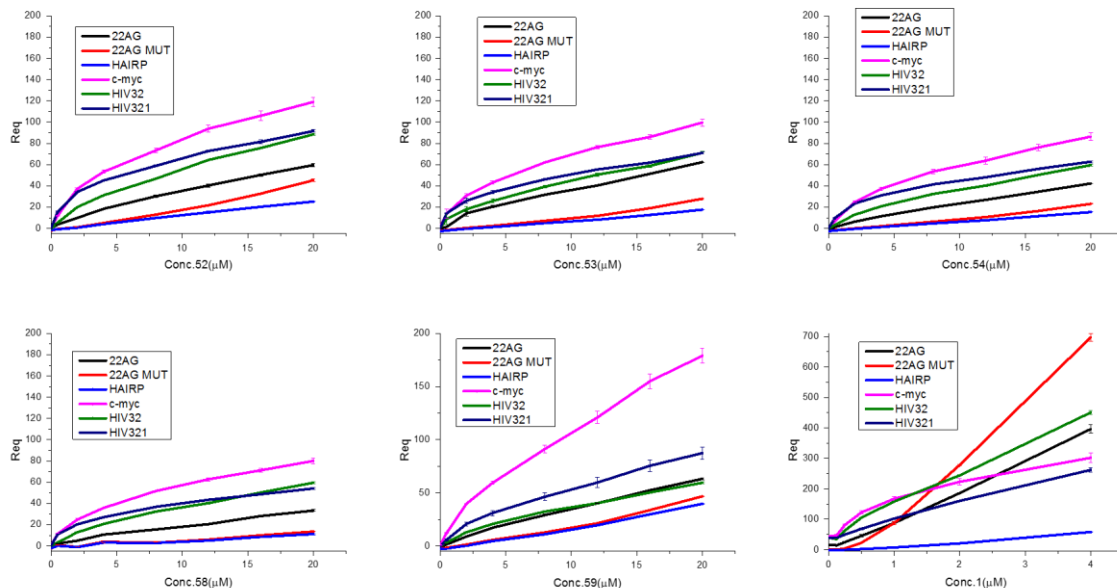
Figure 2.13. Titration curves obtained with the asymmetric carb-PTDZ ligands in 20 mM potassium phosphate pH 6.9 running buffer supplemented with 70 mM potassium chloride and 0.05% (V/V) Tween®. PTDZ ligands assayed: **53** (glcC2-*m*-PTDZ), **52** (malt-*m*-PTDZ), **54** (fuc-*m*-PTDZ), **59** (malt-*p*-PTDZ), **58** (aglycone-PTDZ) and **1** (Neidle's control).

The common trend in all of them was the unusual titration shape curves, not characteristic of specific binding ligands since the saturation area was not reached. Neither they corresponded to an unspecific binding because they did not resemble a straight line. Instead, they showed a mixed shaped curve with specific beginning and an unspecific ending. Only experiments with the duplex and mutated sequences exhibited a straight curve.

The PTDZ ligands **52**, **53**, **54** and the aglycone **58** behaved very similarly, with the following signal response intensities order: c-myc>HIV321>HIV32>22AG>mutated 22AG>Hairpin. The glcC2-PTDZ **53** ligand showed lightly higher response difference between the G4-sequences and the non-G4 sequences, which may be explained by higher selectivity

towards G4-structures. The *p*-PTDZ ligand **59** changed slightly the general trend with signal response in the order: c-myc>HIV321>HIV32=22AG>mutated 22AG>Hairpin. The response signal with c-myc was significantly larger than for the rest of the *m*-PTDZ ligands, becoming the most selective one for this quadruplex forming sequence. On the contrary, our reference compound **1** provided actual different response order profile: mutated 22AG>HIV32>22AG>c-myc>HIV>321>Hairpin. It showed higher signal responses at lower concentration range than the rest of the PTDZ ligands, showing the highest response with the mutated 22AG sequence. These results make the ligand **1** non-selective for G4-structures. In addition, its curves were more typical of non-selective than selective ligand, except for the c-myc sequence.

In general, the results were not as expected especially when we compared with the FRET melting results, since the selectivity profiles were dissimilar between both techniques. FRET assays showed the asymmetric ligands bind more strongly to hTel while SPR preferred the c-myc sequence. In any case, both techniques FRET and SPR showed the carbohydrate PTDZ ligands display low binding for the dsDNA hairpin control. This effect is more clear using FRET than measuring binding through SPR.

It is possible to determine the **K$_d$** from fitted titration curves. Depending on the curves shape the fitting is made by a typical Michaelis-Menten model (specific ligands) or a modified/mixed model (mixed specific+non-specific behavior). Curves showed a mixed specific and non-specific shape, therefore we tried to fit the curves with both models obtaining better fitting results for the second one. The equations applied to fit the curves through a specific and a modified model were:

$$ES = \frac{Emax * s}{Kd + s}$$

$$ES = \frac{Emax * s}{Kd + s} + R * s$$

Where:

ES: response (Req)

E$_{max}$: maximum concentration

S: concentration

R: non-specific component constant

K$_d$: dissociation constant

In some cases, it was not possible to fit the curve with any model since the shape did not resemble neither a specific nor a modified model, so we specified NA (not available) on the summarize table (table **2.4** and figure **2.14**). In general, the non-specific model provided lower K$_d$ values than the specific ones, but with better R results. In both cases the K$_d$ values remained in the micromolar range.

52: malt-*m*-PTDZ

	Kd SPECIFIC	ST	R	Kd MIXED	ST	R
22AG	30.89426	4.69924	0.99752	4.63719	1.1519	0.99966
22AGMUT	NA	NA	NA	NA	NA	NA
HAIRP	NA	NA	NA	NA	NA	NA
c-myc	8.44	1.37	0.99086	2.49	0.48243	0.99878
HIV32	18.75	3.29172	0.99434	3.26	1.04835	0.99878
HIV321	5.49	1.25371	0.97676	1.1	0.14144	0.99919

53: glcC2-*m*-PTDZ

	Kd SPECIFIC	ST	R	Kd MIXED	ST	R
22AG	25.6	8.426	0.98672	2.23	1.85751	0.99463
22AGMUT	NA	NA	NA	NA	NA	NA
HAIRP	NA	NA	NA	NA	NA	NA
c-myc	8.2	1.66116	0.98498	1.61539	0.37666	0.99817
HIV32	14.5	3.97757	0.98154	0.89106	0.3017	0.99813
HIV321	5.38	1.40046	0.96904	0.78375	0.15237	0.99818

54: fuc-*m*-PTDZ

	Kd SPECIFIC	ST	R	Kd MIXED	ST	R
22AG	55.27042	15.660066	0.99619	3.76961	3.93161	0.99762
22AGMUT	NA	NA	NA	NA	NA	NA
HAIRP	NA	NA	NA	NA	NA	NA
c-myc	9.64	1.62233	0.99102	2.73	0.52639	0.99898
HIV32	19.71	4.06542	0.99254	2.85	0.87292	0.99897
HIV321	5.43	1.18253	0.97897	1.25321	0.10228	0.99967

59: malt-*p*-PTDZ

	Kd SPECIFIC	ST	R	Kd MIXED	ST	R
22AG	61.22	14.27523	0.99777	4.94561	4.20739	0.99872
22AGMUT	NA	NA	NA	NA	NA	NA
HAIRP	NA	NA	NA	NA	NA	NA
c-myc	24.01052	6.00141	0.99084	1.62628	0.5537	0.99901
HIV32	19.71	4.06542	0.99254	2.85	0.87292	0.99897
HIV321	19.12	4.61063	0.98941	1.87814	0.56526	0.99895

58: aglycone-PTDZ

	Kd SPECIFIC	ST	R	Kd MIXED	ST	R
22AG	44.76938	18.64	0.9885	1.25	1.71754	0.99342
22AGMUT	NA	NA	NA	NA	NA	NA
HAIRP	NA	NA	NA	NA	NA	NA
c-myc	7.72	1.33463	0.98873	2.01	0.46259	0.99806
HIV32	19.71	4.06542	0.99254	2.85	0.87292	0.99897
HIV321	4.92	1.10241	0.97628	1	0.21523	0.99745

1: control

	Kd SPECIFIC	ST	R	Kd MIXED	ST	R
22AG	NA	NA	NA	NA	NA	NA
22AGMUT	NA	NA	NA	NA	NA	NA
HAIRP	NA	NA	NA	NA	NA	NA
c-myc	1.07	0.33496	0.94212	0.41	0.33219	0.94823
HIV32	6.5	3.35056	0.96793	0.19595	0.33822	0.97999
HIV321	3.3	1.72368	0.92358	0.1	0.2167	0.94529

Table 2.6. Summarize tables with K_d (μM) obtained from both specific and modified models. ST (standard deviation).NA (not available).
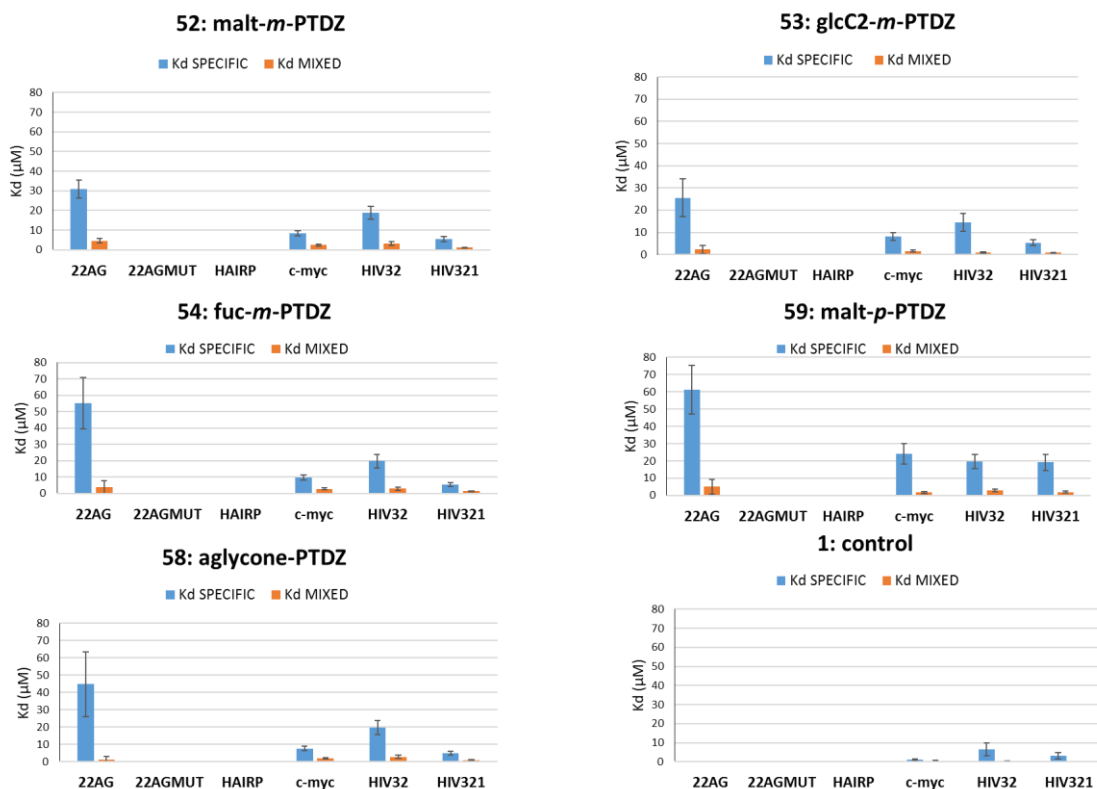
Figure 2.14. Summarize graphs for the Kd (μM) obtained from both models.

The SPR technique is not well optimized to measure ligand binding to G-quadruplexes yet, so we have to take into account several limitations before concluding anything about the results:

• These equations are accurate only when there is just one binding site and at 1:1 stoichiometry complexes. If the observed RU_{max} (maximum response) is consistent with the calculated from the equation, the stoichiometry is 1:1 and the resulting K_d by the modified model is reliable.

• The relation between the molecular weight and the maximum response of the ligand and quadruplex must be:

$$\frac{RUmax\ ligand}{RUmax\ G4} = \frac{Mw ligand}{Mw G4}$$

When we applied this relation, sometimes it was pretty close to one, but in other cases we found higher or lower ratios. In all *m*-derivatives and the HIV321 sequence relation remained 1:1, however there was variability for the rest. Tables with this relation for each ligand are shown below (table **2.5**).

52	RUmax	RUobserved	RUobs/RUmax
22AG	41.953186	20.5	0.488639886
cmyc	53.2499206	67.6	1.269485461
HIV321	46.4620348	45.6	0.981446469
HIV32	40.8388139	35.2	0.861925131

53	RUmax	RUobserved	RUobs/RUmax
22AG	35.3719116	19.5	0.551284879
cmyc	44.8965064	46.5	1.035715331
HIV321	39.1734489	31.9	0.814327073
HIV32	34.4323531	19.4	0.56342359

54	RUmax	RUobserved	RUobs/RUmax
22AG	32.0252647	8.05	0.251364042
cmyc	40.648708	47.1	1.158708416
HIV321	35.4671268	32.6	0.919161008
HIV32	31.1746008	21.05	0.675229175

58	RUmax	RUobserved	RUobs/RUmax
22AG	26.3356864	4.8	0.182262194
cmyc	33.4270969	43.2	1.2923647
HIV321	29.1660705	27.8	0.95316234
HIV32	25.6361507	21.05	0.821106112

59	RUmax	RUobserved	RUobs/RUmax
22AG	41.953186	13.2	0.314636414
cmyc	53.2499206	44	0.826292312
HIV321	46.4620348	26.3	0.566053555
HIV32	40.8388139	21.05	0.515441023

1	RUmax	RUobserved	RUobs/RUmax
22AG	35.9342374		
cmyc	45.6102497	191.7	4.203002638
HIV321	39.7962098	54.3	1.364451547
HIV32	34.9797423	71.2	2.035463823

Table 2.7. RUmax/RU observed relation for each compound assayed.

• We have to consider that it is difficult to know the amount of G4 attached to the surface and properly folded, what could lead to lower responses and therefore to false negatives.

Taking these cosiderations into account, we can conclude:

- The asymmetric PTDZ ligands analyzed bound preferentially to the G4-forming sequences in comparison to DNA duplex or another non-quadruplex structures (22AG mutated), whilst Neidle's control ligand (**1**) did not show such selectivity.
- The binding behavior was not conclusive due to the abnormal titration curves, so the calculated K_d values are only an approximation.

3.2.4. MASS SPECTROMETRY.

Mass spectrometry can help us to confirm G4-ligand binding and provide the ligand-target complex stoichiometry. For this reason, we examined one asymmetric PTDZ derivative (**53**, glcC2-*m*-PTDZ) with several DNA sequences: the human telomeric 22AG d(AG$_3$(T$_2$AG$_3$)$_3$), a duplex sequence d(CGCGA$_2$T$_2$CGCG)$_2$, a parallel tetramolecular G4 d(TG$_4$T)$_4$, and a bimolecular antiparallel sequence d(G$_4$T$_4$G$_4$)$_2$. These experiments were performed by Dr. Frédéric Rosu during my short stay in the IECB (Bordeaux).

We analyzed 1:1 ligand:DNA complexes by injecting them in the corresponding running buffer. With TMAA (trimethylammonium acetate) buffer, the free 22AG and the complexes were detected at the 5$^-$ charge state, with 1:1 and 1:2 stoichiometry complexes (figure **2.15a**). At 4$^-$ charge state, we identified the same complexes, as well as the 3:1, though in lower proportion (figure **2.15b**).
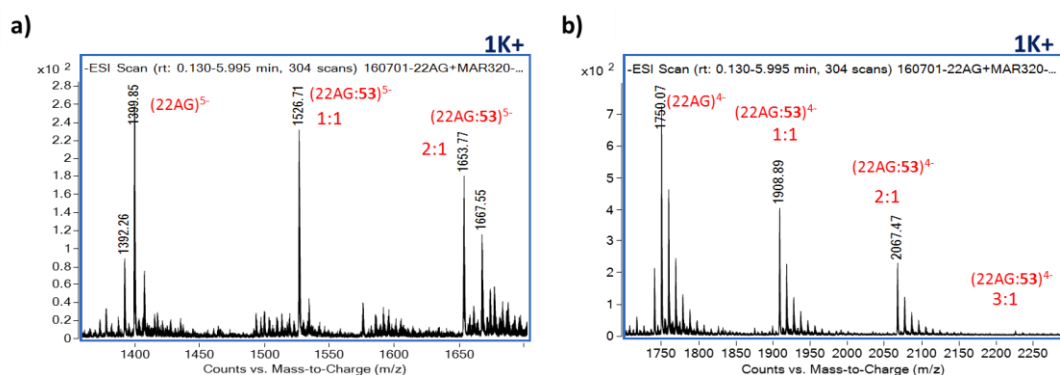
Figure 2.15. Mass spectra for the potassium at 5$^-$ (a) and 4$^-$ (b) charge stages. Samples were prepared with 10 μM DNA concentration and 10 μM ligand **53** concentration in 100 mM TMAA running buffer supplemented with 1 mM KCl.

22AG and the rest of the sequences were studied in ammonium acetate buffer as 1:1 complexes solution (figure **2.16**) with compound **53** (glcC2-*m*-PTDZ). The different topologies of these sequences allowed to study the ligand selectivity for a certain structural configuration.
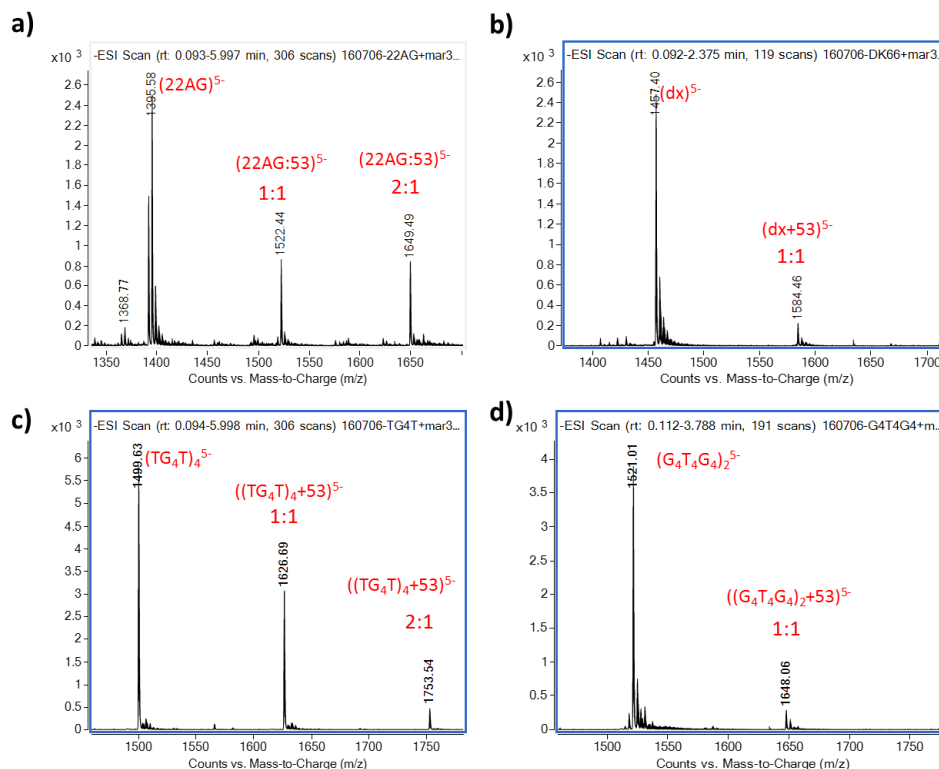
Figure 2.16. Mass spectra in ammonium acetate buffer, with **53** and different DNA-sequences at 5⁻ charge stage: a) 22AG. b) duplex sequence. c) the tetramolecular parallel sequence. d) the bimolecular antiparallel sequence. Samples were prepared with 10 μM DNA concentration and 10 μM ligand **53** concentration in 100 mM AcNH$_4$ running buffer.

The spectrum with the DNA duplex sequence showed a small peak of the ligand:DNA complex indicating worse binding affinity than with the G4-forming sequences (figure **2.16a** and **b**). When we compared the spectra of different G4s, the **53** ligand appeared to have a preference for binding G4 parallel topologies. In fact, the antiparallel structure possesses two diagonal loops which may hinder the ligand access to the G4-tetrads, what could explain the ligand preference.

3.2.5. NMR EXPERIMENTS.

The previous biophysical techniques we have used provided information about the affinity, selectivity, as well as stoichiometry and kinetics of the new carb-PTDZ derivatives for different G4-structures. Alternatively, NMR spectroscopy can provide more structural details of the corresponding complexes such as the quadruplex affected areas due to ligand binding. We carried out titration experiments following the binding by ^{1}H-NMR spectroscopy. We observed chemical shift changes and differences on the shape of protons of the G4 in absence and in the presence of the ligands. We investigated the binding of the carb-PTDZ ligands on four G4-sequences: two HIV sequences and two from the human telomeric sequence, the last ones have been studied in more detail.

3.2.5.1. NMR studies with **HIV-ADR** sequence.

The HIV-ADR sequence d(AG$_3$AG$_2$TGTG$_2$C$_2$TG$_3$CG$_3$) belongs to the HIV-PRO2 DNA.[96] It is very similar to the FHIV32T sequence examined in the FRET melting experiments. Firstly,

we performed a simple ^{1}H-NMR experiment with the HIV-ADR folded sequence, followed by addition of two equivalents of the asymmetric **52** (malt-*m*-PTDZ) or **54** (fuc-*m*-PTDZ) ligands. The spectrum showed twelve imino single peaks corresponding to the formation of three quartets, and an additional signal at lower field which belongs to a base pair. The complete structure resolution has not been reported yet, but according to the previous CD experiments, this G4 folds into a hybrid topology. Ligand addition leads to signal broadening, where G21, G11, G17, G4, G9 and G18 almost fully disappear and all of them show a general signal shift (figure **2.17a**). The base pair was also affected. Regarding these spectral changes, both compounds seem to interact with the G4 and affect its global structure. It is interesting to note that both compounds (containing maltose and fucose) provided almost the same changes, what seems to point out that the carbohydrate nature was not important for the binding. A stepwise titration experiment was carried out with another asymmetric PTDZ compound (**55**, glcNAc-*m*-PTDZ) (figure **2.17b**). We observed a similar shifting behavior as in the case of **52** or **54** with G16 and G20 as the least shifted signals.
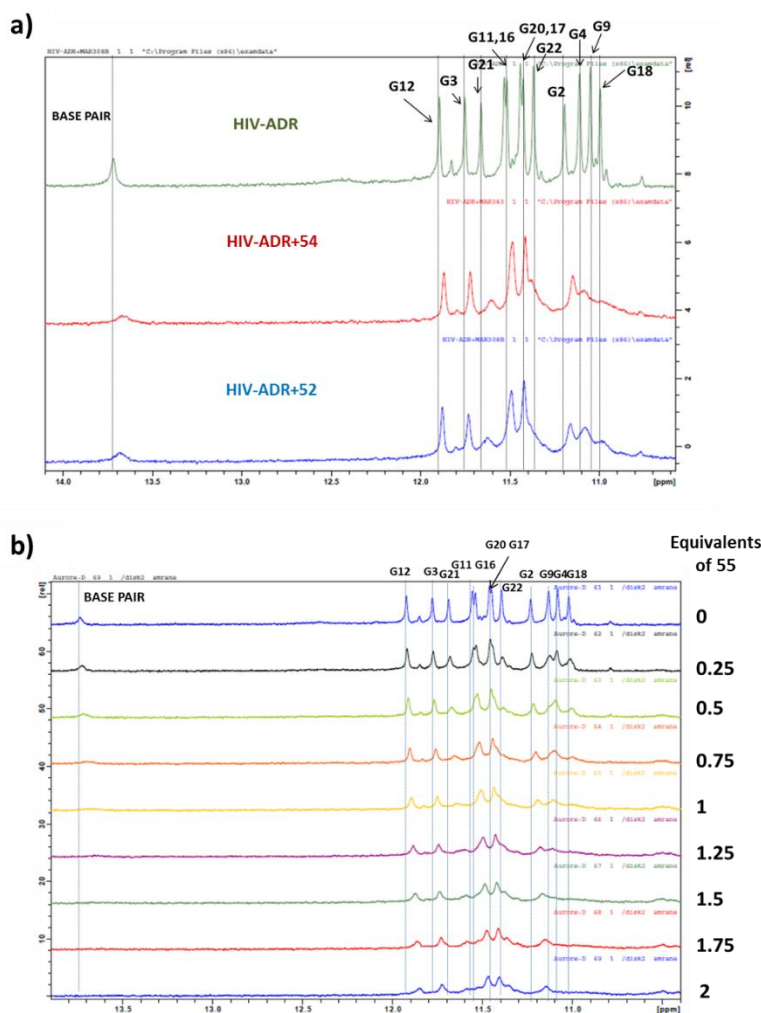
Figure 2.17. Titration experiments with HIV-ADR sequence. a) Spectra with sequence HIVADR alone (green) and two equivalents of **54** (fuc-*m*-PTDZ, red) or **52** (malt-*m*-PTDZ, blue). b) Stepwise titration with **55** (glcNAc-*m*-PTDZ).

3.2.5.2. NMR studies with **HIV321** sequence.

We followed the same experimental approach with the HIV321 sequence. Its preliminary solved structure was recently reported by Amrane *et al.*[96] As we previously described in the general introduction, this quadruplex is formed by two quartets (eight imino signals) with an additional base pair. We firstly carried out a simple titration with 2 equivalents of asymmetric PTDZ ligands with different sugar motifs: **52** (malt-*m*-PTDZ), **53** (glcC2-*m*-PTDZ), **54** (fuc-*m*-PTDZ) and **55** (glcNAc-*m*-PTDZ), in order to explore the possible differences due to sugar motifs (figure **2.18**). However, we observed very similar spectra for all the ligands:

- Four imino signals completely disappeared (G2, G18, G13 and G9) after ligands addition. Only the compound **55** (GlcNAc-*m*-PTDZ) behaved slightly different, affecting more the G12 peak, and signals were more shifted.
- G17 and G8 were shifted upfield.
- G12 and G3 showed smaller changes.
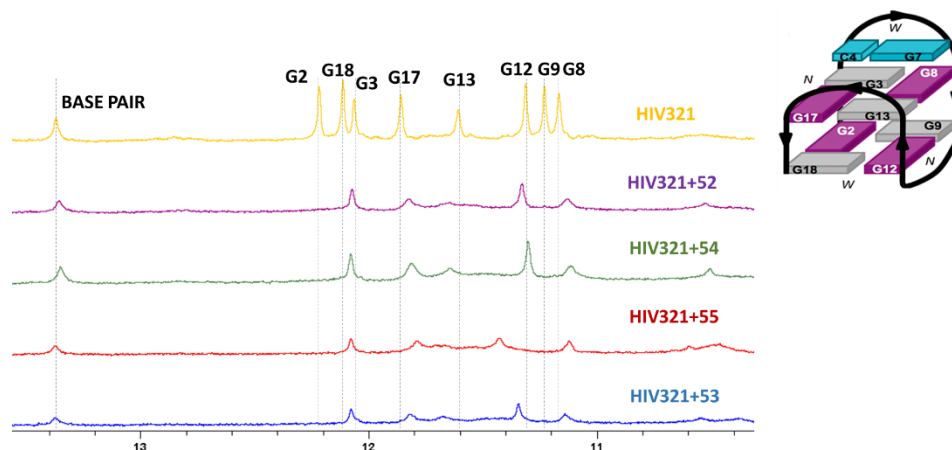
Figure 2.18. HIV321 spectra alone (yellow) and with 2 equivalents of **52** (malt-*m*-PTDZ, purple), **54** (fuc-*m*-PTDZ, green), **55** (glcNAc-*m*-PTDZ, red) or **53** (glcC2-*m*-PTDZ, blue).

Since the partial structure resolution for the HIV321 sequence is known, we decided to explore more in detail the ligand binding. For this reason, we performed stepwise titrations with every type of ligands: **54** (fuc-*m*-PTDZ ligand), **53** (glcC2 *m*-PTDZ ligand), **60** (glc *p*-PTDZ ligand) and our reference compound **1**.

An example of *m*-derivatives titration results with **54** are shown in figure **2.19**. Changes were noticed from the first addition step, some signals were disappearing and shifting, while G2 and G18 signals fully disappeared at one equivalent of added ligand. The bidirectional shift indicated this effect was not due to factors such as temperature, but due to ligand binding. G3, G13, G17 and G9 shifted and progressively disappeared, and G8 decreased and shifted with compound addition. On the contrary, G3 and G12 were nearly unaffected. At 1.5 equivalents of ligand G13H1 started to appear again. Changes were detected also in the aromatic signals, which disappeared and shifted. An aliphatic signal disappeared in that region, maybe as a result of the interaction with the methyl group from a thymine located in the loop.
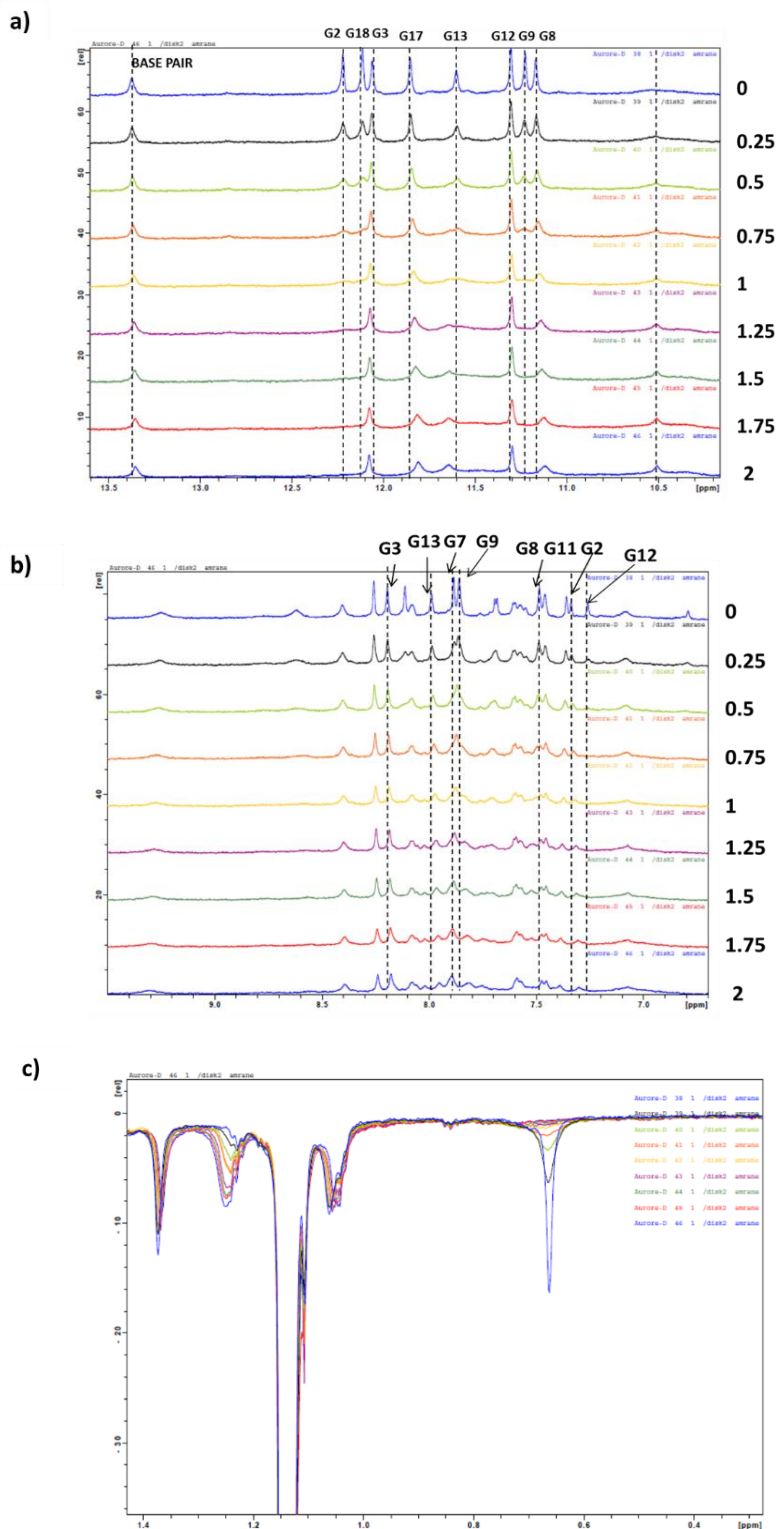
Figure 2.19. Example of titration experiment with HIV321 and **54** (glcNAc-*m*-PTDZ). a) Imino region. b) Aromatic region. c) a signal disappeared in the aliphatic region.

A summarized graph shows the chemical shift perturbation (Δδ in absolute value) for imino and aromatic proton signals in all performed titrations with the four selected compounds (figure

2.20). The experiments showed that all the *m*-PTDZ ligands **54** (glcNAc-*m*-PTDZ) and **53** (glcC2 *m*-PTDZ ligand) presented similar chemical shift changes in the imino region. When proton signal disappeared, it was not possible to calculate Δδ and they are highlighted with red squares in the graph.
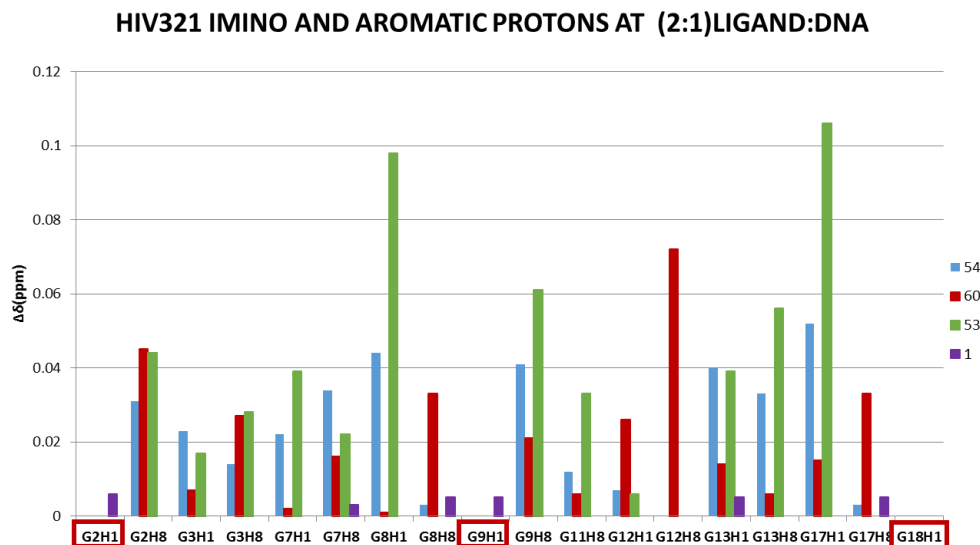
Figure 2.20. Bar graph with Δδ (ppm) for each imino and aromatic proton in the 2:1 (ligand:HIV) complexes. Disappeared signals are highlighted in red squares.
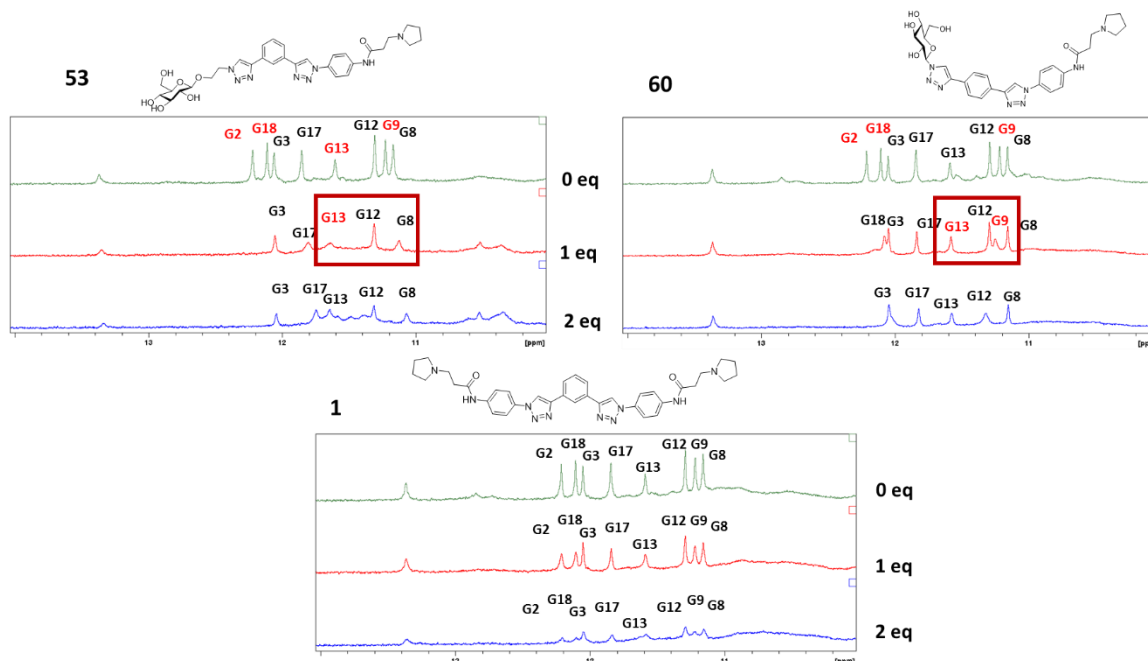
Figure 2.21. Titration experiments with the PTDZ **53**, **60** and control **1**. Red squares highlight the imino peaks difference between the *m*-(**53**) and *p*-(**60**) PTDZ ligands.

To summarize, the most affected signals (disappeared peaks) were G2H1, G9H1 and G18H1 from one tetrad, and G13H1 from the upper tetrad. **53** (glcC2-PTDZ) was the compound

with the largest Δδ of the series, which suggests that the linker may provide more flexibility to the sugar to interact with the quadruplex.

Titration with the control compound **1** provided completely different results: compound addition progression lead to a general broadening in all signals spectrum, however any of them were shifted or disappeared (figure **2.21**).

The *p*-PTDZ ligand **60** did not show such changes on G9H1 and G13H1 (red squares in figure **2.21**). This is an interesting result because guanine 9 is on the opposite side of the G-tetrad and guanine 13 was the only one which belongs to the upper tetrad. On the contrary, compound **60** is the second ligand with largest Δδ of this series and additionally shifted the G12H1 and G12H8 protons while the *m*-derivatives barely affected them. These results indicate the different binding mode of the *m*- and *p*-PTDZ derivatives with this G4-sequence. Figure **2.22** shows the PDB HIV321 structure where guanines are colored depending on their changes upon ligand binding: red represents guanine residues which imino signals disappeared; green represents guanine residues which imino signals show the highest chemical shift; and blue represents guanines which imino signals are barely shifted.
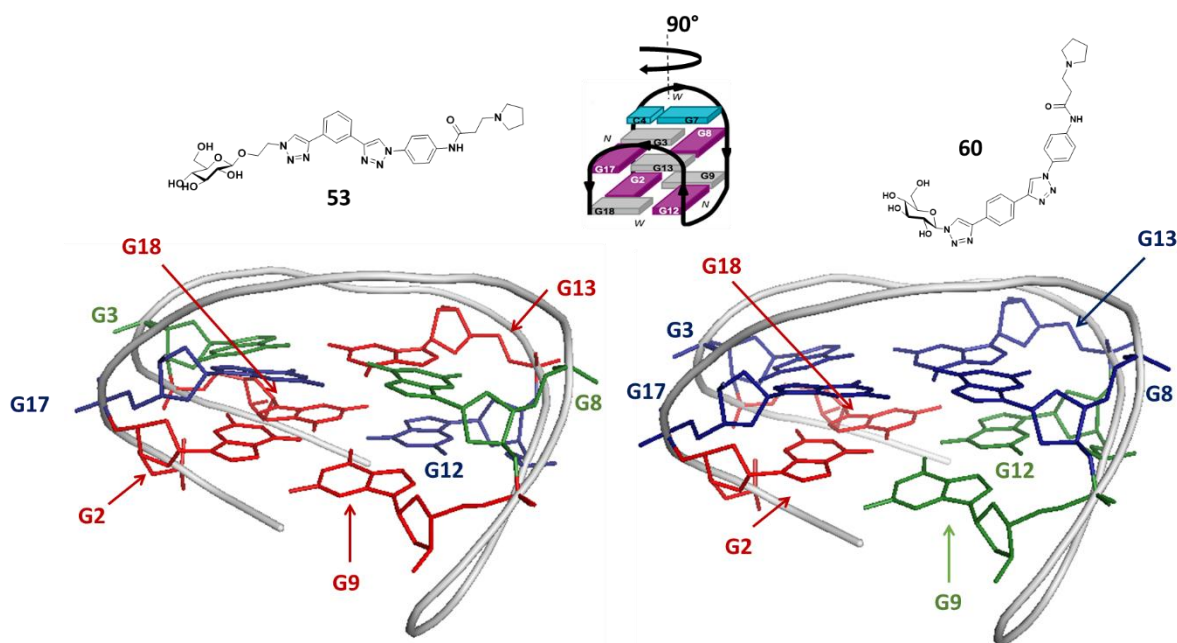
Figure 2.22. HIV321-PDB schematic pictures with the two asymmetric PTDZ ligands **53** and **60** examined. Guanines are colored according to the changes observed in the imino signals upon ligand binding: The red represents the guanines with disappeared imino protons, green represents the guanines with the most shift imino protons and blue represents the guanines with less affected imino protons.

3.2.5.3.   NMR studies with **24TTA** sequence.

Next, we studied the interaction of the carbohydrate modified PTDZ ligands with two sequences of the human telomere: 24TTA d($T_2G_3(T_2AG_3T_2AG_3)_3A$) and 22AG d($G_3(T_2AG_3)_3$).

The G4-hTel has been repeatedly studied by NMR, both in sodium and potassium buffer conditions. We find a complex spectrum when it is recorded at potassium buffer, what seems to

indicate the sequence is a mixture of conformers. However, some terminal residues can favor or disfavor a particular form by interacting with the loops, hence a 24 nt sequence has been used to favor and characterize the (3+1) form 1 in solution and a 25 nt sequence to the form 2.

We worked with the 24 nt 24TTA sequence in potassium buffer to explore the binding of the G4-ligands to the form 1. 24TTA presents an intramolecular (3+1) core fold with distributed G-tetrads as follows: one tetrad *anti·syn·syn·syn* and the other two tetrads *syn·anti·anti·anti*.[37] The disposition of the guanines inside the three quartets is G9·G3·G21·G17, G10·G16·G22·G4, and G11·G15·G23·G5, respectively (figure **2.23**). The folded structure has one double-chain reversal (T6-T7-A8) and two edgewise (T18-T19-A20 and T12-T13-A14) loops, thus three G-tracts are oriented in one direction (G9-G11, G3-G5, and G21-G23) and the fourth (G15-G17) in the opposite direction.
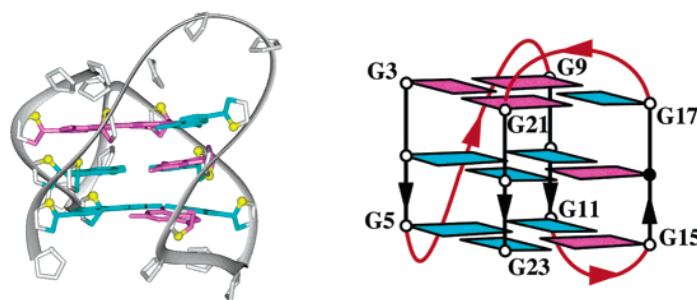
Figure 2.23. Ribbon and schematic views of a representative telomeric 24 TTA sequence refined structure.

We performed titrations with **53** (glc *m*-PTDZ) and with **59** (malt *p*-PTDZ), to compare a G4 ligand with the *m*-substitution at the central benzene with another with a *p*-substitution. We focused exclusively on the chemical shift of the guanine-imino protons already assigned by Phan *et al.*[37] (figure **2.24**).
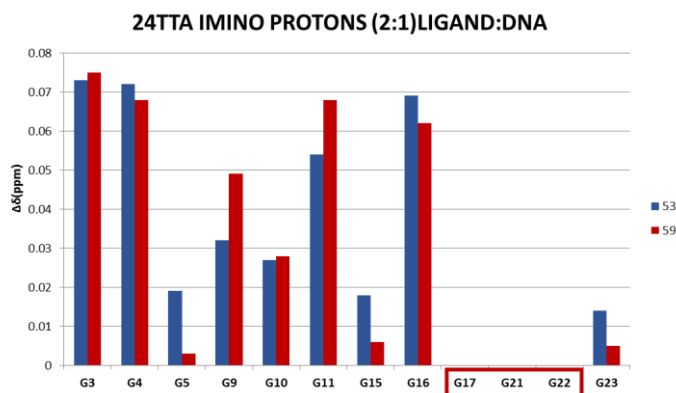
Figure 2.24. Graph with Δδ (ppm in absolute values) imino proton provided by 2:1 (ligand:24TTA) complexes. Disappeared imino protons are surrounded by a red square.

Most imino signals tend to broaden after compound addition. The most noticeable effect was presented over G17, G21 and G22 imino protons, which completely disappeared at 2 equivalents of both ligands. G3 and G4 showed a high shifted upfield. The most affected imino

protons (G17, G21 and G22) belong to the first and second tetrads, which made us suspect ligand binding could be opening the G4-structure. On the other side, comparing both ligands, the difference between the *m-* and *p-* derivatives was quite small in this sequence, simply **59** displayed larger $\Delta\delta$ on G9 and G11 than **53**. Therefore, the binding to the 24TTA topology seems to be quite similar for both ligands independently of the substitution differences. We highlighted the most affected imino protons on the 24TTA PDB solved structure (figure **2.25**).
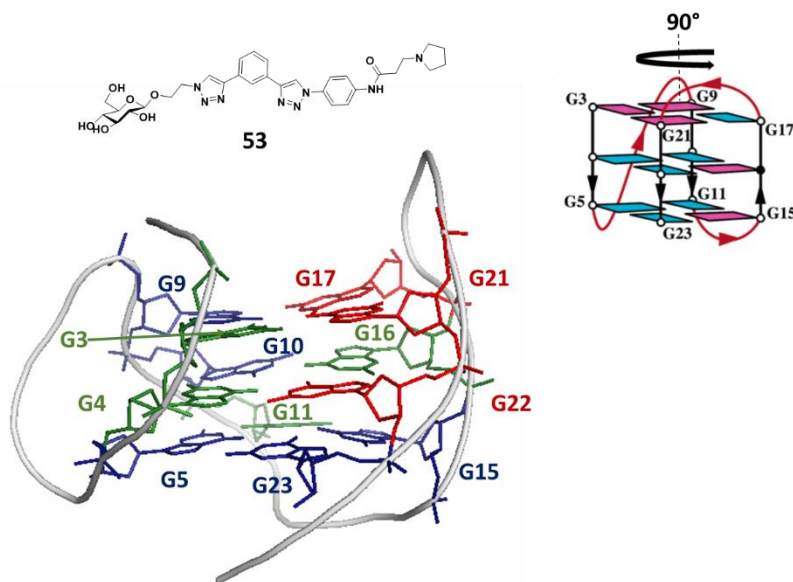
Figure 2.25. 24TTA-PDB:2GKU structure with the guanines affected by **53** ligand binding. Guanines are colored according to the changes observed in the imino signals upon ligand binding: The red represents the guanines with disappeared imino protons, green represents the guanines with the most shift imino protons and blue represents the guanines barely affected imino protons.

3.2.5.4.   NMR studies with the **22AG** sequence.

The 22AG sequence displays a complex ^{1}H-NMR spectrum when it is carried out with potassium as predominant cation, due to its polymorphism at such conditions. Nevertheless, at sodium buffer conditions it shows 12 well-resolved imino proton peaks corresponding to the 12 implicated guanines in the quadruplex core. NMR-characterization of 22AG at sodium rich solution has been previously reported[35] which confirmed the existence of a unique conformation.

According to the glycosidic torsion angle, the described alignments for each tetrad are *anti·syn·syn·anti*, *syn·anti·anti·syn*, and *anti·syn·syn·anti*, corresponding to G2·G10·G22·G14, G3·G9·G21·G15, and G4·G8·G20·G16 guanines distribution, respectively (figure **2.26**). G2-G3-G4 and G14-G15-G16 are aligned adjacent and parallel to each other and directed towards the lateral loops. In contrast, G8-G9-G10 and G20-G21-G22 are also aligned adjacent and parallel to each other but directed towards the diagonal loop. In addition, the hydrogen bonding directionalities around individual G-tetrads also alternate clockwise (G2·G10·G22·G14 and G4·G8·G20·G16) and counter-clockwise (G3·G9·G21·G15) orientations. The authors suggested sodium stabilizing cation sites between G-tetrad planes. The final structure observed is a **basket-type** with two medium width grooves and on opposite faces, one wide and one narrow groove on either side. It contains a central TTA diagonal loop and two TTA lateral loops pointing in opposite

direction respect to the diagonal one. The wide groove is formed by the antiparallel aligned G2-G3-G4 and G8-G9-G10 strands while the antiparallel aligned G14-G15-G16 and G20-G21-G22 strands constitute the narrow one. The medium width grooves are defined by the parallel aligned G2-G3-G4 and G14-G15-G16 strands; and G8-G9-G10 and G20-G21-G22 parallel strands.
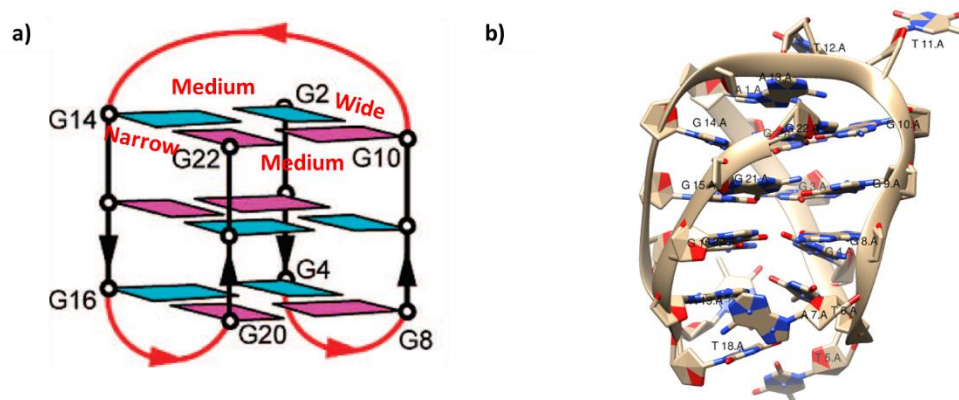
Figure 2.26. 22AG structure resolution at sodium rich buffer conditions. a) Schematic picture reported by the authors. b) PDB resolved structure: 143D.

We carried out titration experiments with different carbohydrate PTDZ ligands, including Neidle's reference control using ^{1}H-NMR spectroscopy. We followed the same procedure as above. We studied changes in the different proton signals (imino and aromatic ones) when the G4 ligand (2 equivalents) was added stepwise to the DNA sample. The compounds examined were **53** (glc *m*-PTDZ), **59** (malt *p*-PTDZ), **58** (aglycone *m*-PTDZ) and **1** (control compound).

Differences in chemical shift ($\Delta\delta$ in absolute values) were plotted for the 2:1 (ligand:22AG) complexes (figure **2.27**). In general, we observed most imino signals remained sharp, in contrast to their behavior during the titration experiment with the HIV321 sequence and 24TTA in which most peaks were highly broaden. The most evident change was on the chemical shift for some imino protons (G2, G3, G4, G16) and some aromatic protons (G3, G4, A7, G8, T18, A19). From one to two equivalents, G4 and G16 imino protons experienced a strong shift with most ligands except for compound **1**. Both guanines belong to the same lower G-tetrad, which suggest the ligands may interact with that region.

When we compare the *m*- and *p*-derivatives, the *p*-derivative (**59**) increased the $\Delta\delta$ of G2, G3, G9 and G20 imino protons, as well as the G14 aromatic proton with respect to the *m*-derivatives (**53**, **58** and control **1**). All these signals are located in the three G4-tetrads. In addition, it affected to the same G4 and G16 imino protons as the *m*-derivatives did, hence **59** may present more binding modes to this G4 structure than the corresponding *m*-derivatives. It is important to note that A7H8 aromatic proton experimented the highest $\Delta\delta$ for **53** and **58**, both containing a C2-linker in their structure. This results agree with the proton chemical shifts of G4 and G16 imino protons, placed on the contiguous tetrad from that loop, locating the ligands interaction on that region.
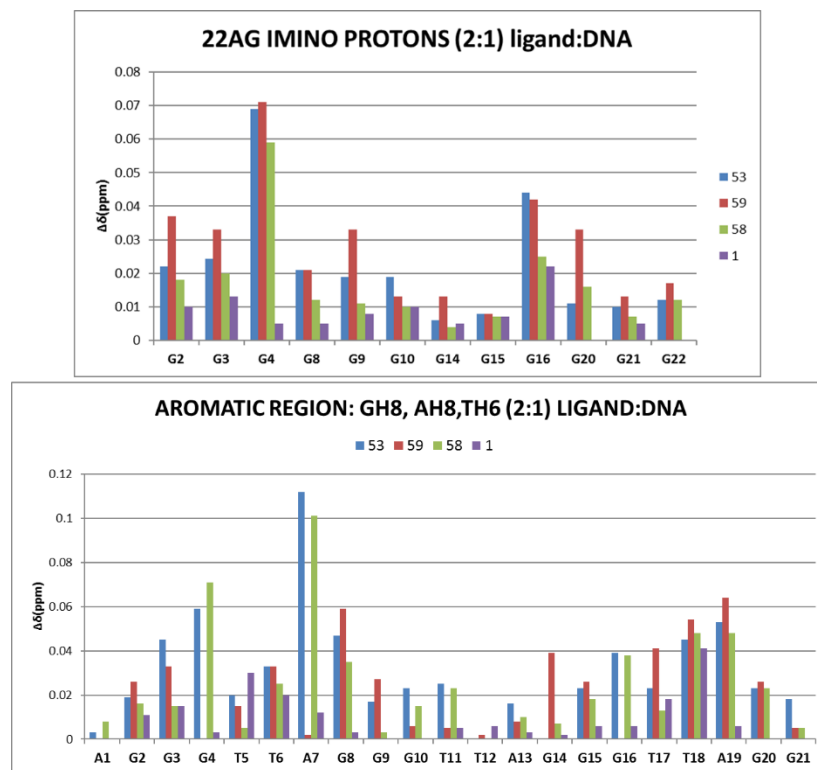
Figure 2.27. Bar graphs with Δδ (ppm) of imino and aromatic protons in 2:1 (ligand:DNA) complexes.

Neidle's ligand **1** barely affected the chemical shift of the analyzed protons, only the T5H6 was slightly shift. This hindered the location of compound **1** when binding the 22AG hTel sequence.

The PDB structures highlighting the most affected protons are shown in figure **2.28**.
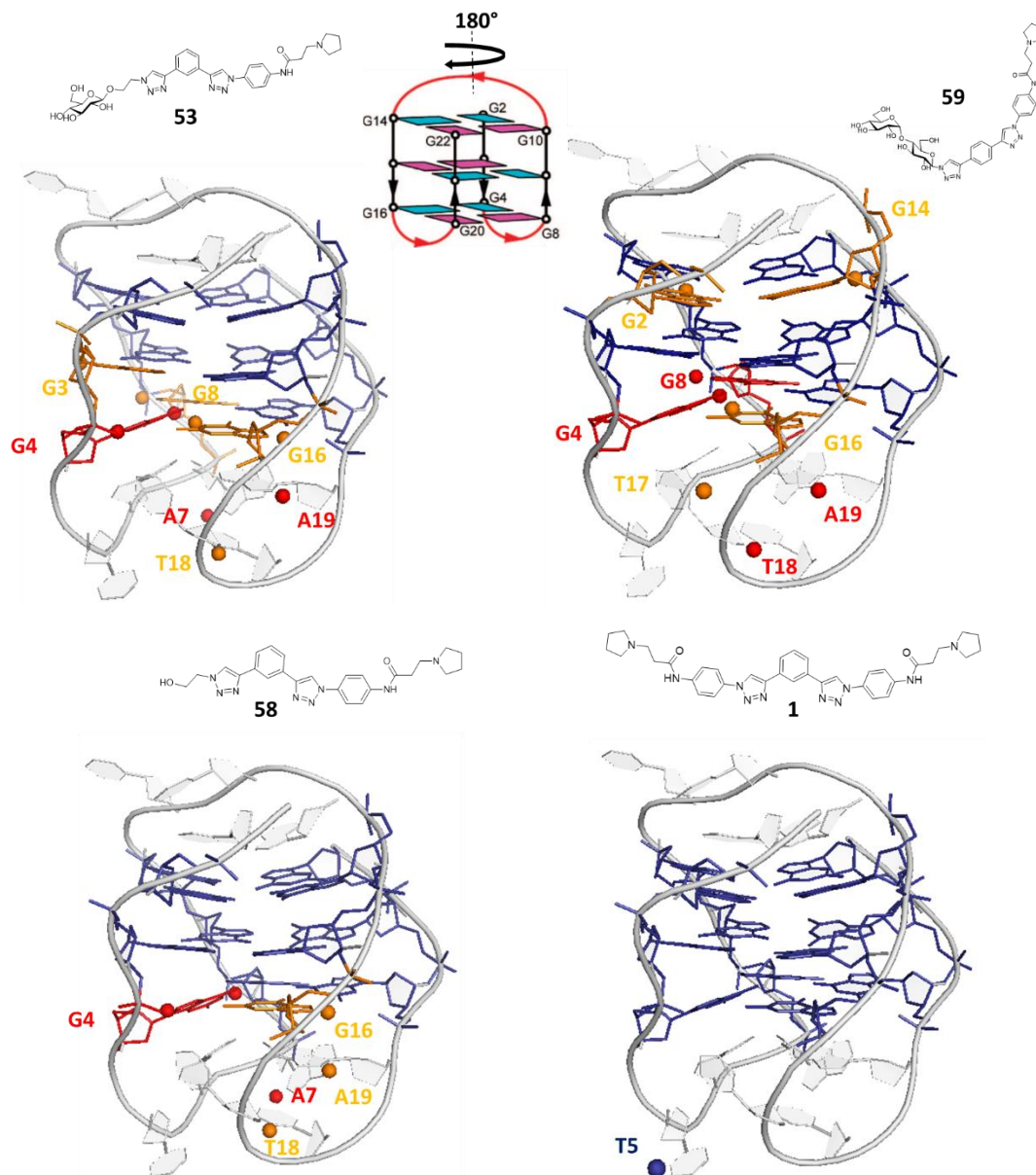
Figure 2.28.22AG-PDB:143D structures with the areas affected by the ligands binding. Guanines conforming the G-tetrads are represented by sticks. Imino and aromatic protons shifted the most by the binding are highlighted as spheres and colored according to the Δδ value with the following color code: red>orange. Guanines whose imino and/or aromatic protons are shifted the most are colored as well.

Detailed NMR analysis of **53**:22AG complex

We selected the complex **53**-22AG to examine in more detail the interaction using bidimensional NMR experiments. They may provide the "hot-spots" binding regions between both entities, which is important to understand the ligand binding properties as well as to find the way to improve them. To do so, we carried out two NOESY experiments: a) 1 mM folded DNA sample at sodium buffer conditions (20 mM sodium phosphate buffer supplemented with 70 mM NaCl), and b) a 1 mM 22AG sample + 2 mM of **53** ligand (glcC2-*m*-PTDZ), added just before running the experiment.

The complete sequence assignment was already reported. We used the assignments recently reported by Dr. Abdelaziz[260] using Sparky software. Most DNA signals were identified in the first experiment, but others were too broad to assign them. We overlapped both spectra to obtain the chemical shift differences ($\Delta\delta$) for each peak (see graph in figure **2.29**). Some examples of the NOESY imino-imino and imino-aromatic correlation are shown in figure **2.30**.
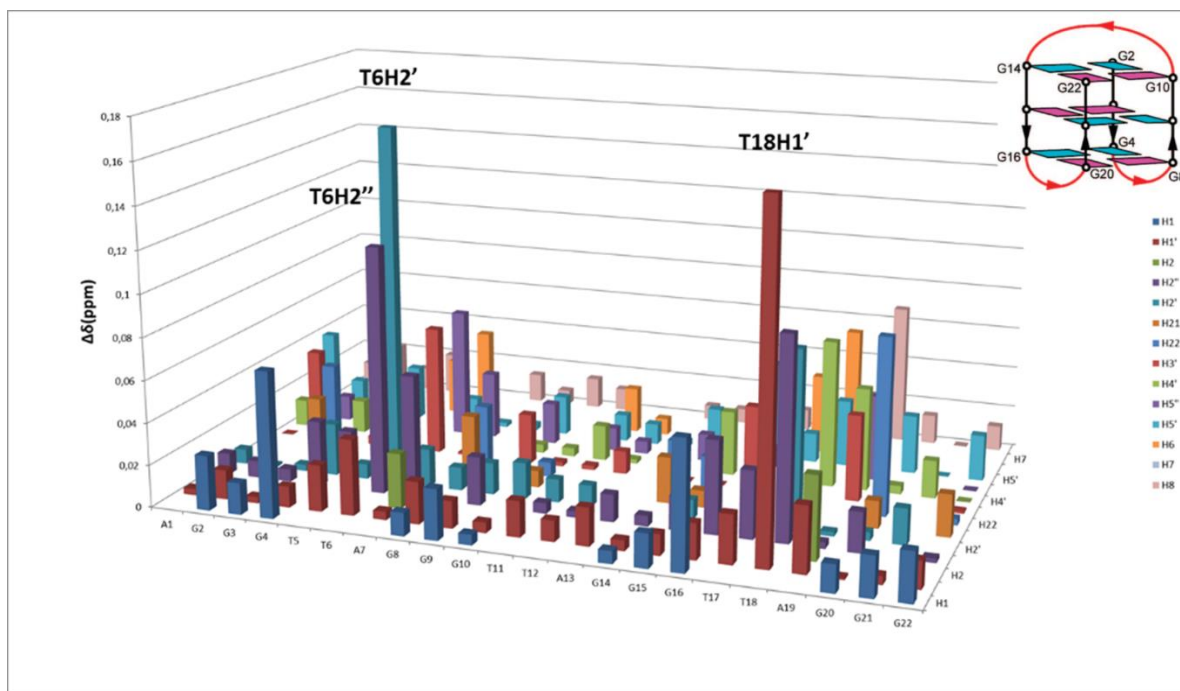
Figure 2.29. Bar graphs with the $\Delta\delta$ (ppm) found for each assigned proton of the 22AG human telomeric G4-sequence.
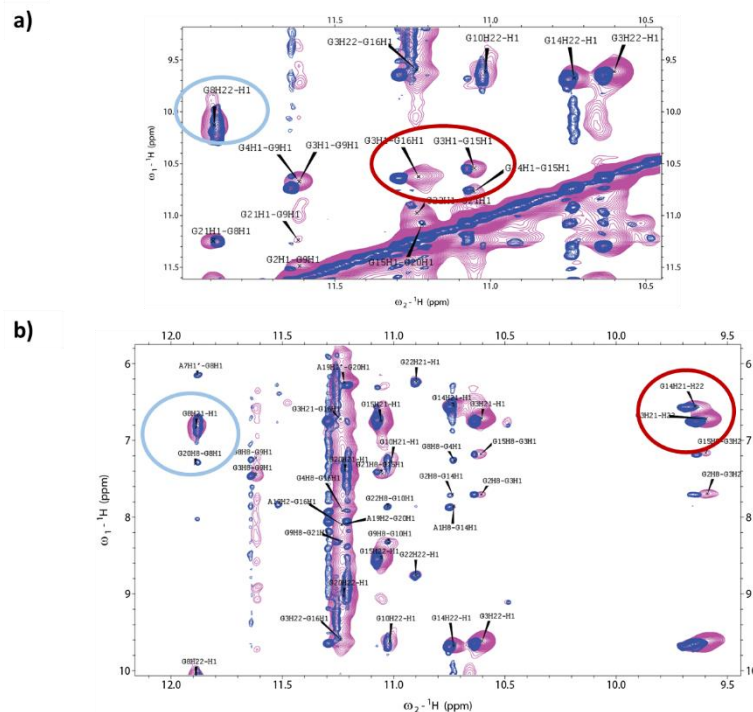
Figure 2.30. Overlapping spectra with 1 mM 22AG sequence sample (blue spectrum) and 1mM 22AG plus 2 mM **53** sample (pink spectrum). Examples of shifted peaks are surrounded in red and not affected signals are surrounded in light blue. a) Imino-imino correlation area. b) Imino-aromatic correlation range.

Two regions of the G4-22AG were mostly affected by ligand binding: from **G4** to **A7** and from **G16** to **G20**, both corresponding to the lateral lower loops. Specifically, **T6H2'** and **T18H1'** were the most shifted NOE signals, from the lateral loops as well. Most of them are sugar protons, located in the outer region of the 22AG structure. These results agreed with the obtained in the previous titration experiments, where most changed imino signals were the corresponding to G6 and G14, from the lower tetrad. On the contrary, the protons placed in the diagonal loop (G10-T11-T12-A13-G14) were barely affected, and therefore the ligand is not binding on this region of the quadruplex.

We classified the protons depending on their Δδ (ppm) value as: most affected (>0.05), less affected (0.05-0.036), and barely or not affected (<0.035). We applied a color code to visualize the most affected areas in the 22AG PDB structure 143D. The mapping locates most of the interaction in the lower loops and the first tetrad. Only the interaction placed in **A1** is out of the "hot spots" region (figure **2.31**).
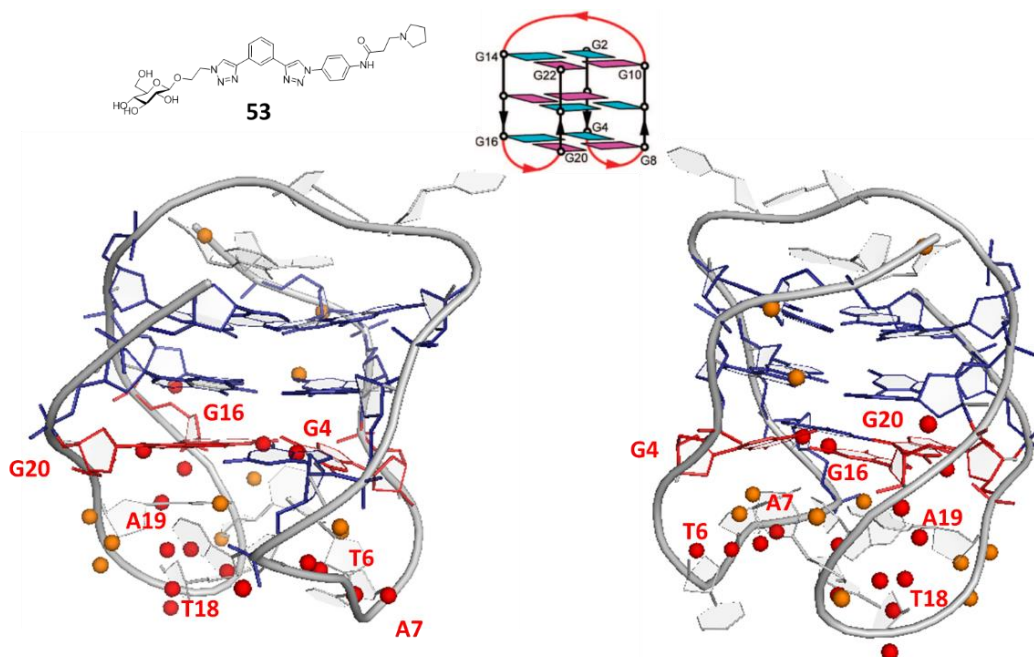
Figure 2.31.22AG-PDB:143D representations guanines in sticks. Affected proton are highlighted in colored spheres, according to the Δδ with the following color code: red>orange. Most affected guanines are colored in red.

The most accurate mode to locate the binding sites is by identifying intermolecular NOEs between the ligand and the 22AG-fold. However, we only were able to assign intramolecular NOEs, since the signals were too broad and vague for assigning intermolecular NOEs. For this reason, we could not find the exact interaction mode, only the main region of contact between the ligand and the G4-structure.

From the NMR experiment we can conclude:

- In general, the new asymmetric carb-PTDZ derivatives induce changes in chemical shift of the G4 protons and in some cases the disappearance of the imino signals is observed. All these data are indicative of ligand binding.
- With HIV321 sequence, we observed differences between the *m-* and *p-* asymmetric PTDZ derivatives. The first ones affected guanines from both tetrads, while the *p-* derivatives guanines from only one tetrad. We did not appreciate these clear differences in the other quadruplex sequences examined.
- We were not able to locate accurately the binding site between 22AG and **53** (glcC2-*m-*PTDZ) even using NOESY experiments. However, considering the affected signals, the interaction seems to occur around the edgewise loops and the adjacent quadruplex quartet.

3.3.   Cytotoxicity of the carbohydrate modified PTDZ ligands.

We investigated the antiproliferative activity and selectivity of the new carb-PTDZ ligands over the human cancer cell lines MCF-7 (breast cancer), HT-29 (colon cancer) and LN-229 (glioblastoma); and the non-tumoral cell lines HEK-293 (epithelial from human kidney embryo) and MRC-5 (fetal lung fibroblasts).

We selected the **MTT** (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) **assay** from the common methods to evaluate cell cytotoxicity because of its efficiency and reproducibility.[261] It is based on the MTT enzymatic metabolism in mitochondria. Mitochondrial enzymes reduce MTT by reaction with the NADH in order to form formazan purple crystals while cell is alive (reaction in figure **2.32**). The resultant purple color can be measured by UV-spectroscopy at 575 nm.
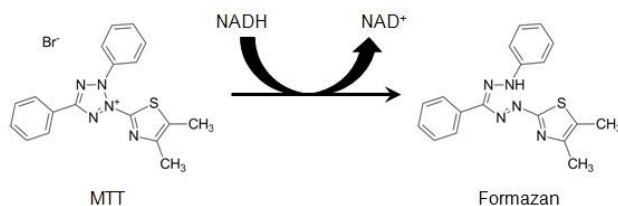
Figure 2.32. MTT reduction by mitochondrial enzymes to produce formazan purple crystals.

The purple crystal UV-absorbance is proportional to the mitochondrial activity, theoretically directly related to the number of living cells. Therefore, it provides the relative cell population compared to a control (not treated cells control). The assay is carried out in a 96 well-plate, hence it is possible to measure several compounds at different concentrations in the same plate. To perform the experiments, cells are first seeded in the well plate. The number of cells per will depends on the cell line (according to their growth gradient) and the experimental time. Adherent cells are kept overnight to be ready and after complete adhesion. Then, compounds are added at serial dilutions to the wells and the plate is incubated for the desired time (normally 24, 48 or 72 h). When the experiment is finished, MTT solution at 5 mg/mL concentration is added to the plate and it is incubated for four hours allowing the MTT to be metabolized by the alive cells. Formazan crystals are insoluble in the media and precipitate. Cell medium is removed (in adherent cells) and a solubilizing solution (DMSO, SDS or an acidic isopropanol solution) is added to dissolve the formazan crystals. Plates are kept at 37 °C at least for half an hour to facilitate the correct crystals solution. Finally, UV-absorbance measurement is read at 575 nm in a plate reader spectrophotometer. Signal intensity is proportional to the amount of alive cells, in percentage, respect to the non-treated control. The compound concentration at which 50% of cells are alive correspond to the IC_{50}, for the experimental time assayed (24, 48, 72 h…).

We performed MTT assays at 72 h with MCF-7, HT-29, LN-229, HEK-293 and MRC-5 cell lines to obtain the new compounds IC_{50} both in tumoral and non-tumoral cell lines. We included the control compound (**1**). The results are shown in table **2.6**.

	Compound	HT29	MCF-7	LN229	HEK-293T	MRC5
Symmetric PTDZ	**43 (β-glc-PTDZ)**		>100	>100	>100	
	44 (β-malt-PTDZ)		>100	>100	>100	
	45 (β-fuc-PTDZ)	>100	>100	>100	>100	
	46 (β-glcNAc-PTDZ)	>100	9.4	>100	>100	
	47 (α-manC2-PTDZ)	>100	>100		>100	>100
	48 (β-glcC2-PTDZ)	>100	>100		>100	>100
Asymmetric m-PTDZ	**51 (β-glc-*m*-PTDZ)**		>100	>100	>100	
	52 (β-malt-*m*-PTDZ)	>100	>100		>100	>100
	53 (β-glcC2-*m*-PTDZ)	>100	>100		>100	>100
	54 (β-fuc-*m*-PTDZ)		>100	>100	>100	
	55 (β-glcNAc-*m*-PTDZ)		>100	>100	63.8	
	56 (α-manC2-*m*-PTDZ)	>100	>100			
	57 (β-lac-*m*-PTDZ)	>100	>100			
	58 (aglycone-PTDZ)	>100	58.7			
	59 (β-malt-*p*-PTDZ)		>100			
	60 (β-glc-*p*-PTDZ)		>100			
Control	1	24.25±15	7.5	2.7±0.91	3.9±0.53	11.25±0.21

Table 2.8. IC_{50} (μM) values of the new compounds in HT-29, MCF-7, LN-229, HEK-293T and MRC-5 cell lines after 72 h of incubation. The MTT assays were performed by triplicate.

The results showed that all the ligands displayed high IC_{50} values for every cell line, both tumoral and non-tumoral ones. Neidle's compound **1** exhibited lower IC_{50} values for MCF-7 and LN-229 cell lines, 7.5 and 2.7 μM, respectively, same as the reported values.[229] Nonetheless, for non-tumoral cells it was cytotoxic at similar concentrations (3.9 μM for HEK-293T and 11.25 μM for MRC-5).

We knew our carbohydrate modified PTDZ ligands are only low to medium G4-binders. Nevertheless, the low cytotoxicity values obtained could be due to a variety of factors: cell uptake, product metabolism, access to the target, and the possibility that the compound does not act through the hypothetical mechanism of action. We tried to understand the low cytotoxicity of the new ligands and checked their **stability in aqueous medium**. We carried out NMR proton experiments in deuterium oxide at 37 °C and at different times (0, 3, 6, 12, 24 and 48 h) with some of the ligands. We chose a representative carbohydrate motif for each type: **44** (symmetric malt-PTDZ), **47** (symmetric manC2-PTDZ), **53** (asymmetric glcC2-*m*-PTDZ), **56** (asymmetric mannC2-*m*-PTDZ) and **1** (Neidle's compound). We included Neidle's compound also to discard any triazole instability. Even at 48 h, we did not appreciate any modification on the spectra pattern,

neither in the aromatic nor the carbohydrate signals. Hence, we concluded that the compounds were stable in water.

We also investigated their **stability in cell media** since the presence of enzymes could hypothetically metabolize the compounds and render them ineffective. To verify this possibility, we performed several experiments with the help of the PhD. student Efres Belmonte and Dr. Pablo Peñalver. We examined the compounds: **53**, **48**, **52** and **1**; at different conditions: in water, medium, and in the presence of MCF-7 and HT-29 cells (supernatant and lysates), with incubation times of 0, 3, 24 and 48 h and 100 µM ligand concentration. Samples were analyzed by UPLC-MS. The area of these peaks were integrated, and the percentage of remnant compound was calculated according to the corresponding calibration curves and the normalized values in water. Results are summarized in figure **2.33**.
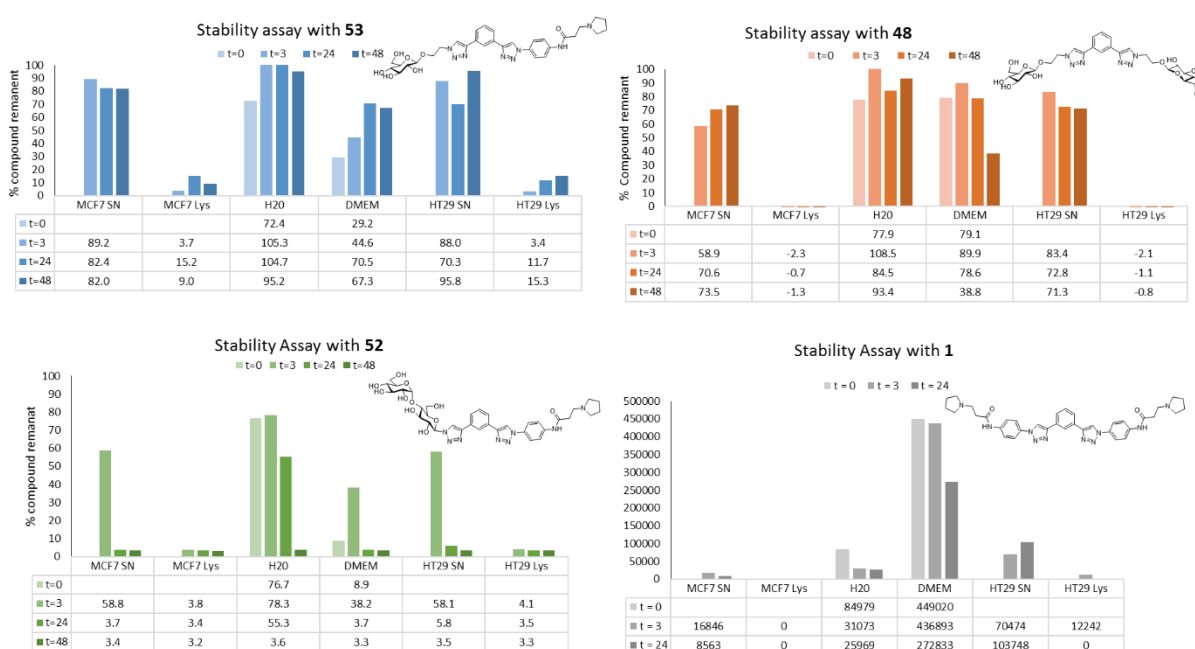
Figure 2.33. Bar graphs and tables with the stability results in PTDZ derivative percentage respect to their concentration in DMEM, except for **1**, which was calculated respect to its concentration in water.

All of them were stable in water, confirming the previous NMR experiments results. Only **52** showed lower proportion after 48 h in water. In DMEM samples, **53** and **48** (both glucose with C2-linker) remained stable, but with **52** we were not able to find its mass peak, then maltose is apparently less stable than glucose with C2-linker. In MCF-7 the supernatant (SN) samples, **53** and **48** were stable up to 72 h. Compound **52**, on the contrary, at 24 h was not found, which agreed with its low stability in the DMEM sample. The MCF-7 lysates (Lys) did not show traces of any of the compounds. Their concentration was high in the supernatants even after 48 h for **53** or **48** and absent in lysates, what could be due to a low cell uptake of the ligands. These results could explain the low IC_{50} values obtained for this family of ligands. Similar results were observed in the experiments on HT-29 cells.

The control **1** was insoluble in water at high concentration, thus it was not possible to make a calibration curve and refer the remnant compound in percentage. Its concentration was low both in MCF7 and HT29 supernatant samples, as well as in lysates.

To summarize, compounds were stable during MTT assays, except **52**, and they were not found inside the cells. Their low cytotoxicity may be due to their poor cell uptake, hindered by the sugar motif and only one positive charge. The cell uptake is facilitated in the case of **1**, which allowed to set out towards its biological target.

From biological experiments we can conclude:

- The new carbohydrate modified PTDZ ligands show low cytotoxicity on the assayed human cancerous cell lines and the non-cancerous cells.
- Carbohydrate modified PTDZ ligands seem to be stable in aqueous solution and cell media as checked by NMR and UPLC-MS, except for the maltose PTDZ derivative **52.**
- The low cytotoxicity of the carbohydrate modified PTDZ ligands may be due to their low cellular uptake.

4. CONCLUSIONS

- The click chemistry reaction has been successfully optimized to prepare symmetric and asymmetric carbohydrate modified PTDZ ligands. We could prepare asymmetric compounds by two consecutive click reactions.

- The biophysical assays showed the new ligands bind selectively to the quadruplexes *versus* duplex DNA, although with low affinity in comparison to other classical G4-ligands. If we compare to Neidle's control compound, the carbohydrate motif decreased the affinity but increased the selectivity. The human telomeric sequence seems to be their preferential G4-forming sequence target.

- The relative triazole position in the phenyl central core (*m-* or *p-*) may be important in the interaction with the HIV321 sequence, according to the NMR titration assays. Different binding modes could possibly be inferred.

- NMR titration and NOESY experiments seem to locate ligand **53** binding the edgewise loops and adjacent tetrad of quadruplex 22AG (hTel).

- The new carbohydrate-based ligands did not display anticancer activity in the MCF-7, HT-29 and LN-229 cancer cell lines and, it may be due to their poor cell uptake.

5. EXPERIMENTAL DETAILS.

5.1. Synthesis of the new carbohydrate-PTDZ conjugates ligands.

5.1.1. MATERIAL AND METHODS.

All reagents and solvents were purchased to Sigma Aldrich, Carbosynth, Fluka or Merck, and used without further purification. All reactions were monitored by TLC, on silica gel plates (60F254 (Merck), and visualized with UV light or with mostain (500 mL of 10% H_2SO_4, 25 g of $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 1 g $Ce(SO_4)_2$ $4H_2O$), sulphuric acid (10% H_2SO_4 in ethanol), anisaldehyde (450 mL ethanol, 25 mL anisaldehyde, 25 mL H_2SO_4 and 1 mL AcOH), or ninhydrin (0.25g ninhydrin, 100 mL ethanol) as staining agents. Products were purified by flash chromatography with silica gel 60 (0.5-0.2 mm, 0.2-0.063 mm, and 0.040-0.015 mm) or alumina (Al_2O_3 basis) type WN6.

NMR spectra were recorded on either a Bruker AVANCE 300 or ARX 400 MHz [300 or 400 MHz (^{1}H), 75 or 100 (^{13}C)] spectrometer, at room temperature for solutions in $CDCl_3$, CD_3OD, D_2O or DMSO-d_6. Chemical shifts were referred to the solvent signal and were expressed in ppm. 2D NMR experiments (COSY, TOCSY and HMQC) were carried out when necessary to assign the corresponding signals of the new compounds Mass spectra were carried out in an Esquire 6000 Bruker-Daltonics (ESI) and Micromass Autospec-Q (EMAR-FAB) and a QTRAP (EMAR-IQ) in the Chemical Research Institute of Seville, and CITIUS (Center of Research, Technology and Innovation and of the University of Seville), respectively.

Final asymmetric compounds were purified by HPLC-UV with a semi-preparative C18 Nucleosil 100 (250 x 4 mm, 5 µm) Phenomenex® column on a Waters Alliance 2790 HPLC, or on an Agilent 1100 instrument. Mobile phases: A (H_2O+0.1%TFA), B (ACN+0.1%TFA). 55 min methods at 3 mL/min flow. Method **I**: 0 min (85% A, 15% B), 30 min (20% A, 80% B), 35 min (100% B), 45 min (85% A, 15% B). Method **II**: 0 min (85% A, 15% B), 30 min (80% A, 20% B), 35 min (100% B), 45 min (85% A, 15% B). Method **III**: 0 min (85% A, 15% B), 30 min (75% A, 25% B), 35 min (100% B), 45 min (85% A, 15% B). Method **IV**: 0 min (80% A, 20% B), 30 min (75% A, 25% B), 35 min (100% B), 45 min (80% A, 20% B). Method **V**: 0 min (85% A, 15% B), 30 min (70% A, 30% B), 35 min (100% B), 45 min (85% A, 15% B).

5.1.2. SYNTHESIS OF CARBOHYDRATE-AZIDO DERIVATIVES.

General method for the synthesis of the azide carbohydrate derivatives

The azido derivatives were prepared in two consecutive reactions:

1) The corresponding peracetylated sugar derivative (1 eq) was dissolved in anhydrous dichloromethane. Hydrogen bromide solution 33 wt. % in acetic acid (2 mL/g starting material) was added under argon atmosphere.[233] The reaction was stirred at room temperature until disappearance of the starting material. Then, it was diluted with dichloromethane and washed with ice, saturated $NaHCO_3$ solution (3x50 mL) and distilled water (50 mL). The organic phase was dried over $MgSO_4$, filtered off and solvents were evaporated under reduced pressure. The crude was used in the following step without further characterization.

2) To a solution of the corresponding bromide sugar crude (1eq) in a mixture of acetone/ water (5:1), sodium azide (4 eq) was added. The reaction was stirred at room temperature until complete conversion of the starting material. Solvents were removed under reduced pressure and the mixture was dissolved in water. The organic phase was extracted with EtOAc (3x30 mL), dried over anhydrous MgSO$_4$, filtered off and solvents were evaporated under reduced pressure. The crude was purified by silica gel column chromatography to afford the desired product.

2,3,4,6-Tetra-*O*-acetyl-1-azido-1-deoxy-β-D-glucopyranose (**3**)

3
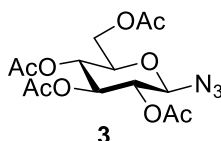
The general procedure was applied to 1,2,3,4,6-penta-*O*-acetyl-β-D-glucopyranose (7.00 g, 17.93 mmol). The azidation reaction was completed after 16 h. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (5:1 to 1:1) to give **3** (2.81 g, 42%) as a white solid, previously reported.[262] LRMS (ES$^+$) calculated for C$_{14}$H$_{19}$N$_3$NaO$_9$, (M+Na): 396.1; found: 396.0.

Synthesis of 2,3,4-tri-*O*-acetyl-1-azido-1-deoxy-β-L-fucopyranose (**4**):

4
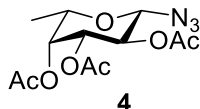
The general procedure was applied to 1,2,3,4-tetra-*O*-acetyl-α-L-fucopyranose (2.8 g, 8.471 mmol). The azidation reaction was completed after 4 h. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (3:1 to 2:1) to yield **4** (1.7 g, 80%) as a white solid, previously described.[263] ^{1}H NMR (400 MHz, CDCl$_3$) δ 5.23 (d, J=3.3 Hz, 1H, H$_4$), 5.16–5.04 (m, 1H, H$_2$), 5.00 (dd, J=10.3, 3.4 Hz, 1H, H$_3$), 4.56 (d, J=8.6 Hz, 1H, H$_1$), 3.88 (q, J=6.4 Hz, 1H, H$_5$), 2.15 (s, 3H,3xCOC$\underline{H}_3$), 2.05 (s, 3H, 3xCOC$\underline{H}_3$), 1.95 (s, 3H,3xCOC$\underline{H}_3$), 1.21 (d, J=6.4 Hz, 3H,3xCC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.58 ($\underline{C}$OCH$_3$), 170.10 ($\underline{C}$OCH$_3$), 169.50 ($\underline{C}$OCH$_3$), 88.25 (C$_1$), 71.58 (C$_5$), 71.20 (C$_3$), 69.99 (C$_4$), 68.26 (C$_2$), 20.75 (COC$\underline{H}_3$), 20.68 (COC$\underline{H}_3$), 20.61 (COC$\underline{H}_3$), 16.06 (C$_6$). LRMS (ES$^+$) calculated for C$_{12}$H$_{17}$N$_3$NaO$_7$, (M+Na): 338.1; found: 338.0.

Synthesis of 4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-2,3,6-tri-*O*-acetyl-1-azido-1-deoxy-β-D-glucopyranose (**5**):

5
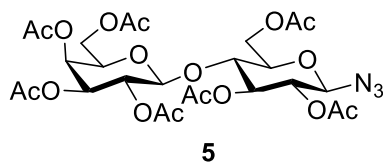
The general procedure was applied to 4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-1,2,3,6-tetra-*O*-acetyl D-glucopyranose (4.00 g, 5.89 mmol). The azidation reaction was completed after 3 h. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 1:1) to give **5** (2.56 g, 73%) as a white solid, as previously described.[264] ^{1}H NMR (500 MHz, CDCl$_3$) δ 5.34 (d, J=3.8 Hz, 1H, H$_{4B}$), 5.20 (t, J=9.2 Hz, 1H, H$_{3A}$), 5.10 (dd, J=10.4, 7.9 Hz, 1H, H$_{2B}$), 4.95 (dd, J=10.4, 3.5 Hz, 1H, H$_{3B}$), 4.85 (t, J=9.2 Hz, 1H, H$_{2A}$), 4.62 (d, J=8.8 Hz, 1H, H$_{1A}$), 4.53–4.45 (m, 2H, H$_{6A}$, H$_{1B}$), 4.16–4.04 (m, 3H, H$_{6'A}$, H$_{6B}$, H$_{6'B}$), 3.87 (t, J=6.8 Hz, 1H, H$_{5B}$), 3.81 (t, J=9.5 Hz, 1H, H$_{4A}$), 3.70 (ddd, J=10.0, 5.2, 2.0 Hz, 1H, H$_{5A}$), 2.14 (d, J=6.9 Hz, 6H, 6xCOC$\underline{H}_3$), 2.05 (dd, J = 10.0, 2.6 Hz, 15H, 15xCOC$\underline{H}_3$), 1.96 (s, 3H, 3xCOC$\underline{H}_3$). ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.45 ($\underline{C}$OCH$_3$), 170.41 ($\underline{C}$OCH$_3$), 170.22 ($\underline{C}$OCH$_3$), 170.15 ($\underline{C}$OCH$_3$), 169. 73 ($\underline{C}$OCH$_3$), 169.60 ($\underline{C}$OCH$_3$), 169.18 ($\underline{C}$OCH$_3$), 101.25 (C$_{1B}$), 87.84 (C$_{1A}$), 75.92 (C$_{4A}$), 74.95 (C$_{5A}$), 72.67 (C$_{3A}$), 71.14 (C$_{2A}$), 71.07 (C$_{3B}$), 70.91 (C$_{5B}$), 69.20 (C$_{2B}$), 66.72 (C$_{4B}$), 61.87 (C$_{6B}$), 60.92 (C$_{6A}$), 20.94 (COC$\underline{H}_3$), 20.93 (COC$\underline{H}_3$), 20.87 (COC$\underline{H}_3$), 20.76 (COC$\underline{H}_3$), 20.74 (COC$\underline{H}_3$), 20.71 (COC$\underline{H}_3$), 20.63 (COC$\underline{H}_3$).

Synthesis of 4-*O*-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)-2,3,6-tri-*O*-acetyl-1-azido-1-deoxy-β D-glucopyranose (**6**).

6
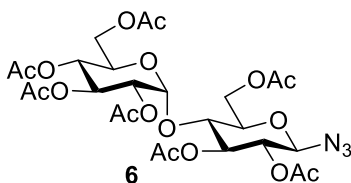
The general procedure was applied to 4-O-(2,3,4,6-Tetra-*O*-acetyl-α-D-glucopyranosyl)-1,2,3,6-tetra-*O*-acetyl D-glucopyranose (1 g, 1.47 mmol). The azidation reaction was completed after 2.5 h and the crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (3:1 to 1:1) to give **6** (577 mg, 60%) as a white solid, previously reported.[265] ^{1}H NMR (400 MHz, CDCl$_3$) δ 5.39 (d, J=4.0 Hz, 1H, H$_{1B}$), 5.34 (t, J=10.0 Hz, 1H, H$_{3B}$), 5.25 (t, J=9.0 Hz, 1H, H$_{3A}$), 5.04 (t, J=9.9 Hz, 1H, H$_{4B}$), 4.84 (dd, J=10.5, 4.0 Hz, 1H, H$_{2B}$), 4.77 (t, J=8.9 Hz, 1H, H$_{2A}$), 4.69 (d, J=8.7 Hz, 1H, H$_{1A}$), 4.49 (dd, J=12.2, 2.5 Hz, 1H, H$_{6A}$), 4.28–4.19 (m, 2H, H$_{6B}$, H$_{6B'}$), 4.14–3.88 (m, 3H, H$_{4A}$, H$_{5B}$, H$_{6A'}$), 3.77 (ddd, J=9.6, 4.3, 2.6 Hz, 1H, H$_{5A}$), 2.14 (s, 3H, 3xCOC$\underline{H}_3$), 2.09 (s, 3H, 3xCOC$\underline{H}_3$), 2.05–1.97 (m, 15H, 15xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.63 ($\underline{C}$OCH$_3$), 170.52 ($\underline{C}$OCH$_3$), 170.21 ($\underline{C}$OCH$_3$), 170.05 ($\underline{C}$OCH$_3$), 169.60 ($\underline{C}$OCH$_3$), 169.52 ($\underline{C}$OCH$_3$), 95.82 (C$_{1B}$), 87.58 (C$_{1A}$), 75.20 (C$_{3A}$), 74.36 (C$_{5A}$), 72.47 (C$_{4A}$), 71.60(C$_{2A}$), 70.10 (C$_{2B}$), 69.37 (C$_{3B}$), 68.75 (C$_{5B}$), 68.07 (C$_{4B}$), 62.65 (C$_{6B}$), 61.58 (C$_{6A}$), 20.96 (COC$\underline{H}_3$), 20.88 (COC$\underline{H}_3$), 20.79 (COC$\underline{H}_3$), 20.68 (COC$\underline{H}_3$). LRMS (ES$^+$) calculated for C$_{26}$H$_{35}$N$_3$NaO$_{17}$, (M+Na): 684.2; found: 684.2.

Synthesis of 2-(acetamido)-2-deoxy-3,4,6-tri-*O*-acetyl-α-D-glucopyranosyl chloride (**7**).

7
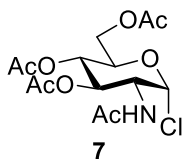
2-Acetamido-2-deoxy-D-glucopyranoside (2 g, 9.04 mmol, 1 eq) was dissolved in acetyl chloride (4 mL, 6.2 eq), and the mixture was refluxed until pink color appeared in the reaction mixture (~1h). Then, it was stirred at room temperature for 16 h. When the starting material had disappeared, the reaction mixture was diluted with dichloromethane and poured into ice. The organic phase was washed with a saturated $NaHCO_3$ solution (3x50 mL), dried over anhydrous $MgSO_4$ and filtered off. Solvent was evaporated under reduced pressure and the crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 0:1) to give **7** (1.4g, 43%) as a white solid, according to the reference.[236] ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.17 (d, J=3.7 Hz, 1H, H$_1$), 5.91 (d, J=8.7 Hz, 1H, CON$\underline{H}$), 5.36–5.26 (m, 1H, H$_3$), 5.24–5.15 (m, 1H, H$_4$), 4.52 (ddd, J=10.6, 8.7, 3.7 Hz, 1H, H$_2$), 4.34–4.21 (m, 2H, H$_5$, H$_6$), 4.17–4.04 (m, 1H, H$_{6'}$), 2.09 (s, 3H, 3xCOC$\underline{H}_3$), 2.03 (d, J=0.7 Hz, 6H, 6xCOC$\underline{H}_3$), 1.97 (s, 3H, 3xNHCOC$\underline{H}_3$). LRMS (ES$^+$) calculated for $C_{14}H_{20}ClNNaO_8$, (M+Na): 388.1; found: 388.0.

Synthesis of 2-acetamido-3,4,6-tri-*O*-acetyl-1-azido-1,2-dideoxy-β-D-glucopyranose (**8**)

8
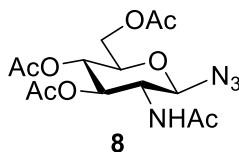
To a solution of the derivative **7** (1744 mg, 4.8 mmol, 1 eq) in anhydrous THF, trimethylsilyl azide (2 mL, 14.3 mmol, 3 eq) and TBAF (4.2 mL, 14.3 mmol, 3 eq) were added. The mixture was heated at 65°C and stirred at reflux during 16 h, until complete consumption of the starting material. Then, it was quenched with methanol and solvents were evaporated under reduced pressure. The concentrate was dissolved in EtOAc, washed with distilled water (3x50 mL), dried over anhydrous $MgSO_4$ and filtered off. Solvent was evaporated under reduced pressure and the crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:5 to 0:1) to yield **8** (1.2 g, 67%), according to the literature.[237] ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.00 (d, J=7.6 Hz, 1H, N$\underline{H}$COCH$_3$), 5.36–5.15 (m, 1H, H$_3$), 5.08 (t, J=9.7 Hz, 1H, H$_4$), 4.78 (d, J=9.3 Hz, 1H, H$_1$), 4.36–4.20 (m, 1H, H$_6$), 4.20–4.11 (m, 1H, H$_{6'}$), 3.91 (dd, J=19.5, 9.2 Hz, 1H, H$_2$), 3.80 (ddd, J=10.0, 4.7, 2.3 Hz, 1H, H$_5$), 2.05 (dt, J=17.6, 13.2 Hz, 12H, 12xCOC$\underline{H}_3$). LRMS (ES$^+$) calculated for $C_{14}H_{20}N_4NaO_8$, (M+Na): 395.1; found: 395.0.

Synthesis of azido derivatives with C2-linker.

Synthesis of 2-bromoethyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranose **9**:
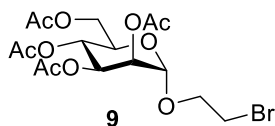
1,2,3,4,6-Penta-*O*-acetyl-α-D-mannopyranose (5 g, 13 mmol, 1 eq) was dissolved in anhydrous dichloromethane (20 mL). 2-bromoethanol (1.15 mL, 16.3 mmol, 1.25 eq) was added dropwise under argon atmosphere, followed by boron trifluoride diethyl etherate (8 mL, 65 mmol, 5 eq). The reaction was kept stirring at 0°C for 2h. It was allowed to warm to room temperature and stirred during 16h at room temperature. The mixture was poured into ice, the organic phase was extracted with dichloromethane (3x30 mL), then washed with a saturated NaHCO$_3$ solution (50 mL) and distilled water (50 mL). Organic phases were dried over MgSO$_4$, filtered off, and solvent was removed under reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (4:1 to 2:1) to give **9** (1.95 g, 33%) as a white solid, as previously described.[238] ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.38–5.22 (m, 3H, H$_3$, H$_4$, H$_2$), 4.87 (d, J=1.8 Hz, 1H, H$_1$), 4.33–4.21 (m, 1H, H$_6$), 4.18–4.09 (m, 2H, H$_{6'}$, H$_5$), 4.02–3.83 (m, 2H, 2xOC<u>H</u>$_2$-), 3.51 (td, J=5.9, 1.3 Hz, 2H, 2xC<u>H</u>$_2$Br), 2.15 (d, J=1.2 Hz, 3H, 3xCOC<u>H</u>$_3$), 2.10 (d, J=1.3 Hz, 3H, 3xCOC<u>H</u>$_3$), 2.04 (d, J=1.3 Hz, 3H, 3xCOC<u>H</u>$_3$), 1.99 (d, J=1.3 Hz, 3H, 3xCOC<u>H</u>$_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.72 (<u>C</u>OCH$_3$), 170.13 (<u>C</u>OCH$_3$), 169.97 (<u>C</u>OCH$_3$), 169.87 (<u>C</u>OCH$_3$), 97.88 (C$_1$), 69.55 (C$_2$), 69.15 (C$_3$), 69.07 (C$_5$), 68.61 (C$_4$), 66.13 (O<u>C</u>H$_2$), 62.54 (C$_6$), 29.74 (<u>C</u>H$_2$Br), 21.00 (CO<u>C</u>H$_3$), 20.88 (CO<u>C</u>H$_3$), 20.84 (CO<u>C</u>H$_3$), 20.80 (CO<u>C</u>H$_3$).

Synthesis of 2-azidoethyl 2,3,4,5-tetra-*O*-acetyl-α-D-mannopyranose, **10**:
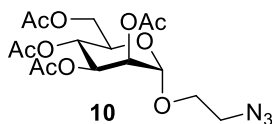
The bromide derivative **9** (1.9 g, 4.2 mmol, 1 eq) was dissolved in anhydrous DMF, and sodium azide (2.2 g, 33.8 mmol, 8 eq) was added to the solution. The reaction was heated at 60°C for 48 h, followed by dilution in EtOAc. The organic phase was washed with distilled water (3x50 mL), dried over MgSO$_4$, filtered off, and the solvent was removed under reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (3:1 to 1:1) to afford **10** (1.5g, 84%) as a white solid, as described previously.[266] ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.35–5.30 (m, 1H, H$_3$), 5.30–5.23 (m, 2H, H$_4$, H$_2$), 4.86 (d, J=1.8 Hz, 1H, H$_1$), 4.28 (dd, J=12.3, 5.3 Hz, 1H, H$_6$), 4.11 (dd, J=12.3, 2.4 Hz, 1H, H$_{6'}$), 4.03 (ddd, J=9.5, 5.3, 2.4 Hz, 1H, H$_5$), 3.86 (ddd, J=10.6, 6.6, 4.1 Hz, 1H, OC<u>H</u>$_2$-), 3.66 (ddd, J=10.5, 5.8, 3.7 Hz, 1H, OC<u>H</u>$_2$), 3.46 (td, J=6.2, 3.8 Hz, 2H, 2xC<u>H</u>$_2$N$_3$), 2.15 (s, 3H, 3xCOC<u>H</u>$_3$), 2.09 (s, 3H, 3xCOC<u>H</u>$_3$), 2.04 (s, 3H, 3xCOC<u>H</u>$_3$), 1.98 (s, 3H, 3xCOC<u>H</u>$_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.70 (<u>C</u>OCH$_3$), 170.10 (<u>C</u>OCH$_3$), 169.89 (<u>C</u>OCH$_3$), 169.85 (<u>C</u>OCH$_3$), 97.85 (C$_1$), 69.50 (C$_2$), 68.96 (C$_3$, C$_5$), 67.15 (O<u>C</u>H$_2$), 66.11 (C$_4$), 62.57 (C$_6$), 50.47 (<u>C</u>H$_2$N$_3$), 20.98 (CO<u>C</u>H$_3$), 20.85 (CO<u>C</u>H$_3$), 20.82 (CO<u>C</u>H$_3$), 20.77 (CO<u>C</u>H$_3$).

Synthesis of 2,3,4,6-tetra-*O*-acetyl-D-glucopyranose (**11**):
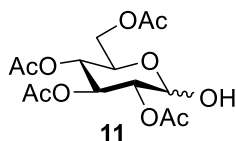
Hydrazine monohydrate (746 µL, 15.4 mmol, 1.2 eq) and acetic acid (880 µL, 15.4 mmol, 1.2 eq) were added to a solution of 1,2,3,4,6-penta-*O*-acetyl-D-glucopyranose (5 g, 12.8 mmol,1 eq) in DMF (15 mL), and the mixture was stirred at room temperature for 30 min. Once reaction was completed, it was diluted with EtOAc and washed with distilled water (2x50 mL) and brine (50 mL). The organic phase was dried over anhydrous MgSO$_4$, filtered off, and solvent was removed under reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 1:2) to give **11** (2.9 g, 65%) as a colorless oil, as it was described previously.[267]

Synthesis of 2,3,4,6-tetra-*O*-acetyl-α-glucopyranosyl trichloroacetimidate (**12**):
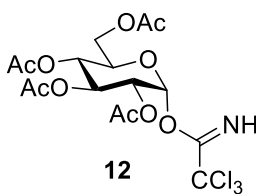
The previous derivative **11** (2.9 g, 8.3 mmol, 1 eq) was dissolved in anhydrous dichloromethane (20 mL). Under argon atmosphere, trichloroacetonitrile (8.4 mL, 83.3 mmol, 10 eq) and DBU in catalytic amounts were added to the solution. The reaction was stirred at room temperature for one hour. When the starting material was consumed, solvents were removed under reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc +5%TEA (2:1 to 3:2) to give **12** (2.96 g, 72%) as a white solid, as it was described previously.[240]

Synthesis of 2-bromoethyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranose (**13**):
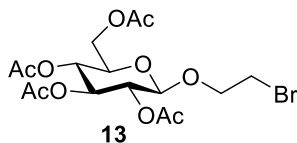
The trichloroacetimidate derivative **12** (2.963 g, 6.0 mmol, 1 eq) was dissolved in anhydrous dichloromethane (20 mL). 2-bromoethanol (470 µL, 6.6 mmol, 1.1 eq) was added dropwise under argon atmosphere, followed by boron trifluoride diethyl etherate (745 µL, 6.0 mmol, 1eq). The mixture was stirred for 20 min at room temperature and the reaction was quenched with a 10% aqueous Na$_2$CO$_3$ solution. The organic phase was washed with the same solution (2x50 mL) and distilled water (50 mL). The organic layer was dried over anhydrous MgSO$_4$, filtered off, and solvent was removed under reduced pressure. The crude was purified by silica gel column

chromatography using as eluent Hex/EtOAc (9:1 to 7:3) to afford **13** (1.47 g, 53%) as a white solid, as previously reported.[268] ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.17 (t, J=9.5 Hz, 1H, H$_3$), 5.03 (t, J=9.7 Hz, 1H, H$_4$), 4.98–4.89 (m, 1H, H$_2$), 4.53 (d, J=7.9 Hz, 1H,H$_1$), 4.21 (dd, J=12.3, 4.8 Hz, 1H, H$_5$), 4.16–4.03 (m, 2H, H$_6$, H$_6'$), 3.85–3.70 (m, 1H,OC$\underline{H}_2$-), 3.69–3.62 (m, 1H, OC$\underline{H}_2$-), 3.47–3.33 (m, 2H,2x-C$\underline{H}_2$Br), 2.00 (dd, J=18.7, 6.1 Hz, 12H, 12xC$\underline{H}_3$CO). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.63 ($\underline{C}$OCH$_3$), 170.24 ($\underline{C}$OCH$_3$), 169.63 ($\underline{C}$OCH$_3$), 169.42 ($\underline{C}$OCH$_3$), 101.04 (C$_1$), 72.67, 71.99, 71.10 (C$_2$,C$_3$,C$_4$,C$_5$), 69.82 (OC$\underline{H}_2$-), 68.41 (-$\underline{C}$H$_2$Br), 61.91 (C$_6$), 29.92 (CO$\underline{C}$H$_3$), 20.77 (CO$\underline{C}$H$_3$), 20.77 (CO$\underline{C}$H$_3$), 20.64 (CO$\underline{C}$H$_3$), 20.62 (CO$\underline{C}$H$_3$).

Synthesis of 2-azidoethyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranose (**14**):
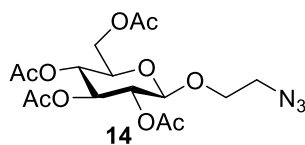
In a round bottomed flask, the bromide derivative **13** (1.47 g, 3.22 mmol, 1 eq) was dissolved in anhydrous DMF, and sodium azide (1.7 g, 25.8 mmol, 8 eq) was incorporated to the solution. The reaction was heated at 60°C and stirred for 48h. Then, it was diluted with EtOAc and the organic layer was washed with distilled water (3x50 mL), dried over anhydrous MgSO$_4$, filtered off, and the solvent was removed under reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (4:1 to 0:1) to give **14** (769 mg, 57%), as previously reported.[269] ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.32–5.14 (m, 1H, H$_4$), 5.13–4.94 (m, 2H, H$_3$, H$_2$), 4.57 (d, J=7.9 Hz, 1H, H$_1$), 4.23 (dd, J=12.3, 4.6 Hz, 1H, H$_6$), 4.13 (dd, J=12.3, 2.5 Hz, 1H, H$_6'$), 4.01 (ddd, J=10.6, 4.8, 3.3 Hz, 1H, H$_5$), 3.75–3.61 (m, 2H, 2x-C$\underline{H}_2$N$_3$), 3.47 (ddd, J=13.6, 8.3, 3.4 Hz, 1H, OC$\underline{H}_2$-), 3.26 (ddd, J=13.5, 4.8, 3.2 Hz, 1H, OC$\underline{H}_2$-), 2.17-1.89 (m, 12H, 12xCOC$\underline{H}_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.72 ($\underline{C}$OCH$_3$), 170.33 ($\underline{C}$OCH$_3$), 169.47 (2x$\underline{C}$OCH$_3$), 100.73 (C$_1$), 72.84, 72.03, 71.12, 68.65, 68.37, 61.89 (C$_2$,C$_3$,C$_4$,C$_5$,C$_6$,OC$\underline{H}_2$-), 50.58 ($\underline{C}$H$_2$N$_3$), 20.82 (CO$\underline{C}$H$_3$), 20.77 (CO$\underline{C}$H$_3$), 20.68 (2xCO$\underline{C}$H$_3$).

Synthesis of 2-azidoethanol (15)
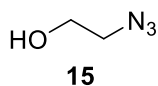
To a solution of sodium azide (2.6 g, 40.21 mmol, 2 eq) in milliQ water, 2-bromoethanol (1.5 mL, 20.10 mmol, 1 eq) and *n*-Bu$_4$NBr (130 mg, 0.402 mmol, 0.02 eq) were added. The mixture was stirred for 24 h at 80°C. The reaction was cooled to room temperature, and extracted with diethyl ether (3x50 mL). The organic phase was dried over MgSO$_4$, filtered off and the solvent evaporated under reduced pressure. A colorless liquid was obtained as product (1.02 g, 58%), and used without further purification, as described previously.[242,266] ^{1}H NMR (300 MHz, CDCl$_3$) δ 3.67 (q, J=5.2, 4.5 Hz, 2H, 2xC$\underline{H}_2$OH), 3.39–3.28 (m, 2H, 2xC$\underline{H}_2$N$_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 60.99 ($\underline{C}$H$_2$OH), 53.18 ($\underline{C}$H$_2$N$_3$).

Synthesis of 1-azido-1-deoxy-6-methylimidazole-β-D-glucopyranose

Synthesis of 1-azido-1-deoxy-6-methanesulfonate-β-D-glucopyranose (**21**)

21
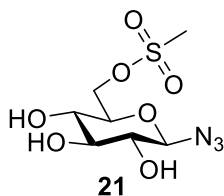
To a solution of 1-azido-1-deoxy- β-D-glucopyranose (788 mg, 3.838 mmol, 1 eq) in anhydrous pyridine (10 mL) methanesulfonyl chloride (297 µL, 3.838 mmol, 1 eq) was added dropwise at -10 °C over 30 min. The mixture was kept in agitation at such temperature during 2 h. Then, reaction was filtered off and solvent was removed. The filtrate was concentrated in vacuum before hexane (10 mL) was added, and sonicated for 5 min, the solvent was then decanted off and followed by addition of Et$_2$O (10 mL) before being sonicated and decanted again.[270] The resulting residue was purified by column chromatography (10:0 to 9:1 CH$_2$Cl$_2$ / MeOH) to yield the product as a colourless solid (696 mg, 64%). ^{1}H NMR (400 MHz, CD$_3$OD) δ 4.84 (s, 1H, H$_1$), 4.80 (dd, J = 11.5, 2.0 Hz, 1H, H$_6$), 4.67 (dd, J = 11.5, 5.3 Hz, 1H, H$_{6'}$), 3.97 – 3.88 (m, 1H, H$_5$), 3.66 (dq, J = 18.6, 9.1 Hz, 2H, H$_3$, H$_4$), 3.45 (t, J = 8.8 Hz, 1H, H$_2$), 3.39 (s, 3H, 3xSC$\underline{H}_3$). ^{13}C NMR (101 MHz, CD$_3$OD) δ 91.31 (C1), 77.32, 76.66, 74.12 (C$_3$, C$_4$, C$_5$), 70.15(C$_2$), 70.12 (C$_6$), 37.38 (S$\underline{C}$H$_3$). HRMS (FAB$^+$): calculated for C$_7$H$_{13}$N$_3$NaO$_7$S (M+Na): 306.0372; found: 306.0366.

Synthesis of 1-azido-6-(3-methyl-1H-imidazol-3-ium)-1,6-dideoxy-2,3,4-tri-*O*-acetyl-β-D-glucopyranoside tetrafluoroborate (**20**)

20
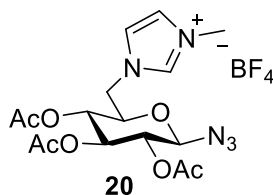
To a solution of sugar **21** (821 mg, 2.898 mmol, 1 eq) in acetonitrile (4 mL) and water (2 mL) 1-methylimidazole (462 µL, 5.796 mmol, 2 eq) and KBF$_4$ (1.46 g mg, 9.784 mmol, 4 eq) were added. The reaction mixture was stirred and heated to reflux for 48 h. After cooling at room temperature, the reaction mixture was filtered and concentrated under reduced pressure, dried under high vacuum and washed with hexane (5 mL), then diethyl ether (3 x 5 mL) with sonication.[247] After decantation, the oil was dried under vacuum, and purified by Sephadex G10 MeOH/H$_2$O (1:9). The product obtained was dissolved in pyridine (20 mL) and acetic anhydride (6 mL, 54.8 mmol, 12 eq) and DMAP were added. The mixture was stirred at room temperature for 16 h. Once the starting material was consumed, the reaction was dissolved with dichloromethane, washed with aqueous HCl 5% solution (3x40 mL) and with distilled water (50 mL). The organic phase was dried over anhydrous MgSO$_4$, filtered off and solvents were evaporated under reduced pressure. The crude was purified by silica gel column chromatography using as eluent CH$_2$Cl$_2$/MeOH (10:0 to 7:3). The fraction obtained was purified again by Sephadex

G25 using as eluent H_2O/MeOH to get the product **20** as a yellow oil (112 mg, 10%). ^1H NMR (300 MHz, D_2O) δ 8.83 (s, 1H, NC$\underline{H}$N), 7.53 (dt, J = 8.1, 2.1 Hz, 2H, NC$\underline{H}$C$\underline{H}$N), 5.43 (t, J = 9.3 Hz, 1H, H_3), 5.10 (d, J = 8.9 Hz, 1H, H_1), 5.00 (q, J = 9.3 Hz, 2H, H_2,H_4), 4.62 (dd, J = 14.9, 2.8 Hz, 1H, H_6), 4.47 (dd, J = 14.9, 6.9 Hz, 1H, $H_{6'}$), 4.29 (ddd, J = 9.8, 6.8, 2.8 Hz, 1H,H_5), 3.95 (s, 3H, 3xNC$\underline{H}_3$), 2.25–2.03 (m, 9H, 9xCOC$\underline{H}_3$). ^{13}C NMR (75 MHz, D_2O) δ 173.24 ($\underline{C}$OCH$_3$), 172.99($\underline{C}$OCH$_3$), 172.78($\underline{C}$OCH$_3$), 137.52(NCHN), 123.89(N$\underline{C}$HCHN), 123.71(NCH$\underline{C}$HN), 87.55 (C_1), 73.16, 72.97, 70.96, 69.19(C_2,C_3,C_4,C_5), 49.30(C_6), 36.25(NCH$_3$), 20.55(CO$\underline{C}$H$_3$), 20.41(CO$\underline{C}$H$_3$), 20.39(CO$\underline{C}$H$_3$). HRMS (FAB$^+$): calculated for $C_{16}H_{22}N_5O_7^+$(M$^+$): 396.1514; found: 396.1509.

5.1.3. SYNTHESIS OF SYMMETRIC PHENYLDITRIAZOLE LIGANDS.

General procedure for the synthesis of symmetric phenylditriazole ligands by "click chemistry"

Inside a 2.5 mL microwave reactor, the corresponding azide peracetylated sugar derivative (3 eq) was dissolved in 1.5 mL of THF:H_2O (1:1), and the copper salt CuSO$_4$·5H$_2$O (0.05 eq), sodium ascorbate (0.5 eq) and 1,3-diethynylbenzene (1 eq) were added to the solution. The mixture was stirred and heated in microwave reactor at 130°C for 30 min. Then, Quadrasil® MP was added to chelate the copper salt by stirring the suspension for 5 min at room temperature. The mixture was filtered off and concentrated. The crude was purified by silica gel column chromatography.

Synthesis of the peracetylated glucose symmetric PTDZ derivative **26**:

26
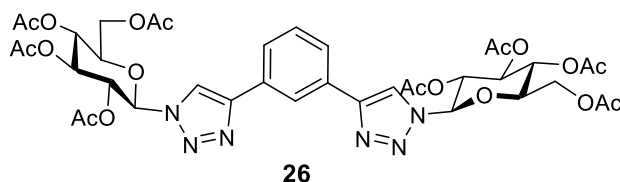
The general procedure was applied to the 2,3,4,6-tetra-*O*-acetyl-1-azido-1-deoxy-β-D-glucopyranose **3** (137 mg, 0.37 mmol, 3 eq). The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 0:1) to give **26** (108 mg, 100 %) as a white solid. ^1H NMR (400 MHz, CDCl$_3$) δ 8.24 (s, 1H, H$_{AR}$), 8.17–8.11 (m, 2xH$_{TRIAZOLE}$), 7.80 (dd, J=8.2, 3.2 Hz, 2H, 2xH$_{AR}$), 7.51–7.38 (m, 1H, H$_{AR}$), 6.00–5.93 (m, 2H, 2xH$_1$), 5.56–5.47 (m, 2H, 2xH$_2$), 5.44 (ddd, J=11.2, 6.6, 2.9 Hz, 2H, 2xH$_3$), 5.31–5.23 (m, 2H, 2xH$_4$), 4.32 (dt, J=13.1, 4.0 Hz, 2H,2xH$_6$), 4.15 (d, J=12.7 Hz, 2H,2xH$_{6'}$), 4.06 (dt, J=8.3, 3.3 Hz, 2H, 2xH$_5$), 2.07–1.99 (m, 18H, 18xCOC$\underline{H}_3$), 1.87–1.84 (m, 6H, 6xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.61 ($\underline{C}$OCH$_3$), 169.99 ($\underline{C}$OCH$_3$), 169.44 ($\underline{C}$OCH$_3$), 169.04 ($\underline{C}$OCH$_3$), 147.99 (2xC$_{TRIAZOLE}$), 130.57 (2xC$_{AR}$), 129.54 (CH$_{AR}$), 125.94 (2xCH$_{AR}$), 123.23 (CH$_{AR}$), 118.36 (2xCH$_{TRIAZOLE}$), 85.84 (2xC$_1$), 75.13 (2xC$_5$), 72.76 (2xC$_3$), 70.38 (2xC$_2$), 67.77 (2xC$_4$), 61.66 (2xC$_6$), 20.74 (CO$\underline{C}$H$_3$), 20.59 (CO$\underline{C}$H$_3$), 20.57 (CO$\underline{C}$H$_3$), 20.23 (CO$\underline{C}$H$_3$). HRMS (FAB$^+$): calculated for $C_{38}H_{44}N_6NaO_{18}$, (M+Na): 895.2610; found: 895.2554.

¹H-NMR

¹³C-NMR
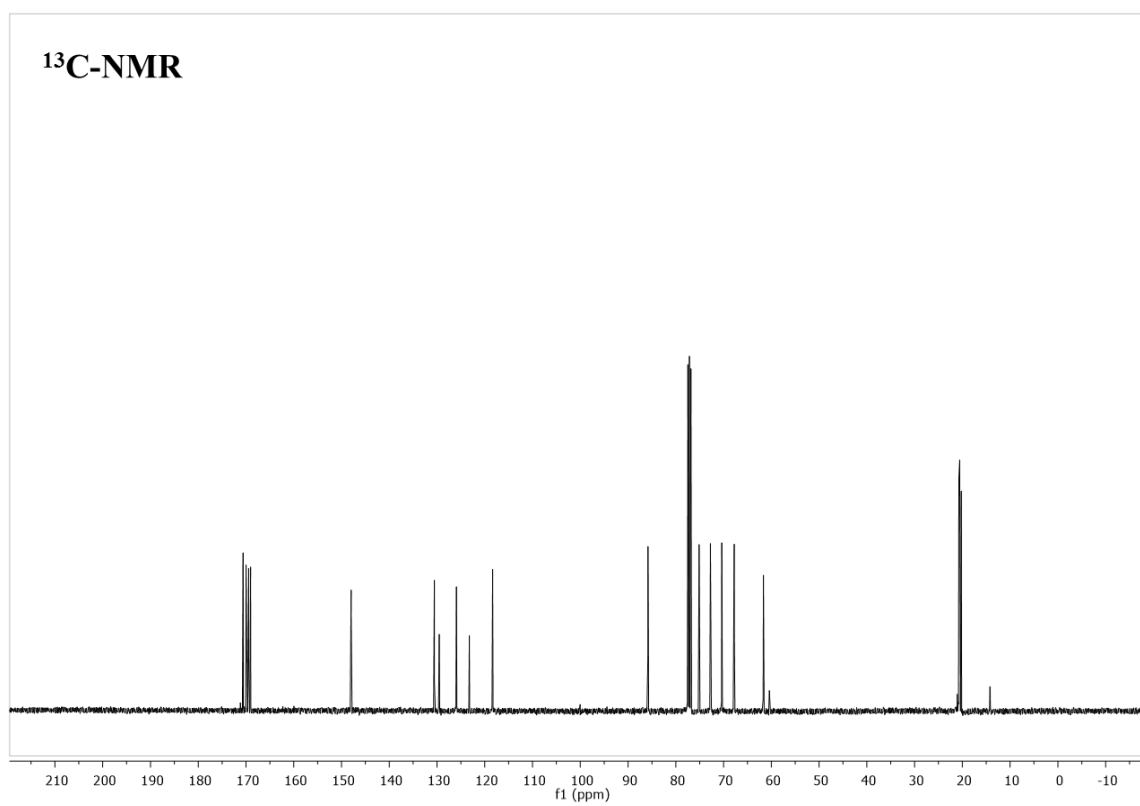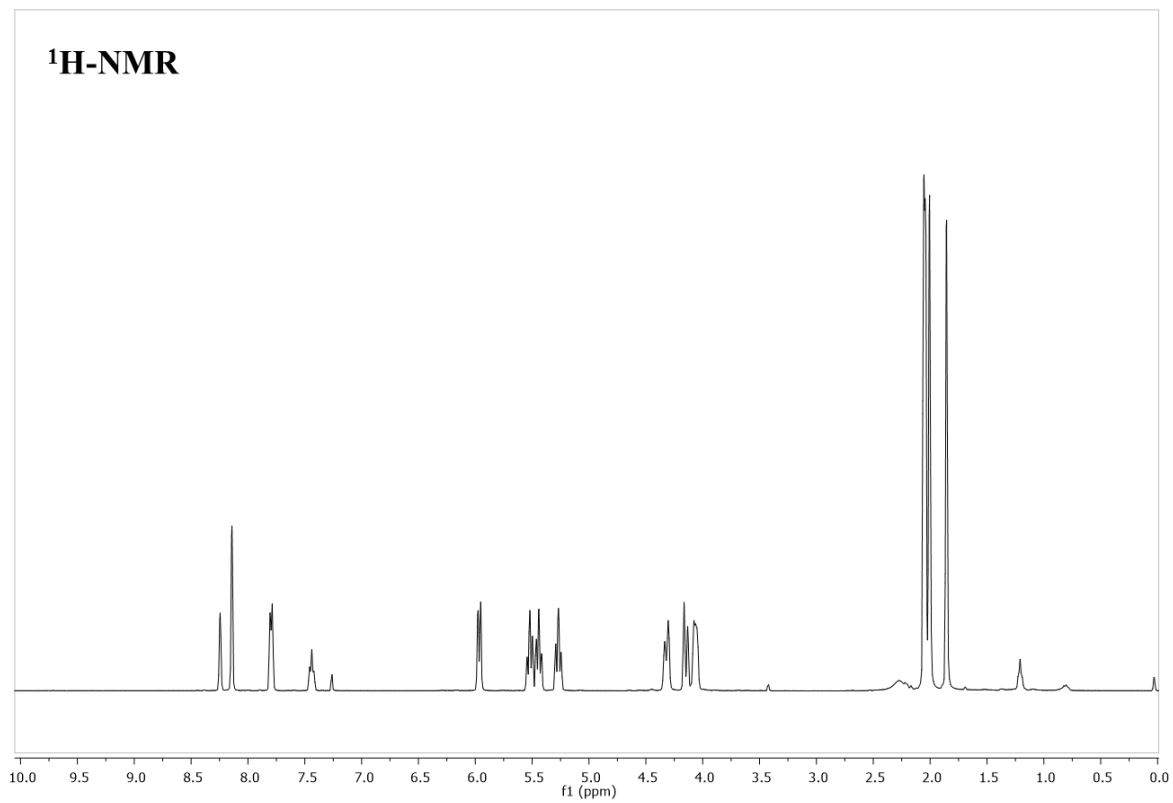
Synthesis of the peracetylated *N*-acetyl-glucosamine symmetric PTDZ derivative **27**:
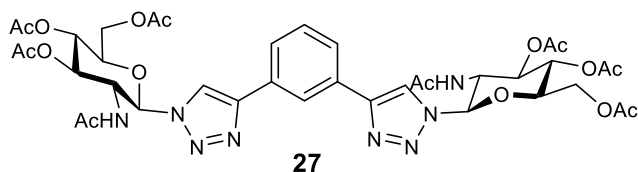
The general procedure was applied to the derivative **8** (447 mg, 1.2 mmol, 3 eq). The crude was purified by silica gel column chromatography using as eluent EtOAc/MeOH (1:0 to 5:1) to give **27** (90 mg, 26%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 8.97 (s, 1H, N$\underline{H}$CO), 8.39 (s, 1H, H$_{AR}$), 8.14 (d, J=9.1 Hz, 2H, 2xH$_{TRIAZOLE}$), 7.82 (d, J=7.6 Hz, 2H, 2xH$_{AR}$), 7.58 (t, J=7.7 Hz, 1H, H$_{AR}$), 6.19 (d, J=9.8 Hz, 2H, 2xH$_1$), 5.42 (t, J=9.8 Hz, 2H, 2xH$_3$), 5.13 (t, J=9.7 Hz, 2H, 2xH$_4$), 4.68 (q, J=9.8 Hz, 2H, 2xH$_2$), 4.34–4.28 (m, 2H, 2xH$_5$), 4.20 (dd, J=12.6, 5.1 Hz, 2H, 2xH$_6$), 4.11 (d, J=11.5 Hz, 2H, 2xH$_{6'}$), 2.04–1.96 (m, 18H,18xCOC$\underline{H}_3$), 1.61 (s, 6H,6xNHCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 170.06 (2x$\underline{C}$OCH$_3$), 169.60 (2x$\underline{C}$OCH$_3$), 169.52 (2x$\underline{C}$OCH$_3$), 169.39 (2x$\underline{C}$OCH$_3$), 146.22 (C$_{TRIAZOLE}$), 130.92 (C$_{AR}$), 129.79 ($\underline{C}$H$_{AR}$), 124.90 ($\underline{C}$H$_{AR}$), 121.87 ($\underline{C}$H$_{AR}$), 120.59 ($\underline{C}$H$_{AR}$), 84.99 (2xC$_1$), 73.43 (2xC$_5$), 72.29 (2xC$_3$), 68.09 (2xC$_4$), 61.83 (2xC$_6$), 52.43 (2xC$_2$), 22.33 (2xCO$\underline{C}$H$_3$), 20.55 (2xCO$\underline{C}$H$_3$), 20.44 (2xCO$\underline{C}$H$_3$), 20.30 (2xCO$\underline{C}$H$_3$). LRMS (ES$^+$) calculated for C$_{38}$H$_{46}$N$_8$NaO$_{16}$, (M+Na): 893.3; found: 893.6. HRMS (FAB$^+$): calculated for C$_{38}$H$_{46}$N$_8$NaO$_{16}$, (M+Na): 893.2929; found: 893.2930.

¹³C-NMR
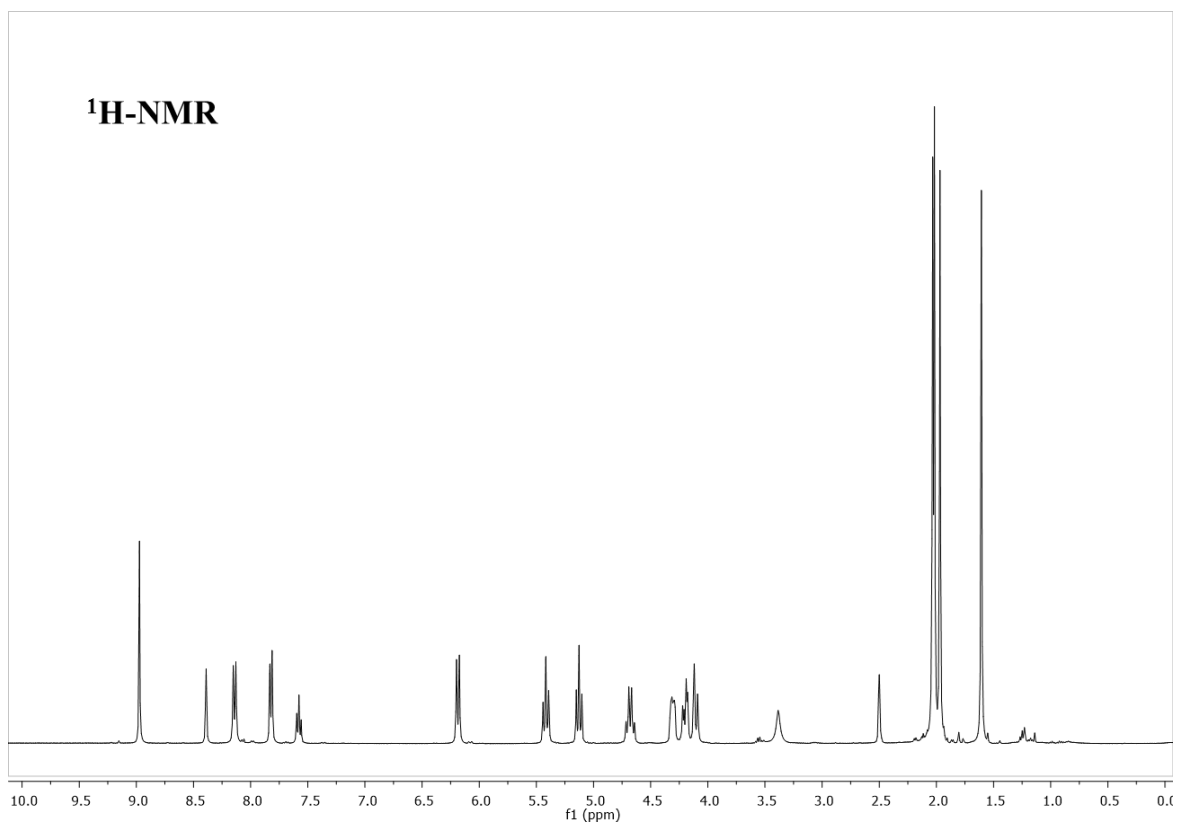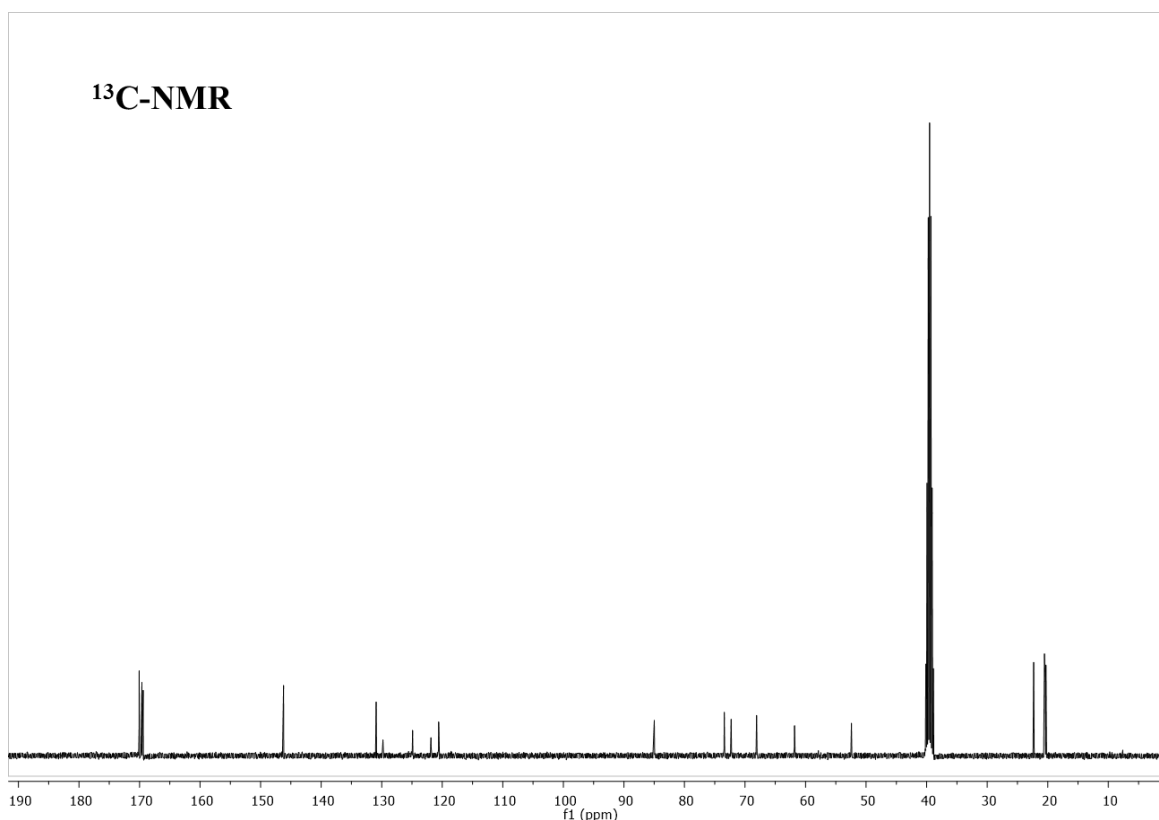
Synthesis of the peracetylated fucose symmetric PTDZ derivative **28**:

28
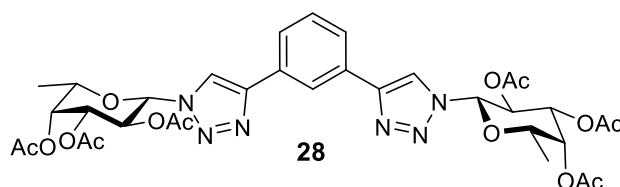
The general procedure was applied 2,3,4-tri-O-acetyl-1-azido-1-deoxy-β-L-fucopyranose **4** (378 mg, 1.2 mmol, 3 eq). The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 2:3) to afford **28** (297 mg, 98%) as a white solid. ¹H NMR (400 MHz, CDCl₃) δ 8.33 (d, J=1.4 Hz, 1H, H$_{AR}$), 8.17 (d, J=1.0 Hz, 2H,2xH$_{TRIAZOLE}$), 7.88–7.83 (m, 2H, 2xH$_{AR}$), 7.49 (t, J=7.8 Hz, 1H,H$_{AR}$), 5.87 (dd, J=9.3, 1.0 Hz, 2H, 2xH₁), 5.58 (t, J=9.7 Hz, 2H, 2xH₂), 5.41 (d, J=3.3 Hz, 2H, 2xH₄), 5.30–5.24 (m, 2H, 2xH₃), 4.15 (q, J=6.6 Hz, 2H, 2xH₅), 2.28 (d, J=0.9 Hz, 6H, 6xCOC$\underline{H}$₃), 2.03–1.99 (m, 6H, 6xCOC$\underline{H}$₃), 1.89 (d, J= 1.0 Hz, 6H, 6xCOC$\underline{H}$₃), 1.29 (d, J=6.3 Hz, 6H, 6xCC$\underline{H}$₃). ¹³C NMR (101 MHz, CDCl₃) δ 170.53 (2x$\underline{C}$OCH₃), 170.01 2x($\underline{C}$OCH₃), 169.32 (2x$\underline{C}$OCH₃), 147.94 (C$_{TRIAZOLE}$), 130.79 (C$_{AR3,5}$), 129.53 ($\underline{C}$H$_{AR}$), 125.88 (2x$\underline{C}$H$_{AR}$), 123.22 (CH$_{AR}$), 118.32 (CH$_{TRIAZOLE}$), 86.56 (2xC₁), 72.89 (2xC₅), 71.33 (2xC₃), 70.01 (2xC₄), 68.14 (2xC₂), 20.84 (2xCO$\underline{C}$H₃), 20.67 (2xCO$\underline{C}$H₃), 20.41(2xCO$\underline{C}$H₃), 16.21 (2xC$\underline{C}$H₃). LRMS (ES⁺) calculated for C₃₄H₄₀N₆NaO₁₄, (M+Na): 779.3; found: 779.5. HRMS (FAB⁺): calculated for C₃₄H₄₀N₆NaO₁₄, (M+Na): 779.2500; found: 779.2495.

¹H-NMR

¹³C-NMR
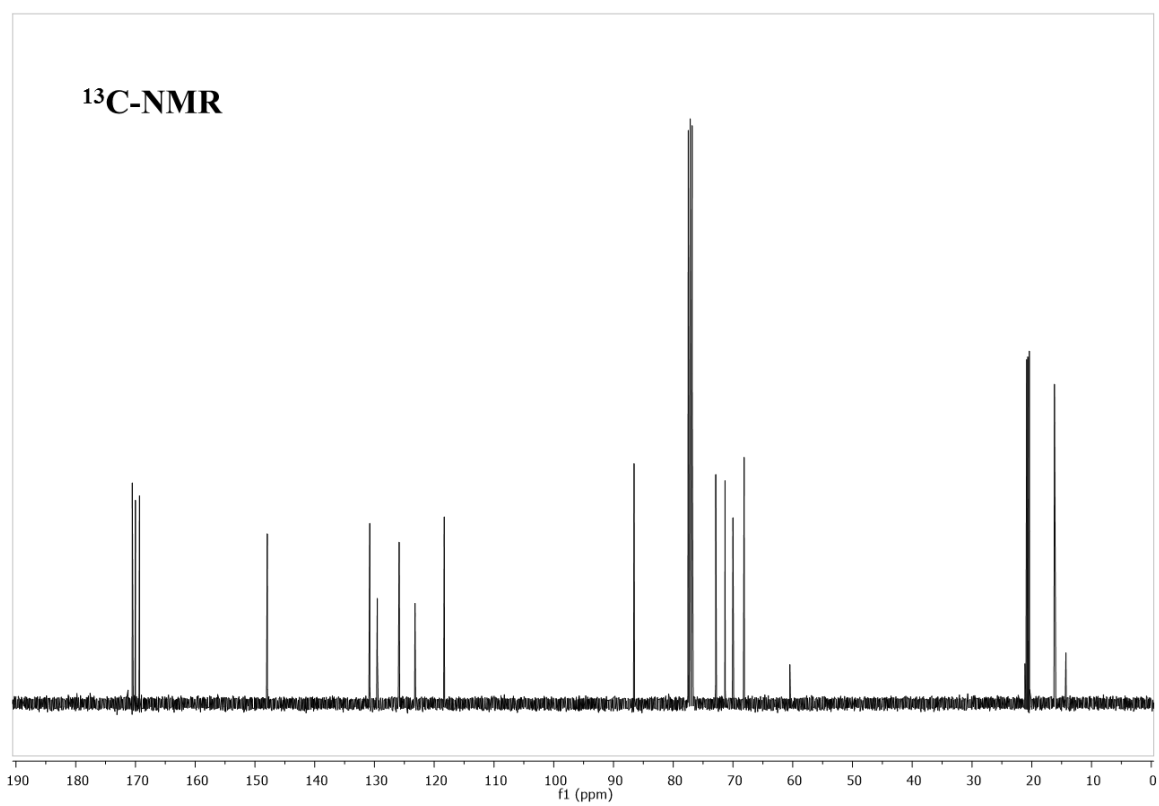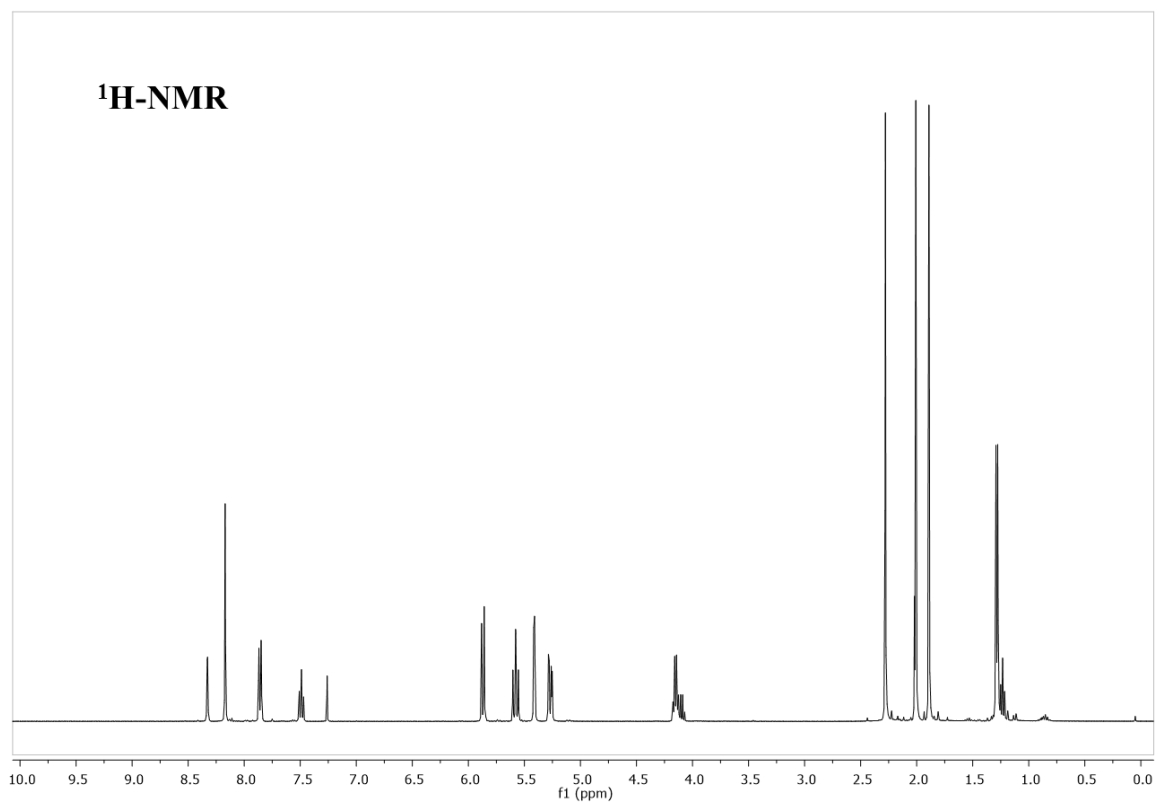
Synthesis of the peracetylated maltose symmetric PTDZ derivative **29**:
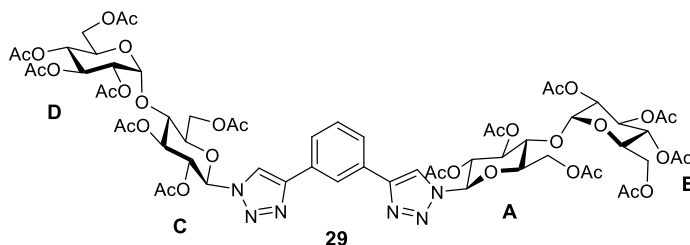
The general procedure was applied to the 2,3,6-tri-*O*-acetyl-1-azido-1-deoxy-4-*O*-(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)-β-D-glucopyranose **6** (397 mg, 0.6 mmol, 3 eq). The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 0:1) to yield **29** (227 mg, 80 %) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.25 (s, 1H,H$_{AR}$), 8.02 (s, 2H, 2xH$_{TRIAZOLE}$), 7.80 (dd, J=7.9, 1.1 Hz, 2H, 2xH$_{AR}$), 7.47 (t, J=7.8 Hz, 1H, H$_{AR}$), 5.94 (d, J=8.9 Hz, 2H, H$_{1A}$,H$_{1C}$), 5.51–5.43 (m, 4H, 4xH$_{SUGAR}$), 5.42–5.33 (m, 4H,2xH$_{2A}$,2xH$_{2C}$), 5.06 (t, J=9.8 Hz, 2H, 2xH$_{SUGAR}$), 4.89–4.88 (m, 1H, H$_{SUGAR}$), 4.87–4.85 (m, 1H, H$_{SUGAR}$), 4.48 (dd, J=12.4, 2.4 Hz, 2H, 2xH$_{SUGAR}$), 4.30–4.20 (m, 4H, 4xH$_{SUGAR}$), 4.16 (t, J=9.1 Hz, 2H, 2xH$_{SUGAR}$), 4.12–4.05 (m, 2H, 2xH$_{SUGAR}$), 4.05–3.94 (m, 4H, 4xH$_{SUGAR}$), 2.12 (d, J=0.8 Hz, 6H,6xCOCH$_3$), 2.09 (d, J=0.8 Hz, 6H, 6xCO-CH$_3$), 2.05 (d, J=0.9 Hz, 6H, 6xCOCH$_3$), 2.03–1.98 (m, 18H,18xCOCH$_3$), 1.84 (d, J=0.9 Hz, 6H, 6xCOCH$_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.64 (2xCOCH$_3$), 170.57 (2xCOCH$_3$), 170.42 (2xCOCH$_3$), 170.01 (2xCOCH$_3$), 169.98 (2xCOCH$_3$), 169.49 (2xCOCH$_3$), 169.34 (2xCOCH$_3$), 147.95 (2xC$_{TRIAZOLE}$), 130.60 (2xC$_{AR}$), 129.61 (CH$_{AR}$), 125.98 (CH$_{AR}$), 123.28 (CH$_{AR}$), 118.27 (2xCH$_{TRIAZOLE}$), 96.00 (C$_{1A,1C}$), 85.45 (C$_{1B,1D}$), 77.36, 75.50, 75.25, 72.57, 71.01, 70.10, 69.29, 68.86, 68.01 (9xC$_{SUGAR}$), 62.63 (C$_{6A,6B}$), 61.54 (C$_{6C,6D}$), 21.12 (2xCOCH$_3$), 20.92 (2xCOCH$_3$), 20.88 (2xCOCH$_3$), 20.78 (2xCOCH$_3$), 20.68 (2xCOCH$_3$), 20.66 (2xCOCH$_3$), 20.29 (2xCOCH$_3$). LRMS (ES$^+$) calculated for C$_{62}$H$_{76}$N$_6$NaO$_{34}$, (M+Na): 1471.4; found 1471.9. HRMS (FAB$^+$): calculated for C$_{62}$H$_{76}$N$_6$NaO$_{34}$, (M+Na): 1471.4300; found: 1471.4333.

¹H-NMR

¹³C-NMR
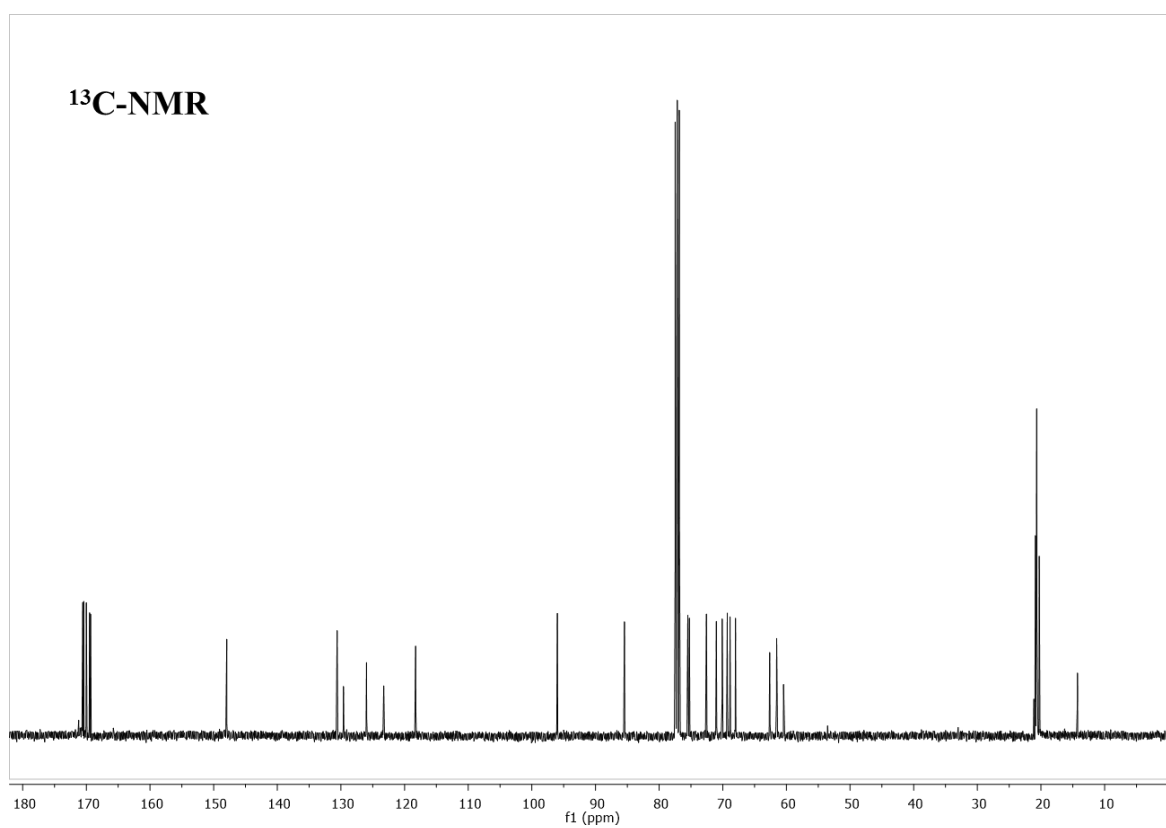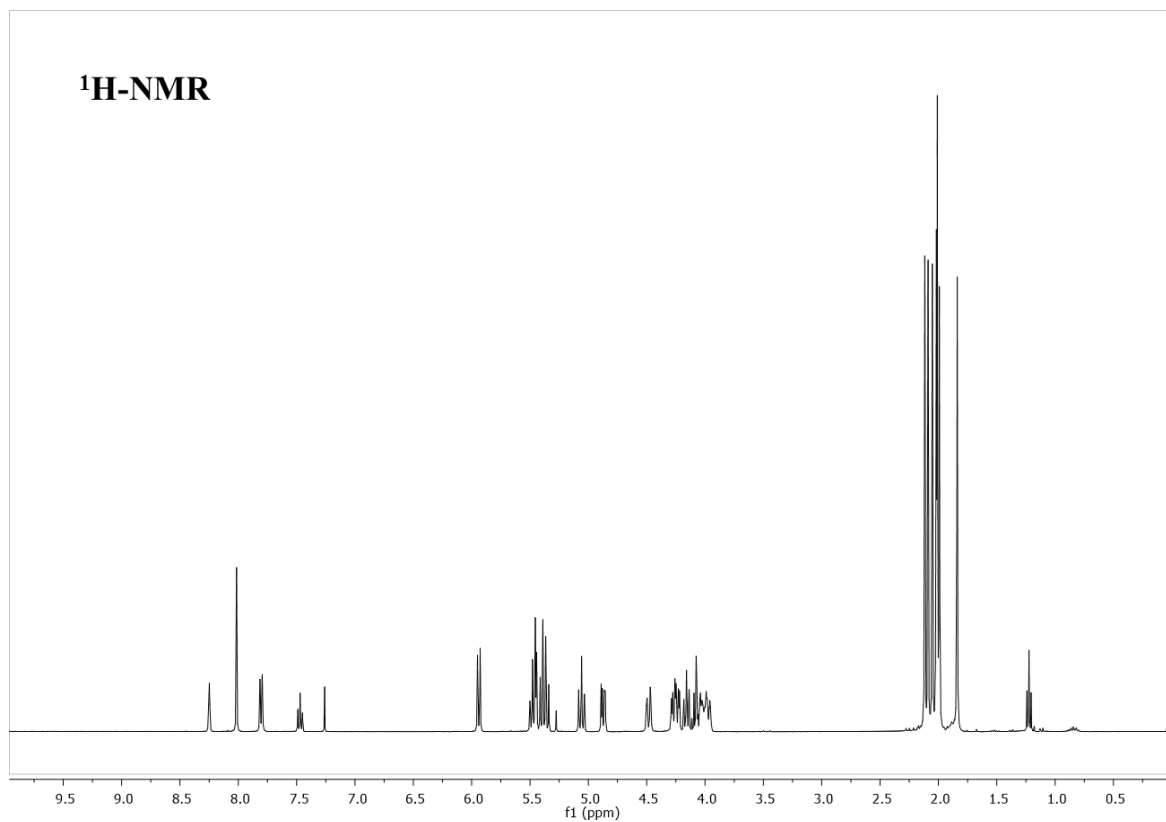
Synthesis of the peracetylated symmetric mannose C2-linker PTDZ derivative **30**:

30
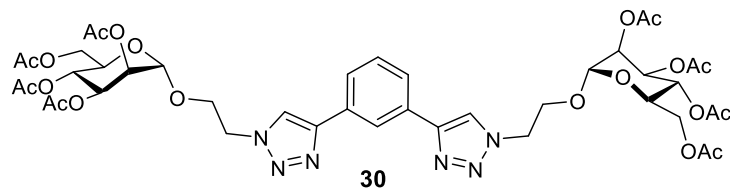
The general procedure was applied to the derivative **10** (200 mg, 0.48 mmol, 3 eq). The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 0:1) to give **30** (131 mg, 85%) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.15 (d, J=2.7 Hz, 3H, 2xH$_{TRIAZOLE}$, H$_{AR}$), 7.90 (dd, J=7.7, 1.7 Hz, 2H, 2xH$_{AR}$), 7.44 (t, J=7.8 Hz, 1H, H$_{AR}$), 5.22–5.16 (m, 4H, 2xH$_2$, 2xH$_3$), 5.15–5.07 (m, 2H, 2xH$_4$), 4.81 (d, J=1.6 Hz, 2H, 2xH$_1$), 4.68 (dt, J=14.4, 4.0 Hz, 2H,2xC$\underline{H}_2$N), 4.58 (ddd, J=14.4, 8.6, 3.7 Hz, 2H,2xCH$_2$N), 4.15–4.00 (m, 4H, 2xH$_6$, 2xOC$\underline{H}_2$), 3.96–3.82 (m, 4H, 2xH$_{6'}$, 2xOC$\underline{H}_2$), 3.16 (ddd, J=9.8, 4.9, 2.3 Hz, 2H, 2xH$_5$), 2.07 (s, 6H, 6xCOC$\underline{H}_3$), 2.02 (s, 6H, 6xCOC$\underline{H}_3$), 1.91 (s, 6H, 6xCOC$\underline{H}_3$), 1.64(s, 6H, 6xCOC$\underline{H}_3$) ^{13}C NMR (101 MHz, CDCl$_3$) δ 171.09 (2x$\underline{C}$OCH$_3$), 170.48 (2x$\underline{C}$OCH$_3$), 169.97 (2x$\underline{C}$OCH$_3$), 169.63 (2x$\underline{C}$OCH$_3$), 147.53 (C$_{AR}$), 131.05 (C$_{AR}$), 129.46 ($\underline{C}$H$_{AR}$), 125.39 (2x$\underline{C}$H$_{AR}$), 123.17 (C$_{AR}$), 121.70 (2x$\underline{C}$H$_{AR}$), 96.94 (2xC$_1$), 69.19 (2xC$_2$), 68.85 (2xC$_5$), 65.70 (2xC$_3$), 65.43 (2xC$_4$), 62.05 (O$\underline{C}$H$_2$), 60.35 (2xC$_6$), 49.77 (2x$\underline{C}$H$_2$N), 21.03 (2xCO$\underline{C}$H$_3$), 20.81 (2xCO$\underline{C}$H$_3$), 20.70 (2xCO$\underline{C}$H$_3$), 20.17 (2xCO$\underline{C}$H$_3$). LRMS (ES$^+$) calculated for C$_{42}$H$_{52}$N$_6$Na0$_{20}$ (M+Na): 983.3, found: 983.5 HRMS (FAB$^+$) Calculated for C$_{42}$H$_{52}$N$_6$Na0$_{20}$(M+Na):983.3134; found: 983.3096.
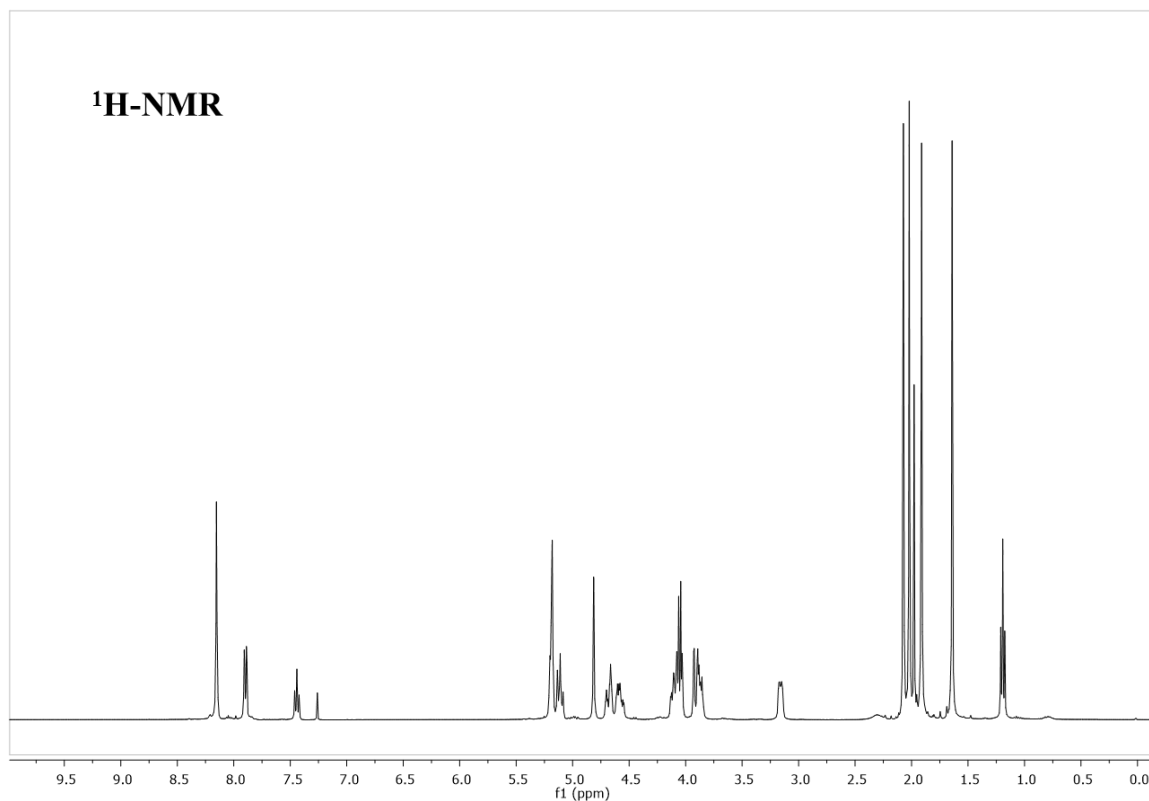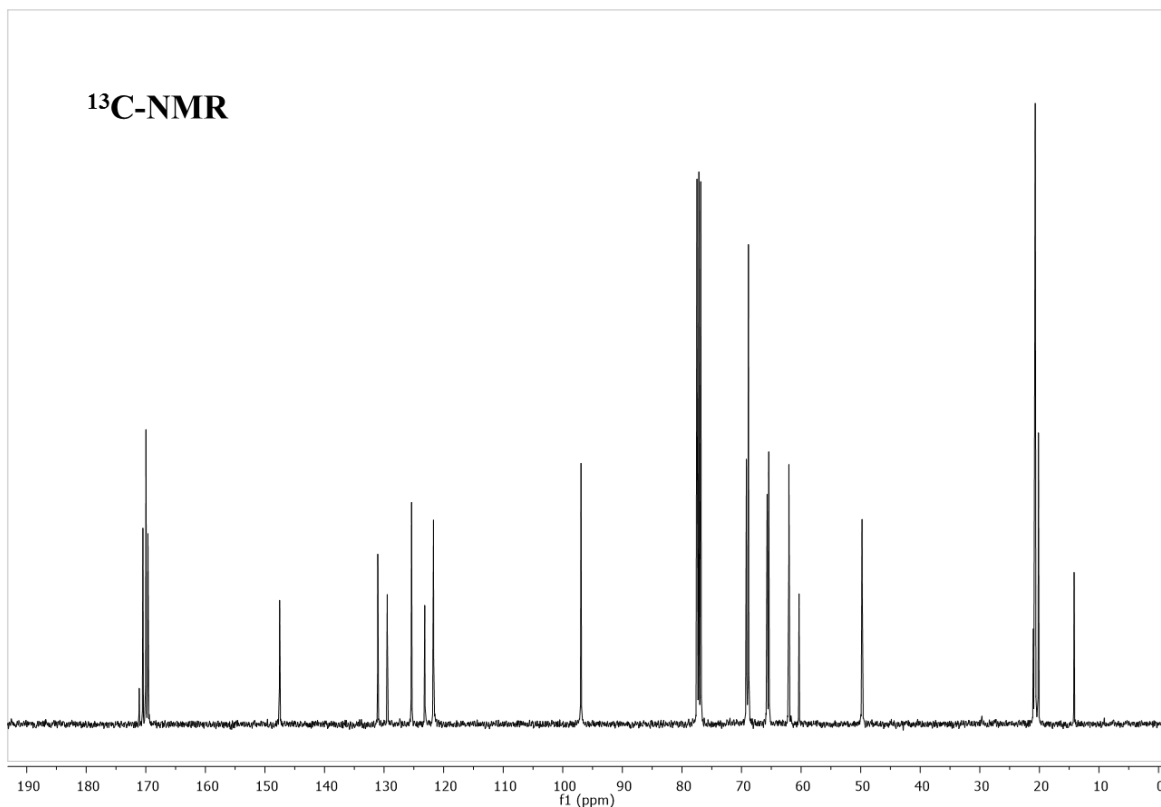
Synthesis of the peracetylated glucose C2-linker symmetric PTDZ derivative **31**:
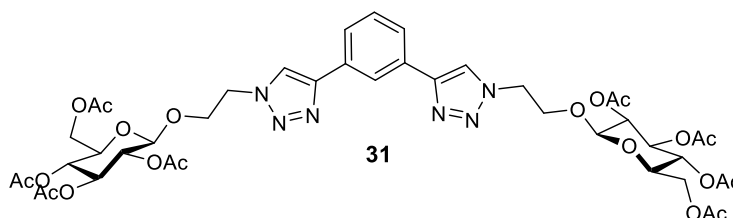
The general procedure was applied to the derivative **14** (200 mg, 0.48 mmol, 3 eq). The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (3:2 to 0:1) to give **31** (127 mg, 83%) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.29 (s, 1H, H$_{AR}$), 7.95 (s, 2H, 2xH$_{AR}$), 7.85 (d, J=7.8 Hz, 2H, 2xH$_{AR}$), 7.47 (t, J=7.7 Hz, 1H, H$_{AR}$), 5.14 (t, J=9.5 Hz, 2H, 2xH$_3$), 5.03 (dt, J=15.2, 9.5 Hz, 4H, 2xH$_4$, 2xH$_2$), 4.73–4.62 (m, 2H, 2xNC$\underline{H}_2$), 4.53 (ddd, J=14.2, 9.2, 3.4 Hz, 2H, 2xNC$\underline{H}_2$), 4.47 (d, J=7.9 Hz, 2H, 2xH$_1$), 4.31–4.20 (m, 4H, 2xOC$\underline{H}_2$, 2xH$_6$), 4.14–4.07 (m, 2H, 2xOC$\underline{H}_2$), 3.96–3.89 (m, 2H, 2xH$_{6'}$), 3.69 (ddd, J=10.0, 4.5, 2.3 Hz, 2H, 2xH$_5$), 2.05 (s, 6H, 6xCOC$\underline{H}_3$), 1.98 (s, 6H, 6xCOC$\underline{H}_3$), 1.95 (s, 6H, 6xCOC$\underline{H}_3$), 1.71 (s, 6H, 6xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.66 (2x$\underline{C}$OCH$_3$), 170.13 (2x$\underline{C}$OCH$_3$), 169.50 (2x$\underline{C}$OCH$_3$), 169.47 (2x$\underline{C}$OCH$_3$), 147.27 (2xC$_{AR}$), 131.13 (2xC$_{AR}$), 129.55 ($\underline{C}$H$_{AR}$), 125.42 ($\underline{C}$H$_{AR}$), 122.94 ($\underline{C}$H$_{AR}$), 121.80 ($\underline{C}$H$_{AR}$), 100.62 (2xC$_1$), 72.52 (2xC$_3$), 72.05 (2xC$_5$), 70.92 (2xC$_2$), 68.30 (2xC$_4$), 67.85 (2xC$_6$), 61.79 (2xOC$\underline{H}_2$), 50.18 (2xNC$\underline{H}_2$), 21.12 (2xCO$\underline{C}$H$_3$), 20.80 (2xCO$\underline{C}$H$_3$), 20.63 (2xCO$\underline{C}$H$_3$), 20.45 (2xCO$\underline{C}$H$_3$). HRMS (FAB$^+$): calculated for C$_{42}$H$_{52}$N$_6$NaO$_{20}$ (M+Na): 983.3134; found: 983.3099.
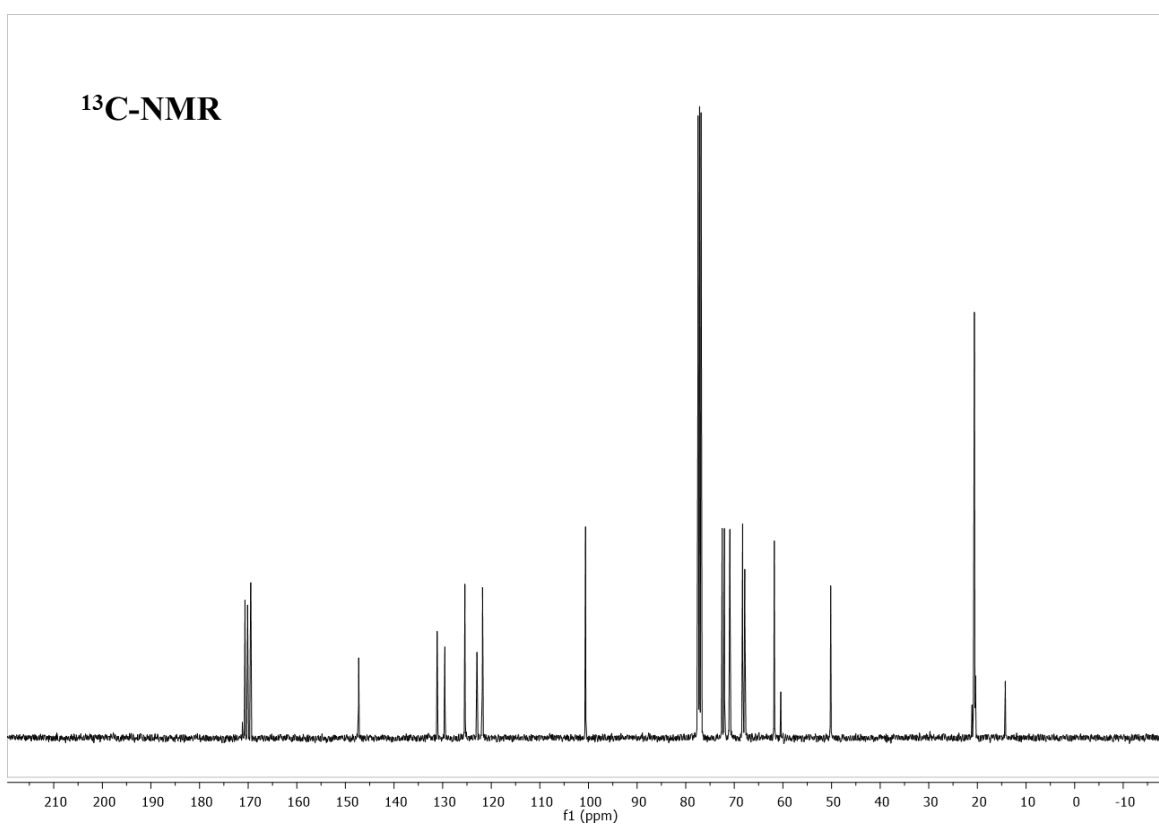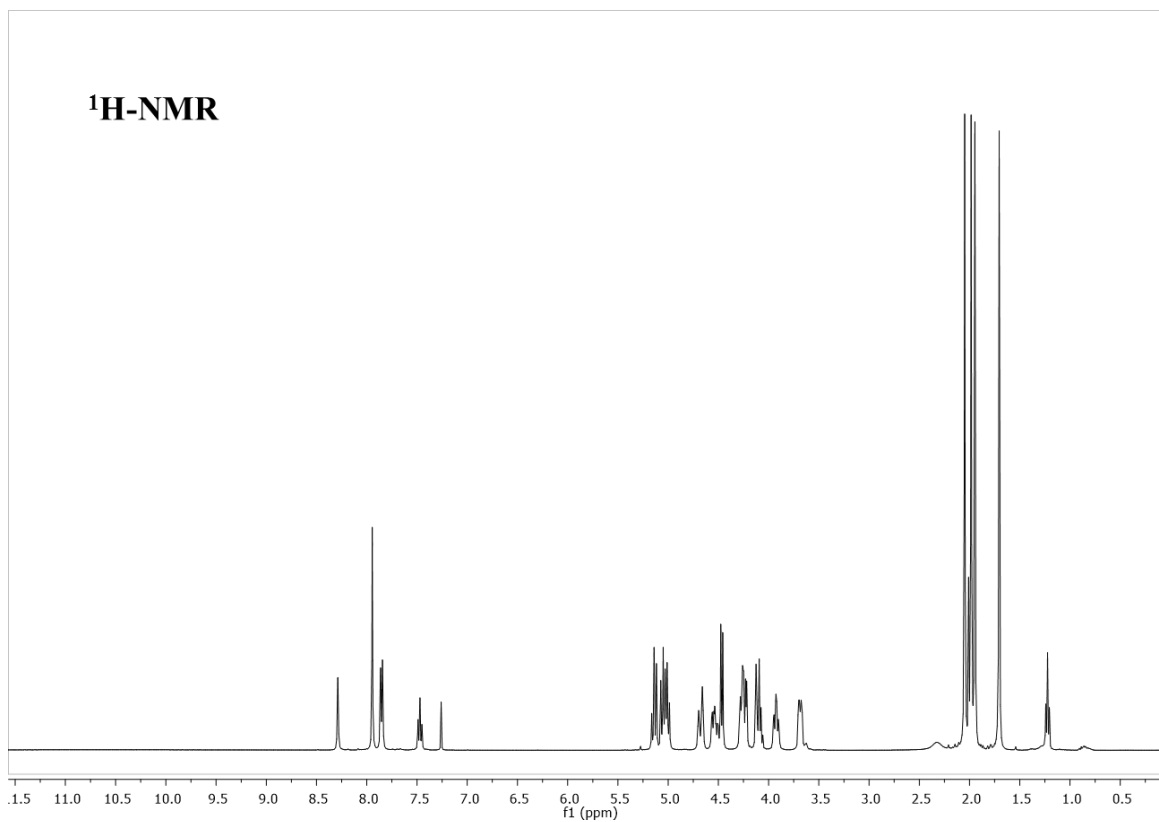
General procedure of deprotection for the peracetylated symmetric PTDZ ligands

Sodium methylate solution in MeOH (1 N, 1 mL/g starting material) was added to the corresponding acetylated derivative in dry MeOH. The reaction was stirred at room temperature until total starting material consumption (~1h). Amberlite IR®120 ion-exchange resin was then added to neutralize the mixture. The resin was filtered off, washed with MeOH, and the filtrate was evaporated. Solvent was removed to afford the final symmetric compound.

<u>Synthesis of the glucose symmetric PTDZ derivative **43**</u>:
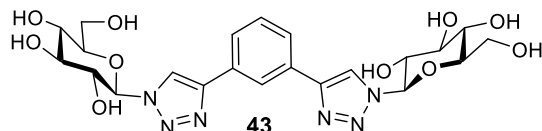
The general procedure was applied to the acetylated derivative **26** (104 mg, 0.119 mmol). A white solid was obtained (64 mg, 100%). ^{1}H NMR (400 MHz, D$_2$O) δ 7.89 (s, 2H, 2xH$_{TRIAZOLE}$), 7.04–7.00 (m, 2H, 2xH$_{AR}$), 6.97 (d, J=6.1 Hz, 2H, 2xH$_{AR}$), 5.58 (d, J=9.2 Hz, 2H, 2xH$_1$), 4.00–3.91 (m, 4H, 2xH$_2$, 2xH$_6$), 3.84 (dd, J=12.7, 5.1 Hz, 2H, 2xH$_5$), 3.73 (t, J=9.1 Hz, 4H, 2xH$_3$, 2xH$_{6'}$), 3.68–3.61 (m, 2H, 2xH$_4$). ^{13}C NMR (101 MHz, D$_2$O) δ 146.14 (2xC$_{AR}$), 129.02 (<u>C</u>H$_{AR}$), 128.60 (2x<u>C</u>H$_{AR}$), 124.82 (2x<u>C</u>H$_{AR}$), 121.53 (2xC$_{AR}$), 120.82 (<u>C</u>H$_{AR}$), 87.46 (2xC$_1$), 78.84 (2xC$_5$), 75.89 (2xC$_3$), 72.40 (2xC$_2$), 68.96 (2xC$_4$), 60.60 (2xC$_6$). LRMS (ES$^-$) calculated for C$_{22}$H$_{27}$N$_6$O$_{10}$, (M-H): 535.1, found 535.1. HRMS (FAB$^+$) calculated for C$_{22}$H$_{27}$N$_6$O$_{10}$, (M+H): 537.1945; found: 537.1919.
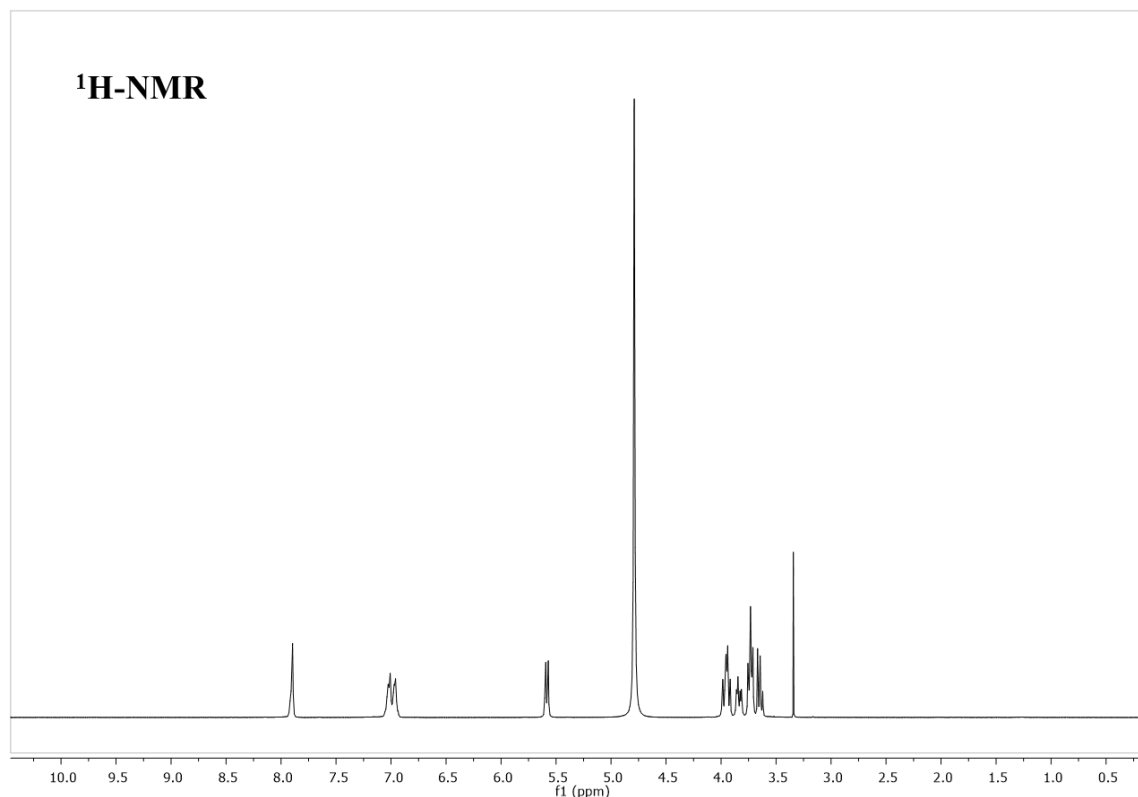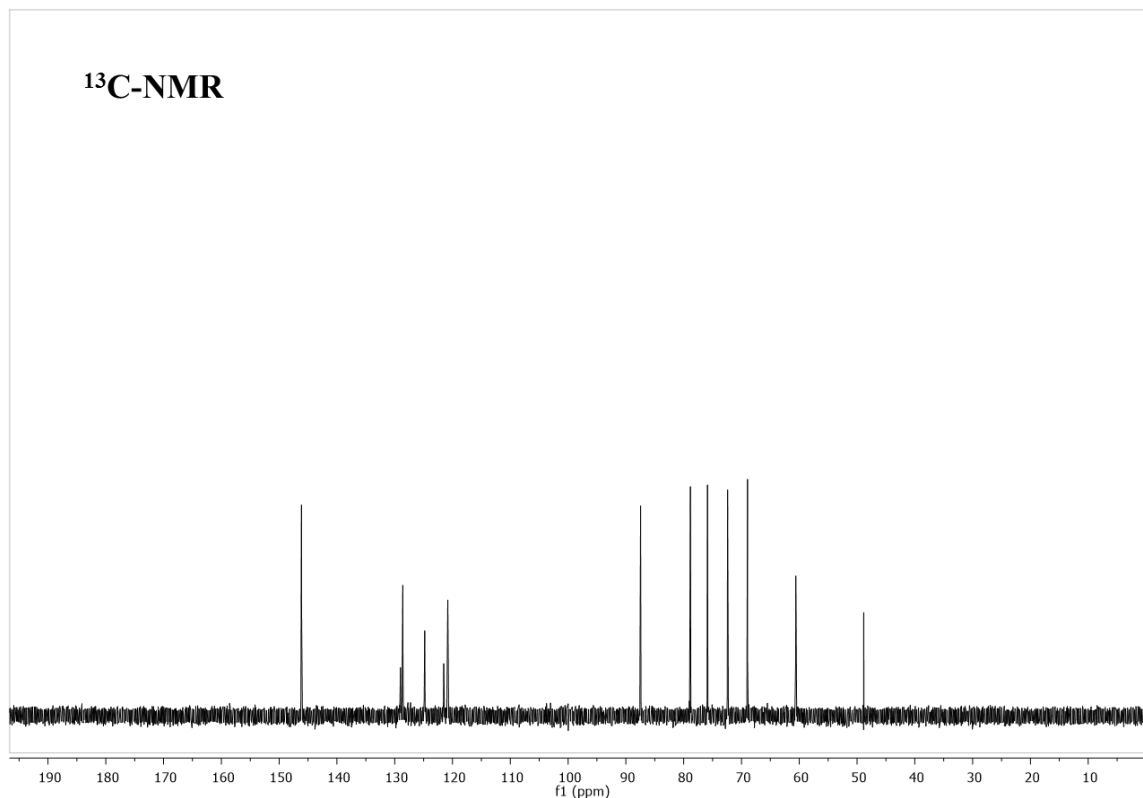
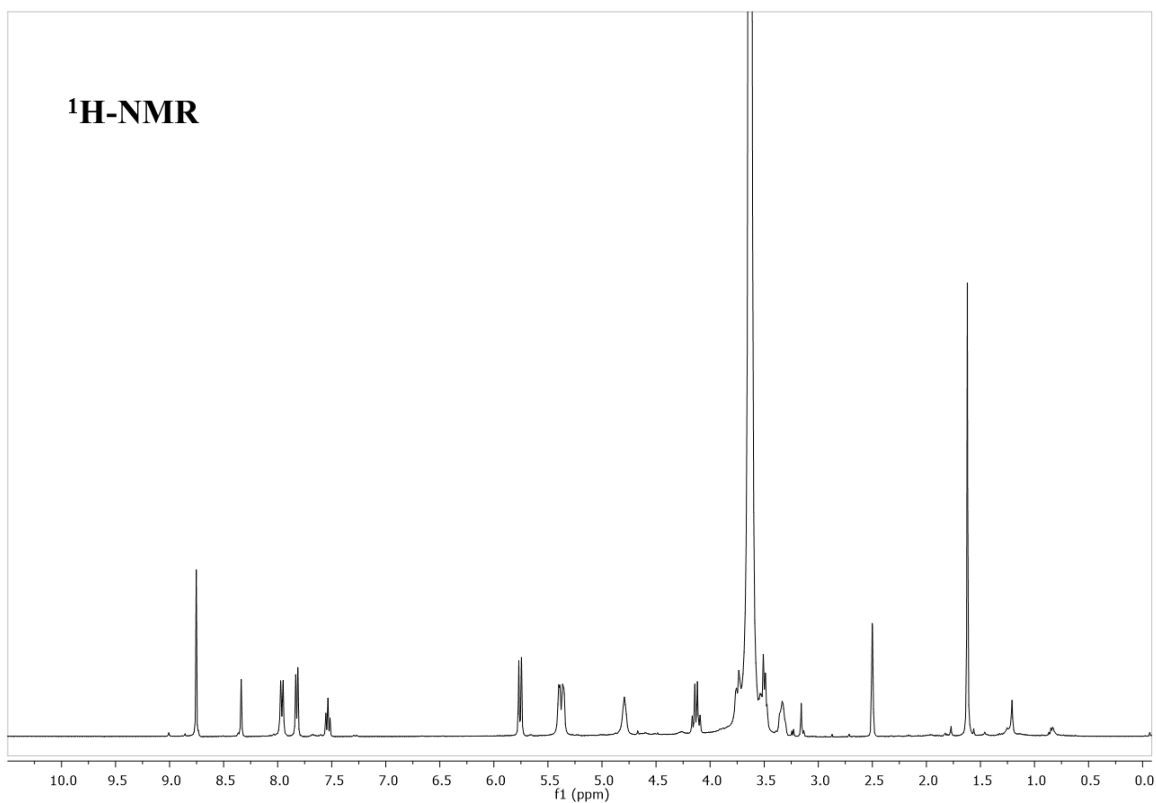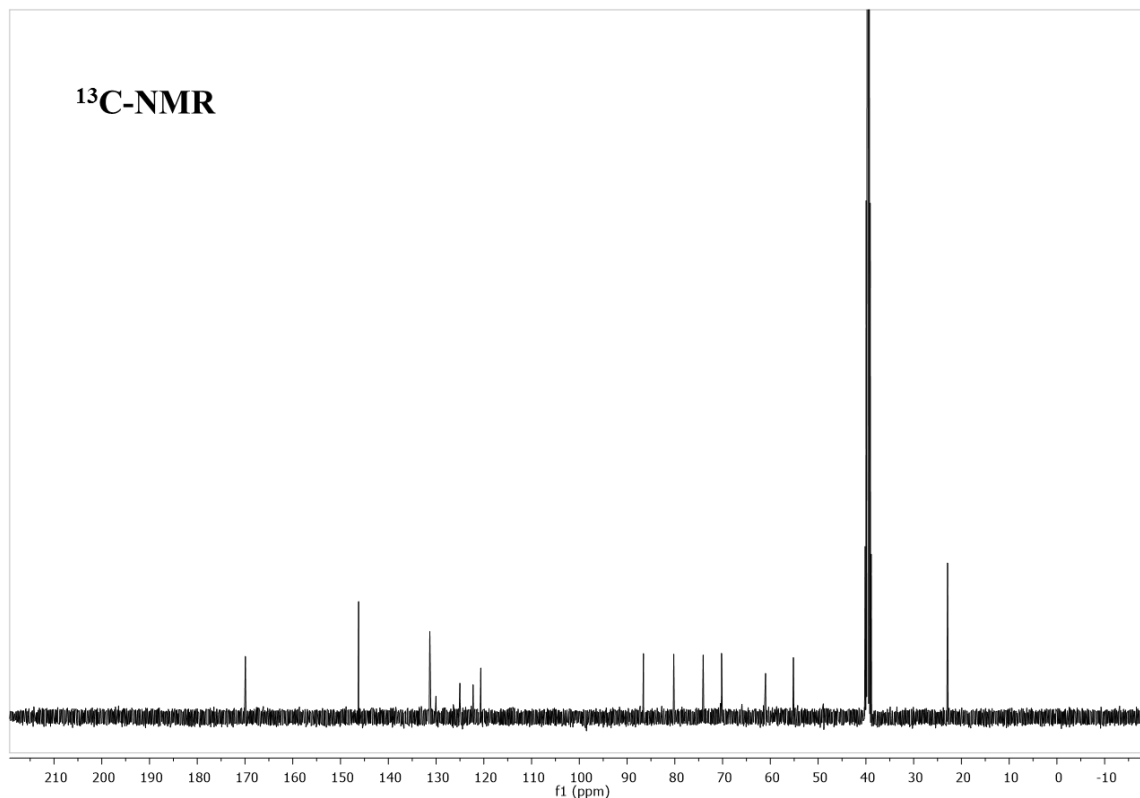Synthesis of the maltose symmetric PTDZ derivative **44**:
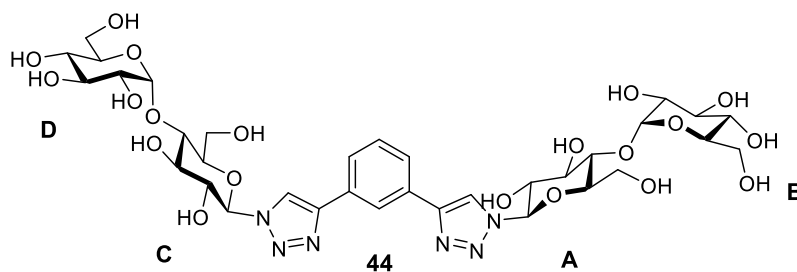
The general procedure was applied to the acetylated derivative **29** (226 mg, 0.156 mmol). A white solid was obtained (137 mg, 99%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 8.93 (s, 2H, 2xH$_{TRIAZOLE}$), 8.45 (s, 1H, H$_{AR}$), 7.87 (d, J=7.8 Hz, 2H, 2xH$_{AR}$), 7.56 (t, J=7.7 Hz, 1H, H$_{AR}$), 5.81 (d, J=3.0 Hz, 2H, 2xO<u>H</u>), 5.69 (d, J=9.1 Hz, 2H, H$_{1A}$, H$_{1C}$), 5.60 (t, J=6.2 Hz, 4H, H$_{3B}$, H$_{3D}$, H$_{3A}$, H$_{3C}$), 5.11 (d, J=3.8 Hz, 2H, H$_{1B}$, H$_{1D}$), 4.95 (d, J=5.2 Hz, 4H, 4xH$_{SUGAR}$), 4.63 (t, J=5.8 Hz, 2H, 2xH$_{SUGAR}$), 4.56 (t, J=5.6 Hz, 2H, 2xH$_{SUGAR}$), 4.11 (d, J=4.8 Hz, 1H, H$_{SUGAR}$), 3.90 (td, J=9.1, 5.5 Hz, 2H, H$_{2A}$, H$_{2C}$), 3.76 (dd, J=8.7, 3.0 Hz, 3H, 3xH$_{SUGAR}$), 3.65 (qd, J=13.1, 11.8, 5.7 Hz, 5H, 5xH$_{SUGAR}$), 3.58–3.40 (m, 8H, 8xH$_{SUGAR}$), 3.28 (dq, J=15.9, 6.4, 5.7 Hz, 2H, H$_{2B}$, H$_{2D}$), 3.17 (d, J=4.0 Hz, 3H, 3xH$_{SUGAR}$), 3.10 (td, J=9.1, 5.1 Hz, 2H, 2xH$_{SUGAR}$). ^{13}C NMR (101 MHz, DMSO-d$_6$) δ 146.24 (2xC$_{AR}$), 131.24 (2xC$_{AR}$), 129.70 (<u>C</u>H$_{AR}$), 124.69 (<u>C</u>H$_{AR}$), 122.02 (<u>C</u>H$_{AR}$), 120.74

(C$\underline{\text{H}}_{\text{AR}}$), 100.96 (C$_{1B,1D}$), 87.46 (C$_{1A,1C}$), 79.20, 78.00, 76.52, 73.63, 73.30, 72.48, 71.90, 69.93, 60.84, 60.32(10xC$_{\text{sugar}}$). LRMS (ES$^-$) calculated for C$_{34}$H$_{47}$N$_6$O$_{20}$, (M-H): 859.3; found 859.3. HRMS (FAB$^+$) calculated for C$_{34}$H$_{48}$N$_6$NaO$_{20}$ (M+Na): 883.2821; found: 883.2820.
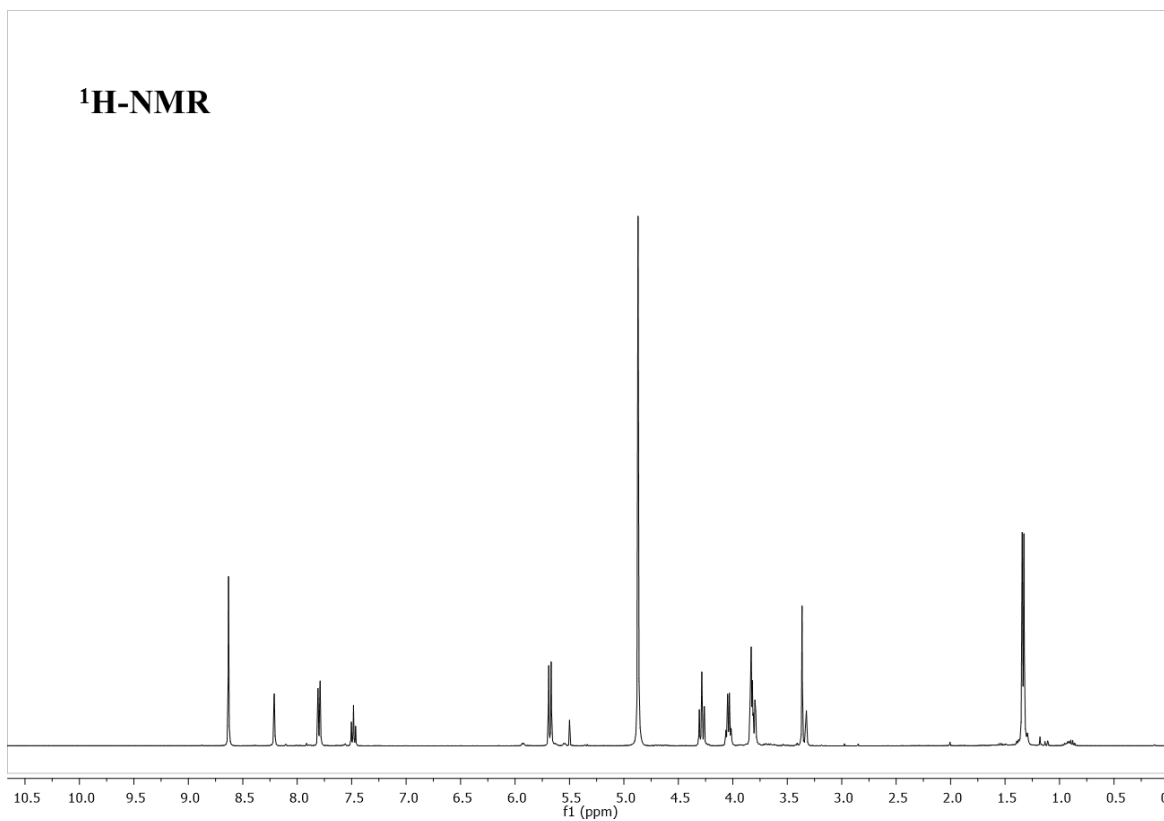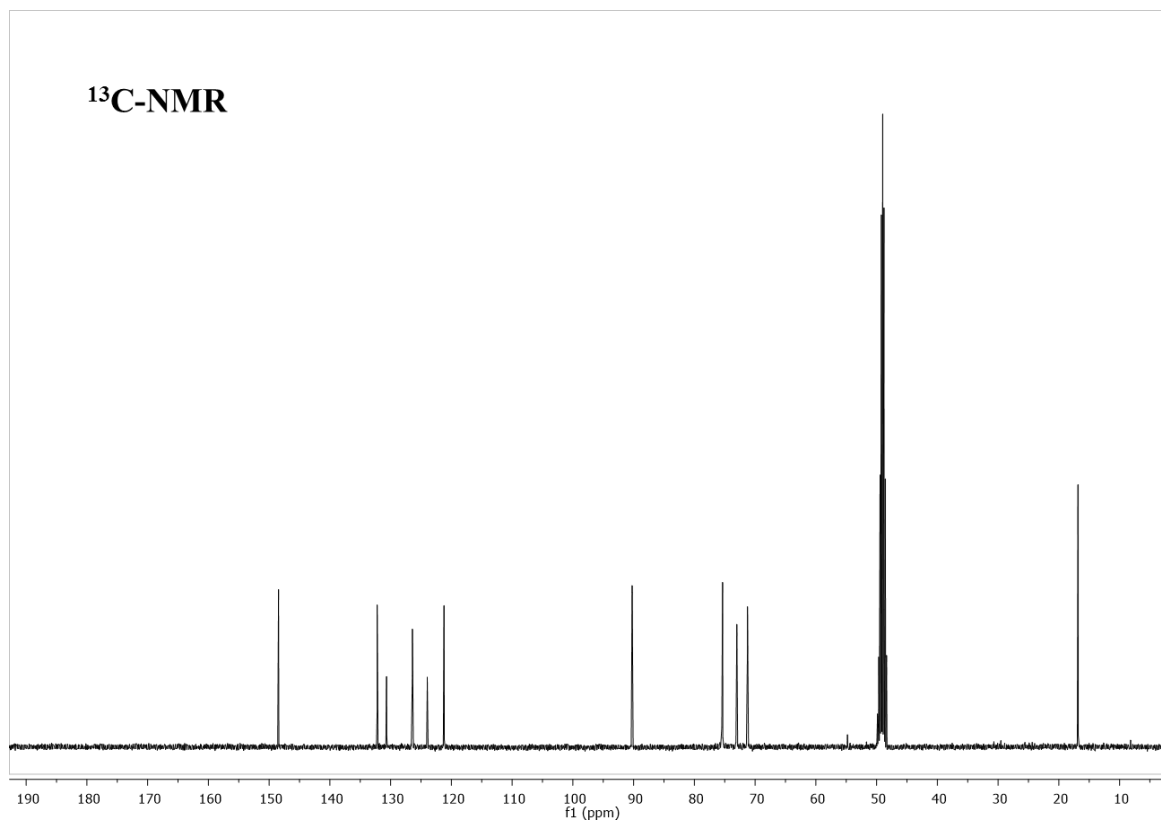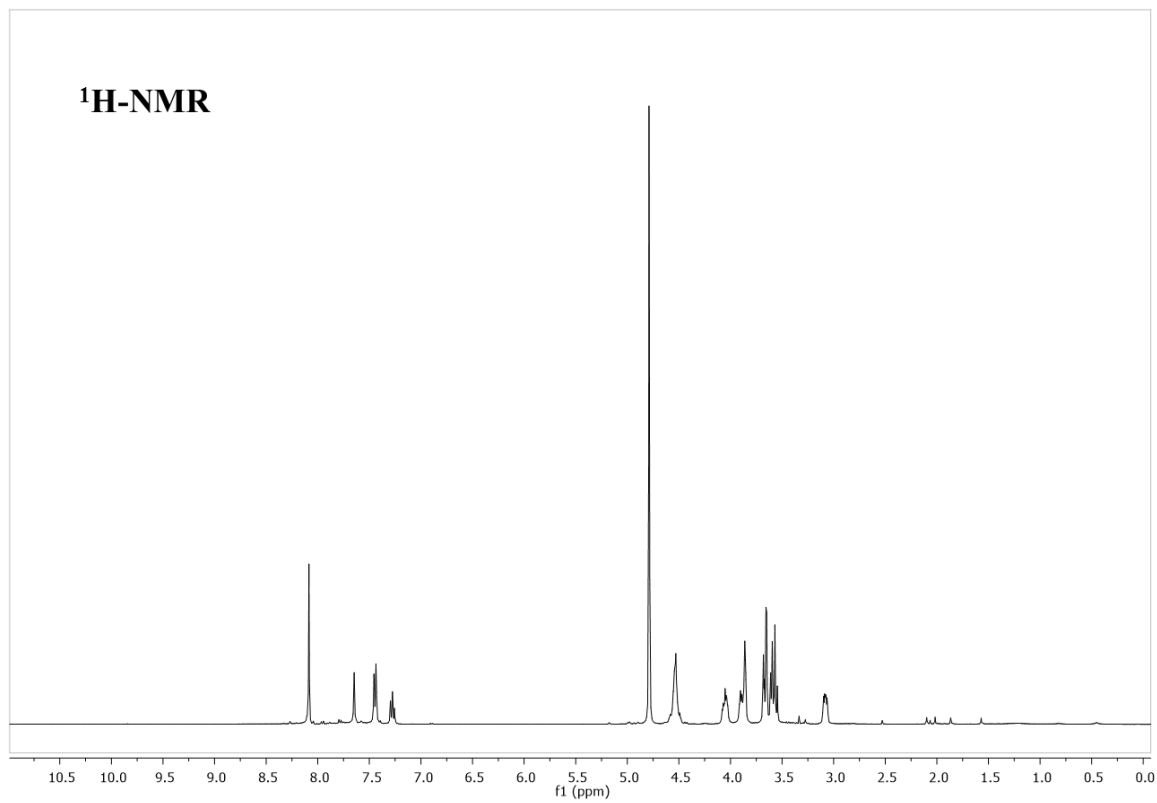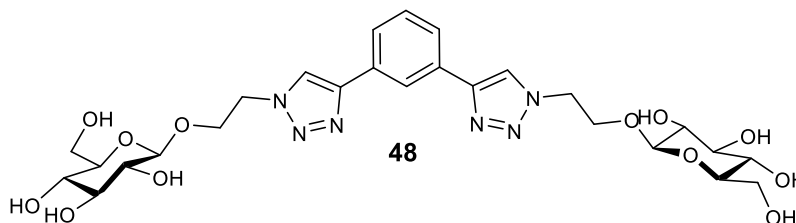
^{1}H-NMR

Synthesis of the fucose symmetric PTDZ derivative **45**:
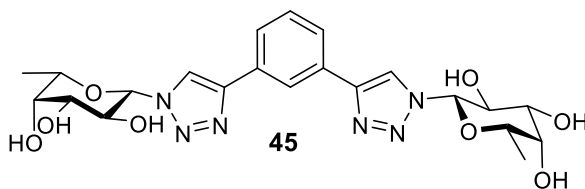
The general procedure was applied to the acetylated derivative **28** (296 mg, 0.391 mmol). A white solid was obtained (197 mg, 99%). ^{1}H NMR (400 MHz, CD$_3$OD) δ 8.63 (d, J=0.9 Hz, 2H, 2xH$_{TRIAZOLE}$), 8.21 (d, J=1.9 Hz, 1H, H$_{AR}$), 7.82–7.78 (m, 2H, 2xH$_{AR}$), 7.48 (t, J=7.8 Hz, 1H, H$_{AR}$), 5.68 (d, J= 9.1 Hz, 2H, 2xH$_1$), 4.29 (t, J=9.2 Hz, 2H, 2xH$_2$), 4.04 (q, J=6.4 Hz, 2H, 2xH$_5$), 3.85–3.78 (m, 4H, 2xH$_3$, 2xH$_4$), 1.34 (d, J=6.4 Hz, 6H, 6xC$\underline{H}_3$). ^{13}C NMR (101 MHz, CD$_3$OD) δ 148.46 (2xC$_{TRIAZOLE}$), 132.22 (2xC$_{AR}$), 130.68 (2xCH$_{AR}$), 126.42 (2xCH$_{AR}$), 123.97 (CH$_{AR}$), 121.21 (2xCH$_{TRIAZOLE}$), 90.24 (2xC$_1$), 75.43 (2xC$_4$), 75.36 (2xC$_3$), 73.02 (2xC$_5$), 71.25 (2xC$_2$), 16.84 (2xC$\underline{H}_3$). LRMS (ES$^+$) calculated for C$_{22}$H$_{27}$N$_6$O$_8$, (M-H): 503.2; found 503.1. HRMS (FAB$^+$) calculated for C$_{22}$H$_{28}$N$_6$NaO$_8$ (M+Na): 527.1866; found: 527.1861.

¹H-NMR

¹³C-NMR
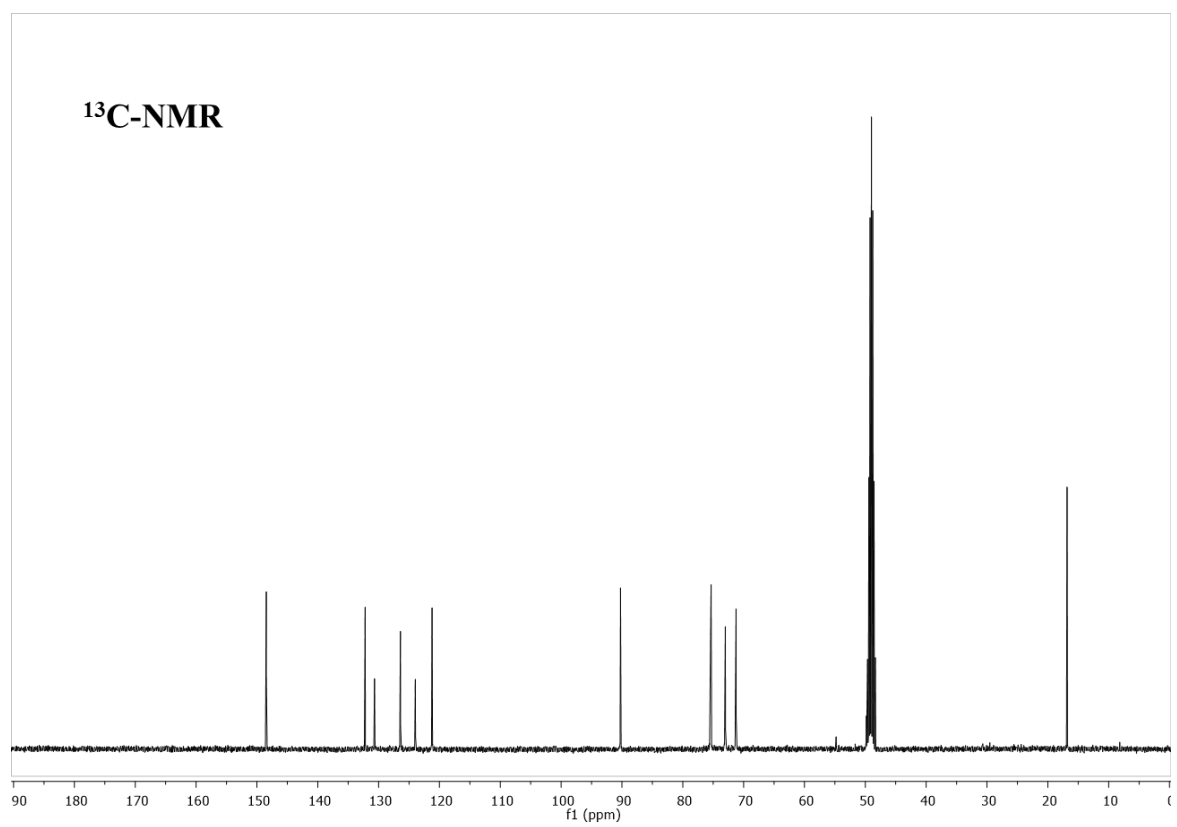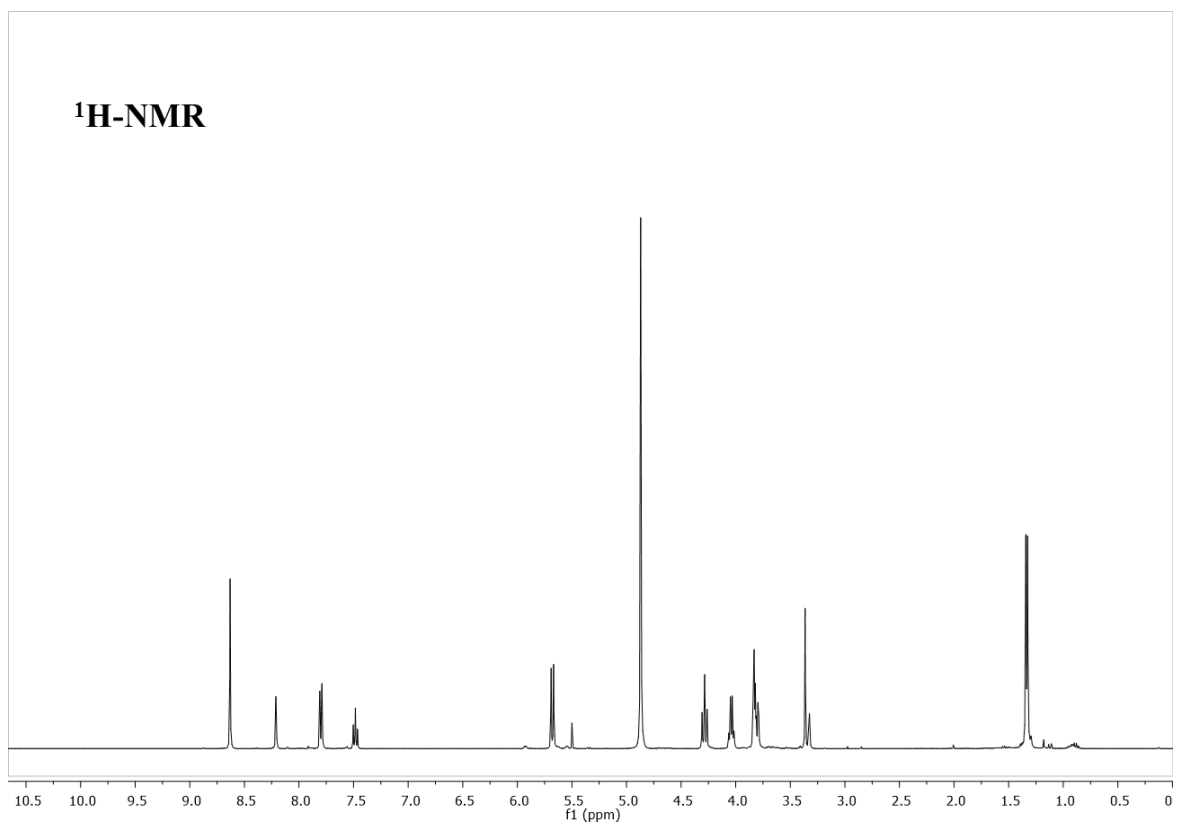
Synthesis of the *N*-acetyl-glucosamine symmetric PTDZ derivative **46**:
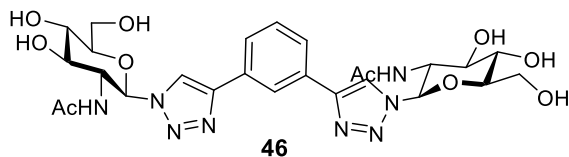
The general procedure was applied to the acetylated derivative **27** (92 mg, 0.106 mmol). A brown solid was obtained (73 mg, 100%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.75 (s, 2H, 2xH$_{TRIAZOLE}$), 8.33 (s, 1H, H$_{AR}$), 7.96 (d, J=9.1 Hz, 2H, 2xO$\underline{H}$), 7.82 (d, J=7.9 Hz, 2xH$_{AR}$,), 7.53 (t, J=7.7 Hz, 1H, H$_{AR}$), 5.75 (d, J=9.8 Hz, 2H, 2xH$_1$), 5.37 (dd, J=15.3, 5.3 Hz, 4H, 4xO$\underline{H}$), 4.13 (q, J=9.7 Hz, 2H, 2xH$_2$), 3.79–3.70 (m, 2H, 2xH$_6$), 3.51 (t, J=9.9 Hz, 6H, 2xH$_3$, 2xH$_{6'}$, 2xH$_5$), 3.34 (dd, J=9.2, 4.6 Hz, 2H, 2xH$_4$), 1.62 (s, 6H, 6xNHCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, DMSO-d_6) δ 169.93 (2xNH$\underline{C}$OCH$_3$), 146.22 (2xC$_{AR}$), 131.32 (2xC$_{AR}$), 130.02 ($\underline{C}$H$_{AR}$), 125.01 (2x$\underline{C}$H$_{AR}$), 122.25 ($\underline{C}$H$_{AR}$), 120.63 (2x$\underline{C}$H$_{TRIAZOLE}$), 86.58 (2xC$_1$), 80.24 (2xC$_5$), 74.08 (2xC$_3$), 70.22 (2xC$_4$), 61.02 (2xC$_6$), 55.18 (2xC$_2$), 22.90 (2x$\underline{C}$OCH$_3$). LRMS (ES$^-$) calculated for C$_{26}$H$_{33}$N$_8$O$_{10}$, (M-H): 617.2; found 617.3. HRMS (FAB$^+$) calculated for C$_{26}$H$_{34}$N$_8$NaO$_{10}$ (M+Na): 641.2296; found: 641.1635.

Synthesis of the mannose C2-linker symmetric PTDZ derivative **47**:
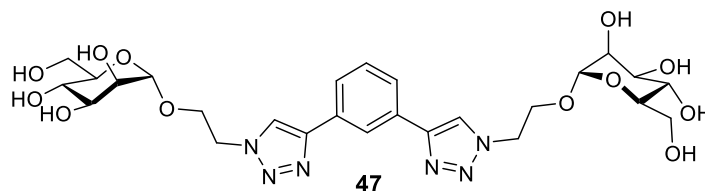
The general procedure was applied to the acetylated derivative **30** (130 mg, 0.135 mmol). A brown solid was obtained (64 mg, 76%). ^{1}H NMR (400 MHz, D$_2$O) δ 8.08 (s, 2H, 2xH$_{\text{TRIAZOLE}}$), 7.69–7.62 (m, 1H, H$_{\text{AR}}$), 7.44 (dd, J=7.7, 1.7 Hz, 2H,2xH$_{\text{AR}}$), 7.28 (t, J=7.8 Hz, 1H, H$_{\text{AR}}$), 4.79 (d, J=3.5 Hz, 2H, 2xH$_1$)4.60–4.48 (m, 4H, 4xNCH$_2$), 4.05 (ddd, J=10.9, 7.1, 4.1 Hz, 2H, 2xOCH$_2$), 3.93–3.83 (m, 4H, 2H$_2$, 2xOCH$_2$), 3.66 (dd, J=9.7, 3.1 Hz, 4H, 2xH$_3$, 2xH$_6$'), 3.63–3.53 (m, 4H, 2xH$_4$, 2xH$_6$), 3.08 (ddd, J=9.8, 5.5, 2.4 Hz, 2H, 2xH$_5$). ^{13}C NMR (101 MHz, D$_2$O) δ 146.57 (2xC$_{\text{TRIAZOLE}}$), 129.58 (CH$_{\text{AR}}$), 125.24 (CH$_{\text{AR}}$), 122.35 (CH$_{\text{TRIAZOLE}}$), 121.90 (CH$_{\text{AR}}$), 99.54 (2xC$_1$), 72.77, 70.43, 69.88, 66.29 (2xC$_2$,2xC$_3$,2xC$_4$,2xC$_5$), 65.38 (2xOCH$_2$), 60.58 (2xC$_6$), 50.10 (2xNCH$_2$). LRMS (ES$^-$) calculated for C$_{26}$H$_{36}$N$_6$O$_{12}$, (M-H): 647.2; found 647.4. HRMS (FAB$^+$), calculated for C$_{26}$H$_{36}$N$_6$NaO$_{12}$ (M+Na):647.2289; found: 647.2263.

¹H-NMR

¹³C-NMR
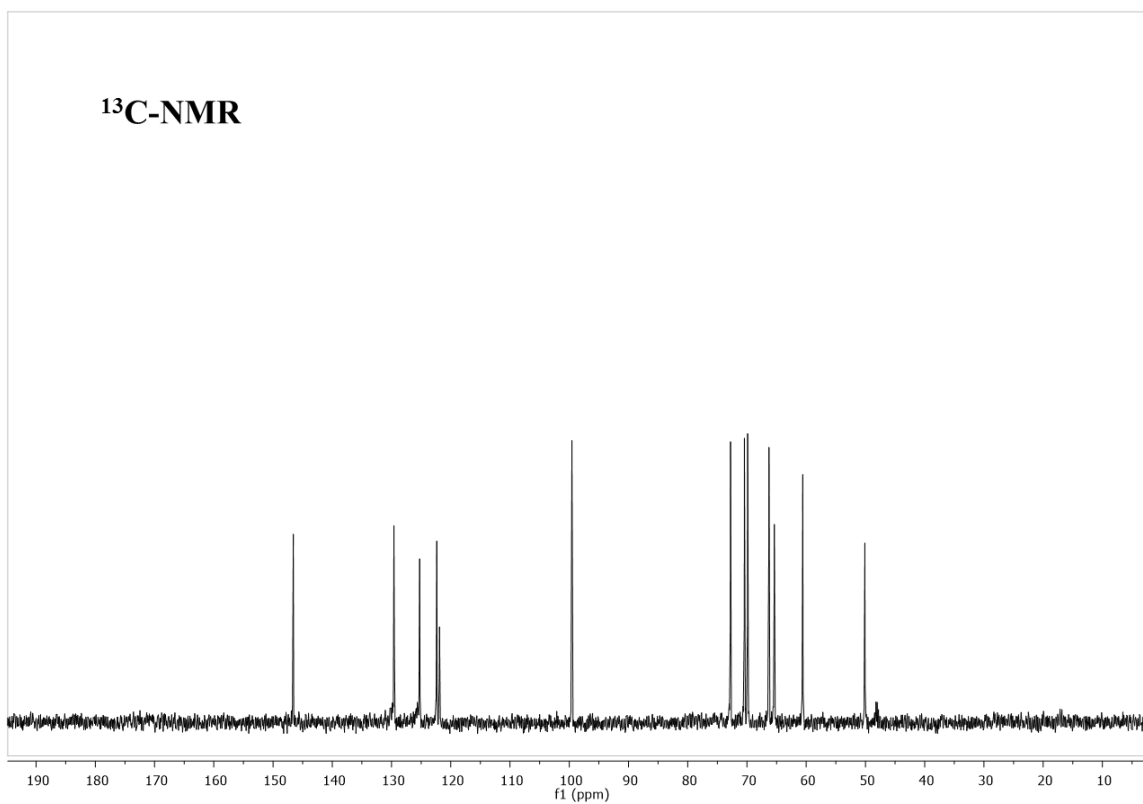
Synthesis of the glucose C2-linker symmetric PTDZ derivative **48**:

The general procedure was applied to the acetylated derivative **31** (125 mg, 0.130 mmol). A brown solid was obtained (33 mg, 41%). ^{1}H NMR (300 MHz, CD$_3$OD) δ 8.54 (s, 2H, 2xH$_{TRIAZOLE}$), 8.24 (s, 1H, H$_{AR}$), 7.86–7.77 (m, 2H, 2xH$_{AR}$), 7.50 (t, J=7.8 Hz, 1H, H$_{AR}$), 4.72 (t, J=4.9 Hz, 4H, 4xNCH$_2$), 4.33 (dt, J=16.2, 6.2 Hz, 4H, 2xH$_1$, 2xCH$_2$O), 4.07 (dt, J=11.1, 5.1 Hz, 2H, 2xCH$_2$O), 3.88 (d, J=11.8 Hz, 2H, 2xH$_6$), 3.75–3.59 (m, 4H, 2xH$_{6'}$, 2xH$_5$), 3.43–3.18 (m, 6H, 2xH$_3$, 2xH$_2$, 2xH$_4$). ^{13}C NMR (75 MHz, CD$_3$OD) δ 132.40, 126.43, 123.89, 123.84, 123.82, 118.90, 113.11(CH$_{AR}$), 104.50 (2xC$_1$), 78.04, 77.99, 74.96, 71.50 (2xC$_2$, 2xC$_3$, 2xC$_4$, 2xC$_5$), 69.05 (2xCH$_2$O), 62.67 (2xC$_6$), 51.80 (2xCH$_2$N). LRMS (ES$^+$) calculated for C$_{26}$H$_{36}$N$_6$NaO$_{12}$ (M+Na):647.2; found 647.3. HRMS (FAB$^+$), calculated for C$_{26}$H$_{36}$N$_6$NaO$_{12}$ (M+Na):647.2289; found: 647.2265.

^{13}C-NMR
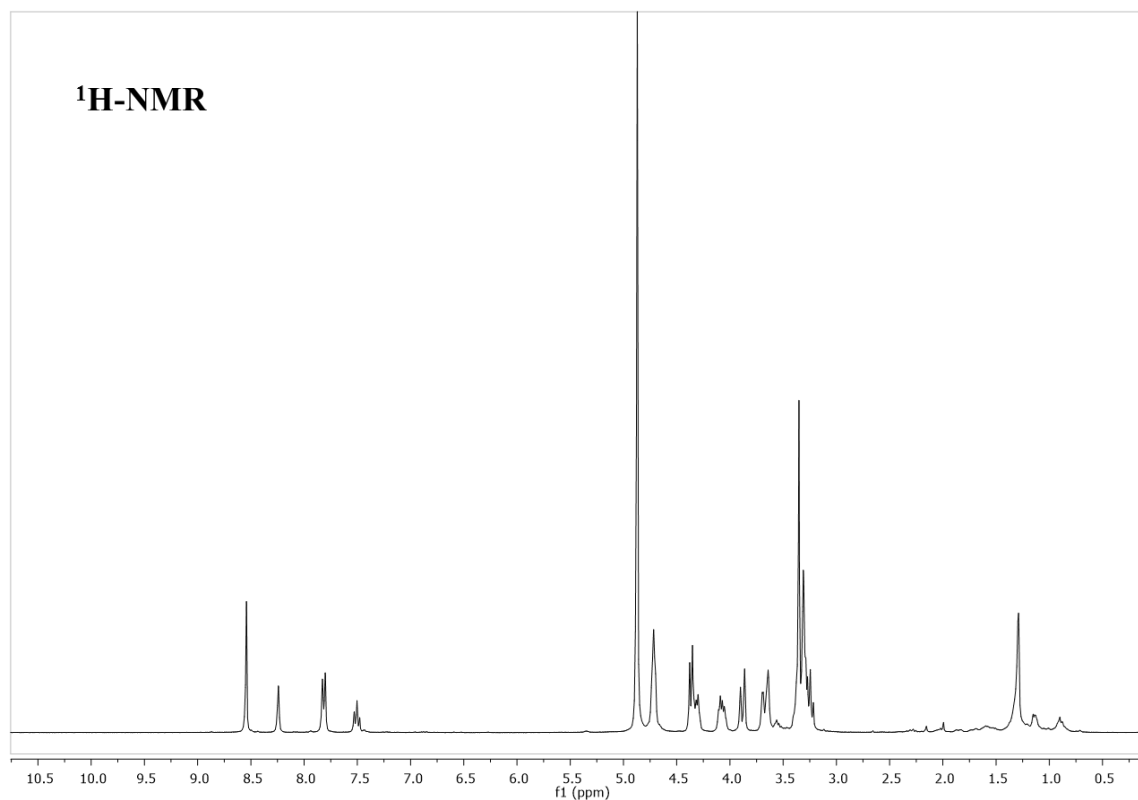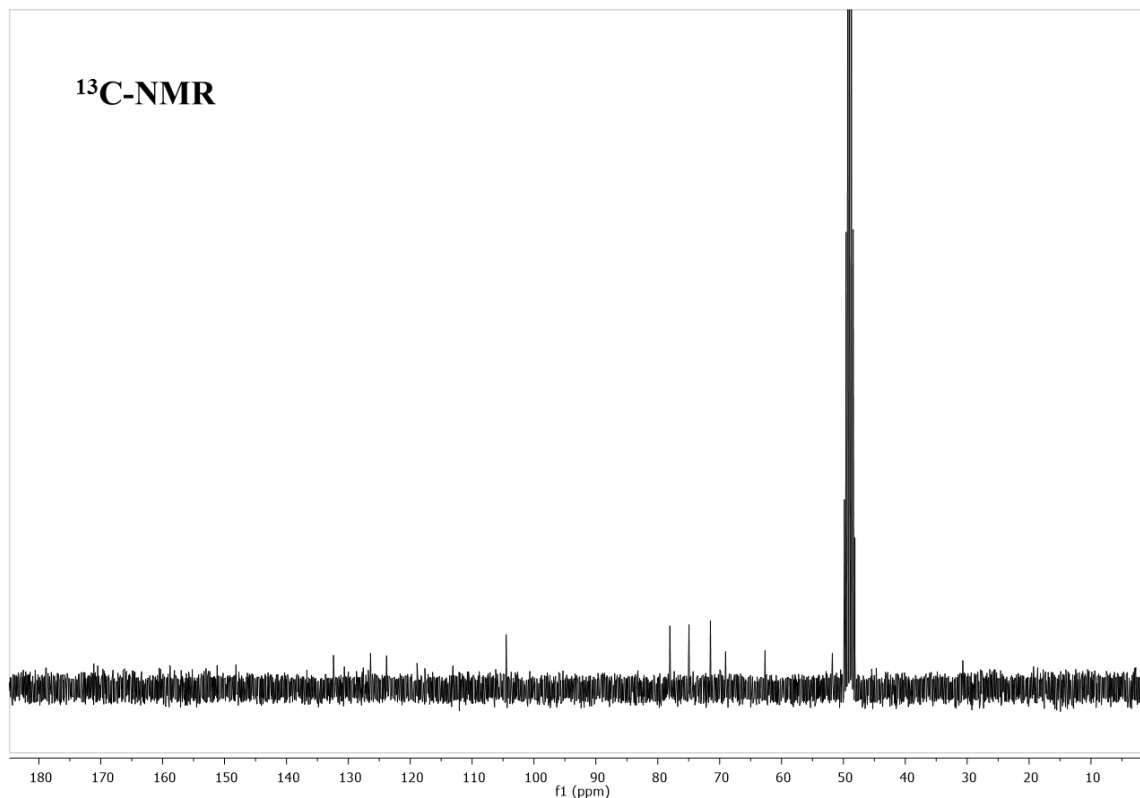
5.1.4. SYNTHESIS OF ASYMMETRIC PHENYLDITRIAZOLE

General procedure for synthesis of the monotriazole derivatives via click chemistry

Inside a 2.5 mL microwave reactor, the corresponding peracetylated azido-sugar derivative (1 eq) was dissolved in 1.5 mL of THF:H$_2$O (1:1) mixture, and CuSO$_4$·5H$_2$O (0.1 eq), sodium ascorbate (0.4 eq) and 1,3-diethynylbenzene or 1,4- diethynylbenzene (1 eq) were added. This solution was stirred and heated in the microwave reactor at 60°C, and the reaction completion was monitored by TLC each 15 min. When the 1,3- or 1-4-diethynylbenzene was consumed with more proportion of the monotriazole than the bitriazole derivative, the reaction was quenched by adding Quadrasil® MP to chelate the copper salt. The mixture was filtered and the solvents were evaporated at reduced pressure. The crude was purified by silica gel column chromatography.

Synthesis of 1-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl)-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**32**):
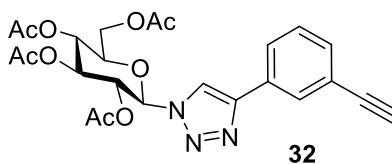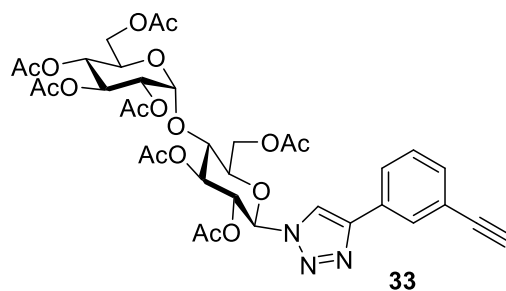
The general procedure was applied to the derivative 2,3,4,6-tetra-*O*-acetyl-1-azido-1-deoxy-β-D-glucopyranoside **3** (672 mg, 1.8 mmol, 2 eq) and 1,3-diethynylbenzene. The reaction was stopped after 65 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 1:1) to give **32** (224 mg, 50%) as a white solid. ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.04 (s, 1H, H$_{AR}$), 7.92 (s, 1H, H$_{TRIAZOLE}$), 7.81 (d, J=7.7 Hz, 1H, H$_{AR}$), 7.44 (d, J=7.7 Hz, 1H, H$_{AR}$), 7.35 (t, J=7.7 Hz, 1H, H$_{AR}$), 5.94 (d, J=8.7 Hz, 1H, H$_1$), 5.47 (p, J=9.3 Hz, 2H, H$_2$, H$_3$), 5.25 (t, J=9.4 Hz, 1H, H$_4$), 4.31 (dd, J=12.6, 5.0 Hz, 1H, H$_6$), 4.16–4.09 (m, 1H, H$_{6'}$), 4.07–3.99 (m, 1H, H$_5$), 3.11 (s, 1H, C≡CH), 2.03 (d, J=11.7 Hz, 9H, 9xCOCH$_3$), 1.86 (s, 3H, 3xCOCH$_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.54 (COCH$_3$), 169.96 (COCH$_3$), 169.44 (COCH$_3$), 169.05 (COCH$_3$), 147.57 (C$_{TRIAZOLE}$), 132.11 (CH$_{AR}$), 130.23 (C$_{AR}$), 129.54 (CH$_{AR}$), 129.01 (CH$_{AR}$), 126.28 (CH$_{AR}$), 122.83 (C$_{AR}$), 118.23 (CH$_{TRIAZOLE}$), 85.85 (C$_1$), 83.19 (C≡CH), 77.90 (C≡CH), 75.19 (C$_5$), 72.70 (C$_2$), 70.32 (C$_3$), 67.77 (C$_4$), 61.63 (C$_6$), 20.74 (COCH$_3$), 20.60 (COCH$_3$), 20.58 (COCH$_3$), 20.23 (COCH$_3$) HRMS (FAB$^+$), calculated for C$_{24}$H$_{25}$N$_3$O$_9$ (M): 499.1591; found: 499.1589.
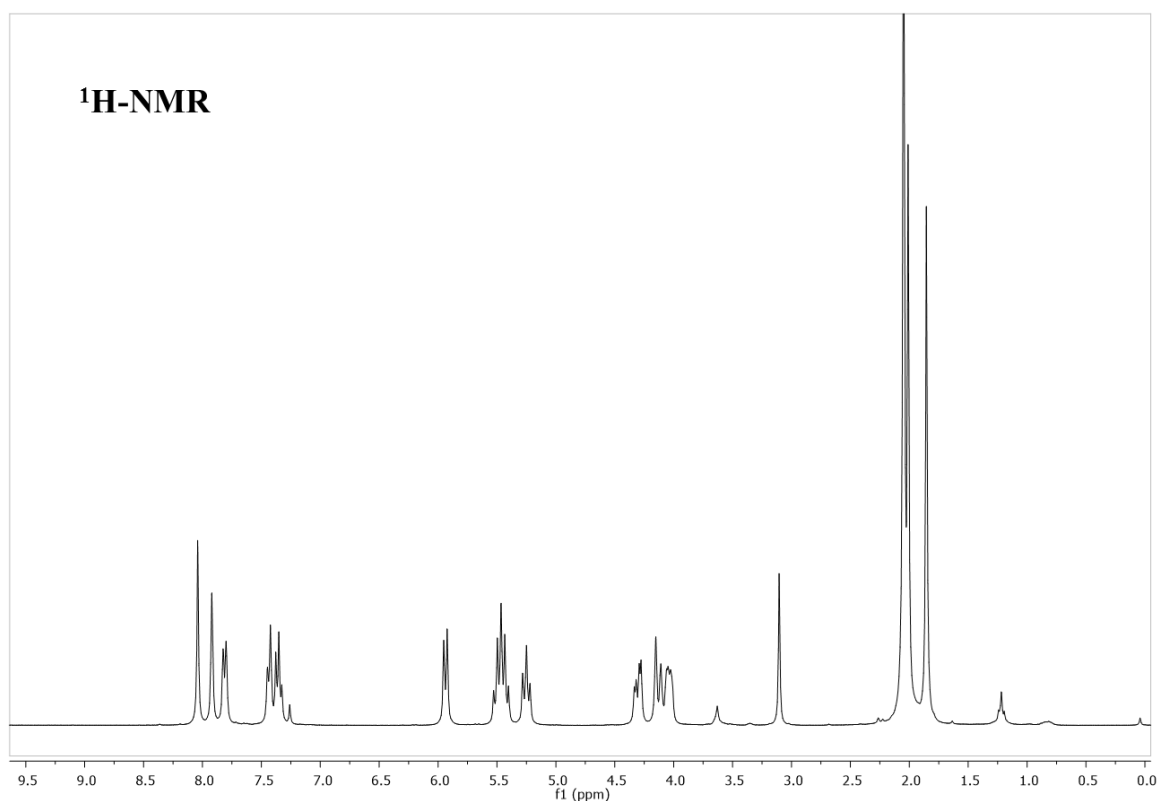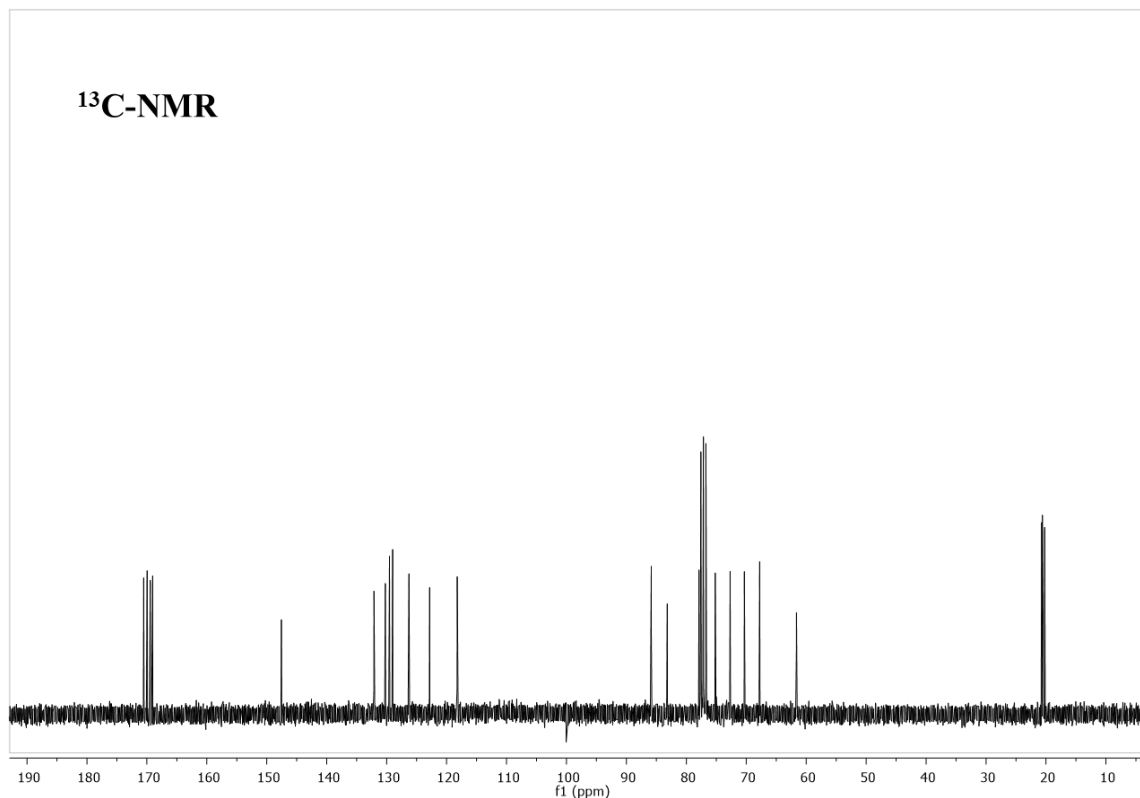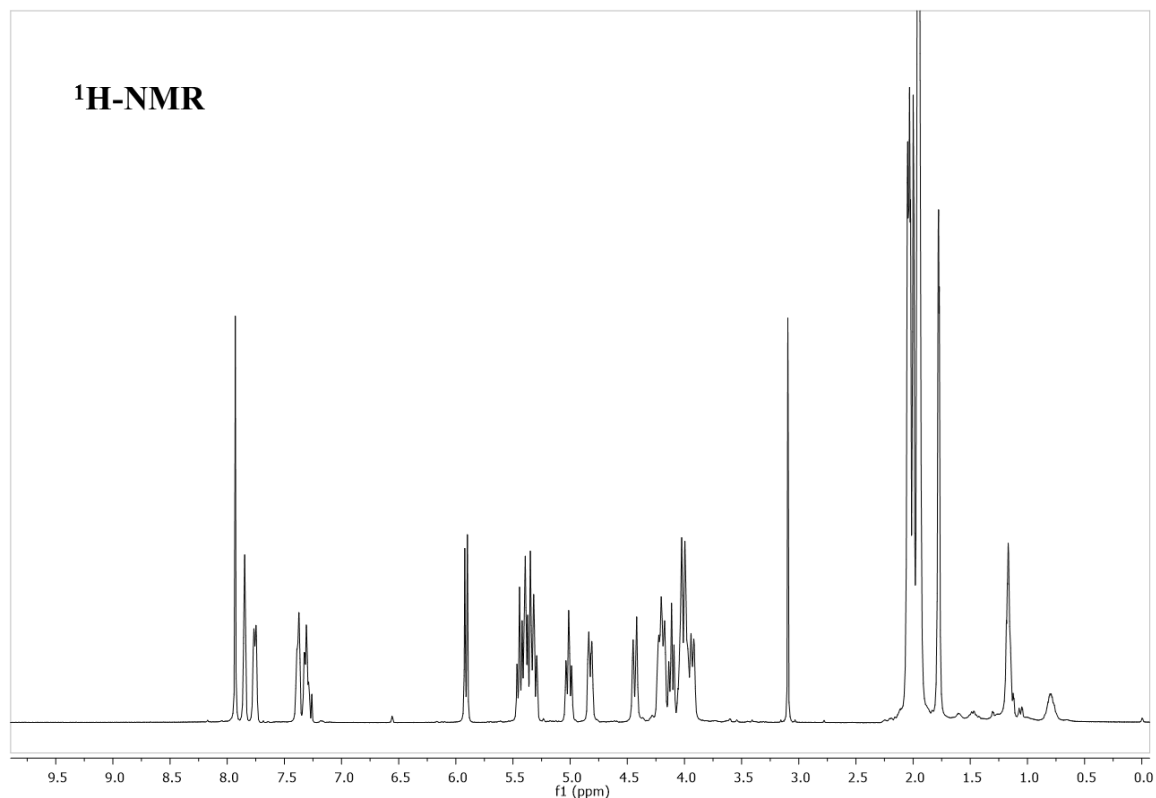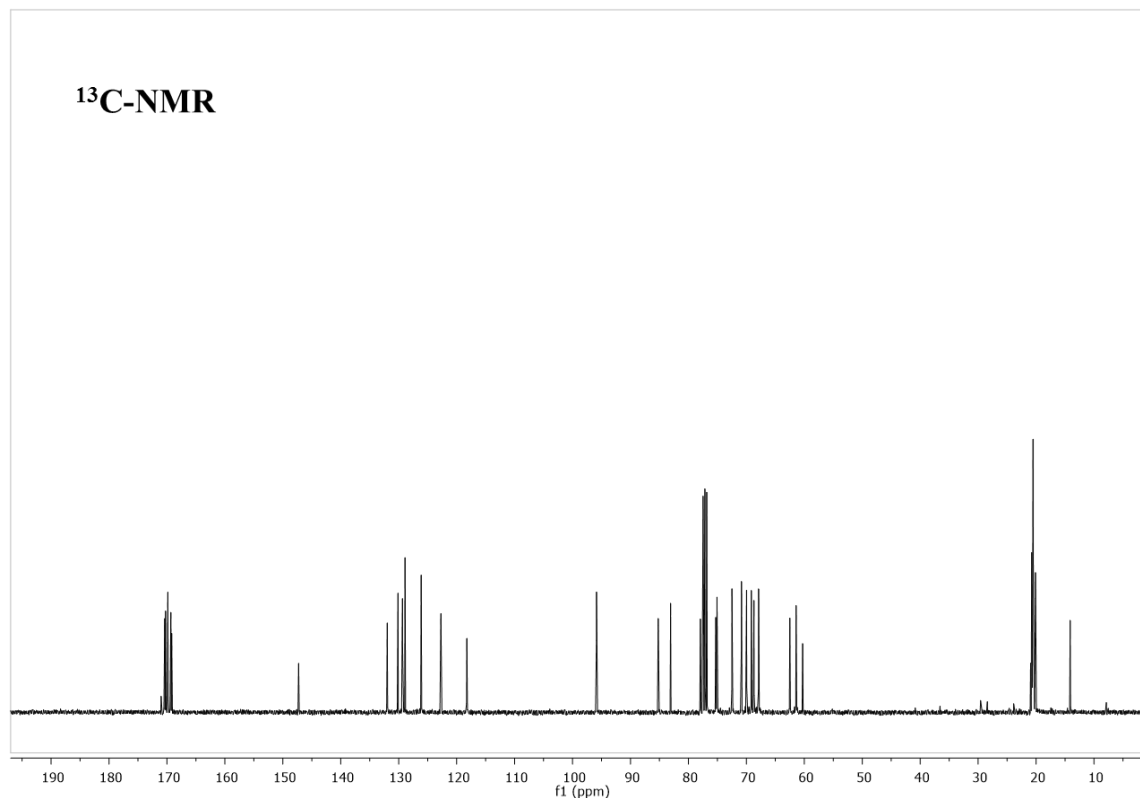

^{1}H-NMR

Synthesis of 1-(2,3,6-tri-*O*-acetyl-4-*O*-(2',3',4',6'-tetra-*O*-acetyl-α-D-glucopyranosyl)-β-D-glucopyranosyl)-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**33**):
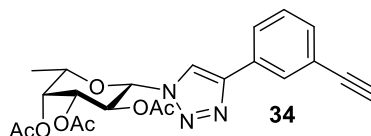

The general procedure was applied to the derivative 2,3,6-tri-*O*-acetyl-4-*O*-(2',3',4',6'-tetra-*O*-acetyl-α-D-glucopyranosyl)-1-azido-1-deoxy-β-D-glucopyranoside **6** (508 mg, 0.77 mmol, 3 eq) and 1,3-diethynylbenzene. The reaction was stopped after 30 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (5:1 to 1:1) to give **33** (204 mg, 80%) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.93 (s, 1H, H$_{TRIAZOLE}$), 7.86–7.83 (m, 1H, H$_{AR}$), 7.78–7.73 (m, 1H, H$_{AR}$), 7.38 (dd, J=7.8, 3.5 Hz, 1H, H$_{AR}$), 7.30 (td, J=7.6, 3.3 Hz, 1H, H$_{AR}$), 5.91 (d, J=9.2 Hz, 1H, H$_{1A}$), 5.48–5.28 (m, 4H, H$_{1B}$, H$_{2A}$, H$_{3A}$, H$_{3B}$), 5.01 (td, J=9.9, 2.7 Hz, 1H, H$_{4A}$), 4.82 (dt, J=10.9, 3.1 Hz, 1H, H$_{2B}$), 4.44 (dd, J=12.6, 2.5 Hz, 1H, H$_{6B}$), 4.20 (dq, J=12.1, 3.6, 3.0 Hz, 2H, H$_{5B}$, H$_{6'B}$), 4.11 (t, J=9.2 Hz, 1H, H$_{6A}$), 4.06–3.90 (m, 3H, H$_{4B}$, H$_{5A}$, H$_{6'A}$), 3.10 (s, 1H, C≡CH), 2.04 (dd, J=7.5, 3.8 Hz, 6H, 6xCOCH$_3$), 2.00 (d, J=3.2 Hz, 3H, 6xCOCH$_3$),

1.95 (q, J=2.8, 2.4 Hz, 9H, 9xCOC$\underline{H}_3$), 1.79–1.76 (m, 3H, 3xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.46 ($\underline{C}$OCH$_3$), 170.39 ($\underline{C}$OCH$_3$), 170.23 ($\underline{C}$OCH$_3$), 169.85 ($\underline{C}$OCH$_3$), 169.81 ($\underline{C}$OCH$_3$), 169.33 ($\underline{C}$OCH$_3$), 169.18 ($\underline{C}$OCH$_3$), 147.27 (C$_{TRIAZOLE}$), 131.98 (CH$_{AR}$), 130.13 (C$_{AR}$), 129.35 (CH$_{AR}$), 128.91 (CH$_{AR}$), 126.12 (CH$_{AR}$), 122.72 (C$_{AR}$), 118.24 (CH$_{TRIAZOLE}$), 95.86 (C$_{1B}$), 85.21 (C$_{1A}$), 83.08 ($\underline{C}$≡CH), 77.94 (C≡$\underline{C}$H), 75.28 (C$_{4B}$), 75.07 (C$_{2A}$), 72.49 (C$_{5B}$), 70.85 (C$_{3A}$), 69.97 (C$_{3B}$), 69.14 (C$_{2B}$), 68.70, 67.89 (C$_{4A}$, C$_{5A}$), 62.50 (C$_{6B}$), 61.42 (C$_{6A}$), 20.95 (CO$\underline{C}$H$_3$), 20.76 (CO$\underline{C}$H$_3$), 20.69 (CO$\underline{C}$H$_3$), 20.61 (CO$\underline{C}$H$_3$), 20.52 (CO$\underline{C}$H$_3$), 20.49 (CO$\underline{C}$H$_3$), 20.11 (CO$\underline{C}$H$_3$). LRMS (ES$^+$) calculated for C$_{36}$H$_{41}$N$_3$NaO$_{17}$ (M+Na): 810.2; found 810.5. HRMS (FAB$^+$), calculated for C$_{36}$H$_{41}$N$_3$NaO$_{17}$ (M+Na): 810.2334; found: 810.2303.
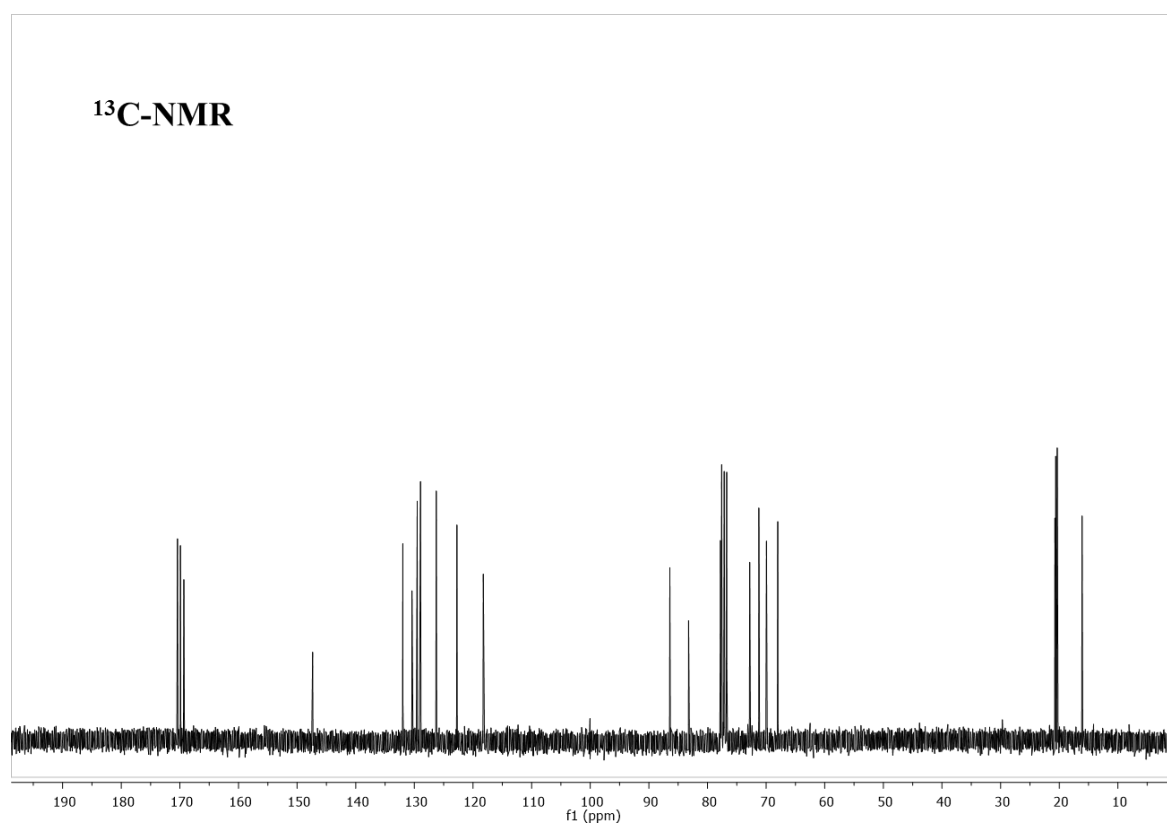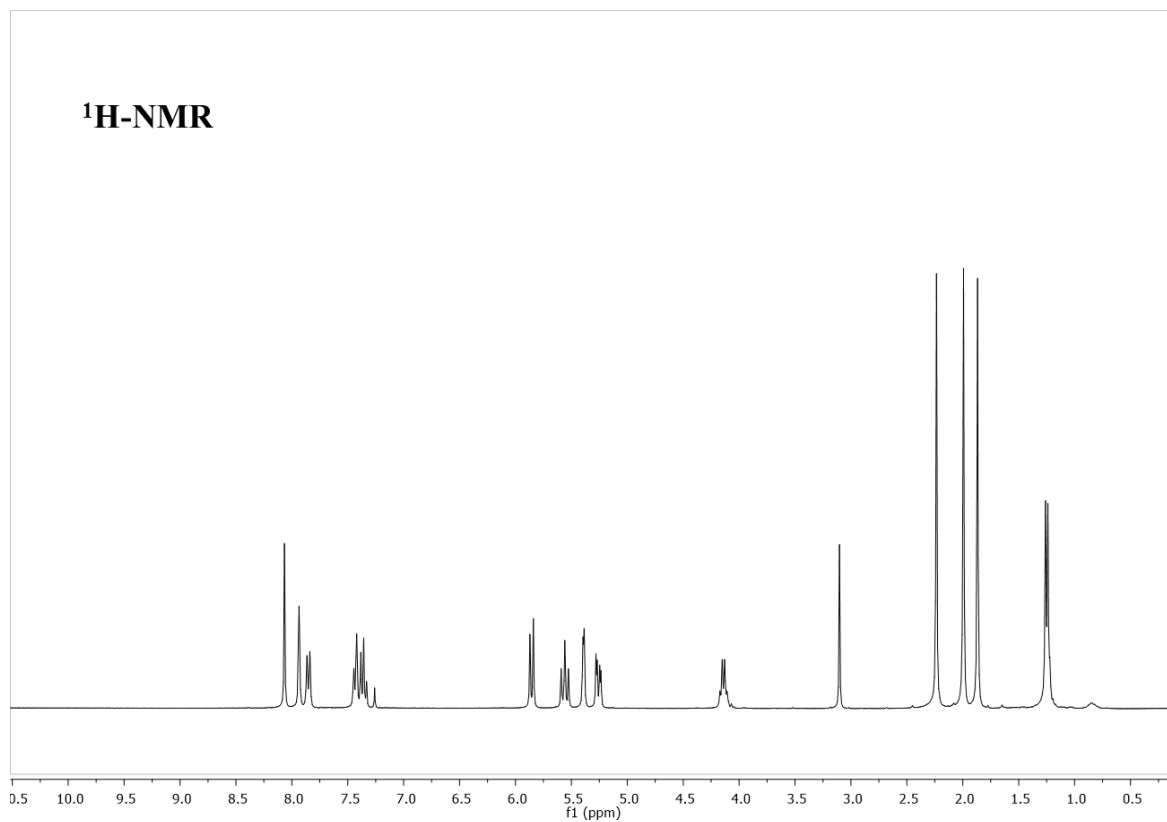
Synthesis of 1-(2,3,4-tri-*O*-acetyl-β-D-fucopyranosyl)-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**34**)
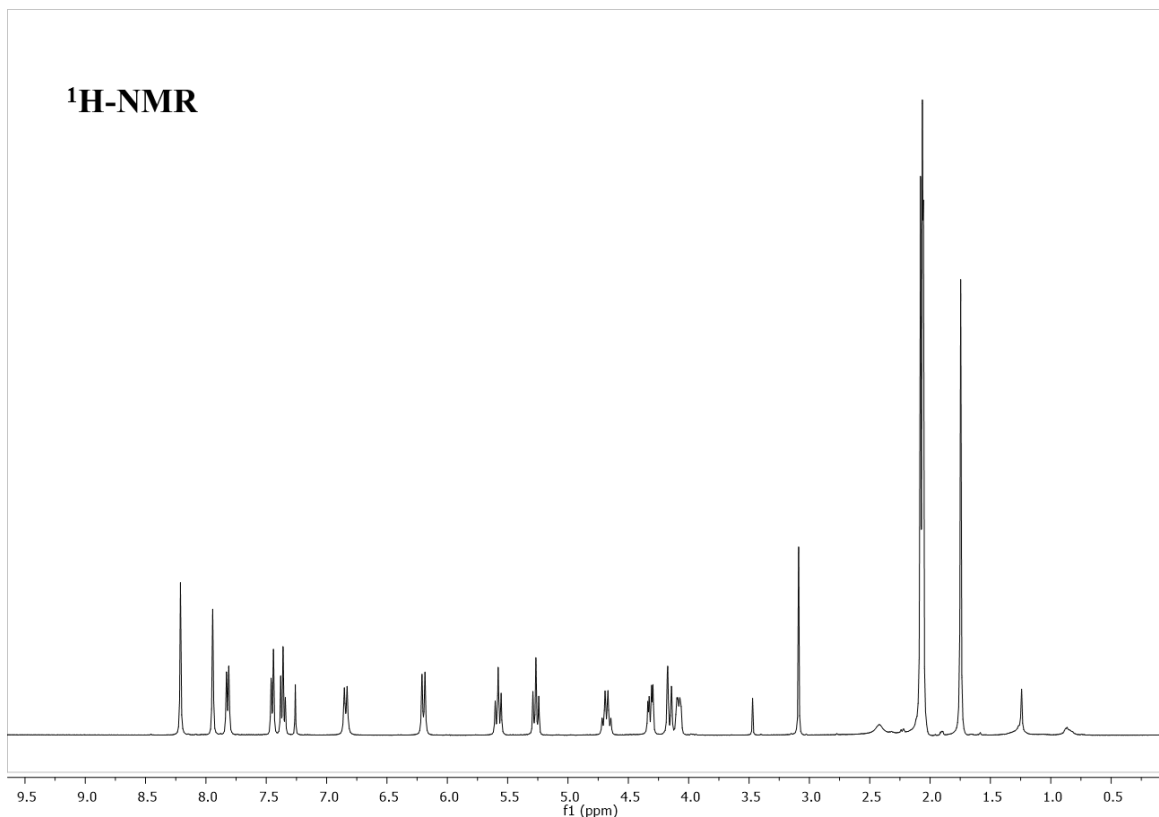

The general procedure was applied to the derivative 2,3,4-tri-*O*-acetyl-1-azido-1deoxy-β-D-fucopyranoside **4** (200 mg, 0.63 mmol, 1 eq) and 1,3-diethynylbenzene. The reaction was stopped after 15 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (4:1 to 2:1) to give **34** (149 mg, 53%) as a white solid. ^{1}H NMR (300 MHz, CDCl$_3$) δ 8.07 (s, 1H, H$_{TRIAZOLE}$), 7.95–7.92 (m, 1H, H$_{AR}$), 7.85 (dt, J=7.8, 1.6 Hz, 1H, H$_{AR}$), 7.46–7.41 (m, 1H, H$_{AR}$), 7.36 (t, J=7.7 Hz, 1H, H$_{AR}$), 5.86 (d, J=9.3 Hz, 1H, H$_1$), 5.56 (t, J=9.7 Hz, 1H, H$_2$), 5.41–5.37 (m, 1H, H$_4$), 5.26 (dd, J=10.2, 3.4 Hz, 1H, H$_3$), 4.19–4.10 (m, 1H, H$_5$), 3.10 (s, 1H,C≡CH̲), 2.24 (s, 3H, 3xCOCH̲$_3$), 1.99 (s, 3H, 3xCOCH̲$_3$), 1.87 (s, 3H, 3xCOCH̲$_3$), 1.25 (d, J=6.3 Hz, 3H, 3xCCH̲$_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.41 (C̲OCH$_3$), 169.92 (C̲OCH$_3$), 169.32 (C̲OCH$_3$), 147.36 (C$_{TRIAZOLE}$), 131.97 (CH$_{AR}$), 130.41 (C$_{AR}$), 129.52 (CH$_{AR}$), 128.96 (CH$_{AR}$), 126.27 (CH$_{AR}$), 122.75 (C$_{AR}$), 118.25 (CH$_{TRIAZOLE}$), 86.42 (C$_1$), 83.24 (C̲≡CH), 77.80 (C≡C̲H), 72.81 (C$_5$), 71.24 (C$_3$), 69.95 (C$_4$), 68.02 (C$_2$), 20.74 (COC̲H$_3$), 20.61 (COC̲H$_3$), 20.35 (COC̲H$_3$), 16.12 (CC̲H$_3$). LRMS (ES$^+$) calculated for C$_{22}$H$_{23}$N$_3$NaO$_7$ (M+Na): 464.1; found 464.2. HRMS (FAB$^+$), calculated for C$_{22}$H$_{24}$N$_3$O$_7$ (M+H): 442.1614; found: 442.1572.
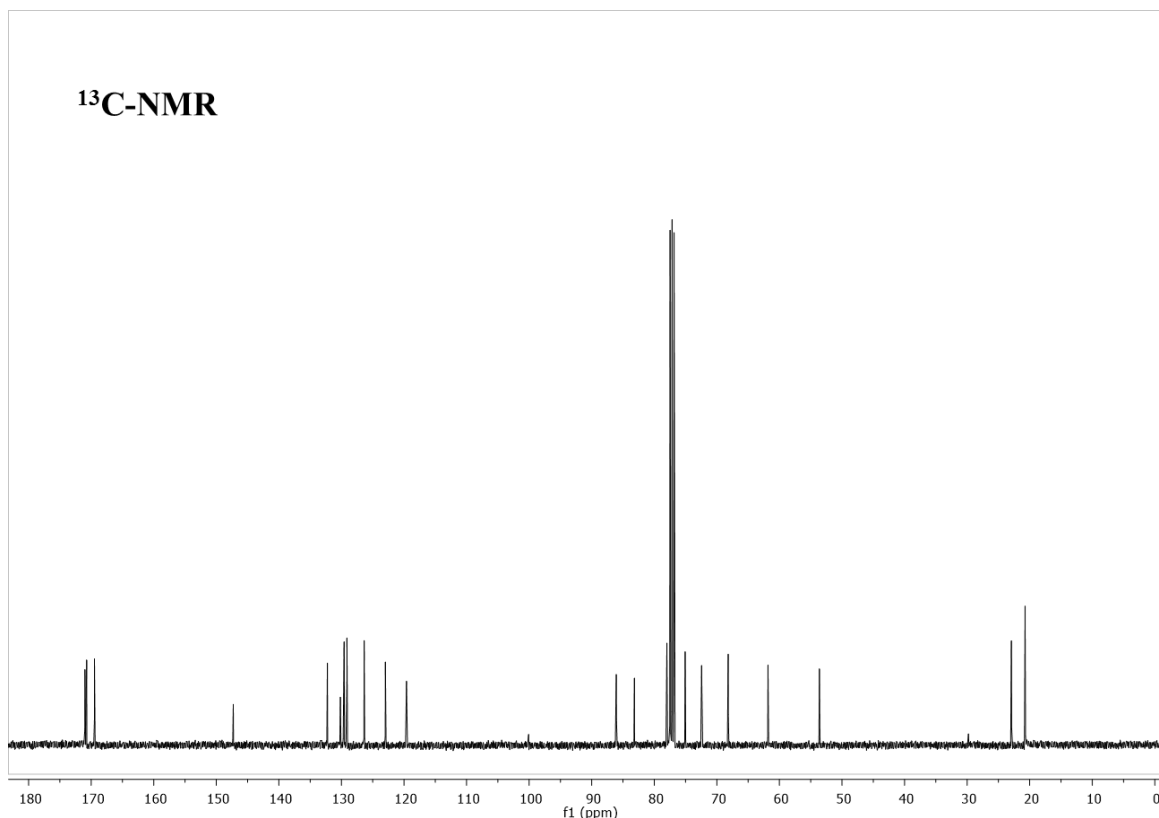
¹H-NMR

¹³C-NMR

<u>Synthesis of 1-(2-acetamide-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranosyl)-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**35**):</u>
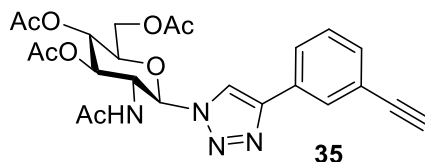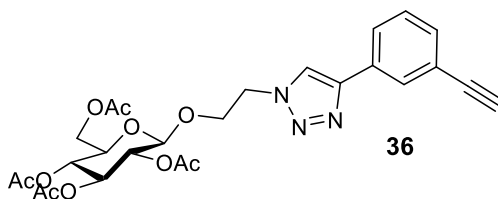

The general procedure was applied to the derivative 2-acetamide-3,4,6-tri-*O*-acetyl-1-azido-1,2-dideoxy-β-*D*-glucopyranoside **8** (200 mg, 0.54 mmol, 1 eq) and 1,3-diethynylbenzene. The reaction ended after 15 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 1:3) to give **35** (116 mg, 43%) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.21 (s, 1H, H$_{TRIAZOLE}$), 7.94 (s, 1H, H$_{AR}$), 7.82 (d, J=7.7 Hz, 1H, H$_{AR}$), 7.45 (d, J=7.6 Hz, 1H, H$_{AR}$), 7.36 (t, J=7.7 Hz, 1H, H$_{AR}$), 6.84 (d, J=9.2 Hz, 1H, NHCO), 6.20 (d, J=9.8 Hz, 1H, H$_1$), 5.58 (t, J=9.9 Hz, 1H, H$_3$), 5.27 (t, J=9.7 Hz, 1H, H$_4$), 4.68 (q, J=9.8 Hz, 1H, H$_2$), 4.32 (dd, J=12.7, 4.6 Hz, 1H, H$_{6A}$), 4.16 (dd, J=12.4, 2.3 Hz, 1H, H$_{6'}$), 4.08 (dq, J=10.1, 2.0 Hz, 1H, H$_5$), 3.09 (s, 1H, C≡CH), 2.11–2.03 (m, 9H, 9xCOCH$_3$), 1.75 (s, 3H, 3xNHOCH$_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 171.00 (COCH$_3$), 170.81 (COCH$_3$), 170.73 (COCH$_3$), 169.48 (COCH$_3$), 147.30 (C$_{TRIAZOLE}$), 132.25 (CH$_{AR}$), 130.17 (C$_{AR}$), 129.58 (CH$_{AR}$), 129.13 (CH$_{AR}$), 126.37 (CH$_{AR}$), 122.98 (C$_{AR}$), 119.61 (CH$_{TRIAZOLE}$), 86.08 (C$_1$), 83.18 (C≡CH), 77.99 (C≡CH), 75.08 (C$_5$), 72.44 (C$_3$), 68.19 (C$_4$), 61.83 (C$_6$), 53.62 (C$_2$), 22.93 (COCH$_3$), 20.81 (COCH$_3$), 20.78 (COCH$_3$), 20.74 (COCH$_3$). LRMS (ES$^+$) calculated for C$_{24}$H$_{26}$N$_4$NaO$_8$ (M+Na): 521.2, found: 521.3. HRMS (FAB$^+$), calculated for C$_{24}$H$_{27}$N$_4$O$_8$ (M): 498.1751; found: 498.1738.
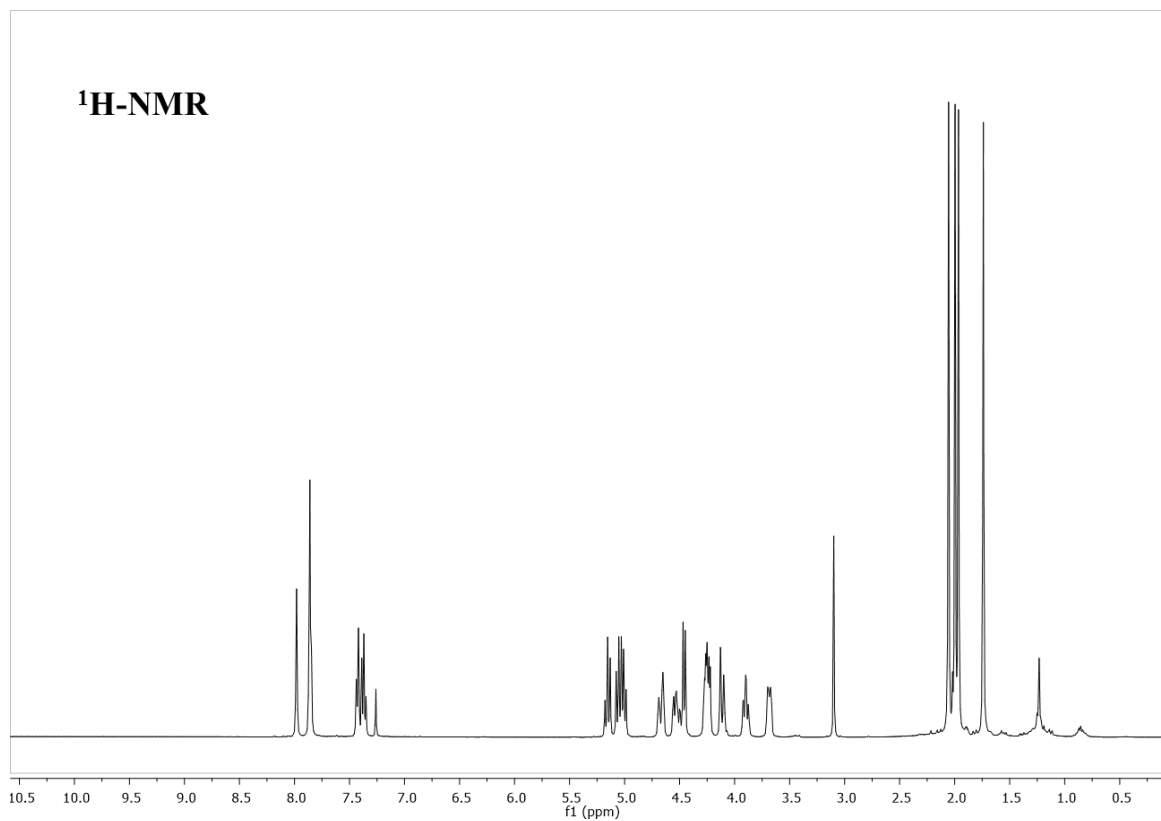

^{1}H-NMR

Synthesis of 1-[2-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl) ethyl]-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**36**):

The general procedure was applied to the derivative 2-azidoethyl 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranoside **14** (200 mg, 0.48 mmol, 1 eq) and 1,3-diethynylbenzene. The reaction was completed after 25 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 2:3) to give **36** (130 mg, 50%) as a white solid. ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.98 (s, 1H, H$_{TRIAZOLE}$), 7.85 (d, J=5.8 Hz, 2H, 2xH$_{AR}$), 7.43 (d, J=7.6 Hz, 1H, H$_{AR}$), 7.37 (t, J=7.6 Hz, 1H, H$_{AR}$), 5.15 (t, J=9.5 Hz, 1H, H$_3$), 5.10–4.97 (m, 2H, H$_2$, H$_4$), 4.67 (dt, J=14.4, 3.5 Hz, 1H, C$\underline{H}_2$N), 4.59–4.48 (m, 1H, C$\underline{H}_2$N), 4.46 (d, J=7.9 Hz, 1H, H$_1$), 4.30–4.20 (m, 2H, H$_6$,OC$\underline{H}_2$), 4.11 (dd, J=12.3, 2.5 Hz, 1H, OC$\underline{H}_2$), 3.90 (ddd, J=11.5, 9.2, 2.9 Hz, 1H, H$_{6'}$), 3.68 (ddd, J=10.0, 4.5, 2.3 Hz, 1H, H$_5$), 3.10 (s, 1H, C≡C$\underline{H}$), 2.06 (s, 3H, 3xCOC$\underline{H}_3$), 1.98 (d, J=12.5 Hz, 6H, 6xCOC$\underline{H}_3$), 1.74 (s, 3H, 3xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 170.67 ($\underline{C}$OCH$_3$), 170.15 ($\underline{C}$OCH$_3$), 169.54 ($\underline{C}$OCH$_3$), 169.51 ($\underline{C}$OCH$_3$), 146.75 (C$_{AR}$), 131.79 (C$_{AR}$), 130.86 (CH$_{AR}$), 129.31 (CH$_{AR}$), 129.00 (CH$_{AR}$), 126.11 (CH$_{AR}$), 122.82 (C$_{AR}$), 121.79 (CH$_{AR}$), 100.60 (C$_1$), 83.33 (C≡CH), 77.74 (C≡CH), 72.48 (C$_5$), 72.07 (C$_3$), 71.01 (C$_2$), 68.29 (C$_4$), 67.85 (C$_6$), 61.76 (O$\underline{C}$H$_2$),

50.23 ($\underline{C}H_2$-N), 20.81 (CO$\underline{C}H_3$), 20.65 (CO$\underline{C}H_3$), 20.64 (CO$\underline{C}H_3$), 20.49 (CO$\underline{C}H_3$). LRMS (ES$^+$) calculated. for $C_{26}H_{29}N_3NaO_{10}$ (M+Na): 566.2; found: 566.3. HRMS (FAB$^+$) calculated for $C_{26}H_{29}N_3NaO_{10}$ (M+Na): 566.1751; found: 566.1724.
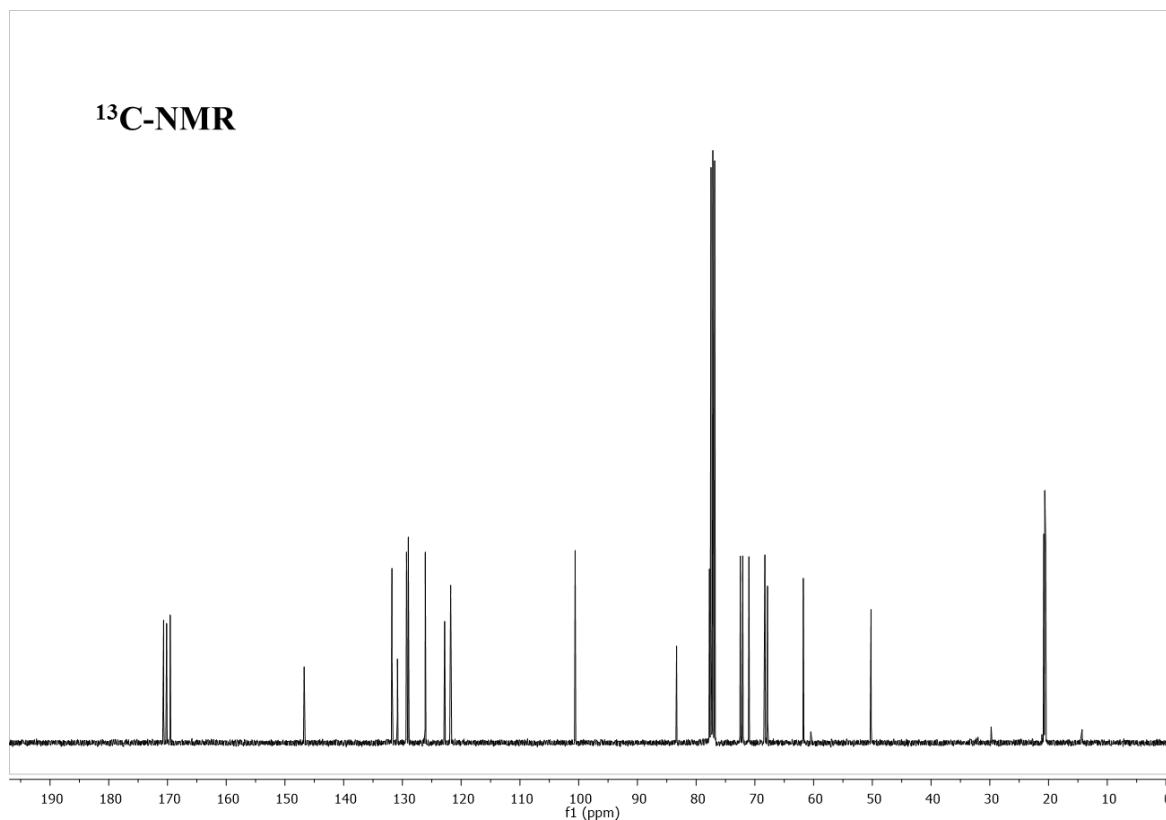
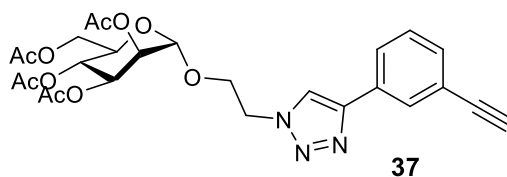Synthesis of 1-[2-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl) ethyl]-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**37**):

The general procedure was applied to the derivative 2-azido ethyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside **10** (600 mg, 1.44 mmol, 1 eq) and 1,3-diethynylbenzene. The reaction was finished after 15 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 0:1) to give **37** (322.7 mg, 41%) as a white solid. ^{1}H NMR (300 MHz, CDCl$_3$) δ 7.93 (d, J=3.8 Hz, 2H, H$_{TRIAZOLE}$, H$_{AR}$), 7.89–7.81 (m, 1H, H$_{AR}$), 7.34 (dt, J=15.2, 7.5 Hz, 2H, 2xH$_{AR}$), 5.27–5.05 (m, 3H, H$_2$, H$_3$, H$_4$), 4.79 (s, 1H, H$_1$), 4.61 (tdd, J=19.5, 12.5, 7.1 Hz, 2H, 2xCH$_2$N), 4.11-3.97 (m, J=10.4, 4.1 Hz, 3H, H$_5$, H$_6$, OCH$_2$), 3.93–3.77 (m, 2H, H$_6$', OCH$_2$), 3.08 (d, J=5.4 Hz, 1H, C≡CH), 2.11–1.90 (m, 9H, 9xCOCH$_3$), 1.68 (s, 3H, 3xCOCH$_3$) ^{13}C NMR (75 MHz, CDCl$_3$) δ 170.40 (COCH$_3$), 169.97 (COCH$_3$), 169.92 (COCH$_3$), 169.56 (COCH$_3$), 146.90 (C$_{AR}$), 131.63 (C$_{AR}$), 130.77 (CH$_{AR}$), 129.36 (CH$_{AR}$), 128.86 (CH$_{AR}$), 126.15 (CH$_{AR}$), 122.67 (C$_{AR}$), 121.52 (CH$_{AR}$), 96.77 (C$_1$), 83.25 (C≡CH), 77.74 (C≡CH), 69.15, 68.95, 68.82, 65.58 (C$_2$, C$_3$, C$_4$, C$_5$), 65.29 (OCH$_2$), 62.03 (C$_6$), 49.80 (CH$_2$N), 20.76 (COCH$_3$), 20.65 (COCH$_3$), 20.61

(CO$\underline{C}$H$_3$), 20.15 (CO$\underline{C}$H$_3$). HRMS (FAB$^+$) calculated for C$_{26}$H$_{30}$N$_3$O$_{10}$(M+H): 544.1931; found: 544.1929.

^{1}H-NMR
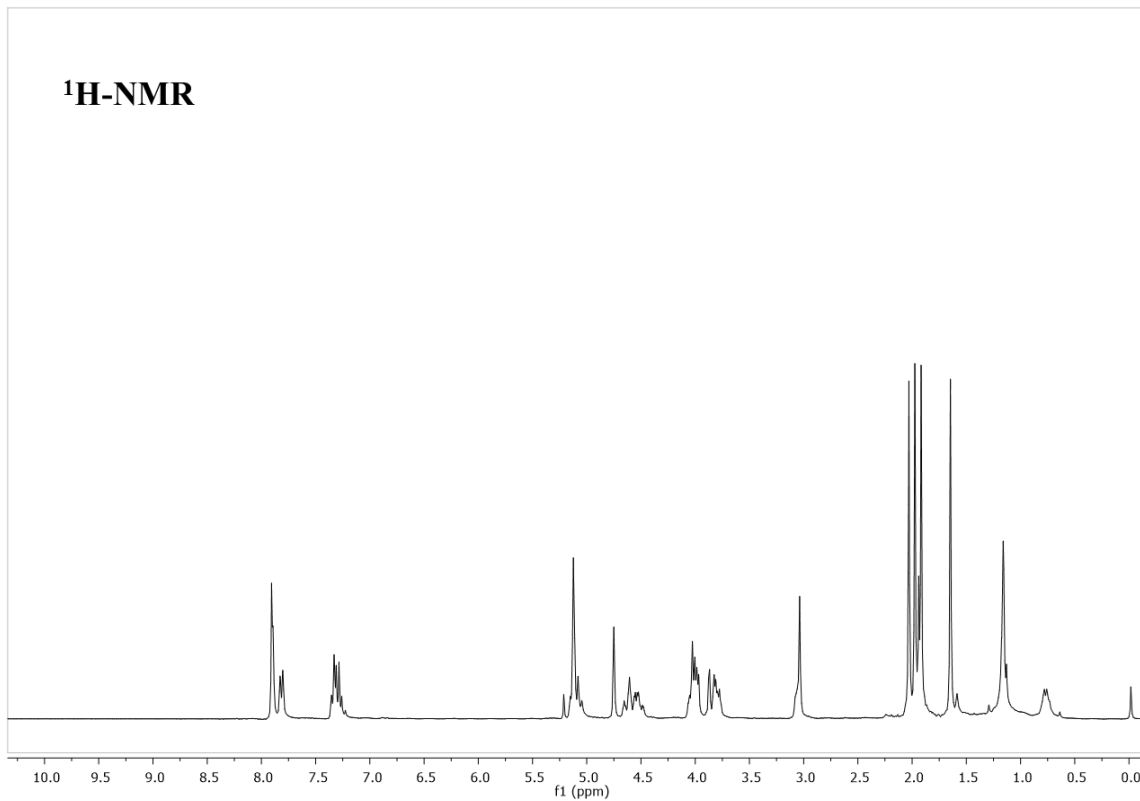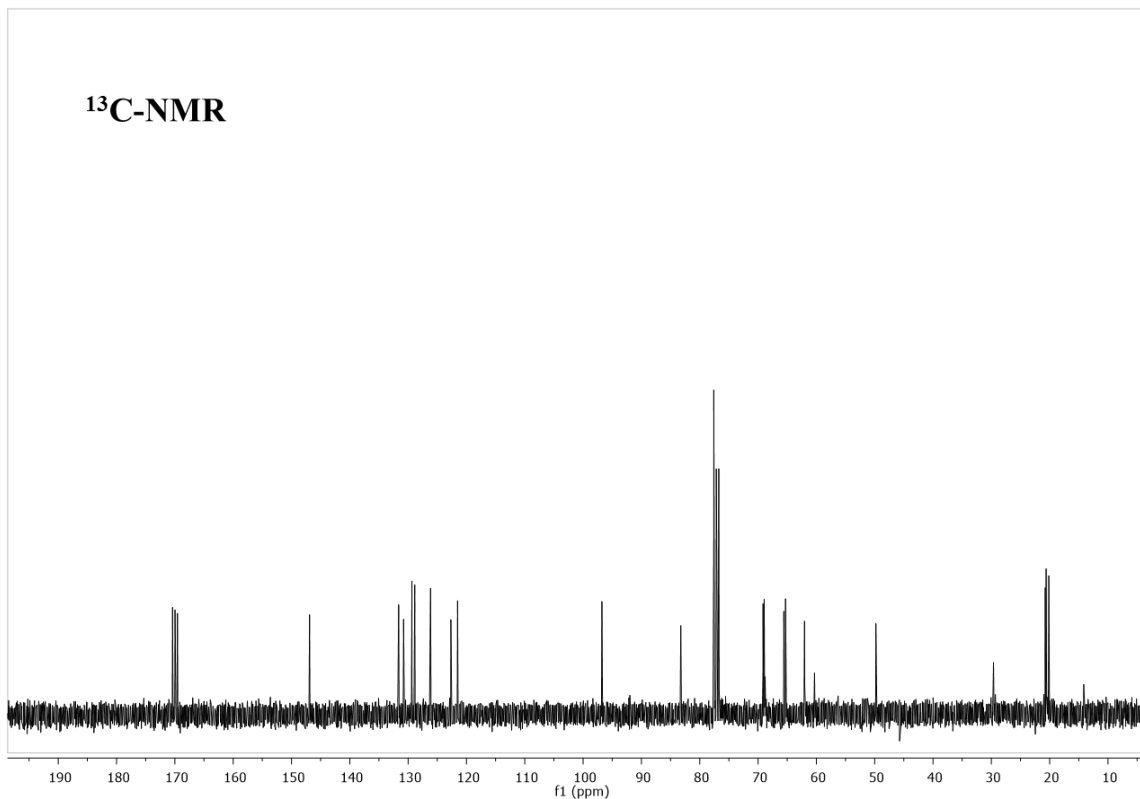
Synthesis of 1-(4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-2,3,6-tri-*O*-acetyl-β-D-glucopyranosyl)-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole (**38**):
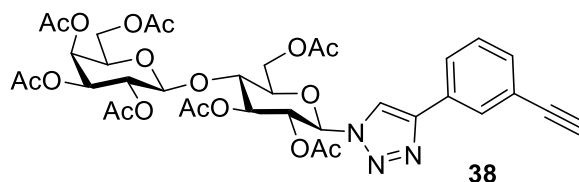
The general procedure was applied to the derivative 4-*O*-(2,3,4,6-tetra-*O*-acetyl-β-D-galactopyranosyl)-2,3,6-tri-*O*-acetyl-1-azido-1-deoxy-β-D-glucopyranose **5** (800 mg, 1.21 mmol, 1 eq) and 1,3-diethynylbenzene. The reaction was completed after 30 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 0:1) to give **38** (290.0 mg, 30%) as a white solid. ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.95 (s, 1H, H$_{TRIAZOLE}$), 7.92 (t, J=1.7 Hz, 1H, H$_{AR}$), 7.82 (dt, J=7.7, 1.5 Hz, 1H, H$_{AR}$), 7.45 (dt, J=7.7, 1.4 Hz, 1H, H$_{AR}$), 7.38 (td, J=7.8, 0.6 Hz, 1H, H$_{AR}$), 5.90–5.85 (m, 1H, H$_{1A}$), 5.49–5.39 (m, 2H, H$_{2A}$, H$_{3A}$), 5.36 (dd, J=3.4, 1.1 Hz, 1H, H$_{4B}$), 5.13 (dd, J=10.4, 7.9 Hz, 1H, H$_{2B}$), 4.97 (dd, J=10.4, 3.4 Hz, 1H, H$_{3B}$), 4.53 (d, J=7.9 Hz, 1H, H$_{1B}$), 4.48 (dd, J=12.3, 1.7 Hz, 1H, H$_{6B}$), 4.20–4.12 (m, 2H, H$_{6A}$, H$_{6B'}$), 4.09 (dd, J=11.2, 7.2 Hz, 1H, H$_{6A'}$), 4.00–3.88 (m, 3H, H$_{4A}$, H$_{5A}$,H$_{5B}$), 3.11 (s, 1H, C≡CH), 2.15 (s, 3H, 3xCOCH$_3$), 2.10 (s, 3H, 3xCOCH$_3$), 2.08–2.04 (m, 9H, 9xCOCH$_3$), 1.96 (s, 3H, 3xCOCH$_3$), 1.87 (s, 3H, 3xCOCH$_3$). ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.45(COCH$_3$), 170.31(COCH$_3$), 170.19(COCH$_3$), 170.14(COCH$_3$), 169.58(COCH$_3$), 169.34(COCH$_3$), 169.17(COCH$_3$), 147.52(C$_{AR}$), 132.18(CH$_{AR}$), 130.28(C$_{AR}$), 129.58(CH$_{AR}$), 129.06(CH$_{AR}$), 126.31(CH$_{AR}$), 122.90(C$_{AR}$), 118.23

(CH$_{TRIAZOLE}$), 101.21(C$_{1B}$), 85.73(C$_{1A}$), 83.22(C≡CH), 77.88(C≡CH), 76.11, 75.77, 72.73, 71.02, 70.97, 70.60(C$_{2A}$, C$_{3A}$, C$_{3B}$, C$_{4A}$, C$_{5A}$, C$_{5B}$), 69.17(C$_{2B}$), 66.72(C$_{4B}$), 61.91(C$_{6B}$), 60.95(C$_{6A}$), 20.89(COCH$_3$), 20.81(COCH$_3$), 20.77(COCH$_3$), 20.73(COCH$_3$), 20.73(COCH$_3$), 20.60(COCH$_3$), 20.34(COCH$_3$). HRMS (FAB$^+$) calculated for C$_{36}$H$_{42}$N$_3$O$_{17}$(M+H): 788.2514; found: 788.2505.
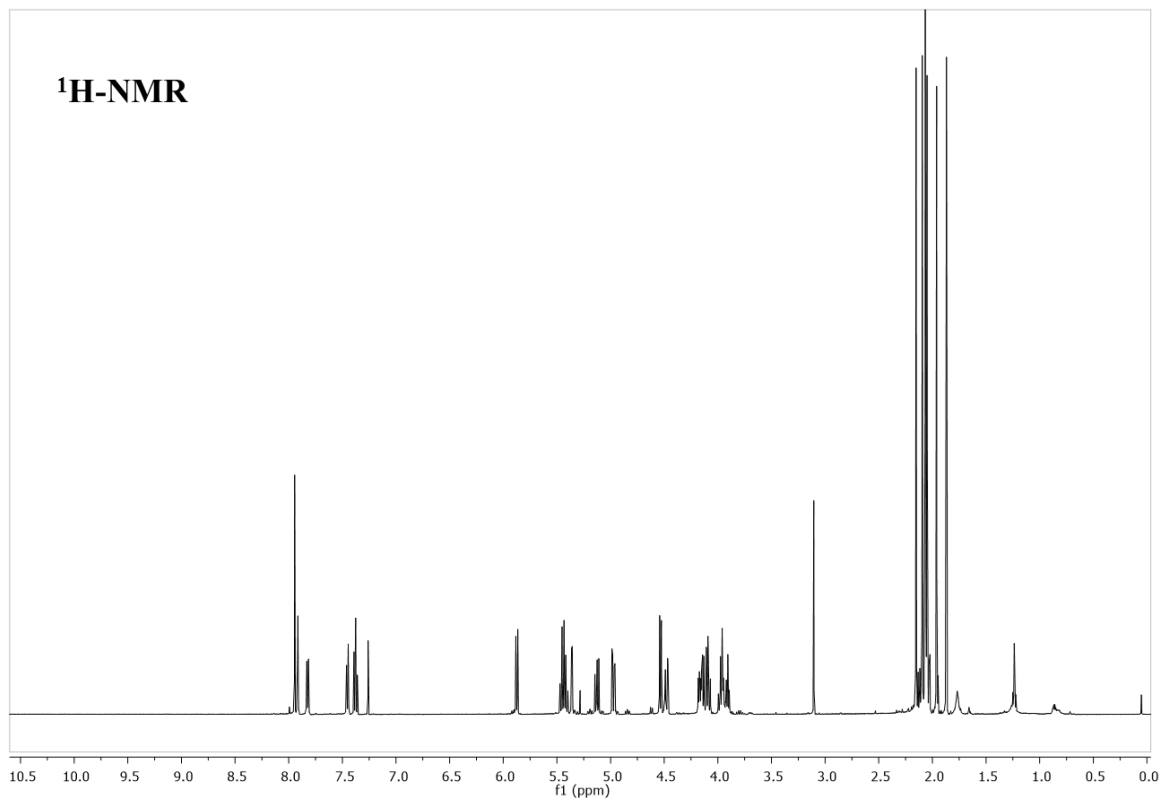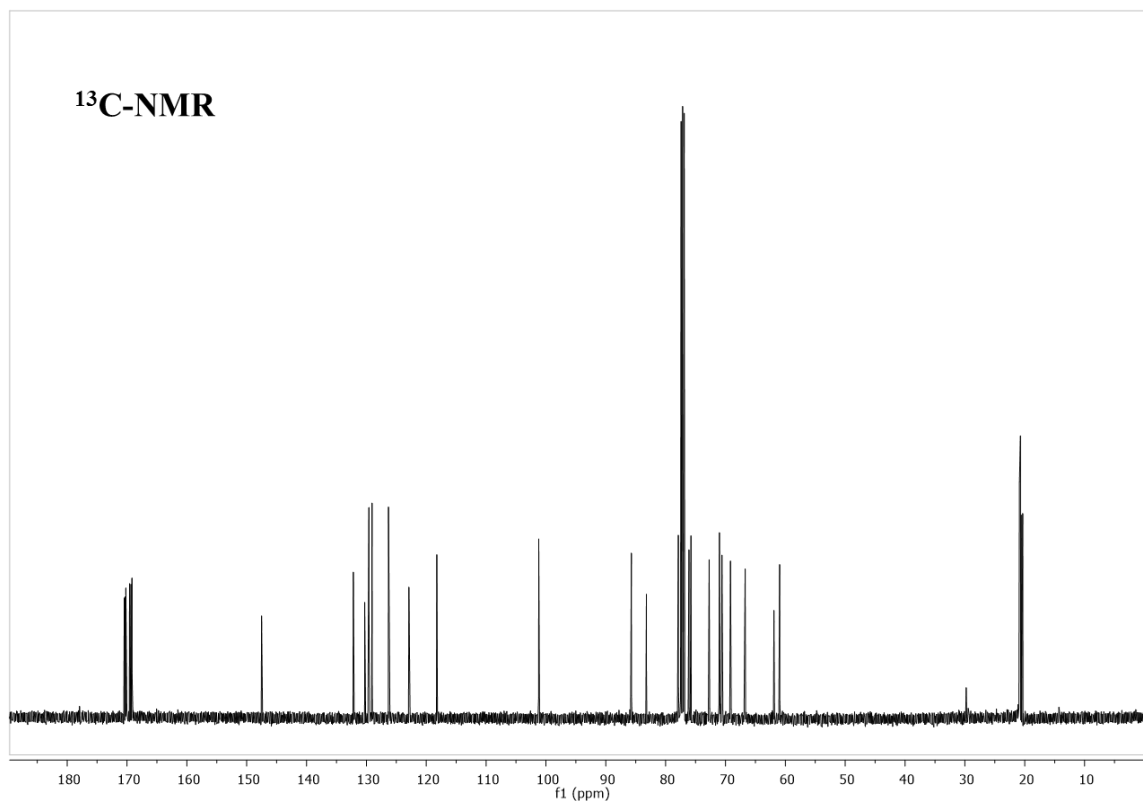
Synthesis of 1-(2,3,6-tri-*O*-acetyl-4-*O*-(2',3',4',6'-tetra-*O*-acetyl-α-D-glucopyranosyl)-β-D-glucopyranosyl)-4-(4-ethynylphenyl)-1*H*-1,2,3-triazole (**39**)
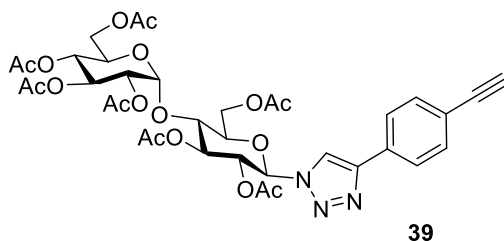
The general procedure was applied to the derivative 2,3,6-tri-*O*-acetyl-4-*O*-(2',3',4',6'-tetra-*O*-acetyl-α-D-glucopyranosyl)-1-azido-deoxy-β-D-glucopyranose **6** (610 mg, 0.92 mmol, 1 eq) and 1,4-diethynylbenzene. The reaction was quenched after 30 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (3:1 to 1:1) to give **39** (111 mg, 15%) as a white solid. ^{1}H NMR (500 MHz, CDCl$_3$) δ 7.94 (s, 1H, H$_{TRIAZOLE}$), 7.80–7.74 (m, 2H, 2xH$_{AR}$), 7.57–7.50 (m, 2H, H$_{AR}$), 5.93 (d, J=9.3 Hz, 1H, H$_{1A}$), 5.52–5.43 (m, 2H, H$_{1B}$, H$_{3A}$), 5.43–5.34 (m, 2H, H$_{2A}$, H$_{3B}$), 5.11–5.03 (m, 1H, H$_{4B}$), 4.88 (dd, J=10.5, 4.0 Hz, 1H, H$_{2B}$), 4.49 (dd, J=12.4, 2.4 Hz, 1H, H$_{6A}$), 4.30–4.21 (m, 2H, H$_{6B}$, H$_{6'A}$), 4.16 (dd, J=9.7, 8.6 Hz, 1H, H$_{4A}$), 4.12–4.05 (m, 1H, H$_{6'B}$), 3.99 (dddd, J=17.0, 10.4, 4.3, 2.5 Hz, 2H, H$_{5A}$,H$_{5B}$), 3.13 (s, 1H, C≡CH̲), 2.12 (s, 3H, 3xCOCH̲$_3$), 2.10 (s, 3H, 3xCOCH̲$_3$), 2.06 (s, 3H, 3xCOCH̲$_3$), 2.03 (d, J=2.2 Hz, 6H, 6xCOCH̲$_3$), 2.00 (s, 3H, 3xCOCH̲$_3$), 1.85 (s, 3H, 3xCOCH̲$_3$). ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.67 (C̲OCH$_3$), 170.59 (C̲OCH$_3$), 170.41 (C̲OCH$_3$), 170.01 (C̲OCH$_3$), 170.00 (C̲OCH$_3$), 169.51

($\underline{C}OCH_3$), 169.42 ($\underline{C}OCH_3$), 147.71 ($C_{TRIAZOLE}$), 132.77 (CH_{AR}), 130.31 (C_{AR}), 125.81 (CH_{AR}), 122.33 (C_{AR}), 118.27 ($\underline{C}H_{TRIAZOLE}$), 96.05 ($C_{1B}$), 85.47 ($C_{1A}$), 83.43 ($\underline{C}\equiv CH$), 78.33 ($C\equiv\underline{C}H$), 75.57 ($C_{5B}$), 75.28 ($C_{3A}$), 72.60 ($C_{4A}$), 70.99 ($C_{3B}$), 70.14 ($C_{2B}$), 69.33 ($C_{2A}$), 68.92 ($C_{5A}$), 68.06 ($C_{4B}$), 62.65 ($C_{6A}$), 61.58 ($C_{6B}$), 20.94 ($CO\underline{C}H_3$), 20.88 ($CO\underline{C}H_3$), 20.80 ($CO\underline{C}H_3$), 20.71 ($CO\underline{C}H_3$), 20.69 ($CO\underline{C}H_3$), 20.32 ($CO\underline{C}H_3$). HRMS ($FAB^+$) calculated for $C_{36}H_{41}N_3O_{17}$ (M+H): 788.2514; found: 788.2521.

^{13}C-NMR
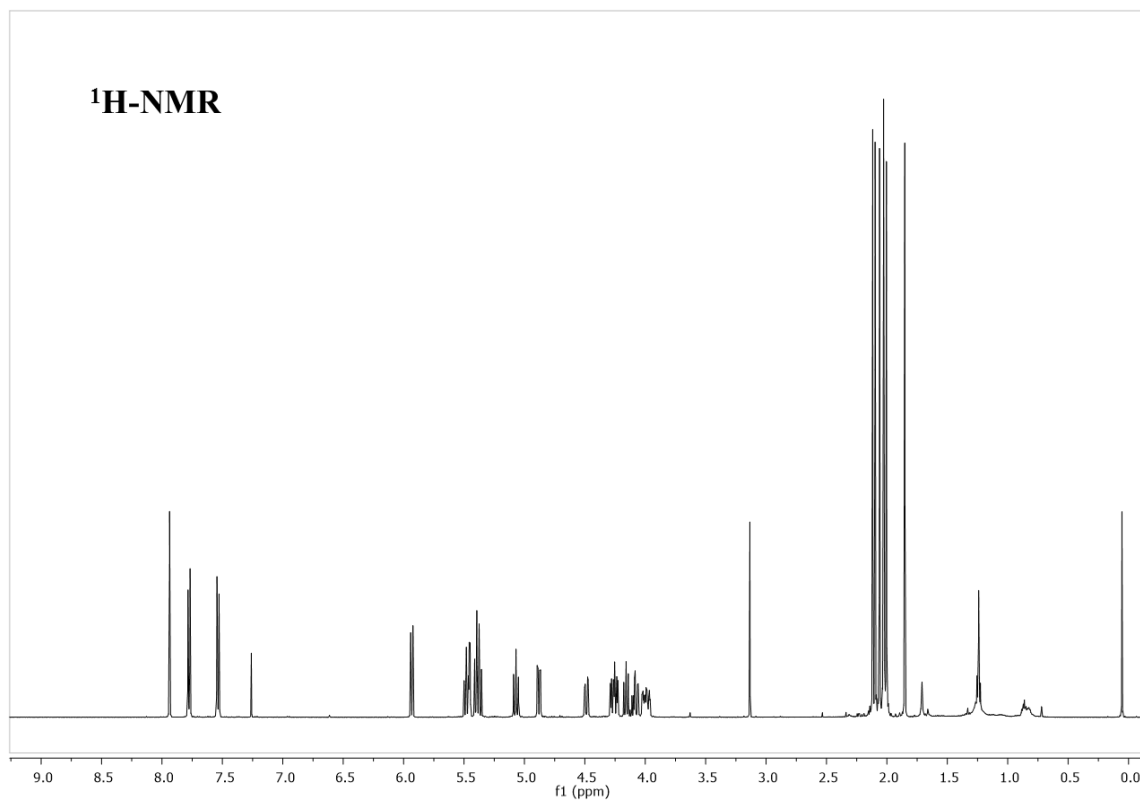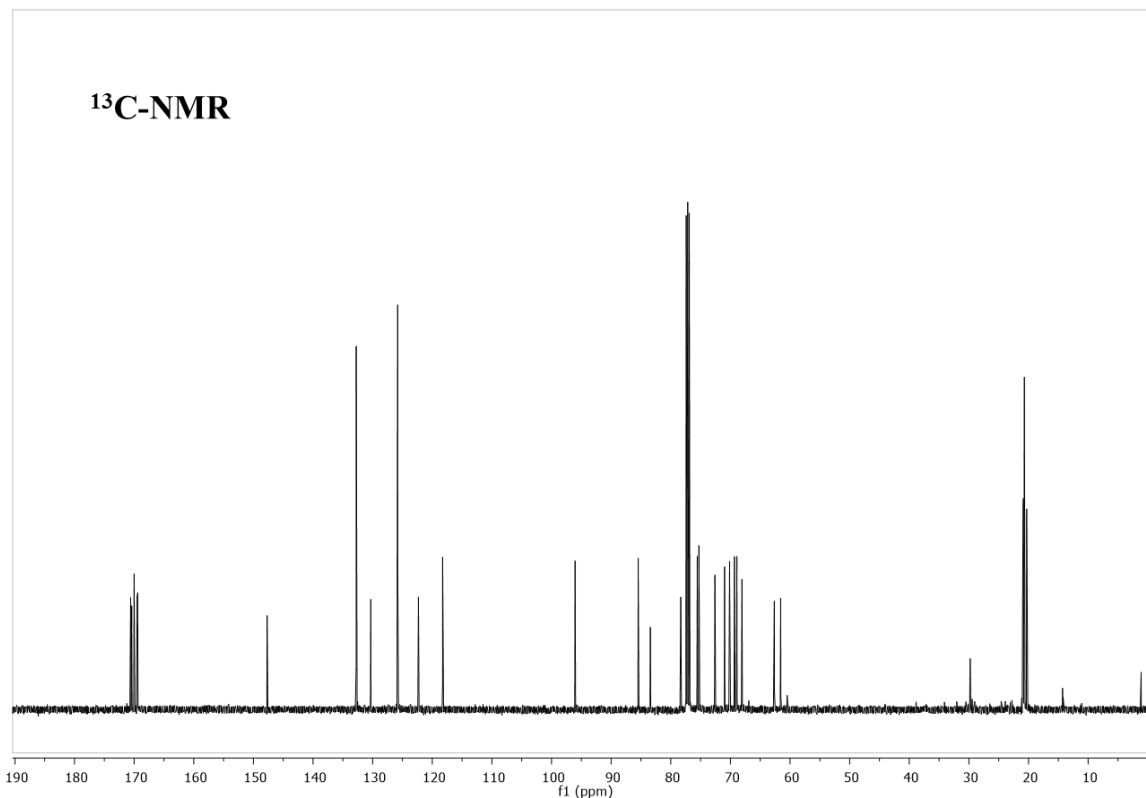
Synthesis of 1-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl)-4-(4-ethynylphenyl)-1*H*-1,2,3-triazole (**40**):
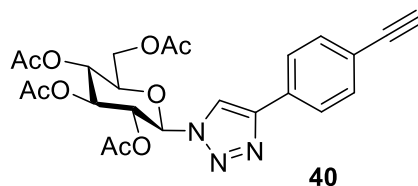
40

The general procedure was applied to the derivative 2,3,4,6-tetra-*O*-acetyl-1-azido-1-deoxy-β-D-glucopyranoside **3** (448 mg, 1.20 mmol, 1 eq) and 1,4-diethynylbenzene. The reaction was stopped after 45 min. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 1:1) to give **40** (216 mg, 36%) as a white solid. ^{1}H NMR (500 MHz, CDCl$_3$) δ 8.02 (s, 1H, H$_{TRIAZOLE}$), 7.83–7.78 (m, 2H, 2xH$_{AR}$), 7.57–7.53 (m, 2H, 2xH$_{AR}$), 5.93 (d, J=9.2 Hz, 1H, H$_1$), 5.50 (t, J=9.4 Hz, 1H, H$_2$), 5.44 (t, J=9.4 Hz, 1H, H$_3$), 5.27 (dd, J=10.2, 9.2 Hz, 1H, H$_4$), 4.33 (dd, J=12.7, 5.1 Hz, 1H, H$_6$), 4.16 (dd, J=12.6, 2.1 Hz, 1H, H$_{6'}$), 4.03 (ddd, J=10.1, 5.1, 2.1 Hz, 1H, H$_5$), 3.14 (s, 1H, C≡CH), 2.08 (d, J=4.2 Hz, 6H, 6xCOCH$_3$), 2.04 (s, 3H, 3xCOCH$_3$), 1.89 (s, 3H, 3xCOCH$_3$) ^{13}C NMR (126 MHz, CDCl$_3$) δ 170.59 (COCH$_3$), 170.02(COCH$_3$), 169.49(COCH$_3$), 169.15(COCH$_3$), 147.87 (C$_{TRIAZOLE}$), 132.79 (CH$_{AR}$), 130.36 (C$_{AR}$), 125.8 (CH$_{AR}$), 122.35 (C$_{AR}$), 118.25 (CH$_{TRIAZOLE}$), 86.00 (C$_1$), 83.48 (C≡CH), 78.29 (C≡CH), 75.38 (C$_5$), 72.81 (C$_3$), 70.38 (C$_2$), 67.86 (C$_4$), 61.70 (C$_6$), 20.83 (COCH$_3$), 20.68

(CO$\underline{C}$H$_3$), 20.66 (CO$\underline{C}$H$_3$), 20.33 (CO$\underline{C}$H$_3$). HRMS (FAB$^+$) calculated for C$_{24}$H$_{26}$N$_3$O$_9$ (M+H): 500.1669; found: 500.1648.
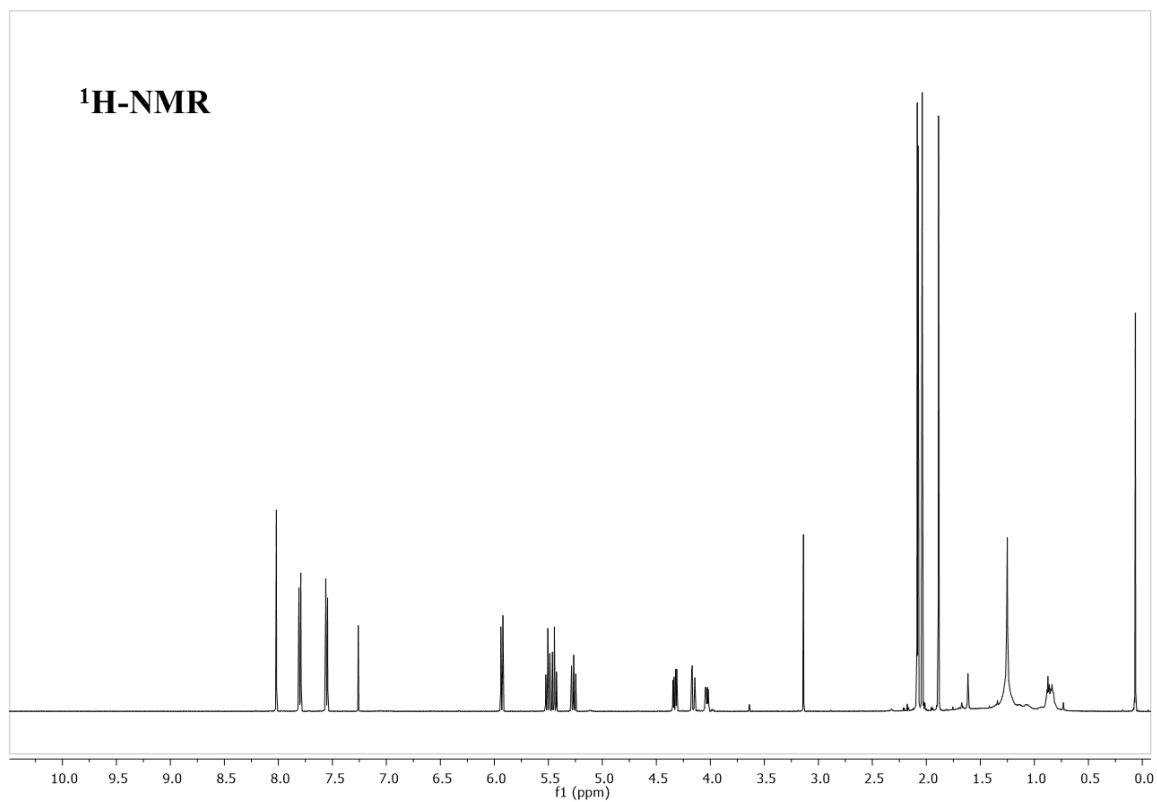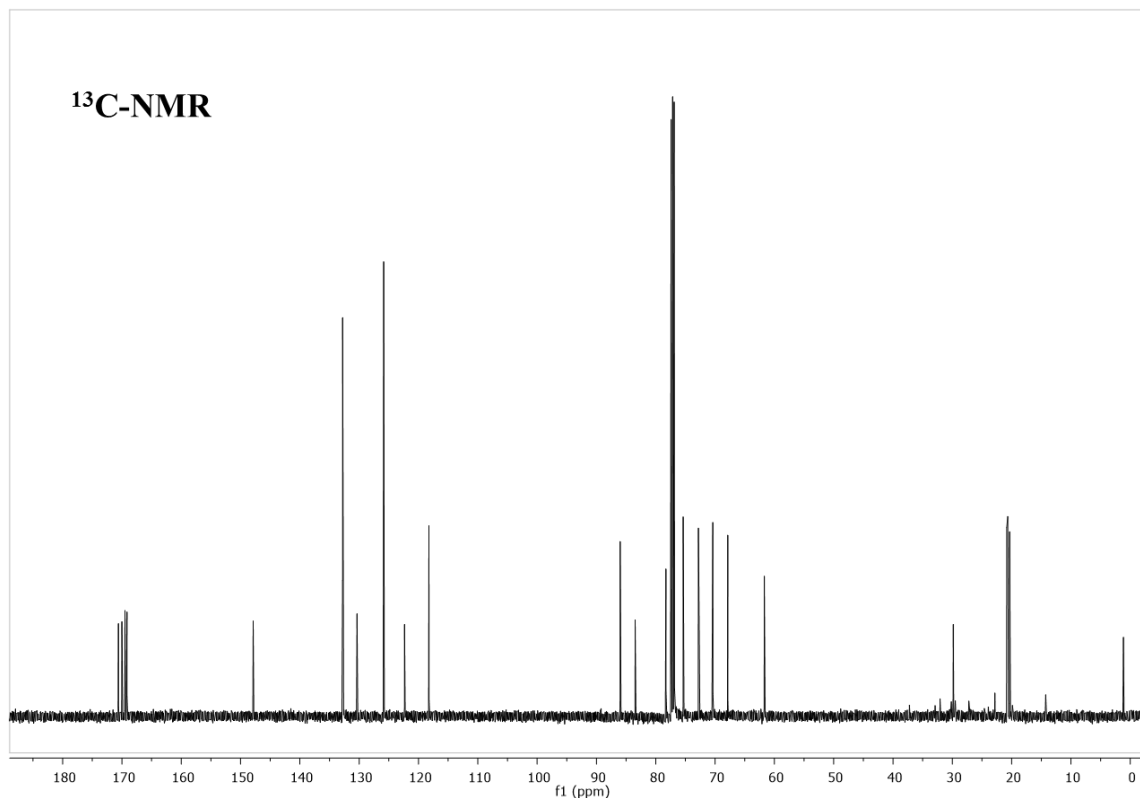
Synthesis of 1-(2-ethanol)-4-(3-ethynylphenyl)-1*H*-1,2,3-triazole **41**:
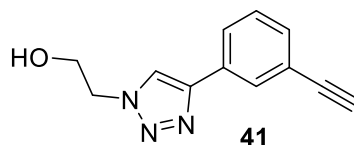
The general procedure was applied to 2-azidoethanol **15** (125 mg, 1.44 mmol, 1 eq) and 1,3-diethynylbenzene. This reaction was stirred at room temperature instead of using microwave. The reaction was stopped after 2h. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:1 to 1:2) to give **41** (81 mg, 26%) as a white solid. ^{1}H NMR (300 MHz, CD$_3$OD) δ 8.36 (s, 1H, H$_{TRIAZOLE}$), 7.94 (d, J=1.7 Hz, 1H, H$_{AR}$), 7.82 (dt, J=6.6, 2.1 Hz, 1H, H$_{AR}$), 7.50–7.36 (m, 2H, 2xH$_{AR}$), 4.55 (t, J=5.2 Hz, 2H, 2xC$\underline{H}_2$OH), 4.00 (t, J=5.2 Hz, 2H, 2xNCH$_2$), 3.56 (s, 1H, C≡CH) ^{13}C NMR (75 MHz, CD$_3$OD) δ 147.63 (C$_{TRIAZOLE}$), 132.60 (CH$_{AR}$), 132.18 (C$_{AR}$), 130.14 (CH$_{AR}$), 129.94 (CH$_{AR}$), 126.89 (CH$_{AR}$), 124.44 (C$_{AR}$), 123.26 (C$_{AR}$), 83.98 (C≡CH), 79.19 (C≡CH), 61.55 (C$\underline{H}_2$OH), 54.0 1(NC$\underline{H}_2$). HRMS (FAB$^+$) calculated for C$_{12}$H$_{12}$N$_3$O (M+H): 214.0980; found: 214.0969.

¹H-NMR

¹³C-NMR
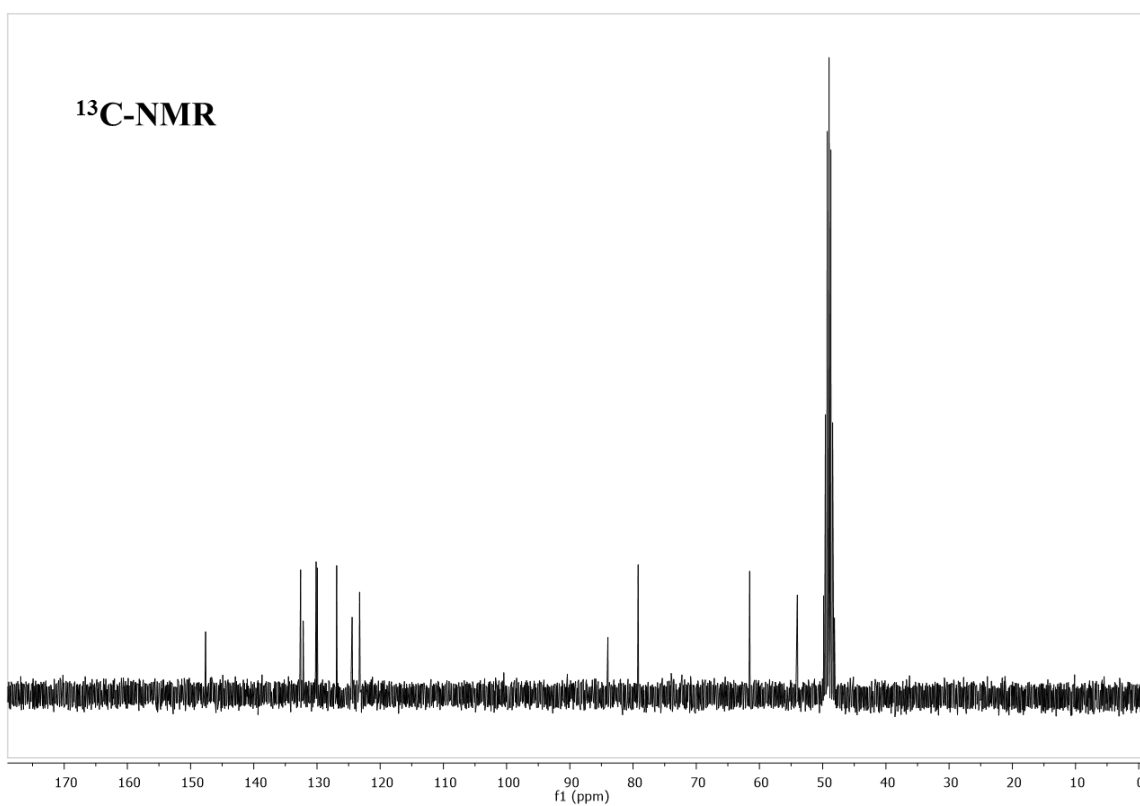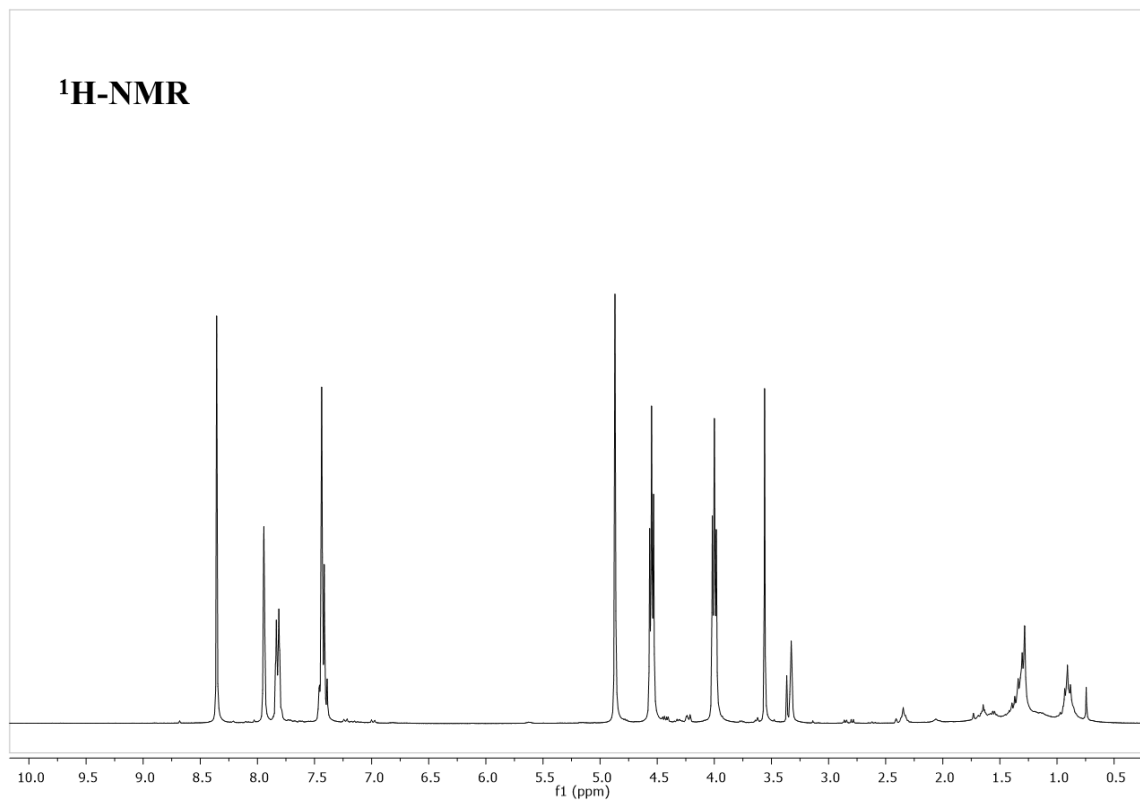
Synthesis of N-(4-azidophenyl)-(3-pyrrolidin-1-yl) propionamide (2)

Synthesis of *N*-(4-nitrophenyl)-(3-pyrrolidin-1-yl) propionamide (**49**):

49
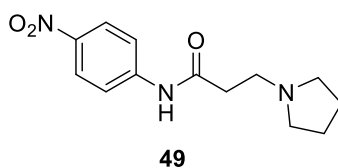
4-Nitroaniline (5 g, 36.2 mmol, 1 eq) was suspended in 50 mL dry THF, and triethylamine (10 mL, 72.4 mmol, 2 eq), and 3-chloropropionylchloride (4.15 mL, 43.4 mmol, 1.2 eq) were added at 0°C. After stirred for 2h at room temperature, pyrrolidine (9.1 mL, 108.6 mmol, 3 eq) was added, and the reaction was stirred for 16 h at room temperature. Reaction was concentrated, the crude dissolved in dichloromethane and washed with $NaHCO_3$ saturated solution (3x80 mL). The organic phase was dried over $MgSO_4$, filtered off and solvent evaporated under reduced pressure. The crude was purified by silica gel column chromatography using as eluent DCM/MeOH (9.9:0.1 to 7:3) to give **49** (8.1 g, 85%) as an orange solid, as previously reported.[227] ^{1}H NMR (300 MHz, CDCl$_3$) δ 11.76 (s, 1H, N<u>H</u>CO) 8.32 (q, J=2.5 Hz, 1H, H$_{AR}$), 7.82 (dddd, J=18.7, 8.1, 2.1, 1.0 Hz, 2H, 2xH$_{AR}$), 7.41 (t, J=8.2 Hz, 1H, H$_{AR}$), 2.86 (dd, J=6.7, 5.1 Hz, 2H, 2xCOC<u>H</u>$_2$), 2.74–2.62 (m, 4H, 4xNC<u>H</u>$_2$), 2.55 (dd, J=6.6, 5.1 Hz, 2H, 2xC<u>H</u>$_2$N), 1.91 (h, J=3.0 Hz, 4H, 4xC<u>H</u>$_2$CH$_2$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 171.47 (NH<u>C</u>O), 148.61(<u>C</u>$_{AR}$NO$_2$), 140.03 (<u>C</u>$_{AR}$NH), 129.68 (CH$_{AR}$), 125.36 (CH$_{AR}$), 118.11 (CH$_{AR}$), 114.32 (CH$_{AR}$), 53.19 (<u>C</u>H$_2$N), 51.21 (2xN<u>C</u>H$_2$), 34.54 (CO<u>C</u>H$_2$), 23.80 (2x<u>C</u>H$_2$CH$_2$).

Synthesis of *N*-(4-Amino-phenyl)-(3-pyrrolidin-1-yl) propionamide (**50**)

50
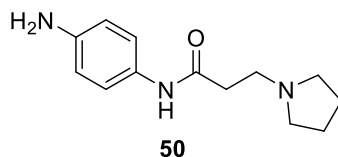
The derivative *N*-(4-nitrophenyl)-(3-pyrrolidin-1-yl) propionamide (**49**)(1.7 g, 6.46 mmol, 1eq) was solved in anhydrous THF, and catalytic amounts of Pd/C were added to the suspension. The reaction was stirred under an atmosphere of hydrogen for 16 h. The mixture was filtered off over celite and solvents were removed. The crude obtained (1.5 g, 90%) as a brown oil was the product used in the following step without any purification.[227] ^{1}H NMR (300 MHz, CDCl$_3$) δ 11.06 (s, 1H, N<u>H</u>CO), 7.18 (t, J=2.1 Hz, 1H, H$_{AR}$), 7.04 (t, J=8.0 Hz, 1H, H$_{AR}$), 6.60 (ddd, J=8.0, 2.0, 0.9 Hz, 1H, H$_{AR}$), 6.38 (ddd, J=8.0, 2.3, 0.9 Hz, 1H, H$_{AR}$), 3.64 (s, 2H, 2xN<u>H</u>$_2$), 2.81 (dd, J=6.7, 5.1 Hz, 2H, 2xCOC<u>H</u>$_2$), 2.71–2.57 (m, J=3.0, 2.1 Hz, 4H, 4xNC<u>H</u>$_2$), 2.50 (dd, J=6.6, 5.1 Hz, 2H, 2xC<u>H</u>$_2$N), 1.87 (h, J=3.0 Hz, 4H, 4xC<u>H</u>$_2$CH$_2$).

Synthesis of *N*-(4-Azidophenyl)-(3-pyrrolidin-1-yl) propionamide (**2**):
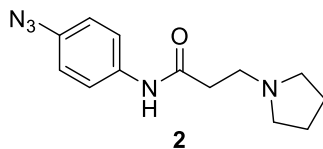
To a solution of *N*-(4-aminophenyl)-(3-pyrrolidin-1-yl) propionamide (**50**)(3 g, 12.86mmol, 1 eq) in anhydrous THF (40 mL) at 0°C conc. HCl (6 mL, 70.72 mol, 5.5 eq), followed by t-BuONO (4 mL, 32.15 mmol, 2.5 eq) was added. The mixture was stirred at 0°C for 1.5 h, then NaN$_3$ (2.5 g, 38.58 mmol, 3 eq) was added, followed by careful addition of H$_2$O until reaction did not produce any bubble. The reaction mixture was cooled to room temperature and stirred for 16h. When the reaction was complete, it was quenched with NaHCO$_3$ saturated solution (30 mL) to neutralize the mixture. The THF was removed under reduced pressure and the remaining aqueous mixture was extracted with dichloromethane (4x50 mL). The organics were dried over anhydrous MgSO$_4$, and concentrated under reduced pressure. The crude was purified by alumina using as eluent DCM/MeOH (1:0 to 10:0.05) to give **2** (1.1 g, 33%) as an orange solid, according to the literature.[227] ^{1}H NMR (300 MHz, CD$_3$OD) δ 10.94 (s, 1H, N*H*CO), 6.95 (t, J=2.1 Hz, 1H, H$_{AR}$), 6.83 (t, J=8.0 Hz, 1H, H$_{AR}$), 6.71 (dd, J=8.5, 1.6 Hz, 1H, H$_{AR}$), 6.33–6.27 (m, 1H, H$_{AR}$), 2.45 (dd, J=6.6, 5.2 Hz, 2H, 2xCOC*H*$_2$), 2.28 (td, J=6.2, 3.4 Hz, 4H, 4xNC*H*$_2$), 2.13 (dd, J=6.7, 5.1 Hz, 2H, 2xC*H*$_2$N), 1.48 (h, J=3.1 Hz, 4H, 4xC*H*$_2$CH$_2$).

Synthesis of N,N 0-{4,40[4,40-(1,3-phenylene)bis(1H-1,2,3-triazole-4,1-diyl)]-bis(4,1-phenylene)}bis(3-pyrolidino propionamide) (1)
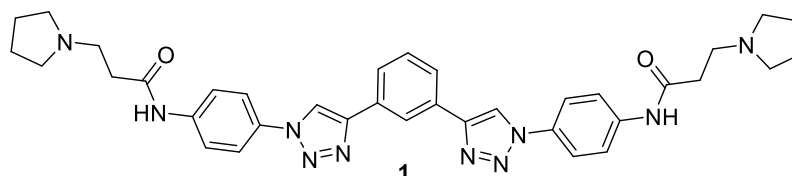
To a solution of N-(4-azidophenyl)-(3-pyrrolidin-1-yl)propionamide **2** (500 mg, 1.93 mmol, 3 eq), in 1 : 1 t-BuOH: H$_2$O (10 mL) was added 1,3-diethynylbenzene (85 µL, 0.64 mmol, 1 eq), sodium ascorbate (63.4 mg, 0.32 mmol, 0.5 eq) and CuSO$_4$·5H$_2$O (7.5 mg, 0.03 mmol, 0.05 eq). This solution was stirred and heated in a microwave at 130 °C, for 45 min with the fixed hold mode on. Then Quadrasil® MP was added to chelate the copper salt, by stirring the suspension for 5 min at room temperature. The mixture was filtered and solvents were evaporated at reduced pressure. The crude was purified by flash chromatography using as eluents CH$_2$Cl$_2$/MeOH (10:0 to 9:1). The obtained fraction was purified again by Sephadex LH20 (DCM/MeOH 1:1) and C18 using MeOH/H$_2$O as eluents (3:7 to 10:0) to obtain **1** as previously described (45 mg, 11%).[227] ^{1}H NMR (400 MHz, CD$_3$OD) δ 8.91 (s, 2H, 2xH$_{TRIAZOLE}$), 8.44 (d, J=24.9 Hz, 3H, 3xH$_{AR}$), 7.98–7.79 (m, 2H, 2xH$_{AR}$), 7.68–7.41 (m, 7H, 7xH$_{AR}$), 4.51 (d, J = 22.6 Hz, 2H, 2xNC*H*$_2$), 3.90–3.37 (m, 6H, 2xNC*H*$_2$, 4xC*H*$_2$N), 3.18 (s, 4H, 4xNC*H*$_2$), 2.97 (t, J = 6.1 Hz, 4H, 4xCOC*H*$_2$), 2.12 (d, J = 46.6 Hz, 8H, 8xC*H*$_2$CH$_2$).

General procedure for the synthesis of the asymmetric phenyl ditriazole compounds via "click chemistry"

The corresponding monotriazole derivative (1 eq), *N*-(4-azidophenyl)-(3-pyrrolidin-1-yl) propionamide **2** (2 eq), CuSO$_4$·5H$_2$O (0.05eq), sodium ascorbate (0.5 eq) and TBTA (0.06 eq) were suspended in 1.5 mL of THF:H$_2$O solvents mixture (1:1) inside a microwave reactor. The reaction was heated at 60°C for 30 min and the reaction was monitored by TLC until starting material disappeared. Then, Quadrasil® MP was added to chelate the copper salt, by stirring the suspension for 5 min at room temperature. The mixture was filtered and solvents were evaporated under reduced pressure. The crude was taken to the deprotection step without further purification. Sodium methylate solution in MeOH (1 N, 1 mL/g starting material) was added to the corresponding acetylated crude solution in dry MeOH. The mixture was stirred at room temperature until no starting material was detected (~1h). Amberlite IR®120 was added to neutralize the mixture, the resin was filtered off, washed with MeOH, and the filtrate was evaporated. The new crude was purified by HPLC following the methods described in the general material and methods section.

Synthesis of the glucose asymmetric PTDZ derivative (**51**):
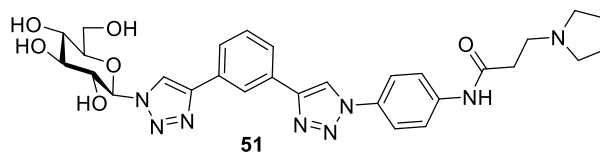
The general procedure was applied to the derivative **32** (139 mg, 0.278 mmol). The crude was purified by HPLC with method I. A brown solid was obtained as final product **51** (12 mg, 7.3%). ^{1}H NMR (400 MHz, D$_2$O) δ 8.03 (s, 1H, H$_{TRIAZOLE}$), 7.65 (s, 1H, H$_{AR}$), 7.08 (s, 2H, 2xH$_{AR}$), 7.05–6.92 (m, 4H, 4xH$_{AR}$), 6.81 (dd, J=27.0, 7.8 Hz, 2H, 2xH$_{AR}$), 5.65 (d, J=9.4 Hz, 1H, H$_1$), 3.98 (q, J=10.4, 9.1 Hz, 2H, H$_2$, H$_6$), 3.86 (dd, J=13.0, 5.3 Hz, 1H, H$_5$), 3.77 (q, J=9.3, 8.3 Hz, 2H, H$_6$', H$_3$), 3.66 (dd, J=9.9, 4.5 Hz, 3H, 2xNC$\underline{H}_2$, H$_4$), 3.35 (t, J=7.1 Hz, 2H, 2xC$\underline{H}_2$N), 3.07 (dt, J=12.6, 7.6 Hz, 2H, 2xNC$\underline{H}_2$), 2.61 (t, J=6.7 Hz, 2H, 2xCOC$\underline{H}_2$), 2.21–2.09 (m, 2H, 2xC$\underline{H}_2$CH$_2$), 2.02 (q, J=6.2, 5.2 Hz, 2H, 2xC$\underline{H}_2$CH$_2$). ^{13}C NMR (101 MHz, D$_2$O) δ 168.73 (CO), 146.35 (C$_{TRIAZOLE}$), 146.16 (C$_{TRIAZOLE}$), 137.84 (C$_{AR}$), 135.04 (C$_{AR}$), 129.80 (C$_{AR}$), 128.78 (C$_{AR}$), 128.54 (CH$_{AR}$), 124.86 (CH$_{AR}$), 120.96 (CH$_{TRIAZOLE}$), 120.75 (CH$_{AR}$), 117.81 (CH$_{TRIAZOLE}$), 117.65 (CH$_{AR}$), 114.91 (CH$_{AR}$), 109.08 (CH$_{AR}$), 87.46 (C$_1$), 78.94 (C$_3$), 76.08 (C$_4$), 72.33 (C$_2$), 69.11 (C$_5$), 60.66 (C$_6$), 54.38 (2xNCH$_2$), 50.31 ($\underline{C}$H$_2$N), 31.69 (CO$\underline{C}$H$_2$), 22.67 (2x$\underline{C}$H$_2$CH$_2$). LRMS (ES$^+$) calculated. for C$_{29}$H$_{35}$N$_8$O$_6$, (M+H): 591.3; found 591.4. HRMS (FAB$^+$) calculated for C$_{29}$H$_{35}$N$_8$O$_6$, (M+H): 591.2680; found: 591.2652.

¹H-NMR

¹³C-NMR
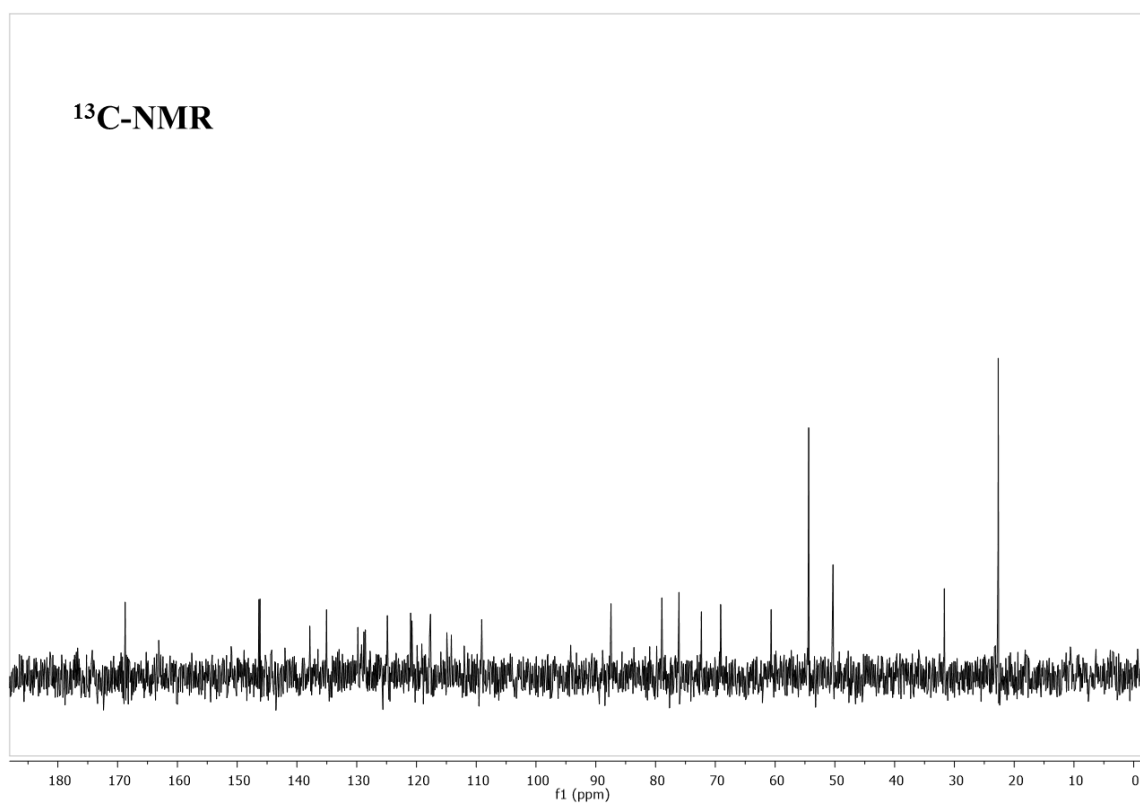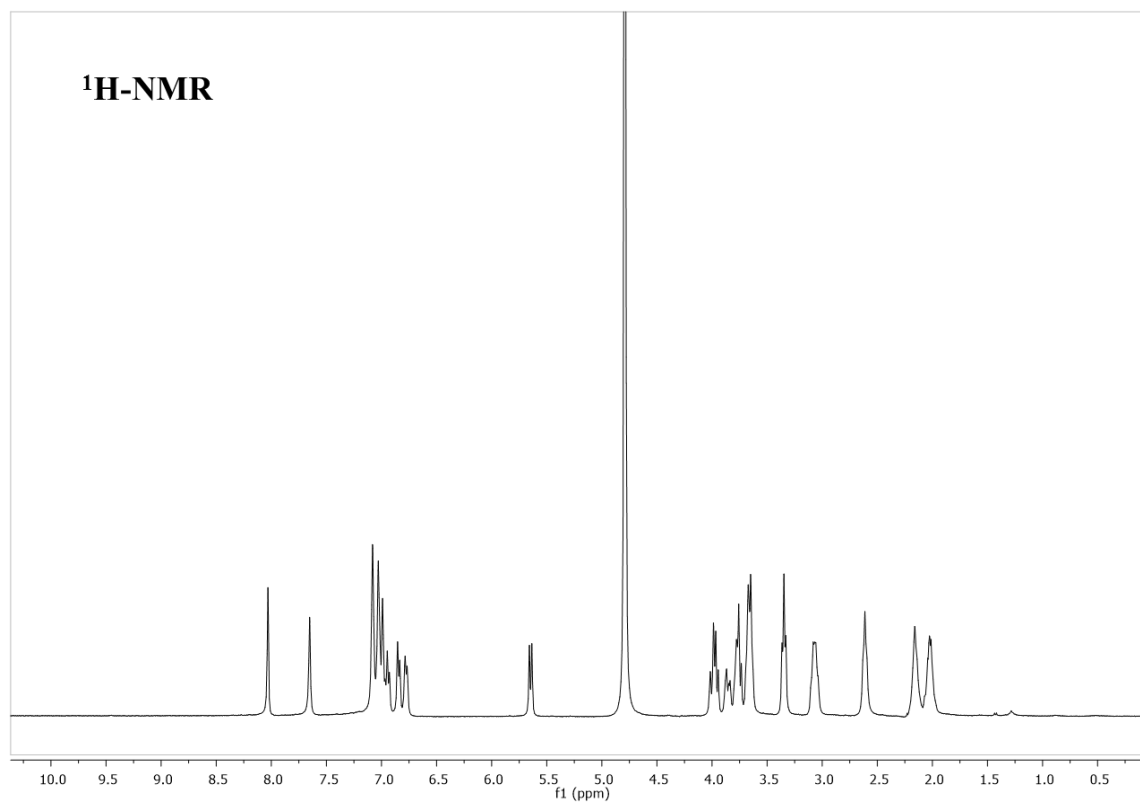
Synthesis of the maltose asymmetric PTDZ derivative (**52**):
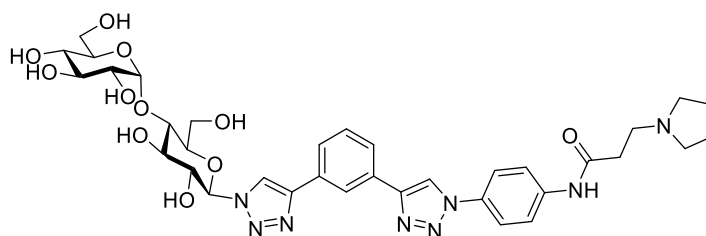
52

The general procedure was applied to the derivative **33** (46 mg, 0.058 mmol). The crude was purified by HPLC with method II. A brown solid was obtained as final product **52** (14 mg, 32%). ^{1}H NMR (400 MHz, CD$_3$OD) δ 8.95 (s, 1H, H$_{AR}$), 8.68 (s, 1H, H$_{AR}$), 8.45 (q, J=1.9 Hz, 2H, 2x H$_{AR}$), 7.98–7.91 (m, 1H, H$_{AR}$), 7.88 (dt, J=7.8, 1.4 Hz, 1H, H$_{AR}$), 7.69–7.53 (m, 4H, H$_{AR}$), 5.74 (d, J=9.1 Hz, 1H, H$_{1A}$), 5.29 (d, J=3.8 Hz, 1H, H$_{1B}$), 4.05 (t, J=9.1 Hz, 1H, H$_{2A}$), 3.96–3.64 (m, 10H, 10xH$_{SUGAR}$, 2xNC$\underline{H}_2$), 3.60 (t, J=6.4 Hz, 2H, 2xC$\underline{H}_2$N), 3.51 (dd, J=9.7, 3.8 Hz, 1H, H$_{2B}$),3.19 (dd, J=16.9, 7.0 Hz, 2H, 2xNC$\underline{H}_2$), 2.98 (t, J=6.4 Hz, 2H, 2xCOC$\underline{H}_2$), 2.12 (dd, J=48.7, 13.6 Hz, 4H, 4xC$\underline{H}_2$CH$_2$). ^{13}C NMR (101 MHz, D$_2$O) δ 170.29(CO), 149.23(C$_{TRIAZOLE}$), 148.48(C$_{TRIAZOLE}$), 141.17(C$_{AR}$), 138.76(C$_{AR}$), 132.45(C$_{AR}$), 132.18(C$_{AR}$), 131.45(CH$_{AR}$), 130.82(CH$_{AR}$), 126.81(CH$_{AR}$), 126.68(CH$_{AR}$), 124.00(CH$_{AR}$), 121.71(CH$_{AR}$), 120.91(CH$_{AR}$), 120.56(CH$_{AR}$), 116.80(CH$_{AR}$), 112.95(CH$_{AR}$), 102.97(C$_{1B}$), 89.62 (C$_{1A}$), 80.33, 79.69, 78.24, 75.10, 74.92, 74.20, 73.76, 71.53(C$_{2A}$, C$_{2B}$, C$_{3A}$, C$_{3B}$, C$_{4A}$, C$_{4B}$, C$_{5A}$, C$_{5B}$), 62.77(C$_{6B}$), 61.88(C$_{61A}$), 55.56 (2xNC$\underline{H}_2$), 52.212xC$\underline{H}_2$N, 32.80(COC$\underline{H}_2$), 24.04 (2xC$\underline{H}_2$CH$_2$). LRMS (ES$^+$) calculated for C$_{35}$H$_{45}$N$_8$O$_{11}$(M+H): 753.3; found: 753.5. HRMS (FAB$^+$), calculated for C$_{35}$H$_{45}$N$_8$O$_{11}$(M+H): 753.3208; found: 753.3167.

¹H-NMR

¹³C-NMR

Synthesis of the glucose C2 linker asymmetric PTDZ derivative (**53**)

The general procedure was applied to the derivative **36** (40 mg, 0.074 mmol). The crude was purified by HPLC with method III. A yellow solid was obtained as final product **53** (15 mg, 30%). ^{1}H NMR (400 MHz, D$_2$O) δ 7.75 (d, J=4.9 Hz, 1H, H$_{TRIAZOLE}$), 7.59 (d, J=19.7 Hz, 1H, H$_{TRIAZOLE}$), 7.00 (p, J=8.9 Hz, 7H, 7xH$_{AR}$), 6.75 (s, 1H, H$_{AR}$), 4.44 (d, J=7.9 Hz, 1H, H$_1$), 4.37 (d, J=6.7 Hz, 2H, 2xC$\underline{H}_2$N$_{TRIAZOLE}$), 4.16 (t, J=7.3 Hz, 1H, OC$\underline{H}_2$), 4.02–3.93 (m, 1H, OC$\underline{H}_2$), 3.94–3.84 (m, 1H, H$_6$), 3.68 (dd, J=12.0, 5.8 Hz, 3H, H$_{6'}$, 2xNC$\underline{H}_2$), 3.53–3.31 (m, 5H, H$_3$, H$_4$, H$_5$, 2xC$\underline{H}_2$N), 3.27 (t, J=8.6 Hz, 1H, H$_2$), 3.08 (ddd, J=12.5, 8.9, 6.0 Hz, 2H, 2xNC$\underline{H}_2$), 2.67 (t, J=6.7 Hz, 2H, 2xCOC$\underline{H}_2$), 2.22–2.08 (m, 2H, 2xC$\underline{H}_2$CH$_2$), 2.09–1.93 (m, 2H, 2xC$\underline{H}_2$CH$_2$). ^{13}C NMR (101 MHz, D$_2$O) δ 168.78 (CO), 145.89 (2xC$_{TRIAZOLE}$), 137.89 (2xC$_{AR}$), 128.41 (C$_{AR}$, CH$_{AR}$), 122.63 (CH$_{AR}$), 120.63 (CH$_{TRIAZOLE}$), 117.75 (CH$_{TRIAZOLE}$), 117.42 (CH$_{AR}$), 114.86 (CH$_{AR}$), 102.37 (C$_1$), 75.97, 75.73, 73.05 (C$_3$, C$_4$, C$_5$), 69.68 (C$_2$), 60.80 (C$_6$), 54.35 (2xNCH$_2$), 50.28 (2xC$\underline{H}_2$N), 31.76 (COC$\underline{H}_2$), 22.64 (2xC$\underline{H}_2$CH$_2$). LRMS (ES$^+$) calculated for C$_{31}$H$_{39}$N$_8$O$_7$(M+H): 635.3; found: 635.4. HRMS (FAB$^+$) calculated for C$_{31}$H$_{39}$N$_8$O$_7$(M+H):635.2942; found: 635.2909.

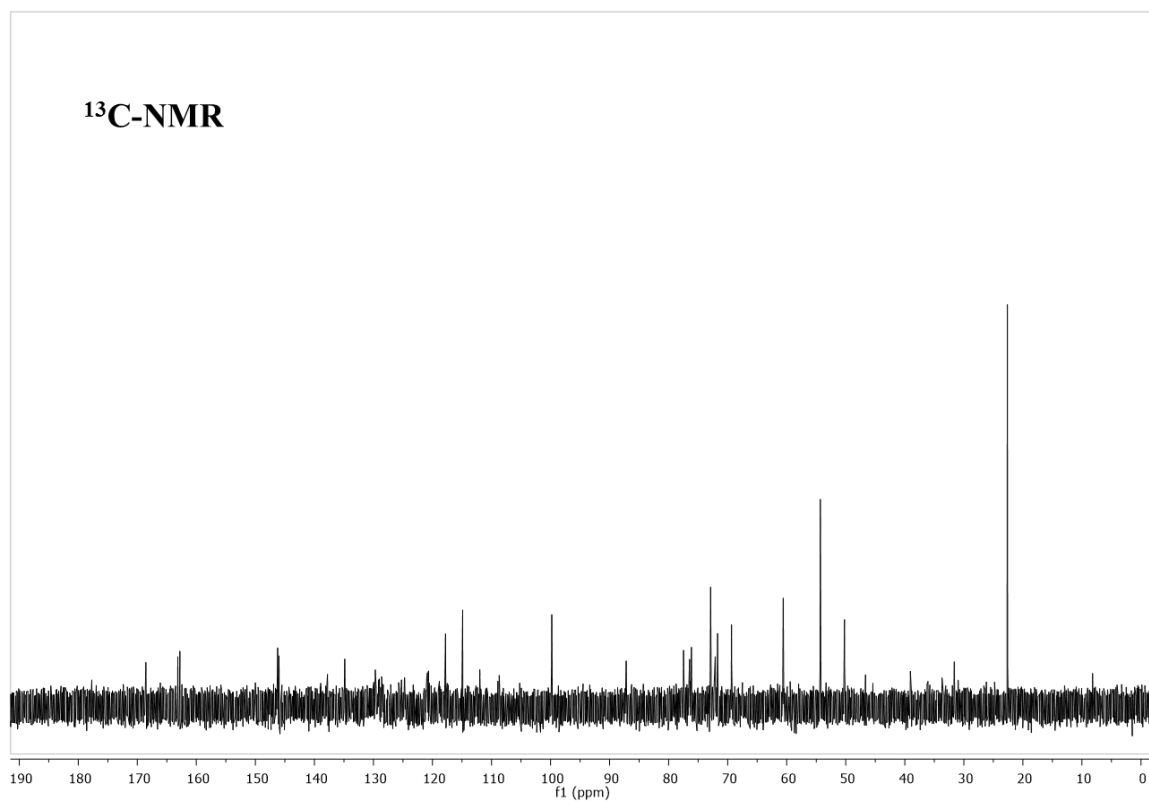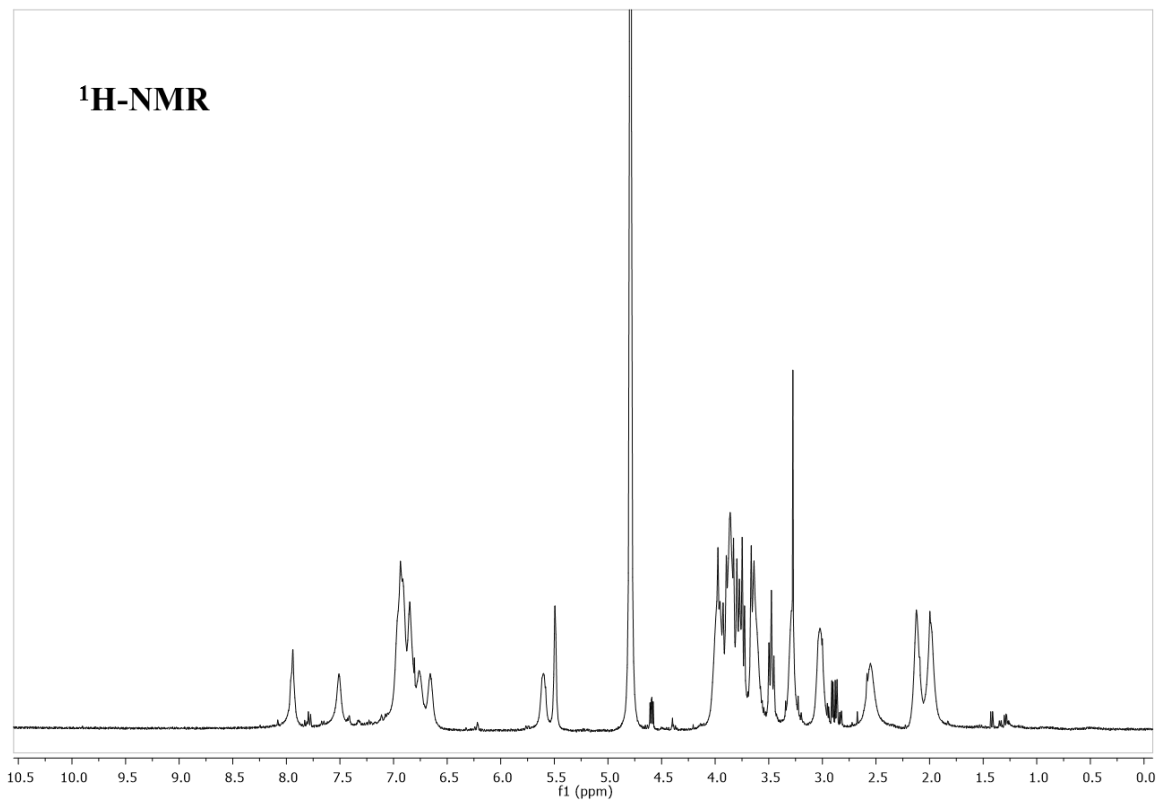^{1}**H-NMR**

¹³C-NMR

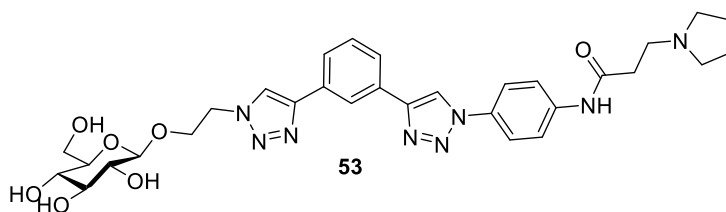## Synthesis of the fucose asymmetric PTDZ derivative (**54**)

54

The general procedure was applied to the derivative **34** (25.5 mg, 0.058 mmol). The crude was purified by HPLC with method IV. A white solid was obtained as final product **54** (7 mg, 21%). ^{1}H NMR (400 MHz, D$_2$O) δ 8.28 (s, 1H, H$_{TRIAZOLE}$), 7.92 (d, J=2.5 Hz, 1H, H$_{TRIAZOLE}$), 7.40 (s, 1H, H$_{AR}$), 7.35–7.16 (m, 4H, 4xH$_{AR}$), 7.16–7.07 (m, 1H, H$_{AR}$), 7.07–6.93 (m, 2H, 2xH$_{AR}$), 5.73–5.63 (m, 1H, H$_1$), 4.33–4.24 (m, 1H, H$_2$), 4.23–4.14 (m, 1H, H$_5$), 4.07–3.95 (m, 2H, H$_3$, H$_4$), 3.74 (d, J=10.0 Hz, 2H, 2xNC$\underline{H}_2$), 3.42 (q, J=5.8, 4.8 Hz, 2H, 2xC$\underline{H}_2$N), 3.20–3.07 (m, 2H, 2xNC$\underline{H}_2$), 2.71 (t, J=7.0 Hz, 2H, 2xCOC$\underline{H}_2$), 2.21 (d, J=7.9 Hz, 2H, 2xCH$_2$C$\underline{H}_2$), 2.15–2.03 (m, 2H, 2xCH$_2$C$\underline{H}_2$), 1.41 (d, J=6.4 Hz, 3H, 3xCC$\underline{H}_3$). ^{13}C NMR (101 MHz, D$_2$O) δ 168.93 (NH$\underline{C}$O), 146.66 (C$_{TRIAZOLE}$), 146.43 (C$_{TRIAZOLE}$), 137.94 (C$_{AR}$), 135.29 (C$_{AR}$), 129.12 (C$_{AR}$), 128.84 (C$_{AR}$), 125.09 ($\underline{C}$H$_{AR}$), 125.05 ($\underline{C}$H$_{AR}$), 120.94 ($\underline{C}$H$_{TRIAZOLE}$), 119.46 ($\underline{C}$H$_{AR}$), 118.04 ($\underline{C}$H$_{AR}$), 117.84 ($\underline{C}$H$_{AR}$), 114.94 ($\underline{C}$H$_{AR}$), 109.59 ($\underline{C}$H$_{AR}$), 88.09 (C$_1$), 74.37 (C$_5$), 73.31 (C$_4$), 71.28 (C$_3$), 69.58 (C$_2$), 54.42 (N$\underline{C}$H$_2$), 50.37 ($\underline{C}$H$_2$N), 31.74 (CO$\underline{C}$H$_2$), 22.70(2xCH$_2$$\underline{C}H_2$), 15.76 (C$_6$). HRMS (FAB$^+$), Calculated for C$_{29}H_{35}N_8O_5$ (M+H): 575.2730; found: 575.2724.
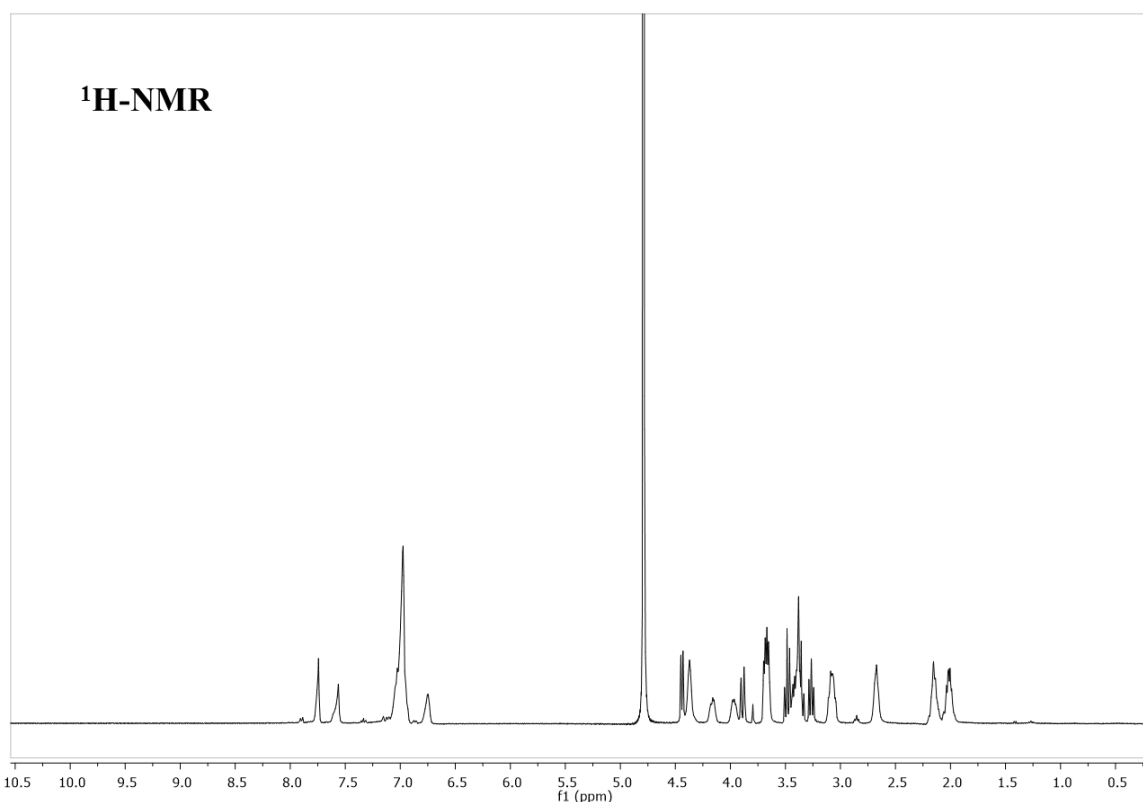
¹H-NMR

¹³C-NMR
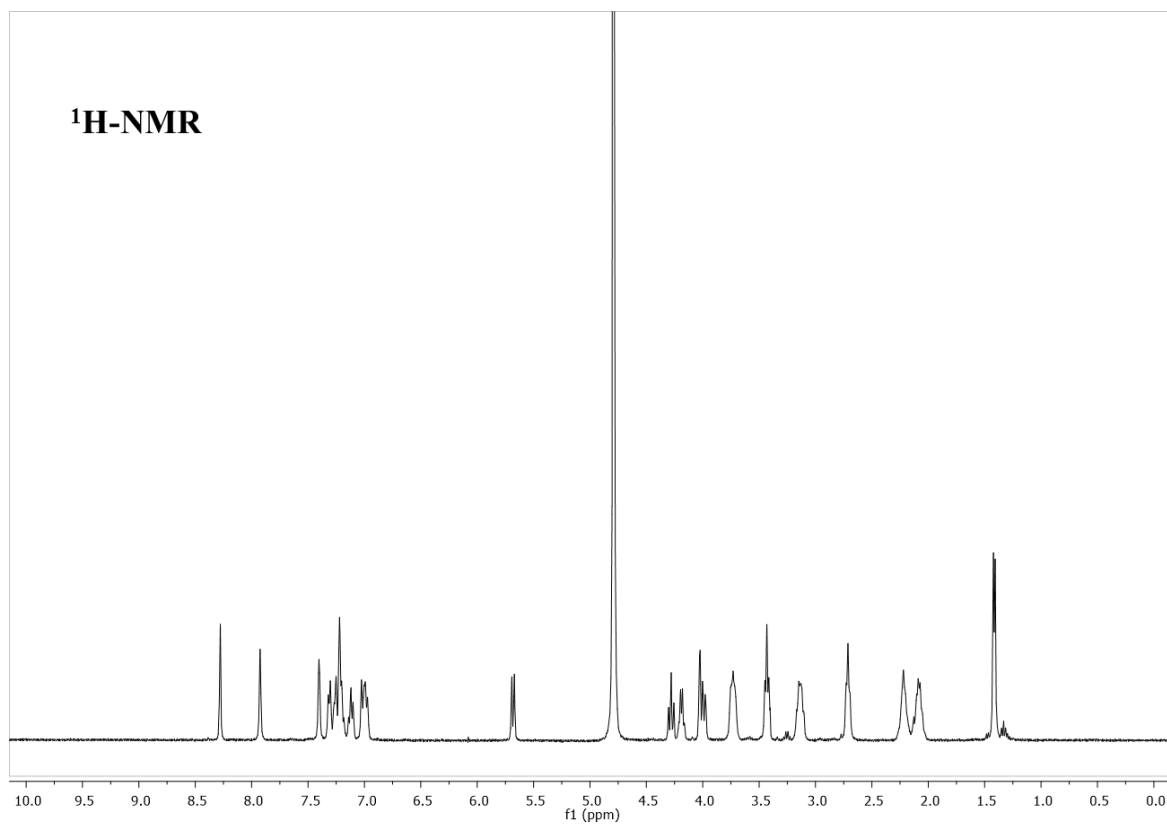
Synthesis of the 2-glucosamine asymmetric PTDZ derivative (**55**)

55
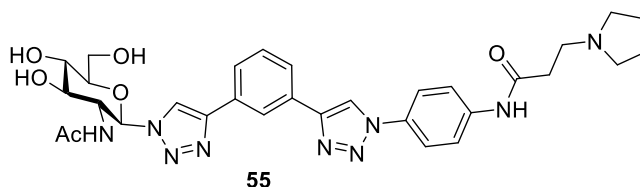
The general procedure was applied to the derivative **35** (29 mg, 0.058 mmol). The crude was purified by HPLC with method V. A white solid was obtained as final product **55** (17 mg, 46%). ^{1}H NMR (400 MHz, D_2O) δ 8.10 (d, J=2.1 Hz, 1H, $H_{TRIAZOLE}$), 7.66 (s, 1H, $H_{TRIAZOLE}$), 7.14 (s, 1H, H_{AR}), 7.09 (d, J=7.2 Hz, 1H, H_{AR}), 7.06–6.93 (m, 3H, 3xH_{AR}), 6.90 (d, J=7.3 Hz, 1H, H_{AR}), 6.84 (d, J=7.9 Hz, 1H, H_{AR}), 6.74 (d, J=7.7 Hz, 1H, H_{AR}), 5.77 (dd, J=9.3, 1.9 Hz, 1H, H_1), 4.35–4.23 (m, 1H, H_2), 4.04 (d, J=12.4 Hz, 1H, H_6), 3.97–3.72 (m, 4H, H_3, H_4, H_5,H_6'), 3.69 (q, J=8.7, 7.1 Hz, 2H, 2xNC$\underline{H}_2$), 3.41–3.30 (m, 2H, 2xC$\underline{H}_2$N), 3.09 (dd, J=11.5, 6.6 Hz, 2H, 2xNC$\underline{H}_2$), 2.62 (t, J=7.0 Hz, 2H, 2xCOC$\underline{H}_2$), 2.16 (d, J=8.2 Hz, 2H, 2xCH$_2$C$\underline{H}_2$), 2.04 (q, J=6.8, 5.8 Hz, 2H, 2xC$\underline{H}_2$), 1.86 (d, J=1.7 Hz, 3H, 3xNHCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, D_2O) δ 173.79 ($\underline{C}OCH_3$), 168.62 (NH$\underline{C}O$), 163.03, 162.68(CF_3COOH), 146.32 (C_{AR}), 146.12 (C_{AR}), 137.88 (C_{AR}), 135.01 (C_{AR}), 129.64 (CH_{AR}), 128.60 (CH_{AR}), 124.77 (CH_{AR}), 120.44 (CH_{AR}), 118.91 (CH_{AR}), 117.82 (CH_{AR}), 108.88 (CH_{AR}), 86.31 (C_1), 78.94 (C_5), 73.71 (C_3), 69.36 (C_4), 60.61 (C_6), 55.27 (C_2),54.36 (N$\underline{C}H_2$), 54.35 ($\underline{C}H_2$), 50.31 ($\underline{C}H_2$N), 31.68 (O$\underline{C}H_2$), 22.66 (2x$CH_2\underline{C}H_2$), 21.97 (CO$\underline{C}H_3$). HRMS (TOF), calculated for $C_{31}H_{38}N_9O_6$ (M+H): 632.2945; found: 632.2964.

^{1}H-NMR

¹³C-NMR

Synthesis of the mannose C2 linker asymmetric PTDZ derivative (**56**)

56

The general procedure was applied to the derivative **37** (51 mg, 0.095 mmol). The crude was purified by HPLC with method IV. A white solid was obtained as final product **56** (42.7 mg, 71%). ^{1}H NMR (500 MHz, CD$_3$OD) δ 8.81 (s, 1H, H$_{TRIAZOLE}$), 8.34 (s, 1H, H$_{TRIAZOLE}$), 8.31 (t, J=2.1 Hz, 1H,H$_{AR}$), 8.25 (t, J=1.7 Hz, 1H, H$_{AR}$), 7.82 (dt, J=7.7, 1.4 Hz, 1H, H$_{AR}$), 7.75 (dt, J=7.7, 1.4 Hz, 1H, H$_{AR}$), 7.54 (dt, J=7.5, 1.8 Hz, 1H, H$_{AR}$), 7.49–7.41 (m, 3H, 3xH$_{AR}$), 4.76 (d, J=1.6 Hz, 1H, H$_1$), 4.71–4.59 (m, 2H, 2xC$\underline{H}_2$N), 4.14 (ddd, J=10.8, 6.5, 4.1 Hz, 1H, H$_{SUGAR}$), 3.96–3.86 (m, 1H, H$_2$), 3.82–3.48 (m, 9H, 2xOC$\underline{H}_2$, 3xH$_{SUGAR}$, 2xNC$\underline{H}_2$, 2xC$\underline{H}_2$N), 3.34–3.28 (m, 1H, H$_{SUGAR}$), 3.18–3.06 (m, 2H, 2xNC$\underline{H}_2$), 2.92 (t, J=6.6 Hz, 2H, 2xCOC$\underline{H}_2$), 2.20–1.94 (m, 4H, 4xCH$_2$C$\underline{H}_2$). ^{13}C NMR (126 MHz, CD$_3$OD) δ 170.20 (NH$\underline{C}$O), 149.07 (C$_{AR}$), 148.40 (C$_{AR}$), 141.04 (C$_{AR}$), 138.56 (C$_{AR}$), 132.47 (C$_{AR}$), 132.00 (C$_{AR}$), 131.33 ($\underline{C}$H$_{AR}$), 130.69 ($\underline{C}$H$_{AR}$), 126.75 ($\underline{C}$H$_{AR}$), 126.45 ($\underline{C}$H$_{AR}$), 124.15 ($\underline{C}$H$_{AR}$), 123.28 ($\underline{C}$H$_{AR}$), 120.82 ($\underline{C}$H$_{TRIAZOLE}$), 120.49 ($\underline{C}$H$_{TRIAZOLE}$), 116.67 ($\underline{C}$H$_{AR}$), 112.74 ($\underline{C}$H$_{AR}$), 101.76 (C$_1$), 75.07, 72.55, 71.94, 68.29 (C$_2$, C$_3$, C$_4$, C$_5$), 66.82 (C$_6$), 62.73 (C$_7$), 55.53 (2xN$\underline{C}$H$_2$), 52.12 (CH$_2$N), 51.46 (C$_8$), 32.94 (CO$\underline{C}$H$_2$), 24.02 (2x$\underline{C}$H$_2$CH$_2$). HRMS (TOF), calculated for C$_{31}$H$_{39}$N$_8$O$_7$ (M+H): 635.2942; found: 635.2961.

¹H-NMR

¹³C-NMR
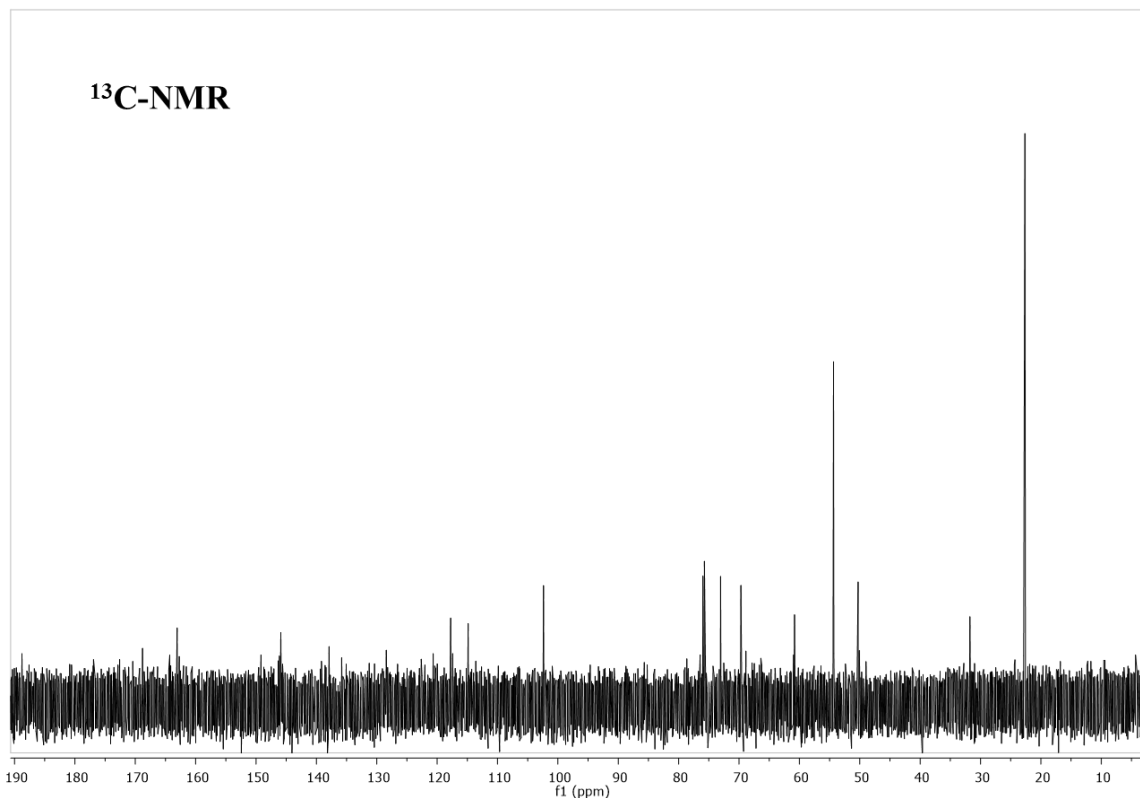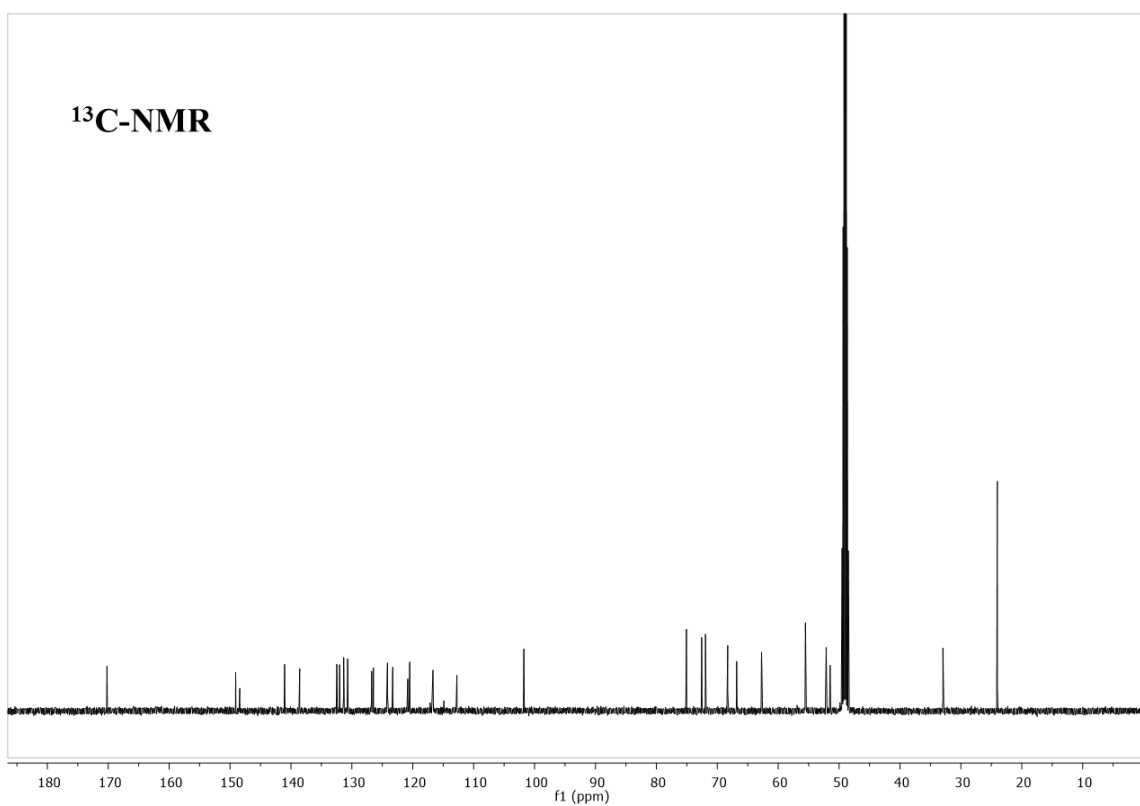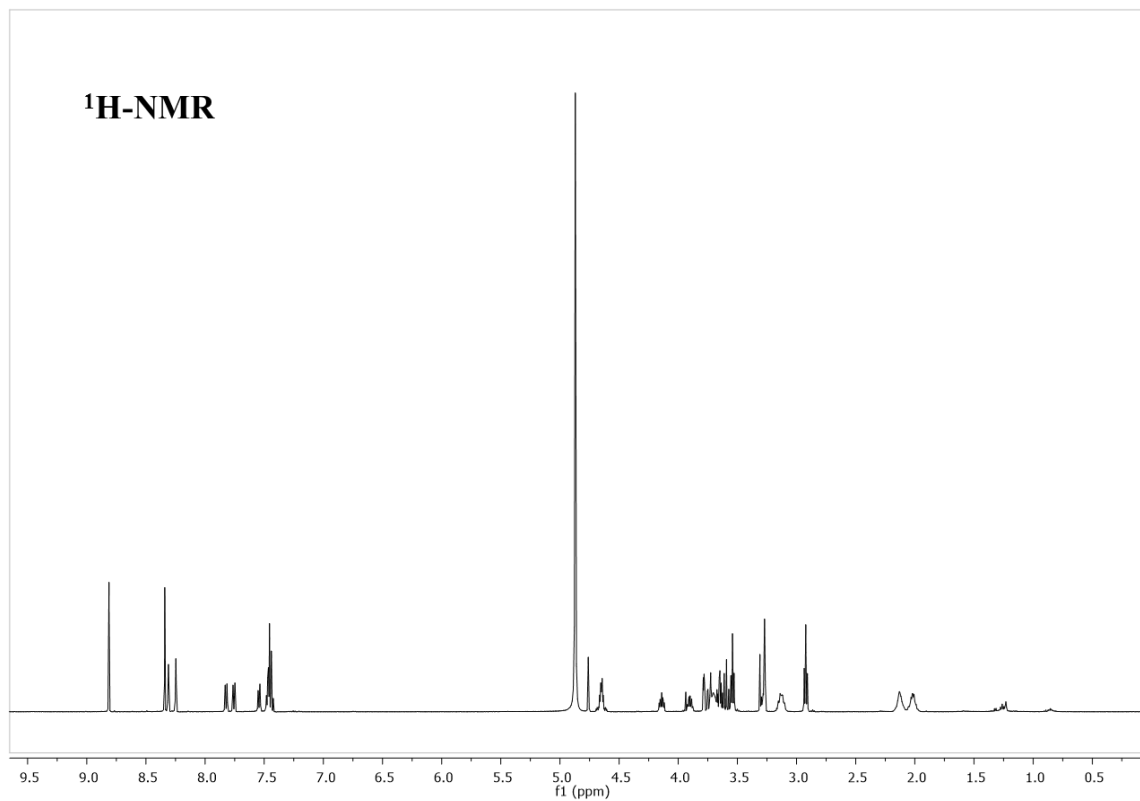
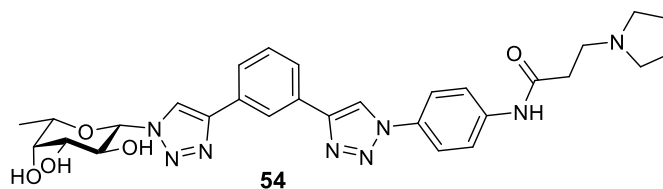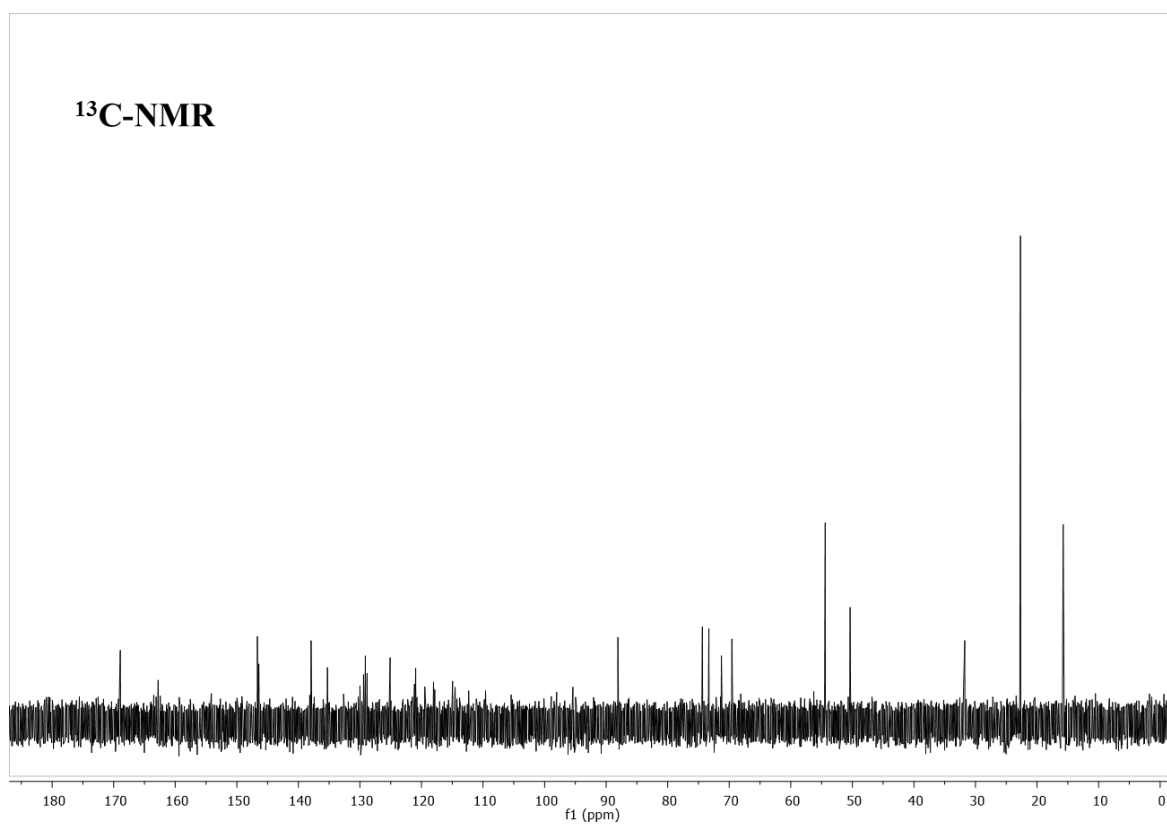Synthesis of the lactose asymmetric PTDZ derivative **57**:
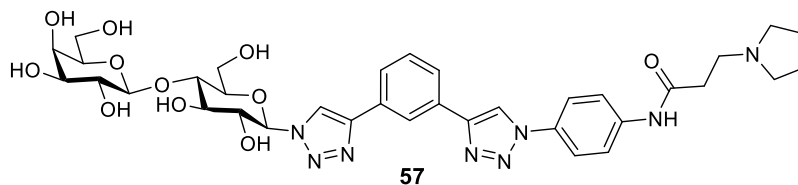
The general procedure was applied to the derivative **38** (100 mg, 0.127 mmol). The crude was purified by HPLC with method III. A red solid was obtained as final product **57** (36 mg, 38%). ^{1}H NMR (500 MHz, CD$_3$OD) δ 8.81 (s, 1H, H$_{TRIAZOLE}$), 8.62 (s, 1H, H$_{TRIAZOLE}$), 8.35 (dt, J=5.8, 1.9 Hz, 2H, 2xH$_{AR}$), 7.82 (ddt, J=26.3, 7.9, 1.3 Hz, 2H, 2xH$_{AR}$), 7.59–7.45 (m, 4H, 4xH$_{AR}$), 5.75 (d, J=9.2 Hz, 1H, H$_{1A}$), 4.45 (d, J=7.7 Hz, 1H, H$_{1B}$), 4.08 (t, J=8.8 Hz, 1H, H$_{2A}$), 4.01–3.91 (m, 2H, H$_6$, H$_{6'}$), 3.87–3.77 (m, 5H, H$_{3A}$, H$_{4A}$, H$_5$, H$_6$, H$_{6'}$), 3.77–3.70 (m, 2H, 2xNC$\underline{H}_2$), 3.69–3.50 (m, 6H, 2xC$\underline{H}_2$N, H$_{2B}$, H$_{3B}$, H$_{4B}$, H$_{5B}$), 3.30 (p, 3H, MeOH), 3.17 (tt, J=10.9, 5.5 Hz, 2H, 2xNC$\underline{H}_2$), 2.95 (t, J=6.6 Hz, 2H, 2xCOC$\underline{H}_2$), 2.25–1.98 (m, 4H, 4xC$\underline{H}_2$CH$_2$). ^{13}C NMR (126 MHz, CD$_3$OD) δ 170.20 ($\underline{C}$ONH), 149.08 (C$_{AR}$), 148.45 (C$_{AR}$), 141.08 (C$_{AR}$), 138.60 (C$_{AR}$), 132.30 (C$_{AR}$), 132.07 (C$_{AR}$), 131.37 (CH$_{AR}$), 130.75 (CH$_{AR}$), 126.72 (CH$_{AR}$), 126.64 (CH$_{AR}$), 123.95 (CH$_{AR}$), 121.78 ($\underline{C}$H$_{TRIAZOLE}$), 120.85 ($\underline{C}$H$_{TRIAZOLE}$), 120.47 (CH$_{AR}$), 116.69 (CH$_{AR}$), 112.80 (CH$_{AR}$), 105.15 (C$_{1B}$), 89.48 (C$_{1A}$), 79.88, 79.61, 77.18, 76.86 (C$_{3A}$, C$_{4A}$, C$_{4B}$, C$_5$), 74.83 (C$_{3B}$), 73.78 (C$_{2A}$), 72.57 (C$_{2B}$), 70.33(C$_5$), 62.59 (C$_6$), 61.60 (C$_6$), 55.56 (NCH$_2$), 52.16 ($\underline{C}$H$_2$N), 32.89 (CO$\underline{C}$H$_2$), 24.03 (2xC$\underline{H}_2$CH$_2$). HRMS (TOF), calculated for C$_{35}$H$_{45}$N$_8$O$_{11}$ (M+H): 753.3208; found: 753.3234.

^{13}C-NMR
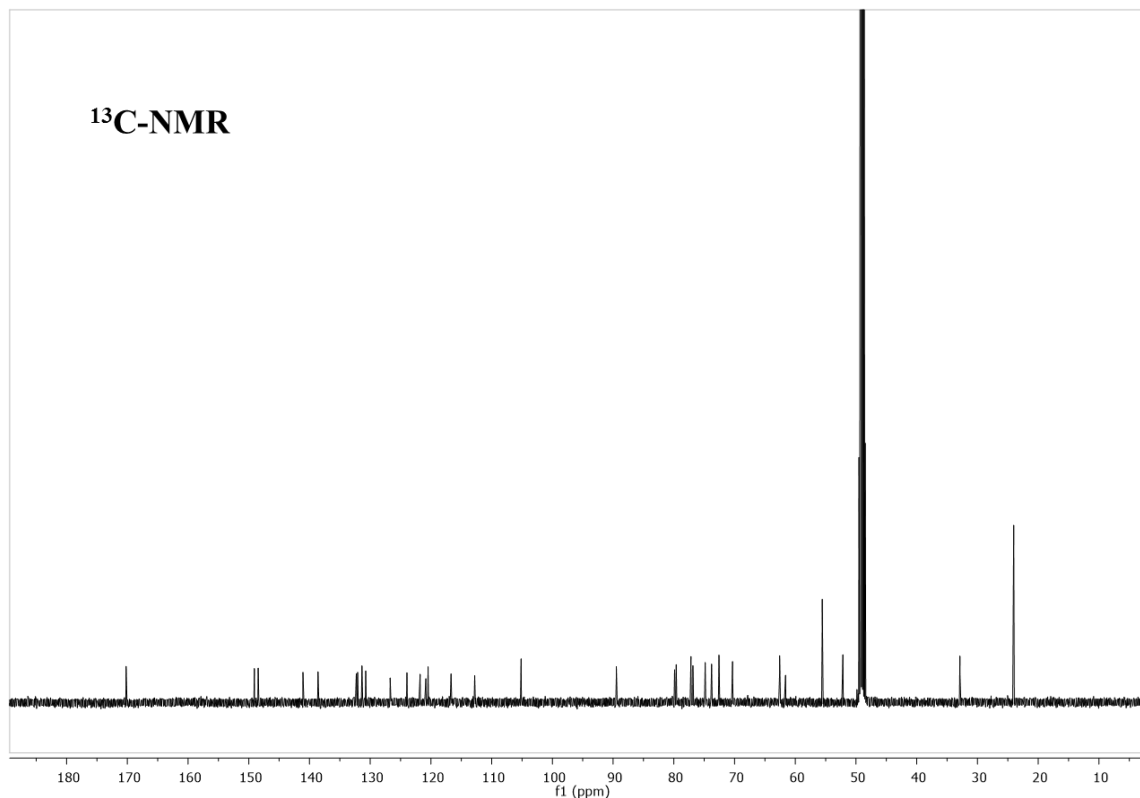
Synthesis of the ethanol asymmetric PTDZ derivative **58**:

58
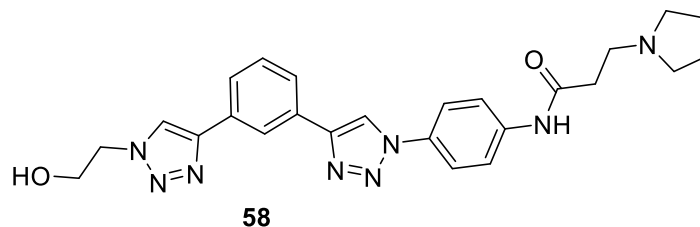
The general procedure was applied to the derivative **41** (81 mg, 0.380 mmol). With this product it was not necessary to carry out the deprotection with MeO$^-$Na$^+$/MeOH. The crude was purified by HPLC with method V. A white solid was obtained as final product **58** (35 mg, 33%). ^{1}H NMR (300 MHz, CD$_3$OD) δ 8.83 (s, 1H, H$_{TRIAZOLE}$), 8.38 (s, 1H, H$_{TRIAZOLE}$), 8.37–8.30 (m, 2H, 2xH$_{AR}$), 7.82 (ddt, J=15.9, 7.9, 1.3 Hz, 2H, 2xH$_{AR}$), 7.60–7.43 (m, 4H, 4xH$_{AR}$), 4.55 (t, J=5.2 Hz, 2H, 2xNC$\underline{H}_2$CH$_2$OH), 4.10–3.91 (m, 2H, 2xC$\underline{H}_2$OH), 3.56 (t, J=6.5 Hz, 2H, 2C$\underline{H}_2$N), 3.29 (p, J=1.6 Hz, 2H, 2xNC$\underline{H}_2$), 3.16 (d, J=.2 Hz, 2H, 2xNC$\underline{H}_2$), 2.94 (t, J=6.5 Hz, 2H, 2xCOC$\underline{H}_2$), 2.21–1.97 (m, 4H, 4xC$\underline{H}_2$CH$_2$). ^{13}C NMR (126 MHz, CD$_3$OD) δ 170.17 ($\underline{C}$ONH), 149.16 (C$_{AR}$), 148.22 (C$_{AR}$), 141.12 (C$_{AR}$), 138.63 (C$_{AR}$), 132.61 (C$_{AR}$), 132.07 (C$_{AR}$), 131.34 (CH$_{AR}$), 130.72 (CH$_{AR}$), 126.67 (CH$_{AR}$), 126.45 (CH$_{AR}$), 123.89 (CH$_{AR}$), 123.27 (CH$_{AR}$), 120.83 (CH$_{AR}$), 120.45 (CH$_{AR}$), 116.67 (CH$_{AR}$), 112.81 (CH$_{AR}$), 61.61 ($\underline{C}$H$_2$OH), 55.53 (N$\underline{C}$H$_2$), 54.09 (N$\underline{C}$H$_2$CH$_2$OH), 52.14 ($\underline{C}$H$_2$N), 32.90 (CO$\underline{C}$H$_2$), 24.03 (2x$\underline{C}$H$_2$CH$_2$). HRMS (FAB$^+$) calculated for C$_{25}$H$_{29}$N$_8$O$_2$ (M+H): 473.2413; found: 473.2433.

¹H-NMR

¹³C-NMR
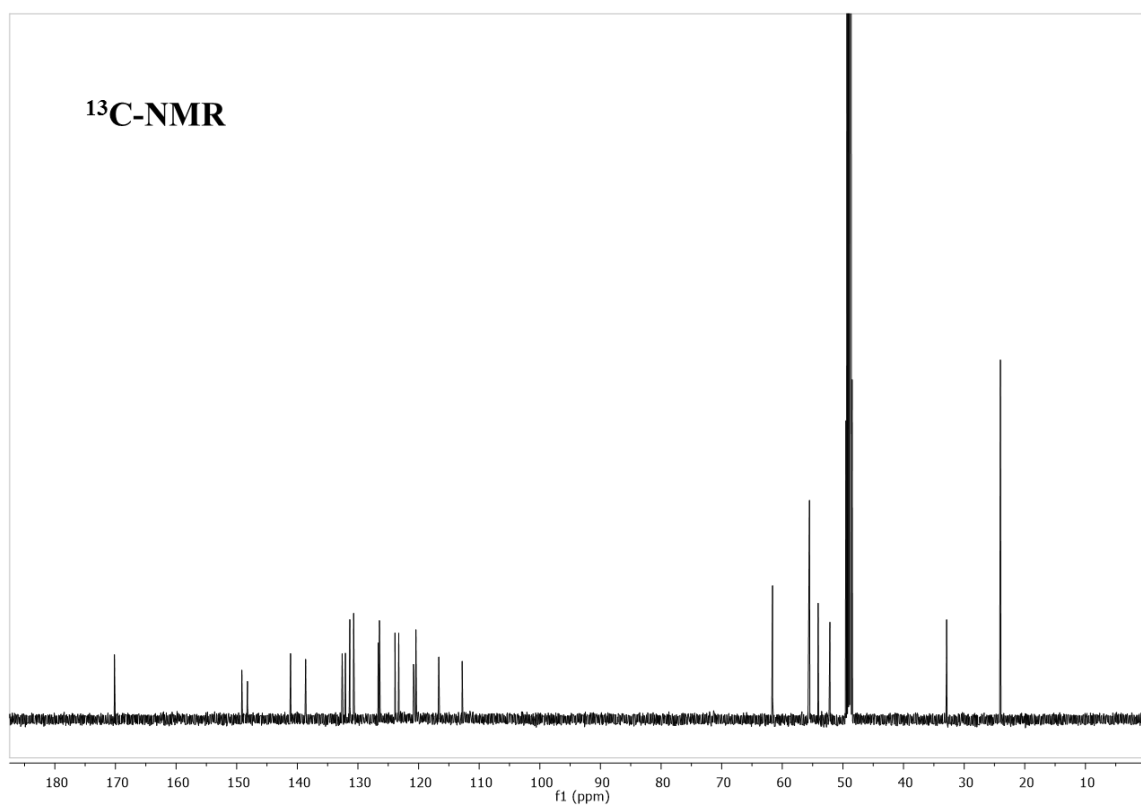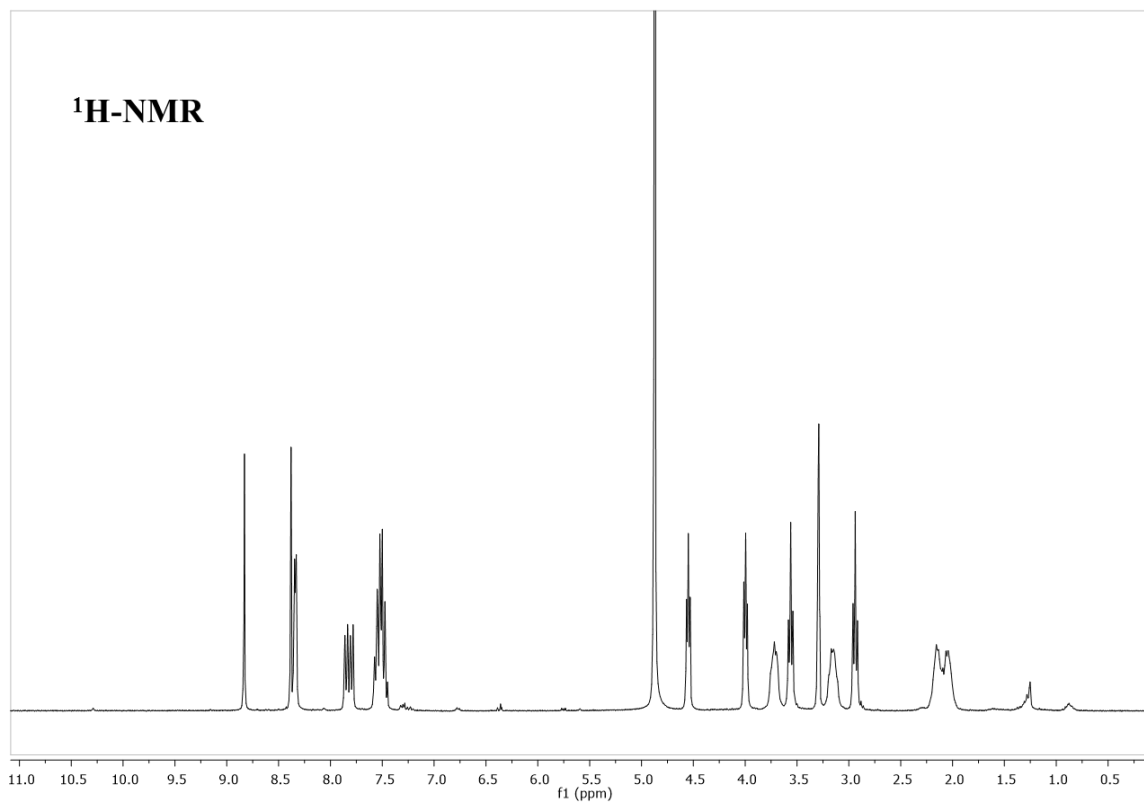
Synthesis of the maltose asymmetric *p*-PTDZ derivative **59**:
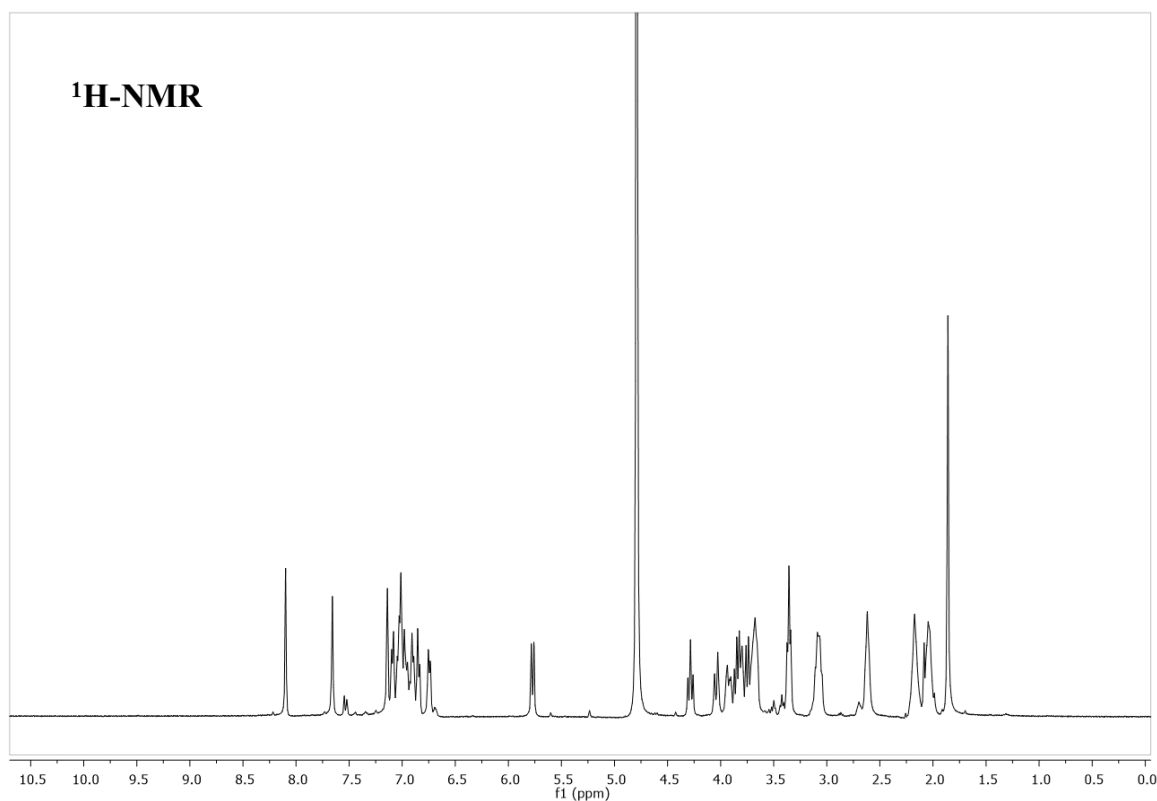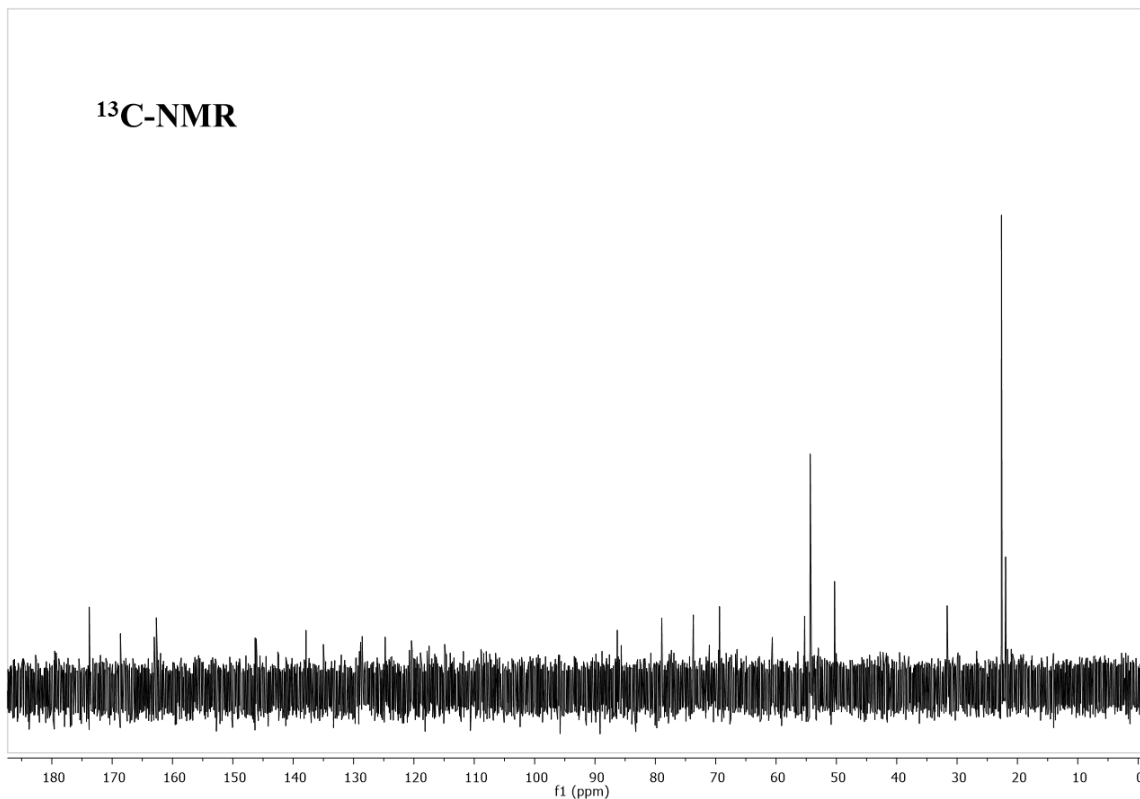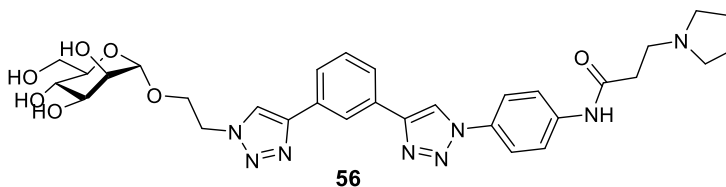

59
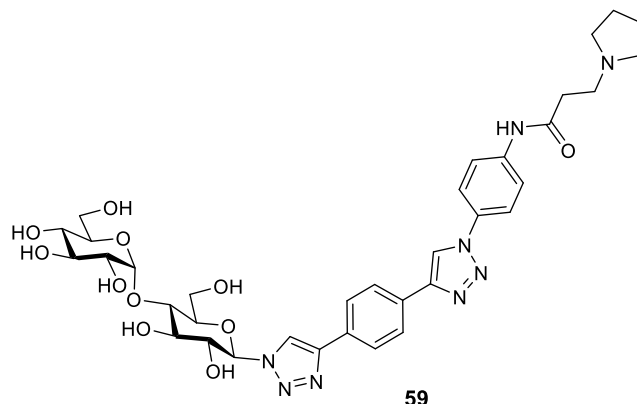
The general procedure was applied to the derivative **39** (31.1 mg, 0.039 mmol). The crude was purified by HPLC with method III. A white solid was obtained as final product **59** (9.7 mg, 33%). ^{1}H NMR (500 MHz, CD$_3$OD) δ 8.92 (s, 1H, H$_{TRIAZOLE}$), 8.64 (s, 1H, H$_{TRIAZOLE}$), 8.43 (t, J=2.1 Hz, 1H, H$_{AR}$), 8.05–7.94 (m, 4H, 4xH$_{AR}$), 7.66–7.52 (m, 3H, 3xH$_{AR}$), 5.73 (d, J=9.1 Hz, 1H, H$_{1A}$), 5.29 (d, J=3.8 Hz, 1H, H$_{1B}$), 4.06 (t, J=9.2 Hz, 1H, H$_{2A}$), 3.96–3.86 (m, 4H, H$_{3A}$, 3xH$_{SUGAR}$), 3.83 (t, J=9.1 Hz, 1H, H$_{SUGAR}$), 3.79–3.73 (m, 2H, 2xNC$\underline{H}_2$), 3.73–3.64 (m, 2H, H$_{3B}$, H$_{5B}$), 3.61 (t, J=6.5 Hz, 2H, 2xC$\underline{H}_2$N), 3.52 (dd, J=9.7, 3.7 Hz, 1H, H$_{2B}$), 3.33(m, 3H, H$_{4B}$, 2xH$_{SUGAR}$) 3.25–3.15 (m, 2H, 2xNC$\underline{H}_2$), 2.98 (t, J=6.5 Hz, 2H, 2xCOC$\underline{H}_2$), 2.14 (dd, J=59.9, 6.5 Hz, 4H, 4xC$\underline{H}_2$CH$_2$). ^{13}C NMR (126 MHz, CD$_3$OD) δ 170.26 ($\underline{C}$ONH), 149.09 (C$_{AR}$), 148.37 (C$_{AR}$), 141.14 (C$_{AR}$), 138.73 (C$_{AR}$), 131.74 (C$_{AR}$), 131.41 (C$_{AR}$), 131.33 (CH$_{AR}$), 127.27 (CH$_{AR}$), 121.63 (CH$_{AR}$), 120.41 (CH$_{AR}$), 116.75 (CH$_{AR}$), 112.91 (CH$_{AR}$), 102.97 (C$_{1B}$), 89.61 (C$_{1A}$), 80.36, 79.69, 78.24, 75.10, 74.92, 74.20, 73.73, 71.54 (C$_{2A}$, C$_{2B}$, C$_{3A}$, C$_{3B}$, C$_{4A}$, C$_{4B}$, C$_{5A}$, C$_{5B}$), 62.78 (C$_6$), 61.88 (C$_6$), 55.56 (N$\underline{C}$H$_2$), 52.21 ($\underline{C}$H$_2$N), 32.82 (CO$\underline{C}$H$_2$), 24.04 (2x$\underline{C}$H$_2$CH$_2$). HRMS (TOF), Calculated for C$_{35}$H$_{45}$N$_8$O$_{11}$ (M+H): 753.3208; found: 753.3234.

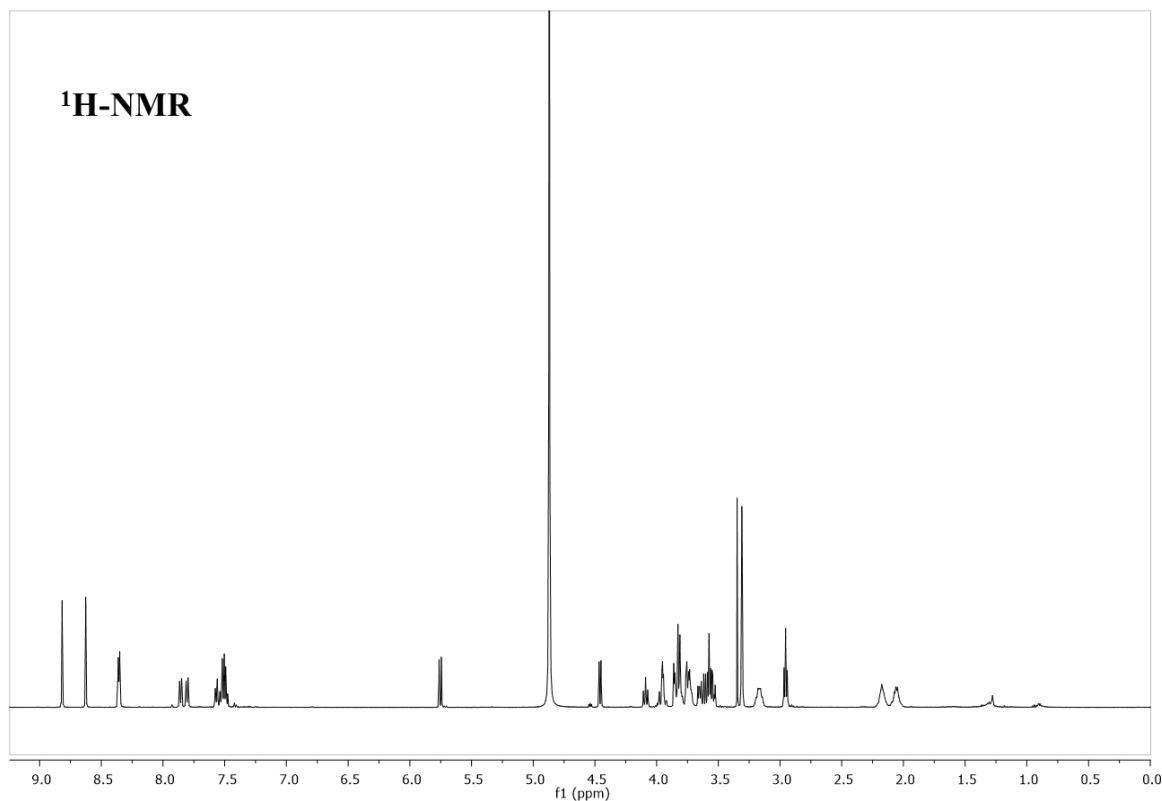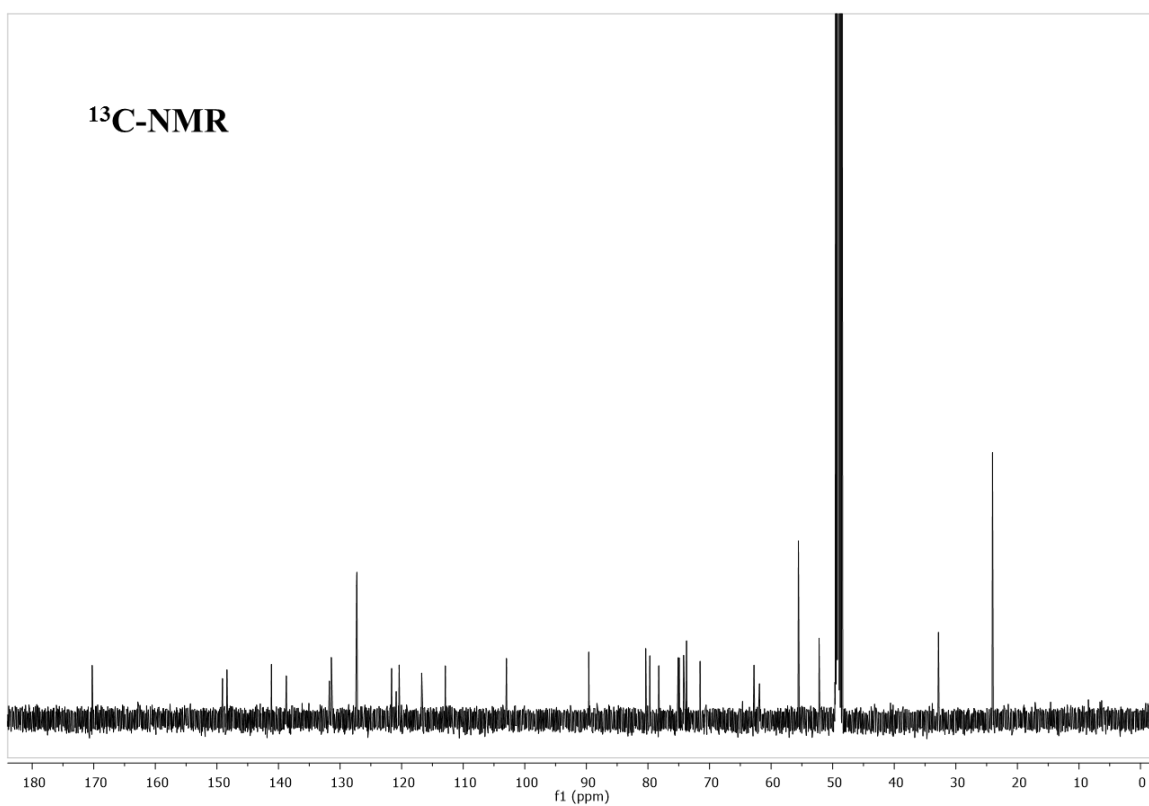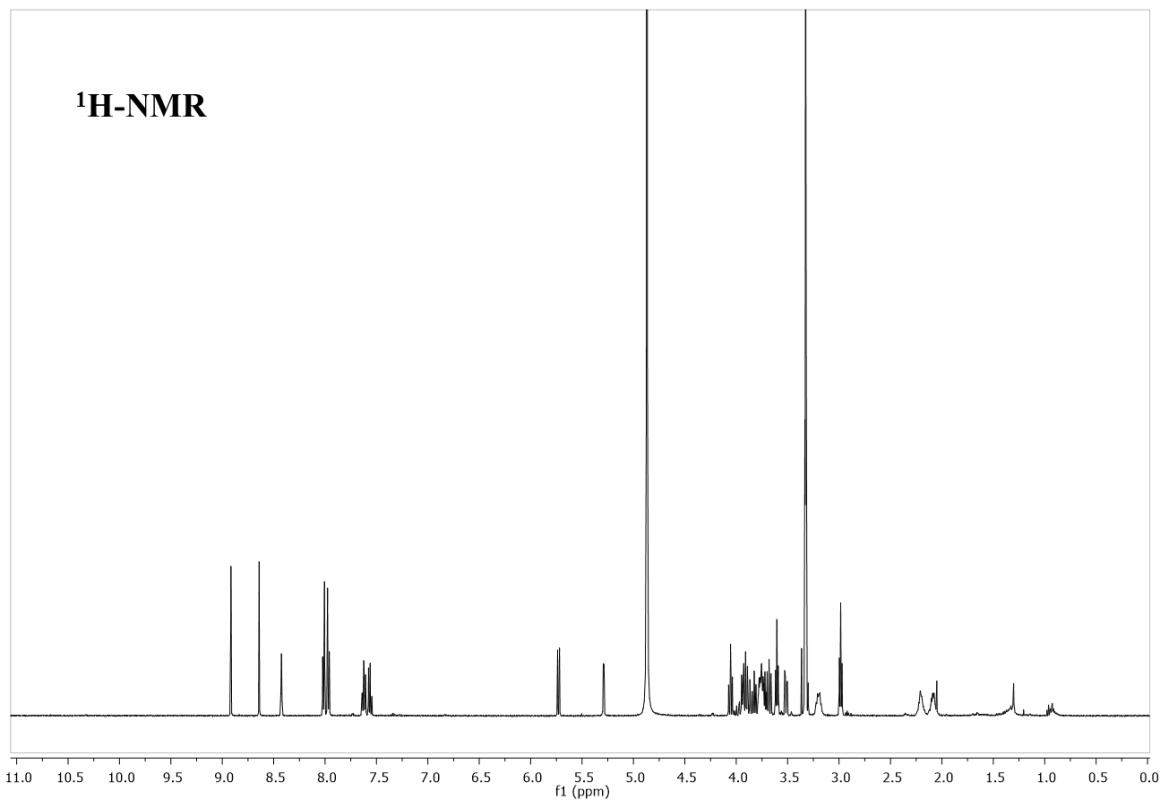Synthesis of the glucose asymmetric *p*-PTDZ derivative **60**:

60
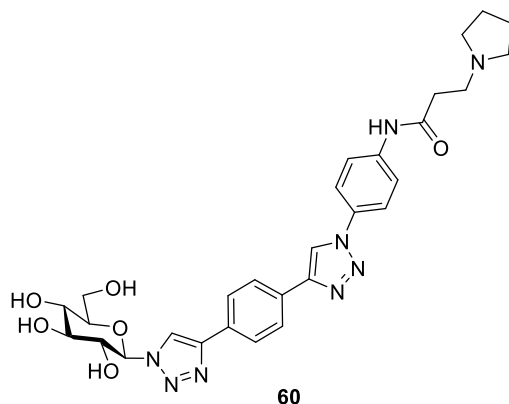
The general procedure was applied to the derivative **40** (100 mg, 0.200 mmol). The crude was purified by HPLC with method III. A white solid was obtained as final product **60** (15.3 mg, 13%). ^{1}H NMR (400 MHz, CD$_3$OD) δ 8.91 (s, 1H, H$_{TRIAZOLE}$), 8.63 (s, 1H, H$_{TRIAZOLE}$), 8.42 (t, J=2.1 Hz, 1H, H$_{AR}$), 8.05–7.93 (m, 4H, 4xH$_{AR}$), 7.67–7.52 (m, 3H, 3xH$_{AR}$), 5.70 (d, J=9.2 Hz, 1H, H$_1$), 4.07–3.91 (m, 2H, H$_2$, H$_6$), 3.86–3.50 (m, 8H, H$_{6'}$, H$_3$, H$_4$, H$_5$, 2xC$\underline{H}_2$N, 2xNC$\underline{H}_2$), 3.18 (d, J=20.8 Hz, 2H, 2xNC$\underline{H}_2$), 2.98 (t, J=6.5 Hz, 2H, 2xCOC$\underline{H}_2$), 2.15 (d, J=45.2 Hz, 4H, 4xCH$_2$CH$_2$). ^{13}C NMR (101 MHz, CD$_3$OD) δ 170.26 ($\underline{C}$ONH), 149.08 (C$_{AR}$), 148.33 (C$_{AR}$), 141.14 (C$_{AR}$), 138.71 (C$_{AR}$), 131.74 (C$_{AR}$), 131.41 (C$_{AR}$), 127.25 (CH$_{AR}$), 121.68 (CH$_{AR}$), 120.87 (CH$_{AR}$), 120.39 (CH$_{AR}$), 116.72 (CH$_{AR}$), 112.87 (CH$_{AR}$), 89.78 (C$_1$), 81.21 (C$_5$), 78.49 (C$_3$), 74.11 (C$_2$), 70.95 (C$_4$), 62.49 (C$_6$), 55.56 ($\underline{C}$H$_2$N), 52.20 (N$\underline{C}$H$_2$), 32.82 (CO$\underline{C}$H$_2$), 24.04 (2x$\underline{C}$H$_2$CH$_2$). HRMS (TOF), calculated for C$_{29}$H$_{35}$N$_8$O$_6$ (M+H): 591.2680; found: 591.2704.

¹H-NMR

¹³C-NMR
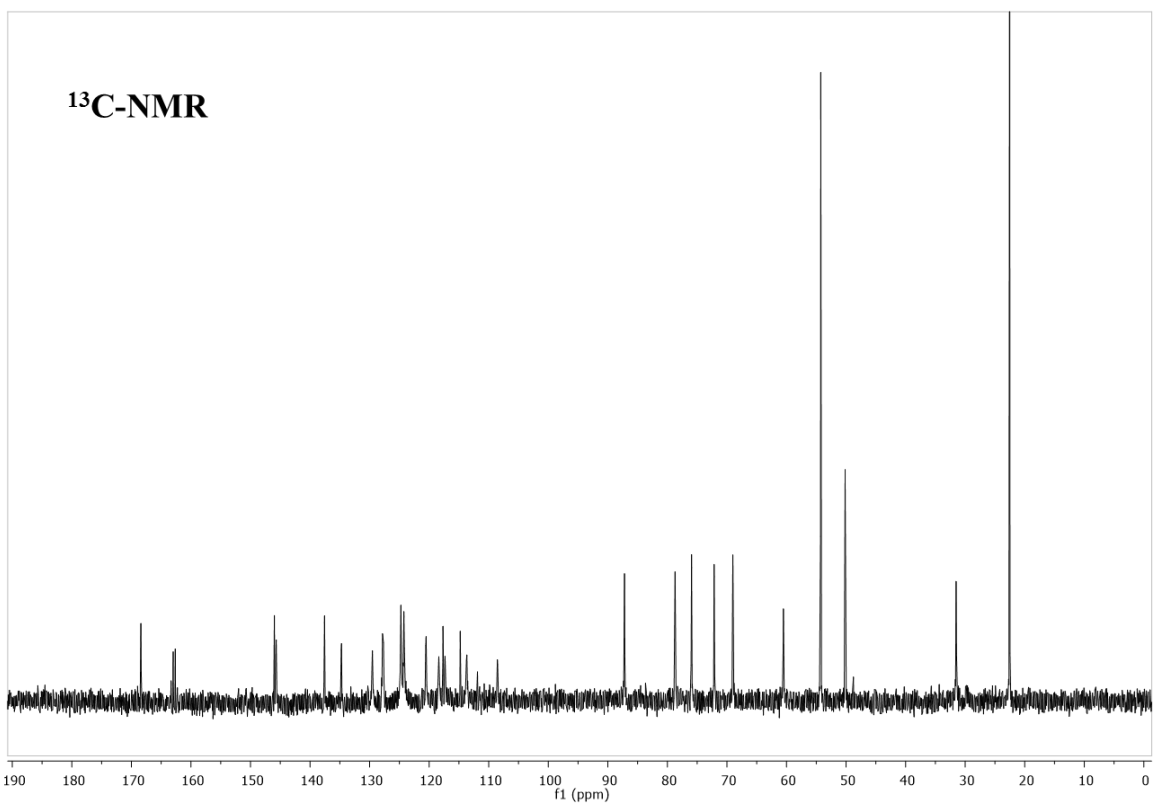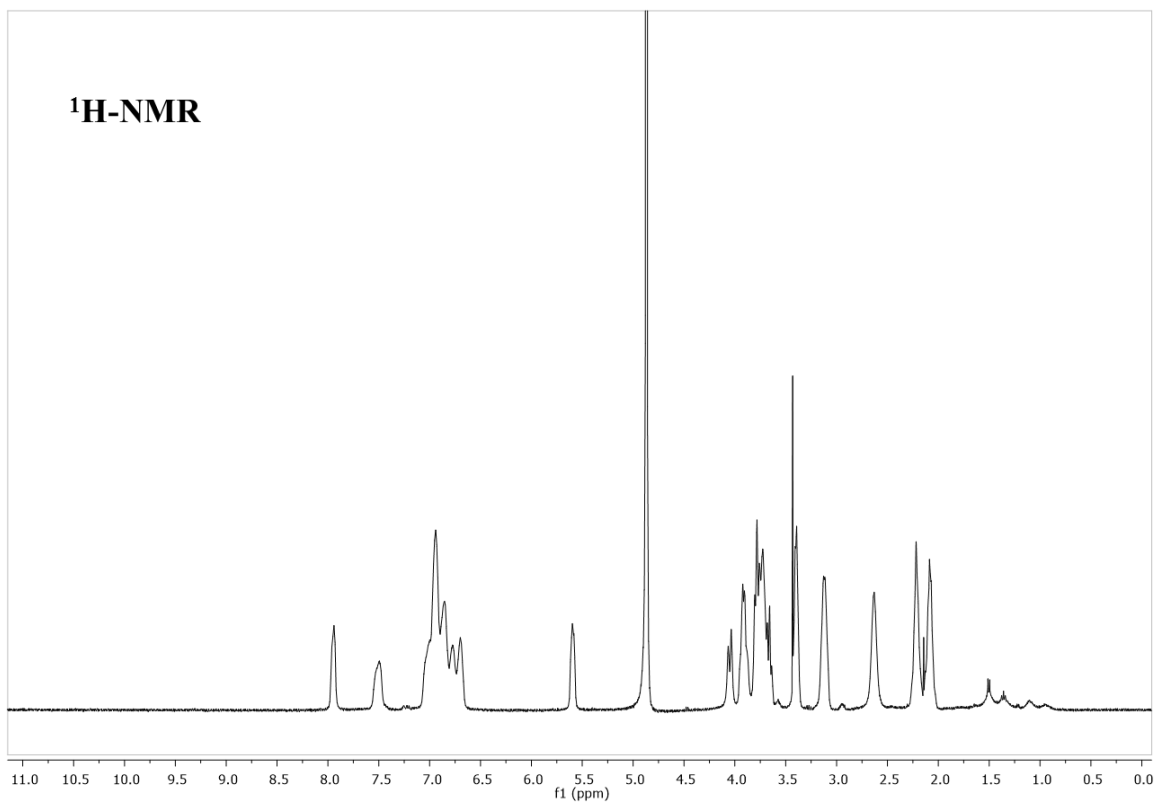
5.2. Biophysical assays.

5.2.1. OLIGONUCLEOTIDES

Lyophilized 200 nmol scale oligonucleotides were purchased from Eurogentec (Seraing, Belgium) and used without further purification (Reverse-Phase Cartridge Gold or PAGE purification), except for titrations and NOESY NMR experiments, when they were purified by ultrafiltration with Vivaspin® 15 TURBO centrifugal concentrators. Stocks were prepared in MilliQ water and stored at -20°C. Concentrations were validated by UV-spectrometry absorption analysis on an Uvikon XS spectrophotometer using quartz cuvettes (Hellma, France) with a path length l=10 mm. Absorbance was measured in the 220-335 nm range where the extinction coefficients were given to its maximum absorption. The final concentration was calculated by Beer-Lambert law application.

5.2.2. FRET-MELTING ASSAY.

Experiments were carried out on a Mx3005P Stratagene real-time PCR by using 96 well plates loaded with 25 μL per well of 200 nM oligonucleotide and varying ligand concentrations. Buffer conditions for all experiments except RNA and HIV DNA were LiCaco 10 mM, 10 mM KCl and 90 mM LiCl (pH 7.2). RNA sequence experiments buffer conditions were LiCaco 10 mM, 1 mM KCl, and 99 mM LiCl (pH 7.2) whilst HIV DNA sequences buffer conditions were LiCaco10 mM and 100 mM KCl (pH 7.2). 6-carboxyfluorescein (FAM) and 6-carboxytetramethylrhodamine (TAMRA) were respectively used as fluorophore donor and acceptor to label the oligonucleotide sequence. The sequences were annealed by heating to 90°C for two minutes and chilling them in ice. The experiment temperatures were maintained 5 min at 25 °C, and then increased 1°C /minute to 95°C. The excitation and detection wavelengths used were 492 and 516 nm, respectively. Each experiment was measured by duplicate on two separate plates and the final analysis of the data were carried out using OriginPro 9. The T_m were determined from the normalized curves whilst ΔT_m were calculated as the difference between the T_m of the oligonucleotide in water and the sample containing the oligonucleotide with the ligand.

5.2.3. CIRCULAR DICHROISM (CD) STUDIES.

CD spectra were recorded using a JASCO J-815 CD spectrometer equipped with a Peltier heating unit. Experiments were performed in isothermal conditions (20 ℃) with 1 cm quartz cuvettes and a sample volume of 0.5 mL. DNA telomeric buffer conditions were 10 mM LiCaCo and 100 mM KCl or 100 mM NaCl (pH 7.2) whilst HIV buffer conditions were 20 mM potassium phosphate with 70 mM KCl. 3 μM oligonucleotides samples in the corresponding buffer were annealed at 90°C for 2 min and then left to cool overnight in the heating block. Parameters of 1-nm bandwidth, 3 accumulations, and a rate of 100 nm/min were used, while readings were measured in the 220-350 nm wavelength range. Table **2.7** describes the DNA sequences used and their origins.

Oligonucleotide name	Sequence	Origin	Oligonucleotide name
22AG	AGGGTTAGGGTTAGGGTTAGGG	Human telomeric sequence	22AG
24TTA	TTGGGTTAGGGTTAGGGTTAGGGA	Human telomeric sequence	24TTA
HIVADR	AGGGAGGTGTGGCCTGGGCGGG	HIV-PRO2	HIVADR
HIV321	TGGCCTGGGCGGGACTGGG	HIV-PRO1	HIV321

Table 2.9. Description of the employed sequences in the CD experiments.

For titration experiments, varying equivalents of the compounds were added step-wise. Initially, buffer samples were measured alone. The oligonucleotide sequence sample was then read, followed by increasing equivalents of the compound. The average measurements were calculated from triplicate readings and the background buffer average was subtracted.

For CD melting experiments, solutions with 3µM 22AG and 24TTA in their buffer solutions (10 mM LiCaCo pH 7.2 + 100 mM NaCl, and 10 mM LiCaCo pH 7.2 + 10 mM KCl + 90 mM LiCl, respectively) were annealed at 90 °C for 5 min and immediately chilled in ice. 3 equivalents of compound **53** were added and samples were kept at 4°C. 220-335 nm range was used for spectra recording at 100 nm/min speed with a 2 nm band width and 3 successive accumulations per spectrum. Samples were heated from 4 to 90°C by 0.4°C/min whilst reading each 2°C and extracting the results at 295 nm for 22AG and 290 nm for 24TTA. Data CD signals (mdeg) at those wavelengths representing each temperature were treated with Origin 9.0 and used to calculate the curve in which T_m was extracted. ΔT_m was calculated by subtracting the T_m of the oligonucleotide sequence to the oligonucleotide sequence with the ligand.

5.2.4. SURFACE PLASMON RESONANCE (SPR) TITRATIONS.

SPR experiments were performed at 23 °C on a BIAcore T200 instrument operating with the BIAcoreT100 evaluation control Software 2.0.1. Firstly, the chip surface was activated with amine coupling NHS (N-hydroxysuccinimide) and EDC (ethyl (dimethylaminopropyl) carbodiimide), in which Streptavidin was immobilized. 40 nM of the 5' biotinylated oligonucleotide was linked to the Streptavidin in 20 mM potassium phosphate and 70 mM potassium chloride buffer (pH 6.9) supplemented with 0.05 % (V/V) of Tween®, to obtain a response of 400-500 RU (response unit). Two chips were employed to configure two different 5-channels systems. The system 1 was composed of: a) streptavidin free; b) hairpin sequence; c) 22AG mutated sequence; d) 22AG sequence channel; and e) running buffer channels. The system 2 was constituted by: a) streptavidin free; b) HIV321 sequence; c) HIV32 sequence; d) c-myc sequence; and e) running buffer channels. Biotinylated sequences are described in table **2.8**. Ligand solutions at seven increasing concentrations were run over the chips at 30 µL/min flow rate. The selected asymmetric ligands for these titrations were **53** (glcC2-*m*-PTDZ), **52** (malt-*m*-PTDZ) **54** (fuc-*m*-PTDZ), **59** (malt-*p*-PTDZ), **58** (aglycone-PTDZ) and **1** (control compound) at varying concentrations of 20, 16, 12, 8, 4 ,2 ,0.5 µM. Exception was **1**, in which -due to a major

compound attachment over the Streptavidin surface- 4, 2 ,1, 0.5, 0.25, 0.125 µM concentrations were used.

The sensorgrams were double-referenced using BiaEval 4.1 software (Biacore) to remove instrument noise and buffer contribution to the signal. Response in equilibrium (Req) was recorded and plotted against ligand concentration for all ligands after correcting with the response unit (RU) of the C-myc sequence. Titration curves (figures **2.34-2.39**) were fitted using Origin® 9.0 with two non-sigmoidal model according to a specifying binding model and a modified model.

Oligonucleotide name	Sequence	Origin	Final Response (RU)
22AG	biotin-5'-AGGGTTAGGGTTAGGGTTAGGG	Human telomere	420
22AGmutated	biotin-5'-AGGTTTAGTGTTATGTTTAGGT	Mutated human telomere	480
Hairpin	Biotin-5'-TATAGCTATA-HEG-TATAGCTATA	Duplex sequence	415
HIV321	biotin-5'-TTGGCCTGGGCGGGACTGGGA	HIV-PRO1	440
HIV32	biotin-5'-CAGGGAGGCGTGGCCTGGGCGGGA	HIV-PRO1	440
C-myc	Biotin-5'-TGAGGGTGGGTAGGGTGGGTAA	c-myc promoter	490

Table 2.10. Sequences assayed in SPR titration experiments.
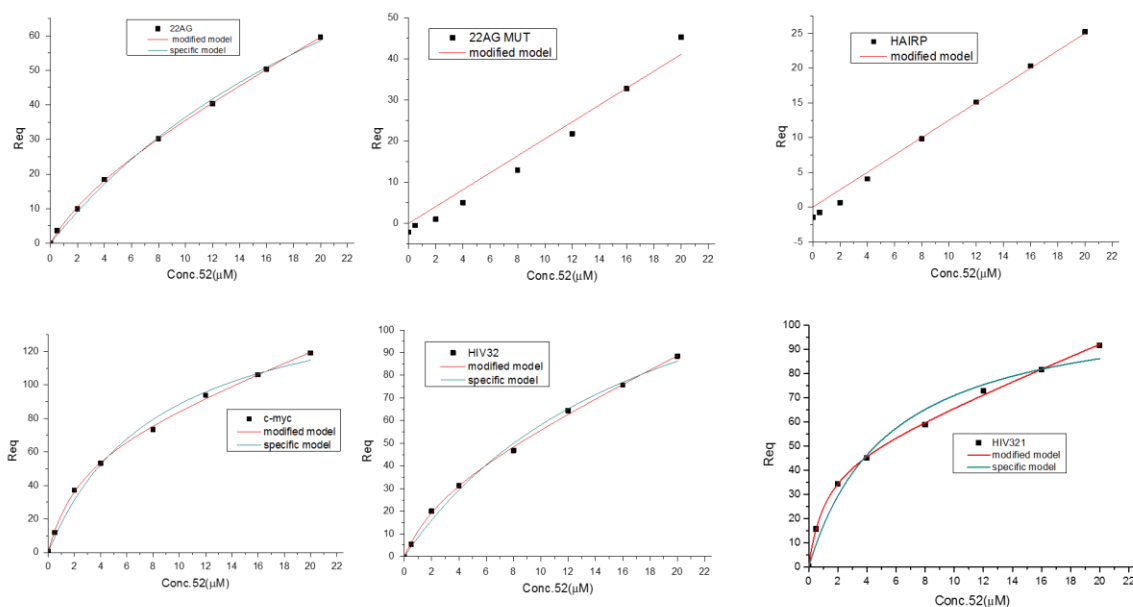
Figure 2.34. SPR fitted titration curves for **52**.
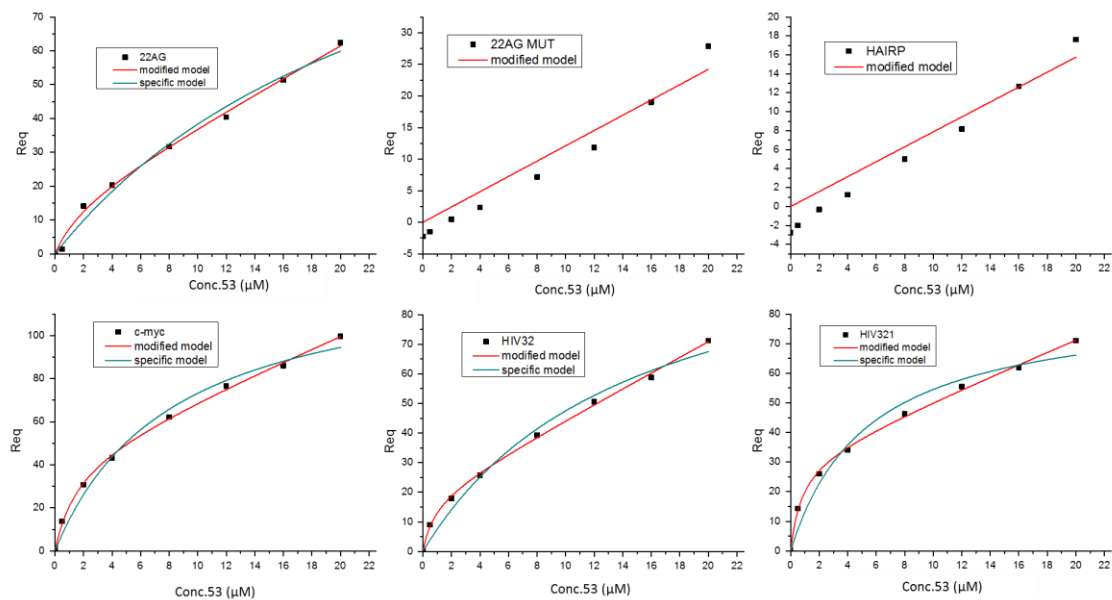
Figure 2.35. SPR fitted titration curves for **53**.
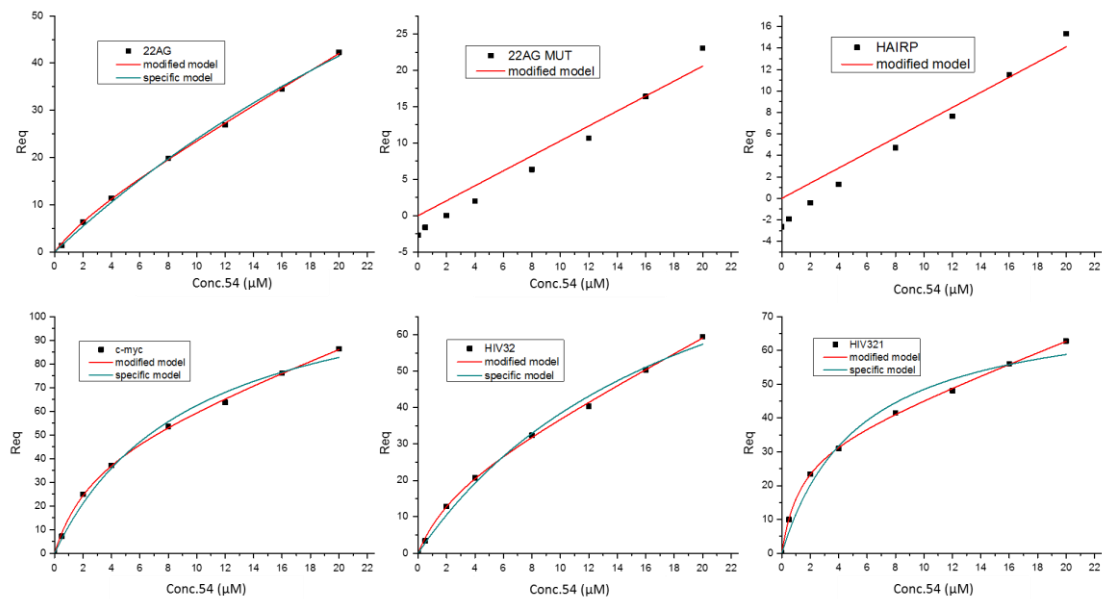
Figure 2.36. SPR fitted titration curves for **54**.
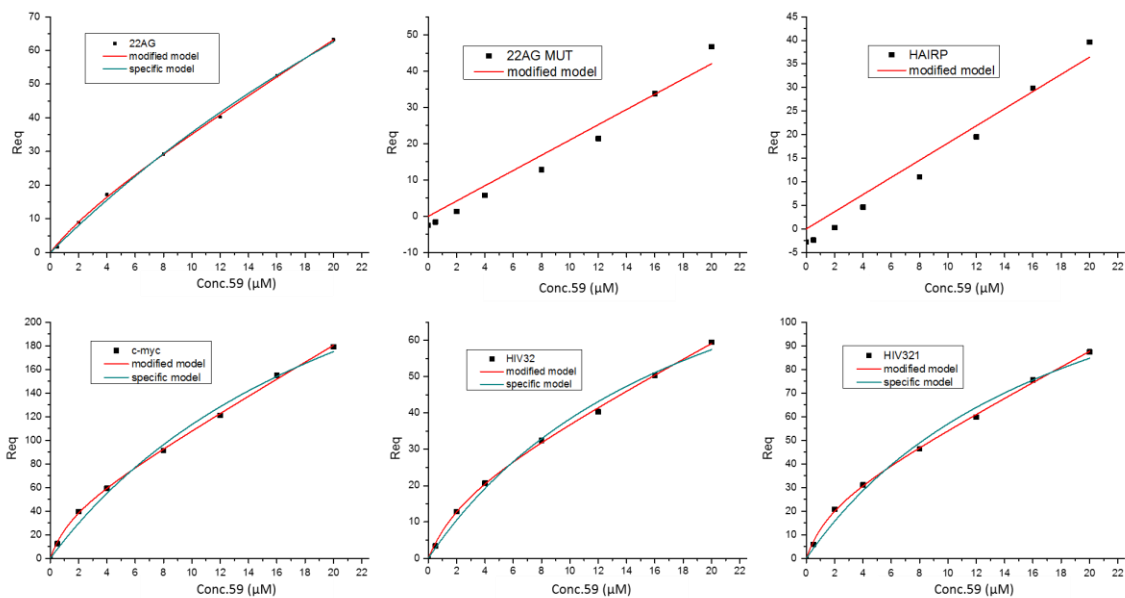
Figure 2.37. SPR fitted titration curves for **59**.
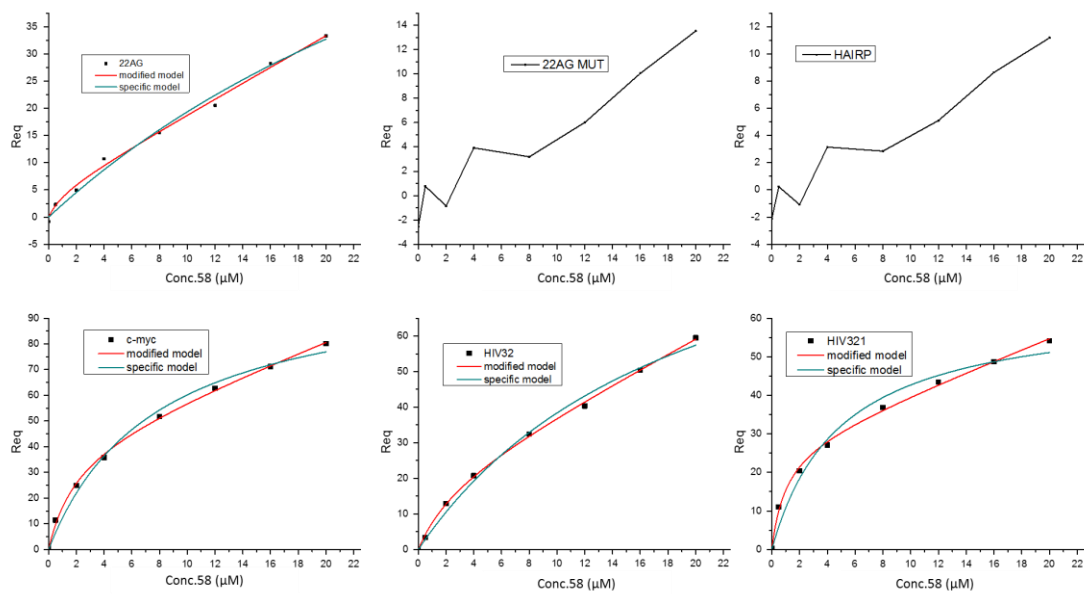
Figure 2.38.SPR fitted titration curves for **58**.
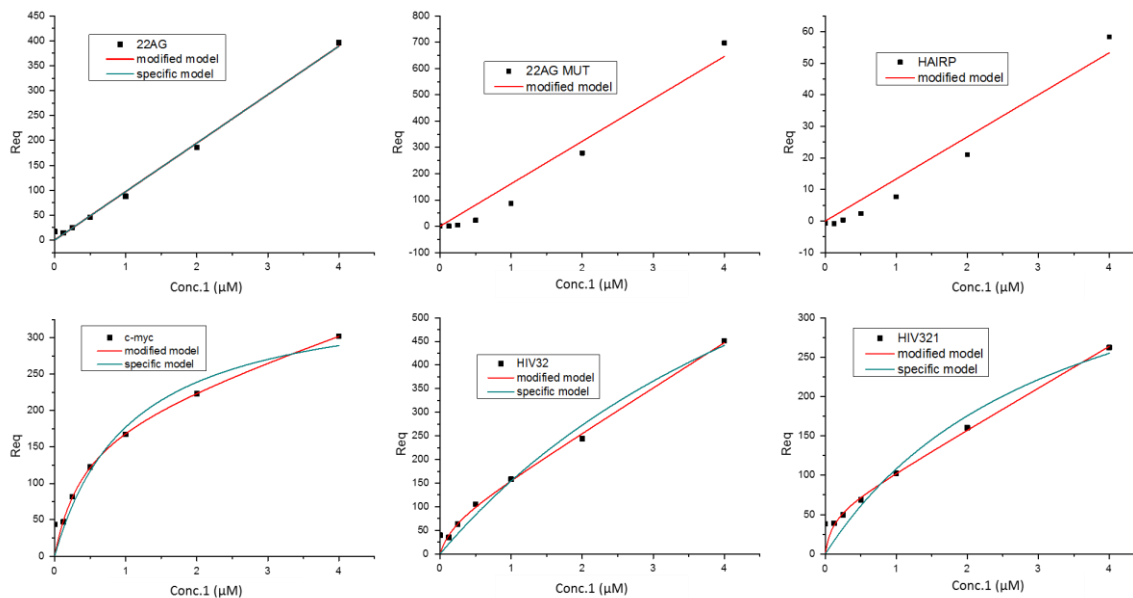
Figure 2.39. SPR fitted titration curves for **1**.

5.2.5. NUCLEAR RESONANCE MAGNETISM (NMR) STUDIES.

NMR experiments were recorded at 20°C on a 700 MHz NMR Bruker spectrometer equipped with cryoprobe or a 400 MHz. 500 µL of the oligonucleotide sample was prepared at a concentration range of 120-1000 µM which contained 10% of D_2O and 90 % of the corresponding buffer solution (buffer conditions are described on table **2.9**). Oligo sample solutions were annealed by heating at 90 °C for two minutes and cooling overnight in the heating block previous to NMR experiments. ^{1}H 1D spectra were obtained by employing the selective excitation sculpting sequence, based on the "double pulsed field gradient spin-echo" which removes water resonance without affecting other resonances including those which are in fast exchange with water.[271] The number of scans for the 1D spectra were 512 on the 700 MHz spectrometer, and 1024 on the 400 MHz; using *hs11echadec* as the pulse program in both spectrometers. For one step-titrations, 2 equivalents of the corresponding compound were added into the 120-250 µM DNA folded sequence sample. Step-wise titration experiments were carried out in eight steps until a total of 2 equivalents were added. ^{1}H-^{1}H-2D NOESY experiments were recorded in 700 MHz spectrometer. *Noesyesgpph* was used as the pulse program with 128 scans and a mixing time of 300 ms to record 1 mM oligo samples in absence or presence of 2 equivalents of the ligand. All spectra were analyzed with TopSpin 3.5pl6, while NOESY spectra manipulation was performed with Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). PDB figures were modified with PyMOL software.

Oligonucleotide name	Sequence	Buffer condition	Origin
22AG	AGGGTTAGGGTTAGGGTTAGGG	20 mM sodium phosphate pH 6.9 70 mM NaCl	Human telomeric sequence
24TTA	TTGGGTTAGGGTTAGGGTTAGGGA	20 mM potassium phosphate pH 6.9 70 mM KCl	Human telomeric sequence
HIVADR	AGGGAGGTGTGGCCTGGGCGGG	20 mM potassium phosphate pH 6.9 70 mM KCl	HIV-PRO1
HIV321	TGGCCTGGGCGGGACTGGG	20 mM potassium phosphate pH 6.9 70 mM KCl	HIV-PRO1

Table 2.11. G4-forming sequences and NMR buffer conditions.

5.3. Biological studies

5.3.1. CELL CULTURE.

MCF-7, MRC-5, HEK-293T, LN-229 and HT-29 have been purchased by Sigma Aldrich (Germany) and from University of Granada. MRC-5 and HEK-293T cells were maintained at 37 °C and 5% CO_2 in 100% of humidity in low glucose DMEM (1 g/L) supplemented with 10% iFBS, 100 U/mL penicillin, 100 mg/mL streptomycin. 1% Non-essential amino acids were also added to HEK293T medium. HT-29, LN-229 and MCF-7 cells were maintained at 37 °C and 5% CO_2 in high glucose DMEM (4.5 g/L) supplemented with 10% iFBS, 100 U/mL penicillin and 100 mg/mL streptomycin.

5.3.2. CYTOTOXICITY ASSAY: MTT

MCF-7, HT-29, LN-229, HEK-293T and MRC-5 cells were harvested by trypsinization (0.25%) and seeded in 96 well plates (5000 cells in 100 µL/well) in the presence of increasing concentrations of compounds. Cellular toxicity was determined using the colorimetric MTT-based assay after incubation at 37ºC for 72 h: after 3 days of incubation, 10 µL of MTT (5 mg/mL in PBS) were added to each well and the plate was incubated for further 4 h (37°C). The resulting formazan was dissolved in 100 µL of DMSO and read at 550 nm in a VersaMax® plate-reader. All determinations were carried out in triplicate. The results are expressed as the concentration of compound that reduce cell growth by 50% versus untreated control cells (IC_{50}).

5.3.3. STABILITY ASSAYS.

HT-29 and MCF-7 cells were seeded (0.5 mL, 50 000 cells/well) on 24 well plates with DMEM High Glucose + 50 mL iBFS + Peniciline/Streptomicine (DMEM), and used 24 h post-seeding. A solution of the compound tested and stored at 10 mM in DMSO was added and incubated for three different times (t = 0 h, t = 3 h, t = 24 h, t = 48 h). Controls were prepared, using only DMEM and water with the compounds (0.5 mL). After incubation time, samples were collected and stored at -20ºC: 0.5 mL of the supernatant (SN); 0.5 mL solution of SDS at 0.4% in PBS with cells contents (LYS); and 0.5 mL of the water and DMEM controls.

A QDA single quadrupole mass spectrometer Waters Acquity UPLC-TM H-class system (Waters, Manchester, UK) equipped with an orthogonal Z-sprayTM electrospray ionization (ESI)
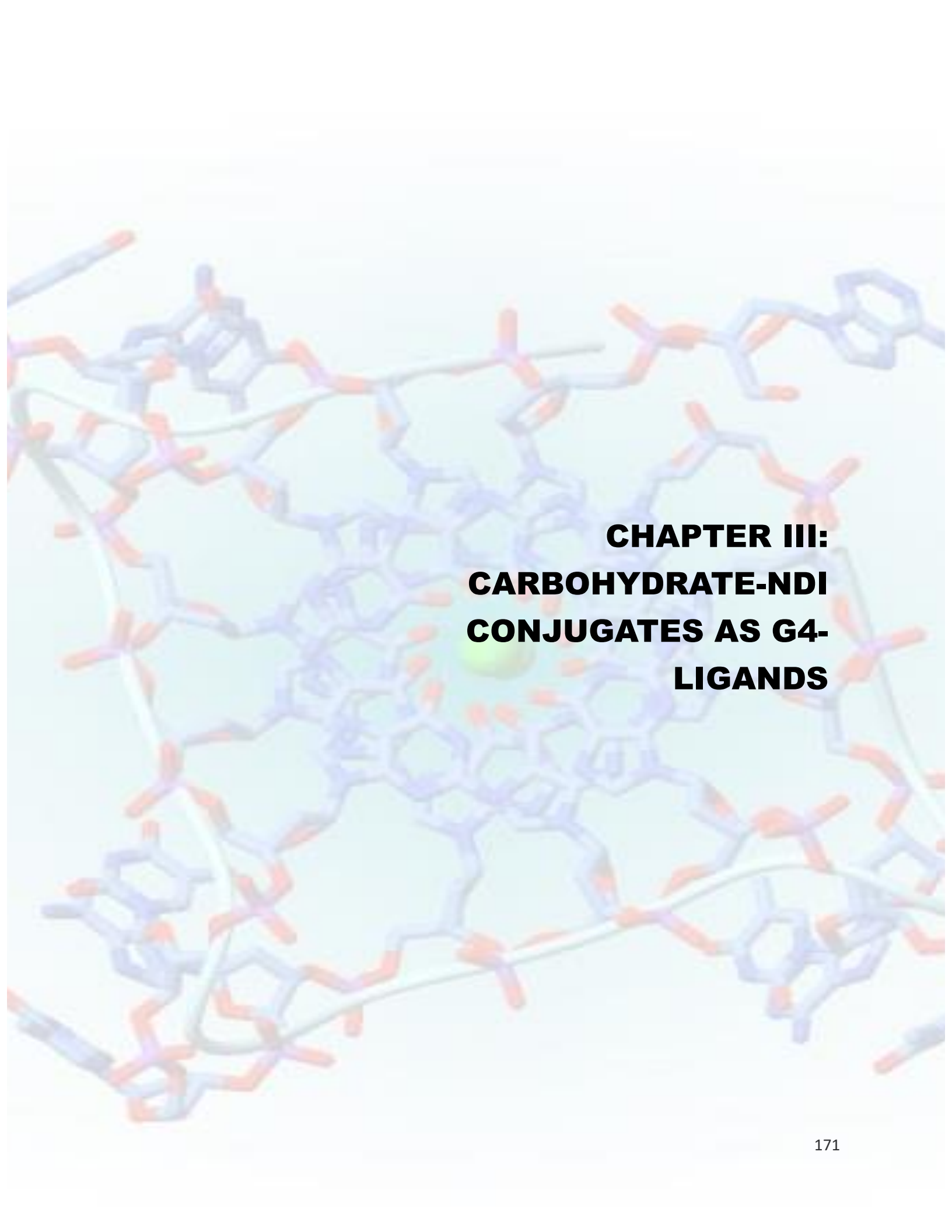
source was used for metabolites detection. The column for chromatographic separation was an Acquity UPLCR BEH C18 (2.1 x 100 mm, 1.7 µM). Empower 3 software was used for instrument control, peak detection and integration. Data analysis was done in Microsoft Excel.

Unfrozen samples were homogenized by vortexing and filtered through a 0.2 µm nylon filter prior to be injected in the instrument. Chromatographic separation was performed using a binary gradient mobile phase consisting of H_2O+ NH_4OAC 0.1% at pH 4.6 (solvent A) and MeOH (solvent B). The flow rate was 0.3 mL/min, the column was maintained at 40 ºC and the injected volume was 10 µL. Gradient conditions were as follows: linear gradient from 90% of A to 0% (total method time 9 min). The mass spectrometer (MS) was operated with electrospray ionization (ESI) in negative ion mode [M-H]$^-$ and the data were collected by selected ion recording mode (SIR). A calibration curve was done at the same time, with the samples in 0.5 mL of DMEM at 80 µM to 2.5 µM. The samples used were the ones previously added for the incubation. For samples analysis, firstly areas of the peaks at the desired mass were integrated; calibration curves were set up and CC of samples were extrapolated from them (CC); and the % of remnant compound found was calculated from the CC difference of the initial and final concentrations (CCsample/80 µM * 100 = %).

Calibration curves:

Chart titled **52**, showing Area vs Concentration (µM) with linear fit:

$y = 37529x - 95581$

$R^2 = 0.9952$
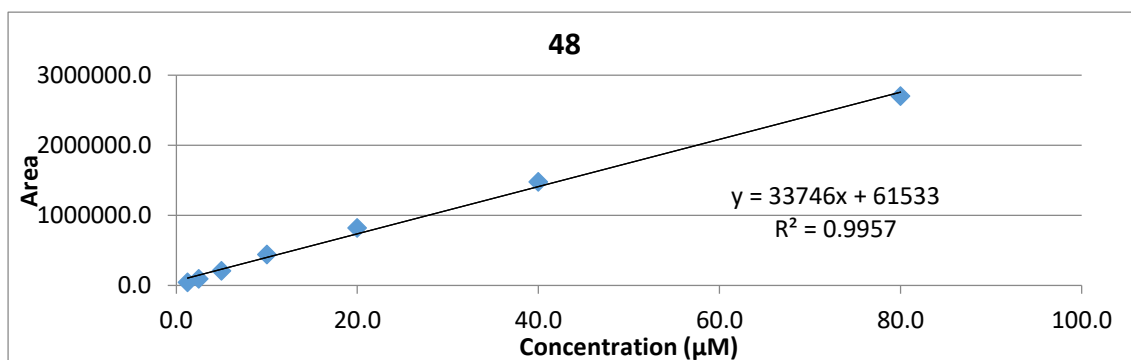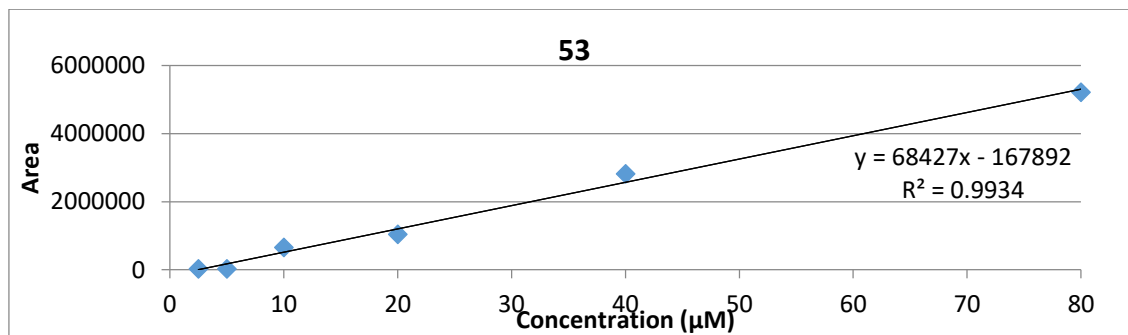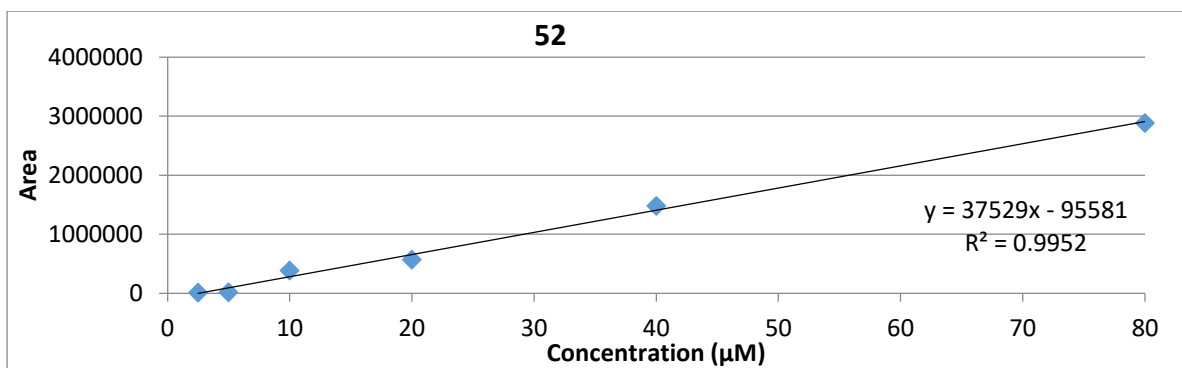
CHAPTER III: CARBOHYDRATE-NDI CONJUGATES AS G4-LIGANDS

1.   INTRODUCTION.

The second compound family of G4-ligands we decided to work with were naphthalene diimide (NDI) derivatives. This part of the work has been carried out in collaboration with Prof. Mauro Freccero (University of Pavia) and Dr. Jean-Louis Mergny (Institut Européen de Chimie et Biologie, University of Bordeaux). The NDI derivatives have the typical quadruplex strong G4-ligands features: large aromatic scaffold and cationic side chains, hence this family has been successfully tested as G4-ligands. We applied our general approach to design carbohydrate-NDI conjugates (**carb-NDI**) as a new G4-ligand series with improved properties.

1.1.   NDIs general chemistry and physical properties.

The 1,4,5,8-Naphthalene diimide structure is constituted by a 4-ring aromatic system including a naphthyl core and two cyclic imides to form a symmetric compound (figure **3.1**). The final structure is neutral, planar, electron-deficient, chemically robust, redox-active and it possesses high melting points. Their chemistry allows substitutions both on the diimide nitrogens and/or the aromatic core with groups whose nature will confer differential features to the whole molecule.
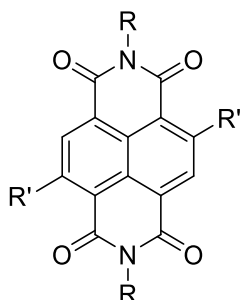
Figure 3.1. General NDI structure.

a.   Synthesis.

The synthesis of symmetric-NDIs **II** is normally accomplished in one step starting from the commercially available 1,4,5,8-napththalene tetracarboxylic dianhydride **I** which is condensed with the corresponding primary amine (scheme **3.1**). A wide variety of functional groups can be introduced at the imides and microwave assistance has been reported to optimize this step. The unsubstituted diimide **III** can be prepared from the same starting material and urea in refluxing 1,3,5-chlorobenzene.[272] Asymmetric NDIs with two different substituents on the diimide nitrogens **V** were reported by Ghadiri *et al.*[273] who applied a two-step pH-dependent method breaking one of the anhydrides and introducing the first amino group **IV** to lately incorporate the second primary amine.

To prepare tri- and tetra-substituted NDIs (3-NDIs and 4-NDIs, respectively), Neidle's team[274] reported a successful methodology starting from the commercially available dibromo derivative **VI**. The corresponding alkylamino chain is covalently linked by a nucleophilic aromatic substitution (S$_N$Ar) both on imido and core positions (scheme **3.2a**). The tri-substituted derivative **VIII** was reached as a byproduct from the same reaction. When different alkylamino moieties were

desired, the dichloro derivative **IX** was chosen as starting material to firstly obtain the diimide **X** and subsequently substitute the chlorines by the second amine (scheme **3.2b**), each step optimized using microwaves at short times.
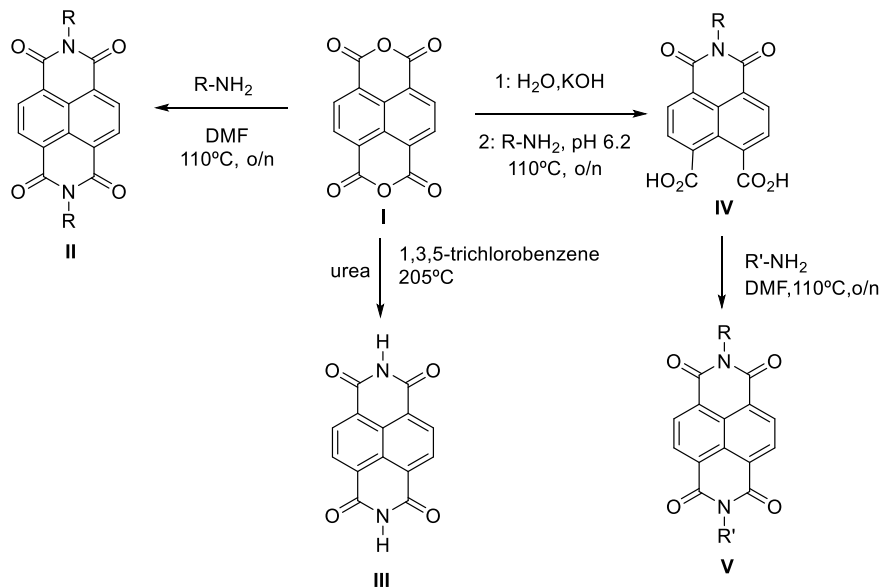
Scheme 3.13. General synthesis method for preparing *N, N*-disubstituted and *N, N'*-disubstituted NDIs.
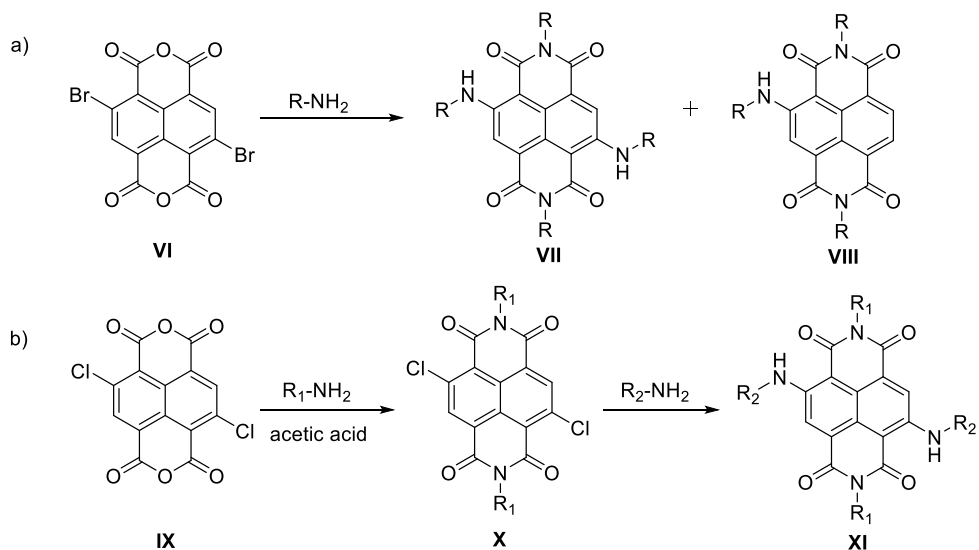
Scheme 3.14. Synthesis of tri- and tetra-substituted naphthalene diimide analogues. a) Reaction with the dibromo derivative. b) reactions stating from the dichloro derivative

b. Solubility.

NDI derivatives generally present poor solubility in any solvent, neither organic nor aqueous, as a result of their aromatic planarity and the presence of π-π aromatic interactions which confer them a strong tendency to aggregate. However, chemical functionalization can solve this

problem.[275] The naphthyl core and four polar carbonyl groups confer solubility in low polarity lipophilic solvents (toluene, DCM, chloroform) and polar aprotic solvents (acetonitrile, DMF, DMSO), depending on the imide substituents. For instance, long and bulky aliphatic substituents avoid aggregation and have higher solubilities in the above mentioned solvents;[276] as well as ionic substituents improve solubility in water.[277] In the solid-state NDIs tend to stack on each other and this tendency can be used for supramolecular applications, but on the other side it reduces the solubility especially when substituents are aryl moieties. Hence, the *N*-substituents play a key role in the global NDI solubility.

c. Physical properties

NDIs undergo single reversible one-electron reduction to form stable radical anions characterized by intense and characteristic visible and near-infrared absorption bands at >450 nm.[278] *N,N*-dialkyl-NDIs display absorption and emission properties which depend on the substituents incorporated on the imides. Substituents on the core system influence also on their fluorescence properties: alkylamino and alkoxy moieties provide brilliant colors and intense fluorescence, in contrast to arylamino groups.[279] Weaker π-donor alkoxy groups yield green/yellow fluorescence while stronger alkylamino π-donors result in a bathochromic shift.

The pH can strongly influence NDI-fluorescence as well. For instance, Fin *et al.*[280] described the ability of hydroquinone core substituted NDIs (cNDIs) to cover all primary colors by sample pH changing (figure **3.2a**) and such panchromatic hypersensitivity also suggests they would be ideal for pattern generation in differential sensing arrays. Solvent may also affect, as Doria *et al.*[281] demonstrated with hydrosoluble tetra-substituted NDIs using NIR (near infrared) absorption (figure **3.2b**).
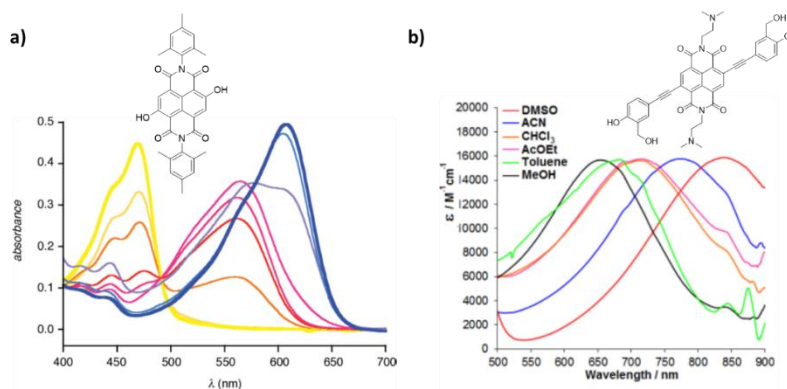
Figure 3.2. a) Changes in the absorption spectra of a cNDI with increasing pH from pH 3 (yellow) to pH 12 (blue) in DMSO/water 4 : 1, 25 °C.[280] b) Solvent dependent absorption of the conjugate bases of the tetrasubstituted NDI in organic solvents.[281]

1.2. NDIs' applications.

NDIs derivatives have been employed in materials chemistry to be incorporated on thin films and other organized structures such as nanotubes or rods. The objective is the preparation of pragmatic molecular based electronic devices such as field-effect transistors (FETs). In this sense, they have been exploited as **DNA-sensors** since their planar aromatic structure results in DNA-intercalation. Takenaka *et al.*[282] used on NDI bearing two redox-active ferrocenyl moieties to

electrochemically sense and discriminate between double- and single-stranded DNA: the dsDNA-NDI complex was more stable than the ssDNA-NDI one.

Among **supramolecular chemistry** applications, their similar electron accepting properties to the natural acceptors of plants or bacterial photosynthetic reaction centers was exploited in molecular-based optoelectronics and in advances in solar energy conversion.[283] These applications were performed by both covalent and non-covalent NDIs systems.

Core substituted naphthalene diimides have been reported as **biomimetic and bioinspired artificial systems** such as electron and energy transfer systems in synthetic multifunctional pores and transmembrane channels.[284–286]

In addition, the NDIs structural and chemical features make them great candidates for **host-guest interactions**, especially in donor-acceptor charge transfer type complexes for two reasons: i. the importance of their electron deficiency and aromatic nature for face-to-face aromatic interactions with electron donors; ii. their rigid, planar structure together with the functionalization variety with different side groups convey strong and well directional assemblies. As host-guest systems, NDIs have been applied as intercalators, foldamers, ion channels, catenanes and rotaxanes.[275] To highlight, self-organized rigid-rod π-helices containing naphthalene diimides have been reported to bear strong channel conformational changes upon intercalation with aromatic compounds such as *p*-octiphenyls (figure **3.3**).[287]
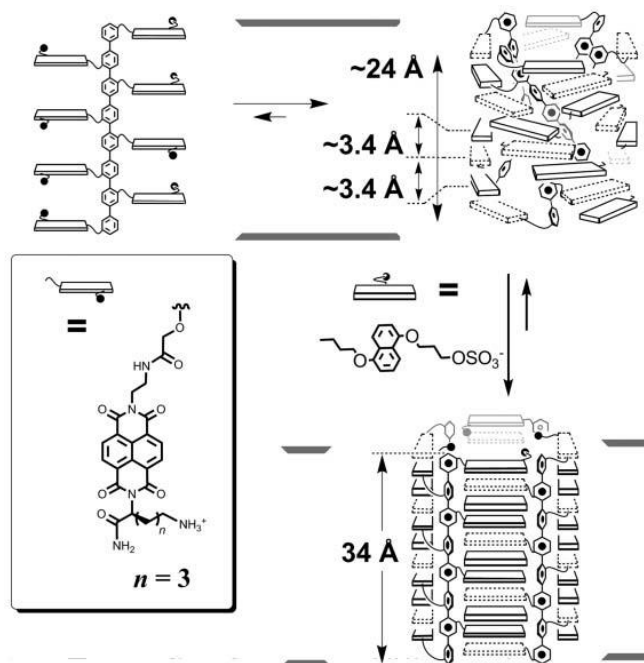
Figure 3.3. Ligand-gated opening of rigid-rod p-helix ion channel by intercalation of DAN ligands into closed scaffolds containing naphthalene diimides.

Taking into account the NDIs inherent fluorescence, redox features and their ability to recognize and stabilize the quadruplex structures (see below), they also were exploited as **G-quadruplex light-up probes**. For instance, Doria *et al.*[288] developed a dimeric NDI conjugating

red and blue dyes to reach a non-fluorescent compound which becomes red/NIR emitting upon G4-binding (figure **3.4a**). The complexes G4/NDI dimer displayed a differential fluorescent lifetime than the corresponding dsDNA/NDI dimer (figure **3.4b**). The same approach was followed by substitution of one NDI unit by a coumarin derivative attached by a linker (figure **3.4c**).[289] It presents a single wavelength emission at 498 nm by itself when free in solution while it displays an additional red/NIR emission when it is bound to G-quadruplex DNA. The coumarin behaves as a light harvesting antenna transferring the energy to the emitting NDI fluorophore and the coumarin emission slightly decreases in the presence of nucleic acids.

Figure 3.4. a) Dimeric NDI structure used as G4-sensor. b) Graph displaying the long lifetime component (τ_{long}, in green) as well as the average lifetime (τ_{av}, in red) of the dimer NDI solutions. c) schematic representation of the developed coumarin-NDI dimer as G4-sensor.

1.3. NDIs as G-quadruplex ligands.

The NDIs meet the stablished requirements for becoming excellent G4-ligands thanks to their aromatic planarity and the possibility to introduce cationic side chains both on the imide and the aromatic core. To date, bi-, tri- and tetra- substituted NDI derivatives have been reported as G4-ligands whose properties depended on the size of the substituents and the number of protonable atoms.

Figure 3.5. Bi-, tri- and tetra- substituted developed NDI derivatives with different moieties.

Bisubstituted NDIs (**2-NDIs**) were screened for the first time in 2007 as G4-ligand without great results due to their low selectivity towards quadruplex structures.[290,291] Neidle's group.[274] designed in 2008 tri- (**3-NDIs**) and tetra- (**4-NDIs**) substituted NDIs (figure **3.5**) to improve such affinity, since human telomeric quadruplex has four targetable grooves to be accessed by the NDIs' substituents. According to the FRET melting assays, 4-NDIs showed the best hTel-stabilization values (up to 35 °C) followed by the 3-NDIs and 2-NDIs. Different binding and selectivity results were observed depending on the specific group attached to the side chains: morpholine and hydroxy derivatives displayed the lowest ΔT_m values, while dimethylamino, diethylamino and pyrrolidine presented the highest. However, the selectivity towards G4 was reversed: morpholine and hydroxyl moieties were the most selective among the side chains used.

The same group developed in 2012 a series of ***N*-methyl piperazine-tetrasubstituted** NDIs with different linker length (3 or 4 carbons) displayed high affinity and selectivity towards hTel according to the FRET melting results.[292] Two crystal structures of the 22 mer hTel G4 sequence and the NDI ligands BMSG-SH3 (linker with 3 atoms) and BMSG-SH4 (linker with 4 atoms) were resolved (figure **3.6**).[145] These compounds favored the parallel-stranded quadruplex topology binding to the 3' surface through extensive π-π contacts. Both NDI:hTel complexes involve two individual quadruplex forming a dimer by stacking at the 5' G-quartet interface, mediated by a coordinating potassium cation. In addition to the aromatic stacking, most protonated side chains are positioned within the G-quadruplex grooves. The charged nitrogen atom of one *N*-methyl-piperazine ring is in direct hydrogen-bonded proximity to a phosphate oxygen group. A network of water molecules intermediates within the interaction between the *N*-methyl piperazine ring and the grooves as well. The only difference between both linkers are the ligand mobility, with BMSG-SH3 ligand being less mobile. The higher mobility in the BMSG-SH4 derivative destabilizes slightly the quadruplex, in concordance to the lower ΔT_m value in FRET melting assays.

Figure 3.6. Representations, projected onto the plane of the terminal G-quartets, of the Gtel22 quadruplex solvent-accessible binding surfaces with the bound ligands BMSG-SH-3 and BMSG-SH-4 shown colored.

In 2013, a new series was developed by substitution of two *N*-methyl-piperazine moieties by more weakly basic groups, such as **morpholine** (less basic than *N*-methyl-piperazine), to improve the pharmacological properties of the previous compounds without affecting G4-afinity.[146]The corresponding crystal structures were also reported, sharing the same interaction pattern than the previous ones. The reduced basicity of the morpholine group leads to a reduced electrostatic contribution to stabilization and consequently slightly lower ΔT_m.

Another approach explored with the NDI derivatives was the development of **G4-alkylating agents**, developed by Prof. Freccero group in 2011.[143,293,294] A class of NDIs were able to reversibly bind and alkylate telomeric DNA through an electrophilic quinone methide moiety (QM), with excellent affinity and selectivity to the hTel G4 sequence (figure **3.7a**). However, the alkylation effect was not as effective as expected and many DNA-adducts appeared. The same group developed a second family of NDI alkylating agents with an enhanced efficiency by the introduction of the oxirane moiety (figure **3.7b**).[295] Results showed a highly selective alkylation of adenines versus guanines in hTel, in agreement with the lower accessibility of the guanines in the structure and the availability of the adenines placed in the loops.

Figure 3.7. NDIs derivatives developed as alkylating agents

Esaki *et al.*[296] explored the synthesis of **cyclic NDIs** to optimize the specificity towards hybrid type quadruplex by connecting the NDI core with 2,2'-cyclohexane-1,1-diyldiacetic acid with linkers of two different lengths (figure **3.8**). This arrangement would reduce the dsDNA binding. Nevertheless, the results still showed some stabilization over duplex DNA structure.
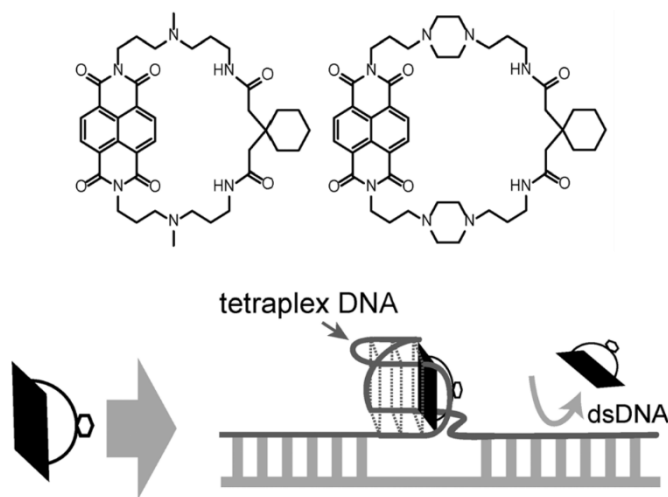
Figure 3.8. Chemical structures of cyclic NDIs and the concept of a G-quartet specific ligand based on cyclic naphthalene diimide.

Most reported NDIs displayed strong G4-stabilization effect, especially the tri- and tetra-substituted derivatives. However, the selectivity is still a limitation factor since they are not completely selective towards quadruplex topologies and bind dsDNA as well.

1.4. Biological studies with NDIs: *in vitro* studies.

The high G4-stabilization ability of the NDI ligands suggested that they could act as potential anticancer agents. Moreover, their inherent fluorescence could be exploited to localize the compounds inside the cells by fluorescence microscopy. Many of the mentioned works included cytotoxicity assays to analyze their activity. On the contrary, only a few studies have been

performed in animal models what hinders the progress to clinical phases. The reported cytotoxic results for NDIs depend on the nature and length of the incorporated side chains.

Neutral bisubstituted NDIs (figure **3.9a**) already displayed cytotoxicity over different cancer cell lines although the reported values were in the low or sub-micromolar range.[291] The NDI containing three methylene side chains and 2,3,4-trimethoxy groups on the two aromatic rings exhibited the best IC$_{50}$ values over a wide number of human cancer cell lines, whilst they did not improve the results from the anticancer agent Vincristine. Additionally, they showed telomerase inhibition activity in some cell lines, but this ability could not be directly related to the anticancer property. It seems they were not selective towards telomeric-G4 and bind dsDNA as well.
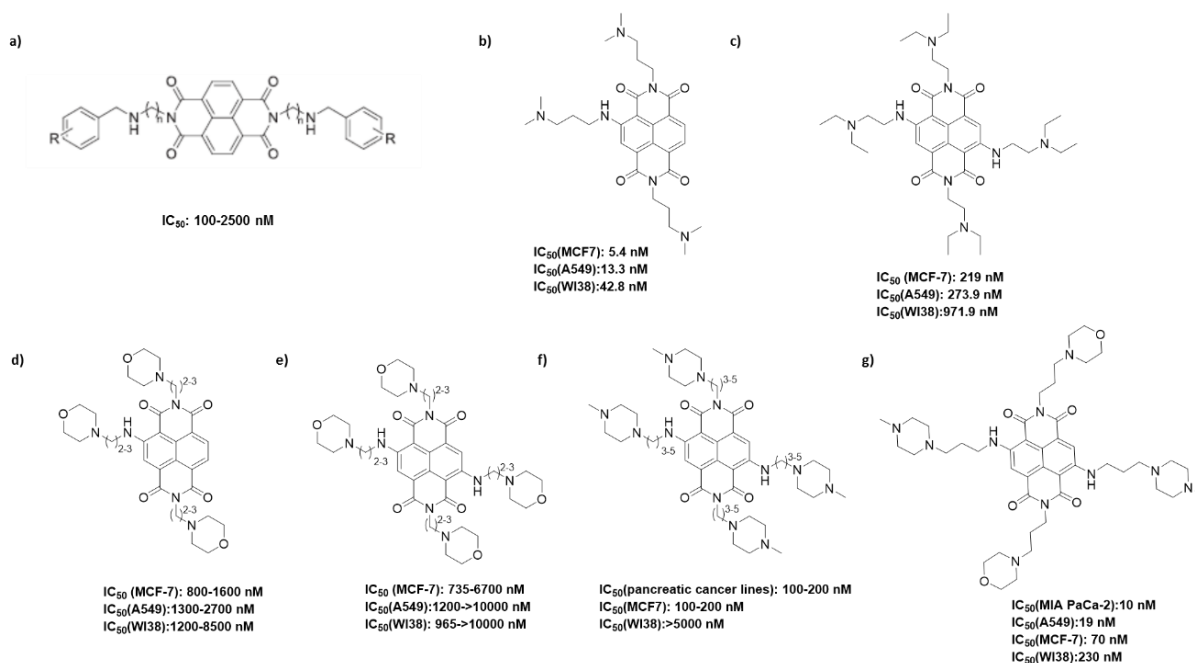
Figure 3.9. Some NDI structures assayed on different cancer cell lines.

The introduction of protonable amino-substituents (dimethylamine, diethylamine, pyrrolidine and piperidine) in **tri- and tetra-substituted NDIs** improved considerably the cytotoxicity effect over MCF-7 and A549 cancer cell lines, with a IC$_{50}$ range of 5-300 nM (figure **3.9b** and **c**).[274] On the contrary, the introduction of morpholine moiety decreased the activity up to the micromolar range due to their low cellular uptake (figure **3.9d** and **e**). In general, the results were in concordance to the reported G4-stability values. However, they could not connect both effects and assure such a mechanism of action. The 3- and 4-NDIs potent fluorescence was applied to locate the compounds inside MCF-7 cells using confocal microscopy. They were found mainly in the nucleus, especially inside the nucleolus.

Next, **tetrapiperazine-NDI derivatives** (figure **3.9f**) were evaluated over some cancer cell lines, including cancer pancreatic cells (MIA-Paca-2, HPAC, and PANC-1), with worse results than the previous ones (IC$_{50}$ values of 100-200 nM), but with improved selectivity: the BxPc-3 normal pancreatic line and WI38 (normal fibroblasts) were around tenfold less sensitive to the compounds than the cancer cell lines.[292] The telomerase inhibition assay with MCF-7 resulted in

a 50% of reduction in telomerase activity.[297] The same tetrapiperazine derivative and two more 4-NDIs with dimethylamine or pyrrolidine groups were located inside MCF-7 cells by confocal microscopy (figure **3.10**) obtaining dissimilar results: while the tetrapiperazine was assayed at high concentration (50 µM) and located in the cytoplasm with low fluorescence and at short incubation times (figure **3.10a**), the other ones required lower concentrations (500 nM) and were placed both in cytoplasm and nucleus with more intense fluorescence (figure **3.10b** and **c**). Their presence in the cytoplasm may indicate the interaction with extranuclear nucleic acids, such as ribosomal RNA, $_m$RNAs or mitochondrial DNA. The author suggested four main mechanisms of action:

- Telomere uncapping with subsequent DNA damage response and senescence.
- Inhibition of transcription/translation of oncogenes.
- Genomic instability by telomeric end-to-end fusions, resulting in mitotic catastrophe and apoptosis.
- Induction of chromosomal instability by telomere aggregate formation.

When two of tetrapiperazine moieties were substituted by morpholine groups (figure **3.9g**) the pharmacological properties improved due to the reduction of the molecular weight and basicity.[146] This change provided selective submicromolar antiproliferative activity in MIA PaCa-2 pancreatic carcinoma and A549 lung adenocarcinoma cell lines (IC_{50} values of 10−20 nM). In contrast with the previous work with MCF-7 cell line this family compounds lacked telomerase inhibitory activity at the cytotoxic concentration range in MIA PaCa-2 cell line, which suggested a more complex mechanism of action. By genome analysis, no significant transcriptional induction of the DNA damage pathways associated with telomere was detected whilst there was a downregulation of some genes encoding DNA-repair proteins (RAD50 and nibrin). In addition, it existed upregulation of another cellular stress markers, indicative of quadruplex-stabilization mediated stress.

Following the previous piperazine and morpholine 4-ND approach, the interchanged position of both substituents were performed to study the differences with their preceding model.[298] The new isomers resulted in improvement of the cytotoxicity profile with a large therapeutic window. Again, they noticed that the telomerase inhibition was not involved in the antiproliferative activity. On the other hand, it seemed longer side chains amino functionality attached to the NDI core resulted in better IC_{50} and ΔT_m values.

Moving to **alkylating agents**, their cytotoxicity has been also evaluated.[293] Even though the NDI cytotoxicity against telomerase-negative human foreskin fibroblasts (HFF) was 2-10 times lower than for the telomerase-positive cancer cell lines, the IC_{50} values remained in the micromolar range (1.4 µM for the best), then they did not improve the results from previous reports. Otherwise, authors maintained the theory that antiproliferative effect was due to specific telomerase/telomere-oriented mechanism of action. The alkylating tetra-substituted version presented very similar cytotoxicity, except for the dicationic derivative, with no effect over the cell lines assayed, possibly by its poor cellular uptake.[294]On the contrary, the pH-dependent dicationic extended-NDI exhibited better IC_{50} over the telomerase-positive human carcinoma cell line, HT-29 (300 nM), and it was located in the nucleus of A549 cells after 30 min incubation taking in advantage of its intense fluorescence (figure **3.10d**).[299]
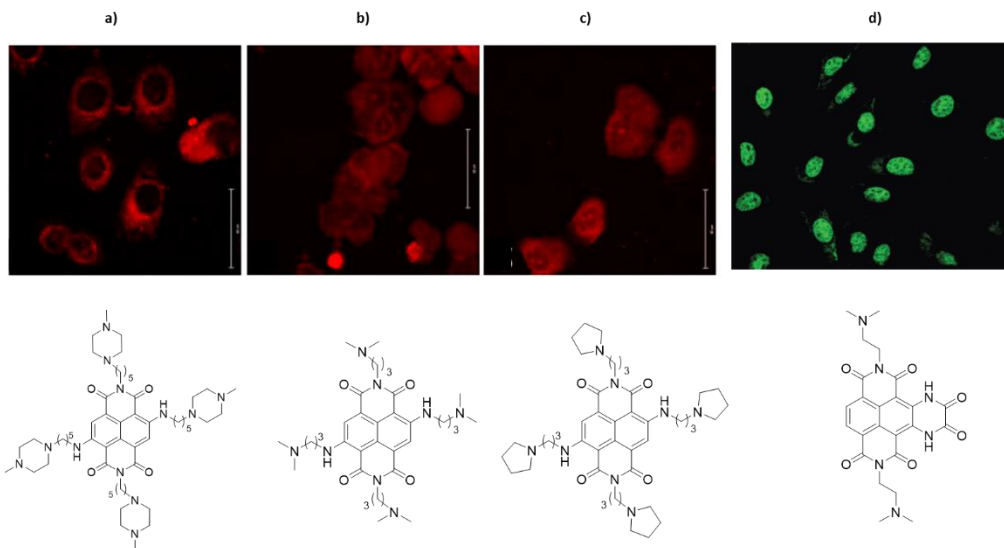
Figure 3.10. Confocal microscopy results locating **a)** tetrapiperazine derivative (50 µM), **b)** dimethylamine derivative (500 nM) and **c)** the pyrrolidine derivative (500 nM) in MCF-7 cancer cell lines; and **d)** extended NDI (1 µM) on A549 cells.

Taking into account the reported studies, NDIs display good antiproliferative activity mostly in the low micromolar or nanomolar range depending on the substituents, which strongly influence their cellular uptake and selectivity. Generally, the tetra-substituted NDIs show better results than the bi- and tri-substituted ones, with protonable terminal tertiary amines such as dimethylamine, diethylamine or piperazine preference over neutral moieties. Concerning their antiproliferative mechanism of action, results were contradictory since some authors defended the telomerase inhibition pathway while others did not assure such mechanism.

1.5.   Biological assays with NDIs: *in vivo* studies.

To date, only three *in vivo* assays have been reported, two in pancreatic cancer models and one on a medullary thyroid cancer (MTC) model. The studies carried out on pancreatic cancer are of especial relevance since it is one of the cancers with highest mortality and resistance, and it is mostly telomerase positive.

The first *in vivo* assay was performed with the tetrapiperazine-NDI derivative **BMSG-SH-3** (figure **3.11a**). Gunaratnam *et al*.[300] reported significant anti-tumor activity for this compound in a pancreatic cancer xenograft model, accompanied by changes in telomerase activity and telomere maintenance. The tumor growth was decreased by ca 50% over the period of the experiment without noticeable animal weight lost. The BMSG-SH-3 fluorescence was exploited to analyze the tumor exposure and ligand biodistribution by spectral fluorescence, which located it mainly in the tumors and pancreas with lower amounts in the lung (figure **3.11a**). Mechanism of action studies showed a 50 % of telomerase inhibition in tumor samples treated with BMSG-SH-3. Protein expression analysis of proteins related to the telomerase enzyme complex (HSP90 and hTERT) exhibited 30 % expression reduction. A small down-regulation (10 %) of bcl-2 protein

were observed, while k-ras protein levels was not affected. These results indicated bcl-2 and k-ras were not BMSG-SH-3 key targets, whereas the telomeres and HSP90 were.
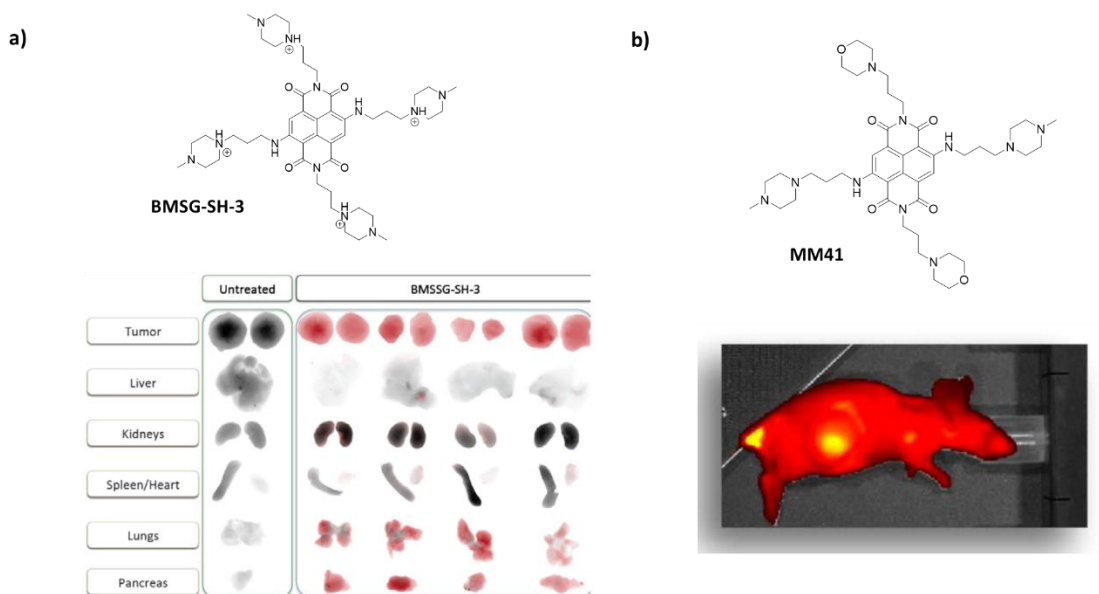
Figure 3.11. a) Structure and biodistribution of BMSG-SH-3 in major organs. b) MM41 structure and (LHS) IVIS whole-body fluorescence image, taken at the MM41 emission wavelength, showing that MM41 is concentrated in the tumor. A substantial amount of MM41 is shown to be retained at the site of administration.

The tetra-substituted NDI derivative **MM41** containing two piperazines and two morpholines moieties was examined as a potential drug the same xenograft pancreatic model (figure **3.11b**).[301] After 40 days mice treatment (15 mg/kg), an average of ca 80% decrease in tumor growth was observed without significant weight reduction or another side effects, improving the BMSG-SH-3 earlier results. No signs of tissue damage were observed. MM41 was located on the nucleus of MIA-PaCa-2 cells after 30 min incubation and mainly located in the tumor and accumulated at the site of administration in the animal after 4 h post-administration (figure **3.11b**). The Bcl-2 and k-ras expression levels were evaluated with Western blots of mice samples, showing a clear expression reduction in comparison to the controls (40 and 30 %, respectively). MM41 did not affect telomere-maintenance pathways, in contrast with the preceding report.

The third *in vivo* study was performed over a medullary thyroid cancer (MTC) xenograft model, treated with a **tri-subtituted and tri-cationic NDI derivative** (figure **3.12a**).[302] The hypothetical target in the experiment was the oncogene RET, which encodes a singlepass transmembrane tyrosine kinase receptor expressed in cells deriving from the branchial arches, the neural crest and the urogenital system. The mutant RET tyrosine kinase receptor controls the proliferation and survival of MTC cells by direct phosphorylation of multiple downstream targets. The proximal region upstream the TSS of RET gene contains two G-rich boxes that are essential for basal promoter activity and are characterized by a typical G4-forming sequence. The authors suggested RET-G4 stabilization by the NDI derivative could control the RET oncogene transcription. According to the cytotoxicity assays, the product displayed IC_{50} values on MTC cells in the low micromolar range (1.8 μM ± 0.3 and 1.7 ± 0.2 μM for MZ CRC-1 and TT cells, respectively). Curiously, the RET expression levels was reduced after cell exposure to the

compound for 96 h and compound stabilized the G4-RET folded sequence by quadruplex T_m increment, which agreed with the hypothesis proposed by the authors. When the compound was tested (12 mg/kg i.p.) on the *in vivo* model, a significant tumor growth delay was observed (50% or 37%, for MZCRC-1 and TT xenografts, respectively) without apparent sign of general toxicity (figure **3.12b**). Furthermore, the biochemical evaluation of the *in vivo* drug-mediated effects showed a marked reduction of the basal and phosphorylated forms of RET protein and an increase in cleavage of caspase-3 in tumors explanted from NDI-treated mice with respect to controls. Nevertheless, c-myc and Bcl-2 protein amounts barely changed, despite of their G4-forming sequences are stabilized by the NDI derivative as well.
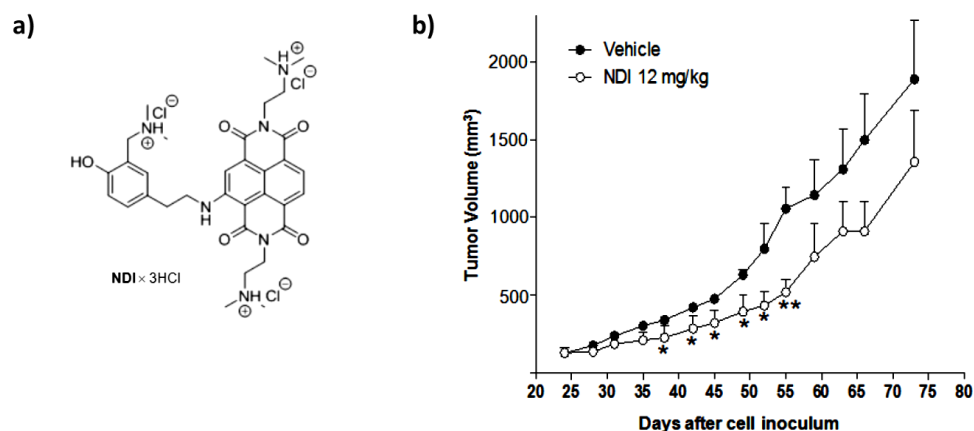
Figure 3.12.a) NDI derivative assayed in the experiments. b) Growth curves of MZ CRC-1 tumors in vehicle-treated mice (●) and upon i.p. administration of 12 mg/kg NDI derivative (○). Data have been reported as average tumor volume (mm³) ± S.E.M.

In summary, none of the *in vitro* or *in vivo* reported assays were able to directly relate the NDI antiproliferative effect with a certain quadruplex stabilization. In fact, it is still not possible to determine how many quadruplex are folded on *in vivo* conditions and which ones can be stabilized by the G4-ligand once it has reached the tumoral cells.

2. AIMS.

- To synthesize a series of NDI-carbohydrate conjugates by incorporating a carbohydrate as side chain to a bisubstituted NDI core (in collaboration with Prof. Mauro Freccero, University of Pavia).

- To characterize the interaction between the new ligands and different G4-forming sequences using different biophysical assays techniques (FRET, CD and NMR spectroscopy) in collaboration with Dr. Jean-Louis Mergny (IECB, University of Bordeaux).

⬩ To evaluate the potential antiproliferative activity of the new ligands: to carry out MTT-assay over several tumoral and non-tumoral cell lines.

⬩ To investigate the cell uptake of the new G4-ligands using flow cytometry, fluorescence spectroscopy and confocal microscopy.

3. RESULTS AND DISCUSSION.

The strong G4-stabilization ability displayed by most reported NDIs inspired us to apply our general approach: to attach carbohydrates moieties to an established G4 ligand. From this conjugation, we expected to increase compound solubility and bioavailability, improve G4 binding through sugar-mediated interactions to the phosphates and G4-DNA grooves, and finally promote selective entrance into tumor cells through GLUT protein transporters. All of this preserving their solubility and the fluorescence properties of the NDI core.

The synthesis of this new family of compounds was carried out in collaboration with the Prof. Mauro Freccero (University of Pavia) whose team prepared the NDI core and optimized the carbohydrate-NDI conjugation step, later we made the final characterization in our laboratory. The biophysical evaluation was performed in Dr. Jean-Louis Mergny laboratory (University of Bordeaux) and the final biological evaluation was carried out in our laboratory with the equal participation of PhD. student Efres Belmonte. The corresponding results of this section have been recently published.[303]

3.1. Synthesis of the NDI carbohydrate conjugates.

Six new NDI carbohydrate conjugates were designed and prepared to be evaluated as G4-ligands and antiproliferative agents. We selected different carbohydrate motifs and two different ways of attachments to explore several sugar presentations: β-glucose (**β-glc-NDI**), ethyl β-glucose (**β-glcC2-NDI**), β-maltose (**β-malt-NDI**), *N*-acetyl-β-glucosamine (**β-glcNAc-NDI**), 6-deoxy-β-glucose (**β-6-dglc-NDI**) and ethyl α-mannose (**α-manC2-NDI**). An additional aglycone NDI (**prop-NDI**) was included as well to study the differential behavior and properties, elucidating the real carbohydrate role (figure **3.13**).
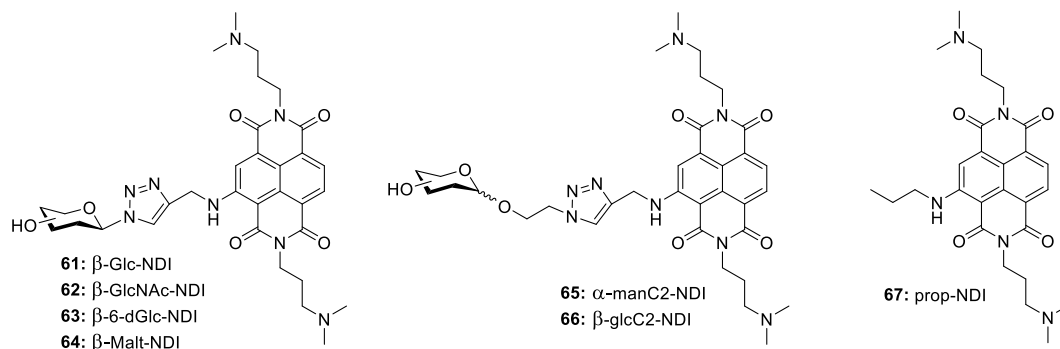
Figure 3.13. General structures of the new NDI carbohydrate conjugates.

The carbohydrate-NDI attachment was achieved by the same click chemistry approach than the phenyltriazole derivatives from the previous chapter. The azido-derivative was represented by the corresponding azido sugar and the dicharged NDI core possessed the alkynyl group.

When Freccero's laboratory optimized the click chemistry coupling, we observed the yields were higher for the deprotected azido sugar than the acetylated. The peracetylated azido sugar derivatives were prepared in our laboratory according to the procedure described in the earlier chapter. Hence, we synthesized β-azido-glucopyranoside (**69**), 2-acetamide-2-deoxy-β-azido-glucopyranoside (**70**), β-azido-maltoside (**72**), 2-azidoethyl-β-glucopyranoside (**74**) and 2-azidoethyl-α-mannopyranoside (**73**). The deprotection of the peracetylated azido sugar derivatives was carried out using standard basic conditions with good yields (scheme **3.3a**).
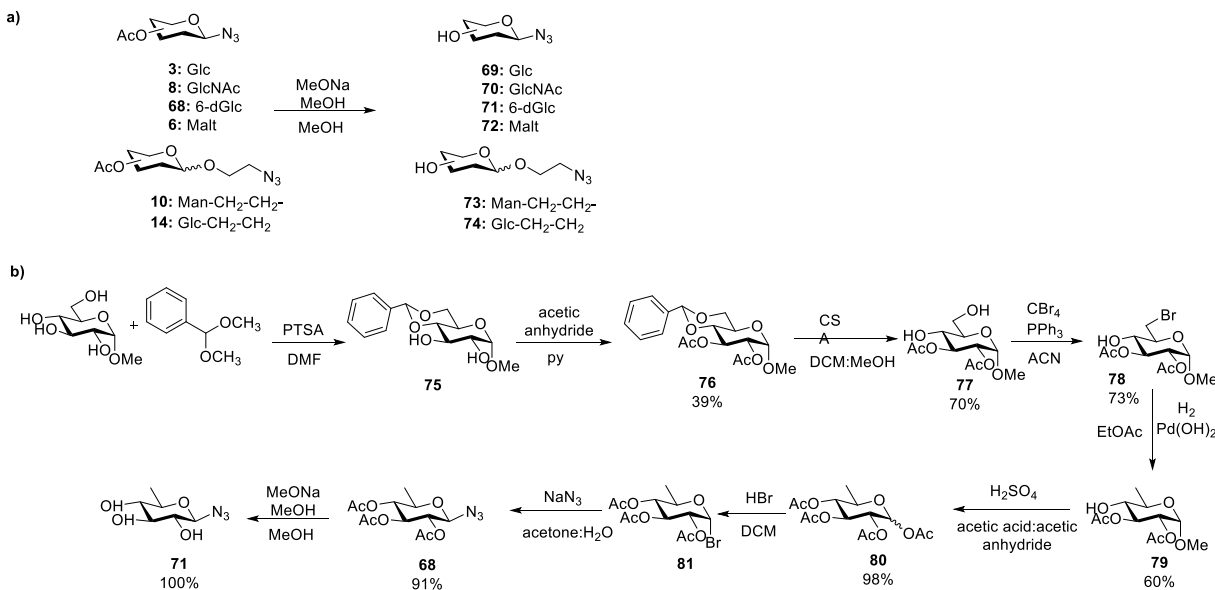
Scheme 3.15. a) general deprotection of **3**, **8**, **68**, **6**, **10** and **14** azido sugar derivatives. b) Synthetic route for the preparation of azido-6-deoxy-β-glucopyranose **71**.

We decided to include a partially hydrophobic sugar to this series: 6-deoxy-β-azido-glucopyranoside (**71**). 6-Deoxy-D-glucose is commercially available but with a prohibitive prize in large amounts. So, we decided to prepare it using classical carbohydrate synthesis (scheme

3.3b). The route was started from methyl α-D-glucopyranoside. The 4,6-diacetal derivative **75** was prepared by reaction with benzaldehyde dimethyl acetal and *p*-toluenesulfonic acid as the proton source catalyst, forming preferentially the six-membered acetal ring.[304] Without further purification, classical acetylation was performed to obtain the 2,3-diacetyl derivative **76**. Then, successful regioselective acetal hydrolysis was carried out at acidic conditions. Once the 4, 6-positions were deprotected we proceeded to bromide the primary hydroxyl with tetrabromomethane[305] yielding the 6-bromo-derivative **78** in good yields (73%). Hydrogenation with Pd(OH)$_2$ lead to the diacetyl 6-deoxyglucose derivative **79**, which after acetylation at acidic conditions yielded the peracetylated 6-deoxy glucose **80**. The corresponding azido sugar was obtain following the methodology described in the previous chapter.

The NDI core was designed to carry two positively charged side chains and the alkynyl group would be attached at the central naphtyl aromatic ring. The synthetic route used by Freccero's group had been already reported (scheme **3.4**).[277,293] They started from the commercially available dibromonaphthalene diimide **VI** which was reacted with *N,N*-dimethyl propylamine in acetic acid to obtain the monobromo NDI derivative **82** and the dibromo NDI derivative **83**. The following nucleophilic aromatic substitution (S$_N$Ar) was carried out with the mono bromo derivative **82** and propargylamine to obtain the corresponding propargyl NDI derivative **84** as major product in good yield (75%).
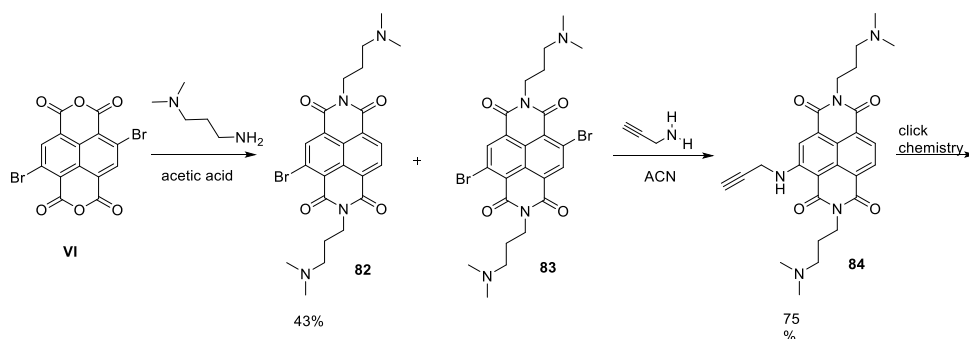
Scheme 3.16. NDI core alkynyl preparation by Filippo Doria.

Next, Freccero's group conjugated the azido sugar to the propargyl NDI derivative **84** (scheme **3.5**). The reaction was optimized including the classical click chemistry reagents: copper (II) sulfate pentahydrate as copper source and (+)-sodium L-ascorbate as copper reducing agent. Finally, a ^{t}BuOH:water solvent mixture was selected and the reaction was performed at room temperature for 2 h. The crude was purified by preparative HPLC to obtain the corresponding carbohydrate-NDI conjugates in moderate yields (35-70%). Aglycone NDI **67** was prepared as a reference ligand to compare the biophysical and biological properties of the new carb-NDI conjugates with a sugar-free NDI analogue. The dibromo derivative **VI** was reacted with n-propylamine in acetonitrile solvent to yield the aglycone NDI **67** (68% yield) (scheme **3.6**).
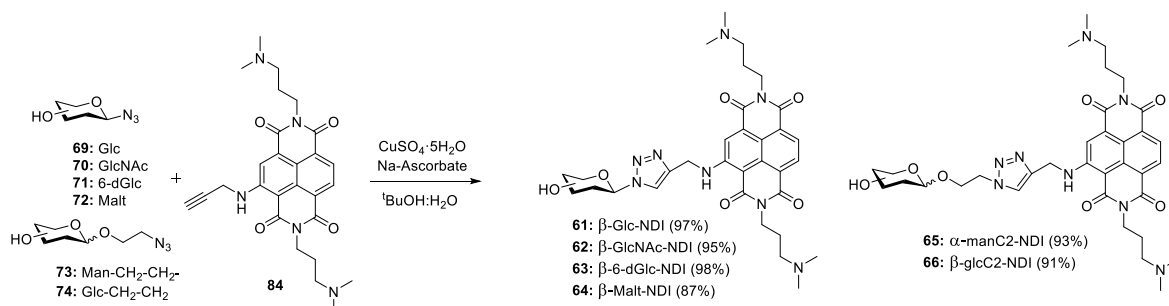
Scheme 3.17. General scheme of the new carbohydrate-based NDI family synthesized by click chemistry.
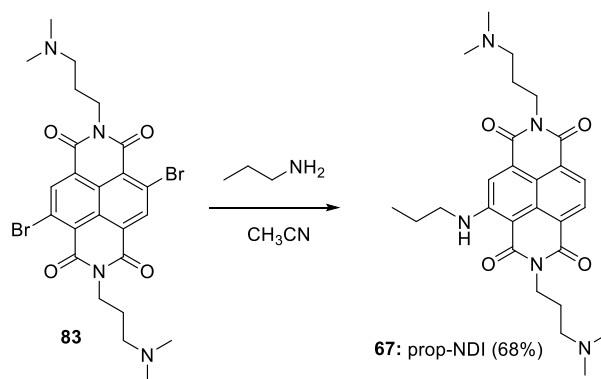
Scheme 3.18. Synthesis of the aglycone NDI derivative **prop-NDI (67)**.

3.2. Biophysical studies of carbohydrate-NDI ligands.

The new NDI carbohydrate conjugates family was studied as potential G4-binders using several biophysical methods: FRET melting assay, CD and NMR. All experiments were performed by Dr. Aurore De-Rache from Dr. Mergny's laboratory (IECB, University of Bordeaux). We expected that the carbohydrate increased the G4-selectivity without sacrificing the characteristic strong affinity of the NDIs.

3.2.1. FRET MELTING ASSAY.

The same methodology applied to the carbohydrate phenyl ditriazole family was used to examine the new carb-NDIs G4-affinity and selectivity. Their affinity towards the G4-human telomeric sequence (hTel) was firstly analyzed in the presence of K^+. The concentration range of the ligand required to notice any change on the oligonucleotide T_m was considerably lower than the needed for the carbohydrate phenyl ditriazole derivatives. All the new carb-NDIs induced an increase in the melting temperature on the hTel sequence (0.2 µM) with values close to 20 °C at only 2 µM ligand concentration (figure **3.14a**). The stabilization was very similar for all of them, both the sugar-derivatives and the control ligand lacking a carbohydrate (**NDI-prop 67**), with the only exception of **β-glcC2-NDI (66)**. Therefore, the introduction of a carbohydrate motif in the NDI structure did not display any detrimental effect on G4-stabilization. At the same time, we did not observe relevant differences on G4-binding among the carb-NDIs examined.

The strong G-stabilization displayed by the NDIs allowed to carry out a competition assay by addition of dsDNA. Stabilization evaluation in the presence of a duplex sequence as competitor provides more accurate selectivity data between G4-structures and DNA duplex. In the presence of 3 µM DNA duplex competitor (15 equiv), the induced stabilization dropped by approximatively 6 °C for all ligands (figure **3.14a**). With 10 µM competitor (50 equiv), the induced stabilization was even lower, but remained significant. These results demonstrate some selectivity of the ligands for G4 against DNA duplexes.

Taking into account the different topology adopted by the human telomeric sequence at sodium buffer conditions, the previous experiment was repeated in the presence of sodium in order to appreciate any topology preference. The resulting ΔT_m was significantly lower, except for the **β-6-dglc-NDI (63)** derivative whose stabilization ability was even better than at potassium conditions (figure **3.14b**). Their general behavior suggested that the compound-induced stabilization depended on the G4-topology. The competition assay results showed again a reduction in ΔT_m value.
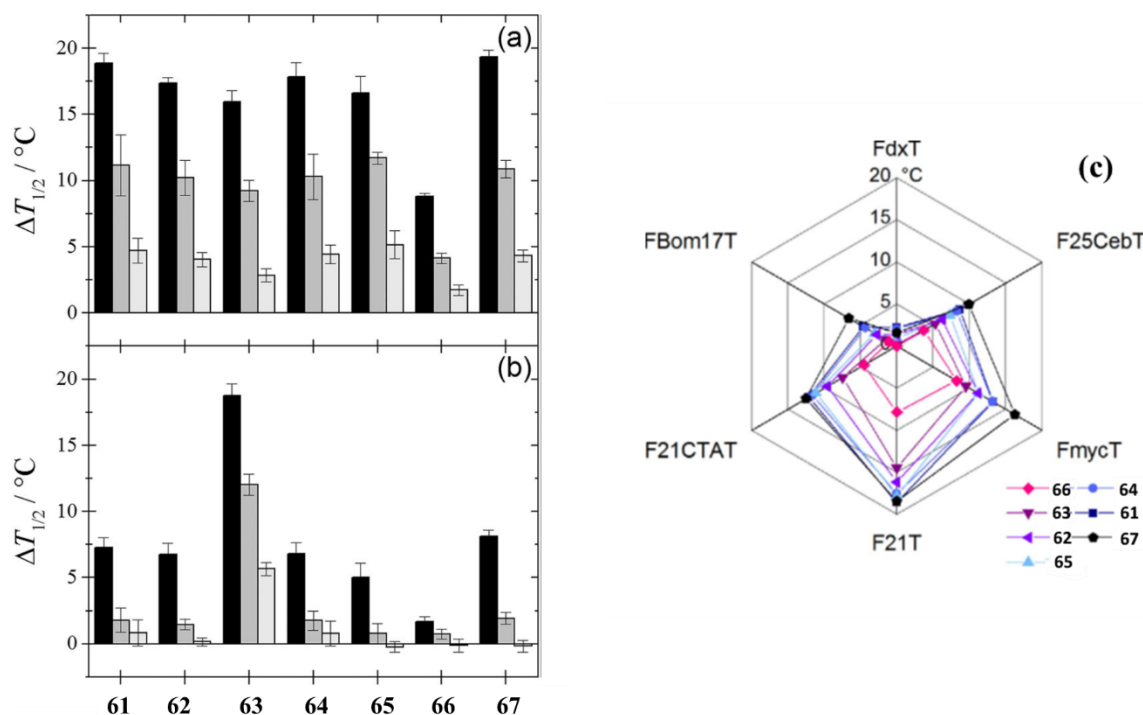
Figure 3.14. FRET melting competition assay results. Thermal stabilization induced by the tested compounds (2 µM) on the hTel quadruplex F21T (0.2 µM) in 10 mM lithium cacodylate pH 7.2 containing: a) 10 mM KCl and 90 mm LiCl, or b) 100 mM NaCl. The duplex competitor (ds26) strand concentrations were 0 (black), 3 (gray), and 10 µM (light gray).c) Selectivity profile results determined by FRET melting stabilization induced by the tested compounds (2 µM) on G4s of various topologies in 10 mM lithium cacodylate, pH 7.2, containing 10 mM KCl and 90 mM LiCl.

To provide further insight into the structural dependencies of ligand binding, FRET melting selectivity profiles (figure **3.14c**) were built from the ΔT_m values obtained with a set of sequences that included a duplex and five G4-forming sequences. Each G4-sequence represents a conformational type; hence we explored the ligands selectivity for the different G4-topologies (see table **2.3** from previous chapter for the type of G4-topology for each sequence). In this experiment

the selectivity towards G4-arrangements *versus* dsDNA was more evident than with the competition assay, they barely increased the melting temperature, which was in agreement with the results obtained in the competitive assay. There was not a clear selectivity for a certain G4-topology, although the compounds showed higher stabilization of the polymorphic human telomeric F21T G4 structures.

Our carbohydrate-NDI series was not among the strongest NDIs G4-ligands reported, still they were the most selective binding quadruplex structures *versus* duplex DNA. It is quite difficult to compare results with other reported trisubtituted NDIs due to the variability on the experimental FRET melting conditions: buffer, DNA and ligand concentration. Nevertheless, previously reported 3-NDI derivatives with the same amino side chain (figure **3.15**) displayed higher ΔT_m value for the hTel (26.5°C) but stabilized up to 12.5 °C the DNA duplex form[274] while our ligands barely affect its T_m still keeping a strong G4-stabilization ability. The different carbohydrate units influenced in the global G-quadruplexes stabilization without altering the overall selectivity profile. As we already discussed for our carbohydrate-PTDZ derivatives in the previous chapter, the ligand G4-selectivity is a crucial point to advance to a potential drug and avoid toxicity since the duplex DNA concentration is much higher than the G4 concentration inside the cell.
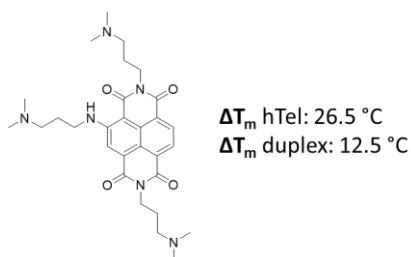
Figure 3.15.3-NDI reported by Cuenca *et al.*[274] with the ΔT_m for the hTel and a duplex DNA sequence.

3.2.2. CD EXPERIMENTS.

We performed CD titrations with four carb-NDI conjugates on the hTel quadruplex 21G d($G_3T_2AG_3T_2AG_3T_2AG_3$) sequence to confirm the strong stabilization observed in FRET melting assays was due to the interaction with the G4 and not due to the intrinsic fluorescence properties of the carb-NDIs. In all cases, the interaction of the ligand with the 21G sequence was evidenced by an intensity increase of the λ=287 nm positive band (typical of [3+1] topology), without any change in G4 folding (figure **3.16**). This effect is very similar to the described for trisubtituted NDIs, for instance the alkylating agent shown in figure **3.17**.[294] In addition, no CD signature for the free ligands were observed at this wavelength thus the observed CD changes was due to an interaction between the ligands and the G4 structure formed by the 21G sequence.
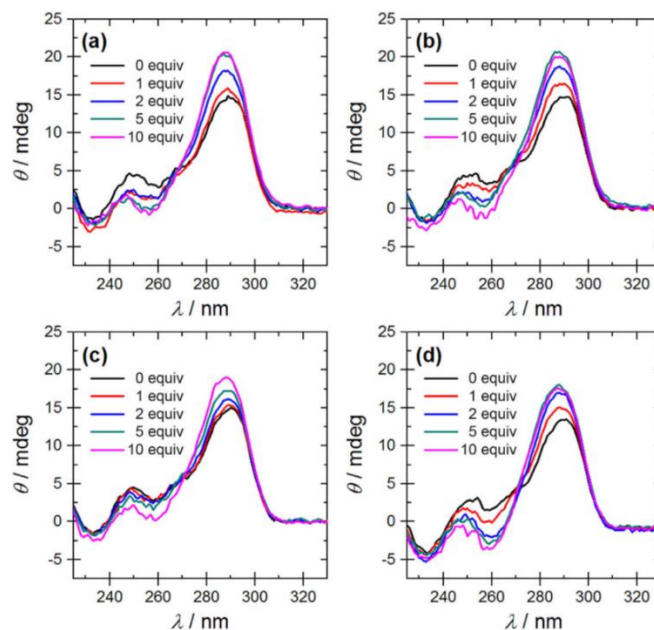
Figure 3.16. CD titration of 21G with compounds: a) **β-glc-NDI (61)**, b) **β-malt-NDI (64)**, c) **β-glcC2-NDI (66)**, and d) **prop-NDI (67)**. The oligonucleotide concentration was 4 µM.
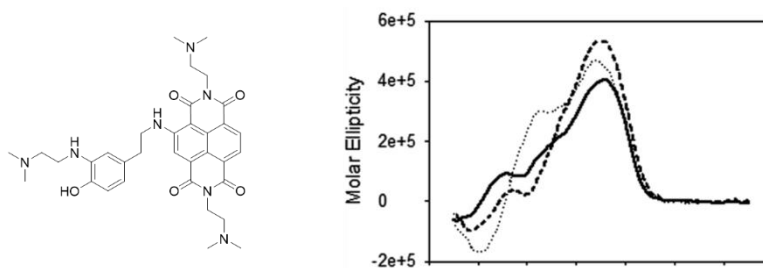
Figure 3.17. 3-NDI reported by Doria *et al.*[294] and the corresponding CD spectra with hTel incubated with after (dashed lines) or before (dotted lines) DNA annealing in the presence of K⁺ 50 mM. The spectrum of hTel DNA annealed in the absence of NDIs is shown as a solid line.

3.2.3. NMR STUDIES.

Dr. De-Rache performed titration experiments using NMR spectroscopy. We studied the interaction of the new carb-NDIs with the hTel quadruplex sequence 24TTA in potassium buffer conditions. As we mentioned previously, the 24TTA sequence folds into an intramolecular (3+1) form with an imino region characterized by twelve sharp signals corresponding to a 3-quartet quadruplex.
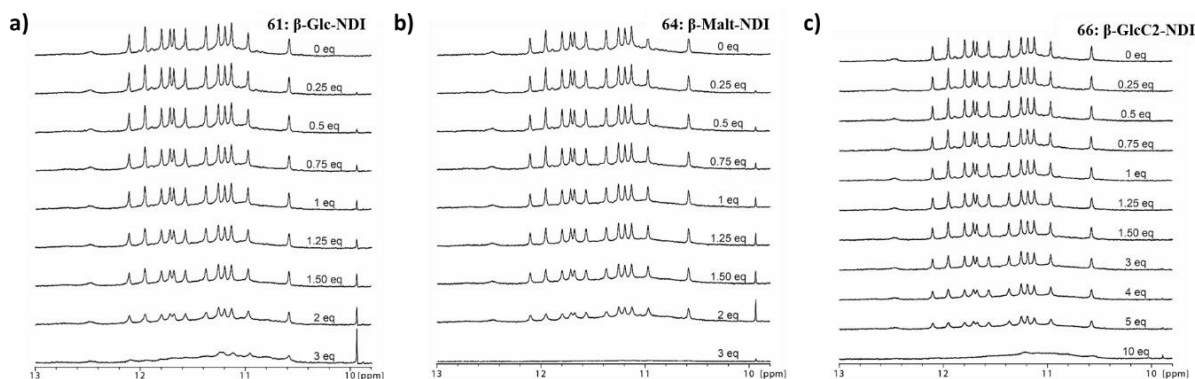
Figure 3.18. NMR titration experiments with some sugar-NDI and 24TTA sequence.

The selected compounds **β-glc-NDI** (**61**), **β-malt-NDI** (**64**) and **β-glcC2-NDI** (**66**) were added to the 24TTA NMR sample stepwise up to 3 equivalents except **β-glcC2-NDI** derivative up to 10 equivalents (figure **3.18**). The spectra of the resulting complexes were considerably different from the obtained with the phenyl ditriazole derivatives. In this case, a gradual decrease and broadening of the NMR peaks was observed without any noticeable signal shift. For the **β-glc-NDI** and **β-malt-NDI** derivatives part of the decrease in intensity can be attributed to a precipitation of the ligand-G-quadruplex complex in the NMR tube at high ligand concentrations. It is difficult to draw conclusions from these experiments due to the precipitation observed for some ligands and the lack of chemical shift changes of the G4 DNA imino signals.

3.3.   Biological studies.

The biological studies were focused on the potential antiproliferative activity of the new carb-NDI series and the potential difference in cell uptake for the carb-NDI.

3.3.1.   *In vitro* ANTIPROLIFERATIVE STUDIES.

The cytotoxicity of the new carb-NDI series has been evaluated over different human cancer cell lines (HT-29, HeLa and MCF-7) by using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) methodology at 48 h, while a non-tumoral cell line (MRC-5) was also included.

The IC_{50} results for each compound (table **3.1**) indicated that the control **prop-NDI** (**67**) displayed the lowest IC_{50} values in all cancerous cell lines (0.13–0.29 µM), closely followed by ethyl-linked mannose and glucose carb-NDIs **β-glcC2-NDI** (**66**) and **α-manC2-NDI** (**65**) (0.15–0.42 µM) and spacer-free **β-6-dglc-NDI** (**63**) (0.35–0.69 µM). Carb-NDIs (**β-glc-NDI 61**, **β-6-dglc-NDI 63**, **α-manC2-NDI 65** and **β-glcC2-NDI** (**66**) displayed less toxicity for the noncancerous cell line within a factor of 1.4- to 5.6-fold. Moreover, the presence of the carbohydrate reduced the toxicity in MRC-5 cells (up to fivefold for **β-malt-NDI**) compared with that of the control **prop-NDI,** which was the most toxic compound of the series. These data could seem worse than the reported for the corresponding dimethylamino trisubtituted NDI (structure in figure **3.15**). But in fact, the comparison is not feasible since such experiments were carried out by

a different methodology (SRB assay) and at longer incubation times (96 h), with obvious lower IC$_{50}$ values (5.4 nM in MCF-7 cell line).[274]

Compound	HT-29	MCF-7	HeLa	MRC-5
β-Glc-NDI (61)	2.92 ± 0.48	1.37 ± 0.81	1.56 ± 0.70	0.89 ± 0.33
β-GlcNAc-NDI (62)	2.26 ± 1.03	1.06 ± 0.01	0.95 ± 0.76	0.51 ± 0.01
β-6-dGlc-NDI (63)	0.40 ± 0.08	0.69 ± 0.71	0.35 ± 0.05	0.91 ± 0.32
β-Malt-NDI (64)	1.85 ± 0.19	1.42 ± 0.38	0.54 ± 0.34	2.04 ± 0.05
α-ManC2-NDI (65)	0.42 ± 0.05	0.15 ± 0.04	0.29 ± 0.14	0.84 ± 0.28
β-GlcC2-NDI (66)	0.36 ± 0.13	0.24 ± 0.16	0.24 ± 0.03	0.73 ± 0.17
Prop-NDI (67)	0.12 ± 0.02	0.07 ± 0.01	0.29 ± 0.18	0.42 ± 0.13

Table 3.12. IC$_{50}$ values (μM ± standard deviation) of the new NDI compounds measured on different cell lines.

On the other hand, the new NDI family improved the results reported for the trisubstituted-NDI alkylating agents shown in figure **3.17**, assayed with the same MTT procedure where IC$_{50}$ remained in the micromolar range for all of them (1-4 μM).[293]

In summary, carb-NDIs present great potential as anticancer drugs despite of their toxicity in MRC-5 cells. Actually, we have to take into account that MRC-5 cells are a modified human cell line with an altered cell division mechanism to become immortal, which is not the faithful reproduction of the *in vivo* model. For this reason, we always have to be cautious with toxicity results on non-cancer cell lines.

3.3.2. CELL UPTAKE STUDIES.

The different cytotoxicity found for the carb-NDI family and the control prop-NDI could be due to their intrinsic targets or due to different cell uptake. Thus, we decided to study in detail the cell entrance of these new compounds. The inherent red fluorescence of the NDI scaffold provides the possibility of exploiting techniques such as flow cytometry, fluorescence emission reading and confocal microscopy to monitor their cell entrance at different conditions, to quantify and to locate the compounds inside the cells, respectively.

We selected three structurally different NDI derivatives: **β-glc-NDI (61)**, with a sugar directly attached to the triazole unit; **α-manC2-NDI (65)**, with a mannose linked through a spacer; and **prop-NDI (67)**, with no sugar attached to the central core; to study their entrance within two cancer cell lines (HT-29 and MCF-7) and a noncancerous cell line (MRC-5).

3.3.2.1. Flow cytometry.

Flow cytometry recounts a certain number of events (cells in our case) with the chosen characteristics and quantifies the total fluorescence emitted by such events. We chose the events with a particular size (forward scattered) and complexity (side scattered), corresponding to each cell line. Although every cell line displays a slight inherent fluorescence (control sample), those which have internalized the corresponding NDI increase considerably their intensity and the global fluorescence measure will rise with the number of events in that condition. For this reason, flow cytometry allows to monitor the cell uptake of the new NDI-derivatives at different experimental conditions.

We incubated HT-29, MCF-7 and MRC-5 cell lines at different NDI concentrations (1 and 5 μM), different times (5 min, 1 h, 2 h, and 4 h), and two temperatures (4 and 37°C). We observed the quick entrance of sugar-free **prop-NDI** (**67**) in all cell lines, independent of the temperature and times assayed (similar uptake at 4°C after 5 min and at 37 °C after 4 h) (figure **3.19**). This suggested the control **prop-NDI** is uptaken by passive diffusion or facilitated transport. In contrast, at 4°C, **β-glc-NDI** (**61**) barely penetrated into the cells, whereas **α-manC2-NDI** (**65**) was incorporated slightly better, being more efficient in its cellular transport. When the temperature was increased to 37°C, **β-glc-NDI** and **α-manC2-NDI** were incorporated much more efficiently than at 4°C; indicating a temperature-dependent mechanism of cell uptake operating in the process. Moreover, **α-manC2-NDI** seems to enter the cells more efficiently than **β-glc-NDI**, so the presence of the linker or the different sugar could be playing a key role. In figure **3.20** flow cytometry results with HT-29 and the three compounds assayed are shown. We hypothesized that the observed temperature-dependence uptake in the sugar derivatives may likely be due to the participation of membrane-proteins in their internalization, such as GLUT transporters, so we planned experiments to confirm this possibility.
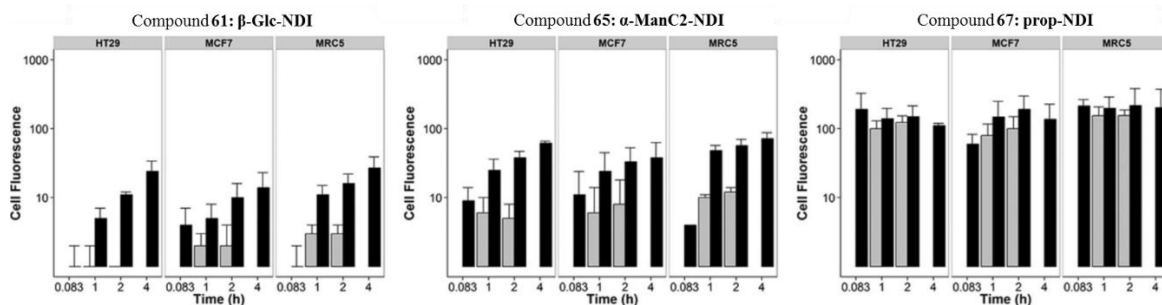
Figure 3.19. The mean fluorescence intensity of **β-glc-NDI**, **α-manC2-NDI**, and **prop-NDI** incubated at a concentration of 5 μM with various cancer and normal cell lines for 5 min, 1, 2, and 4 h, as measured by flow cytometry. Bars in gray, T=4 °C; bars in black, T=37°C.
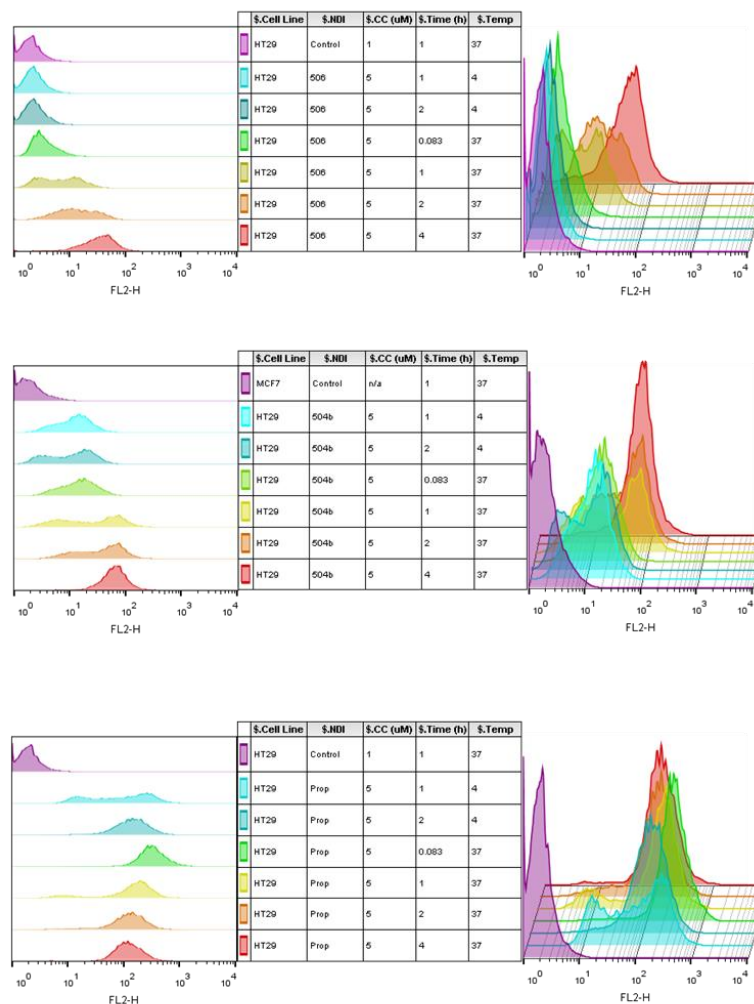
Figure 3.20. Flow cytometry results with HT-29 and **β-glc-NDI** (506), **α-manC2-NDI** (504b) or **prop-NDI** (Prop) incubated at 5 min, 1 h, 2 h and 4 h, and at 4°C and 37°C.

3.3.2.2. Fluorescence quantification.

Flow cytometry provided information about the required time and temperature for the assayed compounds to enter inside each cell line, but not the actual amount of internalized product. To do so, we carried out intracellular quantification of **β-glc-NDI (61)**, **α-manC2-NDI (65)**, **β-glcC2-NDI (66)**, and **prop-NDI (67)** to confirm the apparent differences observed in cellular uptake among the four ligands. Fluorescence spectroscopy measurements were made directly by irradiating the red dye at λ=485 nm and recording its emission at λ=535 nm after incubation with the compounds at 5 µM for 2 h (table **3.2**). As expected from the flow cytometry studies, we observed higher intracellular concentrations of control **prop-NDI** than for the rest of the carbohydrate-NDI ligands in all three cell lines examined. At the same time, cellular uptake of **α-manC2-NDI** and **β-glcC2-NDI** was considerably more efficient (two- to threefold) than that for **β-glc-NDI**. Compounds **α-manC2-NDI** and **β-glcC2-NDI** both comprise a monosaccharide linked through an ethylene glycol spacer to the triazole–NDI scaffold, whereas **β-glc-NDI** has glucose directly linked to the core; this confirms the tendency observed by flow cytometry on HT-29 and

MRC-5 cells, but not on MCF-7 cells (0.8- to 1.4-fold). Our results are in contrast with recent studies by Patra *et al.*,[213] who used glucose–platinum conjugates. Contrary to our findings, they concluded that increasing the length of the linker decreased cell uptake. However, the different sugar position substitution, nature of the core and dimensions may also influence the transport mechanism and could explain these divergent results.

Compound	Cell line		
	HT-29	**MCF-7**	**MRC-5**
β-glc-NDI (61)	453 ± 65 (30%)	526 ± 120 (28%)	236 ± 53 (13%)
α-manC2-NDI (65)	1093± 159 (73%)	439 ± 159 (23%)	545 ± 176 (30%)
β-glcC2-NDI (66)	988 ± 121 (66%)	762 ± 241 (41%)	778 ± 269 (43%)
Prop-NDI (67)	1492 ± 508 (100%)	1858 ± 239 (100%)	1773 ± 98 (100%)

Table 3.13. Compound concentration (nM) per well and percentage of compound entering the cell relative to compound **NDI prop (67)** (values in parentheses) of **β-glc-NDI (61)**, **α-manC2-NDI (65)**, **β-glcC2-NDI (66)**, and aglycone **prop-NDI (67)**, as measured by fluorescence spectroscopy.

3.3.2.3. GLUT inhibition experiments.

We explored the possibility that GLUTs transporters could be the responsible of the sugar-NDI derivatives cell uptake, and their entrance variations may result from the different affinity through the sugar presentations and their accessibility on the NDIs. We decided to use GLUT inhibitors which are capable of blocking the potential entrance of the carb-NDI derivatives. Thus, we examined this possibility by quantifying the entrance (also by fluorescence spectrometry) of carb-NDIs (**β-glc-NDI 61**, **α-manC2-NDI 65**, and **β-glcC2-NDI 66**) and control **prop-NDI 67** into cancerous HT-29 cells in the absence or presence of several GLUT inhibitors, including WZB117, a specific GLUT1 inhibitor;[306] ritonavir, a specific GLUT4 inhibitor;[307] and flavonoids quercetin and genistein, which are reported to be dual GLUT1 and GLUT4 inhibitors.[308] We observed clear cell uptake inhibition for **β-glc-NDI**, **α-manC2-NDI**, and **β-glcC2-NDI** (35–64% inhibition) when quercetin was used as an inhibitor (figure **3.21**). A smaller but similar effect was also perceived with genistein, another GLUT1 and GLUT4 inhibitor, which decreased carb-NDI uptake by 43–44% of **α-manC2-NDI** and **β-glcC2-NDI** compounds. This inhibition was not observed for **β-glc-NDI**, in which the carbohydrate is directly linked to the triazole-NDI scaffold. Unexpectedly, the presence of the specific GLUT1 inhibitor WZB117 yielded no significant cell uptake differences with inhibitor-free control for all carb-NDIs examined. In addition, the control **prop-NDI** showed a small cell uptake inhibition (8–24%) in the presence of the GLUT inhibitors examined, but this small effect was previously reported for other aglycone drugs, such as cyclohexane-1,2-diamine platinum derivatives, when their cell uptake was assayed in the presence of phloretin or 4,5-*O*-ethylidene-α-D-glucose GLUT inhibitors.[213]

HT29-uptake Inhibition (%)

Inhibitor	Glut Target	Prop-NDI 67	β-glc-NDI 61	α-manC2-NDI 65	β-glcC2-NDI 66
Fasentin	Glut 1	18	-9	14	1
ZWB117	Glut 1	20	-19	7	18
Quercetin	Glut 1 & 4	24	54	35	64
Genistein	Glut 1 & 4	9	17	43	44
Ritonavir	Glut 4	8	0	28	59

Inhibition (%)

% < 25	○	10 Small
25 > % > 40	◑	30 Medium
% > 40	●	40 High

MRC5-uptake Inhibition (%)

Inhibitor	Glut Target	Prop-NDI 67	β-glc-NDI 61	α-manC2-NDI 65	β-glcC2-NDI 66
Fasentin	Glut 1	1	5	14	-8
ZWB117	Glut 1	0	-11	-1	-11
Quercetin	Glut 1 & 4	4	9	37	-10
Genistein	Glut 1 & 4	14	17	28	1
Ritonavir	Glut 4	-2	-7	19	-24

Figure 3.21. Cell uptake inhibition in HT-29 and MRC-5 cancer cells of **β-glc-NDI**, **α-manC2-NDI**, **β-glcC2-NDI**, and **prop-NDI** in the presence of several GLUT inhibitors, as determined by intracellular concentration measured by means of fluorescence spectroscopy. Preincubation with inhibitors (100 µM) for 1 h was followed by incubation with NDI compounds (5 µM) for 2 h.

Taking into account the effect of ZWB117 and flavonoids GLUT inhibitors, GLUT1 seems to have a minor role on the carb-NDIs cell uptake. To confirm that GLUT4 played an important role in the cell entrance of **α-manC2-NDI** and **β-glcC2-NDI**, we carried out cell uptake inhibition assays in the presence of the specific GLUT4 inhibitor ritonavir (figure **3.21**). Cellular uptake of **α-manC2-NDI** and **β-glcC2-NDI** compounds decreased in particular without affecting either **β-glc-NDI** or control **prop-NDI**. Our results revealed the main role of GLUT4 translocating compounds **α-manC2-NDI** and **β-glcC2-NDI**, which contain mannose and glucose, respectively, attached through a short linker to the triazole–NDI scaffold. GLUT1, which was previously identified as the main GLUT involved in cellular uptake of other glucose-conjugated drugs, seemed to be playing a minor role, if any, in translocating **α-manC2-NDI** and **β-glcC2-NDI**. It is important to note that the functional groups, as well as the size and shape of the attached drug, may also have a strong influence on the interaction with GLUT and its resulting translocation process. Moreover, other GLUT isoforms could also be involved in the cell uptake of these carbohydrate-conjugated G4 ligands; this is especially pertinent for **β-glc-NDI**, which possesses glucose directly linked to the triazole–NDI structure.

When we examined the influence of GLUT inhibitors in cell uptake of NDI derivatives **β-glc-NDI**, **α-manC2-NDI**, **β-glcC2-NDI**, and **prop-NDI** in noncancerous MRC-5 cells (figure **3.21**), we observed no inhibition with any of the GLUT inhibitors tested, which was in concordance with the supposed lower GLUTs level in these non-tumoral cells. The exception was **α-manC2-NDI** which was partially inhibited by quercetin and genistein (28 and 37 %, respectively).

Whereas important cell uptake differences were observed between control **prop-NDI** and glycoconjugated NDI derivatives, there was no clear correlation with the IC$_{50}$ values. However, when comparing the sugar-derived NDIs, some correlation is observed. Cell uptake of **β-glcC2-NDI** into HT-29 and MCF-7 cells, which was two- to threefold more efficient than that for **β-glc-NDI**, correlated with the 5.7- to eightfold higher cytotoxicity IC$_{50}$ values.

In conclusion, carb-NDI cell uptake correlated to some extent with their cytotoxicity properties, as observed for **β-glcC2-NDI**, which entered the cells more efficiently and was

significantly more toxic in both HT-29 and MCF-7 cells than **β-glc-NDI**. Selectivity towards cancerous cell lines also seemed to be dependent on sugar type and attachment, resulting in punctual selectivity gain compared with that of aglycone **prop-NDI**. Notably, cell uptake studies in the presence of a variety of GLUT inhibitors showed that **α-manC2-NDI** and **β-glcC2-NDI** were transported at least partially through GLUT4.

3.3.2.4. Confocal microscopy studies.

To finalize these studies, we performed confocal microscopy analysis to determine the location of **β-glc-NDI** (**61**), **α-manC2-NDI** (**65**), **β-glcC2-NDI** (**66**), and **prop-NDI** (**67**) compounds in a cancerous (HT-29) and a non-cancerous (MRC-5) cell lines, with the technical support of Dr. Jenny Campos (IPBLN). We assayed different incubation times (5 min, 2h, 4h, 24h) and ligand concentrations (1 and 5 µM). The best experimental conditions were the ones with a vibrant fluorescence emission for all ligands without an evident cytotoxicity. Taking into account this postulation, cells incubation at 5 µM for 2h were the chosen conditions. After incubation at 37°C for 2 h, all four compounds were detected inside the cells and were mostly localized in the nucleus (figure **3.22**).
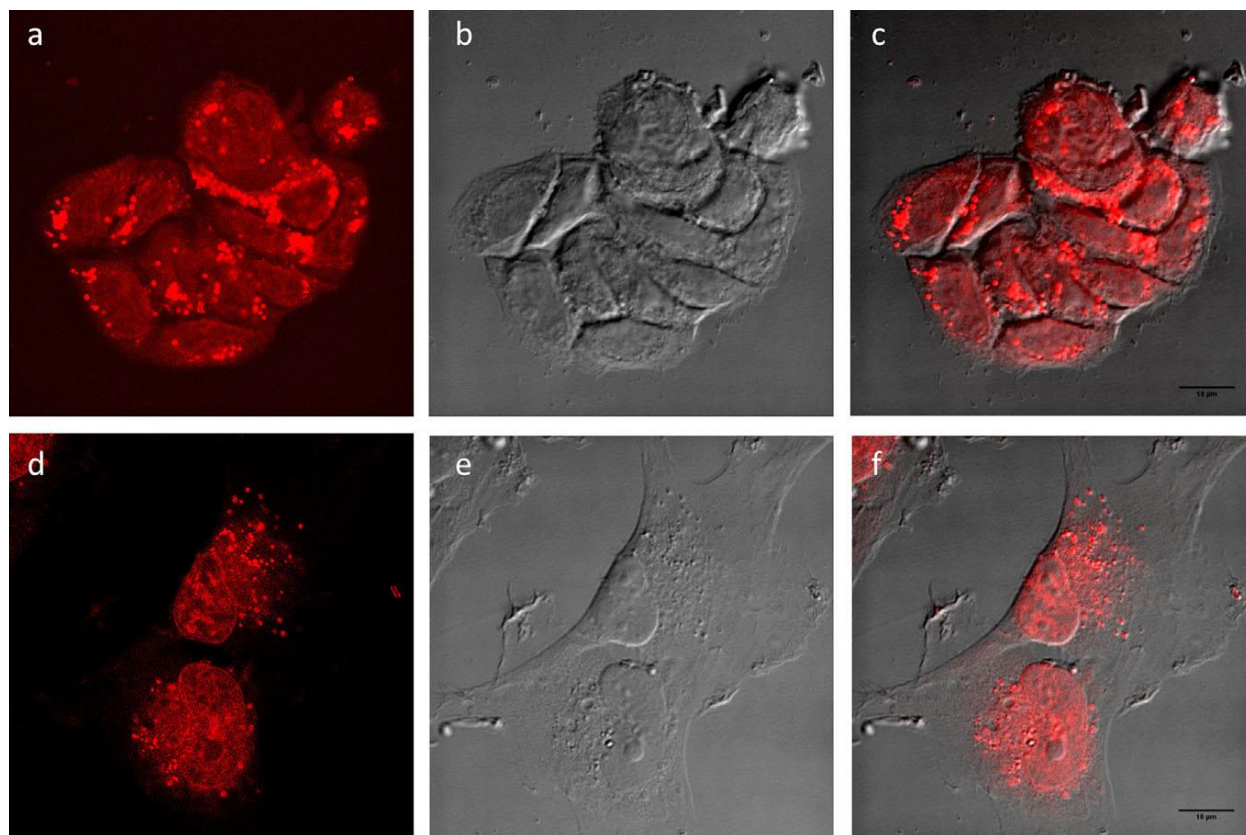
Figure 3.22. Confocal images of HT-29 and MRC-5 cells after co-incubation with **α-manC2-NDI** (**65**) (5 µM) for 2 h. HT-29 cells: a) fluorescence image of 5, b) transmitted light image (differential interference contrast (DIC)), and c) merged image; MRC-5 cells: d) fluorescence image of compound 5, e) transmitted light image (DIC), and f) merged image. Scale bar: 10 µm.

These experiments confirmed that glycoconjugated NDIs may reach their DNA G4 target within the nuclear genome. Similar results were found with control **prop-NDI**, consistent with the

location of other reported NDI derivatives, such as Neidle's tetracharged NDI MM41 in MIA-PaCa-2 cell line.[301] It is important to note that our carbohydrate-modified NDIs possess only two positively charged groups and still locate at the nucleus. So, the four charges substituting the NDI core may not be indispensable for cytotoxicity, two would be enough to have the antiproliferative effect with the possibility to introduce other groups, such as carbohydrates, in order to improve the pharmacological features and the cancer tissue selectivity.

3.3.3. ZEBRA FISH FAST ACUTE TOXICITY ASSAYS.

Next, we measured toxicity of the prepared carb-NDIs in an animal model. We selected the use of Zebra fish due to the significant reduction in time and cost when compared to mammal models. Zebrafish presents extensive homology with human genome and many fundamental cellular and molecular pathways involved in the response to chemicals or stress are conserved between the zebrafish and mammals. We decided to measure the Zebra fish fast acute toxicity assay.[309] The experiments were carried out by ZeClinics® (Barcelona, Spain) through incubation of zebrafish embryos in 5 different logarithmic concentrations (0.01 μM, 0.1 μM, 1μ M, 10 μM and 100 μM) of the NDIs (**β-glc-NDI 61**, **α-manC2-NDI 65**, **β-malt-NDI 64**, **β-glcNAc-NDI 62**, **β-6-dglc-NDI 63**, **β-glcC2-NDI 66** and **prop-NDI 67**) to determine the potential toxic effect after 96 hours of drug incubation. After 96 h Zebra fish embryos incubation with compounds at each concentration, the technicians determined:

- LC50 (median lethal dose), calculated by fitting sigmoidal curve to mortality data.
- NOEC (No observed effect concentration, with mortality score >20% assumed as the effect).
- LOEC (Lowest observed effect concentration, with mortality score >20% assumed as the effect).

Taking into account the IC_{50} values obtained on MRC-5, the results described on table **3.3** were completely unexpected since the LC50 remained higher than 100 μM for all compounds (sometimes too high to be calculable at such concentration range).

Compound	NOEC (μM)	LOEC (μM)	LC50 (μM)	LD50 rodents (μM/kg)	LD50 rodents (mg/kg)
Prop-NDI (67)	10	100	111,35	~111,35	~65.2
β-GlcC2-NDI (66)	100	N.A.	N.A.	~>100	~>96.6
β-Glc-NDI (61)	100	N.A.	N.A.	~>100	~>92.3
α-ManC2-NDI (65)	100	N.A.	N.A.	~>100	~>96.7
β-6-dGlc-NDI (63)	100	N.A.	172,5	~172,5	~156.0
β-GlcNAc-NDI (62)	100	N.A.	N.A.	~>100	~>96.3
β-Malt-NDI (64)	100	N.A.	195,09	~195,09	~211.6

Table 3.14. Summary table of Zebra Fish acute toxicity assay. NA (Not available): when NOEC represents the maximum concentration (1 mM), LOEC is not calculable. Correlation between the LD50 obtained in Zebra fish embryo toxicity assay and LD50 in rodents (mg/kg).

Ali *et al.*[310] reported a correlation between zebrafish embryo Log LC50 (mmol/L medium) and rodent Log LD50 (mmol/kg rodent) for the 60 compounds tested to predict toxicity in rodent,

with very similar values in most of them (figure **3.23**). According to this model, we can converse the LC50 obtained for our carb-NDI in Zebra fish embryo toxicity assay into LD50 in rodents to obtain an approximation of the compounds toxicity in rodents. The results are shown in table **3.3**.
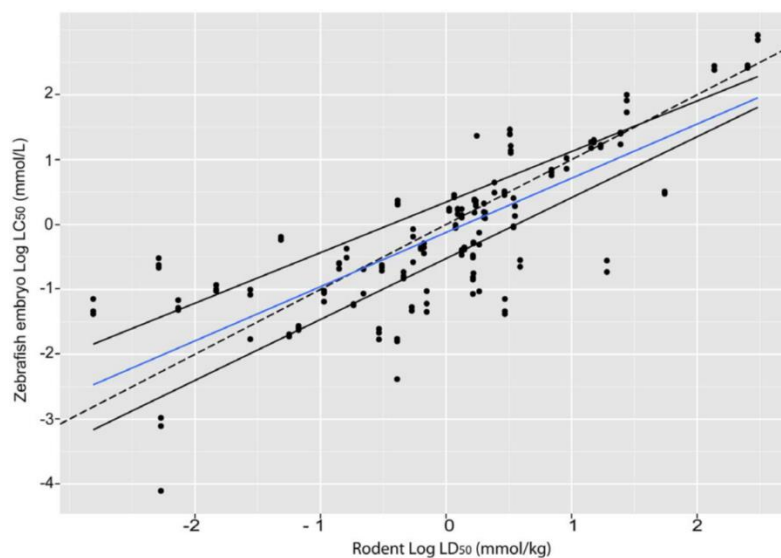
Figure 3.23. Correlation between zebrafish embryo Log LC50 and rodent Log LD50 for the 60 compounds tested in the study.[310] Key: blue, regression line; solid black lines, 0.25 and 0.75 quartiles; dashed line, perfect correlation line. The slope of the regression line (blue) is 0.73403.

Although **prop-NDI** exhibited low NOEC and LOEC values, LC50 was still high and the presence of the carbohydrate in the carb-NDI compounds seem to diminish their toxicity in comparison to the control **prop-NDI**. Pictures of the assay with the compound **β-glcC2-NDI** is shown in figure **3.23**.
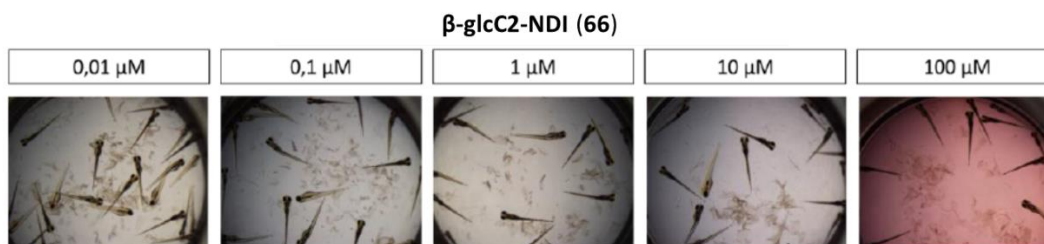
Figure 3.24. Example of Zebra fish acute toxicity assay with **β-glcC2-NDI** (**66**), with pictures of zebra fish embryos after 96 h incubation with five **β-glcC2-NDI** concentrations.

These results suggested the new carb-NDI series was not as toxic as we expected from the *in vitro* cell cytotoxicity experiments. In fact, this low toxicity and the low micromolar IC_{50} values observed on the examined cancer cell lines encourage us to carry out efficiency studies of the carb-NDIs on a cancer animal model.

4. CONCLUSIONS.

- Novel carb-NDI derivatives have been successfully designed and synthesized in collaboration with Mauro Freccero's research group following the click chemistry approach.

- The sugar did not provide any detrimental effect to the G4-binding. The new NDI derivatives displayed strong affinity towards G4-structures, especially the human telomeric sequence, since they significantly increased its T_m. In addition, they were G4-selective since they barely affected duplex DNA melting temperature, although a clear selectivity for a G4-topology was not observed.

- The new carb-NDIs showed cytotoxicity on several cancer cell lines (MCF-7, HeLa and HT-29) with IC_{50} values near the nanomolar range, except for **β-glc-NDI** (**61**), **β-malt-NDI** (**64**) and **β-glcNAc-NDI** (**62**), which remained in the low micromolar range. The lowest IC_{50} value was displayed by the aglycone **prop-NDI** (**67**) derivative while it was the least selective as well. The presence of the sugar increased the selectivity to cancer cell lines and the C2-linker-sugar derivatives exhibited better results than the ones directly attached to the triazole-NDI scaffold.

- We observed a correlation between NDI derivatives cell uptake and their cytotoxicity. In fact, the most toxic compound (**prop-NDI**) was quickly internalized in non-temperature dependent manner, which suggested this compound was uptaken by passive diffusion or facilitated transport. Among the carb-NDIs, the better cell uptake of the C2-linker sugar NDIs also correlated with a higher cytotoxicity than those carb-NDIs with the carbohydrate directly linked to the triazole-NDI core.

- The carb-NDIs were mainly transported through GLUT and dependent on time, temperature, type of carbohydrate, and sugar accessibility. According to the GLUT inhibition experiments, **α-manC2-NDI** (**65**) and **β-glcC2-NDI** (**66**) seemed to be transported, at least partially, through GLUT4.

- Confocal microscopy studies confirmed that carb-NDIs could reach the nucleus where their intended genomic G4 DNA targets are located.

5. EXPERIMENTAL DETAILS.

5.1. Synthesis of the new carbohydrate-NDI derivatives.

5.1.1. <u>MATERIAL AND METHODS.</u>

All reagents and solvents were purchased by Sigma Aldrich, Carbosynth, Fluka or Merck, and used without further purification. All reactions were monitored by TLC, on silica gel plates (60F254 (Merck) and visualized with UV light or stained with mostain (500 ml of 10% H_2SO_4, 25 g of $(NH_4)_6Mo_7O_{24}$ ·$4H_2O$, 1 g $Ce(SO_4)_2$ $4H_2O$), sulphuric acid (10% H_2SO_4 in ethanol), anisaldehyde (450mL ethanol, 25mL anisaldehyde, 25 mL H_2SO_4 and 1 mL AcOH), or ninhydrin (0.25g ninhydrin, 100mL ethanol). Products were purified by flash chromatography with silica gel 60 (0.5-0.2 mm, 0.2-0.063 mm, and 0.040-0.015 mm) or alumina (Al_2O_3 basis) type WN6.

NMR spectra were recorded on either a Bruker AVANCE 300 or ARX 400 MHz [300 or 400 MHz (^{1}H), 75 or 100 (^{13}C)] spectrometer, at room temperature for solutions in $CDCl_3$, CD_3OD,

D$_2$O or DMSO. Chemical shifts were referred to the solvent signal and were expressed in ppm. Mass spectra were performance in an Esquire 6000 Bruker-Daltonics (ESI) and Micromass Autospec-Q (EMAR-FAB) and a QTRAP (EMAR-IQ) in the Chemical Research Institute of Seville, and CITIUS (Center of Research, Technology and Innovation and of the University of Seville), respectively.

The click chemistry reaction was performed in Pavia, by the Dr. Filippo Doria (from Mauro Freccero's team). The characterization of the final NDI derivatives **β-glc-NDI (61), α-manC2-NDI (65), β-malt-NDI (64), β-6-dglc-NDI (63)** and **β-glcNAc-NDI (62)** was carried out in our laboratory.

5.1.2. SYNTHESIS OF CARBOHYDRATE-AZIDO DERIVATIVES.

The acetylated azido-sugar derivatives synthesis was achieved according to the methodology described in the previous chapter.

Synthesis of peracetylated 6-deoxy-glucopyranose

Synthesis of methyl 4,6-*O*-benzylidene-α-D-glucopyranose **75**

75
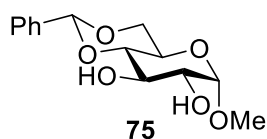
To a solution of 1-*O*-methyl-α-D-glucopyranose (6 g, 20.9 mmol, 1 eq) in anhydrous DMF, benzaldehyde dimethyl acetal (9.3 mL, 61.8 mmol, 2 eq) and *p*-toluenesulfonic acid monohydrate (catalytic amounts) were added. The mixture was stirred at 50°C in a rotary evaporator under diminished pressure for one hour.[304] Then, the solvent was evaporated and the crude (13 g) was used without further purification in the following step.

Synthesis of methyl 2,3-di-*O*-acetyl-4,6-*O*-benzylidene-α-D-glucopyranose (**76**)

76
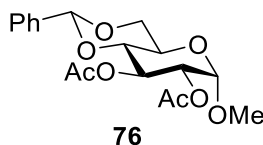
The previous crude was dissolved in dry pyridine (30 mL) and acetic anhydride (11.7 mL, 4 eq) was added. The reaction mixture was stirred for 20 h under argon atmosphere. Then, the mixture was washed with diluted HCl solution and distillated water. The organic phases were dried over anhydrous MgSO$_4$, filtered off and evaporated under reduced pressure. The crude obtained was purified by silica gel column chromatography using as eluent Hex/EtOAc (5:1 to 4:1) to afford **76** as a white solid (4.3 g, 39%, two steps).^{1}H NMR (300 MHz, CDCl$_3$) δ 7.35 (t, J = 9.6 Hz, 2H, 2xH$_{AR}$), 7.27 (d, J = 4.1 Hz, 3H, 3xH$_{AR}$), 5.51 (t, J = 9.5 Hz, 1H, H$_{SUGAR}$), 5.42 (s, 1H, Ph-C$\underline{H}$), 4.87 (s, 1H, H$_{SUGAR}$), 4.83 (d, J = 3.4 Hz, 1H, H$_{SUGAR}$), 4.22 (dd, J = 10.1, 4.6 Hz, 1H, H$_{SUGAR}$), 3.85 (td, J = 9.8, 4.6 Hz, 1H, H$_{SUGAR}$), 3.68 (t, J = 10.2 Hz, 1H, H$_{SUGAR}$), 3.57 (t, J = 9.6 Hz, 1H, H$_{SUGAR}$), 3.32 (s, 3H, 3xOC$\underline{H}_3$), 2.06–1.85 (m, 6H, 6xCOC$\underline{H}_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ

170.14 ($\underline{C}OCH_3$), 169.55 ($\underline{C}OCH_3$), 136.90 (C_{AR}), 128.90 ($\underline{C}H_{AR}$), 128.05 ($2x\underline{C}H_{AR}$), 126.05($2x\underline{C}H_{AR}$), 101.38 (Ph-$\underline{C}H$), 97.47 (C_1), 79.04 (C_4), 71.46, 68.85, 68.65 (C_2, C_3, C_5), 62.23 (C_6), 55.18 (O$\underline{C}H_3$), 20.62 (CO$\underline{C}H_3$), 20.56 (CO$\underline{C}H_3$). HRMS (FAB$^+$), calculated for $C_{18}H_{23}O_8$ (M+H): 367.1393; found: 367.1388.

Synthesis of methyl 2,3-di-*O*-acetyl-α-D-glucopyranose (**77**).

77
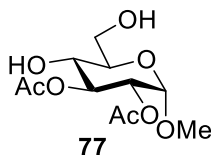
      Methyl 2,3-di-*O*-acetyl-4,6-O-benzylidene-α-D-glucopyranose (**76**) (4.4 g, 11.97 mmol, 1 eq) was solved in a mixture of DCM:MeOH (30:10, 80 mL). Camphor sulfonic acid (2.8 g, 11.97 mmol, 1 eq) was added to the reaction solution and the mixture was stirred at room temperature for one hour. After starting material consumption, the reaction was quenched with TEA and solvents were evaporated at reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (1:2 to 1:4) to give **77** as a white solid (2.3 g, 70%). ^{1}H NMR (300 MHz, CDCl$_3$) δ 5.36–5.19 (m, 1H, H$_3$), 4.87 (d, J=3.7 Hz, 1H, H$_1$), 4.77 (dd, J=10.1, 3.6 Hz, 1H, H$_2$), 3.83 (d, J=2.2 Hz, 2H, H$_5$, H$_6$), 3.74–3.62 (m, 2H, H$_4$, H$_{6'}$), 3.37 (s, 3H, 3xOC$\underline{H}_3$), 2.05 (d, J = 5.7 Hz, 6H, 6xCOC$\underline{H}_3$). ^{13}C NMR (75 MHz, CDCl$_3$) δ 171.53 ($\underline{C}OCH_3$), 170.45 ($\underline{C}OCH_3$), 96.84(C_1), 73.04, 71.26, 70.96, 69.33 (C_2, C_3, C_4, C_5), 61.72 (C_6), 55.22 (O$\underline{C}H_3$), 20.84 (CO$\underline{C}H_3$), 20.71 (CO$\underline{C}H_3$). HRMS (FAB$^+$), calculated for $C_{11}H_{18}O_8$ (M+H): 279.1074; found: 279.1099.

Synthesis of methyl 2,3-di-*O*-acetyl-6-bromo-6-deoxy-α-D-glucopyranose (**78**)

78
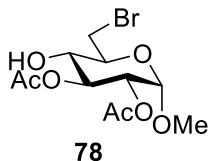
      To a solution of methyl 2,3-di-*O*-acetyl-α-D-glucopyranose **77** (500 mg, 1.8 mmol, 1 eq) in anhydrous acetonitrile, tetrabromomethane (744.8 mg, 2.2 mmol, 1.2 eq) and triphenylphosphine (593.8 mg, 2.26 mmol, 1.26 eq) were added. The reaction mixture was stirred for 72 h at room temperature and when starting material was consumed solvents were evaporated. The crude obtained was solved in DCM and washed with distilled water (3x25 mL). The organic phases were dried over anhydrous MgSO$_4$, filtered off and the solvent was removed under reduced pressure. The crude was purified by silica gel column chromatography using as eluent EtOAc/Hexane (1:1) to afford **78** as a white solid (449 mg, 73%). ^{1}H NMR (400 MHz, CDCl$_3$) δ 5.25 (t, J=9.6 Hz, 1H, H$_3$), 4.91 (d, J=3.6 Hz, 1H, H$_1$), 4.84 (dd, J=10.1, 3.2 Hz, 1H, H$_2$), 3.86–3.76 (m, 1H, H$_5$), 3.74 (dd, J=11.1, 2.1 Hz, 1H, H$_6$), 3.64–3.52 (m, 2H, H$_4$, H$_{6'}$), 3.41 (s, 3H, 3xOC$\underline{H}_3$), 2.07 (d, J=7.1 Hz, 6H, 6xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 172.12 ($\underline{C}OCH_3$), 96.86 (C_1), 73.40 (C_3), 71.68 (C_4), 70.67 (C_2), 70.49 (C_5), 55.49 (OCH_3), 32.82 (C_6), 20.96 (CO$\underline{C}H_3$). HRMS (ES$^+$): calculated for $C_{11}H_{17}BrNaO_7$, (M+Na): 363.0055, found: 363.0056.

Synthesis of methyl 2,3-di-*O*-acetyl-6-deoxy-α-D-glucopyranose (**79**)

79
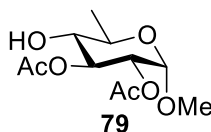
In a hydrogenation reactor, the bromated derivative methyl 2,3-di-*O*-acetyl-6-bromo-6-deoxy-α-D-glucopyranose (**78**) was dissolved in EtOAc under inert atmosphere. Catalytic amounts of Pd(OH)$_2$ were incorporated to the reaction mixture, which was hydrogenated at 4 atm at room temperature for 16 h. The mixture was filtered off over celite and solvents were removed obtaining the desired product **79** as a white solid (272 mg, 60%) without any purification step. ^{1}H NMR (400 MHz, CDCl$_3$) δ 5.19 (t, J=9.5 Hz, 1H, H$_3$), 4.81 (dd, J=5.1, 4.2 Hz, 2H, H$_2$, H$_1$), 3.69 (dd, J=8.7, 6.7 Hz, 1H, H$_5$), 3.35 (s, 3H,3xOC$\underline{H}$$_3$), 3.26 (t, J=9.4 Hz, 1H, H$_4$), 2.05 (d, J=5.6 Hz, 6H, 6xCOC$\underline{H}$$_3$), 1.28 (d, J=6.2 Hz, 3H, 3xCC$\underline{H}$$_3$). ^{13}C NMR (101 MHz, CDCl$_3$) δ 171.97 ($\underline{C}$OCH$_3$), 170.50 ($\underline{C}$OCH$_3$), 96.72 (C$_1$), 75.00 (C$_4$), 73.47 (C$_3$), 71.14 (C$_2$), 67.46 (C$_5$), 55.16 (OCH$_3$), 20.97 (CO$\underline{C}$H$_3$), 20.85 (CO$\underline{C}$H$_3$), 17.43 (CH$_3$). LRMS (ES$^+$): calculated for (M+Na): C$_{11}$H$_{18}$NaO$_7$: 285.1, found: 285.2. HRMS (ES$^+$): calculated for (M+Na): 285.0950, found: 285.0944.

Synthesis of 1,2,3,4-tetra-*O*-acetatyl-6-deoxy-α-D-glucopyranose (**80**)

80
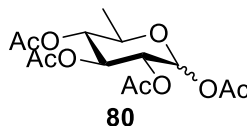
Compound **79** was dissolved in a mixture of acetic acid: acetic anhydride (1:1) and 2-3 drops of H$_2$SO$_4$ were added. The reaction was stirred at room temperature for 24 h until complete consumption of the starting material. Then, the mixture was diluted with diethyl ether and washed with NaHCO$_3$ solution (3x25 mL) and brine (1x25 mL). The organic phase was dried over anhydrous MgSO$_4$, filtered off and the solvent was evaporated at reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (2:1 to 1:1) yielding the desired product **80** (345 mg, 98%) as a white solid, as previously described.[311] LRMS (ES$^+$): calculated for (M+Na): C$_{14}$H$_{20}$NaO$_9$: 355.1, found: 355.3.

Synthesis of 2,3,4-tri-*O*-acetyl-1-azido-1,6-dideoxy- β-D-glucopyranose (**68**)

68
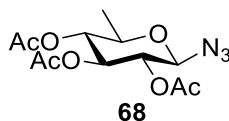
The peracetylated sugar derivative **80** (1026 mg, 3.088 mmol) was dissolved in anhydrous dichloromethane and hydrogen bromide solution 33 wt. % in acetic acid (2 mL/g of starting material) was added under argon atmosphere.[233] The reaction was stirred at room temperature until complete consumption of the starting material (3 h). The mixture was then diluted in dichloromethane and washed with ice, NaHCO$_3$ saturated solution (3x50 ml), and distillated water

(50 mL). The organic phase was dried over $MgSO_4$, filtered off and the solvent was evaporated under reduced pressure. The crude was solved in a mixture of acetone/water (5:1) and sodium azide (822.5 mg, 12.65 mmol, 4 eq) was incorporated to the mixture. The reaction was stirred at room temperature until complete conversion of the starting material. Then, the solvents were removed under reduced pressure and the crude was solved in water. The organic phase was extracted with EtOAc (3x30 mL), dried over anhydrous $MgSO_4$, filtered off and the solvent was evaporated under reduced pressure. The crude was purified by silica gel column chromatography using as eluent Hex/EtOAc (4:1 to 2:1) to give **68** (837 mg, 84%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.16 (t, J = 9.6 Hz, 1H, H$_3$), 4.92 (t, J = 9.3 Hz, 1H, H$_2$), 4.83 (t, J = 9.6 Hz, 1H, H$_4$), 4.58 (d, J = 8.9 Hz, 1H, H$_1$), 3.65 (tt, J = 12.3, 6.1 Hz, 1H, H$_5$), 2.12 – 1.93 (m, 9H, 9xCOC$\underline{H}_3$), 1.27 (d, J = 6.2 Hz, 3H, 3xCC$\underline{H}_3$). ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.31 ($\underline{C}OCH_3$), 169.68 ($\underline{C}OCH_3$), 169.40 ($\underline{C}OCH_3$), 87.80 (C$_1$), 72.99, 72.74, 72.53, 71.20 (C$_2$, C$_3$, C$_4$, C$_5$), 20.72 (CO$\underline{C}H_3$), 17.42 (C$\underline{C}H_3$). LRMS (ES$^+$): calculated for (M+Na): $C_{12}H_{17}N_3NaO7$: 338.1, found: 338.1. HRMS (ES$^+$): calculated for $C_{12}H_{17}N_3NaO_7$, (M+Na): 338.0964, found: 338.0983.

General procedure of deprotection of the peracetylated azido carbohydrates

Sodium methoxide solution in MeOH (1 N, 1mL/g of starting material) was added to the corresponding acetylated derivative in dry MeOH. The reaction was stirred at room temperature until total starting material consumption (~1h). Amberlite IR$^®$120 ion-exchange resin was then added to neutralize the mixture. The resin was filtered off, washed with MeOH, and the filtrate was evaporated. The solvent was removed to afford the final compound without any purification step.

Synthesis of 1-azido-1-deoxy-β-D-glucopyranose (**69**)

69
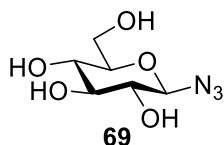
The general procedure is applied to the derivative **3** (1782 mg, 4.773 mmol). A white solid is obtained (971 mg, 99.2%), according to the product already described.[312] ^{1}H NMR (400 MHz, D_2O) δ 4.75 (d, J = 8.8 Hz, 1H, H$_1$), 3.93 (d, J = 12.3 Hz, 1H, H, H$_6$), 3.80 – 3.70 (m, 1H, H$_{6'}$), 3.53 (dd, J = 17.5, 8.8 Hz, 2H, H$_3$, H$_5$), 3.42 (t, J = 9.4 Hz, 1H, H$_4$), 3.27 (t, J = 9.0 Hz, 1H, H$_2$). ^{13}C NMR (101 MHz, D_2O) δ 90.07(C$_1$), 77.85 (C$_5$), 75.70(C$_3$), 72.80(C$_2$), 69.14(C$_4$), 60.52 (C$_6$).

Synthesis of 1-azido-1-deoxy-4-*O*-(α-D-glucopyranosyl)-β-D-glucopyranose (**72**)

72
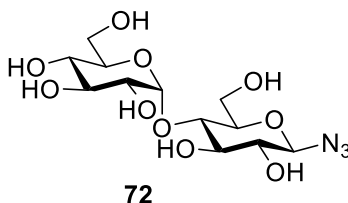
The general procedure was applied to the derivative **6** (577 mg, 0.872 mmol). A white solid was obtained (288 mg, 90%), according to the reported product characterization.[313] ^{1}H NMR (400 MHz, D$_2$O) δ 5.41 (d, J=3.8 Hz, 1H, H$_{1B}$), 4.75 (d, J=8.8 Hz, 1H, H$_{1A}$), 3.94 (d, J=12.2 Hz, 1H, H$_{4A}$), 3.88–3.66 (m, 7H, H$_{5A}$, H$_{5B}$, H$_{3A}$, H$_{6B}$,H$_{6B'}$, H$_{6A}$, H$_{6A'}$), 3.66 (d, J=2.6 Hz, 1H, H$_{3B}$), 3.57 (dd, J=9.9, 3.9 Hz, 1H, H$_{2B}$), 3.41 (t, J=9.4 Hz, 1H, H$_{4B}$), 3.30 (t, J=9.1 Hz, 1H, H$_{2A}$). ^{13}C NMR (101 MHz, D$_2$O) δ 99.56 (C$_{1B}$), 89.86(C$_{1A}$), 76.42, 76.17, 76.13, 72.77 (C$_{3A}$, C$_{3B}$, C$_{4A}$, C$_{5A}$), 72.65(C$_{2B}$), 72.63(C$_5$), 71.60(C$_{4B}$), 69.25(C$_{2A}$), 60.47(C$_{6B}$), 60.44(C$_{6A}$).

Synthesis of 2-acetamido-1-azido-1,2-dideoxy-β-D-glucopyranose (**70**):

70
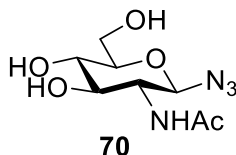
The general procedure was applied to the derivative **8** (594 mg, 1.595 mmol). A white solid was obtained (362 mg, 92%), according to the reported compound.[313] ^{1}H NMR (400 MHz, D$_2$O) δ 4.76 (d, J=9.3 Hz, 1H, H$_1$), 3.94 (dd, J=12.5, 2.2 Hz, 1H, H$_6$), 3.82–3.67 (m, 2H, H$_{6'}$,H$_2$), 3.61-3.43(m,3H, H$_3$, H$_4$, H$_5$), 2.06 (s, 3H, 3xCOC$\underline{H}_3$). ^{13}C NMR (101 MHz, D$_2$O) δ 174.76 (NH$\underline{C}$OCH$_3$), 88.64 (C$_1$), 77.87 (C$_5$), 73.64 (C$_3$), 69.47 (C$_4$), 60.53 (C$_6$), 55.00 (C$_2$), 22.11 (NHCO$\underline{C}$H$_3$).

Synthesis of 1-azido-1,6-dideoxy-β-D-glucopyranose (**71**):

71
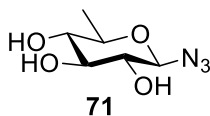
The general procedure is applied to the derivative **68** (837 mg, 2.65 mmol). A white solid was obtained (583 mg, 100%). ^{1}H NMR (500 MHz, D$_2$O) δ 4.76 (d, J=8.8 Hz, 1H, H$_1$), 3.59 (dq, J=9.3, 6.2 Hz, 1H, H$_5$), 3.49 (t, J=9.3 Hz, 1H, H$_3$), 3.29 (t, J=9.1 Hz, 1H, H$_2$), 3.20 (t, J=9.4 Hz, 1H, H$_4$), 1.34 (d, J=6.5 Hz, 3H, 3xC$\underline{H}_3$). ^{13}C NMR (126 MHz, D$_2$O) δ 90.10 (C$_1$), 75.44 (C$_3$), 74.49 (C$_4$), 74.14 (C$_5$), 73.06 (C$_2$), 16.69 (CH$_3$). LRMS (ES$^+$): calculated for (M+Na$^+$): 212.06, found 212.0. HRMS (FAB$^+$) calculated for C$_6$H$_{11}$N$_3$NaO$_4$, (M+Na$^+$): 212.0647, found: 212.0628.

Synthesis of 2-azidoethyl β-D-glucopyranoside (**74**):

74
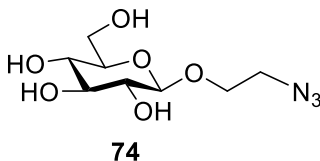
The general procedure was applied to the derivative **14** (550 mg, 1.318mmol). A white solid was achieved (346 mg, 100%) corresponding to the desire product according to literature reference.[314] ^{1}H NMR (600 MHz, CD$_3$OD) δ 4.33 (d, J=7.8 Hz, 1H, H$_1$), 4.03 (dt, J=10.5, 5.1 Hz, 1H, OC$\underline{H}_2$-), 3.88 (dd, J=11.9, 1.6 Hz, 1H, H$_6$), 3.76 (dt, J=10.8, 5.3 Hz, 1H, OC$\underline{H}_2$-), 3.71–3.64 (m, 1H, H$_{6'}$), 3.48 (t, J=5.3 Hz, 2H, 2xC$\underline{H}_2$N$_3$), 3.40–3.35 (m, 1H, H$_4$), 3.33–3.28 (m, 2H, H$_3$, H$_5$),

3.21 (dd, J=9.2, 7.8 Hz, 1H, H$_2$). ^{13}C NMR (151 MHz, CD$_3$OD) δ 104.42 (C$_1$), 77.96 (C$_4$), 77.93 (C$_3$), 75.00 (C$_2$), 71.53 (C$_5$), 69.31 (OC$\underline{H}_2$), 62.73 (C$_6$), 52.02 (C$\underline{H}_2$N$_3$). HRMS (FAB$^+$) calculated for C$_8$H$_{15}$N$_3$NaO$_6$, (M+Na$^+$): 272.0859, found: 272.0848.

General synthesis of the final carbohydrates-NDIs conjugates

To a suspension of azido-glyco-derivatives in 2 mL of ^{t}BuOH, a solution of (+)-sodium L-ascorbate (12 mg, 0.06 mmol), copper(II) sulfate pentahydrate (1.5 mg, 0.006 mmol) and NDI **84** (30 mg, 0.06 mmol) in 2 mL of H$_2$O was added in one portion. The heterogeneous mixture was stirred at room temperature under nitrogen atmosphere for 2 h. The resulting red solution was concentrated under vacuum and a red solid was obtained. The crude product was analysed and purified by HPLC chromatography, according to analytical method B: (CH$_3$CN:H$_2$O 0.1%TFA) 1.4 mL/min, Gradient mobile phase B/mobile phase A (95% eluent A to 60% A, 12 min; isocratic 2 min).

β-Glc-NDI (61):

61
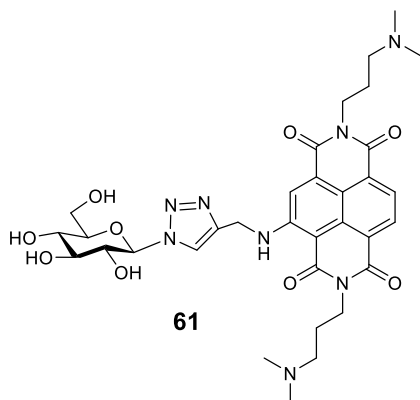
Red Solid, yield = 97%, purity by HPLC analysis, 99.8% (Rt = 6.69 min). m.p. > 275°C.

^{1}H-NMR (400 MHz, CD$_3$OD) δ 8.42 (d, J = 7.8 Hz, 1H, H$_{AR}$), 8.37 (s, 1H, H$_{AR}$), 8.22 (s, 1H, H$_{AR}$), 8.16 (d, J = 7.8 Hz, 1H, H$_{AR}$), 5.68 (d, J = 9.2 Hz, 1H, H$_1$), 4.94 (s, 2H, 2xNHC$\underline{H}_2$), 4.21 (t, J = 6.7 Hz, 4H, 4xC$\underline{H}_2$N), 3.97–3.87 (m, 2H, H$_2$, H$_6$), 3.76–3.71 (m, 1H, H$_6$), 3.62–3.59 (m, 2H, H$_5$, H$_3$), 3.53–3.47 (m, 1H, H$_4$) 3.30–3.24 (m, 4H, 2xNC$\underline{H}_2$), 2.95 (s, 6H, N(C$\underline{H}_3$)$_2$), 2.94 (s, 6H, 6xN(C$\underline{H}_3$)$_2$), 2.24–3.17 (m, 4H, 4xC$\underline{H}_2$).

^{13}C-NMR (101 MHz, CD$_3$OD) δ 167.10(CO), 164.66(CO), 164.39(CO), 164.20(CO), 153.09(C$_{AR}$), 149.71(C$_{AR}$), 132.08(C$_{AR}$), 130.29(C$_{AR}$), 128.98(CH$_{AR}$), 127.30(C$_{AR}$), 125.53(C$_{AR}$), 124.36(CH$_{AR}$), 120.92(CH$_{AR}$), 120.57(CH$_{AR}$), 116.94(C$_{AR}$), 101.02(C$_{AR}$), 89.70(C$_1$), 81.24(C$_3$), 78.50(C$_5$), 74.14(C$_2$), 70.98(C$_4$), 62.38(C$_6$), 56.83(NC$\underline{H}_2$), 56.76(NC$\underline{H}_2$), 43.61(NC$\underline{H}_3$), 39.56(NHC$\underline{H}_2$), 38.12(C$\underline{H}_2$N), 24.63(C$\underline{H}_2$).

HRMS(ES$^+$): Calcd. for C$_{33}$H$_{42}$N$_8$NaO$_9$(M+Na): 717.2972; found: 717.2963.
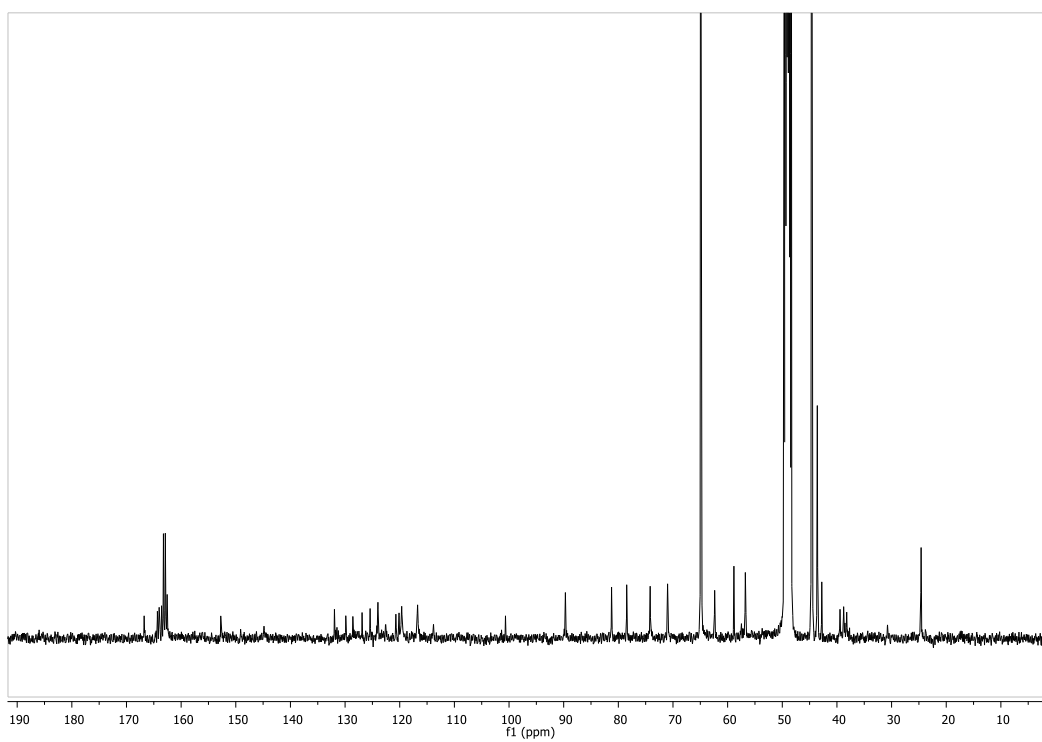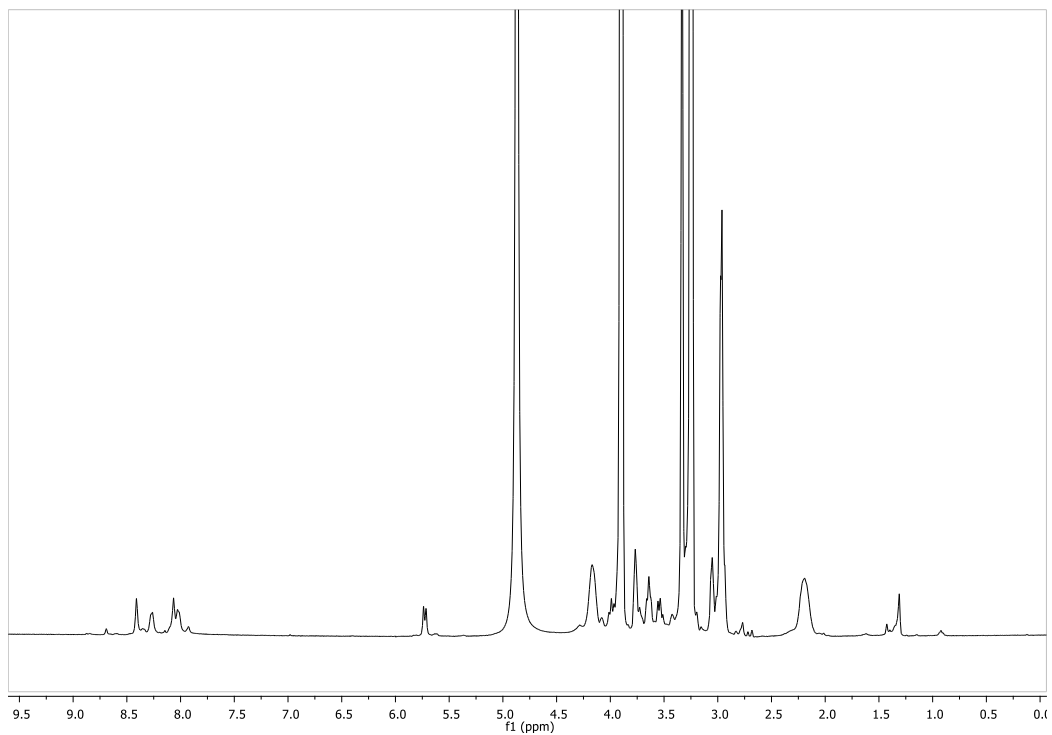
β-GlcNAc-NDI (62):

62
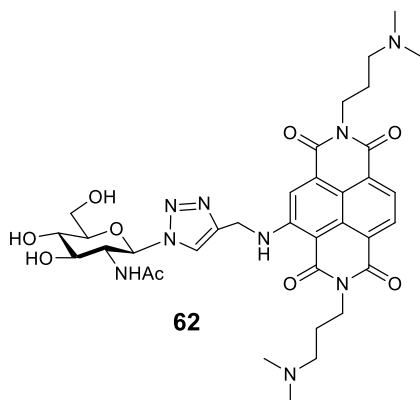
Red Solid, yield = 95%, purity by HPLC analysis, 100% (Rt = 4.83 min). m.p. > 275°C.

^{1}H-NMR (300 MHz, CD$_3$OD) δ 8.89 (s, 1H, NH), 8.70 (d, J = 7.8 Hz, 1H, H$_{AR}$), 8.45 (s, 1H, H$_{AR}$), 8.43 (d, J = 7.8 Hz, 1H, H$_{AR}$), 8.36 (s, 1H, H$_{AR}$), 5.88 (d, J = 9.7 Hz, 1H, H$_1$), 5.08 (s, 2H, 2xNHC$\underline{H}_2$), 4.41–4.34 (m, 5H, 4xC$\underline{H}_2$N, H$_2$), 4.0 (d, J = 11.9 Hz, 1H, H$_5$), 3.89–3.65 (m, 3H, H$_3$, H$_4$, H$_6$), 3.42 (bs, 4H, 4xNC$\underline{H}_2$), 3.05 (s, 12H, 12xN($\underline{CH_3}$)$_2$), 2.34–2.29 (m, 4H, 4xCH$_2$), 1.78 (s, 3H, NHCOC$\underline{H}_3$).

^{13}C NMR (75 MHz, CD$_3$OD) δ 173.6(CO), 167.5(CO), 165.1 (CO), 164. 8(CO), 164.5(CO), 153.5(C$_{AR}$), 145.5(C$_{AR}$), 132.4(C$_{AR}$), 130.7(C$_{AR}$), 129.4(CH$_{AR}$), 127.7(C$_{AR}$), 125.8(C$_{AR}$), 124.7(C$_{AR}$), 123.6(CH$_{AR}$), 121.2(CH$_{AR}$), 121.0(CH$_{AR}$), 101.5(C$_{AR}$), 88.7(C$_1$), 81.6, 76.1, 71.6, 62.6, 57.2 57.1, 56.8, 43.9, 39.6, 38.9, 38.4, 24.9, 22.8(CO$\underline{C}$H$_3$).

HRMS(ES$^+$): Calculated for C$_{35}$H$_{45}$N$_9$NaO$_9$(M+Na): 758.3238; found: 758.3218.
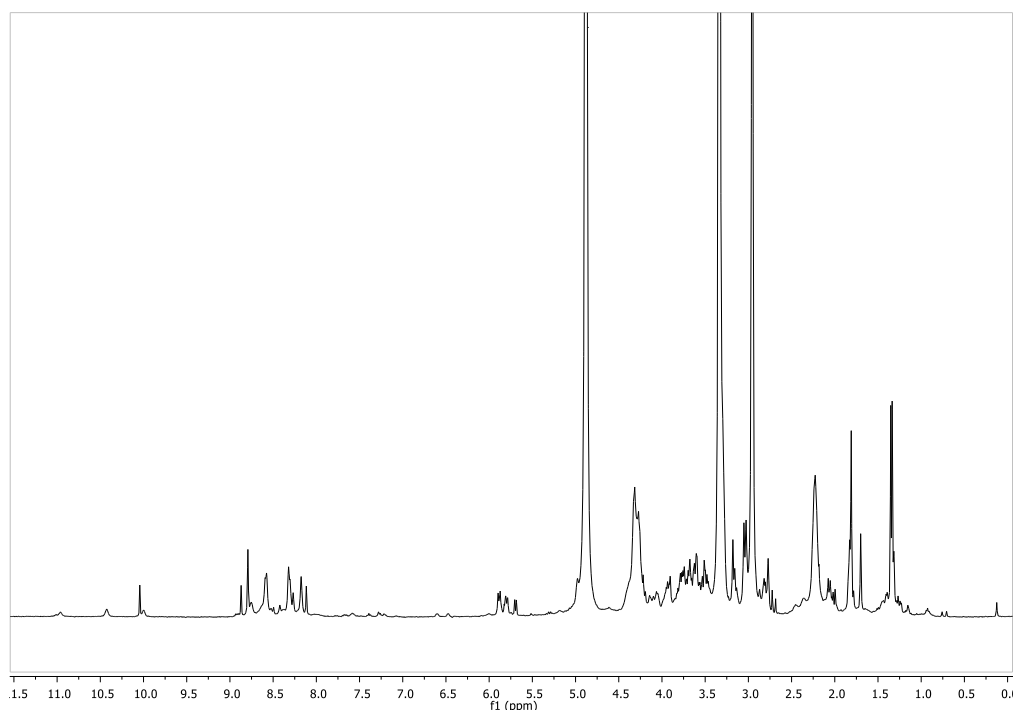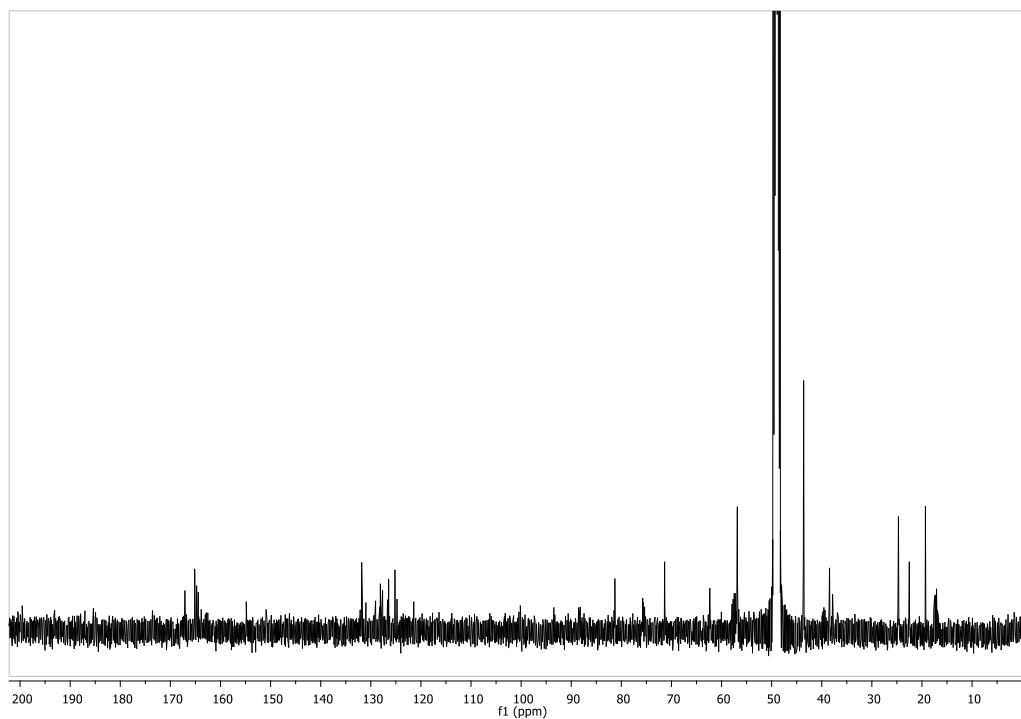
β-6-dGlc-NDI (63):

63
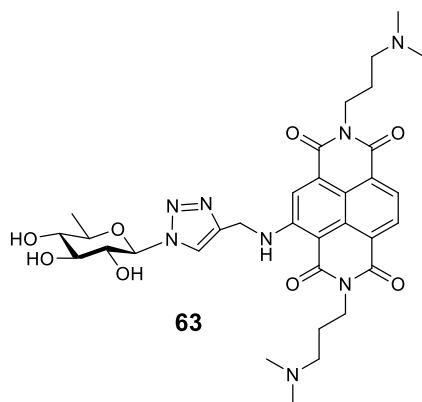
Red Solid, yield = 98%, purity by HPLC analysis, 97.6% (Rt = 7.14 min). m.p. > 275°C.

^{1}H-NMR (400 MHz, CD$_3$OD) δ 10.29–10.22 (m, 1H, N$\underline{H}$), 8.36–8.29 (m, 2H, 2xH$_{AR,}$), 8.12–8.03 (m, 2H, 2xH$_{AR}$), 5.66 (d, J = 9.0 Hz, 1H, H$_1$), 4.89 (s, 2H, 2xNHC$\underline{H_2}$) 4.17 (s, 4H, 4xC$\underline{H_2}$N), 3.98 (t, J = 9.0 Hz, 1H, H$_2$), 3.65 (dd, J = 9.5, 5.8 Hz, 1H, H$_5$), 3.56 (t, J = 9.0 Hz, 1H, H$_3$), 3.30–3.18 (m, 5H, 4xNC$\underline{H_2}$, H$_4$), 2.97 (s, 6H, 6xN(C$\underline{H_3}$)$_2$), 2.95 (s, 6H, N(C$\underline{H_3}$)$_2$), 2.21–2.16 (m, 4H, 4xC$\underline{H_2}$), 1.32 (d, J = 5.8 Hz, 3H, 3xH$_6$).

^{13}C-NMR (101 MHz, CD$_3$OD) δ 166.85(CO), 164.43(CO), 164.12(CO), 163.96(CO), 152.83(C$_{AR}$), 144.87(C$_{AR}$), 131.99(C$_{AR}$), 129.98(CH$_{AR}$), 128.67(C$_{AR}$), 127.00(C$_{AR}$), 125.44(C$_{AR}$), 124.07(CH$_{AR}$), 123.84(C$_{AR}$), 120.77(CH$_{AR}$), 120.22(CH$_{AR}$), 100.73(C$_{AR}$), 89.63(C$_1$), 78.28(C$_3$),

76.69(C_5), 76.51(C_4), 74.34(C_2), 56.80($\underline{C}H_2$-NCH$_3$), 56.73($\underline{C}H_2$-NCH$_3$), 43.63(N$\underline{C}H_3$), 40.04(NH$\underline{C}H_2$), 38.79($\underline{C}H_2$N), 38.22($\underline{C}H_2$N), 24.63(CH_2), 18.15(C_6).

HRMS(ES⁺): Calculated for $C_{33}H_{42}N_8NaO_8$(M+Na): 701.3023; found 701.3007.
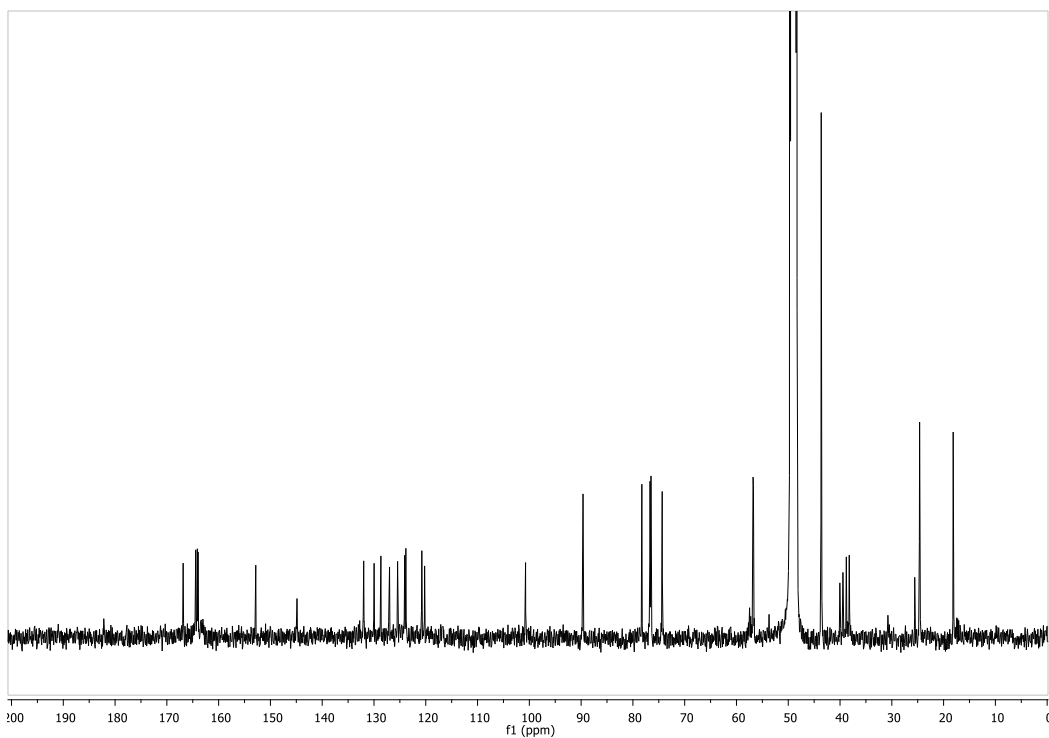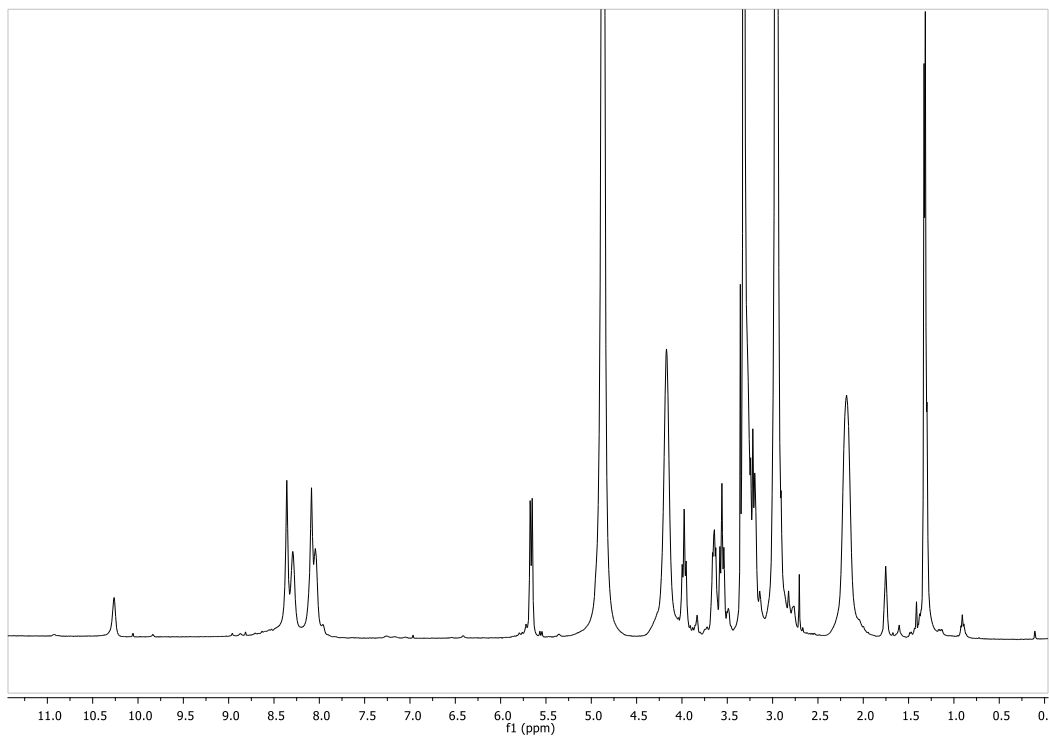
β-Malt-NDI (64):

64
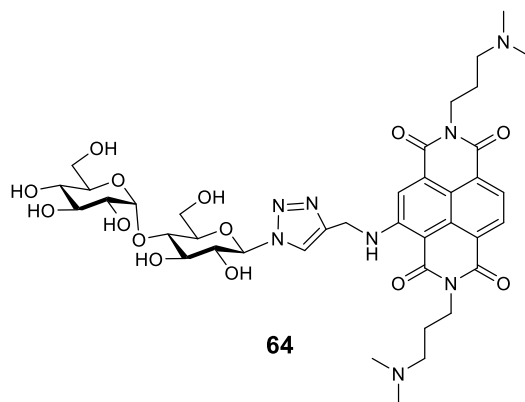
Red Solid, yield = 87%, purity by HPLC analysis, 99.3% (Rt = 6.61 min). m.p. > 275°C.

^{1}H-NMR (400 MHz, CD$_3$OD) δ 8.40 (s, 1H, H$_{AR}$), 8.27 (d, J = 7.8 Hz, 1H, H$_{AR}$), 8.08 (s, 1H, H$_{AR}$), 8.02 (d, J = 7.8 Hz, 1H, H$_{AR}$), 5.74 (d, J = 9.1 Hz, 1H, H$_{1A}$), 5.24 (d, J = 3.8 Hz, 1H, H$_{1B}$), 4.87(s, 2H, 2xNHC$\underline{H}_2$) 4.20–4.10 (m, 4H, 4xC$\underline{H}_2$N), 4.03 (t, J = 9.2 Hz, 1H, H$_{2A}$), 3.94–3.80 (m, 4H, H$_{61B}$, H$_{6'1B}$, H$_{61A}$,H$_{3A}$), 3.80–3.60 (m, 5H, H$_{3B}$, H$_{5B}$, H$_{4A}$, H$_{5A}$, H$_{6'1A}$), 3.49 (dd, J = 9.7, 3.7 Hz, 1H, H$_{2B}$), 3.29–3.23 (m, 5H, H$_{4B}$, 4xNC$\underline{H}_2$), 2.95 (s, 6H, 6xNC$\underline{H}_3$), 2.94 (s, 6H, 6xNC$\underline{H}_3$), 2.25–2.08 (m, 4H, 4xC$\underline{H}_2$).

^{13}C-NMR (101 MHz, CD$_3$OD) δ 166.84(CO), 164.42(CO), 164.11(CO), 163.98(CO), 152.84(C$_{AR}$), 147.78(C$_{AR}$), 132.03(C$_{AR}$), 129.99(C$_{AR}$), 128.67(CH$_{AR}$),127.00(C$_{AR}$), 125.48(C$_{AR}$), 124.08(CH$_{AR}$), 124.02(C$_{AR}$), 120.83(CH$_{AR}$), 120.23(CH$_{AR}$), 102.98(C$_{AR}$), 100.73(C$_{1B}$), 89.52(C$_{1A}$), 80.49(C$_{3B}$), 79.69(C$_{4A}$), 78.24(C$_{3A}$), 75.06(C$_{5B}$), 74.95(C$_{5A}$), 74.12(C$_{2B}$), 73.79(C$_{2A}$), 71.55(C$_{4B}$), 62.79(C$_{6A}$),61.85(C$_{6B}$) 56.80(NC$\underline{H}_2$),43.62(NC$\underline{H}_3$), 39.40(NHC$\underline{H}_2$), 38.75(C$\underline{H}_2$N), 38.18(C$\underline{H}_2$N), 24.59(C$\underline{H}_2$).

HRMS(ES$^+$): Calculated for C$_{39}$H$_{52}$N$_8$NaO$_{14}$(M+Na):879,3501; found: 879.3478.
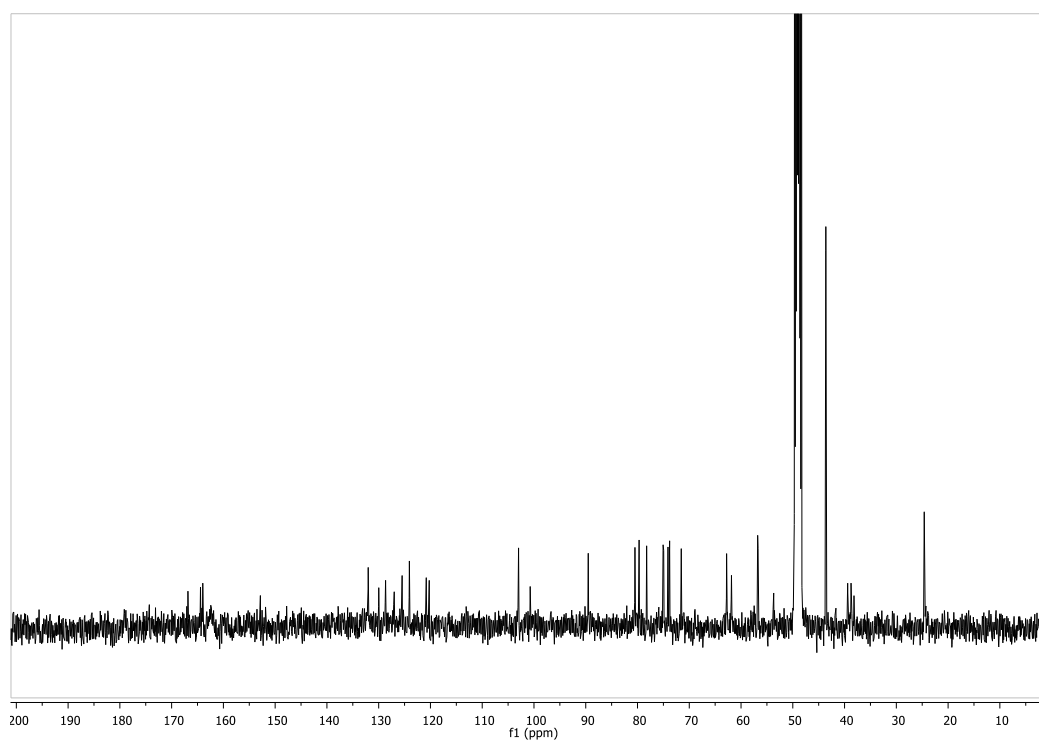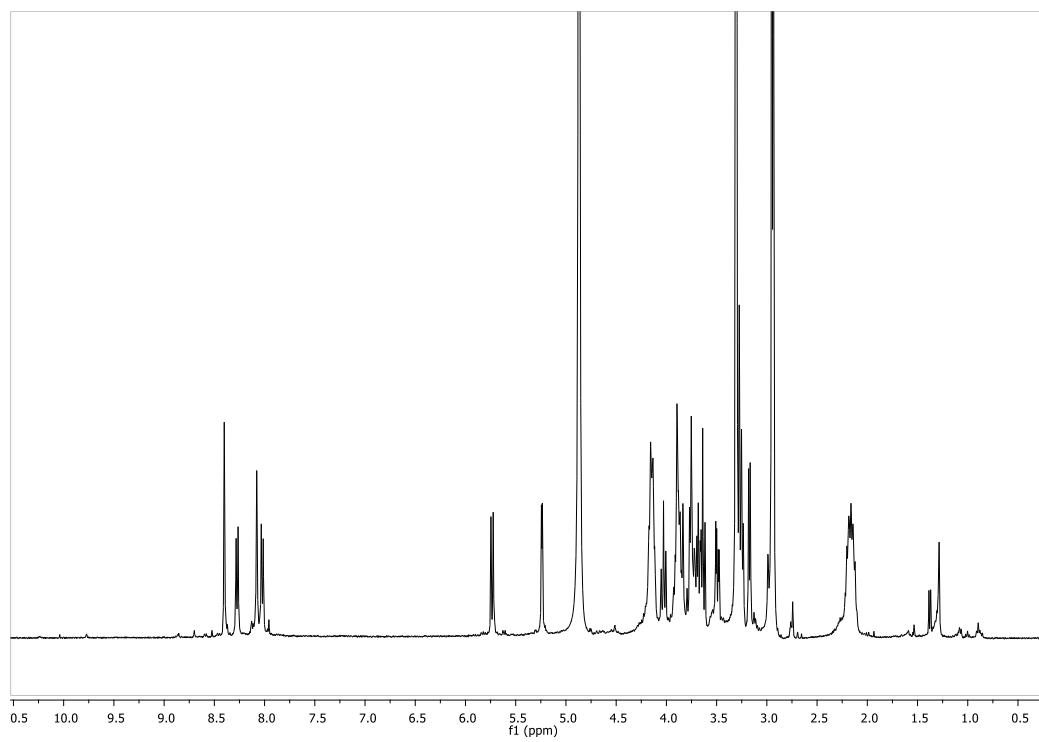
α-ManC2-NDI (65):

65
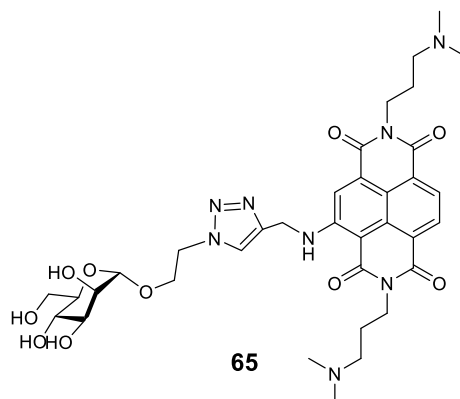
Red Solid, yield = 93%, purity by HPLC analysis, 96.9% (Rt = 6.73 min). m.p. > 275°C.

^{1}H-NMR (400 MHz, CD$_3$OD) δ 10.38(s, 1H, NH), 8.48 (d, J = 7.6 Hz, 1H, H$_{AR}$), 8.22–8.15 (m, 3H, 3xH$_{AR}$), 4.95 (s, 2H, 2xNHC$\underline{H}_2$), 4.83(s, 1H, H$_1$), 4.70 (s, 2H), 4.28–4.23 (m, 4H, 2xCH$_2$N), 4.15–4.12 (m, 1H, H$_2$), 3.88–3.86 (m, 1H, H$_3$), 3.70–3.50 (m, 5H, H$_5$, 2xH$_6$, 2xC$\underline{H}_2$), 3.32–3.22 (m, 4H, 4xNC$\underline{H}_2$), 3.19–3.18 (m, 1H, H$_4$), 2.95 (s, 12H, 12xN(C$\underline{H}_3$)$_2$), 2.23–2.20 (m, 4H, 4xC$\underline{H}_2$).

^{13}C-NMR (101 MHz, CD$_3$OD) δ 167.21(CO), 164.78(CO), 164.48(CO), 164.30(CO), 153.15(C$_{AR}$), 132.08(C$_{AR}$), 130.7(C$_{AR}$), 130.43(C$_{AR}$), 129.12(CH$_{AR}$), 127.44(C$_{AR}$), 125.55(C$_{AR}$), 124.48(CH$_{AR}$), 120.78(CH$_{AR}$), 120.75(CH$_{AR}$), 120.70(C$_{AR}$), 101.61(C$_1$), 101.20(C$_{AR}$), 74.82(C$_4$), 72.48(C$_6$), 71.85(CH$_2$), 68.30(C$_2$), 66.84(C$_3$), 62.59(C$_5$), 56.81(NC$\underline{H}_2$), 51.44(CH$_2$), 43.63(NC$\underline{H}_3$), 39.41(NHC$\underline{H}_2$), 38.67(C$\underline{H}_2$N), 38.15(C$\underline{H}_2$N), 24.70(C$\underline{H}_2$).

HRMS(ES$^+$): Calculated for C$_{35}$H$_{46}$N$_8$NaO$_{10}$ (M+Na): 761.3235; found: 761.3259.
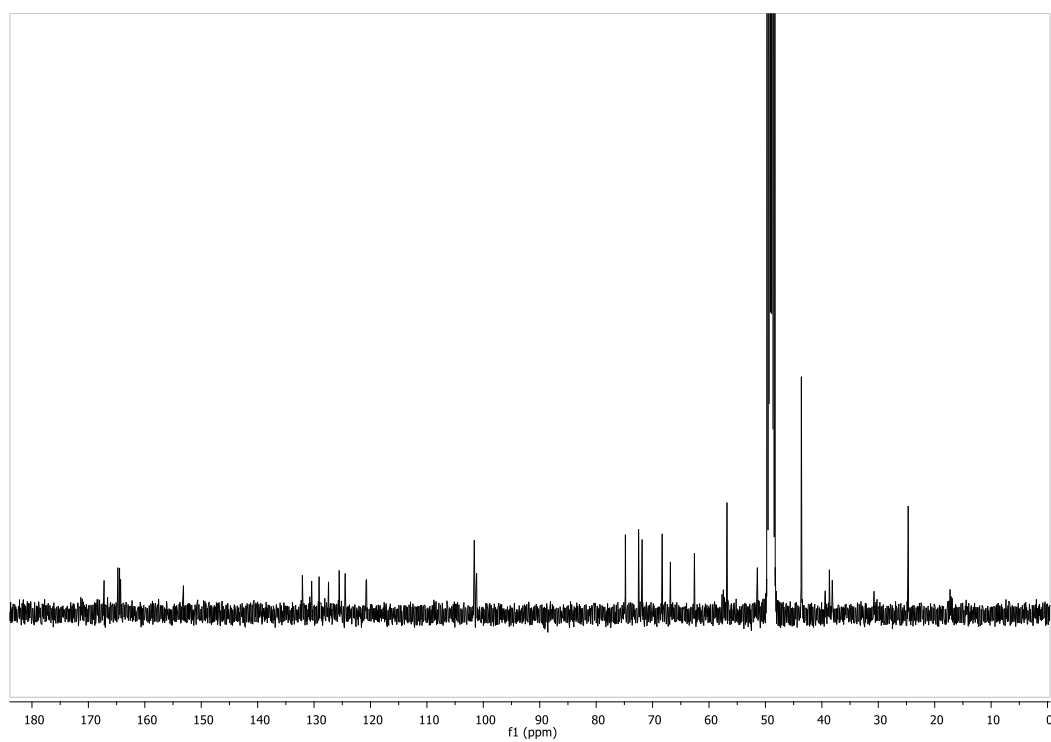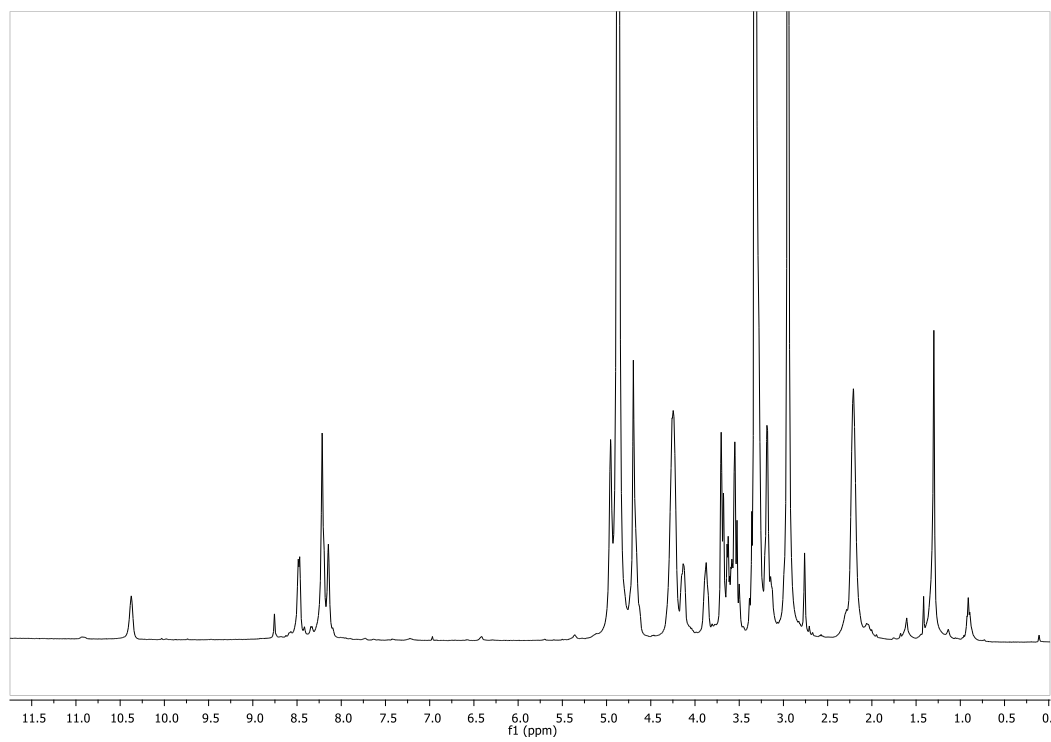
β-GlcC2-NDI (66):

66
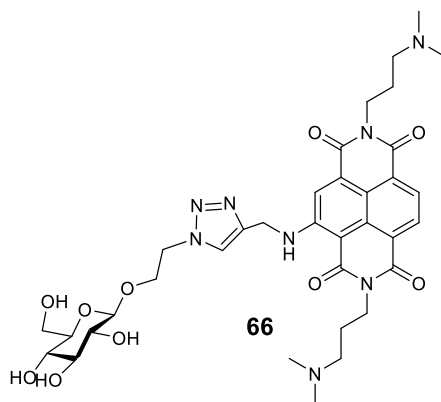
Red Solid, yield = 91%, purity by HPLC analysis, 99.2% (Rt = 7.4 min). m.p. > 275°C.

¹H-NMR (300 MHz, D_2O) δ 8.22 (d, J = 7.6 Hz, 1H, H_{AR}), 8.13 (s, 1H, H_{AR}), 7.97 (d, J = 7.6 Hz, 1H, H_{AR}), 7.87 (s, 1H, H_{AR}), 4.81 (s, 2H, 2xNHC$\underline{H}_2$), 4.64 (s, 2H), 4.28 (d, J = 7.7 Hz, 1H, H_1), 4.23-4.19 (m, 1H), 4.06 (bs, 4H, 4xCH$_2$N), 3.68 (d, J = 12.3 Hz, 1H), 3.47-3.40 (m, 1H), 3.29-3.19 (m, 6H, 2xNC$\underline{H}_2$, OC$\underline{H}_2$), 3.09–2.93 (m, 3H), 2.86 (s, 6H, 6xNC$\underline{H}_3$), 2.85 (s, 6H, 6xNC$\underline{H}_3$), 2.07 (bs, 4H, 4xC$\underline{H}_2$).

¹³C-NMR (75 MHz, D_2O) δ 165.3(CO), 163.8(CO), 163.6(CO), 163.2(CO), 151.4(C_{AR}), 130.9(C_{AR}), 128.2(C_{AR}), 126.7(CH$_{AR}$), 125.2(C_{AR}), 124.9(C_{AR}), 124.8(C_{AR}), 124.5(CH$_{AR}$), 122.3(CH$_{AR}$), 119.7(CH$_{AR}$), 118.5(C_{AR}), 102.2(C_1), 99.3(C_{AR}), 75.7, 75.3, 72.9, 69.4, 67.8, 60.5, 55.2(NC$\underline{H}_2$), 55.1(NC$\underline{H}_2$), 50.4, 42.7(NC$\underline{H}_3$), 37.9(NHC$\underline{H}_2$), 37.6 ($\underline{C}H_2$N), 37.0($\underline{C}H_2$N), 22.8($\underline{C}H_2$).

HRMS(ES⁺): Calculated for $C_{35}H_{47}N_8O_{10}$ (M+): 739.3415; found: 739.3465.
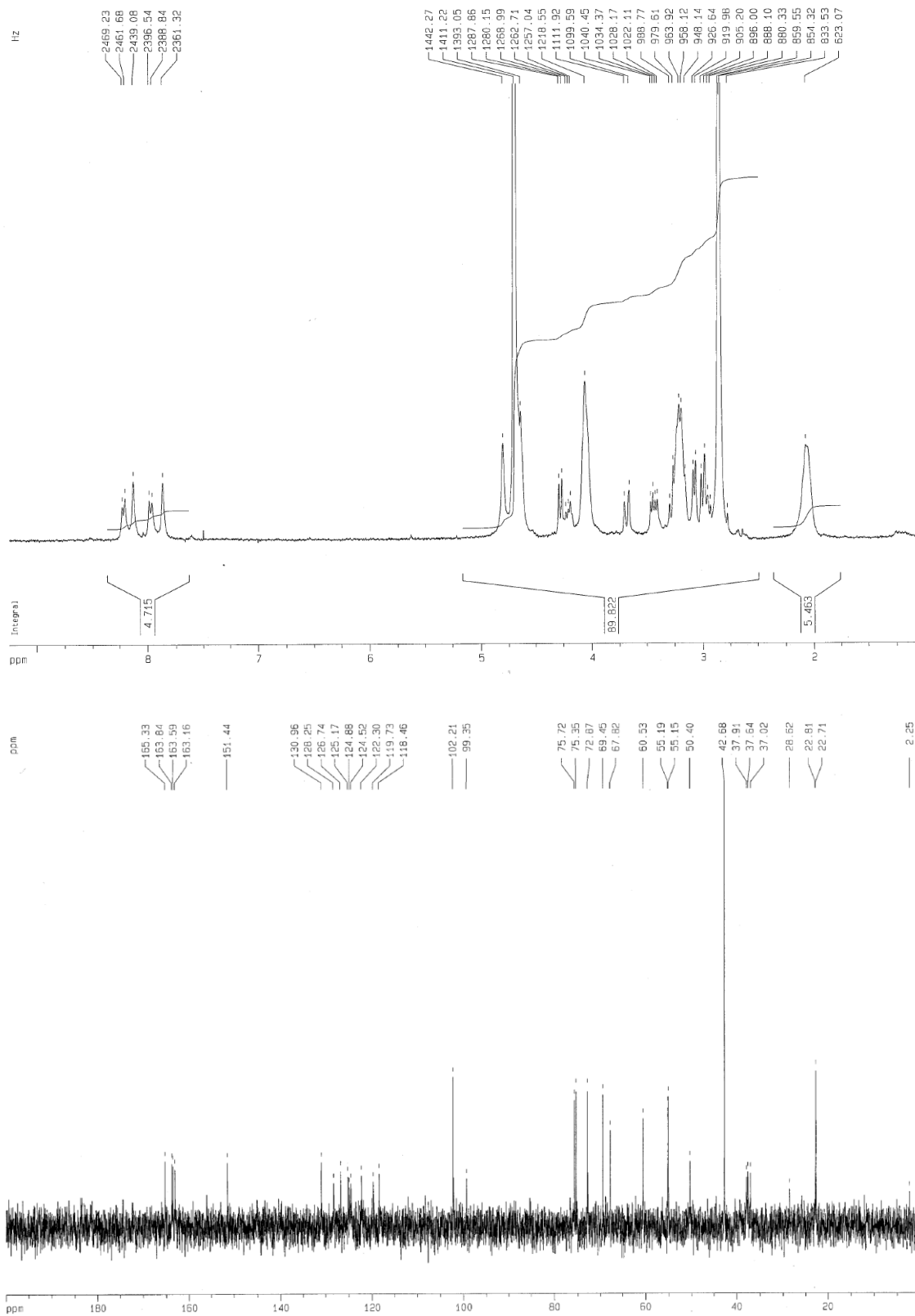
Prop-NDI (67):

67
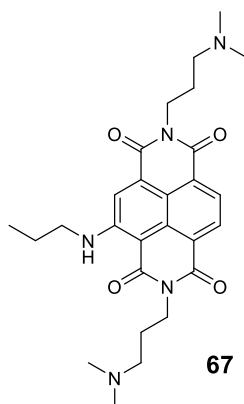
      The dibromo-substituted-NDI **VI** in (0.5 mmols) was dissolved into 40 mL of acetonitrile in a round bottom flask together with 1.5 mmols of n-propylamine, the mixture was stirred at 70°C for 2 h under argon. The resulting red solution was concentrated under vacuum yielding a red solid. The crude product was analyzed and purified by HPLC chromatography according to analytical method B: (CH₃CN:H₂O 0.1%TFA) 1.4 mL/min, gradient mobile phase B/mobile phase A (95% eluent A to 40% A, 8 min; isocratic 4 min).

Red Solid, yield = 68%, purity by HPLC analysis, 100% (Rt = 5.65 min) m.p.>200°C);

^{1}H-NMR (300 MHz, D$_2$O) δ 8.09 (d, J = 7.9 Hz, 1H, HAR), 7.85 (d, J = 7.9 Hz, 1H, HAR), 7.65 (s, 1H, HAR), 4.07-3.98 (m, 4H), 3.37 (t, J = 2.2 Hz, 2H), 3.25-3.05 (m, 4H), 2.86 (s, 6H), 2.84 (s, 6H), 2.10-2.05 (m, 4H), 1.78-1.68 (m, 4H), 1.04 (t, J = 7.3 Hz, 3H).

^{13}C NMR (75 MHz, D$_2$O) δ 165.1(CO), 163.6(CO), 163.3(CO), 163.0(CO), 151.8, 130.6, 128.3, 126.5, 124.9, 123.8, 121.6, 120.0, 118.1, 97.8, 55.1, 44.6, 44.4, 42.6, 37.6, 37.0, 22.6, 21.9, 10.6.

HRMS(ES+): Calculated for C$_{27}$H$_{35}$N$_5$O$_4$Na (M+Na): 516.2587; found: 516.2577.
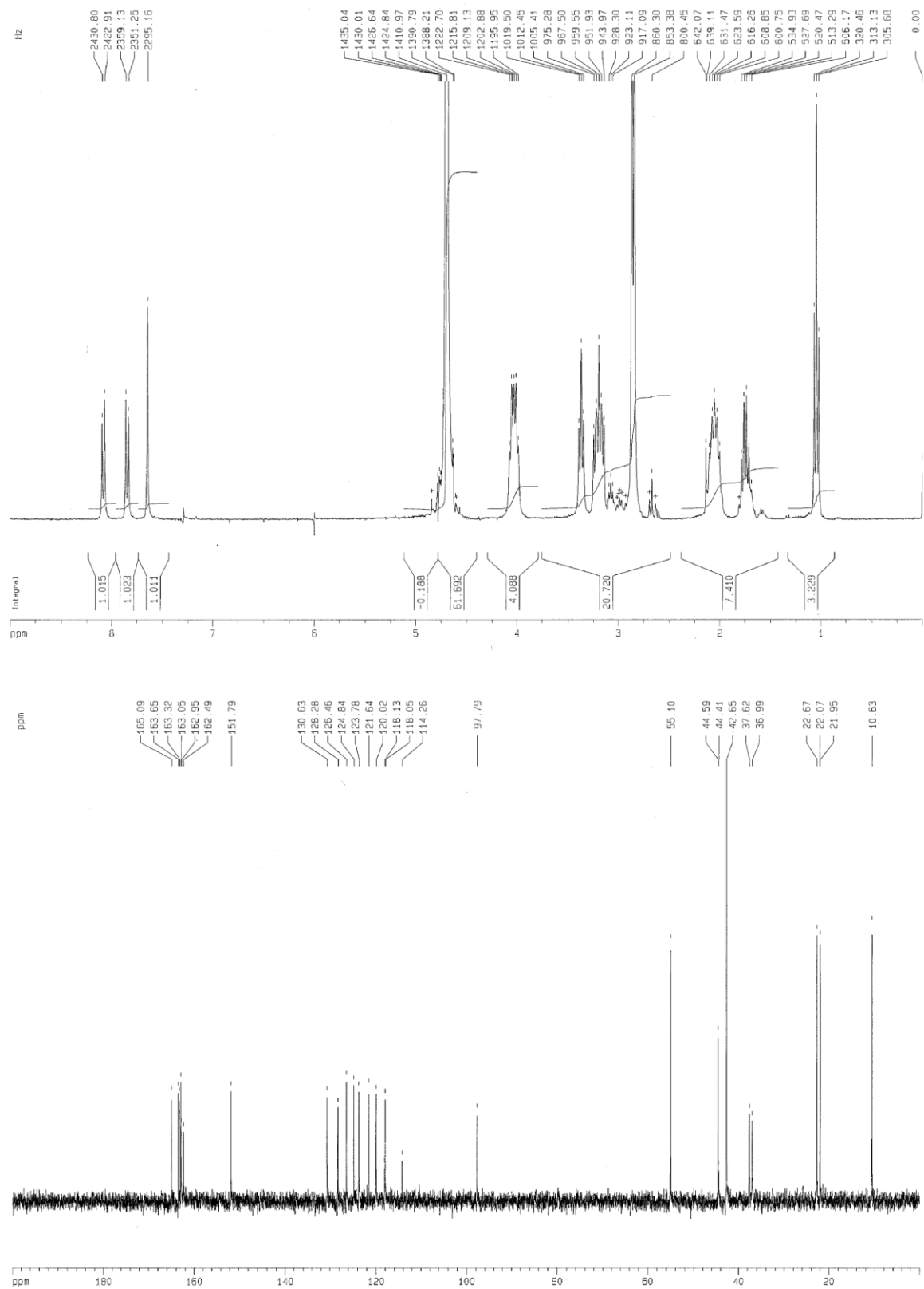
5.2. Biological assays.

Biological assays were performed with the collaboration of PhD. student Efres Belmonte.

5.2.1. CELL CULTURE.

MRC-5 cells were maintained at 37ºC and 5% CO_2 in 100% of humidity in low glucose DMEM (1 g/L) supplemented with 10% iFBS, 100 U/mL penicillin and 100 mg/mL streptomycin. HT-29 and MCF-7 cells were maintained at 37ºC and 5% CO_2 in high glucose DMEM (4.5 g/L) supplemented with 10% iFBS, 100 U/mL penicillin and 100 mg/ml streptomycin.

5.2.2. CYTOTOXICITY ASSAYS.

MCF-7, HT-29 and MRC-5 cells were harvested by trypsinization (0.25%) and seeded in 96 well plates (5000 cells in 100 µL /well) in the presence of increasing concentrations of NDI-compounds. Cellular toxicity was determined using the colorimetric MTT-based assay after incubation at 37 ºC for 48 h (HT-29 and MCF-7) or 72 h (MRC-5). The results are expressed as the concentration of compound that reduce cell growth by 50% versus untreated control cells (EC 50).

5.2.3. FLOW CYTOMETRY.

Determination of cell fluorescence by size and cell complexity (Forward Scatter and Side Scatter) was achieved by flow cytometry. 450,000 MCF-7, HT-29 and MRC-5 cells were harvested by trypsinization and incubated in their medium (1.5 mL) for different times at 4 ºC and 37 ºC. After supernatant removal and washing, the cells were analyzed on an FACScalibur flow cytometer (BD Biosciences, San Jose, CA) simultaneously using a 15mw-Argon and a 635nm diode lasers. Data obtained from detection of orange fluorescence at FL2 channel was organized with FlowJo (Tree Star, Ashland, OR), analyzed with R-Studio V.0.99.891 and represented with ggplot2 package V.1.0.1.

Flow cytometry diagrams of compound **β-glc-NDI (61)** (5 µM) incubated with HT-29, MCF-7 and MRC-5 cell lines for 5 min, 1 h, 2 h and 4 h measured at 4 and 37 °C by flow cytometry:

$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
HT29	Control	1	1	37
HT29	506	5	1	4
HT29	506	5	2	4
HT29	506	5	0.083	37
HT29	506	5	1	37
HT29	506	5	2	37
HT29	506	5	4	37

	$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
	MCF7	Control	n/a	1	37
	MCF7	506	5	1	4
	MCF7	506	5	2	4
	MCF7	506	5	0.083	37
	MCF7	506	5	1	37
	MCF7	506	5	2	37
	MCF7	506	5	4	37

	$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
	MRC5	Control		1	n/a
	MRC5	506	5	1	4
	MRC5	506	5	2	4
	MRC5	506	5	0.083	37
	MRC5	506	5	1	37
	MRC5	506	5	2	37
	MRC5	506	5	4	37

Flow cytometry diagrams of compound **α-manC2-NDI (65)** (5 µM) incubated with HT-29, MCF-7 and MRC-5 cell lines for 5 min, 1 h, 2 h and 4 h measured at 4 and 37 °C by flow cytometry:

	$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
	MCF7	Control	n/a	1	37
	HT29	504b	5	1	4
	HT29	504b	5	2	4
	HT29	504b	5	0.083	37
	HT29	504b	5	1	37
	HT29	504b	5	2	37
	HT29	504b	5	4	37

$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
MCF7	Control	n/a	1	37
MCF7	504b	5	1	4
MCF7	504b	5	2	4
MCF7	504b	5	0.083	37
MCF7	504b	5	1	37
MCF7	504b	5	2	37
MCF7	504b	5	4	37

$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
MRC5	Control		1	n/a
MRC5	504b	5	1	4
MRC5	504b	5	2	4
MRC5	504b	5	0.083	37
MRC5	504b	5	1	37
MRC5	504b	5	2	37
MRC5	504b	5	4	37

Flow cytometry diagrams of compound **prop-NDI (67)** (5 µM), incubated with with HT-29, MCF-7 and MRC-5 cell lines for 5 min, 1 h, 2 h and 4 h measured at 4 and 37 °C by flow cytometry:

$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
HT29	Control	1	1	37
HT29	Prop	5	1	4
HT29	Prop	5	2	4
HT29	Prop	5	0.083	37
HT29	Prop	5	1	37
HT29	Prop	5	2	37
HT29	Prop	5	4	37

$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
MCF7	Control	n/a	1	37
MCF7	Prop	5	1	4
MCF7	Prop	5	2	4
MCF7	Prop	5	0.083	37
MCF7	Prop	5	1	37
MCF7	Prop	5	2	37
MCF7	Prop	5	4	37

$.Cell Line	$.NDI	$.CC (uM)	$.Time (h)	$.Temp
MRC5	Control		1	n/a
MRC5	Prop	5	1	4
MRC5	Prop	5	2	4
MRC5	Prop	5	0.083	37
MRC5	Prop	5	1	37
MRC5	Prop	5	2	37
MRC5	Prop	5	4	37
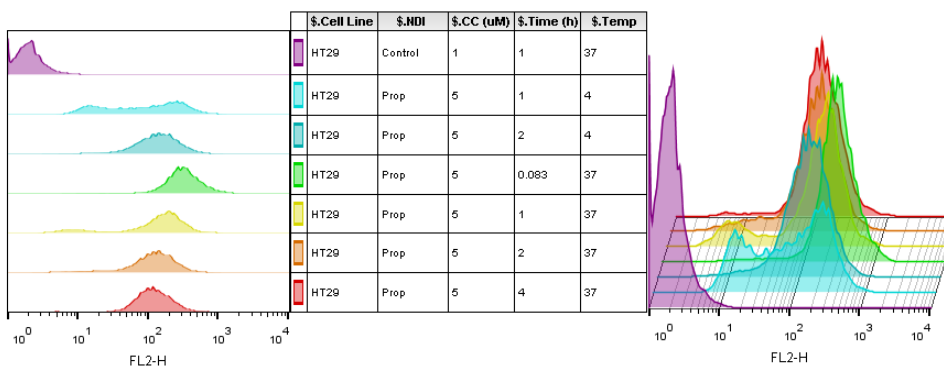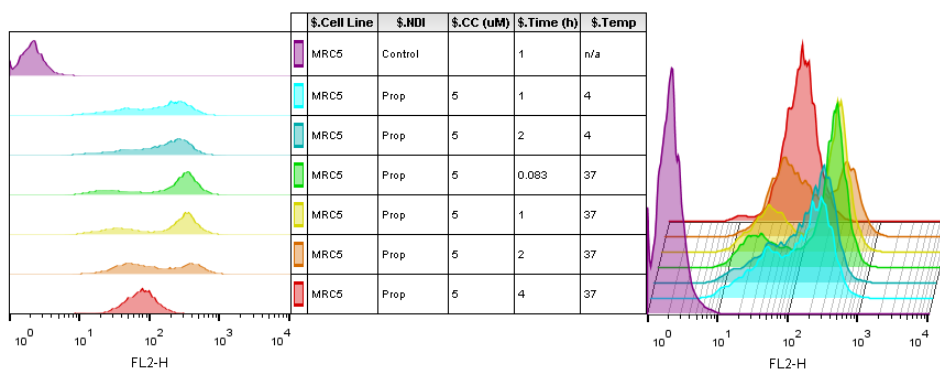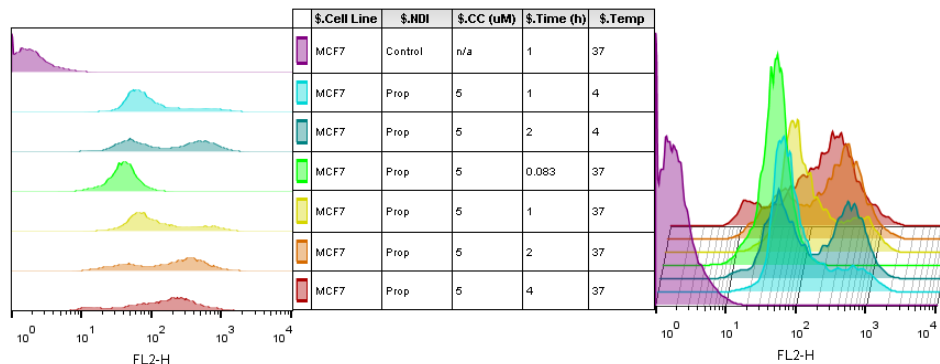
5.2.4. NDI CELL UPTAKE QUANTIFICATION

225,000 MCF-7, HT-29 and MRC-5 cells were harvested by trypsinization and incubated with 5 µM of a NDI-Compound in 0.75 mL of medium for 2 h at 37 ºC. After centrifugation, supernatant removal and washing three times with PBS, the cell pellet was re-suspended in 0.75 mL of SDS 0.4 % and homogenized. Fluorescence was detected with a TECAN infinite F200 fluorescence-intensity multiplate reader (excitation wavelength: 485 nm, emission wavelength: 535 nm). Fluorescence values were normalized via protein quantification assay using Pierce BCA test[315](ThermoFisher Scientific) and concentration values were extrapolated from a fluorescence – NDI-concentration calibration curve.

5.2.5. NDI CELL UPTAKE QUANTIFICATION IN THE PRESENCE OF GLUT INHIBITORS.

225,000 HT-29 and MRC-5 cells were harvested by trypsinization and preincubated in 0.75 mL of medium with 100 µM of a Glut inhibitor (Fasentin, Genistein, Ritonavir, ZWB117 and Quercetin, Sigma Aldrich©) for 1 h at 37 ºC. 5 µM of a NDI-compound was then added and incubated for another 2 h at 37 ºC. Sample treatment and NDI-uptake quantification was achieved following the previously mentioned NDI-uptake quantification protocol. Inhibition values were calculated using NDI-uptake in presence of the Glut inhibitor (A) and NDI-uptake values (B) as inhibitor-free controls using the formula: Inhibition (%) = (B-A)/B x 100.

Graph with cell uptake inhibition percentage for **β-glc-NDI (61)**, **α-manC2-NDI (65)**, **β-glcC2-NDI (66)** and aglycone **prop-NDI (67)** in the presence of GLUT inhibitors shown on figures below.
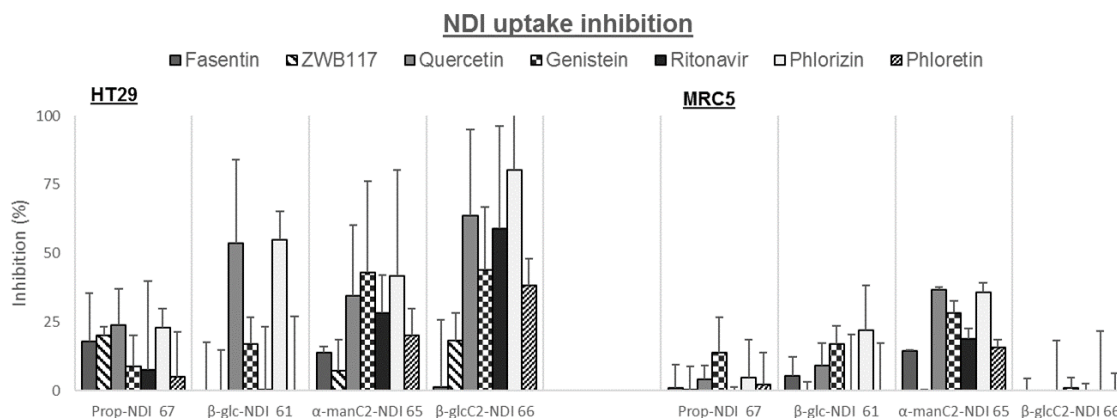
5.2.6. CONFOCAL MICROSCOPY.

MRC-5 and HT-29 cells were incubated with 5 µM of NDI compounds in 0.5 mL of each respective medium for 2 h at 37 ºC and 100 % of humidity. Cells were washed seven times in 500 ml of cold phosphate-buffered saline (PBS), fixed and processed by microscope observation. Images were acquired using a Leica SP5 confocal microscope, while the images were deconvoluted using Huygens Professional image processing software from Scientific Volume Imaging (http://www.svi.nl). The merge of the images were made with Fiji software (https://fiji.sc/).
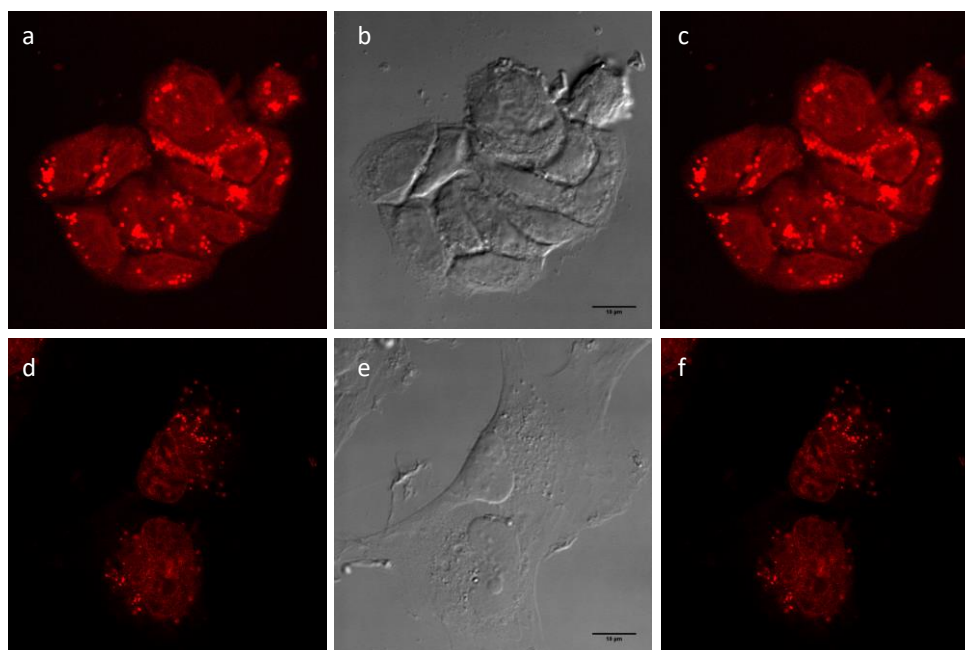
Figure 3.25. Confocal images of HT-29 and MRC-5 cells after co-incubation with **β-glc-NDI (61)** (5 µM) for 2 h. HT-29 cells: a) fluorescence image of compound **1**; b) Transmitted light image; c) merged image; MRC-5 cells: d) fluorescence image of compound **β-glc-NDI (61)**; e) Transmitted light image; f) merged image.
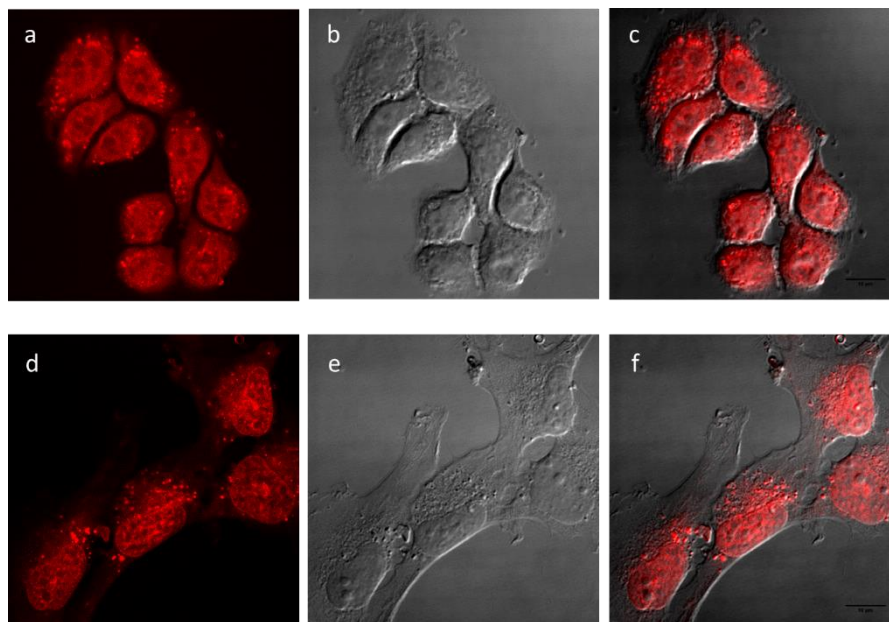
Figure 3.26. Confocal images of HT-29 and MRC-5 cells after co-incubation with **β-glcC2-NDI (66)** (5 µM each) for 2 h. HT-29 cells: a) fluorescence image of compound **6**; b) Transmitted light image; c) merged image; MRC-5 cells: d) fluorescence image of compound **β-glcC2-NDI (66)**; e) Transmitted light image; f) merged image.
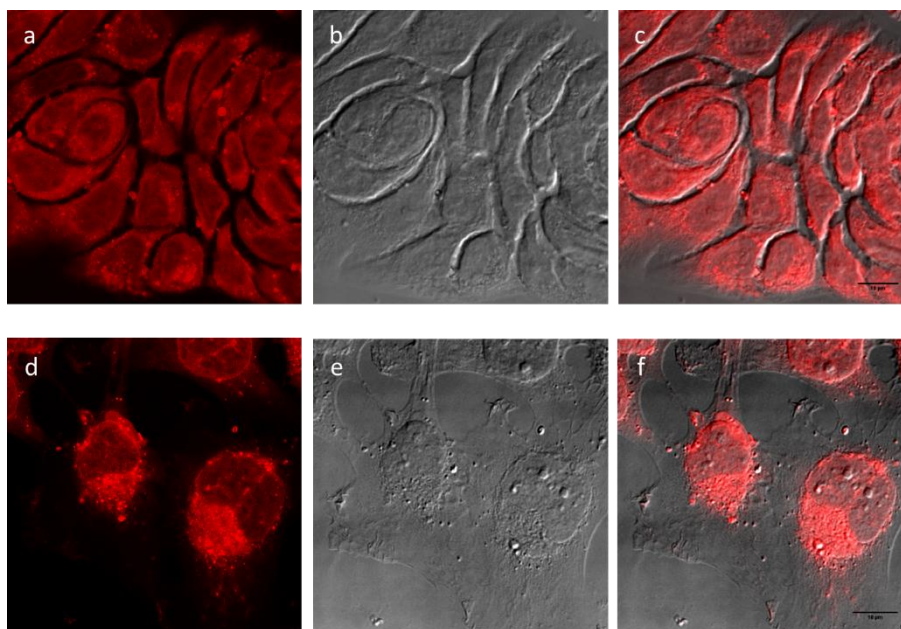
Figure 3.27. Confocal images of HT-29 and MRC-5 cells after co-incubation with **prop-NDI (67)** (5 µM each) for 5 min. HT-29 cells: a) fluorescence image of compound **7**; b) Transmitted light image; c) merged image; MRC-5 cells: d) fluorescence image of **prop-NDI (67)**; e) Transmitted light image; f) merged image.

CHAPTER IV:
G4-LIGANDS AS POTENTIAL ANTIVIRAL AGENTS

1.   INTRODUCTION

In previous chapters, we mentioned the finding of potential G-quadruplex forming sequences in viruses. Actually, we used the HIV-PRO1 sequences in CD and NMR binding studies with our carbohydrate modified phenyl ditriazole ligands. In this section, we focused on the potential application of G4-ligands as antiviral HIV agents since they could be binding DNA/RNA quadruplexes on the HIV genome.

1.1.   HIV and AIDS.

The AIDS (acquired immune deficiency syndrome) is a profound state of immune suppression characterized by the occurrence of opportunistic infections or cancers, caused by the HIV virus (human immunodeficiency virus). The HIV virus infects $CD4^+$ lymphocytes and monocytes/macrophages whose population is dramatically diminished when active viruses spread to the lymphatic tissue. Conversely, latent viruses can remain on the host cells for a long time without syndrome manifestation.

It is still an important global public health problem affecting more than 35 million lives so far and 36.7 million people living with HIV in 2015 (WHO). There is no cure but antiretroviral agents can control the virus and prevent transmission improving patients' quality of life. However, the drugs still present strong side effects and virus resistance, thus new selective agents are needed to improve this condition.

1.1.1.   HIV STRUCTURAL ELEMENTS.

The HIV virus belongs to the *Lentivirus* genus, from the *Retroviridae* family, then the viral particles present its genomic information as RNA. Two types of HIV exist: HIV-1 (the most predominant) and HIV-2. The strains of HIV-1 can be classified into four groups: the "major" group M (majority of infections), the "outlier" group O and two new groups, N and P. HIV-1 strains can be also categorized according to the cell co-receptor they interact with to penetrate inside the host cell:[316]

- **X4 or NL4-3** strains: utilize CXCR4 co-receptor of $CD4^+$T cells, replicate quickly and kill the cells by cytopathogenic effect and by indirect killing $CD8^+$ T cells, which causes significant depletion. On the contrary, they do not infect macrophages. They are associated to progression of the illness.
- **R5 or Bal-L** strains: utilize CCR5 co-receptor in monocytes derived macrophages. The infection with R5 strains of HIV-1 can result in a two-phase process *in vivo*, involving apparently nonpathogenic replication followed by cytopathic replication.

The viral genome consists of two copies of positive single-stranded RNA that codes for nine genes (**gag**, **pol**, **env**, **tat**, **rev**, **vif**, **vpr**, **vpu**, **nef**). When virus is integrated on the human genome is called "provirus or latent virus", with additional **LTR** (long terminal repeat) sequences to control virus transcription (figure **4.1**).[317]
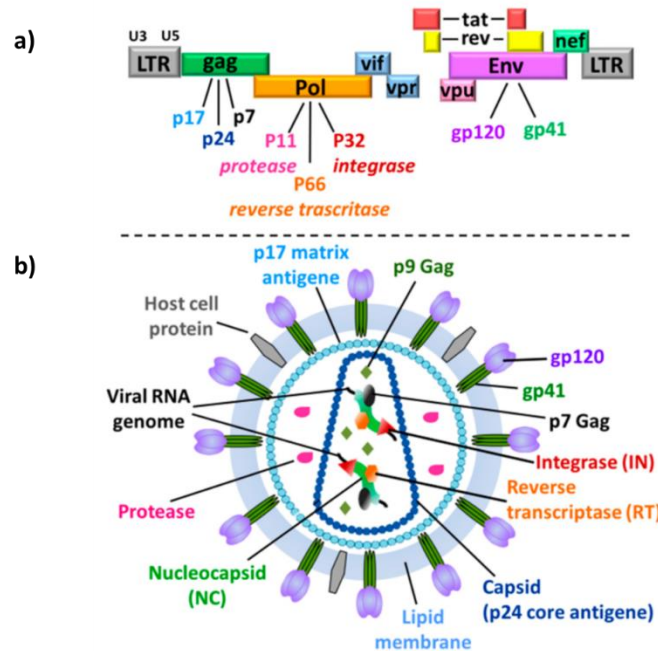
Figure 4.1. a) Schematic view of the HIV genome and the encoding regions. b) HIV structural components.

The main HIV-components are:

- o **LTR:** long terminal repeat, the DNA sequence flanking the genome of integrated proviruses with regulatory regions for transcription initiation and polyadenylation.
- o **GAG:** encodes the capsid proteins. The precursor p55 is processed to p17 (MAtrix), p24 (CApsid), p7 (NucleoCapsid), and p6 proteins. Gag associates with the plasma membrane where the virus assembly takes place.
- o **POL:** encodes the Gag-Pol precursor polyprotein which lead to the enzymes protease, reverse transcriptase and integrase. The <u>protease</u> breaks Gag and Gag-Pol during virus maturation process. The <u>reverse transcriptase</u> presents three key functions on the virus: RNA-dependent DNA polymerase, RNAse H (breaking genomic RNA in RNA-DNA hybrids during DNA synthesis), and DNA-dependent polymerase activity. <u>Integrase</u> cleaves the DNA and inserts the provirus into the genome of the host cell.
- o **ENV:** it encodes the viral glycoproteins gp120 (external glycoprotein) and gp41 (transmembrane glycoprotein). Gp120 contains the binding site for the CD4 receptor, and the seven transmembrane domain chemokine receptors that serve as co-receptors for HIV-1.
- o **TAT:** one of the two essential viral regulatory factor for HIV gene expression (Tat and Rev). Tat binds to the TAR (transactivation responsive element region) RNA element and activates transcription initiation and elongation from the LTR promoter.
- o **REV:** Rev acts by binding to RRE (rev response element) and promoting the nuclear export, stabilization and utilization of the unspliced viral mRNAs containing RRE.
- o **VIF:** viral infectivity factor, is a cytoplasmic protein which promotes the viral particles infectivity.

o **VPR:** viral protein R, it is incorporated to the virion by interaction with p6 Gag part of the p55 Gag precursor. Proposed functions for Vpr include targeting the nuclear import of preintegration complexes, cell growth arrest, transactivation of cellular genes, and induction of cellular differentiation.

o **VPU:** viral protein U, exclusive in HIV-1. It is an integral membrane protein with two main functions: degradation of CD4 in the endoplasmic reticulum and enhancement of virion release from the plasma membrane of HIV-1-infected cells. Vpu is involved in Env maturation.

o **NEF:** Nef is predominantly cytoplasmic and associated with the plasma membrane but has also been identified in the nucleus and linked with the cytoskeleton. It is the most immunogenic accessory protein, indispensable *in vitro*, and for efficient viral spread and disease progression *in vivo*. Nef downregulates CD4, the primary viral receptor and interacts with components of host cell signal transduction. In general, it increases viral infectivity.

o **VPX:** virion protein found in HIV-2 but not HIV-1, a homologous of vpr in HIV-1, but viruses with Vpx carry both vpr and vpx genes. Its function is not fully elucidated.

1.1.2. HIV REPLICATION CYCLE[318–320]

1. The HIV replication cycle begins when the virions recognize the CD4 cell receptors. The viral protein subunits **gp120** (protein SU) and **gp41** (protein TM) binds to the receptor CD4. SU intermediates the CD4 recognition and TM fuse the viral and cell membranes. This step requires cell co-receptors CCR-5 (in macrophages monocytes) and CXCR-4 (in T-cells).

2. Once virions are inside, the **reverse transcriptase** synthesizes dsDNA from the viral RNA in the cytoplasm, which is then carried to the nucleus as a protein-nucleic acid complex (pre-integration complex) and integrated into the host cell DNA by the **integrase enzyme**, leading to the integrated viral DNA (provirus).

3. The provirus is the latent form of the HIV virus and it can remain inactive in the same state for several years, producing few or no copies of the virus. When the host cell receives a signal to become active, the **protein Tat** stimulates the provirus to use the host enzyme RNA polymerase to create copies of the HIV genomic material.

4. The **protein Rev** moves the retroviral RNA to the cytoplasm where viral proteins are synthesized by using the host cell machinery.

5. **Gag** and **pol** capsid proteins assembly with two copies of the viral RNA to give rise immature viral particles. These particles move to the cell membrane to be covered by a lipid bilayer which also includes **Env protein** in a trimeric form.

6. At the same time the enzyme **protease** breaks Gag and Gag-Pol to release the central capsid and the Pol enzymes in their final processed form to configure the mature and infective viral particles.
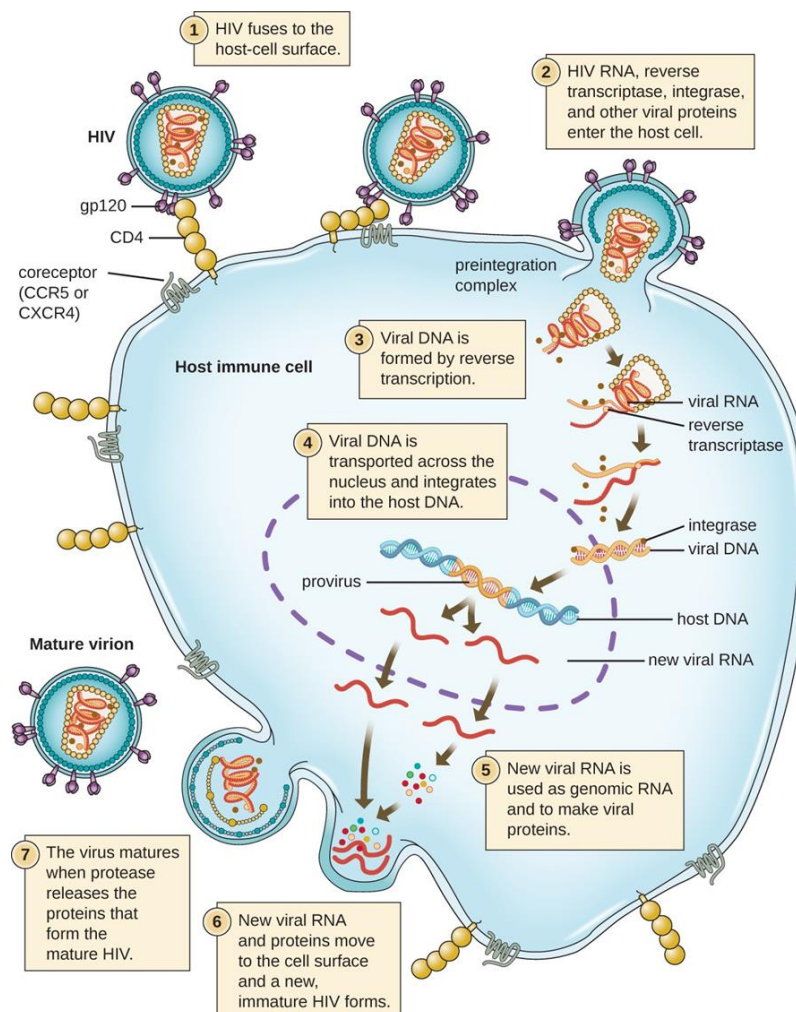
Figure 4.2. HIV replication cycle. Image source: NIAID (National Institute of Allergy and Infectious Diseases).

1.1.3. THERAPEUTIC STRATEGIES AGAINST HIV

The virus takes advantage of the host cell machinery to survive and spread the infection which complicates the therapeutic strategy. Researchers have studied the viral life cycle in detail finding potential targets in each step:

➢ **Virus-receptor interaction and entry:** there are some processes to target during the first stages of the viral replication cycle: the CD4-SU interaction, the SU-chemokine co-receptor interaction, and the TM-mediated virus-cell membrane fusion process.

o CD4-SU interaction: for example the agents Cyanovirin[321] or dextran sulphate.[322]

o The SU-chemokine co-receptor interaction: some chemokines and the co-receptors CCR-5 and CXCR-4 antagonist exist, such as AOP-RANTES,[323] cyclophilin-18[324] or AMD3100.[325]

o TM-mediated virus-cell membrane fusion: such as D-peptides.[326]

➢ **Reverse transcription and proviral integration:** the **reverse transcriptase** is essential for viral replication, then it has been an extensive target. Among reverse transcriptase

inhibitors we find <u>nucleosides</u>, such as Zidovudine,[327] Didanosine, Zalcitabine or Lamivudine; and <u>non-nucleosides</u>, for example Nevirapine, Delaviridine and Efavirenz.[328] In addition, RNA aptamers template analogue reverse transcriptase inhibitors have been developed.[329,330] In the same way, there is not an **integrase** analogue in the human cells so it is a common target as well, which generally involves the usage of oligonucleotides, dinucleotides and different kinds of chemical agents, such as dicaffeoylquinic acids (DCQAs) and 2,4-dioxobutanoic acid analogous.[331,332]

➢ **HIV-transcription:** the Tat/TAR interaction is essential for viral transcription activation thus many strategies have been developed for this process: a mutated **Tat** protein, a single chain Tat antibody (sFv-Tat),[333–335] Tat competitors (TR87),[336] or the use of antisense-, nuclease-, siRNA-based strategies, directed against Tat $_m$RNA.[333,335,337] Besides targeting the Tat encoding regions, siRNAs have been directed against the Gag and Nef encoding regions.[338]

➢ **HIV-1 mRNA splicing and nuclear export:** in this stage, the strategies are based on inhibiting the Rev/RRE interaction. Examples include Rev TNPs (transdominant negative protein such as RevM10),[339] RRE RNA decoys, anti-sense Rev/RRE RNAs, siRNAs, sFvs, nucleases, ribozymes, aptamers, chimeric proteins, and small inhibiting molecules.

➢ **HIV-translation:** anti-sense RNAs against sequences located in the 5' UTR of all HIV-1 mRNAs were designed to hinder the ribosome-complex in completing the translation process and thereby inhibit the protein synthesis.[340]

➢ **Viral assembly, release and maturation:** Strategies directed against this final step in the viral life cycle include RNA decoys, TNPs, chimeric proteins, anti-infectious cellular proteins, sFvs (intracellular single-chain antibodys), nucleases, anti-sense RNAs, ribozymes and peptides (polyarginine as furin inhibitor).[333,334,341] In addition, **protease** plays a key role in virus maturation and inhibitors have been successfully developed, for instance Saquinavir, Indinavir, Ritonavir, Nelfinavir and Amprenavir.[328,342,343]

Many strategies have been tried to fight against HIV virus, however many problems still remain. This virus has high mutation rates leading to more resistant variants facing towards the antiviral agents. This forces to use combinatorial therapies to cover several stages of the virus life cycle and diminish the mutant production. For this reason, new therapeutic strategies are needed specially focused on highly conserved viral sites.

1.2. G4-forming sequences among the HIV genome.

The probable role of the G4 on the regulation of DNA transcription and replication inspired researchers to look for such G-rich sequences on the HIV genome as a new potential strategy against the virus. Effectively, some groups have identified significant PQS in the Nef coding region and in the unique long terminal repeat (LTR) promoter of HIV-1.

In 1993, Sundquist and Heaphy[344] already reported the formation of an interstrand quadruplex on the dimeric RNA of HIV-1 stabilized by guanine (and/or purine)-base tetrads in potassium rich medium. The authors suggested it may mediate the association of genomic HIV-1 RNA *in vivo*. Lyonnais *et al.*[345] described the intermolecular quadruplex formation in the

corresponding central DNA flap and interestingly, the HIV-1 nucleocapsid protein (NCp) showed a high affinity for this quadruplex. They proposed such structures may act as a non-resolvable DNA intermediate trapping the NCp chaperone activity in its transition step, and may be relevant to the biological assembly of mature nucleocapsid in the virion core.

In the viral replication cycle we stated the importance of the **Nef** protein as essential in viral replication and pathogenesis *in vivo*. It alters host cells processes by downregulating CD4 and MHC I expression on the cell surface to escape from the immune system and infectivity or activating CD4[+] cytolytic T lymphocytes to enhance viral replication and infectivity.[346] The Nef coding region is located at the 3'end of the viral genome and partially overlaps with the 3'-long terminal region (LTR). Perrone *et al*.[95] reported the presence of three G4-forming sequences on the Nef coding region, some of them very conserved among most HIV-1 strains (figure **4.3**). They demonstrated these sequences fold into a G4 conformation in the presence of potassium. G4-binders (TMPyP4, BRACO-19 and Piper) incremented the T_m of the corresponding G4 and more importantly, inhibited the viral infectivity. In addition, one of the G4-sequences found was able to pause the polymerase through a G4-ligands non-dependent manner, suggesting that it may naturally function as a molecular switch to assist protein expression. Therefore, the stabilization of this sequence by G4-ligands may impair Nef expression and transcription for production of new copies of the RNA genome.
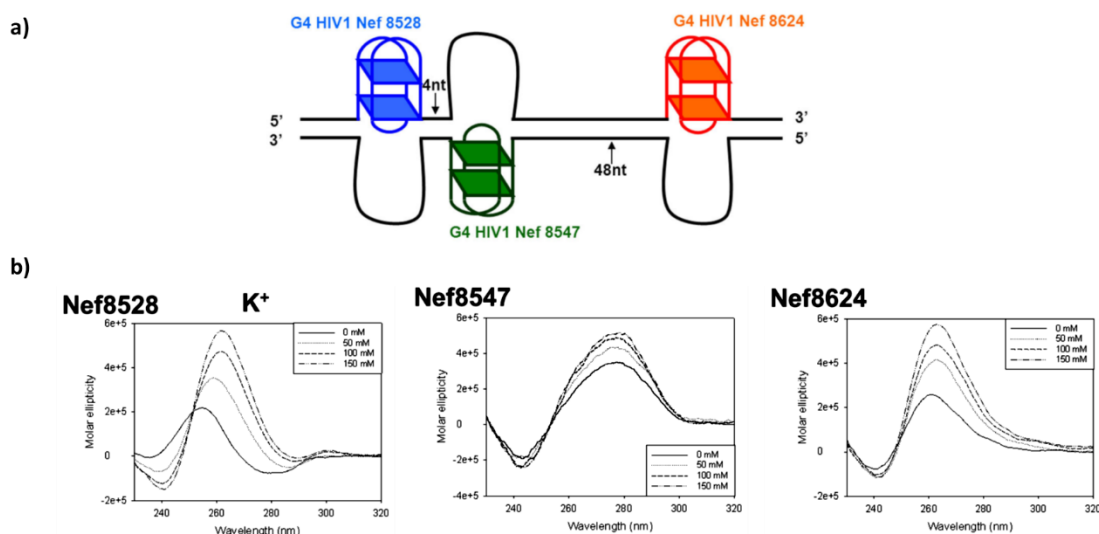
Figure 4.3. a) Scheme of G4 formation within the double-stranded DNA of the nef region: Nef8528, Nef8547, Nef8624 G-4 structures are shown in blue, green and red, respectively. The numbers of nts separating each G4 structure are indicated. The scheme indicates the possibility of formation of a cluster of non-canonical DNA structures within a small portion (112 nts) of the HIV-1 genome. b) CD spectra of the putative G-4 forming oligonucleotides in the nef region in the absence and presence of increasing concentrations (50-150 mM) of K[+] cations.

The integrated virus genome (provirus form) presents two identical **LTR** sequences located in both extremes. These sequences consist of three regions: *U3*, *R*, and *U5*. *U3* in turn is form by three sections:

a) an upstream regulatory element including binding sites for cellular transcription factors,

b) an enhancer with two binding sites for the nuclear factor κB (NF-κB), a
c) nd the core promoter composed of three tandem binding sites for specificity protein 1 (Sp1) and a TATA box (figure **4.4**).

The *R* region contains the unique transactivation response enhancer TAR which modulates Tat activation.[347]
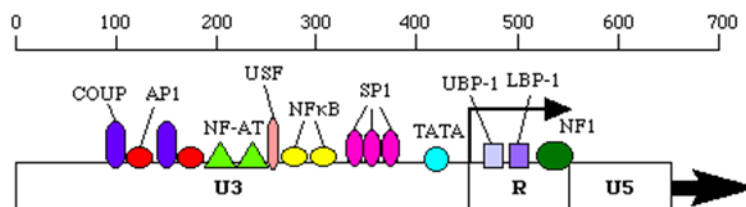
Figure 4.4. Schematic representation of the LTR sequence regions

Once the U3 promoter sequence binds cellular factors, HIV-1 transcription starts and the gene expression considerably increases when the viral protein Tat interacts with the 5'end of nascent viral transcripts.

Prof. Richter's group analyzed the corresponding region to NF-κB and Sp1 sites, and found abundant G-tracts.[20] They identified three main potential G4-forming sequences (LTR-II, LTR-III and LTR-IV) (figure **4.5a**) whose CD spectra showed the typical G4-pattern (figure **4.5b**). This finding suggested the possibility of inhibiting the HIV-1 LTR promoter activity by G -quadruplex interacting small molecules. They assayed the strong G4-binder BRACO as antiviral agent obtaining great results (IC_{50}~3 µM) with minimal effect on cell viability. Later, they achieved anti-HIV activity at nanomolar concentrations with extended core NDI derivatives (figure **4.5c**) with selectivity index (SI) up to 40 for the corresponding bisubsituted NDI derivative.[348] These compounds displayed by FRET melting assay higher stabilization effect on the G4-forming sequence LTR-III than with hTel sequence, with differences of 8 °C on the ΔT_m in some cases. In addition, it was checked the best compound acted after virus entrance and reduced LTR-III activity with considerably decrease in the intensity of EGFP fluorescence over HeLa-tat-III cells (cell line stably transfected with a plasmid embedding the EGFP coding region under the control of the HIV-1 LTR promoter).
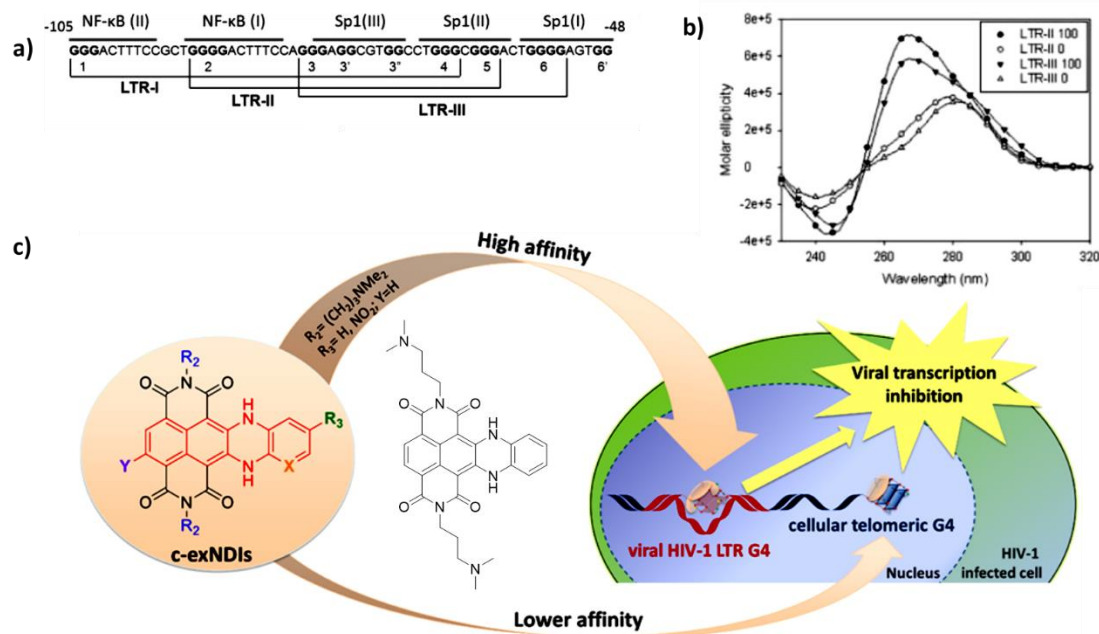
Figure 4.5.a) G-tracts identified as G4-forming sequences. b) CD spectra of the sequences in the absence (0) or presence (100) of K$^+$. c) Scheme of the chemical properties that lead to high-affinity recognition and selectivity of the c-exNDI binders.

At the same time, Amrane *et al.*[96] found a high level of conservation of the eight blocks of guanines on the U3 region which spans the three SP1 binding sites, from the 5′ extremity of SP1-3 to the 3′ extremity of SP1-1. All these sequences had the ability to fold into G4 arrangements *in vitro* and they chose the most represented motif called HIV-PRO1, whose topology was described in the second chapter.

The G4-antiviral approach presents some advantages as new anti-HIV therapy:

- ➢ The G4-binders target viral DNAs and RNAs, the actual source of the disease.
- ➢ The high conservation of these targets across all HIV-1 subspecies suggests that they are important for HIV-1 life cycle and not mutable, decreasing the appearance of resistant mutants.
- ➢ The HIV-PRO1 sequence cannot be found in the Human genome.

On the other hand, this strategy presents some disadvantages as well:

- ➢ Similar G4 sequences from the human genome could also be bound interfering the normal cell function.
- ➢ G4-ligands used to target HIV G4 sequences are not drug-like molecules

2. AIMS.

+ To evaluate the potential activity of the carb-PTDZ ligands as antiviral agents on NL4.3 HIV strain.
+ To study the antiviral activity of the carb-NDI conjugates.
+ To study the potential antiviral activity of new carb-NDI conjugates with extended linkers and tetrasubstituted NDIs.

3. RESULTS AND DISCUSSION.

Taking into account the potential application of the reported G4-ligands as anti-HIV agents we decided to evaluate our carb-PTDZ and carb-NDI derivatives. During a short stay in the laboratory of Prof. Sara Richter (University of Padua) I was able to carry out those experiments on a P3 laboratory (Biosafety level 3). We already described the strong stabilization effect of the carbohydrate-modified NDI derivatives on several G4-forming sequences, and other NDI ligands have been reported to have antiviral activity.[348] We included a new series of carb-NDI derivatives synthesized by Dr. Filippo Doria (University of Pavia) with longer linkers. The carb-PTDZ derivatives did not display extremely strong G4 stabilization ability but they were able to increase the T_m of several DNA-HIV sequences (see **chapter II** section **3.2**).

3.1. Antiviral assay with TZM-bl cells.

Viruses do not have the machinery associated to nucleic acid replication and protein synthesis so they require an alive entity to survive and replicate. For this reason, they cannot be cultured to carry out an antiviral assay as it is normally done for cells, they must be inside host cells. Among the different methodologies available to evaluate the antiviral activity, Prof. Richter's laboratory uses TZM-bl cells.[20] TZM-bl cells are a HeLa cell clone that was engineered to express CD4 and CCR5 receptors and contains integrated reporter genes for firefly luciferase and *E.Coli* β-galactosidase under control of HIV-1 LTR. To activate the reporter genes, viral protein Tat induces the expression soon after infection, then viral particles are needed to promote this process. After 48 h post infection, luciferase activity is quantified by luminescence, whose intensity is proportional to the number of infectious virus particles present in the initial inoculum. If the G4-ligand is able to stabilize the G4 associated to the LTR region, viral Tat does not activate transcription and luciferase is not expressed, which leads to a reduction in luciferase activity respect to the non-treated control.

The experiment is performed on TZM-bl cells, which will be infected with viral particles and immediately treated with the G4-ligands at serial concentrations. The antiviral activity is measured after 48 h post-infection. However, G4-ligands normally act as anticancer agents and TZM-bl are cancer cells, so it is necessary to carry out a MTT assay with non-infected cells concurrently to obtain the cytotoxicity. A decrease in the luminescence may be due to actually a reduction in cell population, thus it is essential to know the products CC_{50} (cytotoxicity). The MTT

assay is performed at the same concentration of ligands and same conditions than the antiviral assay. The selectivity index (**SI**) determines the therapeutic window, dividing the compound CC_{50} into its IC_{50} (CC_{50}/IC_{50}), considering an interesting antiviral candidate that one with SI higher than 10.

TZM-bl cells have the advantage of being adherent cells with high replication rate, whose maintenance does not require any additional supplement or especial incubation conditions. It is a good methodology to have a preliminary idea of the compounds antiviral activity. Nevertheless, although they express CD4 and CCR5 receptors, they are not the natural virus host cells and results are only an approximation. MT-4 cell line (human T cell leukemia) is a more proximal model to the real HIV host cell infection but with more disadvantages: they grow in suspension and their culture is more complex and expensive. For this reason, most antiviral experiments are first carried out on TZM-bl cells, and the best candidates are then assayed over MT-4 cells.

3.1.1. ANTIVIRAL ASSAY FOR THE **CARB-PTDZ** LIGANDS.

The first ligands to assay as potential antiviral agents were the phenyl ditriazole derivatives. They displayed some stabilization ability over the HIV sequences used in the FRET melting assays (see chapter **II** section **3.2.1**). We selected some asymmetric PTDZ derivatives and some symmetric-carb-PTDZ (figure **4.6**). The second group did not present any G4-stabilization properties but still they may act as antiviral agents by an alternative mechanism.
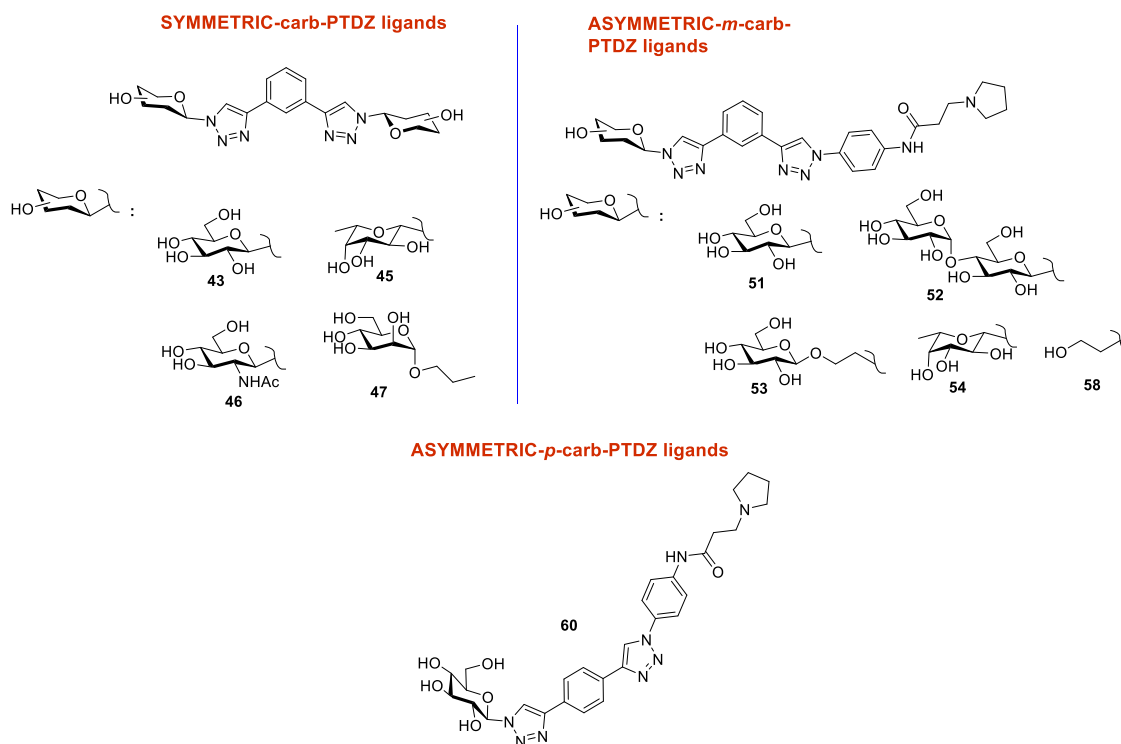
Figure 4.6. PTDZ derivatives assayed as antiviral agents.

The results obtained for these series of ligands using the TZM-bl cells are shown in table **4.1**. They are displayed as IC_{50} for anti-HIV activity and CC_{50} for cytotoxicity over TZM-bl cells.

	PTDZ-ligand	IC$_{50}$ (µM)	CC$_{50}$ (µM)	SI
Symmetric PTDZ	**43** (β-glc-PTDZ)	>100	>100	1.00
	45 (β-fuc-PTDZ)	>100	>100	1.00
	46 (β-glcNAc-PTDZ)	>100	>100	1.00
	47 (α-manC2-PTDZ)	40.0	>100	2.50
Asymmetric PTDZ	**51** (β-glc-*m*-PTDZ)	57.1	>100	1.75
	52 (β-malt-*m*-PTDZ)	>100	>100	1.00
	53 (β-glcC2-*m*-PTDZ)	>25	>25	1.00
	54 (β-fuc-*m*-PTDZ)	>100	>100	1.00
	58 (aglycone-PTDZ)	97.8	>100	1.02
	60 (β-glc-*p*-PTDZ)	61.37	93.34	1.52

Table 4.15. Antiviral activity (IC$_{50}$), cytotoxicity (CC$_{50}$) and selectivity index (SI) of the carbohydrate modified-PTDZ ligands.

Neither of the assayed carb-PTDZ ligands showed any antiviral activity, asymmetric nor symmetric PTDZ derivatives. The **symmetric mannose-PTDZ** gave the best result with a IC$_{50}$ of 40 µM, when it begins to present toxicity on cells. Two examples of the experiments run are shown in figure **4.7**.
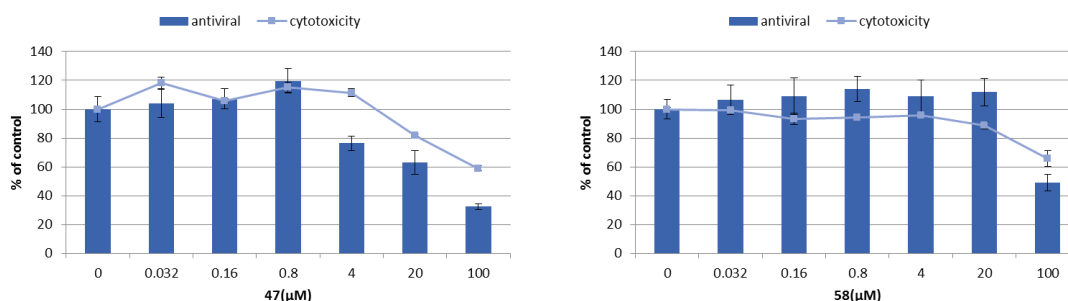
Figure 4.7. Graphs with results of compounds **47** (mannose symmetric PTDZ) and **58** (aglycone asymmetric PTDZ).

3.1.2. ANTIVIRAL ASSAY FOR THE **CARB-NDIs** LIGANDS

The same experiments to measure anti HIV activity were carried out for carb-NDIs, the ones described in the previous chapter and a new series provided by the Dr. Filippo Doria (University of Pavia). The second series of carb-NDIs possesses an extended spacer with the triazole ring in the middle (all compounds are shown in figure **4.8**).
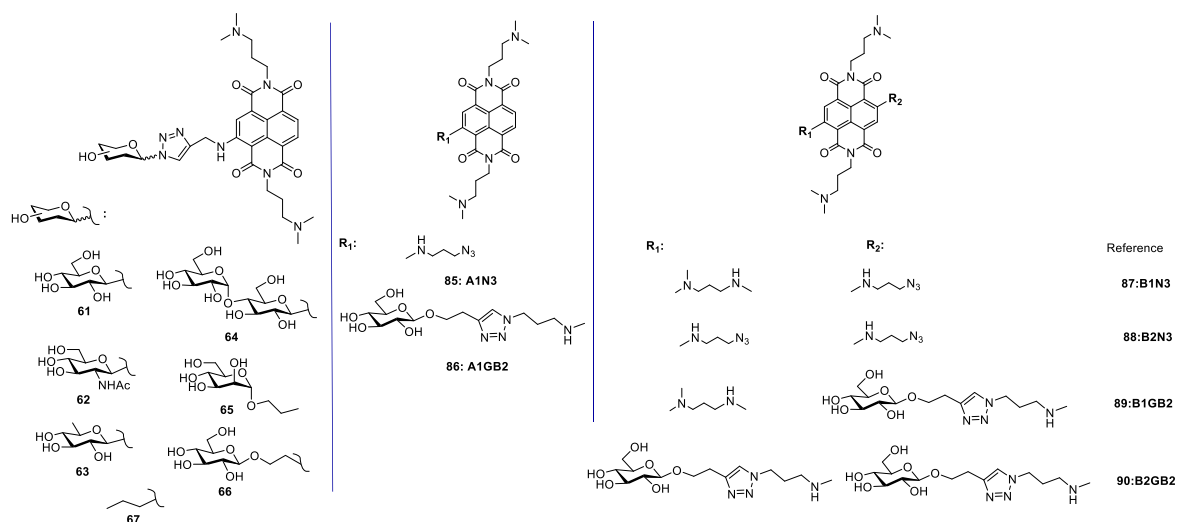
Figure 4.8. NDI structures assayed as antiviral agents.

Since several NDI derivatives had been reported previously to possess excellent antiviral activity and selectivity index, we expected that the carb-NDIs could also exhibit antiviral properties. Antiviral and cytotoxicity assays results are shown in table **4.2**.

NDI derivative	IC$_{50}$ (nM)	CC$_{50}$ (nM)	SI
61:β-Glc-NDI	>500	>500	1.00
62: β-GlcNAc-NDI	>500	>500	1.00
63: β-6-dGlc-NDI	>500	>500	1.00
64: β-Malt-NDI	>500	>500	1.00
65:α-ManC2-NDI	137.4	>500	3.64
66: β-GlcC2-NDI	855.5	848.5	1.01
67:Prop-NDI	127.6	152.3	1.19
85:A1N3	138.9	131.29	0.95
87:B1N3	156.9	141.8	0.9
88:B2N3	>500	>500	1.00
86:A1GB2	20895	>25000	1.20
89:B1GB2	5410	17595	3.25
90:B2GB2	>25000	>25000	1.00

Table 4.16. antiviral activity(IC$_{50}$), cytotoxicity (CC$_{50}$) and selectivity index (SI) of the carbohydrate modified-NDI ligands.

According to the highest SI, the best candidate was **α-manC2-NDI** (**65**), although still too low to be considered as a good antiviral agent. We observed that the NDIs derivatives with carbohydrates in their structure presented lower cytotoxicity than the aglycones **prop-NDI**, **A1N3** and **B1N3**, which maintained their CC$_{50}$ value in the nanomolar range with very low selectivity index. The exception was **B2N3**, which displayed similar results to our carb-NDI derivatives described in chapter III, both with only two charged side chains. The new carb-NDI ligands with longer linkers than the previous ones showed lower toxicity, increasing the CC$_{50}$ range to the micromolar range. However, they did not present any antiviral activity either. The results for compound **B1GB2** are quite surprising. In spite of presenting three charged side chains, its cytotoxicity and antiviral activity were very low, even worse than **β-GlcC2-NDI**, with two charged

chains and the sugar linked through a shorter linker. It seems that the spacer between the triazole and NDI core have a detrimental effect on the antiviral activity of the NDI derivatives.

4. CONCLUSIONS.

➢ The carbohydrate-modified phenyl ditriazole ligands did not exhibit any anti-HIV activity since they did not decrease the viral particles production on TZM-bl cells.

➢ The carb-NDIs presented some antiviral activity, but at concentrations which they were cytotoxic for TZM-bl cells, so they did not display good features as antiviral agents.

➢ The extension of the linker in the new carb-NDIs produced a detrimental effect on both cytotoxicity and antiviral activity.

5. EXPERIMENTAL DETAILS.

5.1. Compounds stock

Compound stocks were prepared in DMSO at 10 or 20 mM concentration and maintained at -20 °C. The biological experiments were carried out with subsequent dilution in the corresponding cell culture medium.

5.2. Cell culture.

TZM-bl cell line (obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH, from Dr. J.C. Kappes, Dr. X. Wu, and Tranzyme Inc.) was maintained in DMEM high glucose with 10% fetal bovine serum and supplementation of L-glutamine and pyruvate. Cells were incubated at 37 °C and 5% CO_2.

5.3. Antiviral assay.

Viral stocks NL4.3 were previously prepared by Prof. Sara Richter's group, by transfection of HEK293T cells with wild type and and Nef-defective (ΔNef) proviral genomes and amplified in the T cell line MT2. For the experiment, TZM-bl cells were seeded on a 96 wells plate (10^4 cells per well). After 24 h, the medium is removed and virus particles solution in medium are added to the plate at a MOI of 0.1. A not-treated control was included as 100% viral production reference. Compounds are then added to the plate at serial dilutions by triplicate. The plate was incubated at 37 °C and 5% CO_2 for 48h. After incubation HIV production was measured by following the long terminal repeat (LTR) luciferase signal using the Britelite[TM] plus Reporter Gene Assay System (Perkin Elmer, Waltham, MA) according to the manufacturer's protocol. The 50% inhibitory concentration (IC_{50}) was defined as the concentration of test compound that inhibit 50% of LTR-Luciferase signal.

5.4. Cytotoxicity assay.

The cytotoxicity assay was carried out concurrently to the antiviral assay. TZM-bl cells were seeded in 96-well plates (10^4), grown overnight and incubated with compounds in DMSO carrier solvent. After 48 h of incubation at 37 °C and 5 % CO_2, cytotoxicity was assessed using the Cell Titer Blue reagent. The 50% cytotoxic concentration (CC_{50}) was defined as the concentration of test compound that was able to reduce the absorbance of cells by 50%.

The selectivity index (SI) is the relative effectiveness of the tested compound in inhibiting viral replication compared to inducing cell death (CC_{50} value/IC_{50} value).

CONCLUSIONES

- Se han sintetizado derivados de fenil ditriazoles en los que se ha incorporado carbohidratos en las cadenas laterales mediante "click chemistry" (carb-PTDZ). Se han optimizado las condiciones de reacción para sintetizar derivados simétricos, en los que ambas cadenas están conformadas por azúcares neutros. Por otro lado, se ha optimizado las condiciones de reacción para sintetizar compuestos asimétricos en los que existe un carbohidrato como cadena lateral y un residuo con una amina terciaria cargada a pH fisiológico como segunda cadena lateral.

- El estudio de las características de unión de los carb-PTDZ a diferentes ADN en cuádruplex (G4) se ha efectuado mediante diversas técnicas biofísicas: FRET, CD, SPR, espectrometría de masas y RMN.

- Los estudios de FRET han revelado que los compuestos simétricos no son capaces de estabilizar ninguna de las secuencias de ADN en cuádruplex ensayadas. Por el contrario, los derivados de fenil ditriazol asimétricos estabilizan dichas estructuras mediante un incremento moderado de la T_m. Se requiere, por tanto, de al menos una cadena cargada para observar estabilización de G4. Los resultados no muestran una mejora en la afinidad respecto a los ligandos de cuádruplex previamente publicados. Sin embargo, cabe destacar la gran selectividad que presentan hacia las estructuras de G4 frente a secuencias de dúplex de ADN. Los estudios de SPR confirmaron esta tendencia.

- En los estudios de unión realizados por RMN de ^{1}H se observó que los ligandos carb-PTDZ asimétricos produjeron cambios en las señales de los protones imino de las secuencias de ADN en cuádruplex HIV-ADR, HIV321 y 24TTA. Se observó cambio en el desplazamiento químico de las señales y ensanchamiento de las mismas hasta su desaparición en algunas ocasiones. Cabe destacar las diferencias observadas entre los compuestos asimétricos con los triazoles situados en *meta-* o en *para-* respecto al fenilo central con la secuencia de HIV-321. La adición de los derivados en *meta-* sobre la secuencia de HIV321 produjo cambios en mayor número de protones imino que el derivado en *para-*.

- Estudios bidimensionales de RMN (NOESY) permitieron localizar los cambios producidos por el ligando asimétrico **53** (***m*-glcC2-PTDZ**) sobre la secuencia del telómero

humano 22AG en presencia de sodio. Los cambios en el desplazamiento químico de los protones de la secuencia se localizaron principalmente en los protones situados en los loops paralelos y la tétrada de guaninas contigua.

- Los estudios de citotoxicidad de los conjugados de fenil ditriazoles con carbohidratos sobre diferentes líneas celulares tumorales humanas (MCF-7, HT-29, HeLa y LN-229) mostraron una capacidad muy moderada de actividad antiproliferativa, con valores superiores a 100 µM en la mayoría de los casos. Los estudios de estabilidad de los compuestos en agua y medio de cultivo no manifestaron inestabilidad en los derivados ensayados, pero sí ausencia de ellos en las muestras de lisados celulares. Se sospecha que la ausencia de actividad antiproliferativa se debe a una baja o nula entrada celular de dichos ligandos.

- Se han sintetizado con éxito conjugados de carbohidratos con NDIs (carb-NDI) en colaboración con el Prof. Mauro Freccero (Universidad de Pavía). El acoplamiento se ha llevado a cabo a través de "click chemistry".

- Los conjugados carb-NDI estabilizaron la secuencia del telómero humano hasta 20 °C a tan solo una concentración de 2 µM de los derivados. Además, de acuerdo con los resultados del ensayo de competitividad, son bastante selectivos de las secuencias formadoras de ADN en cuádruplex. La presencia del carbohidrato no produjo un efecto perjudicial en la afinidad hacia los ADN en cuádruplex, en comparación con el compuesto control sin azúcar **prop-NDI 67**.

- Los valores de IC$_{50}$ obtenidos en los ensayos de citotoxicidad sobre diferentes líneas tumorales (MCF-7, HT-29 y HeLa) se encuentran dentro del rango nanomolar o bajo micromolar, por lo que presentan gran capacidad antiproliferativa. El más toxico de la serie fue el compuesto control sin azúcar **67 (prop-NDI)**, seguido del compuesto **65 (α-mannC2-NDI)** y **66 (β-glcC2-NDI).**

- Según los estudios de entrada celular realizados mediante citometría de flujo y cuantificación por espectrometría de fluorescencia, el compuesto control **prop-NDI 67** entra rápidamente en la célula de una forma no dependiente de la temperatura o el tiempo de incubación, por lo que parece que este compuesto entra principalmente por difusión pasiva.

Por el contrario, la entrada de los conjugados carb-NDI fue dependiente de la temperatura, tiempo de incubación, tipo de carbohidrato y su accesibilidad, lo que apunta a un posible mecanismo de entrada celular a través de transportadores de membrana de tipo GLUT.

- Se pudo relacionar parcialmente la citotoxicidad con las diferencias observadas en la entrada celular de los conjugados carb-NDIs. El derivado **66 (β-glcC2-NDI)** presentó una mayor citotoxicidad y una entrada celular superior al correspondiente derivado cuyo carbohidrato se encuentra conectado directamente al triazol (**61: β-glc-NDI**).

- Los estudios de inhibición de transportadores GLUT mostraron que los compuestos **65 (α-mannC2-NDI)** y **66 (β-glcC2-NDI)** son transportados, al menos parcialmente, por GLUT4.

- Los ensayos de microscopía confocal localizaron los conjugados carb-NDI principalmente en el núcleo celular, donde se encuentra su potencial diana farmacológica, los ADN en cuádruplex.

- Por último, se llevaron a cabo ensayos de actividad antiviral frente a VIH de los carb-PTDZ y de los carb-NDI. Ambas familias presentaron una actividad baja como antivirales.

REFERENCES

1.   Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737–738.

2.   Zimmerman, S. B. *Annu. Rev. Biochem.* **1982**, *51*, 395–427.

3.   Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907–916.

4.   Bacolla, A.; Collins, J. R.; Gold, B.; Chuzhanova, N.; Yi, M.; Stephens, R. M.; Stefanov, S.; Olsh, A.; Jakupciak, J. P.; Dean, M.; Lempicki, R. A.; Cooper, D. N.; Wells, R. D. *Nucleic Acids Res.* **2006**, *34*, 2663–2675.

5.   Bang, I. *Biochem. Zeitschrift* **1910**, *26*, 293–311.

6.   Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013–2018.

7.   Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D. *J. Am. Chem. Soc.* **1978**, *100*, 3625–3627.

8.   Wong, A.; Wu, G. *J. Am. Chem. Soc.* **2003**, *125*, 13895–13905.

9.   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402–5415.

10.   Kim, J.; Cheong, C.; Moore, P. B. *Nature* **1991**, *351*, 331–332.

11.   Todd, A. K. *Nucleic Acids Res.* **2005**, *33*, 2901–2907.

12.   Huppert, J. L. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.

13.   Bedrat, A.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res.* **2016**, *44*, 1746–1759.

14.   Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J.-L.; Phan, A. T. *J. Am. Chem. Soc.* **2012**, *134*, 5807–5816.

15.   Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.-M.; Lemaitre, J.-M. *Nat. Struct. Mol. Biol.* **2012**, *19*, 837–844.

16.   Bugaut, A.; Balasubramanian, S. *Nucleic Acids Res.* **2012**, *40*, 4727–4741.

17.   Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. *Science* **2007**, *318*, 798–801.

18.   Smargiasso, N.; Gabelica, V.; Damblon, C.; Rosu, F.; De Pauw, E.; Teulade-Fichou, M.-P.; Rowe, J.; Claessens, A. *BMC Genomics* **2009**, *10*, 362.

19.   Du, X.; Wojtowicz, D.; Bowers, A. A.; Levens, D.; Benham, C. J.; Przytycka, T. M. *Nucleic Acids Res.* **2013**, *41*, 5965–5977.

20.   Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palù, G.; Richter, S. N. *J. Med. Chem.* **2013**, *56*, 6521–6530.

21.   Frees, S.; Menendez, C.; Crum, M.; Bagga, P. S. *Hum. Genomics* **2014**, *8*, 8.

22.   Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci.* **2001**, *98*, 8572–8577.

23.   Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182–186.

References

24. Biffi, G.; Tannahill, D.; Miller, J.; Howat, W. J.; Balasubramanian, S. *PLoS ONE* **2014**, *9*, e102711.

25. Granotier, C. *Nucleic Acids Res.* **2005**, *33*, 4182–4190.

26. Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825–829.

27. Harley, C. B.; Futcher, A. B.; Greider, C. W. *Nature* **1990**, *345*, 458–460.

28. Greider, C. W.; Blackburn, E. H. *Cell* **1985**, *43*, 405–413.

29. Smith, J. S.; Chen, Q.; Yatsunyk, L. A.; Nicoludis, J. M.; Garcia, M. S.; Kranaster, R.; Balasubramanian, S.; Monchaud, D.; Teulade-Fichou, M.-P.; Abramowitz, L.; Schultz, D. C.; Johnson, F. B. *Nat. Struct. Mol. Biol.* **2011**, *18*, 478–485.

30. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science* **1994**, *266*, 2011–2015.

31. Makarov, V. L.; Hirose, Y.; Langmore, J. P. *Cell* **1997**, *88*, 657–666.

32. de Lange, T. *Cold Spring Harb. Symp. Quant. Biol.* **2010**, *75*, 167–177.

33. Zaug, A. J.; Podell, E. R.; Cech, T. R. *Proc. Natl. Acad. Sci.* **2005**, *102*, 10864–10869.

34. Valton, A.-L.; Prioleau, M.-N. *Trends Genet.* **2016**, *32*, 697–706.

35. Wang, Y.; Patel, D. J. *Struct. Lond. Engl. 1993* **1993**, *1*, 263–282.

36. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. *Nature* **2002**, *417*, 876–880.

37. Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.

38. Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2006**, *34*, 5715–5719.

39. Xu, Y.; Sugiyama, H. *J. Am. Chem. Soc.* **2004**, *126*, 6274–6279.

40. Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.* **2005**, *33*, 4649–4659.

41. Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14629–14634.

42. Ourliac-Garnier, I.; Elizondo-Riojas, M.-A.; Redon, S.; Farrell, N. P.; Bombard, S. *Biochemistry (Mosc.)* **2005**, *44*, 10620–10634.

43. Schoeftner, S.; Blasco, M. A. *Nat. Cell Biol.* **2008**, *10*, 228–236.

44. de Silanes, I. L.; d'Alcontres, M. S.; Blasco, M. A. *Nat. Commun.* **2010**, *1*, 1–9.

45. Randall, A.; Griffith, J. D. *J. Biol. Chem.* **2009**, *284*, 13980–13986.

46. Martadinata, H.; Phan, A. T. *J. Am. Chem. Soc.* **2009**, *131*, 2570–2578.

47. Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N. *Nucleic Acids Res.* **2010**, *38*, 5569–5580.

48. Cusanelli, E.; Chartrand, P. *Front. Genet.* **2015**, *6*.

49. Arora, R.; Azzalin, C. M. *RNA Biol.* **2015**, *12*, 938–941.

50. Porro, A.; Feuerhahn, S.; Delafontaine, J.; Riethman, H.; Rougemont, J.; Lingner, J. *Nat. Commun.* **2014**, *5*, 5379.

51.   Deng, Z.; Norseen, J.; Wiedmer, A.; Riethman, H.; Lieberman, P. M. *Mol. Cell* **2009**, *35*, 403–413.

52.   Biffi, G.; Tannahill, D.; Balasubramanian, S. *J. Am. Chem. Soc.* **2012**, *134*, 11974–11976.

53.   Martadinata, H.; Phan, A. T. *Biochemistry (Mosc.)* **2013**, *52*, 2176–2183.

54.   Eddy, J. *Nucleic Acids Res.* **2006**, *34*, 3887–3896.

55.   Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2006**, *35*, 406–413.

56.   Smestad, J. A.; Maher, L. J. *BMC Med. Genet.* **2015**, *16*.

57.   Johnson, J. E.; Cao, K.; Ryvkin, P.; Wang, L.-S.; Johnson, F. B. *Nucleic Acids Res.* **2010**, *38*, 1114–1122.

58.   Nguyen, G. H.; Tang, W.; Robles, A. I.; Beyer, R. P.; Gray, L. T.; Welsh, J. A.; Schetter, A. J.; Kumamoto, K.; Wang, X. W.; Hickson, I. D.; Maizels, N.; Monnat, R. J.; Harris, C. C. *Proc. Natl. Acad. Sci.* **2014**, *111*, 9905–9910.

59.   Huang, W.; Smaldino, P. J.; Zhang, Q.; Miller, L. D.; Cao, P.; Stadelman, K.; Wan, M.; Giri, B.; Lei, M.; Nagamine, Y.; Vaughn, J. P.; Akman, S. A.; Sui, G. *Nucleic Acids Res.* **2012**, *40*, 1033–1049.

60.   Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N. *Nat. Chem. Biol.* **2014**, *10*, 313–318.

61.   Mendoza, O.; Bourdoncle, A.; Boulé, J.-B.; Brosh, R. M.; Mergny, J.-L. *Nucleic Acids Res.* **2016**, *44*, 1989–2006.

62.   Lin, W.; Sampathi, S.; Dai, H.; Liu, C.; Zhou, M.; Hu, J.; Huang, Q.; Campbell, J.; Shin-Ya, K.; Zheng, L.; Chai, W.; Shen, B. *EMBO J.* **2013**, *32*, 1425–1439.

63.   London, T. B. C.; Barber, L. J.; Mosedale, G.; Kelly, G. P.; Balasubramanian, S.; Hickson, I. D.; Boulton, S. J.; Hiom, K. *J. Biol. Chem.* **2008**, *283*, 36132–36139.

64.   Vannier, J.-B.; Pavicic-Kaltenbrunner, V.; Petalcorin, M. I. R.; Ding, H.; Boulton, S. J. *Cell* **2012**, *149*, 795–806.

65.   Sun, H.; Karow, J. K.; Hickson, I. D.; Maizels, N. *J. Biol. Chem.* **1998**, *273*, 27587–27592.

66.   Liu, J.; Chen, C.; Xue, Y.; Hao, Y.; Tan, Z. *J. Am. Chem. Soc.* **2010**, *132*, 10521–10527.

67.   Creacy, S. D.; Routh, E. D.; Iwamoto, F.; Nagamine, Y.; Akman, S. A.; Vaughn, J. P. *J. Biol. Chem.* **2008**, *283*, 34626–34634.

68.   Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci.* **2002**, *99*, 11593–11598.

69.   Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589.

70.   Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. *Biochemistry (Mosc.)* **2006**, *45*, 7854–7860.

71.   De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. *Biochemistry (Mosc.)* **2005**, *44*, 16341–16350.

References

72.   Lombardo, C. M.; Welsh, S. J.; Strauss, S. J.; Dale, A. G.; Todd, A. K.; Nanjunda, R.; Wilson, W. D.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5984–5988.

73.   Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 5133–5144.

74.   Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. *Nucleic Acids Res.* **2013**, *41*, 10584–10592.

75.   Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2007**, *129*, 10220–10228.

76.   Cogoi, S.; Paramasivam, M.; Spolaore, B.; Xodo, L. E. *Nucleic Acids Res.* **2008**, *36*, 3765–3780.

77.   Membrino, A.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. *PLoS ONE* **2011**, *6*, e24421.

78.   Ohnmacht, S. A.; Micco, M.; Petrucci, V.; Todd, A. K.; Reszka, A. P.; Gunaratnam, M.; Carvalho, M. A.; Zloh, M.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5930–5935.

79.   Wei, D.; Todd, A. K.; Zloh, M.; Gunaratnam, M.; Parkinson, G. N.; Neidle, S. *J. Am. Chem. Soc.* **2013**, *135*, 19319–19329.

80.   Mitchell, T.; Ramos-Montoya, A.; Di Antonio, M.; Murat, P.; Ohnmacht, S.; Micco, M.; Jurmeister, S.; Fryer, L.; Balasubramanian, S.; Neidle, S.; Neal, D. E. *Biochemistry (Mosc.)* **2013**, *52*, 1429–1436.

81.   Yan, J.; Zhao, X.; Liu, B.; Yuan, Y.; Guan, Y. *Mol. Carcinog.* **2016**, *55*, 897–909.

82.   Basu, S. *Nucleic Acids Res.* **1997**, *25*, 1327–1332.

83.   Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167–1172.

84.   Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. *Nucleic Acids Res.* **2011**, *39*, 9023–9033.

85.   Yang, D.; Hurley, L. H. *Nucleosides Nucleotides Nucleic Acids* **2006**, *25*, 951–968.

86.   Phan, A. T.; Kuryavyi, V.; Gaw, H. Y.; Patel, D. J. *Nat. Chem. Biol.* **2005**, *1*, 167–173.

87.   Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. *J. Am. Chem. Soc.* **2014**, *136*, 1750–1753.

88.   Capra, J. A.; Paeschke, K.; Singh, M.; Zakian, V. A. *PLoS Comput. Biol.* **2010**, *6*, e1000861.

89.   Holder, I. T.; Hartig, J. S. *Chem. Biol.* **2014**, *21*, 1511–1521.

90.   Rawal, P. *Genome Res.* **2006**, *16*, 644–655.

91.   Artusi, S.; Nadai, M.; Perrone, R.; Biasolo, M. A.; Palù, G.; Flamand, L.; Calistri, A.; Richter, S. N. *Antiviral Res.* **2015**, *118*, 123–131.

92.   Tlučková, K.; Marušič, M.; Tóthová, P.; Bauer, L.; Šket, P.; Plavec, J.; Viglasky, V. *Biochemistry (Mosc.)* **2013**, *52*, 7207–7216.

93.   Norseen, J.; Johnson, F. B.; Lieberman, P. M. *J. Virol.* **2009**, *83*, 10336–10346.

94.   Tan, J.; Vonrhein, C.; Smart, O. S.; Bricogne, G.; Bollati, M.; Kusov, Y.; Hansen, G.; Mesters, J. R.; Schmidt, C. L.; Hilgenfeld, R. *PLoS Pathog.* **2009**, *5*, e1000428.

95.   Perrone, R.; Nadai, M.; Poe, J. A.; Frasson, I.; Palumbo, M.; Palù, G.; Smithgall, T. E.; Richter, S. N. *PLoS ONE* **2013**, *8*, e73121.

96.   Amrane, S.; Kerkour, A.; Bedrat, A.; Vialet, B.; Andreola, M.-L.; Mergny, J.-L. *J. Am. Chem. Soc.* **2014**, *136*, 5249–5252.

97.   Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discov.* **2011**, *10*, 261–275.

98.   Lim, K. W.; Lacroix, L.; Yue, D. J. E.; Lim, J. K. C.; Lim, J. M. W.; Phan, A. T. *J. Am. Chem. Soc.* **2010**, *132*, 12331–12342.

99.   Ohnmacht, S. A.; Micco, M.; Petrucci, V.; Todd, A. K.; Reszka, A. P.; Gunaratnam, M.; Carvalho, M. A.; Zloh, M.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5930–5935.

100.   Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. *J. Med. Chem.* **1997**, *40*, 2113–2116.

101.   Hahn, W. C.; Stewart, S. A.; Brooks, M. W.; York, S. G.; Eaton, E.; Kurachi, A.; Beijersbergen, R. L.; Knoll, J. H.; Meyerson, M.; Weinberg, R. A. *Nat. Med.* **1999**, *5*, 1164–1170.

102.   Haider, S. M.; Autiero, I.; Neidle, S. *Biochimie* **2011**, *93*, 1275–1279.

103.   Sabharwal, N. C.; Savikhin, V.; Turek-Herman, J. R.; Nicoludis, J. M.; Szalai, V. A.; Yatsunyk, L. A. *FEBS J.* **2014**, *281*, 1726–1737.

104.   Franceschin, M. *Eur. J. Org. Chem.* **2009**, *2009*, 2225–2238.

105.   Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

106.   Huang, H.-S.; Chou, C.-L.; Guo, C.-L.; Yuan, C.-L.; Lu, Y.-C.; Shieh, F.-Y.; Lin, J.-J. *Bioorg. Med. Chem.* **2005**, *13*, 1435–1444.

107.   Seenisamy, J.; Bashyam, S.; Gokhale, V.; Vankayalapati, H.; Sun, D.; Siddiqui-Jain, A.; Streiner, N.; Shin-ya, K.; White, E.; Wilson, W. D.; Hurley, L. H. *J. Am. Chem. Soc.* **2005**, *127*, 2944–2959.

108.   Gomez, D.; Aouali, N.; Renaud, A.; Douarre, C.; Shin-Ya, K.; Tazi, J.; Martinez, S.; Trentesaux, C.; Morjani, H.; Riou, J.-F. *Cancer Res.* **2003**, *63*, 6149–6153.

109.   Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2463–2468.

110.   Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci.* **2001**, *98*, 4844–4849.

111.   Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. *Mol. Pharmacol.* **2002**, *61*, 1154–1162.

112.   Incles, C. M.; Schultes, C. M.; Kempski, H.; Koehler, H.; Kelland, L. R.; Neidle, S. *Mol. Cancer Ther.* **2004**, *3*, 1201–1206.

113.   Burger, A. M. *Cancer Res.* **2005**, *65*, 1489–1496.

References

114. Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. *J. Am. Chem. Soc.* **2008**, *130*, 6722–6724.

115. Perrone, R.; Butovskaya, E.; Daelemans, D.; Palu, G.; Pannecouque, C.; Richter, S. N. *J. Antimicrob. Chemother.* **2014**, *69*, 3248–3258.

116. Naasani, I.; Seimiya, H.; Yamori, T.; Tsuruo, T. *Cancer Res.* **1999**, *59*, 4004–4011.

117. Kim, M.-Y.; Vankayalapati, H.; Shin-ya, K.; Wierzba, K.; Hurley, L. H. *J. Am. Chem. Soc.* **2002**, *124*, 2098–2099.

118. Wu, H. L.; Hsu, C. Y.; Liu, W. H.; Yung, B. Y. M. *Int. J. Cancer* **1999**, *81*, 923–929.

119. Bazzicalupi, C.; Ferraroni, M.; Bilia, A. R.; Scheggi, F.; Gratteri, P. *Nucleic Acids Res.* **2013**, *41*, 632–638.

120. Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T. *J. Am. Chem. Soc.* **2013**, *135*, 13495–13501.

121. Shin-ya, K.; Wierzba, K.; Matsuo, K.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H. *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263.

122. Shammas, M. A. *Clin. Cancer Res.* **2004**, *10*, 770–776.

123. Binz, N.; Shalaby, T.; Rivera, P.; Shin-ya, K.; Grotzer, M. A. *Eur. J. Cancer* **2005**, *41*, 2873–2881.

124. Tauchi, T.; Shin-ya, K.; Sashida, G.; Sumi, M.; Okabe, S.; Ohyashiki, J. H.; Ohyashiki, K. *Oncogene* **2006**, *25*, 5719–5725.

125. Satyanarayana, M.; Kim, Y.-A.; Rzuczek, S. G.; Pilch, D. S.; Liu, A. A.; Liu, L. F.; Rice, J. E.; LaVoie, E. J. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3150–3154.

126. Rzuczek, S. G.; Pilch, D. S.; Liu, A.; Liu, L.; LaVoie, E. J.; Rice, J. E. *J. Med. Chem.* **2010**, *53*, 3632–3644.

127. Teulade-Fichou, M.-P.; Carrasco, C.; Guittat, L.; Bailly, C.; Alberti, P.; Mergny, J.-L.; David, A.; Lehn, J.-M.; Wilson, W. D. *J. Am. Chem. Soc.* **2003**, *125*, 4732–4740.

128. Jantos, K.; Rodriguez, R.; Ladame, S.; Shirude, P. S.; Balasubramanian, S. *J. Am. Chem. Soc.* **2006**, *128*, 13662–13663.

129. Shirude, P. S.; Gillies, E. R.; Ladame, S.; Godde, F.; Shin-ya, K.; Huc, I.; Balasubramanian, S. *J. Am. Chem. Soc.* **2007**, *129*, 11890–11891.

130. Tera, M.; Ishizuka, H.; Takagi, M.; Suganuma, M.; Shin-ya, K.; Nagasawa, K. *Angew. Chem. Int. Ed.* **2008**, *47*, 5557–5560.

131. Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. *Angew. Chem. Int. Ed.* **2014**, *53*, 999–1002.

132. Ohnmacht, S. A.; Varavipour, E.; Nanjunda, R.; Pazitna, I.; Di Vita, G.; Gunaratnam, M.; Kumar, A.; Ismail, M. A.; Boykin, D. W.; Wilson, W. D.; Neidle, S. *Chem Commun* **2014**, *50*, 960–963.

133. Hamon, F.; Largy, E.; Guédin-Beaurepaire, A.; Rouchon-Dagois, M.; Sidibe, A.; Monchaud, D.; Mergny, J.-L.; Riou, J.-F.; Nguyen, C.-H.; Teulade-Fichou, M.-P. *Angew. Chem. Int. Ed.* **2011**, *50,* 8745–8749.

134. Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759.

135. Jain, S. L.; Bhattacharyya, P.; Milton, H. L.; Slawin, A. M. Z.; Crayston, J. A.; Woollins, J. D. *Dalton Trans.* **2004**, 862.

136. Müller, S.; Sanders, D. A.; Di Antonio, M.; Matsis, S.; Riou, J.-F.; Rodriguez, R.; Balasubramanian, S. *Org. Biomol. Chem.* **2012**, *10*, 6537.

137. Han, H.; Cliff, C. L.; Hurley, L. H. *Biochemistry (Mosc.)* **1999**, *38*, 6981–6986.

138. Sissi, C.; Lucatello, L.; Paul Krapcho, A.; Maloney, D. J.; Boxer, M. B.; Camarasa, M. V.; Pezzoni, G.; Menta, E.; Palumbo, M. *Bioorg. Med. Chem.* **2007**, *15*, 555–562.

139. Rossetti, L.; Franceschin, M.; Bianco, A.; Ortaggi, G.; Savino, M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2527–2533.

140. Rossetti, L.; Franceschin, M.; Schirripa, S.; Bianco, A.; Ortaggi, G.; Savino, M. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 413–420.

141. Franceschin, M.; Pascucci, E.; Alvino, A.; D'Ambrosio, D.; Bianco, A.; Ortaggi, G.; Savino, M. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2515–2522.

142. Franceschin, M.; Alvino, A.; Casagrande, V.; Mauriello, C.; Pascucci, E.; Savino, M.; Ortaggi, G.; Bianco, A. *Bioorg. Med. Chem.* **2007**, *15*, 1848–1858.

143. Di Antonio, M.; Doria, F.; Richter, S. N.; Bertipaglia, C.; Mella, M.; Sissi, C.; Palumbo, M.; Freccero, M. *J. Am. Chem. Soc.* **2009**, *131*, 13132–13141.

144. Peduto, A.; Pagano, B.; Petronzi, C.; Massa, A.; Esposito, V.; Virgilio, A.; Paduano, F.; Trapasso, F.; Fiorito, F.; Florio, S.; Giancola, C.; Galeone, A.; Filosa, R. *Bioorg. Med. Chem.* **2011**, *19*, 6419–6429.

145. Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. *J. Am. Chem. Soc.* **2012**, *134*, 2723–2731.

146. Micco, M.; Collie, G. W.; Dale, A. G.; Ohnmacht, S. A.; Pazitna, I.; Gunaratnam, M.; Reszka, A. P.; Neidle, S. *J. Med. Chem.* **2013**, *56*, 2959–2974.

147. Monchaud, D.; Yang, P.; Lacroix, L.; Teulade-Fichou, M.-P.; Mergny, J.-L. *Angew. Chem. Int. Ed.* **2008**, *47*, 4858–4861.

148. Reed, J. E.; Neidle, S.; Vilar, R. *Chem. Commun.* **2007**, 4366.

149. Wheelhouse, R. T.; Sun, D.; Han, H.; Han, F. X.; Hurley, L. H. *J. Am. Chem. Soc.* **1998**, *120*, 3261–3262.

References

150.	Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S. *J. Med. Chem.* **2008**, *51*, 5641–5649.

151.	Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L.-S.; Johnson, F. B. *Nucleic Acids Res.* **2007**, *36*, 144–156.

152.	Duan, W.; Rangan, A.; Vankayalapati, H.; Kim, M. Y.; Zeng, Q.; Sun, D.; Han, H.; Fedoroff, O. Y.; Nishioka, D.; Rha, S. Y.; Izbicka, E.; Von Hoff, D. D.; Hurley, L. H. *Mol. Cancer Ther.* **2001**, *1*, 103–120.

153.	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. *Cancer Res.* **2009**, *69*, 7653–7661.

154.	Martino, L.; Virno, A.; Pagano, B.; Virgilio, A.; Di Micco, S.; Galeone, A.; Giancola, C.; Bifulco, G.; Mayol, L.; Randazzo, A. *J. Am. Chem. Soc.* **2007**, *129*, 16048–16056.

155.	Kaiser, M.; De Cian, A.; Sainlos, M.; Renner, C.; Mergny, J.-L.; Teulade-Fichou, M.-P. *Org. Biomol. Chem.* **2006**, *4*, 1049–1057.

156.	Rahman, K. M.; Tizkova, K.; Reszka, A. P.; Neidle, S.; Thurston, D. E. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 3006–3010.

157.	Pagano, B.; Amato, J.; Iaccarino, N.; Cingolani, C.; Zizza, P.; Biroccio, A.; Novellino, E.; Randazzo, A. *ChemMedChem* **2015**, *10*, 640–649.

158.	Mergny, J. L.; Maurizot, J. C. *Chembiochem Eur. J. Chem. Biol.* **2001**, *2*, 124–132.

159.	Decian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teuladefichou, M.; Lacroix, L.; Mergny, J. *Methods* **2007**, *42*, 183–195.

160.	De Rache, A.; Mergny, J.-L. *Biochimie* **2015**, *115*, 194–202.

161.	Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. *Angew. Chem. Int. Ed.* **2011**, *50*, 10645–10648.

162.	Randazzo, A.; Spada, G. P.; da Silva, M. W. In *Quadruplex Nucleic Acids*; Chaires, J. B.; Graves, D., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, **2012**; Vol. 330, pp. 67–86.

163.	Cousins, A. R. O.; Ritson, D.; Sharma, P.; Stevens, M. F. G.; Moses, J. E.; Searle, M. S. *Chem Commun* **2014**, *50*, 15202–15205.

164.	Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74–78.

165.	Mergny, J.-L.; Lacroix, L. In *Current Protocols in Nucleic Acid Chemistry*; Beaucage, S. L.; Bergstrom, D. E.; Herdewijn, P.; Matsuda, A., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, **2009**; p. 17.1.1-17.1.15.

166.	Mergny, J.-L. *Nucleic Acids Res.* **2005**, *33*, e138–e138.

167.	Cross, A. D.; Crothers, D. M. *Biochemistry (Mosc.)* **1971**, *10*, 4015–4023.

168.   Feigon, J.; Koshlap, K. M.; Smith, F. W. *Methods Enzymol.* **1995**, *261*, 225–255.

169.   Patel, D. J.; Tonelli, A. E. *Biopolymers* **1974**, *13*, 1943–1964.

170.   Adrian, M.; Heddi, B.; Phan, A. T. *Methods* **2012**, *57*, 11–24.

171.   Randazzo, A.; Galeone, A.; Esposito, V.; Varra, M.; Mayol, L. *Nucleosides Nucleotides Nucleic Acids* **2002**, *21*, 535–545.

172.   Chung, W. J.; Heddi, B.; Hamon, F.; Teulade-Fichou, M.-P.; Phan, A. T. *Angew. Chem. Int. Ed.* **2014**, *53*, 999–1002.

173.   Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T. *J. Am. Chem. Soc.* **2013**, *135*, 13495–13501.

174.   Dai, J.; Carver, M.; Hurley, L. H.; Yang, D. *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.

175.   Li, H.; Liu, Y.; Lin, S.; Yuan, G. *Chem. - Eur. J.* **2009**, *15*, 2445–2452.

176.   Nagesh, N.; Chatterji, D. *J. Biochem. Biophys. Methods* **1995**, *30*, 1–8.

177.   Rosu, F.; De Pauw, E.; Gabelica, V. *Biochimie* **2008**, *90*, 1074–1087.

178.   Lombardo, C. M.; Martínez, I. S.; Haider, S.; Gabelica, V.; De Pauw, E.; Moses, J. E.; Neidle, S. *Chem. Commun.* **2010**, *46*, 9116.

179.   Gabelica, V.; Shammel Baker, E.; Teulade-Fichou, M.-P.; De Pauw, E.; Bowers, M. T. *J. Am. Chem. Soc.* **2007**, *129*, 895–904.

180.   Evans, S. E.; Mendez, M. A.; Turner, K. B.; Keating, L. R.; Grimes, R. T.; Melchoir, S.; Szalai, V. A. *JBIC J. Biol. Inorg. Chem.* **2007**, *12*, 1235–1249.

181.   Linder, J.; Garner, T. P.; Williams, H. E. L.; Searle, M. S.; Moody, C. J. *J. Am. Chem. Soc.* **2011**, *133*, 1044–1051.

182.   Cheng, M.-K.; Modi, C.; Cookson, J. C.; Hutchinson, I.; Heald, R. A.; McCarroll, A. J.; Missailidis, S.; Tanious, F.; Wilson, W. D.; Mergny, J.-L.; Laughton, C. A.; Stevens, M. F. G. *J. Med. Chem.* **2008**, *51*, 963–975.

183.   Mueckler, M.; Thorens, B. *Mol. Aspects Med.* **2013**, *34*, 121–138.

184.   Barron, C. C.; Bilan, P. J.; Tsakiridis, T.; Tsiani, E. *Metabolism* **2016**, *65*, 124–139.

185.   Warburg, O.; Wind, F.; Negelein, E. *J. Gen. Physiol.* **1927**, *8*, 519–530.

186.   Gillies, R. J.; Robey, I.; Gatenby, R. A. *J. Nucl. Med.* **2008**, *49*, 24S–42S.

187.   Vander Heiden, M. G.; Cantley, L. C.; Thompson, C. B. *Science* **2009**, *324*, 1029–1033.

188.   DeBerardinis, R. J.; Lum, J. J.; Hatzivassiliou, G.; Thompson, C. B. *Cell Metab.* **2008**, *7*, 11–20.

189.   Shim, H.; Dolde, C.; Lewis, B. C.; Wu, C. S.; Dang, G.; Jungmann, R. A.; Dalla-Favera, R.; Dang, C. V. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6658–6663.

190.   Zhang, H.; Gao, P.; Fukuda, R.; Kumar, G.; Krishnamachary, B.; Zeller, K. I.; Dang, C. V.; Semenza, G. L. *Cancer Cell* **2007**, *11*, 407–420.

References

191. Kim, J. -w.; Gao, P.; Liu, Y.-C.; Semenza, G. L.; Dang, C. V. *Mol. Cell. Biol.* **2007**, *27*, 7381–7393.

192. Bensaad, K.; Tsuruta, A.; Selak, M. A.; Vidal, M. N. C.; Nakano, K.; Bartrons, R.; Gottlieb, E.; Vousden, K. H. *Cell* **2006**, *126*, 107–120.

193. Selak, M. A.; Armour, S. M.; MacKenzie, E. D.; Boulahbel, H.; Watson, D. G.; Mansfield, K. D.; Pan, Y.; Simon, M. C.; Thompson, C. B.; Gottlieb, E. *Cancer Cell* **2005**, *7*, 77–85.

194. Christofk, H. R.; Vander Heiden, M. G.; Harris, M. H.; Ramanathan, A.; Gerszten, R. E.; Wei, R.; Fleming, M. D.; Schreiber, S. L.; Cantley, L. C. *Nature* **2008**, *452*, 230–233.

195. Szablewski, L. *Biochim. Biophys. Acta BBA - Rev. Cancer* **2013**, *1835*, 164–169.

196. Helmke, B. M.; Reisser, C.; Idzko, M.; Dyckhoff, G.; Herold-Mende, C.; Idzkoe, M. *Oral Oncol.* **2004**, *40*, 28–35.

197. Ishikawa, N.; Oguri, T.; Isobe, T.; Fujitaka, K.; Kohno, N. *Jpn. J. Cancer Res. Gann* **2001**, *92*, 874–879.

198. Medina, R. A.; Owen, G. I. *Biol. Res.* **2002**, *35*, 9–26.

199. Jans, J.; van Dijk, J. H.; van Schelven, S.; van der Groep, P.; Willems, S. H.; Jonges, T. N.; van Diest, P. J.; Bosch, J. L. H. R. *Urology* **2010**, *75*, 786–792.

200. Noguchi, Y.; Marat, D.; Saito, A.; Yoshikawa, T.; Doi, C.; Fukuzawa, K.; Tsuburaya, A.; Satoh, S.; Ito, T. *Hepatogastroenterology.* **1999**, *46*, 2683–2689.

201. Godoy, A.; Ulloa, V.; Rodríguez, F.; Reinicke, K.; Yañez, A. J.; García, M. de los A.; Medina, R. A.; Carrasco, M.; Barberis, S.; Castro, T.; Martínez, F.; Koch, X.; Vera, J. C.; Poblete, M. T.; Figueroa, C. D.; Peruzzo, B.; Pérez, F.; Nualart, F. *J. Cell. Physiol.* **2006**, *207*, 614–627.

202. Taova, T. P.; Semicheva, T. V. *Pediatriia* **1987**, 99–101.

203. Airley, R. E.; Mobasheri, A. *Chemotherapy* **2007**, *53*, 233–256.

204. Carvalho, K. C.; Cunha, I. W.; Rocha, R. M.; Ayala, F. R.; Cajaíba, M. M.; Begnami, M. D.; Vilela, R. S.; Paiva, G. R.; Andrade, R. G.; Soares, F. A. *Clin. Sao Paulo Braz.* **2011**, *66*, 965–972.

205. Hirose, Y.; Kaida, H.; Ishibashi, M.; Kawahara, A.; Kobayashi, M.; Hayabuchi, N. *Jpn. J. Radiol.* **2011**, *29*, 217–221.

206. Gowrishankar, G.; Zitzmann-Kolbe, S.; Junutula, A.; Reeves, R.; Levi, J.; Srinivasan, A.; Bruus-Jensen, K.; Cyr, J.; Dinkelborg, L.; Gambhir, S. S. *PLoS ONE* **2011**, *6*, e26902.

207. Zamora-León, S. P.; Golde, D. W.; Concha, I. I.; Rivas, C. I.; Delgado-López, F.; Baselga, J.; Nualart, F.; Vera, J. C. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 1847–1852.

208. Chandler, J. D.; Williams, E. D.; Slavin, J. L.; Best, J. D.; Rogers, S. *Cancer* **2003**, *97*, 2035–2042.

209. Pohl, J.; Bertram, B.; Hilgard, P.; Nowrousian, M. R.; Stüben, J.; Wießler, M. *Cancer Chemother. Pharmacol.* **1995**, *35*, 364–370.

210.   Liu, D.-Z.; Sinchaikul, S.; Reddy, P. V. G.; Chang, M.-Y.; Chen, S.-T. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 617–620.

211.   Cao, J.; Cui, S.; Li, S.; Du, C.; Tian, J.; Wan, S.; Qian, Z.; Gu, Y.; Chen, W. R.; Wang, G. *Cancer Res.* **2013**, *73*, 1362–1373.

212.   Liu, P.; Lu, Y.; Gao, X.; Liu, R.; Zhang-Negrerie, D.; Shi, Y.; Wang, Y.; Wang, S.; Gao, Q. *Chem. Commun.* **2013**, *49*, 2421.

213.   Patra, M.; Johnstone, T. C.; Suntharalingam, K.; Lippard, S. J. *Angew. Chem. Int. Ed.* **2016**, *55*, 2550–2554.

214.   Patra, M.; Awuah, S. G.; Lippard, S. J. *J. Am. Chem. Soc.* **2016**, *138*, 12541–12551.

215.   Calvaresi, E. C.; Granchi, C.; Tuccinardi, T.; Di Bussolo, V.; Huigens, R. W.; Lee, H. Y.; Palchaudhuri, R.; Macchia, M.; Martinelli, A.; Minutolo, F.; Hergenrother, P. J. *ChemBioChem* **2013**, *14*, 2263–2267.

216.   Lucas, R.; Gómez-Pinto, I.; Aviñó, A.; Reina, J. J.; Eritja, R.; González, C.; Morales, J. C. *J. Am. Chem. Soc.* **2011**, *133*, 1909–1916.

217.   Gómez-Pinto, I.; Vengut-Climent, E.; Lucas, R.; Aviñó, A.; Eritja, R.; González, C.; Morales, J. C. *Chem. - Eur. J.* **2013**, *19*, 1920–1927.

218.   Bugaut, A.; Jantos, K.; Wietor, J.-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. *Angew. Chem. Int. Ed.* **2008**, *47*, 2677–2680.

219.   Zhang, J. -l.; Fu, Y.; Zheng, L.; Li, W.; Li, H.; Sun, Q.; Xiao, Y.; Geng, F. *Nucleic Acids Res.* **2009**, *37*, 2471–2482.

220.   Ranjan, N.; Andreasen, K. F.; Kumar, S.; Hyde-Volpe, D.; Arya, D. P. *Biochemistry (Mosc.)* **2010**, *49*, 9891–9903.

221.   Xue, L.; Ranjan, N.; Arya, D. P. *Biochemistry (Mosc.)* **2011**, *50*, 2838–2849.

222.   Ranjan, N.; Davis, E.; Xue, L.; Arya, D. P. *Chem. Commun.* **2013**, *49*, 5796.

223.   Bai, L.-P.; Ho, H.-M.; Ma, D.-L.; Yang, H.; Fu, W.-C.; Jiang, Z.-H. *PLoS ONE* **2013**, *8*, e53962.

224.   Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E.-A.; Rodriguez, R.; Balasubramanian, S. *Angew. Chem. Int. Ed.* **2012**, *51*, 11073–11078.

225.   Duskova, K.; Sierra, S.; Arias-Pérez, M.-S.; Gude, L. *Bioorg. Med. Chem.* **2016**, *24*, 33–41.

226.   Moorhouse, A. D.; Santos, A. M.; Gunaratnam, M.; Moore, M.; Neidle, S.; Moses, J. E. *J. Am. Chem. Soc.* **2006**, *128*, 15972–15973.

227.   Moorhouse, A. D.; Haider, S.; Gunaratnam, M.; Munnur, D.; Neidle, S.; Moses, J. E. *Mol. Biosyst.* **2008**, *4*, 629.

228.   Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.

References

229.	Gunaratnam, M.; Green, C.; Moreira, J. B.; Moorhouse, A. D.; Kelland, L. R.; Moses, J. E.; Neidle, S. *Biochem. Pharmacol.* **2009**, *78*, 115–122.

230.	Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed Engl.* **2001**, *40*, 2004–2021.

231.	Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. *Org. Lett.* **2004**, *6*, 2853–2855.

232.	Guezguez, R.; Bougrin, K.; El Akri, K.; Benhida, R. *Tetrahedron Lett.* **2006**, *47*, 4807–4811.

233.	Rosevear, P.; VanAken, T.; Baxter, J.; Ferguson-Miller, S. *Biochemistry (Mosc.)* **1980**, *19*, 4108–4115.

234.	Adesoye, O. G.; Mills, I. N.; Temelkoff, D. P.; Jackson, J. A.; Norris, P. *J. Chem. Educ.* **2012**, *89*, 943–945.

235.	Micheel, F.; van de Kamp, F.-P.; Petersen, H. *Chem. Ber.* **1957**, *90*, 521–526.

236.	Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G. *Angew. Chem. Int. Ed.* **2009**, *48*, 7798–7802.

237.	Soli, E. D.; Manoso, A. S.; Patterson, M. C.; DeShong, P.; Favor, D. A.; Hirschmann, R.; Smith, A. B. *J. Org. Chem.* **1999**, *64*, 3171–3177.

238.	Dahmén, J.; Frejd, T.; Grönberg, G.; Lave, T.; Magnusson, G.; Noori, G. *Carbohydr. Res.* **1983**, *116*, 303–307.

239.	Arce, E.; Nieto, P. M.; Díaz, V.; Castro, R. G.; Bernad, A.; Rojo, J. *Bioconjug. Chem.* **2003**, *14*, 817–823.

240.	Schmidt, R. R. *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 212–235.

241.	Excoffier, G.; Gagnaire, D.; Utille, J.-P. *Carbohydr. Res.* **1975**, *39*, 368–373.

242.	Chernyak, A. Y.; Sharma, G. V. M.; Kononov, L. O.; Krishna, P. R.; Levinsky, A. B.; Kochetkov, N. K.; Rama Rao, A. V. *Carbohydr. Res.* **1992**, *223*, 303–309.

243.	Tran, A.-T.; Burden, R.; Racys, D. T.; Carmen Galan, M. *Chem. Commun.* **2011**, *47*, 4526.

244.	Galan, M. C.; Tran, A. T.; Bromfield, K.; Rabbani, S.; Ernst, B. *Org. Biomol. Chem.* **2012**, *10*, 7091.

245.	Galan, M. C.; Jones, R. A.; Tran, A.-T. *Carbohydr. Res.* **2013**, *375*, 35–46.

246.	Maslov, M. A.; Morozova, N. G.; Chizhik, E. I.; Rapoport, D. A.; Ryabchikova, E. I.; Zenkova, M. A.; Serebrennikova, G. A. *Carbohydr. Res.* **2010**, *345*, 2438–2449.

247.	Galan, M. C.; Tran, A. T.; Bernard, C. *Chem. Commun.* **2010**, *46*, 8968.

248.	Meldal, M.; Tornøe, C. W. *Chem. Rev.* **2008**, *108*, 2952–3015.

249.	Liebert, T.; Hänsch, C.; Heinze, T. *Macromol. Rapid Commun.* **2006**, *27*, 208–213.

250.	Wilkinson, B. L.; Bornaghi, L. F.; Houston, T. A.; Innocenti, A.; Supuran, C. T.; Poulsen, S.-A. *J. Med. Chem.* **2006**, *49*, 6539–6548.

251.   Wilkinson, B. L.; Bornaghi, L. F.; Poulsen, S.-A.; Houston, T. A. *Tetrahedron* **2006**, *62*, 8115–8125.

252.   Fernandez-Megia, E.; Correa, J.; Rodríguez-Meizoso, I.; Riguera, R. *Macromolecules* **2006**, *39*, 2113–2120.

253.   Wang, S.-M.; Chen, J.-T. *J. Chin. Chem. Soc.* **1975**, *22*, 179–189.

254.   Zemple´n, G.; Pacsu, E. *Ber. Dtsch. Chem. Ges* **1929**, *62*, 1613–1614.

255.   Liu, Y.-H.; Zhang, L.; Xu, X.-N.; Li, Z.-M.; Zhang, D.-W.; Zhao, X.; Li, Z.-T. *Org. Chem. Front.* **2014**, *1*, 494.

256.   Ko, E.; Liu, J.; Perez, L. M.; Lu, G.; Schaefer, A.; Burgess, K. *J. Am. Chem. Soc.* **2011**, *133*, 462–477.

257.   Juríček, M.; Felici, M.; Contreras-Carballada, P.; Lauko, J.; Bou, S. R.; Kouwer, P. H. J.; Brouwer, A. M.; Rowan, A. E. *J Mater Chem* **2011**, *21*, 2104–2111.

258.   Sharma, A.; Soliman, G. M.; Al-Hajaj, N.; Sharma, R.; Maysinger, D.; Kakkar, A. *Biomacromolecules* **2012**, *13*, 239–252.

259.   Garner, T. P.; Williams, H. E. L.; Gluszyk, K. I.; Roe, S.; Oldham, N. J.; Stevens, M. F. G.; Moses, J. E.; Searle, M. S. *Org. Biomol. Chem.* **2009**, *7*, 4194.

260.   Kerkour, A.; Mergny, J.-L.; Salgado, G. F. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2016**.

261.   Mosmann, T. *J. Immunol. Methods* **1983**, *65*, 55–63.

262.   Thomas, G. B.; Rader, L. H.; Park, J.; Abezgauz, L.; Danino, D.; DeShong, P.; English, D. S. *J. Am. Chem. Soc.* **2009**, *131*, 5471–5477.

263.   Gyorgydeak, Z.; Szilagyi, L. *LIEBIGS ANN. CHEM.* **1985**, *1985*, 103–112.

264.   Holder, A. A.; Wootton, J. C.; Baron, A. J.; Chambers, G. K.; Fincham, J. R. *Biochem. J.* **1975**, *149*, 757–773.

265.   Miyake, H.; Otsuka, C.; Nishimura, S.; Nitta, Y. *J. Biochem. (Tokyo)* **2002**, *131*, 587–591.

266.   Dowlut, M.; Hall, D. G.; Hindsgaul, O. *J. Org. Chem.* **2005**, *70*, 9809–9813.

267.   Khan, R.; Konowicz, P.; Gardossi, L.; Matulova, M.; Degennaro, S. *Aust. J. Chem.* **1996**, *49*, 293.

268.   Quagliotto, P.; Viscardi, G.; Barolo, C.; D'Angelo, D.; Barni, E.; Compari, C.; Duce, E.; Fisicaro, E. *J. Org. Chem.* **2005**, *70*, 9857–9866.

269.   Tagmose, T. M.; Bols, M. *Chem. - Eur. J.* **2006**, *3*, 453–462.

270.   Street, S. T. G.; Chin, D. N.; Hollingworth, G. J.; Berry, M.; Morales, J. C.; Galan, M. C. *Chem. - Eur. J.* **2017**.

271.   Nguyen, B. D.; Meng, X.; Donovan, K. J.; Shaka, A. J. *J. Magn. Reson.* **2007**, *184*, 263–274.

272.   Sessler, J. L.; Brown, C. T.; O'Connor, D.; Springs, S. L.; Wang, R.; Sathiosatham, M.; Hirose, T. *J. Org. Chem.* **1998**, *63*, 7370–7374.

References

273. Horne, W. S.; Ashkenasy, N.; Ghadiri, M. R. *Chem. - Eur. J.* **2005**, *11*, 1137–1144.

274. Cuenca, F.; Greciano, O.; Gunaratnam, M.; Haider, S.; Munnur, D.; Nanjunda, R.; Wilson, W. D.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1668–1673.

275. Bhosale, S. V.; Jani, C. H.; Langford, S. J. *Chem Soc Rev* **2008**, *37*, 331–342.

276. Erten, S.; Posokhov, Y.; Alp, S.; Icli, S. *Dyes Pigments* **2005**, *64*, 171–178.

277. Doria, F.; Manet, I.; Grande, V.; Monti, S.; Freccero, M. *J. Org. Chem.* **2013**, *78*, 8065–8073.

278. Andric, G.; Boas, J. F.; Bond, A. M.; Fallon, G. D.; Ghiggino, K. P.; Hogan, C. F.; Hutchison, J. A.; Lee, M. A.-P.; Langford, S. J.; Pilbrow, J. R.; Troup, G. J.; Woodward, C. P. *Aust. J. Chem.* **2004**, *57*, 1011.

279. Würthner, F.; Ahmed, S.; Thalacker, C.; Debaerdemaeker, T. *Chem. - Eur. J.* **2002**, *8*, 4742–4750.

280. Fin, A.; Petkova, I.; Doval, D. A.; Sakai, N.; Vauthey, E.; Matile, S. *Org. Biomol. Chem.* **2011**, *9*, 8246.

281. Doria, F.; Gallati, C. M.; Freccero, M. *Org. Biomol. Chem.* **2013**, *11*, 7838.

282. Takenaka, S.; Yamashita, K.; Takagi, M.; Uto, Y.; Kondo, H. *Anal. Chem.* **2000**, *72*, 1334–1341.

283. Langford, S. J.; Latter, M. J.; Woodward, C. P. *Photochem. Photobiol.* **2006**, *82*, 1530.

284. Bhosale, S. *Science* **2006**, *313*, 84–86.

285. Tanaka, H.; Litvinchuk, S.; Tran, D.-H.; Bollot, G.; Mareda, J.; Sakai, N.; Matile, S. *J. Am. Chem. Soc.* **2006**, *128*, 16000–16001.

286. Gorteau, V.; Bollot, G.; Mareda, J.; Perez-Velasco, A.; Matile, S. *J. Am. Chem. Soc.* **2006**, *128*, 14788–14789.

287. Talukdar, P.; Bollot, G.; Mareda, J.; Sakai, N.; Matile, S. *Chem. - Eur. J.* **2005**, *11*, 6525–6532.

288. Doria, F.; Oppi, A.; Manoli, F.; Botti, S.; Kandoth, N.; Grande, V.; Manet, I.; Freccero, M. *Chem Commun* **2015**, *51*, 9105–9108.

289. Zuffo, M.; Doria, F.; Spalluto, V.; Ladame, S.; Freccero, M. *Chem. - Eur. J.* **2015**, *21*, 17596–17600.

290. Sissi, C.; Lucatello, L.; Paul Krapcho, A.; Maloney, D. J.; Boxer, M. B.; Camarasa, M. V.; Pezzoni, G.; Menta, E.; Palumbo, M. *Bioorg. Med. Chem.* **2007**, *15*, 555–562.

291. Milelli, A.; Tumiatti, V.; Micco, M.; Rosini, M.; Zuccari, G.; Raffaghello, L.; Bianchi, G.; Pistoia, V.; Fernando Díaz, J.; Pera, B.; Trigili, C.; Barasoain, I.; Musetti, C.; Toniolo, M.; Sissi, C.; Alcaro, S.; Moraca, F.; Zini, M.; Stefanelli, C.; Minarini, A. *Eur. J. Med. Chem.* **2012**, *57*, 417–428.

292. Hampel, S. M.; Sidibe, A.; Gunaratnam, M.; Riou, J.-F.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6459–6463.

293. Nadai, M.; Doria, F.; Di Antonio, M.; Sattin, G.; Germani, L.; Percivalle, C.; Palumbo, M.; Richter, S. N.; Freccero, M. *Biochimie* **2011**, *93*, 1328–1340.

294.   Doria, F.; Nadai, M.; Folini, M.; Di Antonio, M.; Germani, L.; Percivalle, C.; Sissi, C.; Zaffaroni, N.; Alcaro, S.; Artese, A.; Richter, S. N.; Freccero, M. *Org. Biomol. Chem.* **2012**, *10*, 2798.

295.   Doria, F.; Nadai, M.; Folini, M.; Scalabrin, M.; Germani, L.; Sattin, G.; Mella, M.; Palumbo, M.; Zaffaroni, N.; Fabris, D.; Freccero, M.; Richter, S. N. *Chem. - Eur. J.* **2013**, *19*, 78–81.

296.   Esaki, Y.; Islam, M. M.; Fujii, S.; Sato, S.; Takenaka, S. *Chem. Commun.* **2014**, *50*, 5967.

297.   Hampel, S. M.; Pepe, A.; Greulich-Bode, K. M.; Malhotra, S. V.; Reszka, A. P.; Veith, S.; Boukamp, P.; Neidle, S. *Mol. Pharmacol.* **2013**, *83*, 470–480.

298.   Mpima, S.; Ohnmacht, S. A.; Barletta, M.; Husby, J.; Pett, L. C.; Gunaratnam, M.; Hilton, S. T.; Neidle, S. *Bioorg. Med. Chem.* **2013**, *21*, 6162–6170.

299.   Doria, F.; Nadai, M.; Sattin, G.; Pasotti, L.; Richter, S. N.; Freccero, M. *Org. Biomol. Chem.* **2012**, *10*, 3830.

300.   Gunaratnam, M.; Fuente, M. de la; Hampel, S. M.; Todd, A. K.; Reszka, A. P.; Schätzlein, A.; Neidle, S. *Bioorg. Med. Chem.* **2011**, *19*, 7151–7157.

301.   Ohnmacht, S. A.; Marchetti, C.; Gunaratnam, M.; Besser, R. J.; Haider, S. M.; Di Vita, G.; Lowe, H. L.; Mellinas-Gomez, M.; Diocou, S.; Robson, M.; Šponer, J.; Islam, B.; Barbara Pedley, R.; Hartley, J. A.; Neidle, S. *Sci. Rep.* **2015**, *5*, 11385.

302.   Lopergolo, A.; Perrone, R.; Tortoreto, M.; Doria, F.; Beretta, G. L.; Zuco, V.; Freccero, M.; Grazia Borrello, M.; Lanzi, C.; Ricther, S. N.; Zaffaroni, N.; Folini, M. *Oncotarget* **2014**.

303.   Arévalo-Ruiz, M.; Doria, F.; Belmonte-Reche, E.; De Rache, A.; Campos-Salinas, J.; Lucas, R.; Falomir, E.; Carda, M.; Pérez-Victoria, J. M.; Mergny, J.-L.; Freccero, M.; Morales, J. C. *Chem. - Eur. J.* **2017**, *23*, 2157–2164.

304.   Evans, M. E. *Methods in Carbohydrate Chemistry* **1980**, *8*, 313–315.

305.   Cicero, D.; Varela, O.; de Lederkremer, R. M. *Carbohydr. Res.* **1991**, *211*, 295–308.

306.   Liu, Y.; Cao, Y.; Zhang, W.; Bergmeier, S.; Qian, Y.; Akbar, H.; Colvin, R.; Ding, J.; Tong, L.; Wu, S.; Hines, J.; Chen, X. *Mol. Cancer Ther.* **2012**, *11*, 1672–1682.

307.   Vyas, A. K.; Koster, J. C.; Tzekov, A.; Hruz, P. W. *J. Biol. Chem.* **2010**, *285*, 36395–36400.

308.   Strobel, P.; Allard, C.; Perez-Acle, T.; Calderon, R.; Aldunate, R.; Leighton, F. *Biochem. J.* **2005**, *386*, 471–478.

309.   OECD *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*; OECD Guidelines for the Testing of Chemicals, Section 2; OECD Publishing, **2013**.

310.   Ali, S.; Mil, H. G. J. van; Richardson, M. K. *PLoS ONE* **2011**, *6*, e21076.

311.   Zercher, C. K.; Fedor, L. R. *Carbohydr. Res.* **1987**, *165*, 299–305.

312.   Sirion, U.; Kim, H. J.; Lee, J. H.; Seo, J. W.; Lee, B. S.; Lee, S. J.; Oh, S. J.; Chi, D. Y. *Tetrahedron Lett.* **2007**, *48*, 3953–3957.

References

313. Tanaka, T.; Nagai, H.; Noguchi, M.; Kobayashi, A.; Shoda, S. *Chem. Commun.* **2009**, 3378.

314. Sun, X.-L.; Haller, C. A.; Wu; Conticello, V. P.; Chaikof, E. L. *J. Proteome Res.* **2005**, *4*, 2355–2359.

315. Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olson, B. J.; Klenk, D. C. *Anal. Biochem.* **1985**, *150*, 76–85.

316. Berkowitz, R. D.; Alexander, S.; Bare, C.; Linquist-Stepps, V.; Bogan, M.; Moreno, M. E.; Gibson, L.; Wieder, E. D.; Kosek, J.; Stoddart, C. A.; McCune, J. M. *J. Virol.* **1998**, *72*, 10108–10117.

317. Foley, B.; Leitner, T.; Apetrei, C.; Hahn, B.; Mizrachi, I.; Mullins, J.; Rambaut, A.; Korber, B.; Wolinsky, S. HIV Sequence Compendium 2016. *Los Alamos Natl. Lab. Theor. Biol. Biophys.* **2016**.

318. Haseltine, W. A. *J. Acquir. Immune Defic. Syndr.* **1988**, *1*, 217–240.

319. Cann, A. J.; Karn, J. *AIDS Lond. Engl.* **1989**, *3 Suppl 1*, S19-34.

320. Terwilliger, E. F.; Sodroski, J. G.; Haseltine, W. A. *Ann. Emerg. Med.* **1990**, *19*, 233–241.

321. Mori, T.; Boyd, M. R. *Antimicrob. Agents Chemother.* **2001**, *45*, 664–672.

322. Mitsuya, H.; Looney, D. J.; Kuno, S.; Ueno, R.; Wong-Staal, F.; Broder, S. *Science* **1988**, *240*, 646–649.

323. Lederman, M. M. *Science* **2004**, *306*, 485–487.

324. Yarovinsky, F.; Andersen, J. F.; King, L. R.; Caspar, P.; Aliberti, J.; Golding, H.; Sher, A. *J. Biol. Chem.* **2004**, *279*, 53635–53642.

325. Dragic, T. *J. Gen. Virol.* **2001**, *82*, 1807–1814.

326. Eckert, D. M.; Kim, P. S. *Annu. Rev. Biochem.* **2001**, *70*, 777–810.

327. Furman, P. A.; Fyfe, J. A.; St Clair, M. H.; Weinhold, K.; Rideout, J. L.; Freeman, G. A.; Lehrman, S. N.; Bolognesi, D. P.; Broder, S.; Mitsuya, H. *Proc. Natl. Acad. Sci.* **1986**, *83*, 8333–8337.

328. Tomaras, G. D.; Greenberg, M. L. *Curr. Infect. Dis. Rep.* **2001**, *3*, 93–99.

329. Joshi, P.; Prasad, V. R. *J. Virol.* **2002**, *76*, 6545–6557.

330. Osborne, S. E.; Matsumura, I.; Ellington, A. D. *Curr. Opin. Chem. Biol.* **1997**, *1*, 5–9.

331. Zhu, K.; Cordeiro, M. L.; Atienza, J.; Robinson Jr., W. E.; Chow, S. A. *J Virol.* **1999**, *73*, 3309–3316.

332. Nair, V. *Rev. Med. Virol.* **2002**, *12*, 179–193.

333. Gottfredsson, M.; Bohjanen, P. R. *Front. Biosci.* **1997**, *2*, 619–634.

334. Bunnell, B. A.; Morgan, R. A. *Clin Microbiol Rev.* **1998**, *11*, 42–56.

335. Ding, S. *Nucleic Acids Res.* **1998**, *26*, 3270–3278.

336. Hwang, S.; Tamilarasu, N.; Kibler, K.; Cao, H.; Ali, A.; Ping, Y.-H.; Jeang, K.-T.; Rana, T. M. *J. Biol. Chem.* **2003**, *278*, 39092–39103.

337. Coburn, G. A.; Cullen, B. R. *J. Virol.* **2002**, *76*, 9225–9231.

338.   Novina, C. D.; Murray, M. F.; Dykxhoorn, D. M.; Beresford, P. J.; Riess, J.; Lee, S.-K.; Collman, R. G.; Lieberman, J.; Shankar, P.; Sharp, P. A. *Nat. Med.* **2002**.

339.   Kjems, J.; Askjaer, P. *Adv. Pharmacol. San Diego Calif* **2000**, *48*, 251–298.

340.   Lamothe, B.; Joshi, S. *Front. Biosci. J. Virtual Libr.* **2000**, *5*, D527-555.

341.   Kibler, K. V.; Miyazato, A.; Yedavalli, V. S. R. K.; Dayton, A. I.; Jacobs, B. L.; Dapolito, G.; Kim, S. -j.; Jeang, K.-T. *J. Biol. Chem.* **2004**, *279*, 49055–49063.

342.   Dewhurst, S.; Whetter, L. *Front. Biosci. J. Virtual Libr.* **1997**, *2*, d147-159.

343.   Fung, H. B.; Kirschenbaum, H. L.; Hameed, R. *Clin. Ther.* **2000**, *22*, 549–572.

344.   Sundquist, W. .; Heaphy, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3393–3397.

345.   Lyonnais, S. *Nucleic Acids Res.* **2003**, *31*, 5754–5763.

346.   Richter, S.; Frasson, I.; Palu, G. *Curr. Med. Chem.* **2009**, *16*, 267–286.

347.   Feng, S.; Holland, E. C. *Nature* **1988**, *334*, 165–167.

348.   Perrone, R.; Doria, F.; Butovskaya, E.; Frasson, I.; Botti, S.; Scalabrin, M.; Lago, S.; Grande, V.; Nadai, M.; Freccero, M.; Richter, S. N. *J. Med. Chem.* **2015**, *58*, 9639–9652.