
**SYNTHESIS AND STRUCTURAL
STUDIES ON DNA
G-QUADRUPLEXES: FROM
NUCLEIC ACID APTAMERS TO
HIGHER ORDER ASSEMBLIES**

Valentina D'Atri

Dottorato in Scienze Biotechologiche – XXV ciclo
Indirizzo Biotechologie Industriali e Molecolari
Università di Napoli Federico II



Dottorato in Scienze Biotechologiche – XXV ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II



**SYNTHESIS AND STRUCTURAL
STUDIES ON DNA
G-QUADRUPLEXES: FROM NUCLEIC
ACID APTAMERS TO HIGHER ORDER
ASSEMBLIES**

Valentina D'Atri

Dottoranda: Valentina D'Atri

Relatore: Prof. Gennaro Piccialli

Coordinatore: Prof. Giovanni Sannia

A chi mi è stato sempre accanto

INDEX

ABSTRACTS

Short abstract	pag. i
Abstract	pag. ii

CHAPTER 1

“The G-quadruplex structures: general introduction and main applications”

Introduction	pag. 1
1.1 Structural variability of G-quadruplex structures	pag. 1
1.1.1 Topological classification of G-quadruplex structures	pag. 2
1.1.2 G-quadruplex DNA groove widths	pag. 2
1.1.3 Metal ion coordination	pag. 5
1.1.4 Beyond G-tetrads, novel tetrad pairing alignment	pag. 5
1.2 Biological relevance of G-quadruplex structures	pag. 8
1.3 G-quadruplex forming Aptamers	pag. 9
1.4 G-quadruplex structures in Nanotechnology	pag. 10
1.4.1 G-quadruplexes-based Nanostructures	pag. 10
1.4.2 G-quadruplexes-based Biosensors and Nanodevices	pag. 11
1.5 Aim of my research work	pag. 12
1.6 References	pag. 13

CHAPTER 2

“New anti-HIV aptamers based on tetra-end-linked DNA G-quadruplexes”

Introduction	pag. 19
2.1 Results and Discussion	pag. 20
2.2 Conclusions	pag. 28
2.3 Experimental Session	pag. 28
2.3.1 General instrumentations	pag. 28
2.3.2 Synthesis of 3'-O-tert-butyl-diphenylsilyl-propyl-1'-O-(2-cyanoethyl)-N,N'-diisopropyl-phosphoramidite	pag. 28
2.3.3 Synthesis and MALDI-TOF spectra of TEL-ODNs	pag. 29
2.3.4 Preparation of quadruple helices (annealing)	pag. 33
2.3.5 CD spectroscopy measurements	pag. 33
2.3.6 NMR spectroscopy measurements	pag. 33
2.3.7 Biological evaluation assays	pag. 33
2.3.8 Surface Plasmon Resonance (SPR) analysis	pag. 33
2.3.9 Docking protocol	pag. 34

2.4. References	pag. 35
------------------------	---------

CHAPTER 3

“Thrombin-Binding Aptamer analogs with enhanced anticoagulant activity: the role of T₇ and T₁₂ residues on biological properties”

Introduction	pag. 39
---------------------	---------

3.1 Results	pag. 42
--------------------	---------

3.1.1 <i>Synthesis of Monomer 6.</i>	pag. 42
--------------------------------------	---------

3.1.2 <i>Structural Characterization</i>	pag. 43
--	---------

3.1.3 <i>PT Assay</i>	pag. 44
-----------------------	---------

3.1.4 <i>Fibrinogen Assay Using Human and Bovine Thrombin</i>	pag. 45
---	---------

3.1.5 <i>Conformational Search on TBA and Modified Analogues TBA-T_{7b} and TBA-T_{12b}</i>	pag. 47
---	---------

3.1.6 <i>Molecular Modeling Studies on Aptamer Interaction with Human and Bovine Thrombin</i>	pag. 47
---	---------

3.2 Discussion	pag. 53
-----------------------	---------

3.3 Conclusions	pag. 55
------------------------	---------

3.4 Experimental Session	pag. 55
---------------------------------	---------

3.4.1 <i>General Procedure</i>	pag. 55
--------------------------------	---------

3.4.2 <i>Synthesis of Monomer 6</i>	pag. 55
-------------------------------------	---------

3.4.3 <i>Synthesis of Oligomers</i>	pag. 57
-------------------------------------	---------

3.4.4 <i>NMR Experiments</i>	pag. 57
------------------------------	---------

3.4.5 <i>CD Experiments</i>	pag. 57
-----------------------------	---------

3.4.6 <i>Prothrombin (PT) Time</i>	pag. 57
------------------------------------	---------

3.4.7 <i>Purified Fibrinogen Clotting Assay</i>	pag. 58
---	---------

3.4.8 <i>Molecular Modeling</i>	pag. 58
---------------------------------	---------

3.4.8.1 <i>Conformational Analysis</i>	pag. 58
--	---------

3.4.8.2 <i>Structural and Bioinformatics Analysis</i>	pag. 59
---	---------

3.5 References	pag. 60
-----------------------	---------

CHAPTER 4

“DNA-based Nanostructures: dimers of G-quadruplexes via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads”

Introduction	pag. 65
---------------------	---------

4.1 Results	pag. 66
--------------------	---------

4.1.1 <i>dCGGTGGT forms an octamer</i>	pag. 66
--	---------

4.1.2 <i>The octamer contains stacked G-tetrads</i>	pag. 67
---	---------

4.1.3 <i>Kinetics of octamer formation studied by ESI-MS</i>	pag. 68
--	---------

4.1.4 <i>Influence of sample preparation procedure</i>	pag. 69
--	---------

4.1.5 <i>NMR investigation of the octamer structure</i>	pag. 70
---	---------

4.1.6 <i>Investigating the role of the central base on the structuring of the octamer</i>	pag. 73
---	---------

4.2 Discussion	pag. 77
4.3 Conclusions	pag. 79
4.4 Experimental Session	pag. 79
4.4.1 <i>DNA sample preparation</i>	pag. 79
4.4.2 <i>Polyacrylamide gel electrophoresis</i>	pag. 80
4.4.3 <i>Electrospray mass spectrometry</i>	pag. 80
4.4.4 <i>Circular dichroism</i>	pag. 81
4.4.5 <i>Nuclear magnetic resonance</i>	pag. 81
4.4.6 <i>Molecular modelling</i>	pag. 81
4.5 References	pag. 82
Appendix	pag. 87
<i>List of Publications and Communications</i>	pag. 87

ACKNOWLEDGEMENTS

I think that every PhD student requires a reasonably challenging topic to work on, good guidance from a supervisor, and a stimulating work environment. In my case, I was fortunate to have received all of the above. First of all, I would like to express my deepest appreciation to my supervisor, Professor Gennaro Piccialli, for his continued support and infinite patience throughout my entire research project. He was the best supervisor I could hope for. Secondly, I would like to thank my NMR mentor, Dr Nicola Borbone. I am grateful for his efforts, and wealth of knowledge, which helped me to acquire extensive skills in the field of Nuclear Magnetic Resonance, and made this experience both educational and enjoyable. Next, I would like to thank my colleagues, who deserve a special mention – Dr Giorgia Oliviero, Dr Jussara Amato, and Dr Stefano D’Errico. They continuously provided answers to my questions, ranging from the field of organic chemistry to the practical applications of analytical chemistry, and they strongly improved the quality of my knowledge and know-how. Thank you. I doubt I will ever work anywhere quite as lively as in your company. During my PhD I also had the opportunity to be hosted for one year at the Centre of Cancer Research and Cell Biology of the Queen’s University of Belfast, where I learned the basics of molecular docking and dynamic simulations. In this context I would like to thank Dr Shozeb Haider and Dr Miriam Sgobba, who not only gave me the excellent opportunity to learn new software, and improve my skills learning from their experiences, but also provided an enhancement of my research project. Finally, I would like to take this opportunity to thank the Department of Natural Compounds Chemistry, in particular Luisa Cuorvo, who has always been very kind and helped me to apply for funding to attend courses and conferences; and the Biotechnological Sciences PhD, that provided me with the financial support to undertake this course.

Sincerely,
Valentina D’Atri

Short Abstract

Quadruple helices, or G-quadruplexes, are DNA secondary structures found in guanine rich oligonucleotide sequences, having a natural propensity to self-associate in coplanar arrays of four guanines, stabilized by Hoogsteen hydrogen bonding. The scientific interest towards these particular DNA structures is mainly due to the presence of guanine rich domains, potentially able to form G-quadruplexes, in important regions of the human genome, as gene promoters and telomeres, and to the fact that the G-quadruplexes can constitute the scaffold of aptamers. Aptamers are short DNA or RNA fragments capable to bind with high affinity specific proteins, as for example thrombin or HIV-proteins. On these grounds, aptamers-based synthetic oligonucleotides can represent a new class of pharmacologically interesting molecules, characterized by a high selectivity of action. Furthermore, the G-quadruplexes can have a potential use in nanotechnology. As a matter of fact, the overall quadruplex scaffold can exhibit several morphologies through intramolecular or intermolecular organization of G-rich oligonucleotide strands, which can form higher-order assemblies by multimerization between G-quadruplex units.

In these contexts, my research studies have been mainly focused on:

1. Investigation of the most prominent biological properties of selected G-quadruplexes and identification of bioactive sequences able to play a biological role as aptamers.
2. Elucidation of the sequence-specific thermodynamic stability and physical-chemical mechanisms that underlie the G-quadruplexes multimerization.

Specifically, during my PhD research studies, I focused on anti-HIV aptamers, and Thrombin Binding Aptamer (TBA) with the aim to identify the structural features required for their biological activity and reaching the way to improve them.

Furthermore, I identify the CGGXGGT oligonucleotide sequence (where X = A, C, G, or T) able to lead to the formation of higher order G-quadruplex assemblies, that can potentially be used as DNA-based nanostructures.

In order to achieve these objectives, it has been necessary to exploit the synthesis and the structural characterization of quadruplex-forming oligonucleotides, the analysis of the sequence-specific thermodynamic stability, the physical-chemical properties and the structural features of resulting G-quadruplexes. Furthermore, as regard the studies about aptamers, the evaluation of the binding properties to the selected proteins, and bioassays to ascertain the biological properties of the synthesized complexes have been undertaken.

Long Abstract

Le G-quadruplex sono strutture secondarie di acido deossiribonucleico che si formano in sequenze ricche di guanine. Tali guanine hanno una naturale propensione ad auto associarsi in quartetti complanari (o tetradi), e sono in grado di stabilizzarsi mediante legami a idrogeno di tipo Hoogsteen (1). Dalla prima identificazione delle G-quadruplex, avvenuta nel 1962, sono ormai passati cinquant'anni. Nonostante questo, l'interesse scientifico verso le G-quadruplex non si è mai esaurito ma è anzi cresciuto in maniera esponenziale, arrivando al suo apice proprio nei mesi scorsi, quando il gruppo di ricerca del prof. Balasubramanian ha pubblicato un articolo in *Nature Chemistry*, in cui ha descritto un metodo per la visualizzazione quantitativa *in vivo* delle G-quadruplex e fornito quindi prove sostanziali e dirette della loro naturale formazione all'interno del genoma umano (2). A parte questo, molto si è detto e molto si è scritto in questi anni sulle G-quadruplex. È ormai dimostrato che queste strutture siano coinvolte in diversi processi biologici, ed è per questo motivo che sono diventate il principale interesse scientifico di molti gruppi di ricerca specializzati in differenti campi: dalla biologia molecolare a quella cellulare, dalla fisica alla chimica.

La presenza di domini ricchi di guanine, potenzialmente in grado di strutturarsi in G-quadruplex, sono presenti in importanti regioni del genoma umano, come ad esempio la porzione a singolo filamento dei telomeri e diversi promotori genici, suggerendo pertanto un possibile ruolo delle strutture G-quadruplex nella regolazione dei processi replicativi e trascrizionali. Quello che però ha maggiormente stimolato la fantasia dei ricercatori, è stata la possibilità di andare ad interferire con i suddetti processi, al fine di ottenere delle strategie antitumorali valide da spendersi nell'inibizione dello sviluppo o della proliferazione delle cellule neoplastiche. Di seguito, e in breve, la descrizione delle due principali e potenziali strategie antitumorali che vedono il diretto coinvolgimento delle strutture G-quadruplex.

Nelle cellule somatiche umane, la lunghezza dei telomeri diminuisce ad ogni evento di divisione cellulare, mentre l'inversione di questa degradazione, da parte di un enzima specializzato chiamato telomerasi, incrementa la capacità replicativa cellulare, determinando una proliferazione incontrollata; nella maggior parte delle cellule tumorali (85-90%) questo enzima è over-espresso (3). Poiché domini ricchi di guanine in grado di strutturarsi in G-quadruplex sono presenti a livello telomerico, e poiché la formazione delle G-quadruplex inibisce direttamente l'elongazione *in vitro* da parte della telomerasi, ligandi che possano selettivamente riconoscere e stabilizzare le strutture G-quadruplex possono interferire con l'elongazione dei telomeri, agendo come inibitori della telomerasi, e quindi come potenziali agenti antitumorali.

Sulla base degli stessi principi, è riportato che le sequenze nucleotidiche ricche di guanine facenti parte dei promotori sono spesso in equilibrio tra una strutturazione a doppio filamento che favorisce la trascrizione, e una forma ripiegata a quadrupla elica, che invece ne inibisce la trascrizione (4). Quando a dover essere trascritto non è un gene ma un oncogene, in altre parole un gene che codifica per proteine che indirizzano la cellula verso lo sviluppo di un fenotipo neoplastico, la possibilità di poter bloccare il processo trascrizionale mediante la stabilizzazione della strutturazione a quadrupla elica, diventa estremamente interessante dal punto di vista farmacologico. Per questo motivo, negli ultimi anni si è amplificata anche la ricerca di specifici ligandi capaci di legarsi alle sequenze promotrici di importanti oncogeni, quali *c-myc* (5-9), *c-kit* (10, 11), and *bcl-2* (12, 13); tali ligandi, in grado di

riconoscere e stabilizzare le strutture G-quadruplex, determinano l'inibizione della trascrizione degli oncogeni, e potrebbero quindi essere utilizzati come agenti antitumorali.

Oltre a quanto accennato, ci sono altri due campi di applicazione che vedono come protagoniste le G-quadruplex: oligonucleotidi sintetici formanti G-quadruplex e nanostrutture a base di G-quadruplex, entrambe le categorie dalle applicazioni estremamente stimolanti.

L'interesse per gli oligonucleotidi sintetici formanti G-quadruplex deriva principalmente dal fatto che essi possano costituire l'impalcatura degli aptameri. Gli aptameri sono piccole molecole di DNA o RNA a singolo filamento, che assumono specifiche strutturazioni tridimensionali per raggiungere un'altissima affinità di legame nei confronti di proteine specifiche, come ad esempio la trombina o alcune proteine dell'HIV, dimostrando così promettenti proprietà biologiche e possibili applicazioni farmacologiche (14). Una delle classi di aptameri a base G-quadruplex particolarmente studiata è quella degli aptameri leganti la trombina (Thrombin-Binding Aptamer, TBA), capaci di legare una proteina fondamentale coinvolta nel processo di coagulazione, l' α -trombina, prevenendo così il processo patologico della trombosi (15, 16). È stato inoltre dimostrato che in questo caso la strutturazione in G-quadruplex e la sua conseguente stabilità, sono requisiti essenziali all'esplicazione degli effetti biofisici e biologici del TBA (17).

Un altro aptamero a base G-quadruplex è AS1411 (Antisoma, London, UK), un nucleotide 26-mer potenzialmente in grado di inibire la proliferazione delle cellule tumorali, che ha come target specifico la nucleolina, una proteina over-espressa sulla superficie delle cellule tumorali. Purtroppo la sperimentazione clinica di AS1411 in pazienti con leucemia mieloide acuta (AML) è stata bloccata in fase IIb a causa del tracollo finanziario dell'azienda produttrice. [fonte: <http://www.fiercebiotech.com> the biotech industry's daily monitor]

Altra grande famiglia è rappresentata dagli aptameri a base G-quadruplex sviluppati come potenziali farmaci anti-HIV, in grado di inibire le proteine essenziali del virus, coinvolte in diversi stadi del suo ciclo vitale. La ricerca in questo senso ha dato luogo a una vasta gamma di aptameri, principalmente diretti contro:

- la glicoproteina di superficie virale coinvolta nel processo di ancoraggio e fusione del virus, nota come gp120.
- l'enzima trascrittasi inversa (Reverse Transcriptase, RT) coinvolto nel processo di trascrizione inversa, quindi in grado di copiare l'RNA virale in un cDNA a doppio filamento previa integrazione nel genoma della cellula ospite.
- l'enzima integrasi (Integrase, IN) coinvolto nel processo di integrazione del DNA virale all'interno del DNA della cellula ospite.

Tra i vari aptameri a base G-quadruplex ad attività anti-HIV proposti e studiati in questi anni, i più efficienti si sono dimostrati l'oligonucleotide 17-mer GTGGTGGGTGGGTGGGT, anche noto come Zintevir, AR177 o T30177, così come due oligonucleotidi 16-mer (T30695 e 93del) e la sequenza più breve dell'oligonucleotide fosforotioato TTGGGGTT (ISIS 5320) (18). Particolare attenzione è stata richiamata dalla sequenza 6-mer di Hotoda TGGGAG, modificata in 5' mediante l'aggiunta di un gruppo dimetossitrietile (19). A partire da questa modificazione chimica, molte altre ne sono seguite con lo scopo di aumentare l'attività dell'aptamero, fino ad arrivare all'R-95288 caratterizzato dalla presenza di un gruppo 3,4-dibenzilossibenzil all'estremità 5' e un gruppo 2-idrossietilfosfato

all'estremità 3', che mostra l'attività biologica più potente e la citotossicità minore (20, 21). Dal momento che la stabilità strutturale delle G-quadruplex formanti aptameri è strettamente correlata all'attività biologica dell'aptamero, diversi approcci sintetici sono stati proposti per migliorarne le caratteristiche chimico-fisiche, sia mediante l'introduzione di larghi sostituenti aromatici (22), sia mediante l'ancoraggio dei filamenti stessi a linker alchilici in grado di veicolare la strutturazione in quadruplex, come proposto da Oliviero et al. (23, 24).

Per quanto riguarda invece il campo nanotecnologico, le G-quadruplex presentano diverse proprietà che le rendono particolarmente interessanti. Tenendo presente che il fiorente campo delle nanotecnologie si basa sulla costruzione controllata di molecole su scala nanometrica o micrometrica, le G-quadruplex sono risultate le concorrenti ideali a questo scopo, grazie alla loro capacità di auto-assemblarsi in maniera ordinata e controllabile (25, 26). In quest'ambito le G-quadruplex sono state utilizzate sia per la costruzione di nanostrutture a base di DNA, quali G-wires, Frayed Wires e Synapses, sia per la costruzione di biosensori e nano-dispositivi. Nello specifico, i G-Wires sono dei polimeri di G-quadruplex identificati per la prima volta da Marsh ed Henderson. Originano dalla disposizione parallela di filamenti aventi sequenza d(G₄T₂G₄), sono straordinariamente stabili ad alte temperature e hanno una lunghezza che può variare dai 10 ai 1000 nm (27). I Frayed Wires sono stati invece individuati da McGregor e collaboratori, si originano dalla sequenza d(G₁₅A₁₅), in cui le guanine vanno a costituire il nucleo principale della struttura, mentre le adenine si irradiano come bracci al di fuori della struttura (28). Sen e collaboratori hanno invece descritto un metodo per costruire nanostrutture a partire dalla connessione (synapsis, da qui il nome Synapses) tra due duplex di DNA (29). In questo caso, ciascun duplex contiene un "dominio sinaptico" costituito da coppie di guanine ingegnerizzate per facilitare la formazione delle G-tetrad e quindi portare alla dimerizzazione dei duplex. Per quanto riguarda invece i biosensori e le nano-macchine a base di G-quadruplex, il principio di base è che gli aptameri a base G-quadruplex, oltre ad essere utilizzati in campo terapeutico e diagnostico, possono avere anche un'applicazione pratica nel campo della chimica bioanalitica in qualità di biosensori. Difatti, data la loro altissima affinità di legame, gli aptameri a base G-quadruplex possono essere utilizzati come fase stazionaria in cromatografie ed elettroforesi per la separazione di piccole molecole e proteine. McGown e collaboratori, ad esempio, hanno utilizzato questo principio per discriminare tra gli enantiomeri D,L del Trp o della Tyr o per separare pool di proteine simili che si differenziano per pochi amminoacidi (30). Un approccio diverso vede invece l'utilizzo delle sonde a Trasferimento di Energia di Risonanza della Fluorescenza (FRET), un metodo distanza-dipendente utilizzato in quest'ambito per il rilevamento dei cambi conformazionali. Sono implicate due sonde, un fluoroforo donatore e un cromoforo accettore rispettivamente legate alle estremità 3' e 5' di un oligonucleotide; quando le due sonde si trovano vicine nello spazio, la fluorescenza è annullata, quando invece si allontanano, per effetto di un cambio conformazionale, avviene un corrispondente aumento della fluorescenza. Takenaka e collaboratori, ad esempio, hanno utilizzato questo metodo per monitorare il passaggio da singolo filamento a struttura G-quadruplex in base alla concentrazione di ioni K⁺ (31); analogamente, Li e Tan lo hanno utilizzato per monitorare in tempo reale i cambi conformazionali duplex-quadruplex, valutando la suscettibilità del cambio in base alla concentrazione di diversi cationi, o alla presenza di ligandi e proteine (32).

In questo contesto si è inserita la mia attività di ricerca e il mio lavoro di tesi si è focalizzato principalmente sull'identificazione e la caratterizzazione di:

- **Aptameri a base G-quadruplex dotati di elevata attività anti-HIV**

Obiettivo di questo lavoro di ricerca: realizzazione di una piccola libreria di aptameri a base G-quadruplex, al fine di valutare l'effetto della sequenza di basi sull'attività anti-HIV.

Nel 2010 Oliviero e collaboratori hanno proposto come nuovi aptameri anti-HIV a base G-quadruplex, degli oligonucleotidi di sequenza dTGGGAG, legati ad un linker tetraramificato (tetra-end-linked, TEL) (23). In questo lavoro, i TEL-aptameri sono stati preparati e analizzati per investigare l'influenza della posizione e della lunghezza del TEL, e la presenza di gruppi lipofili in posizione 5', sulle proprietà strutturali e biologiche. I risultati riportati hanno dimostrato che la presenza del TEL in posizione 3' e di un gruppo terz-butil-difenil-silil (TBDPS) in posizione 5' migliorano notevolmente la stabilità termica della G-quadruplex e la sua attività come aptamero anti-HIV.

Partendo da questi presupposti, è stata effettuata la sintesi di TEL-ODN, aventi sequenza dTGGGXG (dove X = G, C o T), coniugati con un gruppo terz-butil-difenil-silil (TBDPS) all'estremità 5' e al TEL all'estremità 3'. Lo scopo del lavoro è stato quello di sondare l'influenza dei diversi nucleotidi sulle proprietà strutturali e biologiche delle G-quadruplex risultanti, ed è stato accompagnato da studi di modellistica molecolare con lo scopo di valutare l'influenza dei diversi nucleotidi sulle interazioni tra gli aptameri e il loop V3 della proteina gp120 dell'HIV-1. I risultati ottenuti evidenziano che i nuovi aptameri hanno maggiore stabilità termica, maggiore velocità di strutturazione e valori di EC_{50} contro l'HIV-1 compresi in un intervallo di 0,04-0,15 μ M, nonché un'affinità per la proteina gp120 dell'HIV-1 dello stesso ordine di grandezza (33).

- **Analoghi dell'Aptamero Legante la Trombina (Thrombin-Binding Aptamer, TBA) dotati di aumentata attività anticoagulante**

Obiettivo di questo lavoro di ricerca: aumento dell'attività anticoagulante del TBA mediante la modificazione di singole nucleobasi.

Un nucleotide modificato, contenente un ciclo di cinque membri fuso sull'anello pirimidinico, è stato introdotto in posizione 7 o 12 nell'oligonucleotide formante il TBA, di sequenza dGGTTGGTGTGGTTGG. La caratterizzazione strutturale degli aptameri risultanti, TBA-T₇b e TBA-T₁₂b, effettuata mediante spettroscopia NMR e CD, ha evidenziato la loro capacità di ripiegarsi nella tipica strutturazione del TBA, ovvero una G-quadruplex antiparallela di tipo sedia. Le temperature di fusione apparenti registrate al CD indicano che l'introduzione del residuo aciclico, principalmente in posizione 7, migliora la stabilità termica della risultante G-quadruplex rispetto al TBA. L'attività anticoagulante delle nuove molecole è stata poi valutata mediante saggio di tempo di protrombina (prothrombin time, PT), ed è risultato che l'analogo TBA-T₇b fosse più potente rispetto al TBA nel prolungare il tempo di coagulazione. D'altra parte, nel saggio di fibrinogeno purificato, l'attività inibitoria della trombina di entrambe le sequenze modificate è risultata essere inferiore rispetto a quella del TBA, usando enzima umano, ma superiore con enzima bovino. Inoltre, le relazioni struttura-attività sono state indagate mediante studi computazionali e, nel loro insieme, nell'interazione aptamero-proteina, i risultati ottenuti evidenziano il ruolo attivo dei residui T₇ e T₁₂ del TBA e la rilevanza di alcuni aminoacidi situati nell'esosito I anionico della trombina (34).

▪ **Nanostrutture a base di DNA, costituite da dimeri di G-quadruplex**

Obiettivo di questo lavoro: realizzazione di una piccola libreria di nanostrutture costituite da dimeri di G-quadruplex.

Mediante l'utilizzo di diverse tecniche analitiche, quali Risonanza Magnetica Nucleare (NMR), Spettrometria di Massa (MS), Dicroismo Circolare (CD), UV ed Elettroforesi su Gel di Poli-Acrilamide (PAGE) si è dimostrato che la sequenza dCGGXGGT (dove X = T, A, C o G) è in grado di strutturarsi in una nano-struttura di circa 4 nm, costituita da otto filamenti ben organizzati. Di fatto la formazione dell'ottamero avviene mediante un processo di dimerizzazione, che prevede l'impilamento 5'-5' di due subunità G-quadruplex tetramolecolari, mediante l'insolita formazione di un'ottade di tipo G(:C):G(:C):G(:C):(G:C) (35).

Riferimenti bibliografici

- (1) M. L. Bochman, K. P., and V. A. Zakian. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews* 13, 770-780.
- (2) G. Biffi, D. T., J. McCafferty, and S. Balasubramanian (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry* 5, 182-186.
- (3) J. L. Mergny, J. F. R., M. P. Teulade-Fichou, and E. Gilson (2002) Natural and pharmacological regulation of telomerase. . *Nucleic Acids Research* 30, 839-865.
- (4) S. Cogoi, a. L. E. X. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Research* 34, 2536-2549.
- (5) A. Siddiqui-Jain, C. L. G., D. J. Bearss, and L. H. Hurley. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences* 99, 11593-11598.
- (6) C. L. Grand, H. H., R. M. Munoz, S. Weitman, D. D. Von Hoff, L. H. Hurley, and D. J. Bearss. (2002) The cationic porphyrin TMPyP4 down-regulates c-myc and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Molecular Cancer Therapeutics* 8, 565-573.
- (7) J. Seenisamy, E. M. R., T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain, and L. H. Hurley. (2004) The dynamic character of the G-quadruplex element in the c-myc promoter and modification by TMPyP4. *Journal of the American Chemical Society* 126, 8702-8709.
- (8) T. Simonsson, P. P., and M. Kubista. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Research* 26, 1167-1172.
- (9) V. Gabelica, E. S. B., M. P. Teulade-Fichou, E. De Pauw, and M. T. Bowers. (2007) Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands. *Journal of the American Chemical Society* 129, 895-904.
- (10) A. T. Phan, V. K., S. Burge, S. Neidle, and D. J. Patel. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *Journal of the American Chemical Society* 129, 4386-4392.
- (11) S. Rankin, A. P. R., J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian, S. Neidle. (2005) Putative DNA quadruplex

- formation within the human c-kit oncogene. *Journal of the American Chemical Society* 127, 10584-10589.
- (12) J. Dai, T. S. D., D. Chen, M. Carver, A. Ambrus, R. A. Jones, and D. Yang. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *Journal of the American Chemical Society* 128, 1096-1098.
 - (13) T. S. Dexheimer, D. S., and L. H. Hurley. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *Journal of the American Chemical Society* 128, 5404-5415.
 - (14) D. M. Held, J. D. K., J. T. Patterson, D. G. Nickens, and D. H. Burke. (2006) HIV-1 inactivation by nucleic acid aptamers. *Frontiers in Bioscience* 11, 89-112.
 - (15) K. Y. Wang, S. M., R. G. Shea, S. Swaminathan, and P. H. Bolton. (1993) A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry* 32, 1899-1894.
 - (16) L. C. Bock, L. C. G., J. A. Latham, E. H. Vermaas, and J. J. Toole. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355, 564-566.
 - (17) B. Pagano, L. M., A. Randazzo, and C. Giancola. (2008) Stability and binding properties of a modified thrombin binding aptamer. *Biophysical Journal* 94, 562-569.
 - (18) S. H. Chou, K. H. C., and A. H. Wang. (2005) DNA aptamers as potential anti-HIV agents. *Trends in Biochemical Sciences* 30, 231-234.
 - (19) H. Hotoda, K. M., H. Furukawa, T. Nakamura, M. Kaneko, S. Kimura, and K. Shimada. (1994) Biologically Active Oligodeoxyribonucleotides-II: Structure Activity Relationships of anti-HIV-1 pentadecaceoxyribonucleotides bearing 5'-end-modifications. *Nucleosides, Nucleotides, and Nucleic Acids* 13, 1375-1395.
 - (20) M. Koizumi, R. K., H. Hotoda, K. Momota, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura, and K. Shimada. (1997) Biologically active oligodeoxyribonucleotides-IX: synthesis and anti-HIV-1 activity of hexadeoxyribonucleotides, TGGGAG, bearing 3'- and 5'-end-modification. *Bioorganic and Medicinal Chemistry* 5, 2235-2243.
 - (21) M. Koizumi, R. K., H. Hotoda, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura, and K. Shimada. (1998) Biologically Active Oligodeoxynucleotides. Part 11: The laeast phosphate-modification of quadruplex-forming hexadeoxyribonucleotide TGGGAG, bearing 3'- and 5'-end-modification, with anti-HIV-1 activity. *Bioorganic and Medicinal Chemistry* 6, 2469-2475.
 - (22) G. Di Fabio, J. D. O., M. Chiapparelli, B. Hoorelbeke, D. Montesarchio, J. Balzarini, and L. De Napoli. (2011) Discovery of novel anti-HIV active G-quadruplex-forming oligonucleotides. *Chemical Communications* 47, 2363-2365.
 - (23) G. Oliviero, J. A., N. Borbone, S. D'Errico, A. Galeone, L. Mayol, S. Haider, O. Olubiyi, B. Hoorelbeke, J. Balzarini, and G. Piccialli. (2010) Tetra-end-linked oligonucleotides forming DNA G-quadruplexes: a new class of aptamers showing anti-HIV activity. *Chemical Communications* 46, 8971-8973.

- (24) G. Oliviero, N. B., J. Amato, S. D'Errico, A. Galeone, G. Piccialli, M. Varra, and L. Mayol. (2009) Synthesis of quadruplex-forming tetra-end-linked oligonucleotides: effects of the linker size on quadruplex topology and stability. *Biopolymers* 91, 466-477.
- (25) Niemeyer, C. M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology meets materials science. *Angewandte Chemie International Edition* 40, 4128-4158.
- (26) Niemeyer, C. M. (2002) Nanomechanical Devices based on DNA. *Angewandte Chemie International Edition* 41, 3779-3783.
- (27) T. C. Marsh, E. H. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry* 33, 10718-10724.
- (28) E. Protozanova, a. R. B. J. M. (1996) Frayed wires: a thermally stable form of DNA with two distinct structural domains. *Biochemistry* 35, 16638-16645.
- (29) E. A. Venczel, a. D. S. (1996) Synapsable DNA. *Journal of Molecular Biology* 257, 219-224.
- (30) R. B. Kotia, L. L., and L. B. McGown. (2000) Separation of nontarget compounds by DNA aptamers. *Analytical Chemistry* 72, 827-831.
- (31) H. Ueyama, M. T., and S. Takenaka. (2002) A novel potassium sensing in aqueous media with a synthetic oligonucleotide derivative. Fluorescence resonance energy transfer associated with guanine quartet-potassium ion complex formation. *Journal of the American Chemical Society* 124, 14287-14287.
- (32) J. J. Li, a. W. T. (2002) A single DNA molecule nanomotor. *Nano Letters* 2, 315-318.
- (33) V. D'Atri, G. O., J. Amato, N. Borbone, S. D'Errico, L. Mayol, V. Piccialli, S. Haider, B. Hoorelbeke, J. Balzarini, and G. Piccialli. (2012) New anti-HIV aptamers based on tetra-end-linked DNA G-quadruplexes: effect of the base sequence on anti-HIV activity. *Chemical Communications* 48, 9516-9518.
- (34) N. Borbone, M. B., G. Oliviero, E. Morelli, J. Amato, V. D'Atri, S. D'Errico, V. Vellecco, G. Cirino, G. Piccialli, C. Fattorusso, M. Varra, L. Mayol, M. Persico, and M. Scutto. (2012) Investigating the role of T7 and T12 residues on the biological properties of thrombin-binding aptamer: enhancement of anticoagulant activity by a single nucleobase modification. *Journal of Medicinal Chemistry* 55, 10716-10728.
- (35) N. Borbone, J. A., G. Oliviero, V. D'Atri, V. Gabelica, E. De Paw, G. Piccialli, and L. Mayol. (2011) d(CG GTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Research* 39, 7848-7857.

Preface

My PhD thesis project involved the identification and the structural characterization of several biomolecules, such as DNA aptamers and higher order assemblies, whose core architectures are based on G-quadruplexes. I mainly applied the nuclear magnetic resonance (NMR) techniques to determine the high resolution structure and to understand the factors contributing to the stability of G-quadruplex complexes formed by guanine-rich oligonucleotides. The structural characterizations were often followed by molecular docking studies and/or biological assays done in collaboration with different European research groups. Results of my independent intellectual and practical contributions in my thesis research area will be discussed in the forthcoming chapters, and have been published in the papers listed below:

- *Chemical Communication*, (2012), 48, 9516-9518.
Identification of DNA G-quadruplexes forming aptamers endowed with anti-HIV activity. Notably, I performed molecular docking studies between synthetic DNA-based aptamers and the gp120 protein of HIV-1 virus, during the period that I spent at the Centre of Cancer Research and Cell Biology (CCRCB) of the Queen's University of Belfast, UK.
- *Journal of Medicinal Chemistry*, (2012), 55, 23, 10716-10728;
I characterized Thrombin Binding Aptamer analogues with enhanced anticoagulant activity.
- *Nucleic Acids Research*, (2011), 39, 17, 7848-7857;
I characterized a synthetic higher order G-quadruplex assembly that can potentially be used as a DNA-based nanostructure. I also presented the results of this research at the *Second training school on G-quadruplexes: self-assembled guanosine structures for molecular electronic devices* (oral communication & poster session), and at *Giornate Scientifiche: Polo delle Scienze e delle Tecnologie per la Vita* (oral communication).

Beyond my PhD thesis project, I was also involved in several scientific projects focused on the identification and the structural characterization of small organic molecules and modified nucleoside analogues. Furthermore, I achieved further publications in molecular modelling field, thanks to the scientific collaborations established during the period I spent at Centre of Cancer Research and Cell Biology.

The list of the papers published in these fields, which will not be discussed in the forthcoming chapters of the thesis, is reported below.

- *Proceedings of the National Academy of Sciences of the United States of America*, (2013), 110, 7, 2605-2610.
- Chapter in *Guanine Quartets: Structure and Application*. (2013), pages 194-211.
- *European Journal of Organic Chemistry*, (2012), 23, 4293-4305.
- *Journal of Chemical Crystallography*, (2012), 42, 4, 360-365.
- *Tetrahedron*, (2011), 67, 34, 6138-6144.
- *Bioorganic and Medicinal Chemistry Letters*, (2011), 21, 19, 5835-5838.

Chapter 1

The G-quadruplex structures: general introduction and main applications

Nucleic acids are highly flexible molecules that can adopt different conformational structures. While DNA is largely double helical and RNA is single stranded, guanine-rich sequences can exist in alternative structural forms known as G-quadruplexes. G-quadruplexes are unusual nucleic acid structures that can form within specific repetitive G-rich DNA or RNA sequences. They are characterized by the stacking of units known as G-tetrads, consisting of four guanines organized in a square planar arrangement. Each guanine is linked to the two adjacent ones by Hoogsteen hydrogen bonds, involving the amino and imino protons and the nitrogen N7 and oxygen O6 atoms (Figure 1). The stacking of G-tetrads is stabilized by monovalent cations that occupy the central cavities between the stacks, neutralizing the electrostatic repulsion of inwardly pointing guanine oxygens. (1)

More than 50 years have passed since the disclosure of the G-quadruplex motif. Despite this, the scientific interest towards G-quadruplexes not only has never ended, but it rather has increased exponentially, reaching its peak just a few months ago, when the research group led by Prof. Balasubramanian published an article in *Nature Chemistry* describing a method for the quantitative visualization of G-quadruplexes *in vivo*, thus making substantial and direct evidence of their natural formation in the human genome. (2) Several reports suggest that G-quadruplex structures may have important applications in several and different areas, ranging from medicinal chemistry to the nanotechnologies field, due to their possible formation in important G-rich regions of the human genome such as telomeres and gene promoters, and to the fact that they may be used as building blocks for novel DNA-based nanostructures, biosensors and nanodevices.

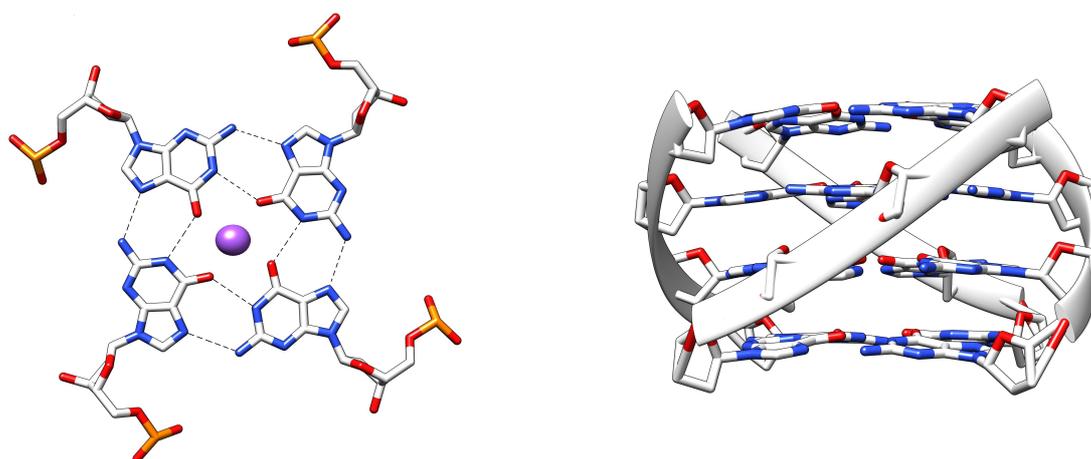


Figure 1. Graphical representation of a G-tetrad and a G-quadruplex.

1.1. Structural variability of G-quadruplex structures

Although the core of G-quadruplexes always respect the G-tetrad formation, the global scaffold is subjected to high structural variability, due to the molecularity and the orientation of the strands, the number and type of possible connecting loops, and glycosidic conformation of the sugar bases (which can be either *syn* or *anti* type).

1.1.1. Topological classification of G-quadruplex structures

The several G-quadruplex topologies can be classified into various groups depending on: i) the number of their strands, and thus classified into monomolecular (one strand), bimolecular (two strands) or tetramolecular (four strands); ii) the orientation of their strands, and thus they can be parallel, antiparallel or hybrids thereof; iii) the orientation of the connecting loops, if any, that can join adjacent strands (lateral loop), opposite strands (diagonal loop) or the bottom G-tetrad with the top G-tetrad (propeller loop) (Figures 2 and 3).

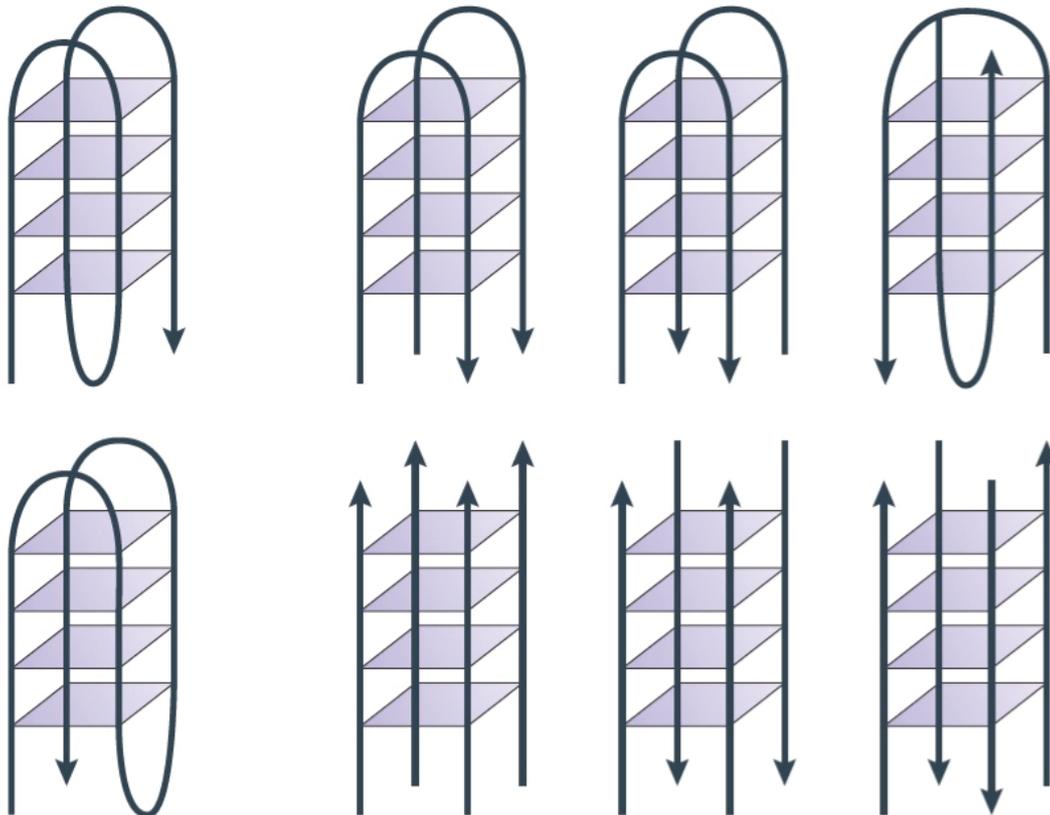


Figure 2. Representation of some possible topologies for monomolecular (left), bimolecular (upper right), and tetramolecular (lower right) G-quadruplex structures. The arrowheads indicate the direction of the strands.

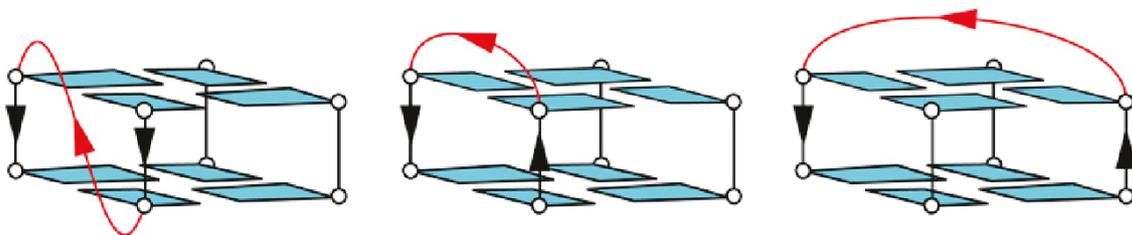


Figure 3. From left to right: propeller loop, lateral loop, and diagonal loop.

1.1.2. G-quadruplex DNA groove widths

The G-quadruplex structures are not stacked linearly, but adopt a right-handed helix exposing four grooves, defined as the cavities bounded by the phosphodiester

backbones. The variation of groove-widths along the side of the quadruplex is due to the spacing between the DNA strands of the G-quadruplex, which depends by the conformation of base χ glycosidic torsion angles (3). This dihedral angle linkage χ is defined as O4'-C1'-N9-C4 for purines and O4'-C1'-N1-C2 for pyrimidine bases (Figure 4a). Glycosidic torsion angles in duplex DNA/RNA fall into two tightly confined conformations. The related orientations adopted by the stacked bases have significant consequences for the overall shape and depth of the DNA/RNA grooves, altering the width, shape, and hydrogen-bonding pattern. Two of the most common conformations observed in folded DNA structures are *syn* and *anti*, shown in Figure 4b, where *syn* results from $0 < \chi < 90^\circ$ and *anti* from $-120 < \chi < 180^\circ$ (4).

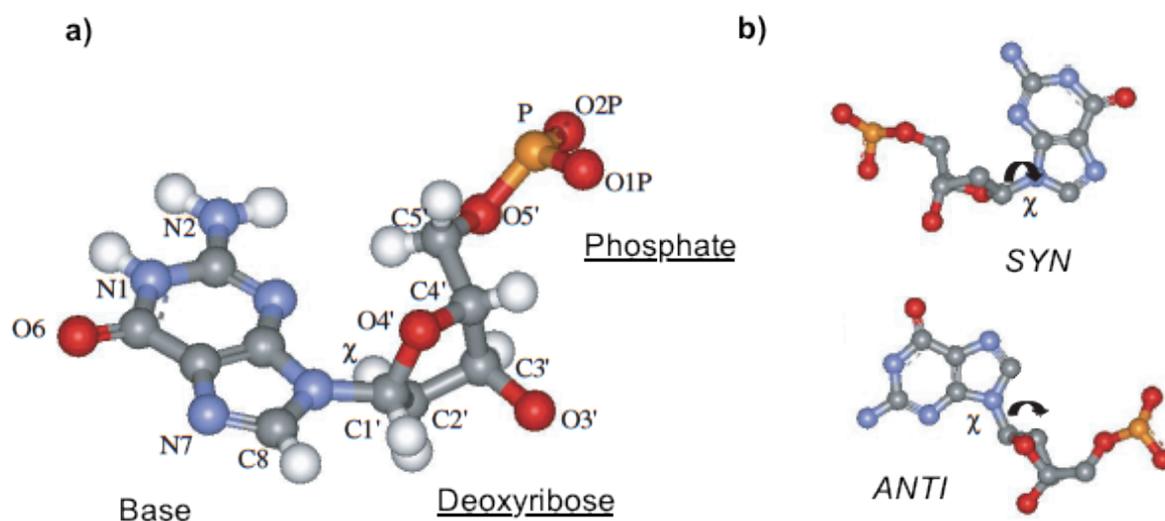


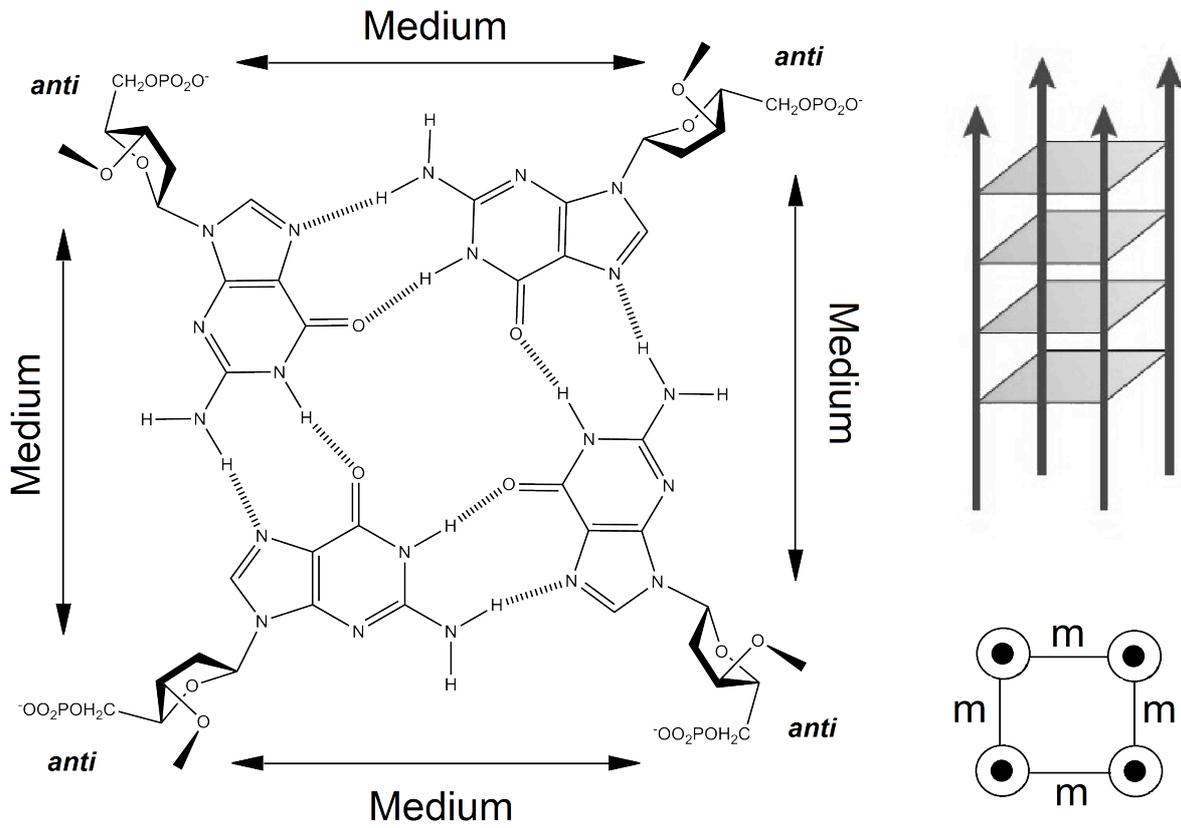
Figure 4. (a) A ball and stick representation of a guanine nucleotide highlighting the three building blocks of DNA, the base, sugar (deoxyribose), and the phosphate group. (b) Guanine nucleotides showing both *syn* and *anti* glycosidic torsion angles χ [O4'-C1'-N9-C4]. The *anti* conformation is the lower energy arrangement.

Notably, when a *syn* guanosine donates hydrogen bonds to a neighbouring *anti* guanosine, the groove formed between the two is extremely narrow; on the other hand, when the hydrogen bonding polarity between adjacent *syn* and *anti* guanosines is reversed, a very wide groove is formed (5). An intermediate width groove results when adjacent guanosines adopt the same glycosidic conformation. For instance, in a parallel four-stranded quadruplex, where all the guanine bases are in the *anti* conformation, the four grooves have approximately the same medium width (Figure 5a). In a monomolecular structure having lateral loops, the guanosines adopt a *syn-anti-syn-anti* conformation, with each *syn* guanosine donating hydrogen bonds to an adjacent *anti* guanosine, and accepting hydrogen bonds from the other adjacent guanosine (6), thus resulting in grooves of alternating wide-narrow-wide-narrow widths (Figure 5b). Conversely, in a bimolecular quadruplex having diagonal loops, the guanosines adopt a *syn-syn-anti-anti* conformation, resulting in wide-medium-narrow-medium widths (Figure 5c).

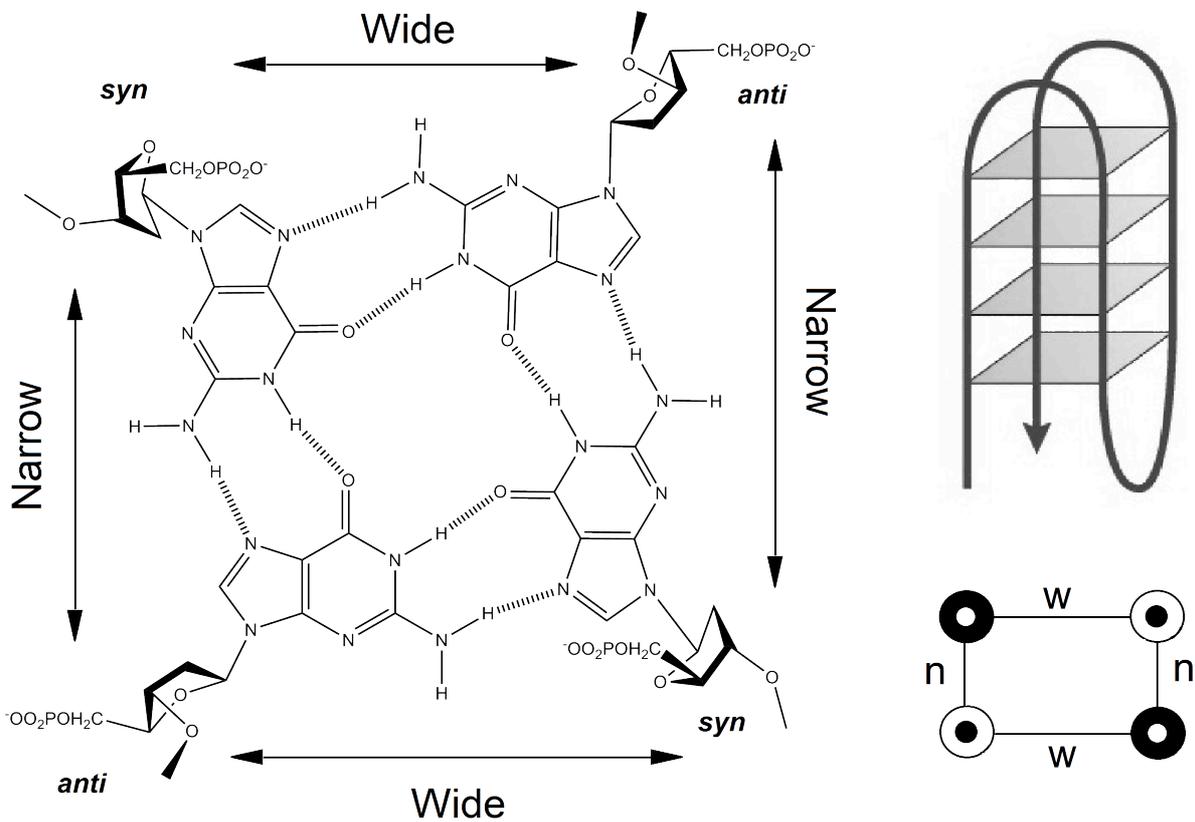
1.1.3 Metal ion coordination

The thermodynamic stability of the G-quadruplex structures in the solution state is mainly due to the interaction of the G-quartets with specific metal ion.

a)



b)



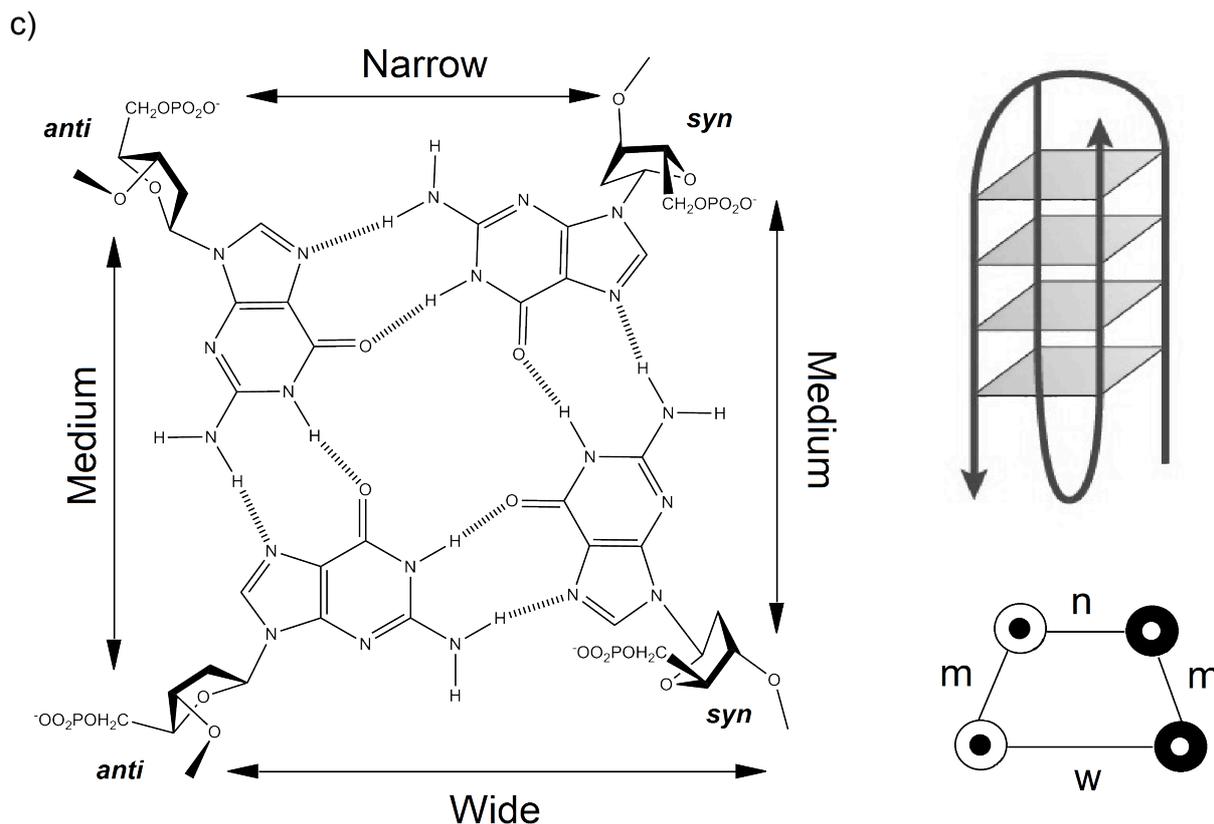


Figure 5. Variations in glycosidic torsional angles and their effects on groove width in G-quadruplex structures. a) all *anti*, all medium widths b) *syn-anti-syn-anti*, wide-narrow-wide-narrow widths and c) *syn-syn-anti-anti*, wide-medium-narrow-medium widths.

Although the structure of the G-quartets appears to be stabilized by the hydrogen bonds, generally, these structures do not form in the absence of such metal ions. In effect, the central core of the G-quartets produces specific geometric arrangements of lone pairs of electrons from the four O6 of the guanines, which can coordinate a monovalent ion of the correct size, such as sodium or potassium, whose positioning depends on its nature. For instance, the smaller sodium ion can sit exactly at the center of the quartet, in the same plane formed by the oxygen atoms; on the contrary, the larger potassium requires a non planar component, and it lies between two such G-quartets, coordinating eight oxygen atoms with a bi-pyramidal antiprismatic geometry (7).

1.1.4 Beyond G-tetrads, novel tetrad pairing alignment

A variety of alternative bases beyond guanines have been observed in G-quadruplex structures, and considerable effort has gone into defining these alignments. Keeping in mind that the standard view of G-quadruplex formation involves a scaffold stabilized by stacked G-tetrads, further non-guanine bases tetrads or mixed tetrads can coexist within the same structure (Figure 6) (8). For instance, a T tetrad has been found in a parallel tetraplex formed by the DNA sequence d-TGGTGGC containing two repeats of *Saccharomyces cerevisiae* telomere DNA (9). An A tetrad has been found in the truncated human telomeric DNA sequence d-AGGGT (10), and a C tetrad (11) has been observed in d-TGGGCGGT containing a repeat sequence from the SV40 viral genome. GCGC tetrads have been found in DNA tetraplexes with the repeat sequence d-GGGCT₄GGGC that is observed in adeno-

associated viral DNA and in human chromosome 19 and with the sequence d-GGGCT₃GGGC found in the fragile X syndrome triplet repeat (12). From a topological point of view, base tetrads may exhibit a cyclic, linear, or branched base-base interaction pattern.

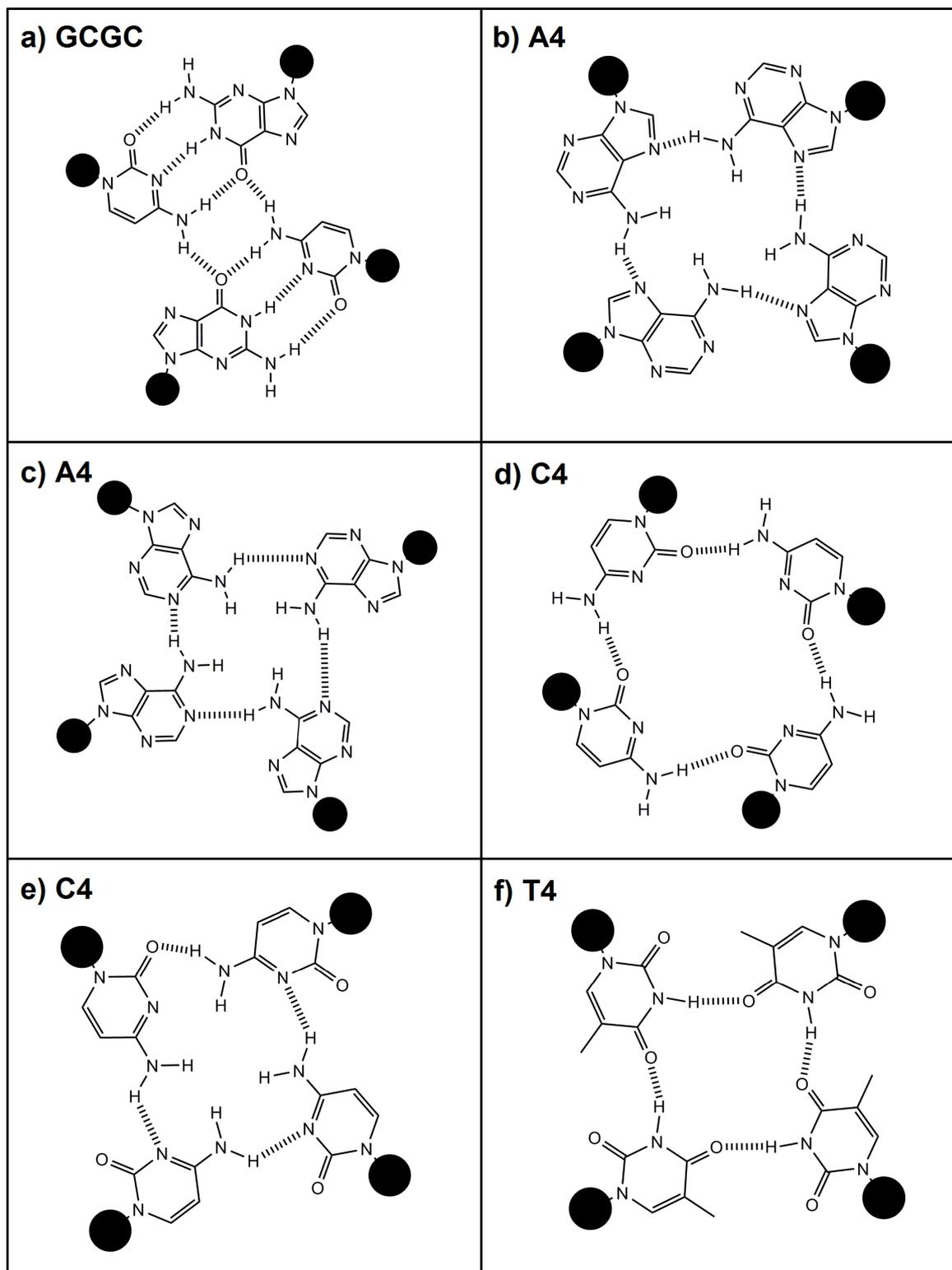


Figure 6. H-bond patterns of cyclic base tetrads found in experimental tetraplex structures. a) GCGC, b) A4: N6-H-N7 H-bonds, c) A4: N6-H-N1 H-bonds, d) C4: N4-H-O2 H-bonds, e) C4: N4-H-N3 H-bonds, f) T4. The full circles indicate the links to the sugar-phosphate backbone in nucleic acid tetraplex.

Almost all base tetrads known thus far adopt a cyclic structure (Figure 5). NMR-based investigations of G-quadruplexes have also identified formation of A-(G-G-G-G) pentads (13, 14), A-(G-G-G-G)-A hexads (15) and heptads (16, 17) (Figure 7). Such alignments essentially are composed of G-(A-G) triads, where one or more A residue(s) align(s) along one or more minor groove edge(s) of a G-tetrad.

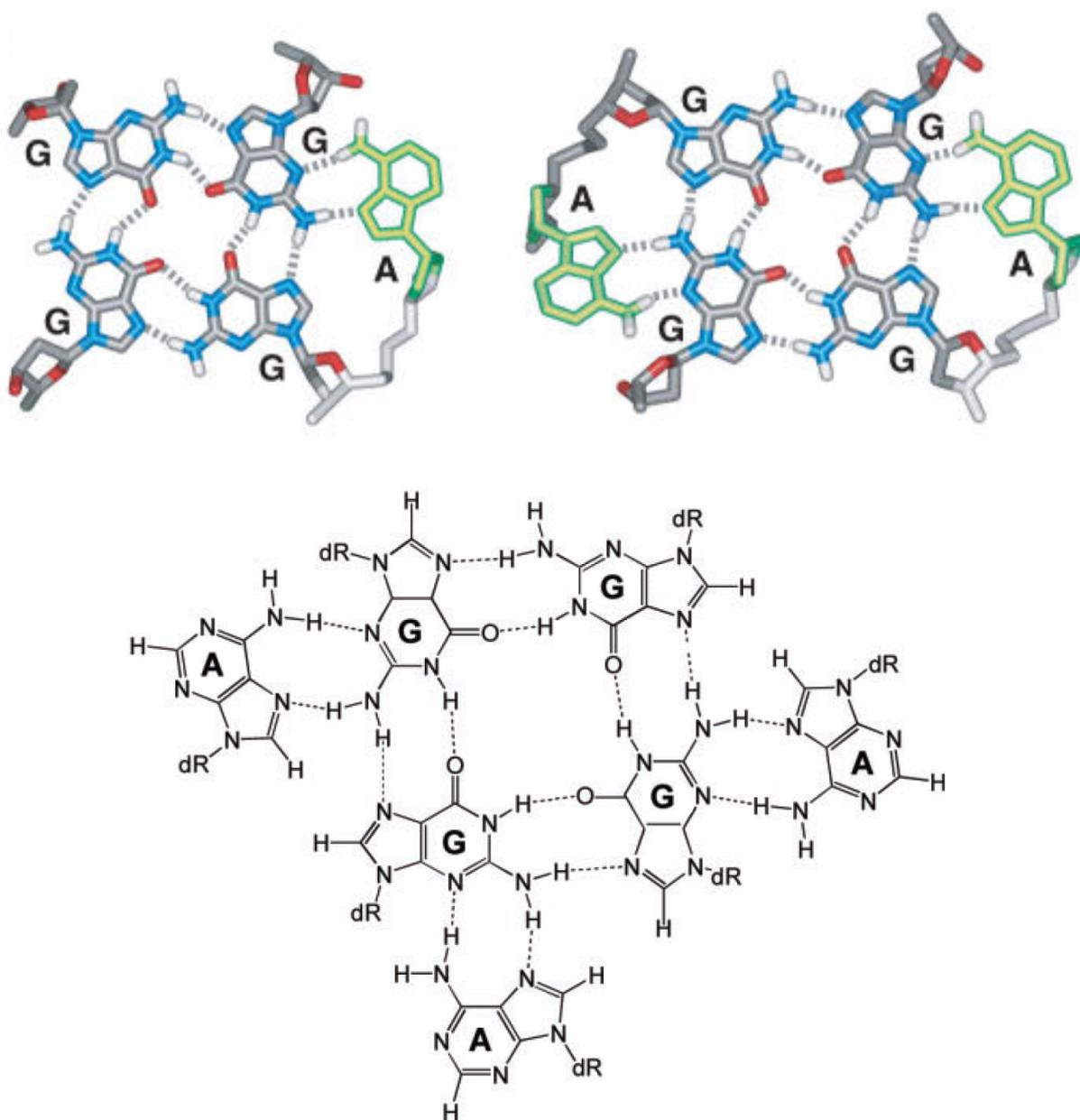


Figure 7. H-bond patterns of pentad (upper left), hexad (upper right), and heptad (lower) founded in experimental G-quadruplex structures.

1.2 Biological relevance of G-quadruplex structures

The presence of guanines-rich domains, potentially able to form G-quadruplex structures, are present in important regions of the human genome, such as the single-stranded portion of telomeres and different gene promoters, thus suggesting a possible role of the G-quadruplexes in the regulation of transcriptional and replicative

processes. What has stimulated more the imagination of researchers has been the possibility of going to interfere with the just mentioned processes, by exploiting small organic molecules able to recognize and bind the guanines-rich domains, with the aim of promote and stabilize the G-quadruplex formation. In this way, interesting and potential anticancer strategies have come to light.

For instance, in human somatic cells, telomere length decreases with each cell division event, while the inversion of this degradation, by a specialized enzyme called telomerase, increases the cell replicative capacity, resulting in an uncontrolled proliferation; in most cancer cells (85-90%), this enzyme is over-expressed (18). Since guanines-rich domains able to form G-quadruplex structures are present in the telomere, and since the formation of G-quadruplexes directly inhibits *in vitro* the telomerase activity, ligands that can selectively recognize and stabilize the G-quadruplexes, can interfere with the elongation of telomeres, by acting as inhibitors of telomerase, and thus as potential antitumor agents (Figure 8).

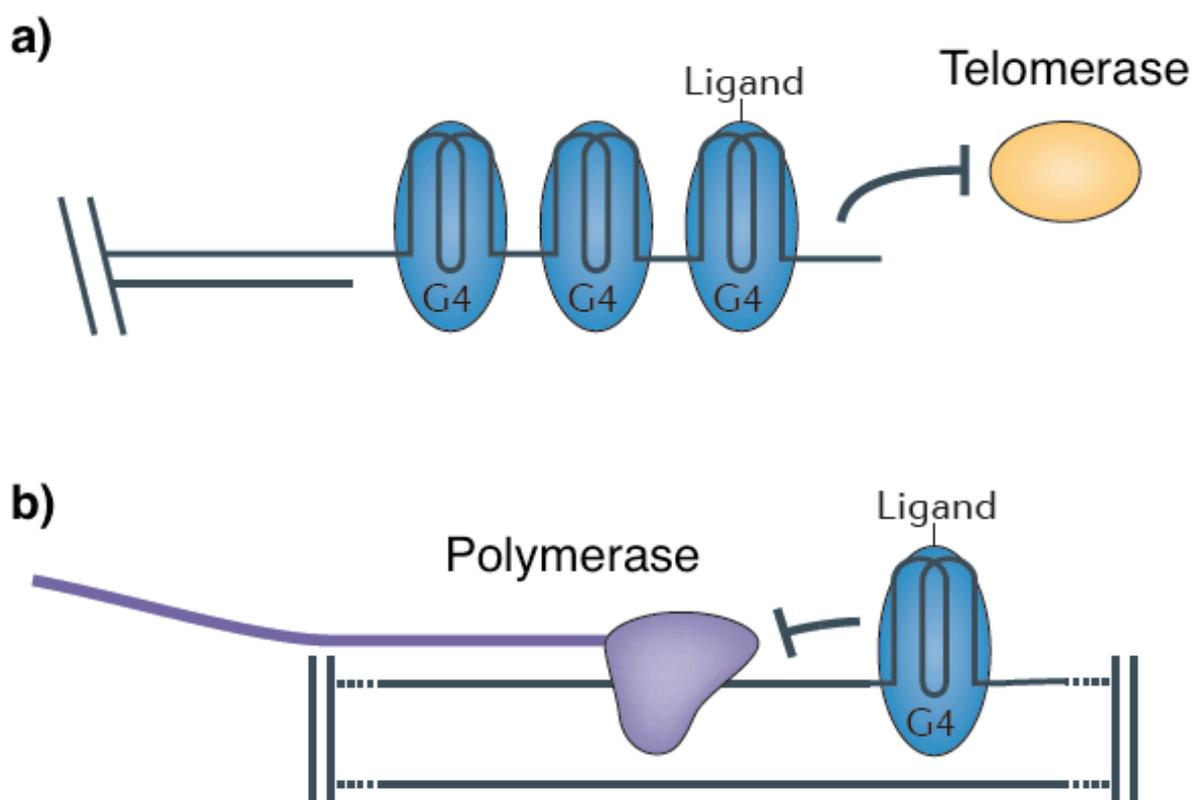


Figure 8. a) Ligands (blue) that bind to telomeric G-quadruplex (G4) are currently being analysed for their ability to influence telomere length by altering telomerase (yellow) activity. b) Ligands (blue) that bind to G-quadruplex (G4) of promoter sequences are postulated to block transcription by inhibiting polymerase (purple).

On the basis of the same principles, it is reported that the guanines-rich domains in the nucleotide sequences of promoters are often in equilibrium between a double stranded structure, that promotes the transcription, and a quadruplex-folded form, that instead inhibits the transcription (19). When the transcription process does not involve a gene but an oncogene, that is a gene which codes for proteins that direct the cell towards a neoplastic phenotype, the possibility of being able to block the transcription process by stabilizing the quadruplex-folded form, becomes extremely interesting from a pharmacological point of view. For this reason, ligands able to recognize and stabilize the G-quadruplex structures in promoter sequences of

important oncogenes, such as *c-Myc* (20-24), *c-kit* (25, 26), and *bcl-2* (27, 28), may cause the inhibition of transcription, and may therefore be used as antineoplastic agents.

1.3 G-quadruplex forming Aptamers

Another extremely interesting field of application of G-quadruplex structures is that of synthetic G-quadruplex forming oligonucleotides. The interest towards this kind of oligonucleotides mainly derives from the fact that they may serve as the scaffold of aptamers. Aptamers are small DNA or RNA molecules able to assume specific three-dimensional structuring in order to achieve a very high binding affinity towards specific proteins, such as thrombin or some HIV proteins, thus demonstrating promising biological properties and possible pharmacological applications (29). Particularly, one of the most studied G-quadruplex-based aptamer class is that of thrombin-binding aptamers (TBAs), which bind to α -thrombin, a key protein involved in the clotting process, preventing the pathological process of thrombosis (30, 31). It was also demonstrated that in this case the structuring in the G-quadruplex and its consequent stability, are essential requirements for the explication of biophysical and biological effects of TBA (32).

The aptamer AS1411 (Antisoma, London, UK) is another synthetic G-quadruplex forming nucleotide potentially able to inhibit the proliferation of cancer cells. The specific target protein of AS1411 is the nucleolin, a protein located on the surface of cancer cells. Unfortunately, clinical trials of AS1411 in patients with acute myeloid leukaemia (AML) have been locked in phase IIb, due to the financial collapse of the producer company. [Source: <http://www.fiercebiotech.com> the biotech industry's daily monitor]

Another large G-quadruplex-based aptamers class is that of potentially anti-HIV aptamers, that can inhibit the essential proteins of the HIV virus at different stages of its life cycle. Research in this sense has given rise to a wide range of aptamers, mainly directed against:

- the viral surface glycoprotein involved in the process of anchoring and fusion of the virus, known as gp120.
- the reverse transcriptase enzyme (Reverse Transcriptase, RT) involved in the process of reverse transcription, thus able to copy the viral RNA into a double-stranded cDNA before its integration into the genome of the host cell.
- the integrase enzyme (Integrase, IN) involved in the process of integration of viral DNA into the DNA of host cells.

Among the several G-quadruplex-based aptamers proposed and studied for their potentially anti-HIV activity, the most efficient were the 17-mer oligonucleotide GTGGTGGGTGGGTGGGT, also known as Zintevir, AR177 or T30177, as well as two 16-mer oligonucleotides (T30695 and 93del) and the shorter phosphorothioate oligonucleotide sequence TTGGGGTT, also known as ISIS 5320 (Figure 9a) (33). Particular attention has been called by the Hotoda 6-mer TGGGAG sequence, modified to bear a dimethoxytrityl group linked to its 5' hydroxyl function (34). Subsequently, other modified ODNs were prepared and tested and the 6-mer bearing a 3,4-dibenzyloxybenzyl group at 5'-end and a 2-hydroxy-ethylphosphate at the 3'-end, also known as R-95288 (Figure 9b), showed the most potent and less cytotoxic activity (35, 36).

Since the structural stability of G-quadruplex-forming aptamers is closely related to the biological activity of the aptamer, different synthetic approaches have been proposed to improve the chemical-physical characteristics of the G-quadruplex

structures, either by the introduction of large aromatic substituents at the 3'- or 5'-end (37), either by the anchor of the filaments to alkyl tetra-end-linker (TEL) (38) able to direct the structuring of the G-quadruplex, as proposed for the aptamer TEL-GAGGGT-TBDPS by Oliviero et al. in 2010 (Figure 9c) (39).

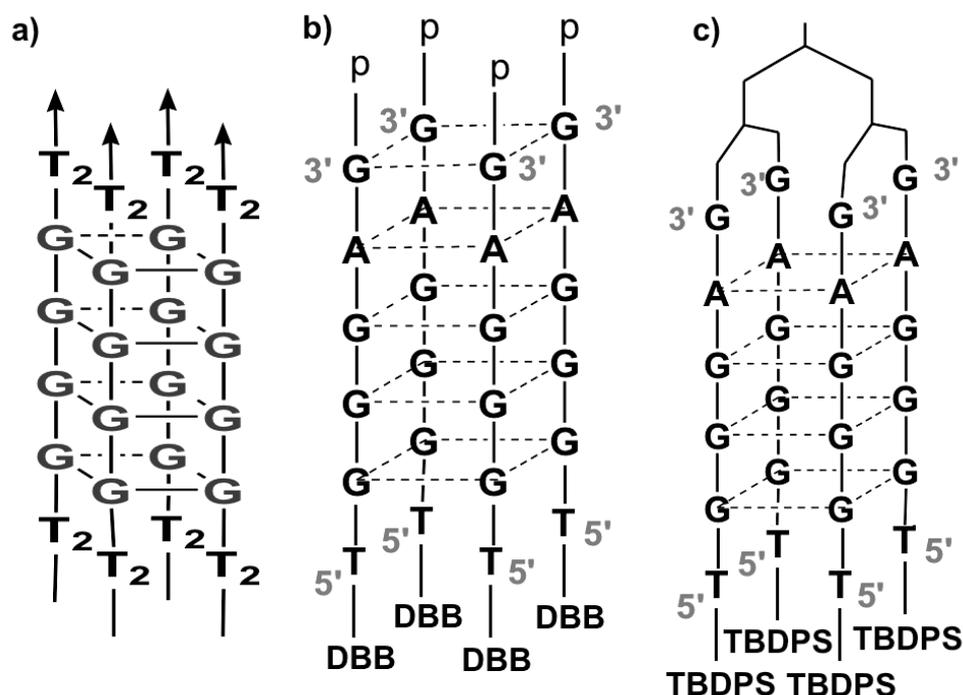


Figure 9. Some of reported anti-HIV aptamers. a) Phosphorothioate ISIS 5320; b) R-95288, where p = 2-hydroxyethylphosphate and DBB= 3,4-dibenzyloxybenzyl; c) TEL-GAGGGT-TBDPS, where TBDPS= terz-buthyldiphenyl-silil

1.4 G-quadruplex structures in Nanotechnology

The G-quadruplex structures have several properties that make them particularly interesting in the use in nanotechnology. Keeping in mind that the burgeoning field of nanotechnology is based on the controlled construction of molecules on micrometric or nanometric scale, the G-quadruplex represent the ideal competitors for this purpose, due to their ability of orderly and controllable self-assembly (40, 41). In this context, the G-quadruplex were used both for the construction of DNA-based nanostructures, such as G-wires, Frayed Wires and Synapses, both for the construction of biosensors and nano-devices.

1.4.1 G-quadruplexes-based Nanostructures

For duty of record, the G-Wires are G-quadruplex polymers, identified by Marsh and Henderson for the first time in 1994. G-wires originate from the parallel arrangement of strands having sequence $d(G_4T_2G_4)$; they are extraordinarily stable at high temperatures and they have a length that can vary from 10 to 1000 nm (42). In 1996, Macgregor and co-workers first identified the Frayed Wires, extended structures originated from the sequence $d(G_{15}A_{15})$, in which the guanines constitute the core of the structure, while the adenines radiate as arms outside of the structure (43). In the

same year, Sen and co-workers described a method to build nanostructures from the synapsis (hence the name Synapses) between two DNA duplexes (44). In this case, each duplex contains a "synaptic domain" consisting of guanines pairs engineered to facilitate the formation of G-tetrads and thus lead to dimerization of the duplexes.

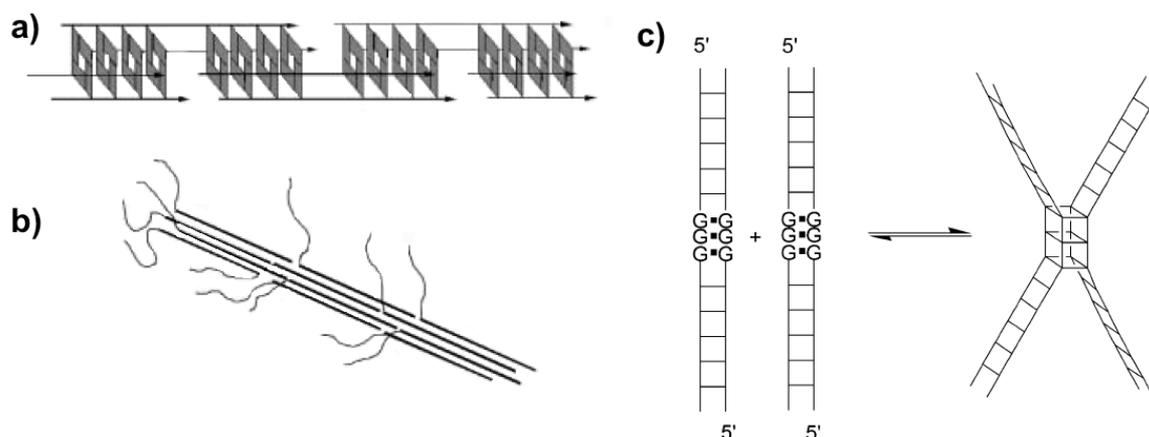


Figure 10. G-quadruplexes-based Nanostructures. a) G-wires formed by association of $d(G_4T_2G_4)$; b) Frayed wires formed by association of $d(G_{15}A_{15})$. The G residues make up the core, while the A_{15} arms radiate from the core; c) Synapsed DNA formed from G-quartet structures between duplex DNA strands containing G-G mismatches.

1.4.2 G-quadruplexes-based Biosensors and Nanodevices

The G-quadruplex-based aptamers, in addition to being used in therapeutic and diagnostic field, they also may have a practical application as biosensors in bioanalytical chemistry. In fact, given their high binding affinity, the G-quadruplex-based aptamers can be used as stationary phase in chromatography and electrophoresis for the separation of small molecules and proteins. For instance, McGown and co-workers have used this principle to discriminate between the enantiomers D, L of Trp or Tyr or to separate pool of similar proteins that differ in a few amino acids (45). On the other hand, a different approach involves the use of Fluorescence Resonance Energy Transfer (FRET) probes, a distance-dependent method used for the detection of conformational changes. Two probes are involved, a donor fluorophore and an acceptor chromophore respectively linked to the 3'-end and 5'-end of an oligonucleotide; when the two probes are close in space, the fluorescence is quenched, when they move apart, as result of a conformational change, there is a corresponding increase in fluorescence. Takenaka and co-workers used this method to monitor the transition from single-stranded DNA into G-quadruplex structure, according to the concentration of K^+ ions (46); similarly, Li and Tan monitored in real time the conformational duplex-quadruplex switching, evaluating the susceptibility of the switch according to the concentration of different cations or in the presence of ligands and proteins (47).

1.5 Aim of my research work

In this context, my research activities involved the identification and the structural characterization of several biomolecules, such as DNA aptamers and nanostructures,

whose core architectures were based on G-quadruplex structures. Notably, I mainly focused my attention on:

- **G-quadruplex-based aptamers endowed with high anti-HIV activity**

The aim of this research was to achieve a small library of G-quadruplex-based aptamers, in order to assess the influence of different nucleotide sequences on the anti-HIV activity.

In 2010, Oliviero and co-workers proposed some tetra-end-linked oligonucleotides (TEL-ODNs) containing the Hotoda's sequence TGGGAG, as new modified aptamers with noteworthy anti-HIV activity (39). In this work, several TEL-ODNs were prepared and analysed in order to probe the influence of the lipophilic moieties and the TEL position on the structural and biological properties, and a complex lacking of the TEL moiety was used as reference for the evaluation of the thermal stability and the anti-HIV activity. The reported results showed that the presence of both the 5'-tert-butyl-diphenyl-silyl (TBDPS) lipophilic groups and the 3'-TEL was necessary to dramatically improve the thermal stability of the aptamers and their anti-HIV-1 activity. Furthermore, it was previously reported that Hotoda's aptamers bind to the glycoprotein 120 hyper variable V3 loop (gp120 V3 loop) of the HIV-1 virus, thus inhibiting cell absorption and penetration by the virus (35). For this reason, the biological activity of TEL-ODNs was mainly related to two different factors: (a) the extent of interaction with the V3 loop and (b) the amount of G4 able to interact with target, that in turn is dependent on the quadruplex thermal stability, the kinetics of formation and the resistance to nucleases. As regards the first factor, molecular modelling studies reported in the above-mentioned work suggested that the V3 loop residues interact with the grooves and the sugar-phosphate backbones of the aptamer. Thus, the higher biological activity of TEL-ODNs was ascribed to both their enhanced thermal stability and to their monomolecular nature that improves the kinetics of formation. Furthermore, it should be taken into account that complexes showing the highest anti-HIV activities were characterized by capped 3'- and 5'-ends that definitively improve their resistance to exonucleases.

Starting from these assumptions, the synthesis of TEL-ODNs, having TGGGXG (where X = G, C or T) sequences, conjugated with tert-butyl-diphenyl-silyl (TBDPS) groups attached to the 5'-end and linked to the TEL moiety on the 3'-end, has been undertaken to probe the influence of the different nucleotides on the structural and biological properties of such anti-HIV aptamers. Notably, during the period that I spent at the Centre of Cancer Research and Cell Biology (CCRCB) of the Queen's University of Belfast, I performed molecular modelling studies useful to reveal the mechanism of action of the new synthesized aptamers, and investigate the influence of different nucleotides on the interactions with V3 loop of HIV.

From the collected data, the new aptamers showed EC_{50} values against HIV-1 in the range of 0.04–0.15 μ M as well as affinities for the HIV-1 gp120 envelope in the same order of magnitude (48).

- **Thrombin-Binding Aptamers (TBAs) Analogs with increased anticoagulant activity**

The aim of this research work was to achieve an increased anticoagulant activity of TBA through the modification of individual nucleobases.

A modified nucleotide, containing a five-member cycle fused on the pyrimidine ring, was introduced at position 7 or 12 of the 15-mer TBA-forming oligonucleotide dGGTTGGTGTGGTTGG. The structural characterization of the resulting aptamers, TBA-T₇b and TBA-T₁₂b, showed their ability to fold into the typical structuring of TBA, an antiparallel chair-like G-quadruplex. The apparent CD melting temperatures indicated that the introduction of the acyclic residue, mainly in position 7, improves the thermal stability of the resulting G-quadruplex with respect to TBA. The anticoagulant activity of the new molecules was then evaluated by the prothrombin time assay (PT assay), and it resulted that TBA-T₇b is more potent than TBA in prolonging clotting time. On the other hand, in purified fibrinogen assay the thrombin inhibitory activity of both modified sequences was lower than that of TBA using human enzyme, whereas the potency trend was again reversed using bovine enzyme. Obtained structure–activity relationships were investigated by structural and computational studies. Taken together, the results reveal the active role of TBA residues T₇ and T₁₂ and the relevance of some amino acids located in the anion binding exosite I of the protein in aptamer–thrombin interaction (49).

▪ DNA-based Nanostructures formed by G-quadruplexes dimers

The aim of this research work was to achieve a small library of DNA-based nanostructures consisting of dimers of G-quadruplexes.

Thanks to several analytical techniques, such as Nuclear Magnetic Resonance (NMR), mass spectrometry (MS), Circular Dichroism (CD), UV and gel electrophoresis of Poly-acrylamide (PAGE), it was proved that the sequence CGGXGGT (where X = T, A, C or G) can form a nanostructure of about 4 nm in length, consisting of eight well organized strands. This octamer formation takes place through a dimerization process, which is allowed by the 5'-5' stacking of two tetramolecular G-quadruplex subunits, and assisted by the formation of an unusual G(:C):G(:C):G(:C):G(:C) octad arrangement, involving C1 and G2 bases of each CGGXGGT strand (50).

1.6 References

- (1) M. L. Bochman, K. P., and V. A. Zakian. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nature Reviews* 13, 770-780.
- (2) G. Biffi, D. T., J. McCafferty, and S. Balasubramanian. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry*.
- (3) Kerwin, S. (2000) G-quadruplex DNA as a target for drug design. *Current Pharmaceutical Design* 6, 441-478.
- (4) Parkinson, G. N. (2006) Fundamentals of Quadruplex Structures, in *Quadruplex Nucleic Acids* pp 1-30, RSC Publishing.
- (5) Patel, Y. W. a. D. J. (1995) Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex. *Journal of Molecular Biology* 251, 76-94.
- (6) C. Kang, X. Z., R. Ratliff, R. Moyzis, and A. Rich. (1992) Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature* 356, 126-131.
- (7) A. N. Lane, J. B. C., R. D. Gray, and J. O. Trent. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research* 36, 5482-5515.
- (8) M. Meyer, C. S., M. Brandl, and J. Suhnel. (2001) Cyclic Adenine-, Cytosine-, Thymine-, and Mixed Guanine-Cytosine-Base Tetrads in Nucleic Acids

- Viewed from a Quantum-Chemical and Force Field Perspective. *The Journal of Physical Chemistry A* 105, 11560-11573.
- (9) P. K. Patel, a. R. V. H. (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Research* 27, 2457-2464.
 - (10) P. K. Patel, A. S. K., and R. V. Hosur. (1999) NMR studies on truncated sequences of human telomeric DNA: observation of novel A-tetrad. *Nucleic Acids Research* 27, 3836-3843.
 - (11) P. K. Patel, N. S. B., and R. V. Hosur. (2000) NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG. *Biochemical and Biophysical Research Communications* 270, 967-971.
 - (12) A. Kettani, S. B., A. Gorin, H. Zhao, R. A. Jones, and D. J. Patel. (1998) Solution structure of a Na cation stabilized DNA quadruplex containing GGGG and GCGC tetrads formed by GGGC repeats observed in adeno-associated viral DNA. *Journal of Molecular Biology* 282, 619-636.
 - (13) N. Zhang, A. G., A. Majumdar, A. Kettani, N. Chernichenko, E. Skipkin, and D. J. Patel. (2001) V-shaped scaffold: a new architectural motif identified in an A(GGGG) pentad-containing dimeric DNA quadruplex involving stacked G(anti)G(anti)G(anti)G(syn) tetrads. *Journal of Molecular Biology* 311, 1063-1079.
 - (14) A. T. Phan, V. K., J. B. Ma, A. Faure, M. L. Andreola, and D. J. Patel. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proceedings of the National Academy of Sciences* 102, 634-639.
 - (15) A. Kettani, A. G., A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. A. Jones, and D. J. Patel. (2000) A dimeric DNA interface stabilized by stacked A(GGGG)A hexads and coordinated monovalent cations. *Journal of Molecular Biology* 297, 627-644.
 - (16) A. Matsugami, K. O., M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi, and M. Katahira. (2001) An intermolecular quadruplex of (GGA)₄ triplet repeat DNA with GGGG tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *Journal of Molecular Biology* 313, 255-269.
 - (17) H. Sotoya, A. M., T. Ikeda, K. Ouhashi, S. Uesugi, and M. Katahira. (2004) Method for direct discrimination of intra- and intermolecular hydrogen bonds, and characterization of the G(:A):G(:A):G(:A):G heptad, with scalar couplings across hydrogen bonds. *Nucleic Acids Research* 32, 5113-5118.
 - (18) J. L. Mergny, J. F. R., M. P. Teulade-Fichou, and E. Gilson. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Research* 30, 839-865.
 - (19) S. Cogoi, a. L. E. X. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Research* 34, 2536-2549.
 - (20) T. Simonsson, P. P., and M. Kubista. (1998) DNA tetraplex formation in the control region of *c-myc*. *Nucleic Acids Research* 26, 1167-1172.
 - (21) A. Siddiqui-Jain, C. L. G., D. J. Bearss, and L. H. Hurley. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proceedings of the National Academy of Sciences* 99, 11593-11598.
 - (22) C. L. Grand, H. H., R. M. Munoz, S. Weitman, D. D. Von Hoff, L. H. Hurley, and D. J. Bearss. (2002) The cationic porphyrin TMPyP4 down-regulates c-

- myc* and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Molecular Cancer Therapeutics* 8, 565-573.
- (23) J. Seenisamy, E. M. R., T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain, and L. H. Hurley. (2004) The dynamic character of the G-quadruplex element in the *c-myc* promoter and modification by TMPyP4. *Journal of the American Chemical Society* 126, 8702-8709.
- (24) V. Gabelica, E. S. B., M. P. Teulade-Fichou, E. De Pauw, and M. T. Bowers. (2007) Stabilization and structure of telomeric and *c-myc* region intramolecular G-quadruplexes: the role of central cations and small planar ligands. *Journal of the American Chemical Society* 129, 895-904.
- (25) S. Rankin, A. P. R., J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian, S. Neidle. (2005) Putative DNA quadruplex formation within the human *c-kit* oncogene. *Journal of the American Chemical Society* 127, 10584-10589.
- (26) A. T. Phan, V. K., S. Burge, S. Neidle, and D. J. Patel. (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-kit* promoter. *Journal of the American Chemical Society* 129, 4386-4392.
- (27) J. Dai, T. S. D., D. Chen, M. Carver, A. Ambrus, R. A. Jones, and D. Yang. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human *BCL-2* promoter region in solution. *Journal of the American Chemical Society* 128, 1096-1098.
- (28) T. S. Dexheimer, D. S., and L. H. Hurley. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the *bcl-2* P1 promoter. *Journal of the American Chemical Society* 128, 5404-5415.
- (29) D. M. Held, J. D. K., J. T. Patterson, D. G. Nickens, and D. H. Burke. (2006) HIV-1 inactivation by nucleic acid aptamers. *Frontiers in Bioscience* 11, 89-112.
- (30) L. C. Bock, L. C. G., J. A. Latham, E. H. Vermaas, and J. J. Toole. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355, 564-566.
- (31) K. Y. Wang, S. M., R. G. Shea, S. Swaminathan, and P. H. Bolton. (1993) A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry* 32, 1899-1894.
- (32) B. Pagano, L. M., A. Randazzo, and C. Giancola. (2008) Stability and binding properties of a modified thrombin binding aptamer. *Biophysical Journal* 94, 562-569.
- (33) S. H. Chou, K. H. C., and A. H. Wang. (2005) DNA aptamers as potential anti-HIV agents. *Trends in Biochemical Sciences* 30, 231-234.
- (34) H. Hotoda, K. M., H. Furukawa, T. Nakamura, M. Kaneko, S. Kimura, and K. Shimada. (1994) Biologically Active Oligodeoxyribonucleotides-II: Structure Activity Relationships of anti-HIV-1 pentadecaceoxyribonucleotides bearing 5'-end-modifications. *Nucleosides, Nucleotides, and Nucleic Acids* 13, 1375-1395.
- (35) M. Koizumi, R. K., H. Hotoda, K. Momota, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura, and K. Shimada. (1997) Biologically active oligodeoxyribonucleotides-IX: synthesis and anti-HIV-1 activity of hexadeoxyribonucleotides, TGGGAG, bearing 3'- and 5'-end-modification. *Bioorganic and Medicinal Chemistry* 5, 2235-2243.

- (36) M. Koizumi, R. K., H. Hotoda, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura, and K. Shimada. (1998) Biologically Active Oligodeoxynucleotides. Part 11: The laeast phosphate-modification of quadruplex-forming hexadeoxyribonucleotide TGGGAG, bearing 3'- and 5'-end-modification, with anti-HIV-1 activity. *Bioorganic and Medicinal Chemistry* 6, 2469-2475.
- (37) G. Di Fabio, J. D. O., M. Chiapparelli, B. Hoorelbeke, D. Montesarchio, J. Balzarini, and L. De Napoli. (2011) Discovery of novel anti-HIV active G-quadruplex-forming oligonucleotides. *Chemical Communications* 47, 2363-2365.
- (38) G. Oliviero, N. B., J. Amato, S. D'Errico, A. Galeone, G. Piccialli, M. Varra, and L. Mayol. (2009) Synthesis of quadruplex-forming tetra-end-linked oligonucleotides: effects of the linker size on quadruplex topology and stability. *Biopolymers* 91, 466-477.
- (39) G. Oliviero, J. A., N. Borbone, S. D'Errico, A. Galeone, L. Mayol, S. Haider, O. Olubiyi, B. Hoorelbeke, J. Balzarini, and G. Piccialli. (2010) Tetra-end-linked oligonucleotides forming DNA G-quadruplexes: a new class of aptamers showing anti-HIV activity. *Chemical Communications* 46, 8971-8973.
- (40) Niemeyer, C. M. (2001) Nanoparticles, Proteins, and Nucleic Acids: Biotechnology meets materials science. *Angewandte Chemie International Edition* 40, 4128-4158.
- (41) Niemeyer, C. M. (2002) Nanomechanical Devixes based on DNA. *Angewandte Chemie International Edition* 41, 3779-3783.
- (42) T. C. Marsh, a. E. H. (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. *Biochemistry* 33, 10718-10724.
- (43) E. Protozanova, a. R. B. J. M. (1996) Frayed wires: a thermally stable form of DNA with two distinct structural dimains. *Biochemistry* 35, 16638-16645.
- (44) E. A. Venczel, a. D. S. (1996) Synapsable DNA. *Journal of Molecular Biology* 257, 219-224.
- (45) R. B. Kotia, L. L., and L. B. McGown. (2000) Separation of nontarget compounds by DNA aptamers. *Analytical Chemistry* 72, 827-831.
- (46) H. Ueyama, M. T., and S. Takenaka. (2002) A novel potassium sensing in aqueous media with a synthetic oligonucleotide derivative. Fluorescence resonance energy transfer associated with guanine quartet-potassium ion complex formation. *Journal of the American Chemical Society* 124, 14287-14287.
- (47) J. J. Li, a. W. T. (2002) A single DNA molecule nanomotor. *Nano Letters* 2, 315-318.
- (48) V. D'Atri, G. O., J. Amato, N. Borbone, S. D'Errico, L. Mayol, V. Piccialli, S. Haider, B. Hoorelbeke, J. Balzarini, and G. Piccialli. (2012) New anti-HIV aptamers based on tetra-end-linked DNA G-quadruplexes: effect of the base sequence on anti-HIV activity. *Chemical Communications* 48, 9516-9518.
- (49) N. Borbone, M. B., G. Oliviero, E. Morelli, J. Amato, V. D'Atri, S. D'Errico, V. Vellecco, G. Cirino, G. Piccialli, C. Fattorusso, M. Varra, L. Mayol, M. Persico, and M. Scutto. (2012) Investigating the role of T7 and T12 residues on the biological properties of thrombin-binding aptamer:enhancement of anticoagulant activity by a single nucleobase modification. *Journal of Medicinal Chemistry* 55, 10716-10728.

- (50) N. Borbone, J. A., G. Oliviero, V. D'Atri, V. Gabelica, E. De Paw, G. Piccialli, and L. Mayol. (2011) d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Research* 39, 7848-7857.

Chapter 2

New anti-HIV aptamers based on tetra-end-linked DNA G-quadruplexes

The first stage of the HIV infection requires the entry of human immunodeficiency virus (HIV) into host cells. This stage involves the sequential interaction of the virion surface glycoprotein gp120 with the CD4 glycoprotein and a chemokine receptor, either CCR5 or CXCR4, on the host cell surface (1). The CD4 glycoprotein is expressed on the surface of T-lymphocytes, monocytes, dendritic cells and brain microglia, and its expression makes these cells a target for HIV in vivo. Furthermore, an interesting function of CD4 binding is to induce conformational changes in gp120 that allow binding to the co-receptor, which is essential for viral entry (2). The third variable region of gp120 (from now on designated V3 loop) is a pivotal component of the co-receptor binding site, typically consisting of a 35 amino acid-loop (range 31 to 39), closed by two cysteines that form a disulphide bridge. The crystal structure of gp120 complexed with the CD4 receptor and a neutralizing antibody (PDB ID 2B4C) revealed that the V3 loop is extended away from the gp120 protein and it is involved in co-receptor binding and selection, acting as a “molecular hook” that organizes associations within the viral spike (3). The importance of gp120 and the role of the V3 loop in HIV-1 entry and pathogenesis have led to the recent pursuit of drugs targeted against it. One of the most studied alternatives is the use of aptamer technology (4). Aptamers are single stranded DNA or RNA molecules selected among large pools of nucleic acid sequences for their ability to bind selectively and with high affinity to a biomedically relevant target. For their properties, aptamers can be considered as the nucleic acid analogues of antibodies. Like the antibodies, several aptamers proved to be valuable diagnostic tools (5) and promising therapeutics (6, 7). In 1994 H. Hotoda (SA-1042) (8) and J. R. Wyatt (ISIS 5320) (9) independently reported the first anti-HIV aptamers targeted against the V3 loop. In both cases the ODN sequences were chemically modified to improve their resistance against nucleases by capping the 5'-ends with DMT groups (SA-1042) or by using the phosphorothioate backbone (ISIS 5320). Subsequently, other analogues of SA-1042 were obtained and tested (10, 11), and the 6-mer TGGGAG ODN (known as “Hotoda’s sequence”) bearing 3,4-dibenzyloxybenzyl groups at the 5'-ends and 2-hydroxyethylphosphates at the 3'-ends (R-95288) showed the highest anti-HIV activity (11). CD investigations on ISIS 5320, R95288 and their analogues suggested their structuration into parallel tetramolecular quadruplexes. Following biophysical studies established that the presence of aromatic (12, 13) moieties at 5'- and/or 3'-ends of TGGGAG dramatically enhances the rate of formation of quadruplex complexes. Moreover, the overall stability of quadruplex complexes was found to correlate with the reported anti-HIV activity (12).

In 2010, with the aim to overcome the unfavourable entropic factor and to stabilize the A-tetrad in Hotoda’s analogues, Oliviero et al. synthesized a series of new monomolecular anti-HIV aptamers by using the Tetra-End-Linker (TEL) strategy proposed by the same group in 2004 (14, 15). Several TEL-(TGGGAG)₄ aptamers were prepared and analysed in order to probe the influence of lipophilic groups and TEL size and position on the structural and anti-HIV properties of the resulting TEL-quadruplexes (16). The results showed that (i) the presence of the TEL at either 5'- or 3'-ends was required for the anti-HIV activity, and (ii) lipophilic tert-butylidiphenylsilyl (TBDPS) groups at the 5'-ends strongly enhanced both the stability of TEL-quadruplexes and their anti-HIV activity. The EC₅₀ of the best aptamer (I, Fig. 1),

bearing TBDPS groups at the 5'-ends and the longer TEL at the 3'-ends was $0.082 \pm 0.04 \mu\text{M}$.

2.1 Results and Discussion

With the aim to further improve the anti-HIV activity of TEL-ODN aptamers, we used **I** as a molecular synthon for the synthesis of a series of new TEL-ODN aptamers embodying the TGGGXXG sequence (**II–IV**, Fig. 1). TEL-ODNs **V** and **VI** (a TEL-(TGGGAG)₄ analogue unable to form a TEL-quadruplex) were also synthesized and studied in order to investigate the role played by the G-tetrad at the 3'-terminus of TEL-quadruplexes **I–IV** and to further corroborate the requirement of quadruplex formation for the inhibition of HIV cytopathicity. In this communication we report the results of biophysical, biological and SAR studies performed on **I–VI**.

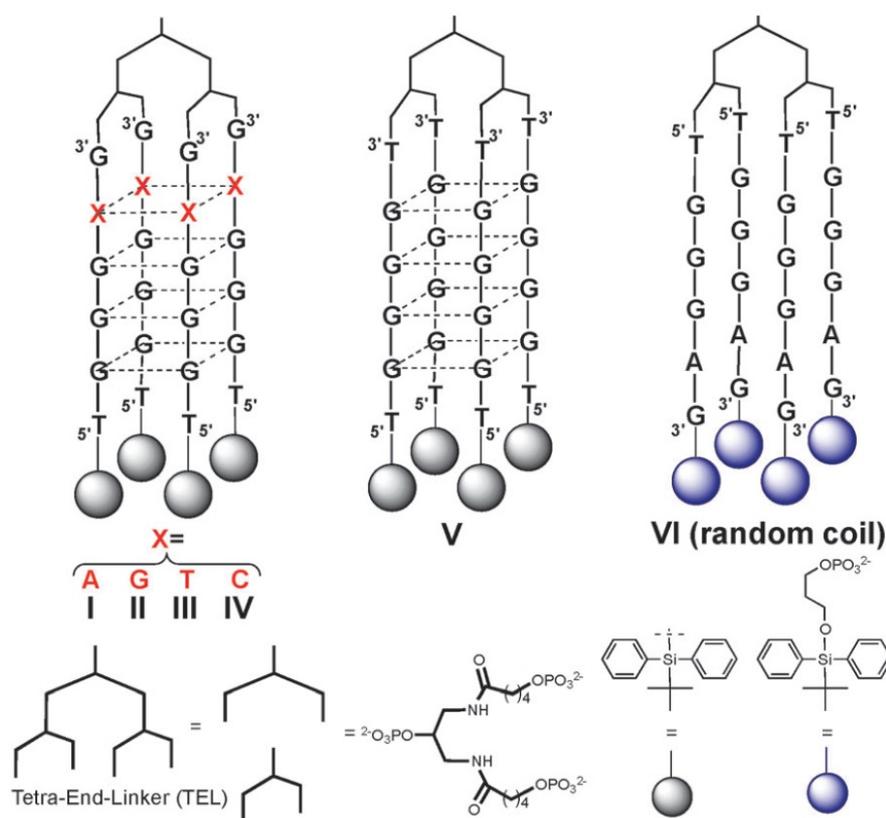


Figure 1. Schematic representation of investigated TEL-ODNs.

All products were synthesized and characterized following the previously described approach (14, 16). In order to verify that the presence of one (**II–IV**) or two (**V**) different bases did not affect the parallel quadruplex arrangement observed for the parent TEL-ODN **I**, CD measurements were performed (Fig. 2). The overall CD profiles, recorded in 100 mM K⁺ buffer, were in agreement with the formation of G-quadruplex structures, showing two positive maxima at 220 and 260 nm and a negative minimum close to 240 nm, which are characteristic of head-to-tail arrangement of guanines, as typically found in parallel G-quadruplexes (17–19). Unexpectedly, CD spectra of **VI** were similar to those of a random coil, thus evidencing that the insertion of a propyl-phosphate moiety between the OH at 3'-ends

and TBDPS groups is detrimental for quadruplex stability. CD melting analyses on **II–IV** confirmed the expected great thermal stability of the new TEL-quadruplexes (Fig. 11 in 2.3.5). In fact, as for **I**, the apparent melting temperature of **II–IV** could not be determined since no derivatizable melting curve was obtained.

The formation of thermally stable TEL-quadruplexes for **II–IV** was also confirmed by ^1H NMR evidence (Fig. 3 and 4). The G-quadruplex diagnostic exchange-protected imino proton signals were observed in all recorded spectra up to $90\text{ }^\circ\text{C}$ (100 mM K^+ buffer; $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1). As expected, the intensity of imino proton signals reflected the stability of the X-tetrad within the G-quadruplex structure. Stronger imino signals were observed for **II** (five stacked G-tetrads), whereas weaker imino signals were observed for **I**, **III** and **IV** due to the presence of the less stable A-, T- or C-tetrad, respectively, that accounts for the faster exchange of imino protons with the solvent.

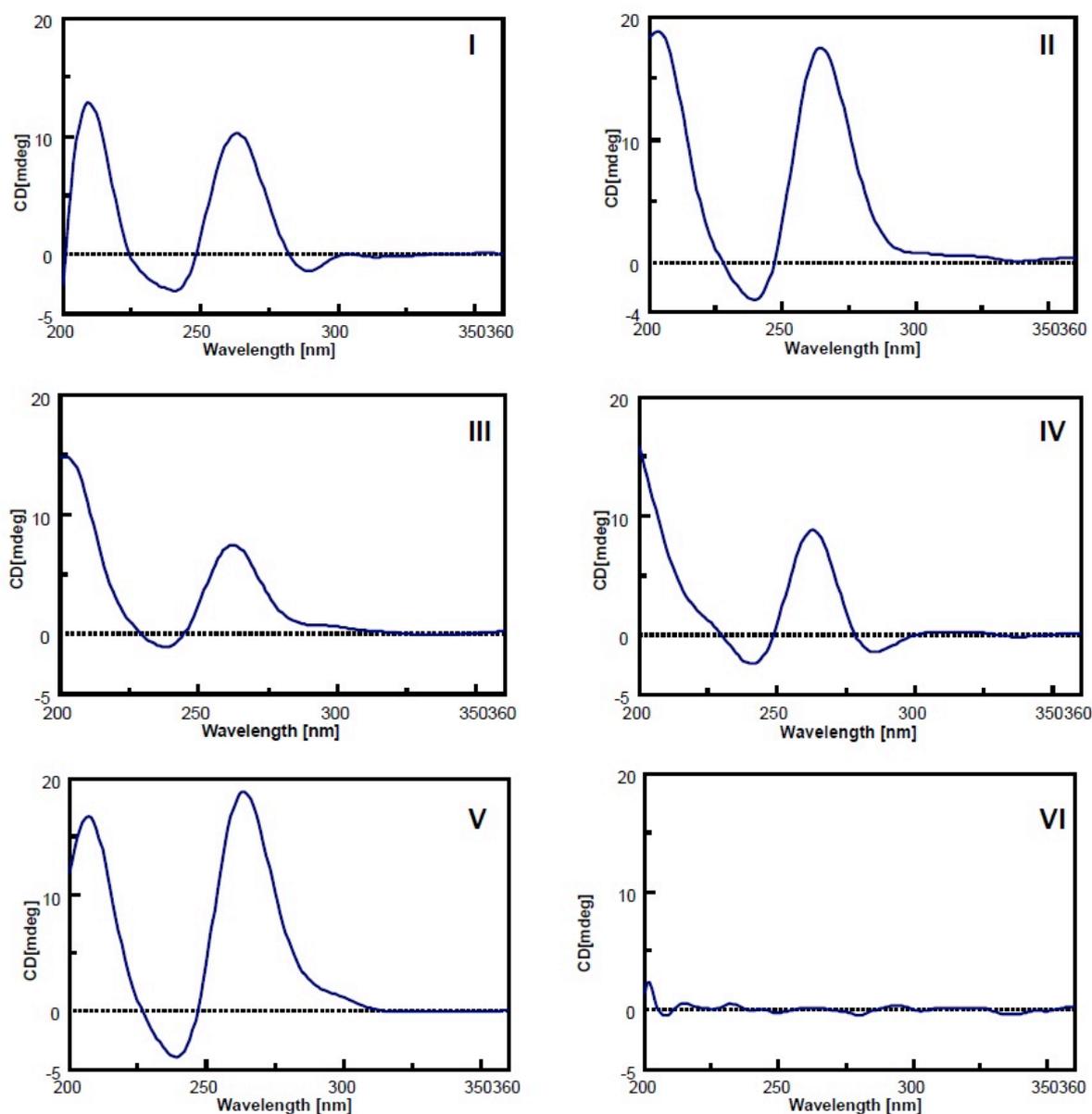


Figure 2. CD profiles of **I–VI** in 100 mM K^+ buffer at $25\text{ }^\circ\text{C}$.

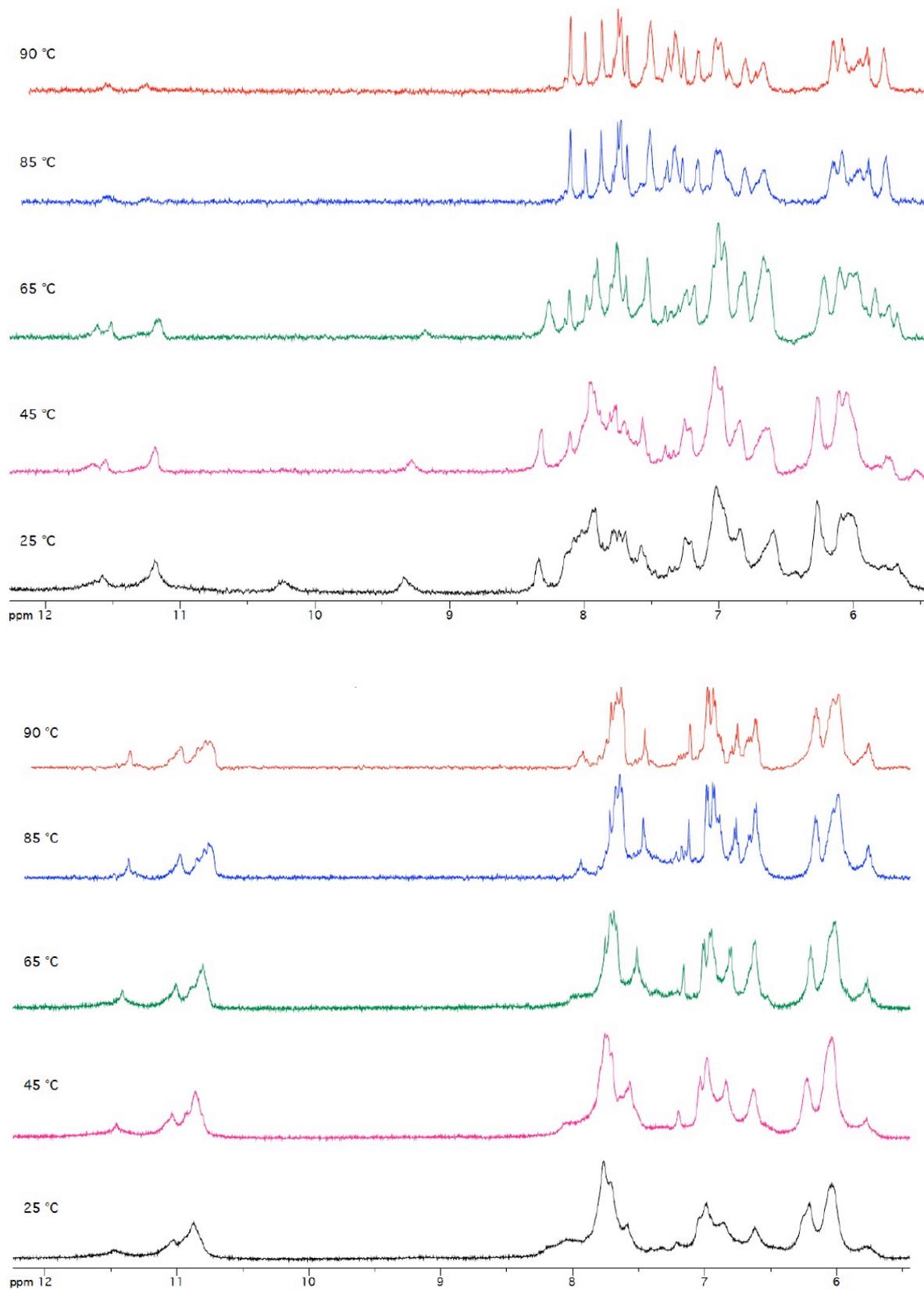


Figure 3. ^1H NMR spectra of **I** (first panel) and **II** (second panel) recorded in the temperature range 25-90 °C (100 mM K^+ buffer; $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1).

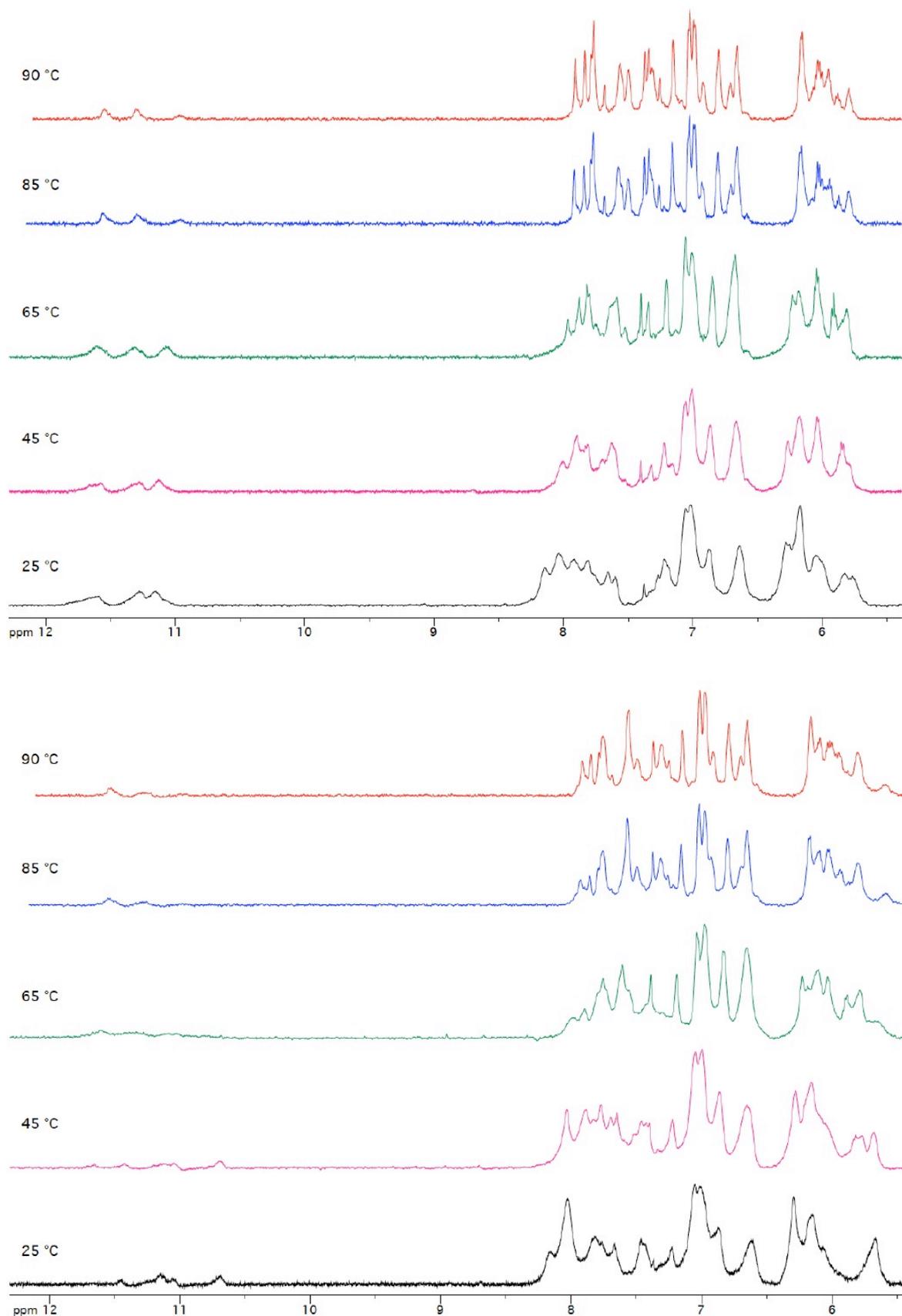


Figure 4. ^1H NMR spectra of **III** (first panel) and **IV** (second panel) recorded in the temperature range 25-90 °C (100 mM K^+ buffer; $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1).

The antiviral activity of **II–IV** against HIV-1 and HIV-2 was determined as previously reported for **I** (16) and is shown in Table 1. The binding properties of active aptamers **II–IV** to HIV-1 (III_B) gp120 were also determined through Surface Plasmon Resonance (SPR) technology (Table 1; Fig. 5). All new TEL-(TGGGXG)₄ aptamers reported (**II–IV**) retained potent anti-HIV-1 activity (nanomolar EC₅₀ values). They displayed comparable on (*k_a*) and off (*k_d*) rates for HIV-1 gp120 binding resulting in very similar gp120 binding affinities (*K_D*) to **I**, thus confirming that the residues of the gp120 V3 loop interact mainly with the grooves and the sugar–phosphate backbone of the aptamers. However, considering that the EC₅₀'s of **II** and **IV** were significantly lower than those of **I** and **III** (0.039–0.041 versus 0.11–0.15 mM), and that **V** proved to be markedly less active notwithstanding its structuration in a stable TEL-quadruplex structure, it has been hypothesized that some specific interactions between the X5G6 bases of **I–IV** and the V3 loop must occur.

ODN	EC ₅₀ ^a (μM)		<i>K_D</i> (nM)	SPR vs. HIV-1 (III _B) gp120 ^b	
	HIV-1	HIV-2		<i>k_a</i> (1/Ms)	<i>k_d</i> (1/s)
I	0.11 ± 0.05	1.4 ± 0.9	256 ± 63	(1.72 ± 0.28) × 10 ⁴	(4.31 ± 0.37) × 10 ⁻³
II	0.041 ± 0.007	≥ 2	183 ± 40	(2.40 ± 0.20) × 10 ⁴	(4.35 ± 0.61) × 10 ⁻³
III	0.15 ± 0.001	1.4 ± 0.87	248 ± 28	(2.55 ± 0.35) × 10 ⁴	(6.29 ± 0.14) × 10 ⁻³
IV	0.039 ± 0.005	0.73 ± 0.16	196 ± 0.16	(2.15 ± 0.40) × 10 ⁴	(4.07 ± 0.71) × 10 ⁻³
V	≥ 2	2	–	–	–

^a EC₅₀ represents the 50% effective concentration required to inhibit virus-induced cytopathicity in CEM cell cultures by 50%. ^b HIV-1 (III_B) gp120 was obtained from recombinant-expressed gp120 in CHO cell cultures.

Table 1. Anti-HIV and SPR studies on **I–V**.

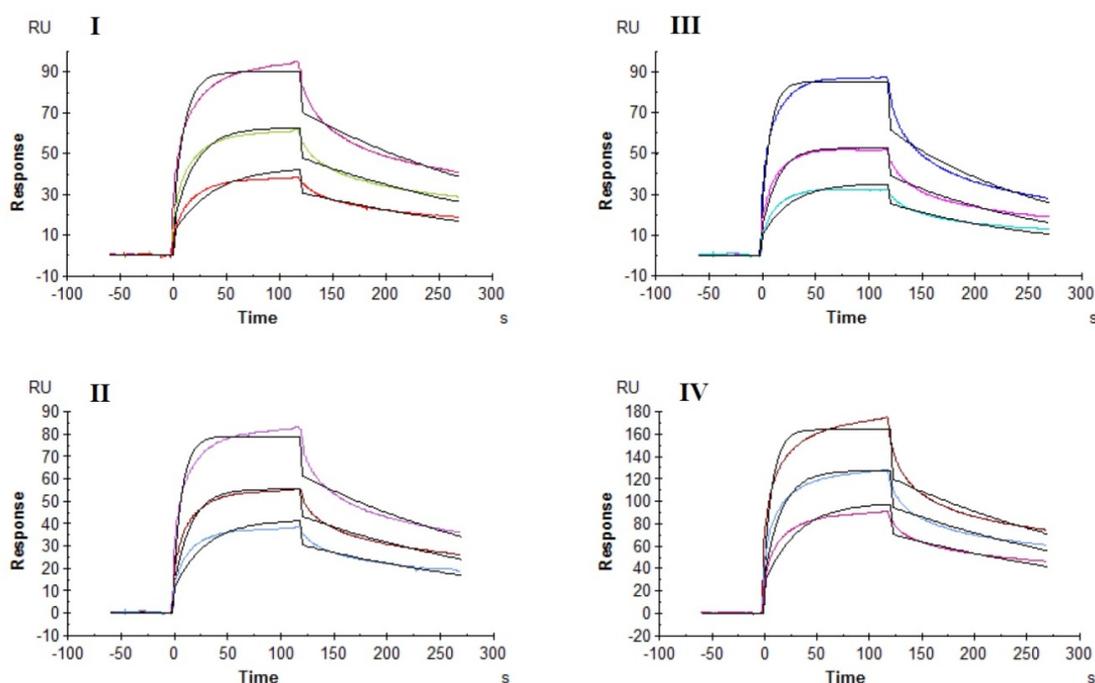
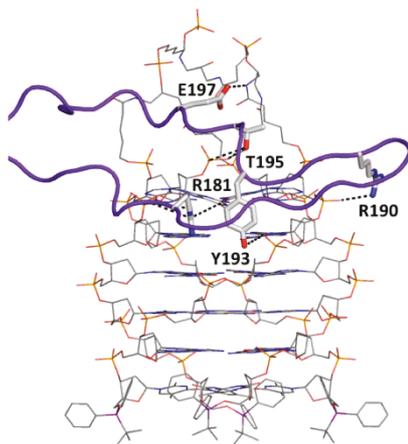


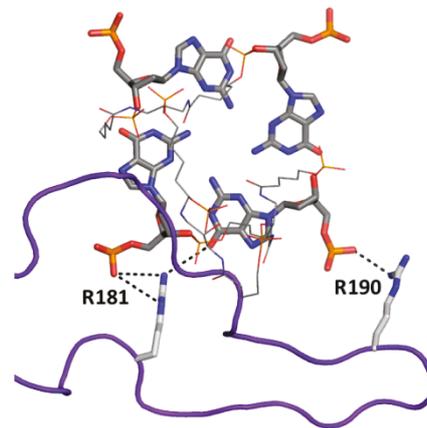
Figure 5. Kinetic analysis of interactions of compounds **I–IV** with recombinant HIV-(III_B) gp120 expressed in CHO cell cultures using SPR technology. The experimental data (coloured curves) were fit using the 1:1 binding model (black lines) to determine the kinetic parameters. The data are a representative example of two independent experiments.

It is well described that binding of agents (i.e. monoclonal antibodies) to the viral envelope does not necessarily result in efficient virus neutralization. Their effect on viral infectivity depends on the molecular epitope that is recognized on gp120 (20). Thus, it seems that **I** and **III** bind nearly as efficiently to gp120 as **II** and **IV** but neutralize about 3-fold less efficiently HIV-1 possibly due to subtle differences in epitope recognition. Alternatively, it cannot be excluded that the different TEL-quadruplexes have slightly different cellular uptake efficiencies or intracellular stability.

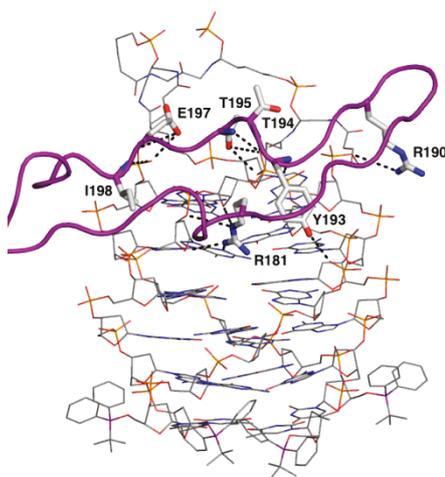
To obtain insight into the nature of the atomistic interactions between the TEL-aptamers **I–IV** and the V3 loop, we carried out molecular modelling studies by docking **I–IV** to the V3 loop of gp120 (PDB ID 2B4C). Our results revealed that (i) the V3 loop interacts with all aptamers with a similar binding mode involving the 3'-end tetrad and the TEL (Fig. 6), (ii) the type of nucleobases at the aptamer-V3 loop interface determines the chemical groups available for the interaction, and (iii) the number and type of interactions between the aptamer and the protein are responsible for the subtle differences in the binding energies (Table 2).



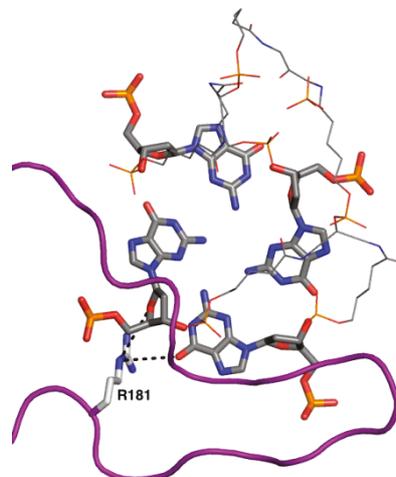
(TBDPS-5'TGGGAG3')₄-TEL (I)



G-tetrad at 3' terminus interactions



(TBDPS-5'TGGGGG3')₄-TEL (II)



G-tetrad at 3' terminus interactions

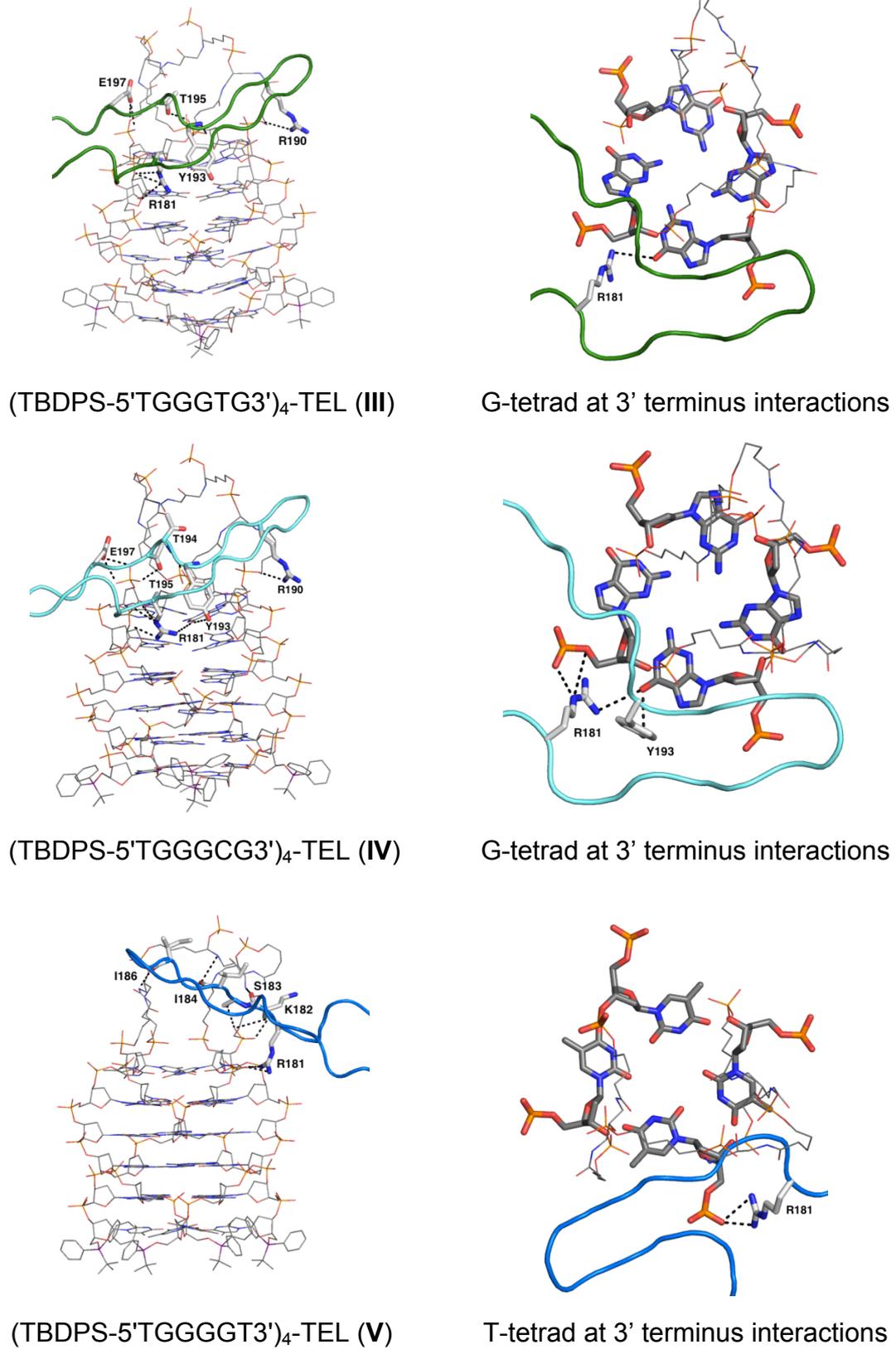


Figure 6. Molecular models of interactions between aptamers I-V and the V3 loop of gp120. The aptamers are drawn as lines, the V3 loop as cartoon and hydrogen bonding interactions between the V3 loop and the aptamers are drawn as dashed lines.

As reported by Honig and co-workers (21), charged residues are important in protein–DNA interactions. In this case, an important role is played by the side chain of residue R181 in the V3 loop that binds into the groove created by the phosphodiester backbone atoms (see Table 2). In particular, in **II** and **IV**, R181 makes multiple interactions with both phosphates and purine bases (G5 and G6 in **II**, G6 in **IV**). Furthermore, the side chains of R190, T195 and E197 and the nitrogen backbone atom of T194 of the V3 loop interact with the oxygen atoms of the TEL, giving additional stability to the complexes. In the **II**–V3 loop complex, the side chain of Y193 and the nitrogen backbone atom of Y193, T195 and I198 established additional interactions with the TEL, thereby resulting in lower interaction energy of this complex with respect to the other ones (Table 2). It should also be noted that when the side chain of Y193 is involved in the interaction (**I**, **II** and **IV**), the resulting complexes are found to have a better biological activity. Thus, the G5 nucleobase in **II** presents additional points for hydrogen bonding with the V3 loop. This is not seen in other bases. Therefore we infer that the differences in activity can arise from the thermal stability of the structures. Furthermore, this is also rationalised by the SPR experiments.

	G ₆ -O6	X ₅ -N3	G ₆ -OP3'	G ₆ -O3'/O4/O5'	T ₆ -OP3'	X ₅ -OP3'	X ₅ -O3'/O4/O5'	TEL-O	TEL-N
Arg181/NH	⊙		⊙◆	⊙⊙			⊙◆		
Arg181/NH2	⊙◆◆	⊙	◆		■		⊙⊙◆◆		
Lys182/N								■	
Lys182/O									■
Ser183/OH								■	
Ile184/O									■
Ile186/O									■
Arg190/NH2			◆					⊙⊙◆	
Tyr193/N								⊙⊙◆	
Tyr193/OH	⊙		◆			⊙			
Thr194/N								⊙⊙	
Thr195/N								⊙	
Thr195/OH								⊙⊙⊙◆◆◆	
Glu197/OH								⊙⊙⊙⊙◆◆◆	◆
Ile198/N								⊙	

Table 2. Table of interactions among the residues of V3 loop and **I** (◆), **II** (⊙), **III** (◆), **IV** (⊙) and **V** (■).

In order to better understand the structural features critical for the biological activity, we also carried out a molecular modelling study between the V3 loop and **V**, a quadruplex structure lacking marked anti-HIV activity. Our results revealed that the V3 loop interacts with **V** by using a different binding mode (Fig. 6). Differently from what was observed in the above-described complexes, **V** interacts with the V3 loop primarily via the TEL atoms and no atom of nucleobases is involved. Furthermore, except for R181, different residues of the V3 loop, such as K182, S183, I184, and I186, are involved in the interactions (Table 2). A plausible reason for the different binding mode of **V** is due to the presence of the T-tetrad at the 3'-terminus. As shown in Fig. 6, when there is a G-tetrad at the 3'-position, the oxygen atoms of guanines are involved in the interaction with the V3 loop through the side chains of R181 (and Y193 in the case of **IV**). In the **V**–V3 loop complex the methyl groups of the thymines are positioned in the groove formed by the phosphodiester backbone atoms, not allowing the formation of H bonds with R181 and Y193 of the V3 loop. All together, the structural evidence suggests that the T-tetrad at the 3'-position markedly affects the biological activity. In accordance with this finding and with the experimental data, the aptamer **II** showed the best docking score and the highest number of hydrogen bonds with the protein (Table 3).

Aptamer	Interactions energy	Number of H bonds
I	-24.11	8
II	-28.45	15
III	-13.74	9
IV	-28.09	11
V	-13.74	8

Table 3. Interaction energies of top docks (Kcal mol^{-1}) and number of H bonds between I–V and the V3 loop of gp120

2.2 Conclusions

The anti-HIV activity against HIV-1 and HIV-2 and the binding properties with HIV gp120 of a small library of new TEL-aptamers (II–IV) have been reported. Results from TEL-ODNs V and VI confirmed that the formation of a quadruplex species by the aptamer is required, but not sufficient to exert the anti-HIV activity. The docking data suggest that the interaction of the V3 loop with both the backbone and the TEL of the aptamers is required. Furthermore, the direct involvement of nucleobases in the interaction with the V3 loop gives additional stability to the complexes and results in a better biological activity. Overall, the here reported results expand our knowledge about anti-HIV G-quadruplexes and provide the rational basis for the design of novel anti-HIV aptamers with improved biological activity.

2.3 Experimental Session

2.3.1 General instrumentations

Automated solid-phase oligonucleotide synthesis was performed on a 8909 DNA-Synthesizer (Applied Biosystems). Mass spectra of TEL-ODNs were performed on a Bruker Autoflex I MALDI-TOF spectrometer using a picolinic/3-hydroxypicolinic acid mixture as the matrix. CD measurements were performed on a JASCO J-715 spectropolarimeter equipped with a Peltier Thermostat JASCO ETC-505T using 0.1 cm path length cuvettes. NMR spectra were acquired on Varian ^{UNITY}INOVA 500 MHz, 700 MHz and Mercury Plus 400 MHz spectrometers and processed using the Varian VNMR software package. For the experiments in H₂O, water suppression was achieved by including a double pulsed-field gradient spin-echo (DPFGSE) module (22, 23) in the pulse sequence prior to acquisition. Chemical shifts are reported in parts per million (δ) relative to the residual solvent signals.

2.3.2 Synthesis of 3'-O-tert-butylidiphenylsilyl-propyl-1'-O-(2-cyanoethyl)-N,N'-diisopropylphosphoramidite

1,3-propandiol (1 g, 13.1 mmol) dissolved in 5 mL of dimethylformamide (DMF) was treated with imidazole (1.07 g, 15.7 mmol) and TBDPS-chloride (3.24 g, 11.8 mmol) at 0°C under stirring at room temperature. After 2 h the reaction was quenched by addition of water and the resulting mixture was taken to dryness, dissolved in ethyl acetate (30 mL) and washed with water. The organic phase was dried over sodium sulphate, evaporated to oil and then purified by column chromatography (50 g Kieselgel 60H) eluted with esane/ethyl acetate 80:20 (v/v), giving 3.3 g (10.5 mmol, 80% yield) of the monosilylated 1,3-(O-TBDPS)-propandiol. ¹H-NMR (400 MHz; CD₃OD): δ 7.67–7.66 (m, 4H, Ph), 7.45–7.37 (m, 6H, Ph), 3.78 (t,

$J = 6.2$ Hz, 2H, CH₂O), 3.70 (t, $J = 6.4$ Hz, 2H, CH₂O), 1.79 (quintuplet, $J = 6.3$ Hz, 2H, CH₂), 1.03 (s, 9H, *tert*-Bu).

To the above product, dissolved in 20 mL of dry CH₂Cl₂, N,N-diisopropylethylamine (2.9 mL, 12.6 mmol) and 2-cyanoethyl-N,N'-diisopropylchlorophosphoramidite (2.81 mL, 12.6 mmol) were added and the mixture was stirred at r.t. for 1h. The mixture was diluted with ethyl acetate (100 mL) and washed with saturated aq. sodium chloride (3 x 50 mL). The organic phase, dried over sodium sulphate and evaporated to an oil, was purified by column chromatography (10 g Kieselgel 60H) eluted with *nesane*/ethyl acetate/lutidine 90:9.5:0.5 (v/v) to give 3'-O-*tert*-butyldiphenylsilyl-propyl-1'-O-(2-cyanoethyl)-N,N'-diisopropylphosphoramidite as a white powder (9.4 mmol, 90% yield).

¹H-NMR (400 MHz; acetone-d₆): δ 7.71-7.70 (m, 4H, Ph), 7.48-7.42 (m, 6H, Ph), 3.88-3.76 (complex signal, 6H, 3 x CH₂O), 3.67-3.58 (complex signal, 2H, 2 x CH), 2.71 (t, $J = 6.0$ Hz, 2H, CH₂CN), 1.88 (quintuplet, $J = 6.1$ Hz, 2H, CH₂), 1.18 (d, $J = 6.8$ Hz, 6H, *iso*Pr), 1.15 (d, $J = 6.8$ Hz, 6H, *i*Pr), 1.04 (s, 9H, *t*Bu); ³¹P-NMR (202 MHz; acetone-d₆): δ 150.4 (s); HRESIMS: m/z 537.2672 [M+Na]⁺, C₂₈H₄₃N₂NaO₃PSi, requires 537.2678.

2.3.3 Synthesis and MALDI-TOF spectra of TEL-ODNs

The syntheses of TEL-ODNs **I-V** were performed using a solid-phase automated DNA synthesizer according to the previously described procedure (16). In the case of TEL-ODN **VI**, in the last coupling step the 3'-O-*tert*-butyldiphenylsilyl-propyl-1'-O-(2-cyanoethyl)-N,N'-diisopropylphosphoramidite was used at 50 mg/mL in acetonitrile solution. After automated synthesis, the oligonucleotides were detached from the support and deprotected by using concentrated aqueous ammonia at 55 °C for 17 h. The combined filtrates and washings were dried, redissolved in water, analyzed and purified by HPLC on an anion exchange column (Macherey-Nagel, 1000-8/46, 4.4x50 mm, 5 μ m) using a linear gradient from 0 to 100% B in 30 min, flow rate = 1 mL/min and detection at 260 nm (buffer A: 20 mM NaH₂PO₄ aq. solution pH 7.0, containing 20% (v/v) CH₃CN; buffer B: 20 mM NaH₂PO₄ aq. solution pH 7.0, containing 1M NaCl and 20% (v/v) CH₃CN).

After HPLC purification, the oligonucleotide samples were desalted on a biogel column (BIORAD) eluted with H₂O/CH₃CH₂OH (9:1 v/v). The isolated oligomers were characterized by MALDI TOF mass:

I calculated: [M] = 9765.3; found [M+H]⁺ = 9764

II calculated: [M] = 9829.3; found [M+H]⁺ = 9830

III calculated: [M] = 9729.3; found [M+H]⁺ = 9730

IV calculated: [M] = 9669.3; found [M+H]⁺ = 9669

V calculated: [M] = 9729.3; found [M+H]⁺ = 9729

VI calculated: [M] = 10316.1; found [M+H]⁺ = 10318

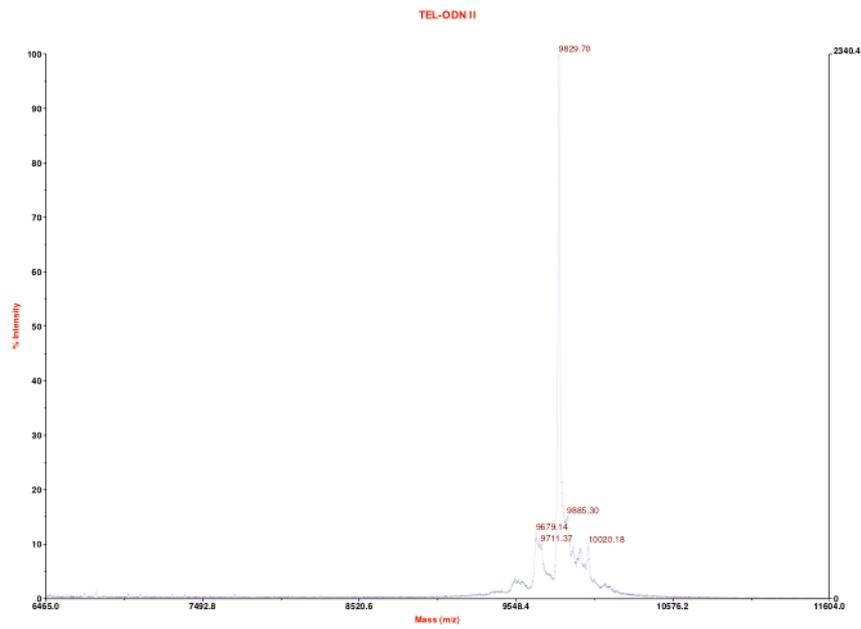
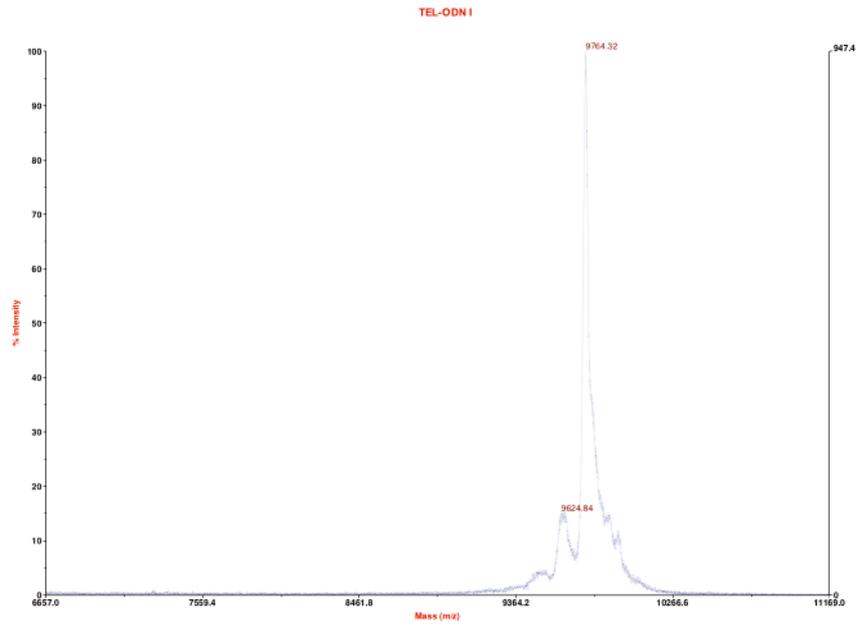


Figure 8. MALDI-TOF spectra (positive mode) of TEL-ODNs I-II.

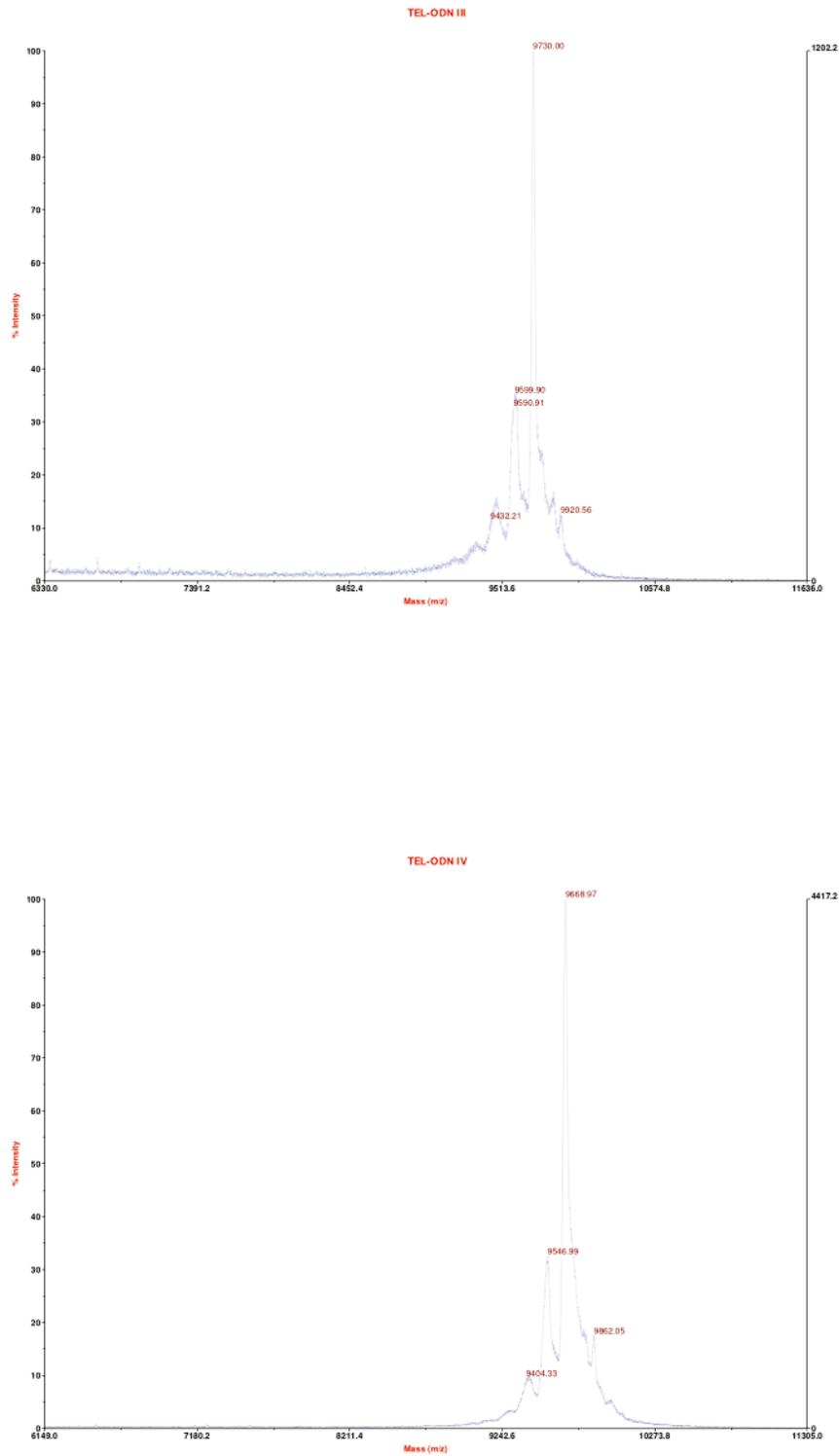


Figure 9. MALDI-TOF spectra (positive mode) of TEL-ODNs III-IV.

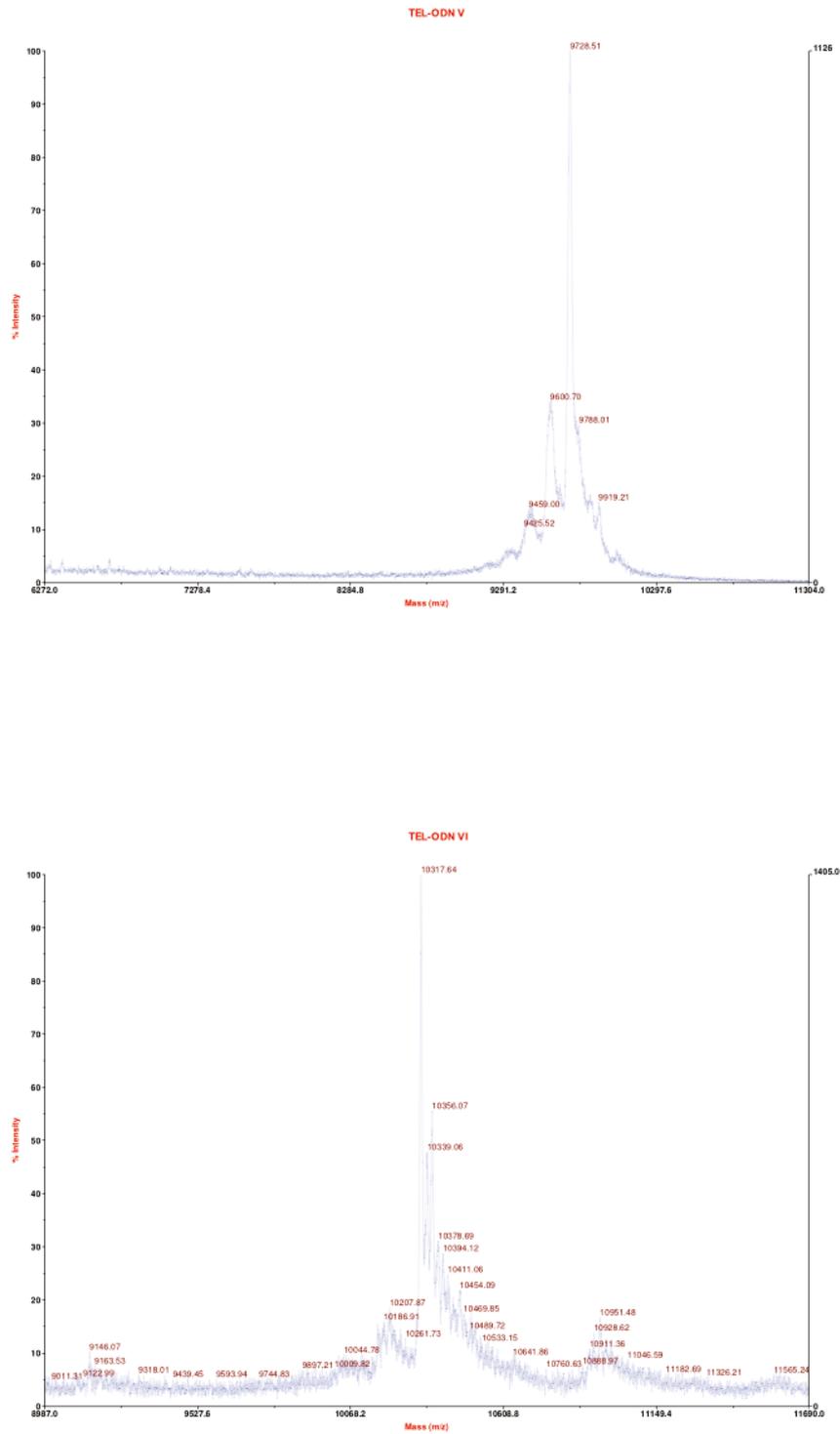


Figure 10. MALDI-TOF spectra (positive mode) of TEL-ODNs V-IV.

2.3.4 Preparation of quadruple helices (annealing)

TEL-ODN samples were dissolved in 100 mM potassium buffer (10 mM KH_2PO_4 aq. solution containing 90 mM KCl, pH 7.0) and the corresponding quadruplexes were formed by heating at 90 °C for 5 min and slowly cooling at room temperature for 12 h. Samples were stored at 4°C for 24 h before measurements. TEL-ODNs concentrations were determined in water by measuring the absorbance at 260 nm at 90 °C, using the nearest-neighbour calculated molar extinction coefficient of d[5'TGGGXG3'] (for **I-IV** and **VI**) or d[5'TGGGGT3'] (for **V**) multiplied by 4 (the number of strands in each TEL-ODN complex). A contribute of $8658 \text{ cm}^{-1}\text{M}^{-1}$ was added for each TBDPS group (24).

I: calculated $\epsilon = 284632 \text{ cm}^{-1}\text{M}^{-1}$

II: calculated $\epsilon = 272232 \text{ cm}^{-1}\text{M}^{-1}$

III: calculated $\epsilon = 266232 \text{ cm}^{-1}\text{M}^{-1}$

IV: calculated $\epsilon = 258232 \text{ cm}^{-1}\text{M}^{-1}$

V: calculated $\epsilon = 265832 \text{ cm}^{-1}\text{M}^{-1}$

VI: calculated $\epsilon = 284632 \text{ cm}^{-1}\text{M}^{-1}$

2.3.5 CD spectroscopy measurements

TEL-ODN samples ($2 \times 10^{-5} \text{ M}$) were annealed as described above and stored at 4°C overnight before measurements. CD spectra were recorded in the 360-200 nm wavelength range as an average of 5 scans (100 nm/min, 1 s response time, 1 nm bandwidth) and normalized by subtraction of the background scan containing only the buffer. Thermal denaturation experiments (Fig. 8) were carried out in the temperature range of 5-90 °C by monitoring CD values at 264 nm at a heating rate of 0.5 °C/min.

2.3.6 NMR spectroscopy measurements

The NMR samples were prepared in 0.2 ml of buffer (90% $\text{H}_2\text{O}/10\% \text{D}_2\text{O}$), containing 0.1 M K^+ at pH 7.0, and NMR spectra were recorded in the temperature range 25-90 °C on Varian ^{UNITY}INOVA 700 MHz spectrometer.

2.3.7 Biological evaluation assays

The methodology of the anti-HIV assays was as follows: $\sim 3 \times 10^5$ human CD_4 T-lymphocyte (CEM) cells were infected with 100 CCID_{50} of HIV(III_B) or HIV-2(ROD)/ml and seeded in 200 μL wells of a 96-well microtiter plate containing appropriate dilutions of the test compounds. After 4 days of incubation at 37°C, HIV-induced syncytia formation in the CEM cell cultures was examined microscopically. Data represent the mean (\pm SD) of 3 independent experiments.

2.3.8 Surface plasmon resonance (SPR) analysis

Recombinant gp120 protein from HIV-1(III_B) (ImmunoDiagnostics Inc., Woburn, MA) (produced by CHO cell cultures) was covalently immobilized on a CM5 sensor chip in 10 mM sodium acetate, pH 4.0, using standard amine coupling chemistry. The chip density was 2345 RU ($\sim 19.5 \text{ fmol}$ of gp120). A reference flow cell was used as a control for non-specific binding and refractive index changes. The experiment was performed at 25°C on a Biacore T200 instrument (GE Healthcare, Uppsala, Sweden).

The test compounds were serially diluted in HBS-P (10 mM HEPES, 150 mM NaCl and 0.05% surfactant P20; pH 7.4), covering a concentration range between 1.17 and 4.69 μM . Samples were injected for 120 sec at a flow rate of 30 $\mu\text{l}/\text{min}$ and the dissociation was followed for 180 sec. One duplicate sample and several buffer blanks were used as a control and as double referencing, respectively. The CM5 sensor chip surface was regenerated with 1 injection of 50 mM NaOH. The experimental data were fit using the 1:1 binding model (Biacore T200 Evaluation software 1.0) to determine the binding kinetics.

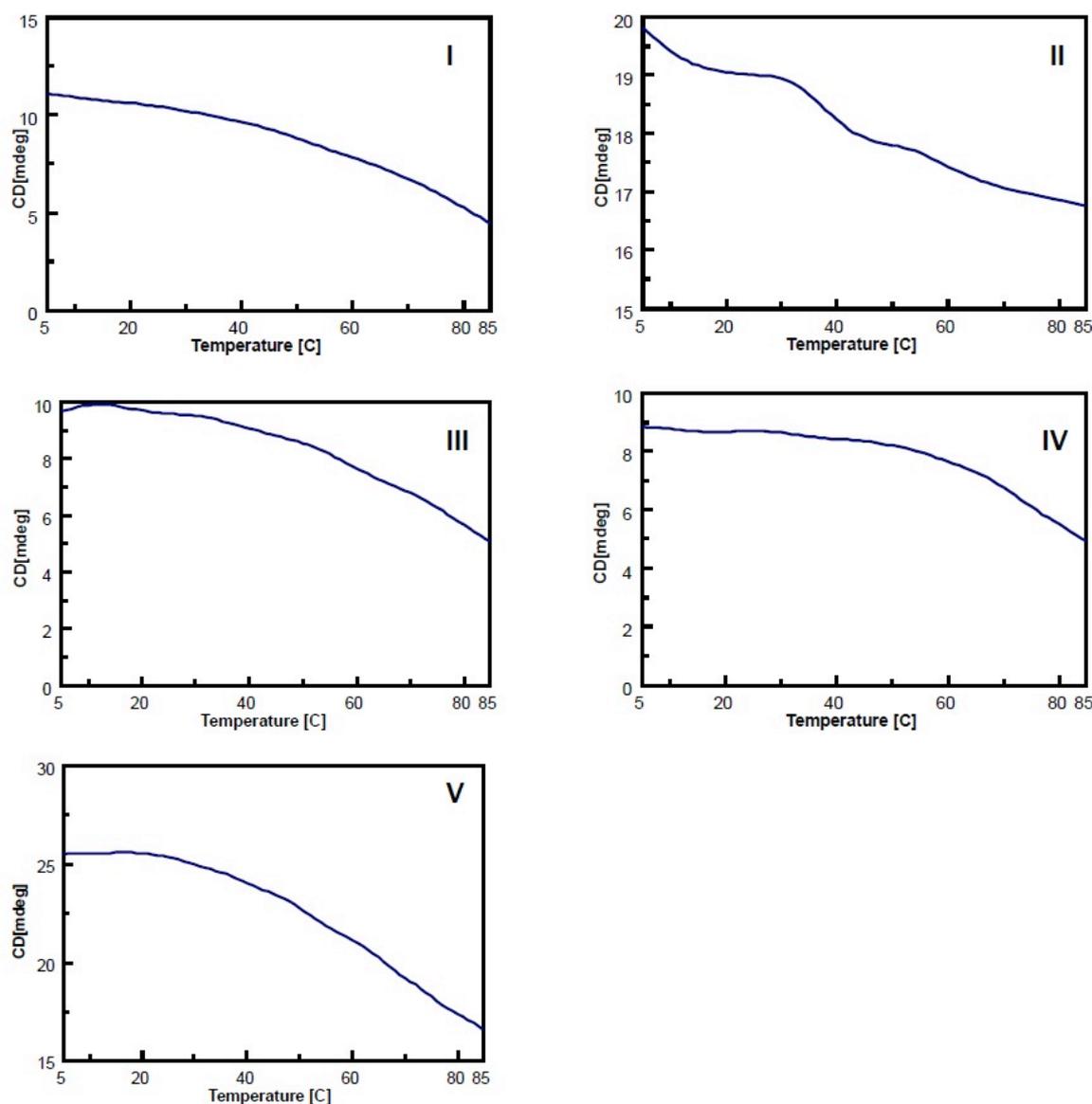


Figure 11: CD-melting curves of I-V recorded at 264 nm in 100 mM K^+ buffer.

2.3.9 Docking protocol

The molecular models of TEL-ODNs were generated using the InsightII suite program (Accelrys, Inc), by employing the Biopolymer module, to make the quadruplexes, and the Builder module, to add the modified linkers at 3'- and 5'-ends. The molecular models were parameterized according to the AMBER force field and

then brought to their energetically minimized structures with Discovery3 module. The minimizations were performed using 1000 cycles of steepest descent method followed by conjugate gradient calculation until convergence at 0.1 rms was reached. A distance-dependent macroscopic dielectric constant of 4.0 and an infinite cut-off for non-bonded interactions were used to partially compensate for the lack of solvent. The Protein Data Bank was used to download the HIV gp120 protein (PDB ID 2B4C). All dockings between the V3 loop in gp120 and TEL-ODN were performed using the ZDOCK server (zdock.umassmed.edu), that evaluates Pairwise shape complementarity, desolvation and electrostatic energies using Fast Fourier Transform algorithm to search multiple binding possibilities for receptor and ligand. The output from the program is a list of transformation values for the docks, from which models can be generated. For each docking, a total of 2000 transformations were calculated, ranked based on internal scoring functions and presented through a global energy term. The best solution from each dock was chosen and further minimised to relieve any steric clashes that might have arisen in docking. No atoms were restrained during the minimization process. The minimized models were then visualized using the Pymol software and the electrostatic interactions between protein and aptamer were defined.

2.4 References

- (1) P. D. Kwong, R. W., J. Robinson, R. W. Sweet, J. Sodroski, W. A. Hendrickson. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648–659.
- (2) R. Wyatt, J. S. (1998) The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science* 280, 1884–1888.
- (3) C. C. Huang, M. T., M.-Y. Zhang, S. Majeed, E. Montabana, R. L. Stanfield, D. S. Dimitrov, B. Korber, J. Sodroski, I. A. Wilson, R. Wyatt, P. D. Kwong. (2005) Structure of a V3-Containing HIV-1 gp120 Core. *Science* 310, 1025–1028.
- (4) S. Chou, K. C. (2005) DNA Aptamers as Potential Anti-HIV Agents. *Trends in Biochemical Science* 30, 231–234.
- (5) A. C. Perkins, S. M. (2007) Radiolabelled aptamers for tumour imaging and therapy. *The Quarterly Journal of Nuclear Medicine and Molecular Imaging* 51, 292–296.
- (6) A. D. Keefe, S. P., A. Ellington. (2010) Aptamers as therapeutics. *Nature Reviews Drug Discovery* 9, 537-550.
- (7) J. Lee, G. S., A. Ellington. (2006) Aptamer therapeutics advance. *Current Opinion in Chemical Biology* 10, 282–289.
- (8) H. Hotoda, K. M., H. Furukawa, T. Nakamura, M. Kaneko, S. Kimura, K. Shimada. (1994) Biologically Active Oligodeoxyribonucleotides - II: Structure Activity Relationships of Anti-HIV-1 Pentadecadeoxyribonucleotides Bearing 5'-End-Modifications. *Nucleosides & Nucleotides* 13, 1375-1395.
- (9) J. R. Wyatt, T. A. V., J. L. Roberson, R. W. Buckheit, T. Klimkait, E. DeBaets, P. W. Davis, B. Rayner, J. L. Imbach, D. J. Ecker. (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. *Proceedings of the National Academy of Sciences* 91, 1356–1360.
- (10) H. Furukawa, K. M., T. Agatsuma, I. Yamamoto, S. Kimura, K. Shimada. (1997) Identification of a phosphodiester hexanucleotide that inhibits HIV-1

- infection in vitro on covalent linkage of its 5'-end with a dimethoxytrityl residue. *Antisense & Nucleic Acid Drug Development* 7, 167–175.
- (11) M. Koizumi, R. K., H. Hotoda, K. Momota, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura, K. Shimada. (1997) Biologically active oligodeoxyribonucleotides—IX. Synthesis and anti-HIV-1 activity of hexadeoxyribonucleotides, TGGGAG, bearing 3'- and 5'-end-modification. *Bioorganic & Medicinal Chemistry* 5, 2235–2243.
 - (12) H. Hotoda, M. K., R. Koga, M. Kaneko, K. Momota, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, J. Sone, S. Tsutsumi, T. Kosaka, K. Abe, S. Kimura, K. Shimada. (1998) Biologically Active Oligodeoxyribonucleotides. 5. 5'-End-Substituted d(TGGGAG) Possesses Anti-Human Immunodeficiency Virus Type 1 Activity by Forming a G-Quadruplex Structure. *Journal of Medicinal Chemistry* 41, 3655–3663.
 - (13) M. Koizumi, R. K., H. Hotoda, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura, K. Shimada. (1998) Biologically active oligodeoxyribonucleotides. Part 11: The least phosphate-modification of quadruplex-forming hexadeoxyribonucleotide TGGGAG, bearing 3'- and 5'-end-modification, with anti-HIV-1 activity. *Bioorganic & Medicinal Chemistry* 6, 2469–2475.
 - (14) G. Oliviero, N. B., A. Galeone, M. Varra, G. Piccialli, L. Mayol. (2004) Synthesis and characterization of a bunchy oligonucleotide forming a monomolecular parallel quadruplex structure in solution. *Tetrahedron Letters* 45, 4869–4872.
 - (15) G. Oliviero, J. A., N. Borbone, A. Galeone, M. Varra, G. Piccialli, L. Mayol. (2006) Synthesis and characterization of DNA quadruplexes containing T-tetrads formed by bunch-oligonucleotides. *Biopolymers* 81, 194–201.
 - (16) G. Oliviero, J. A., N. Borbone, S. D'Errico, A. Galeone, L. Mayol, S. Haider, O. Olubiyi, B. Hoorelbeke, J. Balzarini, G. Piccialli. (2010) Tetra-end-linked oligonucleotides forming DNA G-quadruplexes : a new class of aptamers showing anti-HIV activity. *Chemical Communications* 46, 8971–8973.
 - (17) J.-L. Mergny, A. D. C., A. Ghelab, B. Saccà, L. Lacroix. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Research* 33, 81–94.
 - (18) S. Paramasivan, I. R., P. H. Bolton. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods* 43, 324–331.
 - (19) S. Masiero, R. T., S. Pieraccini, S. De Tito, R. Perone, A. Randazzo, G. P. Spada. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Organic & Biomolecular Chemistry* 8, 2653–2872.
 - (20) L. M. Walker, M. H., K. J. Doores, E. Falkowska, R. Pejchal, J. P. Julien, S. K. Wang, A. Ramos, P. Y. Chan-Hui, M. Moyle, J. L. Mitcham, P. W. Hammond, O. A. Olsen, P. Phung, S. Fling, C. H. Wong, S. Phogat, T. Wrin, M. D. Simek, W. C. Koff, I. A. Wilson, D. R. Burton, P. Poignard. (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477, 466–470.
 - (21) R. Rohs, S. M. W., A. Sosinsky, P. Liu, R. S. Mann, B. Honig. (2009) The role of DNA shape in protein-DNA recognition. *Nature* 461, 1248–1253.
 - (22) T. L. Hwang, A. J. S. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* 112, 275–279.

- (23) Dalvit, C. (1998) Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *Journal of Biomolecular NMR* 11, 437-444.
- (24) D'Onofrio, L. P., E. Erra, L. Martino, G. Di Fabio, L. De Napoli, C. Giancola, D. Montesarchio. (2007) 5'-Modified G-Quadruplex Forming Oligonucleotides Endowed with Anti-HIV Activity: Synthesis and Biophysical Properties. *Bioconjugate Chemistry* 18, 1194–1204.

Chapter 3

Thrombin-Binding Aptamer analogs with enhanced anticoagulant activity: the role of T₇ and T₁₂ residues on biological properties

Thrombin is a sodium-activated allosteric enzyme playing a key role in blood coagulation. (1, 2) The complexity of thrombin function and regulation has gained this enzyme pre-eminence as the prototypic allosteric serine protease. (3-7) It clots blood by converting fibrinogen into fibrin, by activating factors V, VIII, and XIII, the latter being responsible for the cross-linking of fibrin fibers, and by promoting platelet aggregation. Because activation of factor V is required by activated factor X to cut prothrombin into thrombin, the synthesis of thrombin is in part modulated by its own blood concentration, thus providing a fast response to injury. Besides procoagulant stimuli, also anticoagulant stimuli can be triggered by thrombin via activation of protein C, under the allosteric control of the cofactor thrombomodulin. Moreover, thrombin inactivation and clearance in plasma can also be achieved by interaction with the serpin antithrombin and with the heparin cofactor II. Due to its central role in the coagulation cascade, malfunctions in the regulatory mechanism of thrombin activity cause pathological states such as haemorrhage or abnormal clot growth. Thrombosis and connected diseases are among the main causes of mortality in Western countries; (8, 9) thus, the discovery of molecules capable of modulating thrombin activity represents a major target for the development of anticoagulant strategies. (10) Aptamer technology has been efficiently employed to obtain new direct thrombin inhibitors by selecting thrombin-binding oligonucleotides. (11, 12) The term aptamer generally refers to a single-stranded oligonucleotide that binds to a selected protein and specifically inhibits one or more of its functions. The first reported consensus sequence able to inhibit thrombin activity was the 15-mer oligonucleotide (ON) GGTTGGTGTGGTTGG, usually known with the acronym TBA (thrombin-binding aptamer). (12) In the presence of thrombin and/or monovalent cations TBA folds into a specific three-dimensional structure that dictates its thrombin-binding affinity. The structures of TBA alone and in complex with thrombin were determined by NMR (13-15) and X-ray (16-18) methods, respectively. In all experimentally determined structures TBA adopted a monomolecular chair-like G-quadruplex folding topology, consisting of two G-tetrads connected by one TGT loop and two TT loops. First, (17, 18) different TBA–thrombin X-ray complexes were reported, in which TBA adopted either the same folding as derived by NMR (PDB ID 1HAO; Figure 1A) or a different one (PDB IDs 1HUT and 1HAP; Figure 1B), presenting a diverse positioning of the connecting loops. In these X-ray complexes, the aptamer interacts with distinct regions of two thrombin molecules, (i) the fibrinogen exosite (namely anion-binding exosite I, ABE I) and (ii) the heparin exosite (namely, ABE II) near the carboxylate terminal helix of a neighbouring thrombin. Thus, according to the adopted folding and the relative orientation with respect to thrombin, TBA bound to ABE I by using the TT loops (PDB ID 1HAO; Figure 1A) or the TGT loop (PDB IDs 1HUT and 1HAP; Figure 1B). A subsequent re-evaluation (19) of the diffraction data of TBA–thrombin complexes (17, 18) evidenced that the NMR-derived folding (Figure 1A) could fit all diffraction data if alternative aptamer orientations with respect to thrombin were considered. Indeed, according to the electron density maps, it was assumed that the oligonucleotide quartet region (including the DNA backbone) was correct as reported in the crystal structures, whereas, due to the D₂ symmetry of the aptamer core, there are four distinct

possible orientations of the NMR folding that overlap with the core of the X-ray model (Figure 2).

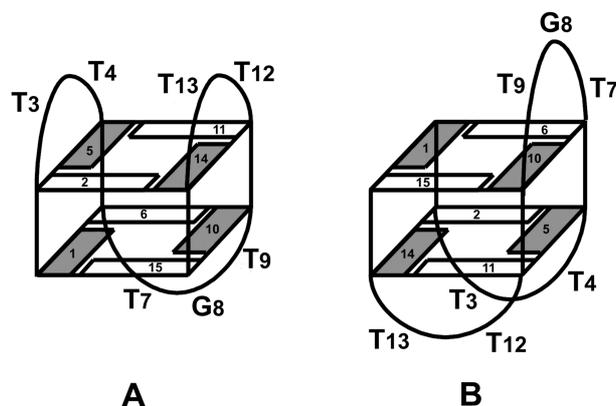


Figure 1. Schematic drawings of the X-ray derived G-quadruplex structures of TBA as reported in PDB IDs 1HAO (A) and 1HUT and 1HAP (B). In (A) TT loops span narrow grooves and TGT spans wide grooves, vice versa in (B). White and grey squares indicate G bases in anti and syn conformation, respectively.

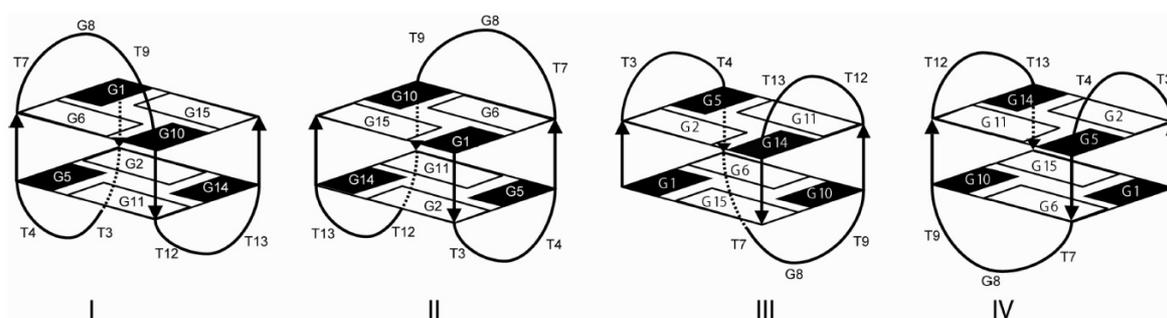


Figure 2. Representation of the four orientations of the NMR derived folding with respect to thrombin corresponding to the D2 symmetry of the quadruplex core assuming the same strand polarity of the X-ray complex.

The re-evaluation of the X-ray complex gave similar results among orientations I–IV, but orientation III showed the best agreement with the experimental data. (19) Accordingly, different TBA–thrombin complexes in which the aptamer can bind ABE I through the TGT loop and ABE II through the TT loops (Figure 2, orientations I and II), or the other way around (Figure 2, orientations III and IV), are possible. Supporting this view, the newly released X-ray structures of thrombin–TBA–Na⁺ and thrombin–TBA–K⁺ complexes (PDB IDs 4DIH and 4DII, respectively) (16) showed the aptamer interacting with the enzyme assuming orientation IV (Figure 2). Moreover, the X-ray structure of human thrombin in complex with a modified TBA containing a 5'-5' inversion of the polarity site (3'GGT5'-5'TGGGTGTGGTTGG3', namely mTBA) has also been reported (PDB ID 3QLP). (20) The X-ray complex revealed that the interaction occurs between the TT loops and ABE I (orientation III in Figure 2), whereas the TGT loop, particularly T₇, is not involved in any relevant interaction with the protein. Despite the fact that mTBA binds to thrombin with higher affinity with respect to TBA, (21) it showed poor anticoagulant activity if compared to TBA in PT assay. (22) On the other hand, Toggle-25t, an RNA aptamer that contains 2'-fluoropyrimidine nucleotides, and a 29-mer single-stranded DNA, designated DNA60-18 or HD22, bind selectively thrombin at ABE II and they also showed limited effect on clotting times. Because of their thrombin allostery, ABE I and ABE II

aptamers were used in combination to test their mixed effect on thrombin activity. (10, 23-26) The obtained results showed that synergistic anticoagulant effects could be achieved by mixing TBA with Toggle-25t or HD22 and by linking TBA to HD22 with an appropriate-sized spacer.

In this scenario, due to the symmetry of TBA and the complexity of the regulatory mechanism of thrombin, which includes a long-range allosteric linkage between ABE I and ABE II, (27, 28) the exact binding mode of aptamers to thrombin and its correlation with the observed biological activity is still a matter of debate. Nevertheless, biological results indicated that TBA exerts its anticoagulant activity mainly competing with fibrinogen at ABE I. (29) The phase I clinical trial for TBA demonstrated its positive pharmacokinetic profile in humans; however, clinical trials were halted after phase I due to suboptimal dosing profiles. (30) To optimize its anticoagulant properties, of particular interest is the enlargement of data gathering about TBA structure–activity relationships (SARs) through the development of new derivatives. In the nucleic acids research area the introduction of suitable modifications on natural oligonucleotide sequences represents a method widely employed to explore the relevance of single nucleotides both on secondary structure folding topology and on binding with target proteins. The explored modifications involve the nucleobases, (31, 32) the sugar–phosphate backbone, (33, 34) and the conjugation to flexible non-nucleotide linkers. (35-37) In our understanding of the bases of TBA–thrombin recognition, we previously developed new single-stranded TBA analogues, in which the acyclic nucleoside **a** (Figure 3) replaced, one at a time, all thymine residues along the aptamer sequence. (38)

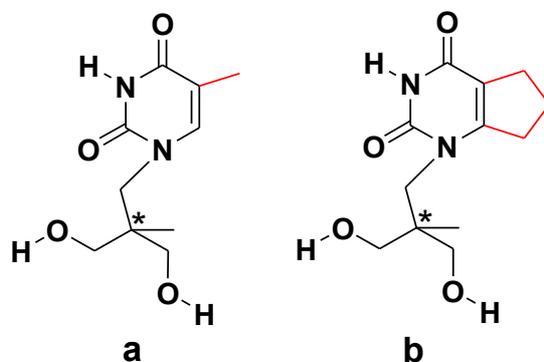


Figure 3. Acyclic nucleosides mimicking T. The presence of the carbon atom marked with an asterisk causes the formation of two diastereomeric ONs when the nucleoside is inserted in the TBA sequence.

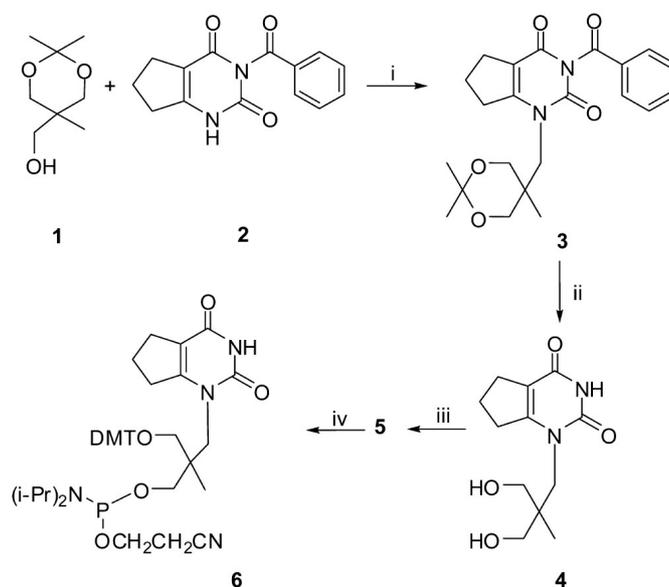
The modification at position 7 in the $T_7G_8T_9$ loop (herein TBA- T_7 **a**) gives rise to an ON that folds into a quadruplex structure that is more stable and active than the quadruplex structure formed by TBA. On the other hand, within the four ONs containing the acyclic nucleoside **a** at one position of TT loops, the ON modified at T_{12} (herein TBA- T_{12} **a**) showed the most significant biological properties. (38) In line with data reported from other authors, (39) our results evidenced that modification at T_7 and T_{12} along the TBA sequence can be fruitful to improve the biological activity. To investigate the role of the modification at T_7 and T_{12} in affecting the structural stability and the biological activity, we synthesized two new modified TBAs, named TBA- T_7 **b** and TBA- T_{12} **b**, which differed from the previously reported analogues only in the thymidine base coupled with the acyclic linker, which was bulked by fusing a

hydrophobic five member cycle with the pyrimidine ring (**b**, Figure 3). The abilities of TBA-T₇**b** and TBA-T₁₂**b** to fold into G-quadruplex structures, as well as their biological properties, were evaluated by CD and ¹H NMR experiments and by PT and purified fibrinogen clotting assays, respectively. Finally, a computational study allowed the analysis of the obtained SARs.

3.1 Results

3.1.1 Synthesis of Monomer **6**.

The synthesis of the modified acyclic nucleoside phosphoramidite building block **6**, used for the automated synthesis of TBA-T₇**b** and TBA-T₁₂**b**, is summarized in Scheme 1. The synthesis of N1-alkylated intermediate **3** was performed via Mitsunobu reaction (40) between **2**, obtained as described under Experimental Section, and **1**. The yields of this coupling reaction were strongly dependent on the initial temperature of the system and on the order in which the reagents were added. We found that temperatures lower than 30 °C were always detrimental to the reaction yield, and that higher yields (45–50%) were obtained by mild heating (30 °C) of a dioxane solution containing **2** and triphenylphosphine followed by addition of di-*tert*-butylazodicarboxylate and finally a dioxane solution of protected linker **1** (100 μL of solvent per 1 mmol). The intermediate **4** was obtained by removing both the linker and nucleobase protecting groups from **3** using a one-pot procedure (see Experimental Section). Finally, **4** was functionalized to phosphoramidite building block **6** using standard procedure. The acyclic nucleoside **6** was inserted as a T mimic at position 7 or 12 along the TBA sequence to obtain TBA-T₇**b** and TBA-T₁₂**b**, respectively.



Scheme 1. Synthesis of Monomer **6**: (i) **1** 0.94 g (5.9 mmol), **2** 1.50 g (5.9 mmol), triphenylphosphine 2.28 g (8.7 mmol), di-*tert*-butylazodicarboxylate 2.1 g (9.1 mmol), dry dioxane (70 mL), 24 h, 30°C, yields 43%; (ii) (a) **3** 1.0 g (2.5 mmol), Dowex H⁺ 450 mg, MeOH/H₂O (9:1 v/v) 100 mL; (b) aqueous NaOH (1.0 M, 5 mL) yields 98%; (iii) **4** 640 mg (2.5 mmol), 4,4'-dimethoxytrytylchloride 541 mg (1.6 mmol), 2,4-dimethylaminopyridine 15 mg (0.12 mmol), pyridine (20 mL), acetonitrile (10 mL), 1.5 h, rt, yields 45%; (iv) **5** 600 mg (1.1 mmol), 2-cyanoethyl-diisopropylchlorophosphoramidite 300 μL (1.2 mmol), DIPEA 600 μL (3.6 mmol), DCM (8 mL), 40 min, rt, yields 99%.

3.1.2 Structural Characterization

TBA-T₇**b** and TBA-T₁₂**b** were characterized by circular dichroism (CD), CD melting, and ¹H NMR for their ability to fold into G-quadruplexes. The introduction of the new acyclic nucleoside **b** at position 7 or 12 does not affect the overall G-quadruplex topology formed in solution. CD profiles of TBA-T₇**b** and TBA-T₁₂**b** are similar to those of TBA in both PBS and K⁺ buffer (left and right panels, respectively, in Figure 4), showing maxima at about 208, 245, and 295 nm and a minimum at 267 nm.

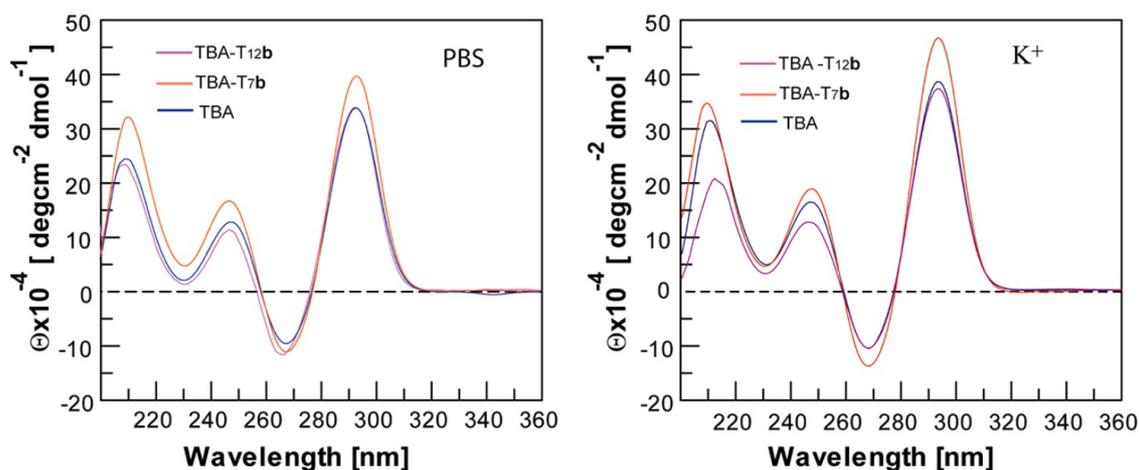


Figure 4. CD profiles in PBS (left) and in potassium phosphate buffer (right); [ON] = 2.0×10^{-5} M.

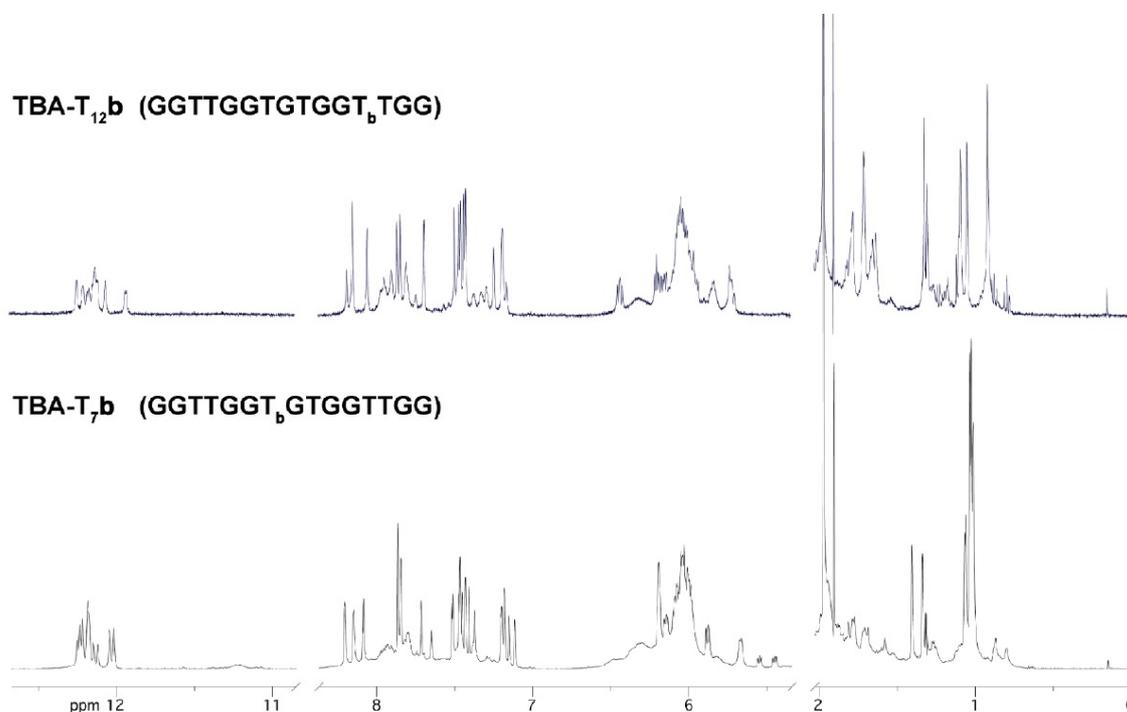


Figure 5. ¹H NMR spectra of TBA-T₇**b** and TBA-T₁₂**b** recorded at 25 °C in K⁺ buffer.

As expected, positive CD bands have higher molar ellipticities in K^+ buffer than in PBS, according to the well-known dependence of G-quadruplex stabilities from the type of cations present in the buffer solution. (41, 42) The 1H NMR analysis, performed at temperatures in the range of 25–50 °C, confirmed that TBA-T₇**b** and TBA-T₁₂**b** form a stable G-quadruplex structure when annealed in K^+ buffer. However, 1H NMR spectra appear more crowded than expected for a single quadruplex species, even at 25 °C (Figure 5), where only 14 aromatic signals and 8 imino proton signals are expected. This finding was previously explained by considering that the replacement of a T residue with the acyclic nucleoside **a** produces two closely related diastereomeric G-quadruplex forming ONs. (38) Although both modified ONs show T_m values higher than that of TBA (Figure 6 and Table 1), the replacement of T₇ with the acyclic nucleoside **b** stabilizes the resulting G-quadruplexes more efficiently than replacement of T₁₂.

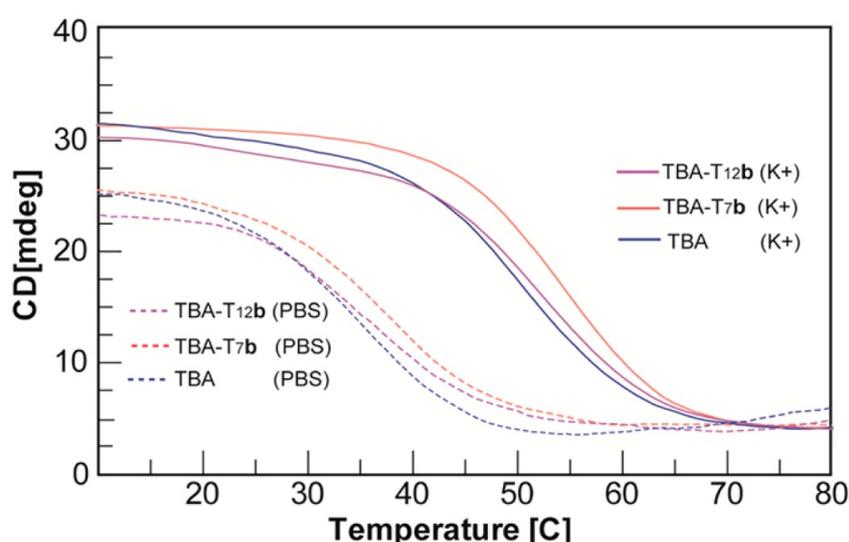


Figure 6. CD melting curves of TBA and its analogues in potassium phosphate (continuous lines) or PBS (dotted lines) buffer. Concentration of each ON was 2.0×10^{-5} M. The curves were obtained by monitoring the variation of absorbance at 295 nm from 10 to 80 °C at 0.5 °C/min.

ON	T_m (± 0.5 °C)	
	PBS	K^+ buffer
TBA	33.0	50.0
TBA-T ₁₂ a	34.0 ^a	51.0 ^a
TBA-T ₁₂ b	34.5	51.5
TBA-T ₇ a	37.5 ^a	54.0 ^a
TBA-T ₇ b	37.5	54.0

^aData taken from ref 38

Table 1. T_m values of TBA and its analogues.

3.1.3 PT Assay

The anticoagulant activity in the presence of all thrombin substrates and cofactors was evaluated by PT assay. PT analyses were performed at [ON] of 2 and 20 μ M after 2 min of incubation with human plasma (Figure 7). To better understand the effects induced by the presence of nucleoside **b** on the anticoagulant activity, we compared the prolonging of clotting time caused by TBA-T₇**b** and TBA-T₁₂**b** with that of TBA and those of the previously reported (38) analogues TBA-T₇**a** and TBA-T₁₂**a** (Table 2). It is noteworthy that TBA-T₇**b** prolonged the basal PT value to a significantly larger extent with respect to TBA and TBA-T₇**a** (Figure 7 and Table 2).

Replacement of nucleoside **a** for **b** was more effective also at position 12 (Table 2). In particular, TBA-T₁₂**b** is less active than TBA at 2 μ M, but showed a greater increase of activity at higher concentration, thus attaining that of TBA at 20 μ M (Figure 7). To exclude that the observed activity of modified aptamers could be due to higher nuclease stability with respect to TBA, we repeated the experiments, changing the incubation time of TBA and analogues with plasma from 30 s to 15 min. PT results were almost unchanged for all sequences (data not shown).

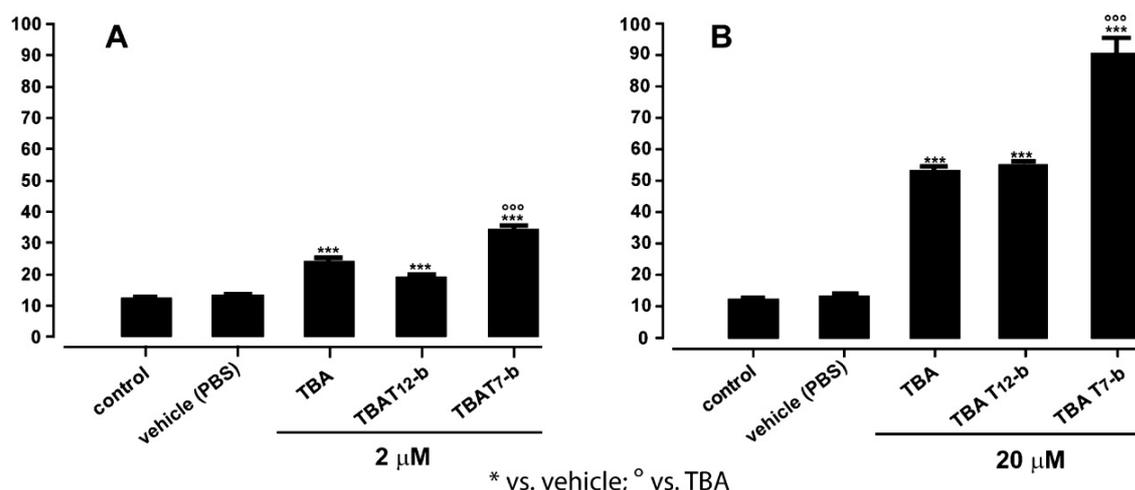


Figure 7. Concentration-dependent response, following 2 min ON incubation with human plasma at 2 (A) or 20 μ M (B) [ON]. PT values are expressed in seconds. Each measurement was performed in triplicate and is shown as the mean \pm SEM. The basal PT time is 13.4 \pm 0.2 s.

ON	PT value ^a (20 μ M)	Fold increase
TBA	53.68 \pm 1.48	4.01 \pm 0.70
TBA-T ₇ a	60.20 \pm 1.38 ^b	4.50 \pm 0.60
TBA-T ₇ b	90.60 \pm 5.26	6.77 \pm 0.24
TBA-T ₁₂ a	44.83 \pm 1.29 ^b	3.35 \pm 0.60
TBA-T ₁₂ b	55.20 \pm 0.47	4.12 \pm 0.20

^aPT values are expressed in seconds. ^bData taken from ref 38

Table 2. PT values measured after 2 min of incubation using 20 μ M ON concentration and fold increases of basal PT time (13.4 \pm 0.2 s).

3.1.4 Fibrinogen Assay Using Human and Bovine Thrombin

The new ONs were subjected to purified fibrinogen clotting assay to evaluate their ability to inhibit the conversion of soluble fibrinogen into insoluble strands of fibrin in the absence of any other thrombin ligands/ effectors (Figure 8 and Tables 3 and 4). The assay was performed using various concentrations of each aptamer. The purified fibrinogen solution was preincubated with the aptamer and the reactions were initiated by the addition of human thrombin and clotting times were recorded. The fibrinogen assay results evidenced that at all ON concentrations tested, the most efficient inhibitor is the unmodified TBA sequence (Figure 8A and Table 3), thus reversing the results gathered from the PT assay. To further investigate this

phenomenon, we performed the fibrinogen assay by using bovine thrombin, which differs from human thrombin in some residues of the fibrinogen-binding site (ABE I) that are crucial for TBA–thrombin interaction (see next paragraphs for details). By changing the source of the target enzyme, both modified sequences showed a substantial increase in their ability to inhibit thrombin activity (Figure 8B and Table 4), whereas the inhibitory activity of TBA was only slightly increased; as a consequence, TBA-T₇b became again the best inhibitor.

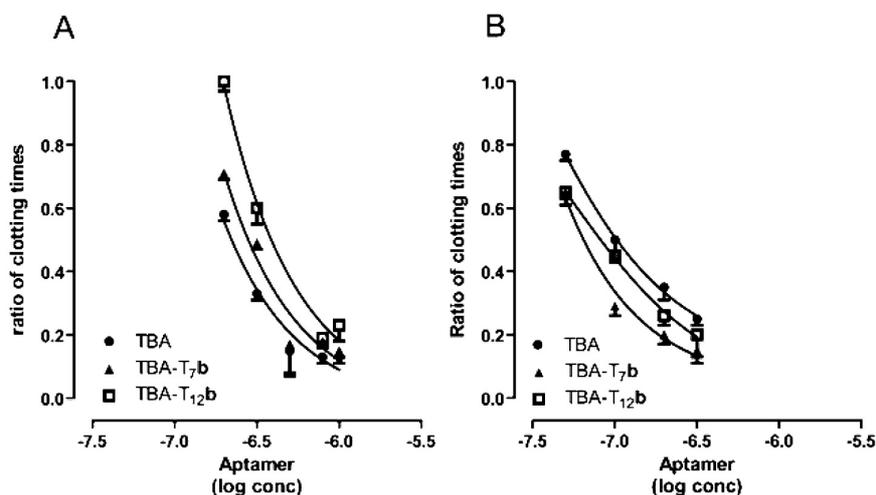


Figure 8. Effects of aptamers on human (A) and bovine (B) thrombin clotting time in fibrinogen solution. The results are expressed as the ratio of clotting times measured in the absence and in the presence of the aptamers. The effects are reported as a function of log [ON]. Each measurement was performed in triplicate and is shown as the mean \pm SD.

ON	Prolonged clotting time (s)				
	200 nM	300 nM	500 nM	750 nM	1000 nM
TBA	18 \pm 2.7	50 \pm 2.3	137 \pm 6.3	170 \pm 24	170 \pm 25
TBA-T ₇ b	10 \pm 2.5	26 \pm 2.3	120 \pm 6.0	120 \pm 22	140 \pm 27
TBA-T ₁₂ b	NA ^a	16.8 \pm 0.71	98.9 \pm 0.7	100 \pm 30	100 \pm 10

^aNot active

Table 3. Prolonged fibrinogen clotting time results in presence of human thrombin. Each value was calculated subtracting the clotting time produced by thrombin alone from that measured in the presence of one of the aptamers. The measurements were performed preincubating for 1 min 200 μ L of purified fibrinogen solution (2.0 mg/mL in PBS) with various concentrations of aptamers. The reaction started by the addition of 100 μ L of human thrombin solution (10 NIH *per* mL) and clotting time was recorded. In absence of any inhibitor, the measured clotting time for human thrombin was 24.9 \pm 0.5 s.

ON	Prolonged clotting time (s)			
	50 nM	100 nM	200 nM	300 nM
TBA	4 \pm 1.23	18 \pm 1.5	36 \pm 1.8	57 \pm 3.7
TBA-T ₇ b	9 \pm 3.3	50 \pm 12	80 \pm 12	120 \pm 16
TBA-T ₁₂ b	9 \pm 1.4	23 \pm 2.5	57 \pm 1.7	81 \pm 3.8

Table 4. Prolonged fibrinogen clotting time results in presence of bovine thrombin. Each value was calculated subtracting the clotting time produced by thrombin alone from that measured in the presence of one of the aptamers. The measurements were performed preincubating for 1 min 200 μ L of purified fibrinogen solution (2.0 mg/mL in PBS) with various concentrations of aptamers. The reaction started by the addition of 100 μ L of bovine thrombin solution (10 NIH *per* mL) and clotting time was recorded. In absence of any inhibitor, the measured clotting time for bovine thrombin was 22.5 \pm 1.2 s.

3.1.5 Conformational Search on TBA and Modified Analogues TBA-T₇b and TBA-T₁₂b

An extensive molecular modelling study, including molecular mechanics (MM) and dynamics (MD) calculations, was undertaken to analyse the SARs of new modified TBAs. In particular, to investigate the conformational properties of TBA and the new modified aptamers, a simulated annealing (SA) procedure followed by MM energy minimization was applied. Following the criteria described under Experimental Section, resulting conformers were analysed and each nucleotide of loops was classified on the basis of (i) χ torsional angle values, to identify the conformation of the glycosidic bond (i.e., *syn*, *anti*, or *s/a*); and (ii) the interatomic distance between its own centroid and that of the two G-tetrads, to investigate the position of loop nucleotides with respect to the guanine planes (i.e., “stacked” or “not-stacked”). Occurrence rates were calculated and compared to those obtained for TBA. Because the replacement of T₇ and T₁₂ residues with nucleoside **b** produced a mixture of diastereoisomers, characterized by *S* or *R* configuration at the acyclic linker, the occurrence rates reported in Tables 5 and 7 are the mean of the values obtained for the two diastereoisomers; single diastereoisomer values are reported in Tables 6 and 8.

3.1.6 Molecular Modelling Studies on Aptamer Interaction with Human and Bovine Thrombin

With the aim of identifying the effects of the new structural modifications on aptamer–thrombin interactions, first, all possible aptamer–thrombin binding modes and the corresponding binding interactions were analysed (PDB IDs 1HAO, 1HAP, 1HUT, 4DIH, and 4DII; see Experimental Section).

	TBA			TBA-T ₇ b			TBA-T ₁₂ b		
	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>
T ₃	28.5	69.5	2.00	37.0	61.5	1.50	47.0	50.5	2.50
T ₄	38.5	56.0	5.50	29.0	65.5	5.50	43.0	52.0	5.00
T ₇ ^a	39.0	59.5	1.50				36.0	62.5	1.50
G ₈	39.5	38.0	22.5	35.0	46.5	18.5	34.5	44.0	21.5
T ₉	41.5	50.0	8.50	34.0	57.0	9.00	42.0	51.0	7.00
T ₁₂ ^a	32.0	66.0	2.00	42.0	56.0	2.00			
T ₁₃	37.0	59.0	4.00	36.0	60.5	3.50	42.0	54.5	3.50

^aT₇ residue in TBA-T₇b, and T₁₂ residue in TBA-T₁₂b are replaced by the acyclic nucleoside **b**.

Table 5. Calculated occurrence rates (percent) of *syn*, *anti*, and *s/a* conformers of TGT and TT glycosidic bonds.

	TBA-T ₇ b (S)			TBA-T ₇ b (R)			TBA-T ₁₂ b (S)			TBA-T ₁₂ b (R)		
	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>
T ₃	40.5	57.5	2.00	33.5	65.5	1.00	46.5	51.0	2.50	47.0	50.0	3.00
T ₄	28.0	66.0	6.00	30.0	65.0	5.00	37.0	57.0	6.00	49.5	47.0	3.50
T ₇ ^a	-	-	-	-	-	-	40.0	57.5	2.50	31.5	67.5	1.00
G ₈	33.5	46.0	20.5	37.0	46.5	16.5	36.5	39.5	24.0	32.5	48.5	19.0
T ₉	36.5	54.5	9.00	31.5	60.0	8.50	39.0	53.0	8.00	44.5	48.5	7.00
T ₁₂ ^a	42.0	55.5	2.50	42.0	56.0	2.00	-	-	-	-	-	-
T ₁₃	40.0	56.0	4.00	32.5	65.0	2.50	43.0	53.0	4.00	41.0	56.0	3.00

^aT₇ residue in TBA-T₇b (S) and TBA-T₇b (R), and T₁₂ residue in TBA-T₁₂b (S) and TBA-T₁₂b (R), are replaced by the acyclic nucleoside **b**.

Table 6. Calculated occurrence rates (%) of *syn*, *anti*, and *s/a* conformers of TGT and TT glycosidic bonds.

	TBA	TBA-T ₇ b	TBA-T ₁₂ b
T ₃	62.5	33.0	39.0
T ₄	46.0	60.0	65.0
T ₇ ^a	14.0	12.0	10.0
G ₈	78.5	50.0	49.0
T ₉	26.0	39.5	42.5
T ₁₂ ^a	33.5	37.0	29.0
T ₁₃	38.0	40.5	29.0

^aT₇ residue in TBA-T₇b, and T₁₂ residue in TBA-T₁₂b are replaced by the acyclic nucleoside **b**.

Table 7. Calculated occurrence rates (percent) of conformers presenting TGT and TT nucleobases “stacked” on the guanine planes.

	TBA-T ₇ b (S)	TBA-T ₇ b (R)	TBA-T ₁₂ b (S)	TBA-T ₁₂ b (R)
T ₃	25.0	40.5	38.5	40.0
T ₄	61.5	58.5	59.5	70.5
T ₇ ^a	6.5	18.0	14.0	6.50
G ₈	25.5	74.5	49.5	49.0
T ₉	52.0	27.0	47.5	37.5
T ₁₂ ^a	36.0	37.5	37.5	20.0
T ₁₃	25.0	56.0	20.5	38.0

^aT₇ residue in TBA-T₇b (S) and TBA-T₇b (R), and T₁₂ residue in TBA-T₁₂b (S) and TBA-T₁₂b (R), are replaced by the acyclic nucleoside **b**.

Table 8. Calculated occurrence rates (%) of conformers presenting TGT and TT nucleobases “stacked” on the guanine planes.

This study revealed that the various binding orientations of TBA with respect to human thrombin (Figures 1 and 2) produced different binding modes in the crystal complexes sharing similar interactions with ABE I but involving different aptamer residues (Figure 9 and Table 9). Indeed, a thymine residue of TBA always interacts with ABE I hydrophobic cleft lined by Ile24, His71, Ile79 and Tyr117, but, depending on the binding mode, the interactions involve T₃ (Figure 9 C, D), or T₇ (Figure 9 A, B), or T₁₂ (Figure 9 E, F). T₇ interacts penetrating into the hydrophobic cleft more than T₁₂ and T₃ (Figure 9 F vs Figure 9 B, D); while the positioning of these latter resulted identical due to the symmetry of the two TT loops (Figure 9 B vs Figure 9 D). By consequence, a second ABE I subsite, including Gln38, Tyr76 and Ile82, interacts with T₁₂ or T₃, depending on the binding orientation (Figure 9 A–D). When the TGT loop binds to ABE I, the interaction with this second subsite is absent (Figure 9 E, F). Second, to identify the structural differences in TBA-binding site, experimentally determined structures of human and bovine thrombin were compared (PDB IDs are listed in the Experimental Section). Only three TBA-binding site residues, Ile24, Asn78, and Ile79, were mutated in bovine thrombin (Table 10), with no significant variation of the backbone structure (Figure 9). Human thrombin Asn78 residue (replaced by a lysine in bovine thrombin) interacts only with the TGT loop, establishing an Hbond interaction with the phosphate group of T₉ (PDB IDs 1HUT and 1HAP; Figure 9 E, F and Table 9). On the contrary, the interactions with human ABE I hydrophobic cleft containing Ile24 and Ile79 are conserved in all crystal complexes and involve T₃ or T₇ or T₁₂ depending on TBA-binding orientation (Figure 9 and Table 9).

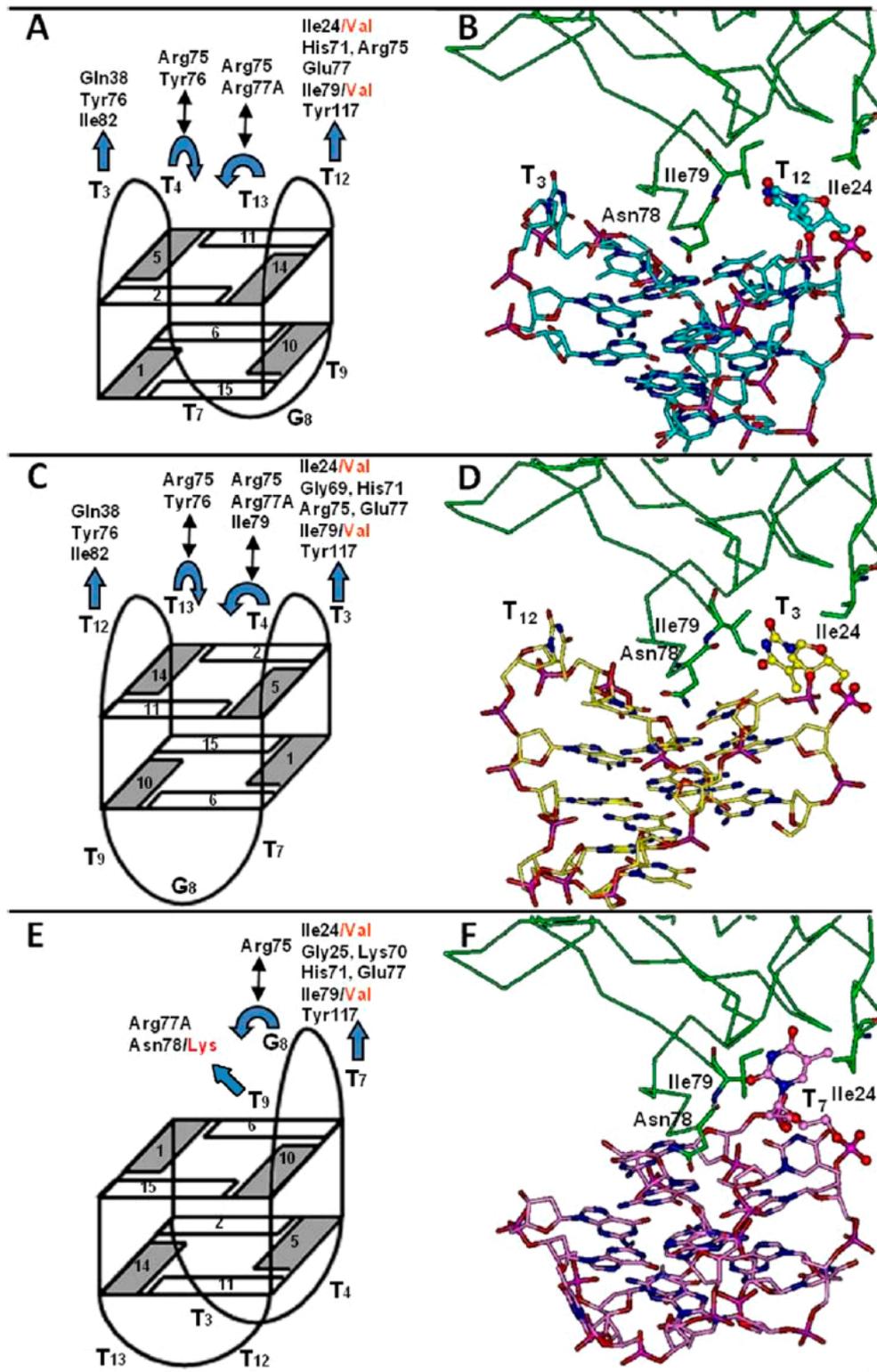


Figure 9. Cartoon (left) and 3D (right) structure of the different binding orientations of TBA with respect to human thrombin ABE I (green). (A, C, and E) Amino acid residues involved in TBA–thrombin interactions are evidenced in black; where mutated, the corresponding bovine residues are reported in red. The blue curve and straight arrows indicate loop nucleotides “stacked” or “not-stacked” on G-tetrad, respectively. (B) PDB ID 1HAO (TBA carbons = cyan); (D) PDB ID 4DIH (TBA carbons = yellow); (F) PDB ID 1HAP (TBA carbons = magenta). Protein carbons are colored in green, residues within TBA binding site that are mutated in bovine thrombin are shown and labeled. The nucleotides T3, T7, and T12 are evidenced as ball and stick. Heteroatoms are colored as follows: O = red; N = blue; P = magenta. Hydrogens are omitted for the sake of clarity.

1HAO TBA-thrombin X-ray complex			4DIH TBA-thrombin X-ray complex		
Aptamer residue	Thrombin residue	Interaction	Aptamer residue	Thrombin residue	Interaction
T3 (O2)	Gln 38	Hydrogen bond ^a	T3 (N3)	Glu 77	Hydrogen bond
T3	Tyr 76	π - π stacking	T3 (O2)	Gly 69	Hydrogen bond ^a
T3	Ile 82	Hydrophobic	T3	Ile 24 Ile 79	Hydrophobic
T4	Tyr 76	T-shape	T3	Hys 71	T-shape
T4 (O2)	Arg 75	Hydrogen bond	T3 (O3')	Tyr 117	Hydrogen bond
T12	Ile 24 Ile 79	Hydrophobic	T3(O4)	Arg 75	Hydrogen bond
T12 (O3/O4)	Tyr 117	Hydrogen bond	T4(O4)	Arg 75	Hydrogen bond
T12 (N3)	Glu 77	Hydrogen bond	T4(O2)	Arg 77A	Hydrogen bond
T12	His 71	T-shape	T4	Ile 79	Hydrophobic
T12 (O4)	Arg 75	Hydrogen bond ^a	G5 (O4')	Arg 77A	Hydrogen bond
T13 (O4)	Arg 75	Hydrogen bond	T12 (O2)	Gln 38	Hydrogen bond ^a
T13 (O2)	Arg 77A	Hydrogen bond	T12	Tyr 76	π - π stacking
			T12	Ile 82	Hydrophobic
			T13(O2)	Arg 75	Hydrogen bond
			T13(O4)	Tyr 76	Hydrogen bond
			T13	Tyr 76	T-shape

4DII TBA-thrombin X-ray complex			1HAP TBA-thrombin X-ray complex		
Aptamer residue	Thrombin residue	Interaction	Aptamer residue	Thrombin residue	Interaction
T3 (O4)	His 71	Hydrogen bond	T7 (N3)	Ile 24 Ile 79	Hydrophobic
T3	Ile 79	Hydrophobic	T7 (O4)	Gly 25 Lys 70	Hydrogen bond ^a
T3	Tyr 117	T-shape	T7	Tyr 117 His 71	π - π stacking
T4 (O4)	Arg 75	Hydrogen bond	T7 (N3)	Glu 77	Hydrogen bond
T4 (O3')	Asn 78	Hydrogen bond	G8 (O1P)	Arg 75	Ion pair
T4	Ile 79	Hydrophobic	T3(O4)	Arg 75	Hydrogen bond
G5 (O4')	Arg 77A	Hydrogen bond	T9 (O3')	Arg 77A	Hydrogen bond
T12	Tyr 76	π - π stacking	T9 (O1P)	Asn 78	Hydrogen bond
T13 (O2)	Arg 75	Hydrogen bond			
T13 (O4')	Tyr 76	Hydrogen bond			
T13	Tyr 76	T-shape			

^aHydrogen bond through a water molecule

Table 9. Binding interactions between TBA and human thrombin ABE I in X-ray complexes.

Human Thrombin	Bovine Thrombin
Ile 24^a	Val 24
Met 26 ^a	Leu 26
Asn 78^a	Lys 78
Ile 79^a	Val 79

^a Amino acids within Å radius from any given TBA atom (PDB IDs: 1HAO, 1HAP, 1HUT, 4DIH and 4DII)

Table 10. Mutated amino acid residues in human and bovine thrombin ABE I. The amino acid residues within TBA binding site are highlighted in bold.

The replacement of these two isoleucine residues in human thrombin by two valine residues in the bovine homologue determines an enlargement of the employable volume within the ABE I cleft. We then calculated the occurrence rates of TBA, TBA-T₇**b**, and TBA-T₁₂**b** SA/MM conformers presenting solvent accessible surface areas of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA binding the ABE I hydrophobic cleft in experimentally determined complexes (Table 11; single diastereoisomer values are reported in Table 12). Interestingly, the replacement of T₇ and T₁₂ with the acyclic nucleoside **b** determined in both cases a greater rate of conformations presenting a solvent-exposed conformation. Moreover, in agreement with the results reported in Tables 5 and 7, T₃ is more exposed to the solvent in the new modified aptamers than in TBA. Finally, we calculated the occurrence rates of bioactive conformations of residues at positions 3, 7, and 12 resulting from SA/MM calculations on TBA, TBA-T₇**b**, and TBA-T₁₂**b** (Figure 10 and Table 13; single diastereoisomer values are reported in Table 14).

ON	T3	T ₇ ^a	T ₁₂ ^b
TBA	16.5	85.0	36.5
TBA-T₇b	21.0	88.0	34.5
TBA-T₁₂b	22.0	76.0	50.0

^aT₇ residue in TBA-T₇**b** is replaced by the acyclic nucleoside **b**.
^bT₁₂ residue in TBA-T₁₂**b** is replaced by the acyclic nucleoside **b**.

Table 11. Occurrence rates (%) of TBA, TBA-T₇**b**, and TBA-T₁₂**b** SA/MM conformers presenting a solvent accessible surface area (Å²) of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA assuming the bioactive conformation at the ABE I hydrophobic cleft.

ON	T3	T ₇ ^a	T ₁₂ ^b
TBA-T₇b (S)	23.5	85.0	33.5
TBA-T₇b (R)	19.0	91.0	35.5
TBA-T₁₂b (S)	22.0	72.0	47.5
TBA-T₁₂b (R)	22.0	80.5	52.0

^aT₇ residue in TBA-T₇**b** (S) and TBA-T₇**b** (R) is replaced by the acyclic nucleoside **b**.
^bT₁₂ residue in TBA-T₁₂**b** (S) and TBA-T₁₂**b** (R) is replaced by the acyclic nucleoside **b**.

Table 12. Occurrence rates (%) of TBA-T₇**b** (S), TBA-T₇**b** (R), TBA-T₁₂**b** (S) and TBA-T₁₂**b** (R) SA/MM conformers presenting a solvent accessible surface area (Å²) of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA assuming the bioactive conformation at the ABE I hydrophobic cleft.

With this aim, all obtained conformers were superimposed on the experimentally determined TBA–thrombin complexes by fitting the guanine tetrads to evaluate the overlap with the residue interacting with the ABE I hydrophobic cleft (see Experimental Section for details). Obtained results show that TBA-T₇**b** presents an increased rate of bioactive conformations if compared to TBA; on the contrary, the substitution of T₁₂ with **b** decreases the rate of bioactive conformations of T₃ and T₇. Interestingly, the replacement of T₇ and T₁₂ with the acyclic nucleoside **b** determined in both cases a greater rate of conformations presenting a solvent-exposed position able to drive the interaction with the ABE I hydrophobic cleft. Nevertheless, likely due to the higher flexibility of the TGT loop with respect to the TT loops, there is a higher occurrence rate of T₇**b** assuming the ABE I bioactive conformation compared to T₃ and T₁₂**b** (Table 13 and Figure 10).

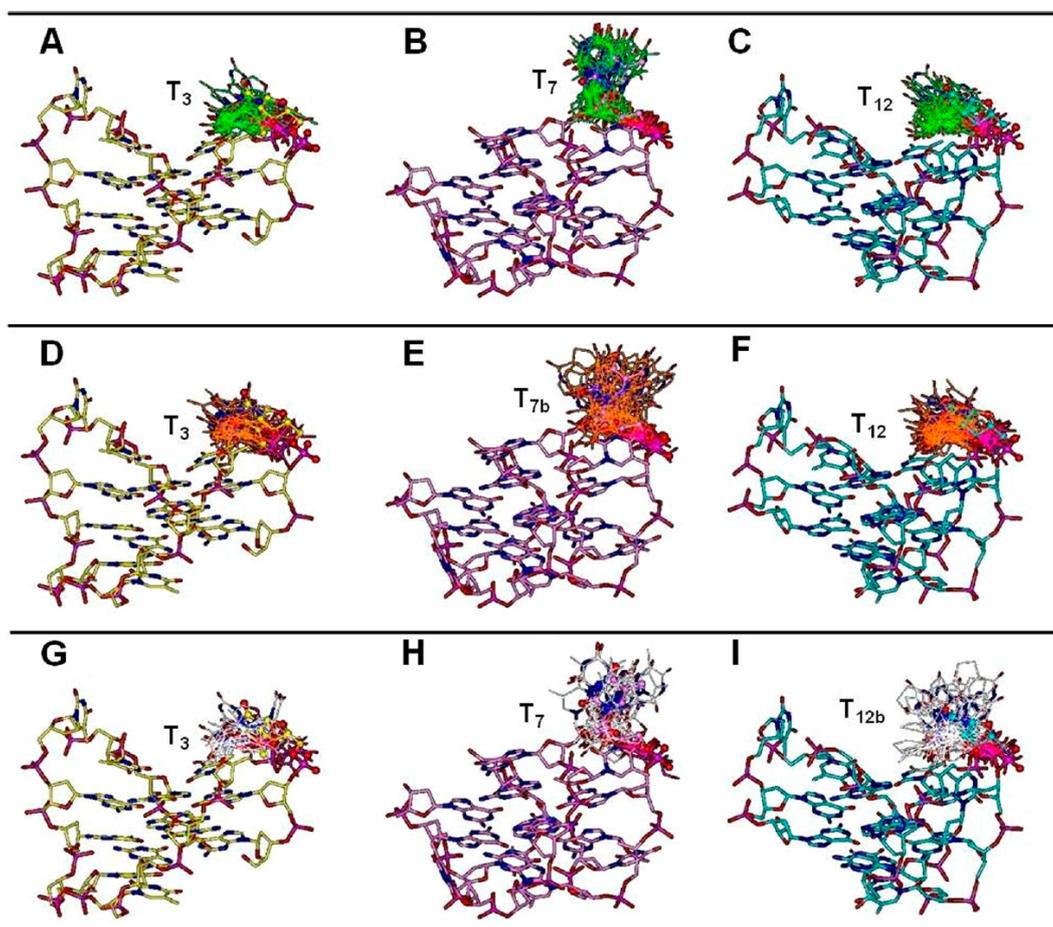


Figure 10. TBA (green; A-C), TBA-T₇**b** (orange; D-F), and TBA-T₁₂**b** (white; G-I) SA/MM conformers presenting residue 3, 7, or 12 assuming ABE I hydrophobic cleft binding conformation. Molecular models are superimposed on 4DIH (A, D, G; yellow), 1HAP (B, E, H; pink), or 1HAO (C, F, I; cyan) by fitting the guanine tetrads. All conformers of TBA and the mean of those obtained for the two diastereoisomers of TBA-T₇**b** and TBA-T₁₂**b** are shown. For the sake of clarity only the residue interacting with the ABE I hydrophobic cleft is displayed and hydrogens are omitted. Heteroatoms are colored as follows: O = red; N = blue; P = magenta.

ON	T3	T ₇ ^a	T ₁₂ ^b
TBA	9.50	12.0	10.0
TBA-T ₇ b	8.00	14.0	13.5
TBA-T ₁₂ b	6.50	6.50	10.0

^aT₇ residue in TBA-T₇**b** is replaced by the acyclic nucleoside **b**.
^bT₁₂ residue in TBA-T₁₂**b** is replaced by the acyclic nucleoside **b**.

Table 13. Occurrence rates (%) of bioactive conformations at the ABE I hydrophobic cleft of residues at positions 3, 7, and 12 resulting from SA/MM calculations.

ON	T3	T ₇ ^a	T ₁₂ ^b
TBA-T ₇ b (S)	9.00	12.0	14.0
TBA-T ₇ b (R)	6.50	16.0	13.0
TBA-T ₁₂ b (S)	5.50	7.00	16.5
TBA-T ₁₂ b (R)	7.50	6.00	4.00

^aT₇ residue in TBA-T₇**b** (S) and TBA-T₇**b** (R) is replaced by the acyclic nucleoside **b**.
^bT₁₂ residue in TBA-T₁₂**b** (S) and TBA-T₁₂**b** (R) is replaced by the acyclic nucleoside **b**.

Table 14. Occurrence rates (%) of bioactive conformations at the ABE I hydrophobic cleft of residues at positions 3, 7, and 12 resulting from SA/MM calculations on of TBA-T₇**b** (S), TBA-T₇**b** (R), TBA-T₁₂**b** (S) and TBA-T₁₂**b** (R).

3.2 Discussion

The introduction of a bulky five member cycle on the pyrimidine ring of the modified thymine residues of TBA-T₇**a** and TBA-T₁₂**a** led to compounds (TBA-T₇**b** and TBA-T₁₂**b**) able to fold into G-quadruplexes characterized by unaffected structural stability but showing different anticoagulant activity with respect to their parent compounds. (38) Because TBA-T₇**b** and TBA-T₇**a** show the same melting profile, with a T_m higher than that of TBA (Table 1), it can be concluded that the sole presence of the acyclic linker unit at position 7 is responsible for the structural stabilizing effect. Conformational analysis results indicate that the introduction of the acyclic linker at position 7 increased the solvent-exposed surface area (Table 13) of the modified residue and the rate of T₄ and T₁₃ conformations stacked on the guanine planes (Table 7). These data account for the nucleobase 7 always positioned outside the G-quadruplex core and not involved in intramolecular interactions, as well as for the reported stabilizing role played by T₄ and T₁₃ on TBA structural stability. (13-15) In line with our results, a similar increase in T_m value has also been observed by replacing the sugar moiety of T₇ with a different acyclic linker. (39) Interestingly, the fact that TBA-T₇**a** and TBA-T₇**b** showed the same enhancement of the T_m value with respect to TBA is not correlated with their diverse anticoagulant activities. Indeed, although the anticoagulant activities of TBA-T₇**a** and TBA-T₇**b** are higher than that of TBA, TBA-T₇**b** is more active than TBA-T₇**a** (Figure 7). Furthermore, PT assay data show that the incorporation of nucleoside **b** was more effective than that of nucleoside **a** also at position 12 of TBA (Table 2). It was somehow surprising that the activity of TBA with respect to the new modified analogues is reversed in fibrinogen assay (Figure 8 A). It could be supposed that the introduction of an acyclic linker in the TBA phosphate backbone enhances the aptamers' stability to nucleases, thus producing an increment of inhibitory effect in plasma tests. However, PT results were almost unaffected upon increasing of the incubation time of all sequences with

plasma (data not shown), thus suggesting that, in the explored conditions, nuclease stabilities and inhibitory activities are uncorrelated and that modifications at T₇ directly affect the interaction of aptamers with thrombin. Supporting this view, the use of bovine thrombin in fibrinogen assay changed again the activity trend. In particular, when tested toward bovine thrombin, the inhibitory activity of TBA-T₇**b** and TBA-T₁₂**b** significantly increases, becoming much higher than that of TBA, which, on the contrary, is almost unchanged in both fibrinogen tests (Figure 8, panel A vs panel B). On this ground, subtle structural differences between human and bovine thrombin that specifically affect the binding with TBA-T₇**b** and TBA-T₁₂**b** must exist. The results from our molecular modelling study using both human and bovine thrombin revealed that the obtained SARs fit with the binding of modified nucleobases, T₇**b** and T₁₂**b** to thrombin ABE I. Indeed, within the TBA binding site, human thrombin differs from the bovine counterpart by only three amino acids located at ABE I (Table 10). Specifically, in the bovine enzyme two isoleucine residues are replaced by valine residues and an asparagine is mutated into a lysine. In the reported human thrombin-TBA complexes, the isoleucine residues are involved in hydrophobic contacts with T₃ or T₇ or T₁₂, depending on the binding orientation of the aptamer (Figure 9). On the other hand, the asparagine residue is involved in H-bond interactions with the phosphate group linking G₈ to T₉ only when the TGT loop binds to ABE I (Tables 9 and 10). In the new TBA derivatives TBA-T₇**b** and TBA-T₁₂**b**, the introduction of the acyclic linker facilitates the extension of the phosphate backbone toward the solvent and the penetration of the modified nucleobase within the ABE I hydrophobic cleft, thus amplifying the steric hindrance caused by the presence of the five member cycle on modified nucleobase **b** (Table 11 and Figure 10). Accordingly, the presence of the smaller Val24 and Val79 residues in bovine thrombin could explain the increased activity of the new modified TBA analogues for this enzyme. The observed higher activity of TBA-T₇**b** is indeed correlated to the calculated higher rate of bioactive conformations at the ABE I hydrophobic cleft with respect to TBA and TBA-T₁₂**b** (Table 13 and Figure 10). This hypothesis is supported by the previously reported activity of TBA-T₇**a** and TBA-T₁₂**a**, (38) which, lacking the steric bulk at the modified nucleoside, was more active in the fibrinogen assay using human thrombin. On the other hand, the charged group at the lysine side chain allows the establishment of an ionic interaction with the modified TGT phosphate backbone of TBA-T₇**b**, likely contributing to the enhancement of inhibitory ability. On these bases, the apparent incongruence of PT and fibrinogen assay results, using human thrombin, can be interpreted by taking into account the effects of different aptamer binding modes on thrombin allostery. (3, 5-7) In the PT assay a number of thrombin ligands/effectors other than fibrinogen that are able to modulate thrombin activity are present, (1-5) (27, 43-45) thus driving different binding modes; the new modified aptamers could differently affect this ligand/effector binding network. In this scenario, it is noteworthy that the 65-84 thrombin loop, involved in specific interactions with the modified nucleobase of the new TBA derivatives according to our binding mode hypothesis, has been proved to be responsible for the allosteric long-range communication among ABE I, the catalytic site, and ABE II. (4, 28) Altogether, our results reveal the absence of a direct correlation between the structural stability and the anticoagulant property in our series of modified TBAs, suggesting that the role played by T₇ (placed in the TGT loop) and T₁₂ (placed in the TT loop close to the 3' terminus) in the biological activity of TBA and its analogues could involve alternative aptamer binding modes and the complex and not yet fully understood allosteric mechanism of action of thrombin.

3.3 Conclusions

Bringing together the so far acquired knowledge of the thrombin-TBA interaction and thrombin allosteric regulatory mechanism with our SAR data, we determined that the G-quadruplexes formed by the modified TBAs can bind the thrombin ABE I hydrophobic cleft by using either TGT or one of the two TT loops. Our results indicate that the incoherency of their behaviours in the two explored biological tests is due to the molecular process through which they exert the anticoagulant activity, which is surely more intricate than a competitive inhibition with the fibrinogen-binding site. Besides their high anticoagulant activity, these molecules represent valuable tools to further explore the complex regulatory mechanism of thrombin during plasma coagulation that is, as yet, not completely clarified.

3.4 Experimental Session

3.4.1 General Procedure

Chemicals and anhydrous solvents were purchased from Fluka-Sigma-Aldrich. TLCs were run on Merck silica gel 60 F254 plates. Silica gel chromatography was performed by using Merck silica gel 60 (0.063-0.200 mm). An API 2000 (Applied Biosystem) mass spectrometer was used to perform the analyses of the intermediates and monomer **6**. NMR experiments were recorded using Varian Mercury Plus 400 MHz and Varian ^{UNITY} INOVA 500 and 700 MHz spectrometers and processed using the Varian VNMR software package. Reagents and phosphoramidites for DNA synthesis were purchased from Glenn Research. ON syntheses were performed on a PerSeptive Biosystem Expedite DNA synthesizer. HPLC purifications and analyses were carried out using a JASCO PU-2089 Plus HPLC pump equipped with a JASCO BS-997-01 UV detector. The purity of the final products was determined as >95% by using a C-18 RP analytical column (C-18 Purospher STAR, Merck) eluted by a gradient of CH₃OH in H₂O (from 0 to 100% in 30 min). CD experiments were performed on a JASCO 715 spectropolarimeter equipped with a PTC-348 temperature controller.

3.4.2 Synthesis of Monomer **6**

3.4.2.1 3-Benzoyl-6,7-dihydro-1H-cyclopenta-[d]-pyrimidine-2,4(3H,5H)-dione (**2**)

Nucleobase 6,7-dihydro-1H-cyclopenta[d]pyrimidine-2,4(3H,5H)-dione was obtained as previously reported by Renault et al. (46) and converted in its N3 benzoyl derivative **2**. (38) ¹H NMR (400 MHz; mixture of CD₃OD/CDCl₃, 3:1 v/v): δ 7.75 (d, 2H, *J* = 8.0 Hz), 7.60 (t, 1H, *J* = 8.0 Hz), 7.43 (t, 2H, *J* = 8.0 Hz), 3.03 (t, 2H, *J* = 7.3 Hz), 2.67 (t, 2H, *J* = 7.3 Hz), 2.10 (q, 2H, *J* = 7.3 Hz) ppm. ¹³C NMR (100 MHz; mixture of CD₃OD/CDCl₃, 3:1 v/v): δ 169.3, 166.3, 155.2, 147.3, 134.2, 132.2, 128.9, 128.3, 127.5, 127.0, 113.4, 34.8, 30.5, 18.0 ppm. ESI Mass (positive mode): 257.1 [M+H]⁺, 279.1 [M+Na]⁺ (calculated 256.1).

3.4.2.2 3-Benzoyl-1-[(2,2,5-trimethyl-1,3-dioxan-5-yl)methyl]-6,7-dihydro-1H-cyclopenta-[d]-pyrimidine-2,4(3H,5H)-dione (**3**)

Compound **2** (1.5 g, 5.9 mmol) was suspended in 70 mL of dry dioxane at 30 °C in the presence of triphenylphosphine (2.28 g, 8.7 mmol) before the addition of di-*tert*-butyl azodicarboxylate (2.1 g, 9.1 mmol). To the resultant mixture, after 10 min of stirring at room temperature, was added 5.9 mmol of **1** dissolved in dry dioxane (300 μL). The reaction mixture was stirred at room temperature for 18 h under argon. The solution was concentrated under reduced pressure and the residue purified by

column chromatography on silica gel eluted with 95:5 Et₂O/CH₂Cl₂ to give **3** as a white solid (yield 43%, R_f 0.70). ¹H NMR (400 MHz; CDCl₃): δ 7.98 (d, 2H, J = 8.0 Hz), 7.60 (t, 1H, J = 8.0 Hz), 7.43 (t, 2H, J = 8.0 Hz), 4.0 (bs, 2H), 3.62 (m, 4H), 3.03 (t, 2H, J = 7.3 Hz), 2.66 (t, 2H, J = 7.3 Hz), 2.09 (q, 2H, J = 7.3 Hz), 1.38 (s, 3H), 1.42 (s, 3H), 0.87 (s, 3H) ppm. ESI Mass (positive mode): 399.2 [M+H]⁺, 421.2 [M+Na]⁺ (calculated 398.2).

3.4.2.3 1-(3-hydroxy-2-(hydroxymethyl)-2-methylpropyl)-6,7-dihydro-1H-cyclopenta-[d]-pyrimidine-2,4(3H,5H)-dione (**4**)

Compound **3** (1.0 g, 2.5 mmol) was suspended in 9:1 MeOH/H₂O (100 mL) containing 450 mg of Dowex 50WX8 (H⁺). After 8 h at room temperature, a NaOH 0.5 M aqueous solution was slowly added to neutralization. The solution was filtered and, in turn, basified with NaOH 1.0 M (5 mL) to remove the N-3 benzoyl group from the nucleobase. After 12 h at room temperature, the pH of the reaction was neutralized, the solvent was evaporated under vacuum, and the residue was dissolved in MeOH and purified by HPLC (C-18 reversephase column, Grace Davison Discovery Sciences, eluted with ACN in H₂O from 0 to 50% in 30 min) to give **4** as a white solid (yield 98%, elution time 15 min). ¹H NMR (700 MHz; CD₃OD): δ 3.80 (bs, 2H), 3.41 (d, 2H, J = 11.4 Hz), 3.38 (d, 2H, J = 11.4 Hz), 3.03 (t, 2H, J = 7.3 Hz), 2.66 (t, 2H, J = 7.3 Hz), 2.09 (q, 2H, J = 7.3 Hz), 0.87 (s, 3H) ppm. ¹³C-NMR (175 MHz; CD₃OD): δ 163.8, 161.4, 155.7, 113.6, 66.6, 49.5, 44.0, 34.2, 28.2, 22.6, 17.8 ppm; NOESY (700 MHz; CD₃OD), fundamental NOE signal between 3.80 and 3.03 ppm that confirms the linker at the 1N position. ESI Mass (positive mode): 255.3 [M+H]⁺, 277.3 [M+Na]⁺, 293.2 [M+K]⁺ (calculated 254.3).

3.4.2.4 3-(4,4'-dimethoxy-trityl-oxy)-2-((2,4-dioxo-2,3,4,5,6,7-hexahydro-1H-cyclopenta-[d]-pyrimidin-1-yl)methyl)-2-methylpropyl-2-cyanoethyl Diisopropylphosphoramidite (**6**)

Compound **4** (640 mg, 2.5 mmol), 4,4'-dimethoxytrytyl chloride (541 mg, 1.6 mmol), and 4-dimethylaminopyridine (15.0 mg, 0.12 mmol) were dissolved in dry pyridine (20 mL) and dry ACN (10 mL). The resulting solution was stirred at room temperature under argon for 1.5 h. Dry methanol (200 μL) was then added to quench the reaction. After 30 min under stirring, the solution was concentrated under reduced pressure and the residue purified by column chromatography on silica gel (eluted with 50:50:1 EtOAc/hexane/Et₃N) to give monodimethoxytritylated **5** as a clear yellow solid (44% yield from **4**; R_f 0.51 in EtOAc/hexane 1:1 v/v). The solid (600 mg, 1.1 mmol) was dried in vacuo overnight before being dissolved in anhydrous DCM (8 mL) and diisopropylethylamine (600 μL, 3.6 mmol) under argon. Three hundred microliters of β-cyanoethyl diisopropylchlorophosphoramidite was then added (1.2 mmol). After 40 min, the reaction was quenched by the addition of dry methanol (100 μL), diluted with ethyl acetate (15 mL), and finally washed with 10% sodium carbonate solution (15 mL) and brine (15 mL). The organic layer was dried on magnesium sulfate and concentrated in vacuo. The residue was purified by silica gel chromatography eluted with DCM, ethyl acetate, and triethylamine (80:10:10). The fractions containing the product were collected and concentrated under vacuum, yielding **6** as a white foam (99% yield; R_f 0.65 in CHCl₃/MeOH/TEA 97:3:0.05 v/v/v). ¹H NMR (700 MHz; CDCl₃): δ 7.90 (2H), 7.38 (3H), 7.35 (4H), 7.12 (1H), 6.85 (4H), 3.90 (2H), 3.70 (6H), 3.65 (2H), 3.60 (2H), 3.45 (2H) 3.39-3.02 (2H), 2.60-2.45 (2H), 1.45(3H) 1.09 (6H), 1.04 (6H) ppm. ¹³C-NMR (175 MHz; CDCl₃): δ 158.6, 151.6, 147.3, 139.4, 135.1, 130.3, 129.1, 127.8, 127.7, 127.1, 113.1, 64.4, 60.4, 55.2, 47.3, 33.6, 29.7, 27.6, 20.9, 19.3,

17.3 ppm. ^{31}P NMR (200 MHz, CDCl_3) δ 146.1 and 145.9 ppm. ESI Mass (positive mode): 757.9 $[\text{M}+\text{H}]^+$, 779.9 $[\text{M}+\text{Na}]^+$ (calculated 756.9).

3.4.3 Synthesis of Oligomers

TBA and analogues were synthesized by using standard solid phase DNA chemistry on controlled pore glass (CPG) support following the β -cyanoethyl phosphoramidite method. The oligomers were detached from the support and deprotected by treatment with an aqueous ammonia solution (33%) at 55 °C overnight. The combined filtrates and washings were concentrated under reduced pressure, dissolved in H_2O , and purified by HPLC using an anionic exchange column eluted with a linear gradient (from 0 to 100% B in 30 min) of phosphate buffer at pH 7.0 (A, 20 mM NaH_2PO_4 aqueous solution containing 20% CH_3CN ; B, 1.0 M NaCl, 20 mM NaH_2PO_4 aqueous solution containing 20% CH_3CN). The oligomers were successively desalted by molecular exclusion chromatography on Biogel P-2 fine. The purity was checked on HPLC by using reverse phase column. The concentrations of the samples used in CD and UV experiments were determined by measuring the absorbance at 260 nm at 80 °C and using the open access program available on <http://basic.northwestern.edu/biotools/OligoCalc.html>. (47)

3.4.4 NMR Experiments

1D NMR spectra were acquired as 16384 data points with a recycle delay of 1.0 s at temperatures in the range of 2-50 °C. Data sets were zero filled to 32768 points prior to Fourier transformation and apodized with a shifted sine bell squared window function. The pulsed-field gradient DPGSE (48, 49) sequence was used for H_2O suppression. NMR samples of TBA-T₇**b** and TBA-T₁₂**b** (0.5 mM single-strand concentration) were prepared in 100 mM K^+ buffer ($\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1 v/v containing 90 mM KCl, 10 mM KH_2PO_4 , and 0.2 mM EDTA).

3.4.5 CD Experiments

To perform CD experiments, each ON was dissolved in the potassium (90 mM KCl, 10 mM KH_2PO_4 , pH 7.0) or PBS (Sigma-Aldrich; 10 mM phosphate buffer, 2.7 mM KCl, 137 mM NaCl, pH 7.4) phosphate buffer at the final ON concentration of 2.0×10^{-5} M and submitted to the annealing procedure (heating at 90 °C and slowly cooling at room temperature). Before each experiment, the samples were equilibrated at 10 °C for 30 min. CD spectra were recorded from 200 to 360 at 100 nm/min scanning rate, 16 s response, and 1.0 nm bandwidth. Each CD profile was obtained by taking the average of three scans from which the spectrum of background buffer was subtracted. CD melting curves were obtained by monitoring the variation of absorbance at 295 nm from 10 to 80 °C. Two melting experiments for each ON were recorded at 0.5 °C/min heating rate.

3.4.6 Prothrombin (PT) Time

PT time was measured by using a Koagulab MJ Coagulation System with a specific kit RecombiPlas Tin HemosIL (Instrumentation Laboratories, Lexington, MA, USA). The procedure was performed according to the manufacturer's instructions. In our experimental protocol a time course of each ON or vehicle incubated with 100 μL of plasma at 37 °C has been performed. For the evaluation of PT at 20.0 μM , in the apposite microtube, 2.0 μL of the corresponding ON solution (1.0×10^{-3} M in PBS) or vehicle was added. The PT at final ON concentration of 2.0 μM was determined by

using 2.0 μL of a diluted ON solution (the initial ON solution 1.0×10^{-3} M in PBS was diluted at a final concentration of 1.0×10^{-4} M). Using six different incubation times from 30 s to 15 min (i.e., 30 s and 1, 2, 5, 10, and 15 min) 200 μL of the kit solution containing Recombiplastin was added with consequent activation of extrinsic pathway. The PT measurement, for each incubation time, was produced in triplicate, and the average and its standard error values were calculated. The basal clotting time was determined by measuring the clotting time in the absence of any ON. The fold increase of basal PT time was calculated as the ratio between the measured PT time in the presence of each ON and the basal PT value (13.4 ± 0.2 s).

3.4.7 Purified Fibrinogen Clotting Assay

ONs were incubated for 1min at 37 °C in 200 μL of buffer (20 mM tris acetate, 140 mM NaCl, 2.7 mM KCl, 1.0 mM MgCl_2 , 1.0 mM CaCl_2 , pH 7.4) containing 2.0 mg/mL of fibrinogen (fibrinogen from human plasma, F 3879, Sigma-Aldrich). One hundred microliters of human (Sigma-Aldrich, T8885, human thrombin suitable for thrombin time test) or bovine (HemosIL, Thrombin Time Kit, Instrumentation Laboratories) thrombin (10 NIH per mL) was then added to the solution containing the fibrinogen and the ON. The time required to clot was measured using a Koagulab MJ Coagulation System. The clotting time of each ON was determined in triplicate at different concentrations. The basal clotting time was determined by measuring the clotting time in the absence of any ONs. Prolonged clotting time was obtained by subtracting the basal clotting value from each ON clotting time. The ratio of basal and prolonged clotting time versus log [ON] was reported.

3.4.8 Molecular Modeling

Molecular modeling calculations were performed on SGI Origin 200 8XR12000 and E4 Server Twin 2 x Dual Xeon 5520, equipped with two nodes. Each node was 2 x Intel Xeon QuadCore E5520, 2.26 GHz, 36 GB RAM. The molecular modeling graphics were carried out on SGI Octane 2 workstations.

3.4.8.1 Conformational Analysis

Experimentally determined structures of TBA alone (PDB ID 148D) and in complex with thrombin (PDB IDs 1HAO, 1HAP, and 1HUT) were downloaded from Protein Data Bank (PDB, <http://www.rcsb.org/pdb/>) and analyzed using the Homology module of Insight 2005 (Accelrys Software Inc., San Diego, CA, USA). Hydrogens were added to all of these structures considering a pH value of 7.4 (Biopolymer Module, Insight 2005). Because the replacement of T₇ and T₁₂ residues with nucleoside **b** produced a mixture of diastereoisomers characterized by *S* or *R* configuration at the acyclic linker, for each new TBA analogue TBA-T₇**b** and TBA-T₁₂**b**, the two diastereoisomers were built by modifying the experimentally determined structure of TBA in complex with thrombin (PDB ID 1HAO; Insight2005 Builder module). Atomic potentials and charges were assigned using the CVFF force field.³⁰ The conformational space of TBA (PDB ID 1HAO) and the new modified analogues was sampled through 200 cycles of simulated annealing (SA) followed by molecular mechanics (MM) energy minimization. During the SA procedure, the temperature is altered in time increments from an initial temperature to a final temperature by adjusting the kinetic energy of the structure (by rescaling the velocities of the atoms). The following protocol was applied: the system was heated to 1000 °K over 2000 fs (time step of 1.0 fs); a temperature of 1000 °K was applied to the system for 2000 fs (time step of 1.0 fs) to surmount torsional barriers; successively, the temperature was

linearly reduced to 300 °K in 1000 fs (time step of 1.0 fs). Resulting conformations were then subjected to MM energy minimization within Insight 2005 Discover_3 module (CVFF force field) until the maximum rms derivative was <0.001 kcal/Å, using a conjugate gradient (50) as the minimization algorithm.

To reproduce the physiological environment where these molecules act and, to evaluate the effects of the implicit solvent, we sampled the conformational space through the combined procedure of SA/MM calculations, using the dielectric constant of the water ($\epsilon = 80r$). Moreover, to allow a complete relaxation of the structures preserving the monomolecular chairlike G-quadruplex folding topology, during the entire course of SA/MM calculations, we applied a tether force of 100 kcal/Å² to the guanine bases of two quartets. All resulting conformers were subsequently analyzed and loop nucleotides were classified on the bases of (i) glycosidic bond χ values, that is, $0^\circ < \chi < 90^\circ = \textit{syn}$; $-60^\circ < \chi < -180^\circ = \textit{anti}$; (51, 52) $90^\circ < \chi < 180^\circ$ and $-60^\circ < \chi < 0^\circ = \textit{s/a}$); (ii) the interatomic distance between the centroid of the ring atoms of the nucleobase of each loop nucleotide and the centroid of the ring atoms of the nucleobases of the two G-tetrads (Pseudo_Atom Define command, Biopolymer Module, Insight 2005). According to the latter parameter, the loop nucleotide was classified as “stacked” when the distance was <8 Å or as “not-stacked” when the distance was >12 Å, whereas a 3D visual inspection was needed to classify the nucleotide as “stacked” or “not-stacked” when the distance was between 8 and 12 Å. A nucleotide termed “stacked” presented at least one nucleobase atom shielded by the G-tetrads; a nucleotide termed “not-stacked” presented no atoms shielded by the G-tetrads. Occurrence rates were calculated for TBA and for each diastereoisomer of the new analogues TBA-T₇**b** and TBA-T₁₂**b**. Because all experimental data refer to the mixture of the two diastereoisomers, the mean of the values obtained for the two diastereoisomers was also calculated.

3.4.8.2 Structural and Bioinformatics Analysis

To analyze the binding modes of TBAs and the corresponding aptamer-thrombin interactions, all of the experimentally determined structures of aptamers in complex with human thrombin were downloaded from Protein Data Bank: 1HAO, 1HAP, 1HUT, 3DD2, 3QLP, 4DIH, and 4DII. On the other hand, to identify the structural differences in TBA binding site between human and bovine thrombin, additional structures of human (PDB IDs 1HXF, 1TB6, 1TMT, 1TMU, 1XMN, 3HTC, and 4HTC) and bovine (PDB IDs 1HRT, 1VIT, and 3PMA) thrombin, sharing similar ABE I and ABE II ligands, were selected and analyzed. Hydrogens were added to all of the PDB structures considering a pH value of 7.4 (Biopolymer Module, Insight 2005). All structures were superimposed by C α atoms, and their sequences were extracted using the Homology module of Insight 2005 (Accelrys). On the other hand, the human (entry P00734) and bovine (entry P00735) prothrombin sequences were downloaded from the UniProt Knowledgebase (<http://www.uniprot.org>), and the sequence alignments were performed using Multiple_Sequence Alignment pulldown in the Insight 2005 Homology module. Moreover, for each ligand/enzyme complex, a subset around the ligands that consisted of all residues and water molecules having at least one atom within a 6 Å radius from any given ligand atom was defined. The created subsets were displayed and analyzed through a 3D visual inspection. The results of this analysis were compared with those obtained through the sequence alignments. Starting from the results obtained from this structural and bioinformatics analysis, to evaluate in detail the possibility of the residues at positions 3, 7, and 12 to interact with thrombin ABE I hydrophobic cleft, the solvent-accessible surface area

of these nucleotides was evaluated by calculating the Connolly surface with a probe radius of 1.4 Å, which approximates the radius of a water molecule (Viewer Module, Insight 2005, Accelrys Software Inc.). In particular, the Connolly surface of the considered nucleotides was calculated for the TBA experimentally determined structures (PDB IDs 1HAO, 1HAP, 4DIH, and 4DII) and for all conformers of TBA and new modified analogues, resulting from SA/MM calculations. The TBA, TBA-T₇**b**, and TBA-T₁₂**b** SA/MM conformers presenting solvent-accessible surface areas of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA when interacting with the ABE I hydrophobic cleft in experimentally determined complexes (i.e., T₃, 191.01 Å²; T₇, 159.35 Å²; and T₁₂, 180.59 Å²) were selected, and their occurrence rates were calculated. Because all experimental data refer to the mixture of the two diastereoisomers, the mean of the values obtained for the two diastereoisomers was also calculated. Finally, with the aim to calculate the occurrence rates of bioactive conformations of residues at position 3, 7, and 12 of TBA and new modified analogues TBA-T₇**b** and TBA-T₁₂**b**, all conformers, resulting from SA/MM calculations, were superimposed on the experimentally determined structures of TBA in complex with thrombin (PDB IDs 1HAO, 1HAP, 1HUT, 4DIH, and 4DII) by fitting heavy atoms of the guanine bases of two quartets, and the overlap with the residue interacting with the ABE I hydrophobic cleft (I1e24, His71, I1e79, and Tyr117) was evaluated. To assess the bioactive conformation of the residue at position 7, the orientation that allows residue 7 to be located in the same position of T₇ of 1HAP or 1HUT crystal structure was considered (orientation II in Figure 2). The conformation of residues at positions 3, 7, and 12 was considered bioactive when the correspondent nucleobase was positioned within the ABE I hydrophobic cleft and presented at least one atom superimposed on T₃ (PDB IDs 4DIH and 4DII) or on T₇ (PDB IDs 1HAP and 1HUT) or on T₁₂ (PDB ID 1HAO), respectively. Occurrence rates of bioactive conformations of residues at positions 3, 7, and 12 were then calculated. Because all experimental data refer to the mixture of the two diastereoisomers, the mean of the values obtained for the two diastereoisomers was also calculated.

3.5 References

- (1) Huntington, J. A. (2005) Molecular recognition mechanisms of thrombin. *Journal of Thrombosis and Haemostasis* 3, 1861–1872.
- (2) Cera, E. D. (2003) Thrombin interactions. *Chest* 124, 11S-17S.
- (3) W. Niu, Z. C., P. S. Gandhi, A. D. Vogt, N. Pozzi, L. Pelc, F. Zapata, E. Di Cera. (2011) Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry* 50, 6301–6307.
- (4) B. C. Lechtenberg, D. J. D. J., S. M. V. Freund, J. A. Huntington. (2010) NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proceedings of the National Academy of Sciences of the United States of America* 107, 14087-14092.
- (5) P. S. Gandhi, Z. C., F. S. Mathews, E. Di Cera. (2008) Structural identification of the pathway of long-range communication in an allosteric enzyme. *Proceedings of the National Academy of Sciences of the United States of America* 105, 1832-1837.
- (6) E. Di Cera, M. J. P., A. Bah, L. A. Bush-Pelc, L. C. Garvey. (2007) Thrombin allostery. *Physical Chemistry Chemical Physics* 9.

- (7) Cera, E. D. (2007) Thrombin as procoagulant and anticoagulant. *Journal of Thrombosis and Haemostasis* 5, 196-202
- (8) Dahlback, B. (2005) Blood coagulation and its regulation by anticoagulant pathways: genetic pathogenesis of bleeding and thrombotic diseases. *Journal of Internal Medicine* 257, 209-223
- (9) Desai, U. R. (2004) New antithrombin-based anticoagulants *Medicinal Research Reviews* 24, 151-181.
- (10) S. M. Nimjee, S. O., Z. Volovyk, K. M. Bompiani, S. B. Long, M. Hoffman, B. A. Sullenger. (2009) Synergistic effect of aptamers that inhibit exosites 1 and 2 on thrombin. *RNA* 15, 2105-2111.
- (11) W. X. Li, A. V. K., G. W. Grant, J. J. Toole, L. L. Leung. (1994) A novel nucleotide-based thrombin inhibitor inhibits clot-bound thrombin and reduces arterial platelet thrombus formation. *Blood* 83, 677-682.
- (12) L. C. Bock, L. C. G., J. A. Latham, E. H. Vermaas, J. J. Toole. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355, 564-566.
- (13) P. Schultze, R. F. M., J. Feigon. (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d-(GGTTGGTGTGGTTGG). *Journal of Molecular Biology* 235, 1532-1547.
- (14) R. Macaya, P. S., F. Smith, J. Roe, J. Feigon. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proceedings of the National Academy of Sciences of the United States of America* 90, 3745-3749.
- (15) K. Y. Wang, S. H. K., N. Bischofberger, S. Swaminathan, P. H. Bolton. (1993) The tertiary structure of a DNA aptamer which binds to and inhibits thrombin determines Activity. *Biochemistry* 32, 11285-11292.
- (16) I. Russo Krauss, A. M., A. Randazzo, E. Novellino, L. Mazzarella, F. Sica. (2012) High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Research* 40, 8119-28.
- (17) K. Padmanabhan, A. T. (1996) An ambiguous structure of a DNA 15-mer thrombin complex. *Acta Crystallographica Section D: Biological Crystallography* 52, 272-282.
- (18) K. Padmanabhan, K. P. P., J. D. Ferrara, J. E. Sadler, A. Tulinsky. (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *Journal of Biological Chemistry* 268, 17651-17654.
- (19) J. A. Kelly, J. F., T. O. Yeates. (1996) Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d-(GGTTGGTGTGGTTGG). *Journal of Molecular Biology* 256 417-422.
- (20) I. Russo Krauss, A. M., C. Giancola, A. Randazzo, L. Mazzarella, F. Sica. (2011) Thrombin-aptamer recognition: a revealed ambiguity. *Nucleic Acids Research* 39, 7858-7867.
- (21) B. Pagano, L. M., A. Randazzo, C. Giancola. (2008) Stability and binding properties of a modified thrombin binding aptamer *Biophysical Journal* 94, 562-569.
- (22) L. Martino, A. V., A. Randazzo, A. Virgilio, V. Esposito, C. Giancola, M. Bucci, G. Cirino, L. Mayol. (2006) A new modified thrombin binding aptamer containing a 5'-5' inversion of polarity site. *Nucleic Acids Research* 34, 6653-6662.

- (23) G. Zhou, X. H., Y. Qu. (2010) Thi binding effect of aptamers on thrombin. *Biochemical Engineering Journal* 52, 117-122.
- (24) D. M. Tasset, M. F. K., W. Steiner. (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *Journal of Molecular Biology* 272, 688-698
- (25) Y. Kim, Z. C., W. Tan. (2008) Molecular assembly for highperformance bivalent nucleic acid inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* 105, 5664-5669.
- (26) J. Müller, D. F., G. Mayer, B. Pöttsch. (2008) Anticoagulant characteristics of HD1-22, a bivalent aptamer that specifically inhibits thrombin and prothrombinase. *Journal of Thrombosis and Haemostasis* 6, 2105-2112.
- (27) N. S. Petrera, A. R. S., B. A. Leslie, C. A. Kretz, J. C. Fredenburgh, J. I. Weitz. (2009) Long range communication between exosites 1 and 2 modulates thrombin function. *Journal of Biological Chemistry* 284, 25620-25629.
- (28) T. M. Sabo, H. D. F., M. C. Maurer. (2006) Conformational analysis of γ peptide (410-427) interactions with thrombin anion binding exosite II. *Biochemistry* 45, 7434-7445.
- (29) M. Tsiang, A. K. J., K. E. Dunn, M. E. Rojas, L. L. Leung, C. S. Gibbs. (1995) Functional mapping of the surface residues of human thrombin *Journal of Biological Chemistry* 270, 16854-16863.
- (30) Schwienhorst, A. (2006) Direct thrombin inhibitors a survey of recent developments. *Cellular and Molecular Life Sciences* 63, 2773-2791.
- (31) M. C. R. Buff, F. S. f., B. Wulffen, J. Müller, B. Pöttsch, A. Heckel, G. Mayer. (2010) Dependence of aptamer activity on opposed terminal extensions: improvement of light-regulation efficiency. *Nucleic Acids Research* 38, 2111-2118.
- (32) G. X. He, S. H. K., S. Swaminathan, R. G. Shea, J. P. Dougherty, T. Terhorst. (1998) N2- and C8-substituted oligodeoxynucleotides with enhanced thrombin inhibitory activity in vitro and in vivo. *Journal of Medicinal Chemistry* 41, 2234-2242.
- (33) B. Saccá, L. L., J. L. Mergny. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Research* 33, 1182-1192.
- (34) G. X. He, J. P. W., M. J. Postich, S. Swaminathan, R. G. Shea, T. Terhorst, V. S. Law, C. T. Mao, C. Sueoka, S. Coutré, N. Bischofberger. (1998) In vitro and in vivo activities of oligodeoxynucleotide-based thrombin inhibitors containing neutral formacetal linkages. *Journal of Medicinal Chemistry* 41, 4224-4231.
- (35) G. Oliviero, J. A., N. Borbone, S. D'Errico, A. Galeone, L. Mayol, S. Haider, O. Olubiyi, B. Hoorelbeke, J. Balzarini, G. Piccialli (2010) Tetraend-linked oligonucleotides forming DNA G-quadruplexes: a new class of aptamers showing anti-HIV activity. *Chemical Communications* 46, 8971-8973.
- (36) G. Oliviero, J. A., N. Borbone, A. Galeone, L. Petraccone, M. Varra, G. Piccialli, L. Mayol. (2006) Synthesis and characterization of monomolecular DNA G-quadruplexes formed by tetra-end-linked oligonucleotides. *Bioconjugate Chemistry* 17, 889-898.
- (37) G. Oliviero, N. B., A. Galeone, M. Varra, G. Piccialli, L. Mayol. (2004) Synthesis and characterization of a bunched oligonucleotide forming a monomolecular parallel quadruplex structure in solution. *Tetrahedron Letters* 45, 4869-4872.
- (38) T. Coppola, M. V., G. Oliviero, A. Galeone, G. D'Isa, L. Mayol, E. Morelli, M. R. Bucci, V. Vellecco, G. Cirino, N. Borbone. (2008) Synthesis, structural studies

- and biological properties of new TBA analogues containing an acyclic nucleotide *Bioorganic & Medicinal Chemistry* 16, 8244–8253.
- (39) A. Pasternak, F. J. H., L. M. Rasmussen, B. Vester, J. Wengel. (2011) Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer. *Nucleic Acids Research* 39, 1155–1164.
- (40) T. Y. S. But, P. H. T. (2007) The Mitsunobu reaction: origin, mechanism, improvements, and applications. *Chemistry – An Asian Journal* 2, 1340–1355.
- (41) A. N. Lane, J. B. C., R. D. Gray, J. O. Trent. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research* 36, 5482–5515.
- (42) S. Neidle, S. B. (2006) (Publishing, R., Ed.) pp 100–130, London, UK
- (43) C. A. Kretz, A. R. S., J. C. Fredenburgh, J. I. Weitz. (2006) HD1, a thrombin-directed aptamer, binds exosite 1 on prothrombin with high affinity and inhibits its activation by prothrombinase. *Journal of Biological Chemistry* 281, 37477–37485.
- (44) P. C. Liaw, J. C. F., A. R. Stafford, A. Tulinsky, R. C. Austin, J. I. Weitz. (1998) Localization of the thrombin-binding domain on prothrombin fragment 2. *Journal of Biological Chemistry* 273, 8932–8939.
- (45) H. V. Jakubowski, M. D. K., W. G. Owen. (1986) The effect of bovine thrombomodulin on the specificity of bovine thrombin. *Journal of Biological Chemistry* 261, 3876–3882.
- (46) J. Renault, D. L., M. Robba. (1994) Synthesis and antiviral study of cyclopentano[d]pyrimidine-2,4-diones and octahydroquinazoline-2,4-diones acyclic nucleosides as potential anti-HIV agents. *Nucleosides, Nucleotides and Nucleic Acids* 13, 891–901.
- (47) Kibbe, W. A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* 35, W43–W46.
- (48) Dalvit, C. (1998) Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *Journal of Biomolecular NMR* 11, 437–444.
- (49) T. L. Hwang, A. J. S. (1995) Water suppression that works. Excitation sculpting using arbitrary wave forms and pulsed field gradients. *Journal of Magnetic Resonance, Series A* | 112, 275–279.
- (50) Fletcher, R. (1980) *Unconstrained optimization*. In *Practical Methods of Optimization*, Vol. 1, Wiley, New York.
- (51) Neidle, S. (2008) *Principles of Nucleic Acid Structures*, Academic Press, London, UK.
- (52) J. Reichert, J. S. (2002) The IMB Jena Image Library of Biological Macromolecules 2002 Update. *Nucleic Acids Research* 30, 253–254.

Chapter 4

DNA-based Nanostructures: dimers of G-quadruplexes via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads

Guanine-rich oligonucleotides can form a variety of G-quadruplex architectures involving stacked planar G-tetrads. The structural polymorphism of G-quadruplex topologies depends on the range of alternate strand orientations, loop connectivities, and *syn-anti* distribution of guanine bases around G-tetrads (1). Additional structural diversity arises from the accommodation of A-, T- or C-tetrads (2-4). Furthermore, several papers describing the higher order packing of intramolecular (5-7) or intermolecular (8, 9) DNA quadruplexes have also been published and higher order pairing alignments involving G-tetrads have been recently identified, such as pentads (10, 11), hexads (12), and heptads (8, 13) for sequences containing guanine stretches separated by adenine residues. One of the structural features required for quadruplex multimerization is the presence of a stretch of guanines at the 5'- or 3'-termini of parallel G-quadruplexes, and various stacking modes (5'-5', 3'-3' and 3'-5') have been proposed as a means for higher-order packaging of G-quadruplexes (14). These types of packaging have been reported to lead to higher ordered G-quadruplexes assemblies by either the association of interlocked slipped strands (15) or by end-to-end stacking (9, 15). Understanding and controlling higher order structures formation is of prime importance for in vitro studies of oligonucleotides derived from biologically relevant sequences, and for the development of DNA-based nanostructures. In this respect, biomolecular self-assembly of the strongly interacting DNA bases provides a promising avenue, and the construction of specific higher order G-quadruplexes as DNA-based nanostructures may represent a key goal in the field of nanotechnologies. The main fundamental step to drive this topic from a laboratory curiosity towards the technologically relevant engineering is the achieving of control of self-folding and self-stimulated formation of these higher order G-quadruplexes in solution as a function of their molecular composition (i.e., base sequence) and solution conditions (e.g., DNA and cations concentration)(16). Different experimental techniques may provide different views on the formation and structuring of higher order G-quadruplexes. For example, Polyacrylamide Gel Electrophoresis (PAGE) (17) and mass spectrometry (18-20) give indications on the molecular sizes of G-quadruplexes, while CD spectra are empirically used to speculate on the relative orientation of the strands constituting the G-tetrad core. So far, only NMR and X-ray crystallography can provide atomic-resolution structures of G-quadruplexes, and for NMR structural analysis of G-quadruplexes, unambiguous and model-independent spectral assignment approaches are required (21).

In this context, the aim of this research work was to achieve a small library of DNA-based nanostructures consisting of dimers of G-quadruplexes. Thanks to several analytical techniques, such as Nuclear Magnetic Resonance (NMR), mass spectrometry (MS), Circular Dichroism (CD), UV and PAGE, it was proved that the sequence CGGTGGT can form a nanostructure of about 4 nm in length, consisting of a 'dimeric G-quadruplex' (from now on **2Q**), which is in fact an octamer of the sequence, coexisting in solution with minor amounts of other DNA secondary structures and with the single-stranded molecule. The octamer formation takes place through a dimerization process, which is allowed by the 5'-5' stacking of two 'monomeric G-quadruplexes' (**1Q**) subunits, and assisted by the formation of an unusual G(:C):G(:C):G(:C):G(:C) octad arrangement, never reported before, involving

C1 and G2 bases of each CGGTGGT strand (22). Starting from these results, a speculation about the main features essentials to the dimerization process has been carried on, focusing the attention on ODN bases sequence and solution conditions.

4.1 Results

4.1.1 dCGGTGGT forms an octamer

The formation of a higher order assembly than a tetramer by the DNA sequence dCGGTGGT was first noticed in native PAGE experiments. Figure 1a shows PAGE experiments on the following DNA sequences: TGGGGT annealed in 0.1 M potassium buffer (lane 1) used as a tetramolecular quadruplex ($[dTGGGGT]_4$) size-marker, CTGTGTT (lane 2) used as a single-stranded 7-mer marker, TGGTGGC (lanes 3 and 4), CGGTGGT (lanes 5 and 6) and CGGGGT (lanes 7 and 8). The sequence TGGTGGC migrates similarly to $[dTGGGGT]_4$; whereas, the sequence CGGTGGT shows a major band migrating much slower. We conclude that TGGTGGC forms prevalently a tetrameric assembly, in agreement with the reported tetramer G-quadruplex formation for this sequence (23), whereas CGGTGGT forms a higher order assembly. The polarity of the sequence therefore appears crucial in the higher order structure formation.

The migration profile of dCGGGGT has also been investigated to determine whether the additional central thymine played any role in the higher order structure formation. Two bands are observed for dCGGGGT, the faster one migrating as the tetrameric $[dTGGGGT]_4$, and the slower one migrating like the structure formed by CGGTGGT.

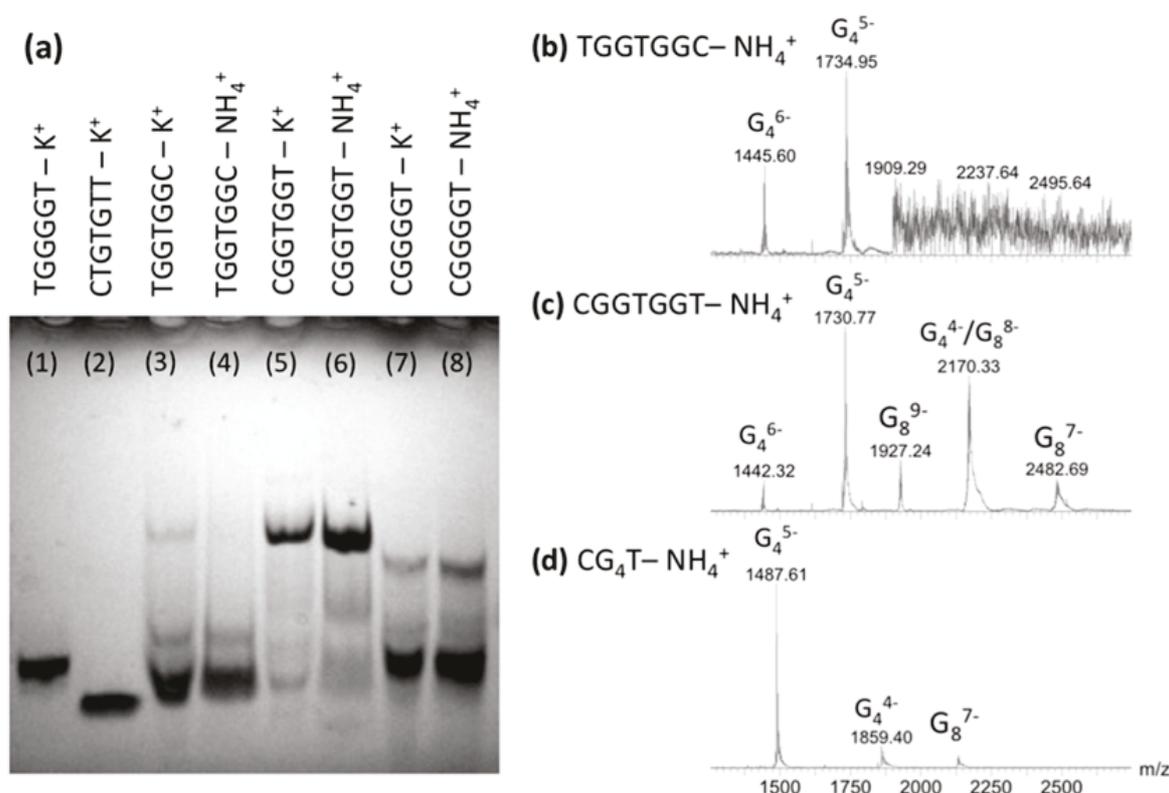


Figure 1. Stoichiometry of self-assemblies determined by (a) PAGE for sequences TGGGGT (lane 1), CTGTGTT (lane 2), TGGTGGC (lanes 3 and 4), CGGTGGT (lanes 5 and 6) and CGGGGT (lanes 7 and 8), and by (b–d) ESI-MS for TGGTGGC (b), CGGTGGT (c) and CGGGGT (d). In the peak annotation G_n^{z-} , n indicates the number of strands and z indicates the charge.

As the PAGE results are similar in NH_4^+ and in K^+ cations, we therefore used ESI-MS, which can only be carried out in NH_4^+ , to determine unambiguously the strand stoichiometry of the higher order assemblies formed by CGGTGGT and CGGGGT. The ESI-MS spectra of TGGTGGC, CGGTGGT and CGGGGT, each separately annealed at 6mM in 150 mM NH_4OAc ($\text{pH} = 7.0$) and diluted to 80 μM final strand concentration, are shown in Figure 1b-d, respectively. All three sequences show peaks corresponding to a tetramolecular assembly (noted G_4), but the spectra of CGGTGGT and CGGGGT also display peaks corresponding to an octameric assembly (noted G_8). We therefore assign the slow migrating PAGE band to an octameric self-assembly. In terms of relative intensities, although native PAGE and ESI-MS experiments were performed on samples obtained by dilution of the same annealed stock solution, the octamer is the darkest band in the gels of CGGTGGT (lanes 5 and 6), whereas the ESI-MS peaks of the octamer have lower intensities than the tetramer peaks. The analysis of the electrospray response factors in the kinetics ESI-MS experiments (Figures 1b-d) indeed reveals the octamer has a lower ESI-MS response than the tetramer in the spectrometer used, and that the octamer is indeed the major species in solution.

4.1.2 The octamer contains stacked G-tetrads

All G-rich sequences studied here are expected to form G-quadruplex structures. Note that 'G-quadruplex' means that the structure contains stacked G-quartets, not necessarily that the assembly is a tetramer. CD was used to probe the relative orientation of the bases in the tetramer $[\text{TGGTGGC}]_4$ and in the octamer $[\text{CGGTGGT}]_8$. The overall CD profiles recorded in K^+ (Figure 2a) and NH_4^+ (Figure 2b) are in agreement with the formation of G-quadruplex structures (24, 25) both in K^+ and NH_4^+ media, showing a positive maximum at about 260 nm and a negative minimum close to 240 nm, which are characteristics of head-to-tail arrangement of guanines, as typically found in parallel oligodeoxynucleotides G-quadruplexes (26-28). Interestingly, the CD spectra of CGGTGGT also show a weak, but clearly visible negative band centered at 290 nm. This minimum suggests that additional base stacking interactions could be involved in the formation of the higher ordered quadruplex species by allowing the quadruplex multimerization via end-to-end stacking. A typical heteropolar stacking in G-quadruplexes formed by strands with no inverted polarity usually results in a positive band at 290 nm (27), but negative bands at 290 nm have been observed in octameric self-assemblies of lipophilic guanosine derivatives (29). The type of stacking present in our octamers could therefore resemble that present in those lipophilic octamers.

Because the CD signature is peculiar and does not allow to conclude unambiguously to a G-quadruplex structure, we decided to probe cation incorporation into the octamer. Indeed, successive G-quartets incorporate monovalent cations (here K^+ or NH_4^+) by coordination between eight guanines. High-resolution ESI-FTICRMS was used to count the number of ammonium cations per tetramer and per octamer (Figure 2c). Cation incorporation was found in all cases, confirming G-quadruplex formation both in the tetramer and the octamer. The major peaks correspond to three cations per tetramer $[\text{dCGGTGGT}]_4$ and eight cations per octamer $[\text{dCGGTGGT}]_8$. The number of cations is indicative of the number of stacked quartets. Interestingly, the tetramer $[\text{dCGGTGGT}]_4$ contains three ammonium cations, whereas the tetramer $[\text{dTGGTGGC}]_4$ contains four (data not shown), suggesting a structural difference depending on the strand polarity. However, the incorporation of eight ammonium

cations in the octamer suggests that within the octamer, each tetrameric subunit is now capable of incorporating four ammonium cations.

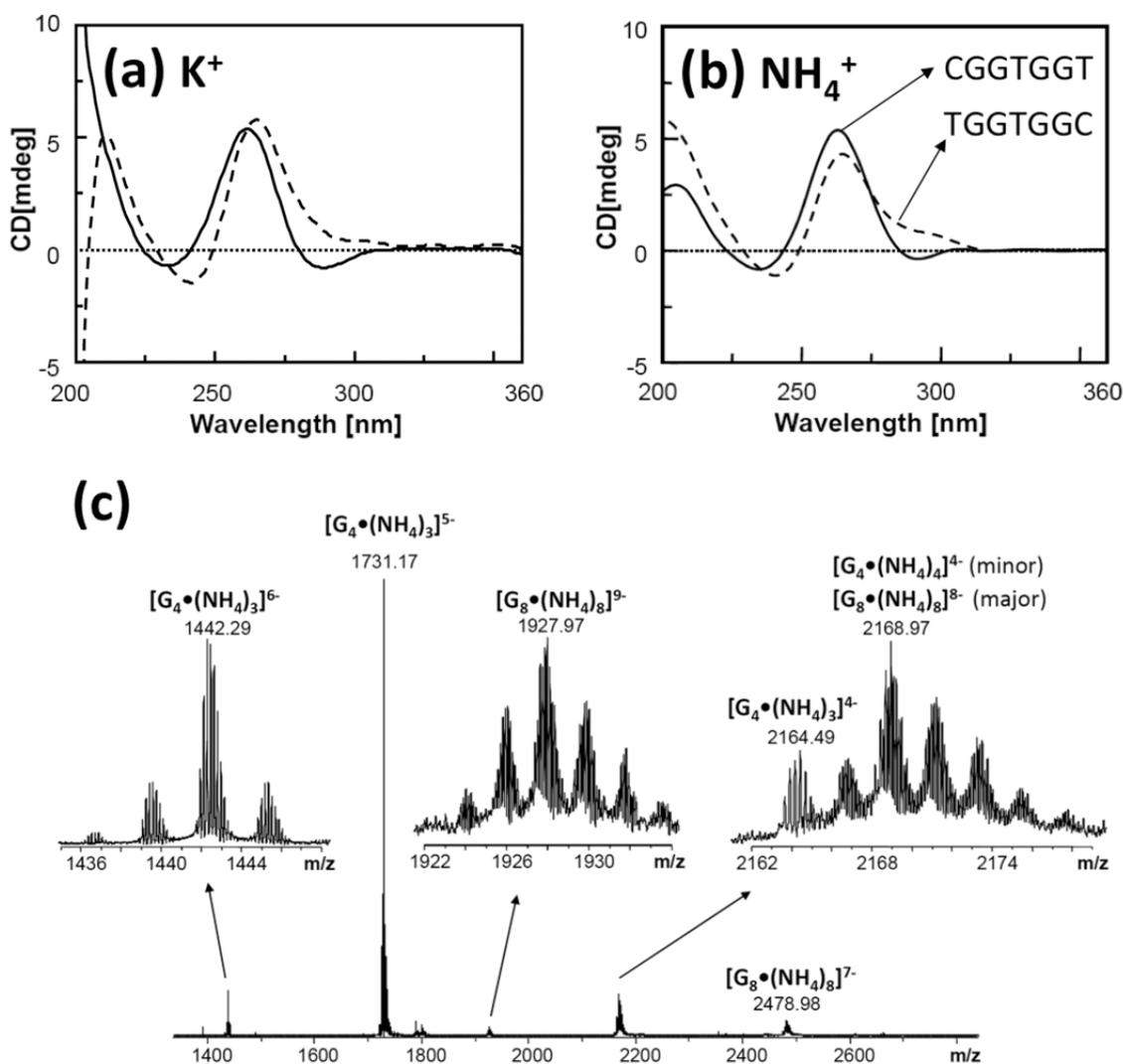


Figure 2. G-quadruplex formation probed by (a and b) CD in 100mM KCl (a) and 150mM NH₄OAc (b) of the sequences CGGTGGT (solid line) and TGGTGGC (dotted line), and by (c) counting the number of ammonium cations incorporated in the tetramer (noted G₄) and the octamer (noted G₈) of CGGTGGT using high resolution ESI-FTICR mass spectrometry.

4.1.3 Kinetics of octamer formation studied by ESI-MS

We followed the kinetics of octamer formation from the CGGTGGT solution to monitor tetramer and octamer formation as a function of the time elapsed after ammonium acetate addition. This experiment is performed with an internal standard, and therefore also allows to ascertain the relative response of each species. Typical ESI-MS spectra recorded after different reaction times (1 h, 24 h and 17 days) are shown in Figure 3 (a-c). The relative intensity of the monomer (G₁) decreases with time, the relative intensity of the tetramer (G₄) first increases, then decreases, while the relative intensity of the octamer (G₈) increases slowly, but nevertheless remains lower than the intensity of the tetramer (Figure 3d). Next, the mass balance equation procedure is used to calculate the relative response factors of the monomer, tetramer and octamer. The dimer and trimer response factors could not be evaluated because

there are not enough independent equations to calculate them (their intensity variation profile is close to that of the monomer). Because they were minor species in the spectrum, they were ignored in the mass balance equation. We found that, in our experimental conditions, the response of the tetramer G_4 (=Q1) is 8-fold larger than the response of the octamer G_8 (=2Q), 2.9 times larger than the response of the monomer G_1 . Figure 3e is obtained from Figure 3d, after correction of the relative intensities by the relative response of each species ($R_{(G_1)}/R_{(G_4)} = 0.348$, $R_{(G_8)}/R_{(G_4)} = 0.125$). Evolution of the concentrations of each species as a function of time shows that the octamer becomes the major species after ≈ 80 h, and is formed at $> 80\%$ after 168 h (7 days).

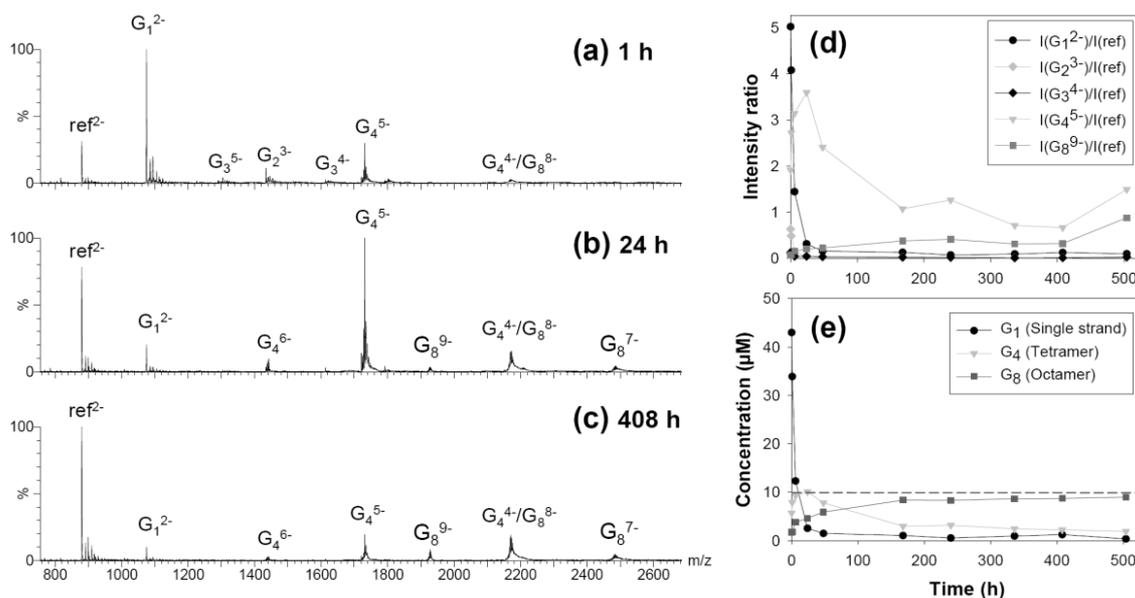


Figure 3. Kinetics of dCGGTGGT octamer formation in ammonium acetate. The total strand concentration in the ageing solution was $6 \mu\text{M}$ (left at RT). The total dCGGTGGT concentration injected was $80 \mu\text{M}$, with $10 \mu\text{M}$ dT₆ as internal reference. (a-c) ESI-MS spectra obtained after 1h, 24h, and 17 days, respectively. (d) Time evolution of the intensity ratio of each species relative to the reference. (e) Time evolution of the concentrations of single strand, tetramer, and octamer, after correction for the relative response of each species ($R_{(G_1)}/R_{(G_4)} = 0.348$, $R_{(G_8)}/R_{(G_4)} = 0.125$).

4.1.4 Influence of sample preparation procedure

In all experiments described above, the octamer was obtained by annealing the strand CGGTGGT at 6 mM single-strand ODN concentration at $\text{pH} = 7.0$ in 100 mM K^+ or 150 mM NH_4^+ . CD experiments carried out using a higher K^+ concentration (1 M) show that higher cation concentration further favours the formation of the structure characterized by the negative band at 290 nm (data not shown). We also tested the influence of the solution pH, to test whether cytosine protonation might favour the formation of the new structure, but no clear influence of pH was found (data not shown). However, because of the octamer stoichiometry, it is anticipated that the strand concentration upon annealing would have a great influence on the formation kinetics. Figure 4a shows the CD spectra recorded on CGGTGGT annealed at 6.0 mM or 0.2 mM ODN concentration in 1 M K^+ . The presence of the negative band at 290 nm was strictly dependent from ODN concentration: it is present when the sample is annealed starting from the 6 mM stock solution, whereas it is not observed when the sample is obtained by annealing the 0.2 mM ODN (Figure 4a). However, the CD spectrum of the species formed from the 0.2 mM ODN solution resembles that of a regular tetramolecular G-quadruplex. The influence of the strand

concentration, therefore, indicates slow formation kinetics for the octamer. To study the kinetics of octamer formation in the 6 mM stock solution, and to probe the nature of the reaction intermediates, we recorded ESI-MS spectra as a function of the time after addition of cation in the ODN solution, at room temperature. This experiment could be carried out only in ammonium acetate (150 mM). The details of the methodology, representative ESI-MS spectra at different reaction times, and the relative intensities are presented in Supplementary Figure 3. Briefly, the single-stranded dT₆ strand (0.75 mM) is added to the 6 mM single-stranded dCGGTGGT in water. Quadruplex formation is initiated by the addition of ammonium acetate (final concentration: 150 mM) at room temperature. At each time point, an aliquot of the sample is diluted to 80 μ M final dCGGTGGT single strand concentration, and injected in the ESI-MS. We therefore assume that, once formed, the octamer resists the dilution step. The signal intensity of the reference is used to normalize the intensity variations of the dCGGTGGT monomer, tetramer and octamer signal, and deduce their respective response factors (Figure 3).

Figure 4b presents the same data as Figures 3e, but showing the time evolution of the relative proportions of single strand, tetramer and octamer, after correction for the relative response of each species, expressed as the fraction of single strand in each form. The single strand is quickly converted into a tetramer, but the tetramer-to-octamer conversion is slower. The dimerization is therefore the rate-limiting step in the octamer formation. The results also show that, although the octamer is not the most abundant peak in the mass spectra (see Figures 1b-d and 2c), it is nevertheless the predominant species. The predominance of the octamer at long time scales also confirms that, once formed, the octamer resists the dilution step.

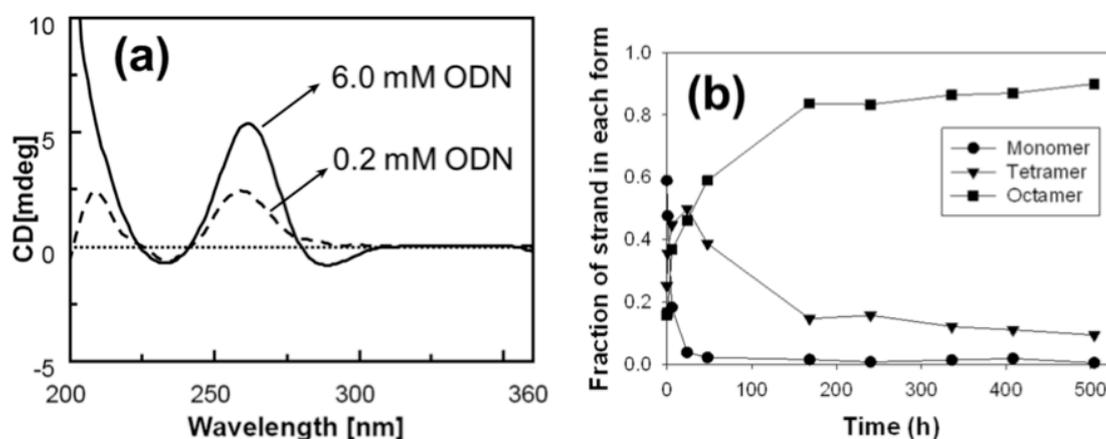


Figure 4. Influence of the sample preparation conditions on the octamer formation. (a) Influence of strand concentration: CD spectra of CGGTGGT in 1.0M K⁺ buffer annealed, respectively, at 6mM (solid line) and 200 mM (dotted line) single-strand concentration. (b) Kinetics of tetramer and octamer formation of 6mM CGGTGGT in 150mM ammonium acetate at room temperature, as determined by time-resolved ESI-MS.

4.1.5 NMR investigation of the octamer structure

The ¹H NMR spectrum of CGGTGGT in 100 mM K⁺ buffer at 25°C recorded 1 hour after the annealing (data not shown) is characterized by the presence of four well-resolved signals in the 11.0-11.6 ppm region attributable to the exchange-protected imino protons involved in the formation of Hoogsteen hydrogen bonds of four G-tetrads (30-32), in close analogy with what was previously observed for the parallel quadruplex structure formed by TGGTGGC (23). Furthermore, the count of

quadruplex NMR signals is the same as what must be expected for the NMR spectrum of the corresponding single strand, thus suggesting the formation of a highly symmetric parallel G-quadruplex possessing a 4-fold symmetry. However, the aromatic region of ^1H NMR spectrum is populated by more than the expected seven signals belonging to the H8/H6 aromatic protons of nucleobases. In order to discriminate whether the minor peaks were due to sample impurities or to the coexistence in solution of alternative minor conformations of CGGTGGT, we recorded the ^1H NMR spectrum also at 65°C, and the high-temperature spectrum was clearly in agreement with the second hypothesis as evidenced by the presence of just seven well-defined signals in the aromatic region of the spectrum. We therefore studied the influence of time elapsed after the annealing, of K^+ concentration, and of pH (data not shown) on the 1D NMR spectra. In agreement with the CD data, the NMR results confirmed that the best medium for the structural study of the new octameric structure formed by CGGTGGT requires 1.0 M K^+ and long folding times (Figure 5a), and that the change of pH from 7.0 to 4.5 does not affect the folding of the structure. The combined analysis of 2D NOESY and TOCSY spectra (700 MHz, 25°C, 1 M K^+ , 15 days after the annealing) allowed us to get the complete assignment (Table 1) of both exchangeable and non-exchangeable protons of CGGTGGT.

	H ₆ /H ₈	H ₅ /CH ₃	NH	NH _{2i}	H1'	H2'	H2''	H3'	H4'	H5'/H5''
C1	7.54	5.48			5.63	2.28	2.10	4.66	4.13	3.87/3.95
G2	8.14		11.56	9.82	6.09	2.72	2.96	5.10	4.49	4.18
G3	7.56		11.13	9.80	6.12	2.55	2.94	5.07	4.50	4.30
T4	7.03	0.72	9.40		5.84	2.24	2.46	4.92	4.34	4.25
G5	8.00		11.26	9.82	5.98	2.72	2.84	5.10	4.49	4.20/4.30
G6	7.73		11.05	9.81	6.27	2.58	2.71	4.98	4.52	4.11/4.25
T7	7.34	1.64			6.08	2.73	2.21	4.51	4.08	4.25

Table 1. Proton chemical shift (ppm, relative to H₂O at 4.78 ppm) for CGGTGGT at 25°C, pH 7.0, 1M K^+ . NH_{2i} refers to the internal (i) H-bonded amino protons of guanosines.

The sequential connectivities between the aromatic protons (Figure 5b) were traced, as well as the self and sequential NOE connectivities between H8/H6 base protons and H1', H2', H2'' and H3' sugar protons. The intensities of intranucleotide H8/H1' NOE cross-peaks (Figure 5c) indicate that all Gs adopt an *anti* glycosidic conformation. This finding together with the observation for each G-tetrad of a G_iH8/G_jN1H NOE cross-peak (Figure 5f), with $i = j$ (where i and j represent the nucleotide labels), and with the observation of sequential G₂N1H/G₃N1H and G₅N1H/G₆N1H cross-peaks (Figure 5e), definitely allowed us to identify the presence of 4-fold symmetric parallel quadruplex moieties in the higher ordered structure formed by CGGTGGT when annealed in K^+ -containing buffer. In close analogy with what observed by Patel and Hosur for the quadruplex structure formed by TGGTGGC (23), we observed the formation of a central T-tetrad as ascertained by the appearance of the diagnostic T4NH/T4Me NOE cross-peak (Figure 5g) and supported by the observation of the strong shielding ring current experienced by T4 methyl protons that are significantly upfield-shifted when compared to the corresponding signals of the 'free' T7 (0.72 versus 1.64 ppm). Although the overall observed NOEs account for a 'classical' tetramolecular parallel quadruplex, two peculiar NOE evidences drove us toward the disclosure of the higher ordered octamer structure.

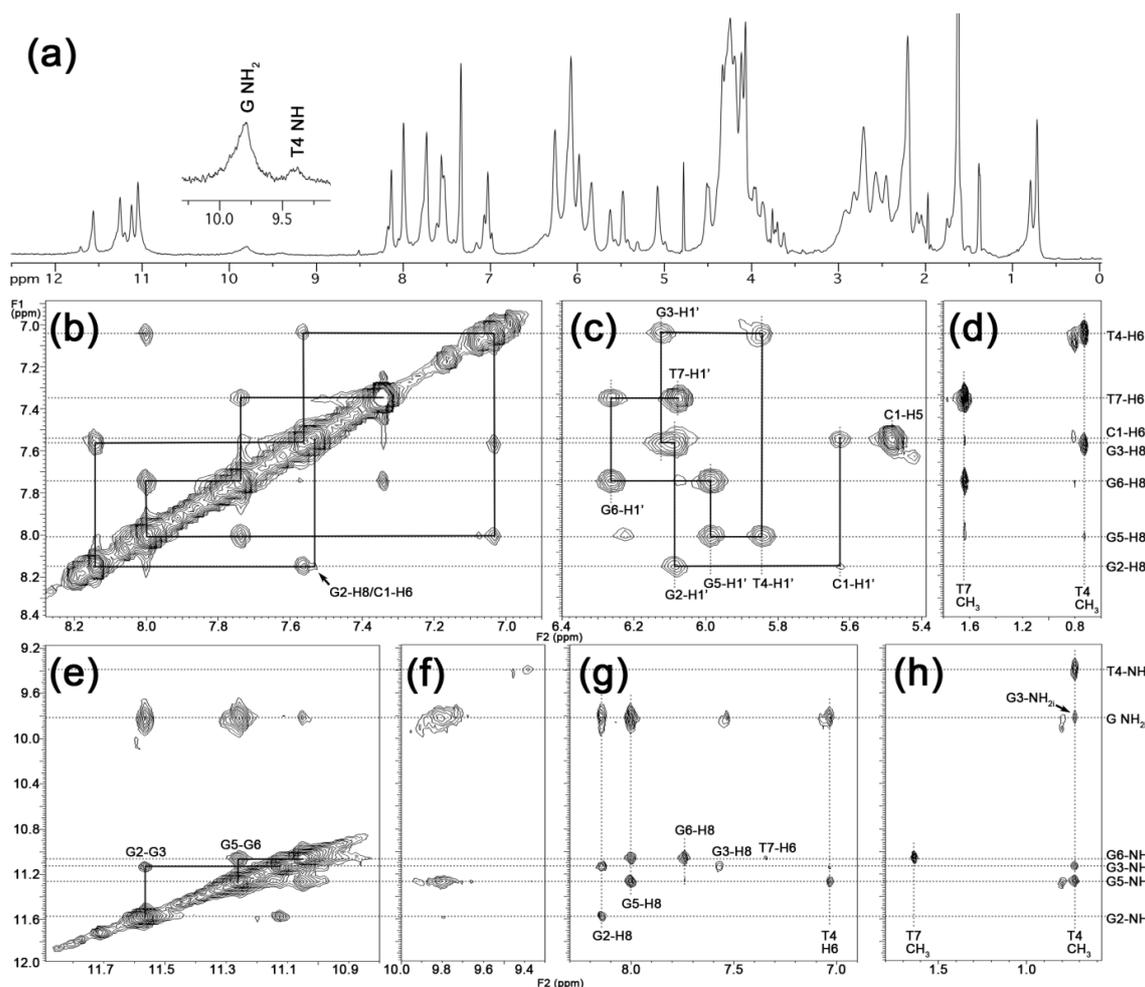


Figure 5. Study of the octamer structure by NMR. (a) ^1H -NMR spectra recorded at 25°C of CGGTGGT in $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1, annealed at 6mM strand concentration in 1M K^+ at pH=7.0. (b–g) Expansions of 2D NOESY spectrum of CGGTGGT (25°C, 200 ms mixing time, 1M K^+ , pH=7.0). The sequential H8–H6 and self-peaks involving H' protons have been labeled in (b) and (c), respectively. (d) and (g) show NOE cross-peaks involving T4/T7 methyl protons and H8/H6 or imino and amino protons, respectively. The T-tetrad diagnostic peaks between T4Me and T4NH and sequential T4Me–G3NH, T4Me–G3NH_{2i} and T4Me–G5NH are observable in panel (g). Imino-imino NOE correlations are shown in (e). The sequential imino-imino NOE correlations have been drawn in (e). NOE correlations between aromatic H8/H6 and imino and amino protons are shown in (f).

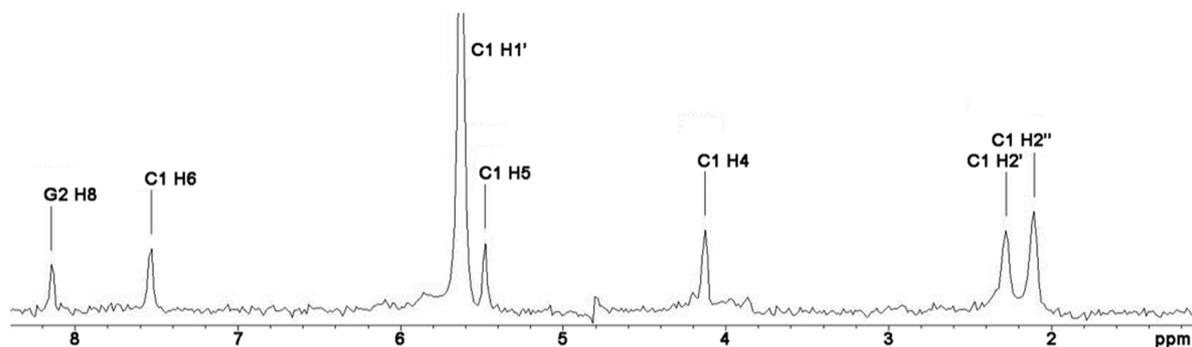


Figure 6. 1D projection of 2D NOESY spectrum (25 °C, 200 ms mixing time) of CGGTGGT taken at 5.66 ppm, corresponding to C1 H1' protons.

In this regard, in Figure 6 it is shown the 1D cross-section of the NOESY spectrum (200 ms mixing time) of CGGTGGT taken at 5.66 ppm (C1H1' protons). This figure shows that the NOE between C1H1' and C1H5 has almost the same intensity than that between C1H1' and C1H6. These NOEs are in turn stronger than the sequential NOE C1H1'-G2H8, but weaker, as expected, than those between C1H1' and self H2'/H2". The relative intensities of the aforementioned NOEs involving C1H1' require that the C1H1'-C1H5 distance is in the same range of the intranucleotide C1H1'-C1H6 distance (2.31-3.74 Å depending on the glycosyl torsion angle). This experimental distance is compatible only with the occurrence of an intermolecular dipolar coupling between C1 bases, because the corresponding intranucleotide C1H1'-C1H5 distance would be in the range 4.64-5.49 Å.

This finding, together with the observation of the very weak sequential NOE between C1H6 and G2H8 (marked with the arrow in Figure 5b), disclosed that, in the octamer, which is in fact a "dimeric G-quadruplex (**2Q**)", each C1 base does not stack over the flanking G2 tetrad, as one may expect, but projects outside the quadruplex scaffold in such a way to allow dipolar couplings with one of the four C1 bases belonging to the other tetramer. This unusual arrangement might be favoured by the formation of additional C1NH2-G2N3 and G2NH2-C1N3 hydrogen bonds, thus allowing the formation of an unusual G2(:C1):G2(:C1):G2(:C1):G2(:C1) octad (Figure 11a) in each of the two tetrameric quadruplexes (**1Q**).

4.1.6 Investigating the role of the central base on the structuring of the octamer.

In summary, the above mentioned results allowed us to establish the optimal solution conditions for the self-folding and self-stimulated formation of the octamer **2Q**: i) higher single-strand ODN concentration for the annealing step (6 mM better than 0.2 mM) and ii) higher K⁺ concentration (1 M better than 0.1 M). Furthermore, we did not observe clear evidences of pH influence on quadruplex structuration in the pH range 4.5-7.0. Starting from these results, a speculation about the role of the base sequence towards the dimerization process was carried out. Assuming the presence of C1 as a cardinal point for the formation of **2Q**, the replacement of the middle base in dCGGTGGT sequence was undertaken, to evaluate its importance on the stability and rate of formation of the putative octamers; thus, dCGGAGGT (**CA**), dCGGCGGT (**CC**), and dCGGGGGT (**CG**) sequences were taken into account. All the samples were obtained by annealing the strands at 6 mM single-strand ODN concentration, and their **structuration** into quadruplex species were monitored by PAGE and spectrometric and spectroscopic measurements at different time-points from the annealing. Again, already at day 1 from the annealing, clear evidences about the formation of higher order assemblies were noticed in native PAGE experiments, as shown in Figure 7a. For this experiment, all sequences were annealed in 1M K⁺ buffer. CGGTGGT (Figure 7a, lane 1) was loaded on the gel at day 15 from the annealing, to be used as a **2Q** marker, and T₆ (lane 5) and TGGGGT (lane 6) were used as a single-stranded 6-mer and a tetramolecular quadruplex size-markers, respectively. All the CGGXGGT sequences taken into account (Figure 7a, lanes 2-4) showed a major band corresponding to the formation of 2Q-like secondary structures. In detail, the migration profile of **CA** showed a unique slow-migrating band (Figure 7a, lane 2), whereas two bands were observed for the migration profiles of **CC** and **CG** (Figure 7a, lanes 3 and 4). To verify unequivocally the strand stoichiometry of the structures, we performed mass spectrometry studies (ESI-MS) (Figure 7b-e). The ESI-MS spectra were recorded one day after the annealing under conditions that preserved non-covalent interactions. Interestingly, we observed mass

peaks corresponding to tetrameric assemblies (noted as **Q**) and also peaks corresponding to octameric assemblies (noted as **2Q**), perfectly in line with what observed by PAGE data. Consequently, all the slower-migrating bands were ascribed to octameric self-assemblies.

Melting profiles recorded by UV spectrophotometry at 295 nm (A_{295}) revealed that all sequences presented a noteworthy thermal stability, with T_m in the range of 63-72 °C (Figure 8, on the left). Moreover, the **CG** sequence resulted to have a so marked thermal stability to not allowing the melting. To probe the relative formation of G-quadruplex structures, as well as the orientation of the bases in the octamers, CD measurements were carried out. The overall CD profiles, recorded in K^+ buffer after 1 day from the annealing (Figure 8, on the right), showed a positive maximum close to 260 nm and a negative minimum at about 240 nm, thus suggesting a head-to-tail arrangement of guanines, which is characteristic of parallel G-quadruplexes.

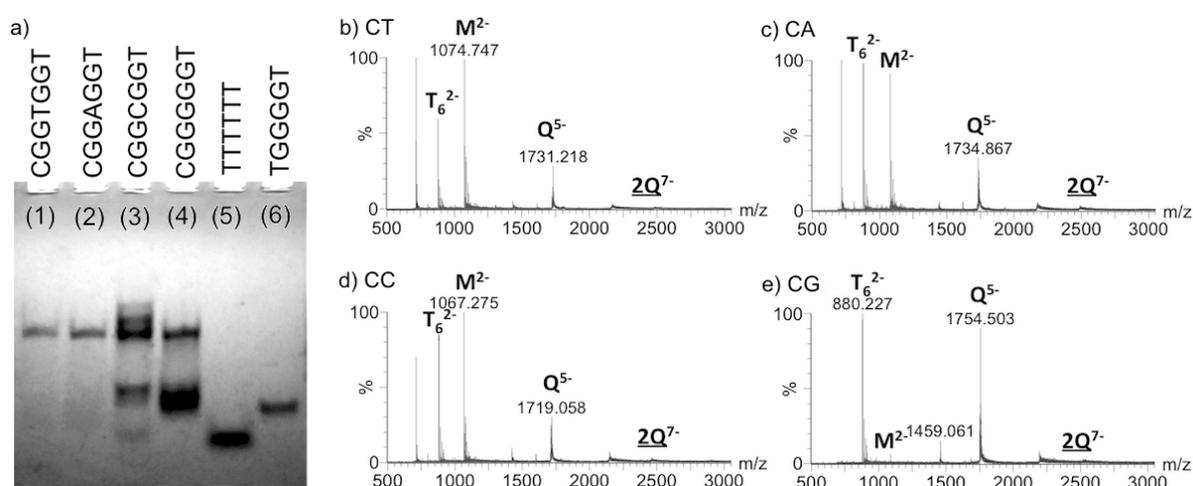


Figure 7. Stoichiometry of self-assemblies determined by a) PAGE experiments after 1 day from the annealing, 1M K^+ buffer for sequences CGGTGGT (lane 1), CGGAGGT (lane 2), CGGCGGT (lane 3), CGGGGGT (lane 4), TTTTTT (lane 5), and TGGGGT (lane 6), and by (b-e) ESI-MS for CGGTGGT (b), CGGAGGT (c), CGGCGGT (d), CGGGGGT (e). In the peak annotation G_n^{z-} , n indicates the number of strands and z indicates the charge.

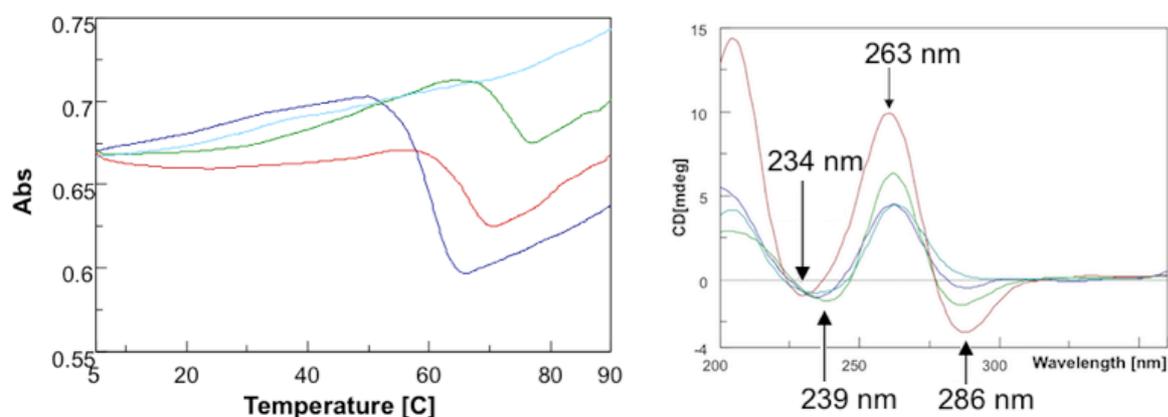


Figure 8. On the left: UV Melting profiles of CGGAGGT (red line, $T_m = 65.2^\circ\text{C}$), CGGCGGT (green line, $T_m = 72.2^\circ\text{C}$), CGGTGGT (purple line, $T_m = 63^\circ\text{C}$), and CGGGGGT (light blue line, no melting). On the right: CD profiles, after 1 day from the annealing, of CGGAGGT (red line), CGGCGGT (green line), CGGTGGT (purple line), and CGGGGGT (light blue line).

Furthermore, the CD profiles of **CT**, **CA**, and **CC** showed a peculiar negative band at about 290 nm. As previously discussed, this particular Cotton effect denotes the G-quadruplex dimerization through a non-covalent 5'-5' stacking allowed by the head-to-head stacking between G₂-tetrads of two different tetramolecular parallel quadruplexes **1Q**. In this context, a similar dimerization process was suggested for **CA** and **CC** sequences. As regards the **CG** sequence, no minimum at about 290 nm was observed, thus we do not exclude that a different dimerization process might occurs for this case.

In order to monitor the structural behaviour of **CA**, **CC**, and **CG** during the time, we reacquired the CD and ¹H NMR spectra at different time points after the annealing procedure (Figure 9).

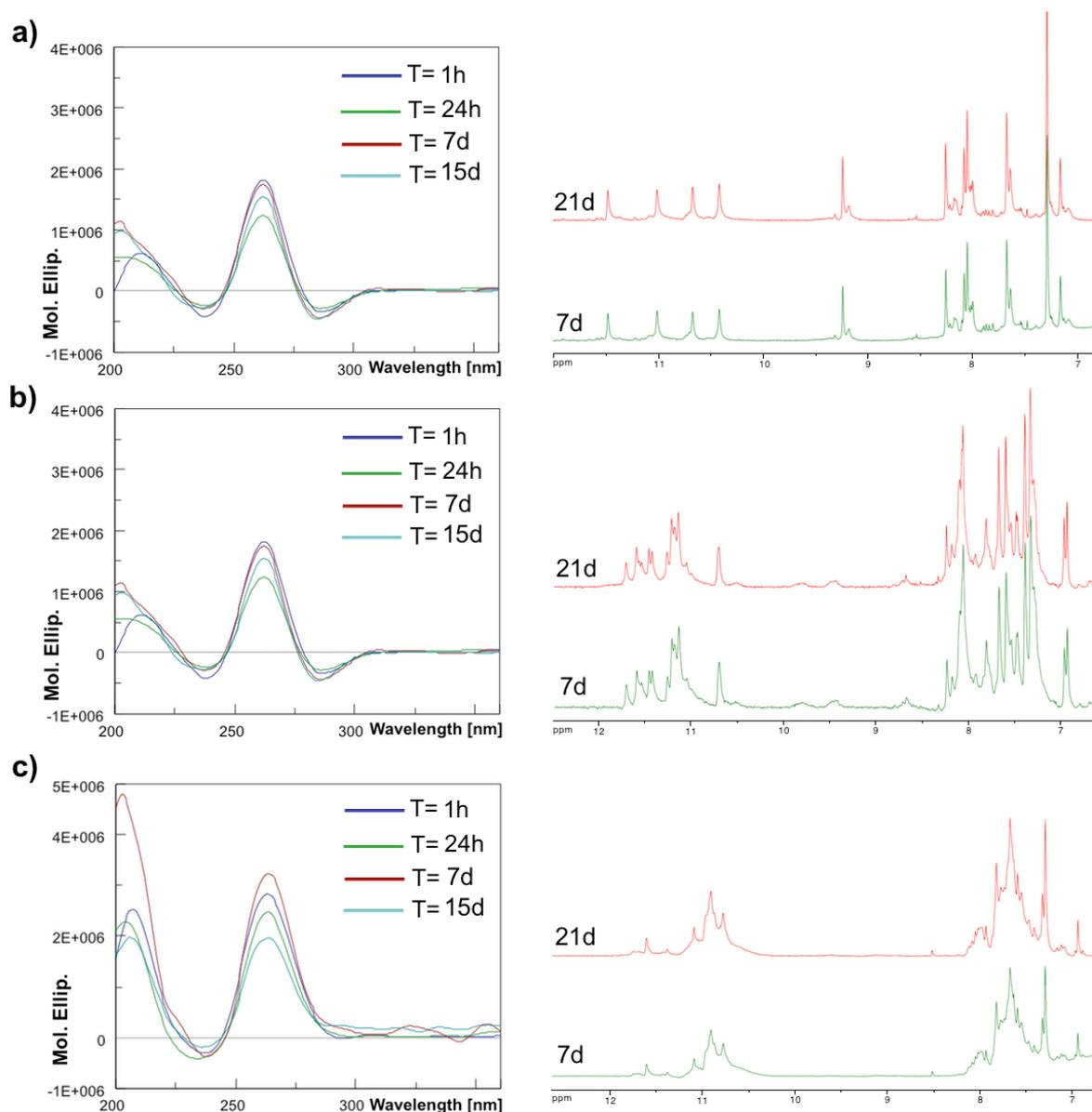


Figure 9: On the left: CD profiles after 1 hour (blue), 1 day (green), 7days (red) and 15 days (cyan) from the annealing of the sequences CGGAGGT (a), CGGCGGT (b), and CGGGGGT (c) sequences in K⁺ 1M. On the right: ¹H NMR spectra recorded at 25°C of the sequences CGGAGGT (a), CGGCGGT (b), and CGGGGGT (c) sequences in H₂O/D₂O 9:1, after 7days (green) and 21 days (red) from the annealing, that occurs at 6mM ODN single-strand concentration in K⁺ 1M at pH = 7.0.

The ^1H NMR spectra of **CA** showed that a single **2Q** species was present in solution, whereas formation of multiple species was observed for **CC** and **CG**. Notably, **CC** also showed multiple bands in PAGE, while **CG** showed exclusively the **Q** and **2Q** bands. In detail, ^1H NMR spectra of **CA** show a unique, and highly symmetric G-quadruplex conformation, characterized by four well-resolved NH resonances in the 10.0-11.5 ppm region (consistent with the presence of four stable G-tetrads), and eight aromatic signals in the 7.0-8.5 ppm region. In this context, the unusual downfield shifted non-exchangeable resonance, detected at 9.20 ppm, accounted for A4-H2 protons, and it was diagnostic of a stabilised A-tetrad with a putative 6-NH₂ to N1 hydrogen bonding geometry (see Figure 6c in Chapter 1), as reported by Searle et al. (3). Moreover, the signal at 9.20 ppm gave strong NOEs with G3NH1, and G5NH1, and appreciable NOEs with G3H8, A4H8, and G5H8 (data not shown), supporting the hypothesis that strong stacking interactions across G3-A4-G5 planes occurred. In addition to the above, the complete assignment (Table 2) of both exchangeable and non-exchangeable protons was obtained by the combined analysis of 2D NOESY and TOCSY spectra. For all nucleotides were traced self and sequential NOE connectivities from H8/H6 base protons to H1', H2'/H2'', and H3' sugar protons. Furthermore, the intensities of intranucleotide H8/H1' NOE cross-peaks showed that all purine glycosidic torsion angles were in *anti* conformation (33, 34). Taken together, the NMR features accounted for the presence of a parallel G-quadruplex structure, as also supported by CD data.

	H ₆ /H ₈	H ₂ /H ₅ /Me	NH	NH _{2i}	H1'	H2'	H2''	H3'	H4'	H5'/H5''
C1	7.23	5.41	-	6.50	5.50	2.05	1.93		4.37	3.75/3.99
G2	8.03	-	11.45	9.43/6.05	5.99	2.61	2.92			
G3	7.11	-	10.97	9.43/5.99	5.99	2.21	2.93			
A4	8.22	9.20	-	-	5.99	2.77	2.60			
G5	8.00	-	10.38	9.34/5.42	6.03	2.59	2.69		4.21	
G6	7.63	-	10.61	9.34	6.14	2.47	2.59	4.89	4.46	3.97/4.17
T7	7.25	1.56	-	-	5.97	2.09	2.09	4.42	3.97	

Table 2. Proton chemical shift (ppm, relative to H₂O at 4.78 ppm) for **CA** at 25°C, pH 7.0, 1M K⁺. NH_{2i} refers to the internal (i) H-bonded amino protons of guanosines.

Furthermore, the analysis of the intensity of NOE correlation peaks between the C1 and G2 bases of **CA**, confirmed that the dimerization pathway of **CA** was similar to that of **CT**, via the formation of the unusual G2(:C1):G2(:C1):G2(:C1):G2(:C1) octad, with the difference that the former took place in a direct and faster manner. As concerned **CC** and **CG** sequences, the ^1H NMR spectra were characterized by the presence of several signals in the 10.5-12.0 ppm region attributable to the exchange-protected imino protons involved in the formation of Hoogsteen hydrogen bonds of four G-tetrads (30-32). Furthermore, the aromatic region of ^1H NMR spectra were populated by more than the expected seven signals belonging to the H8/H6 aromatic protons of nucleobases, due to the coexistence in solution of alternative minor conformations. Thus, given the presence of more than one species in solution, and given that no species were selected as function of time or temperature, it was not possible to achieve the complete assignment of both exchangeable and non-exchangeable protons for **CC** and **CG** species.

4.2 Discussion

The structure adopted by the sequence dCGGTGGT must account for the following observations. It was a highly symmetrical octamer, formed from two parallel-stranded tetramers. In the octamer, each tetramer contained a central T-tetrad and the G-tetrads were stacked in head-to-tail arrangement. However, at the octamer's interface between the two tetramers, the stacking of the G-tetrads was likely to be different, as determined from CD. Interlocked structures like that formed by the (dGCGGTGGT)₄ tetramer (35) were initially taken in consideration to build an octamer model, but were discarded due to NMR inconsistencies and symmetry-related considerations. Unresponsiveness to pH changes also led us to discard structures involving protonated cytosines. The NOEs involving C1H1' point to the formation of a G2(:C1):G2(:C1):G2(:C1):G2(:C1) octad (Figure 10a) in each of the two tetrameric quadruplexes. The formation of the planar octad system would allow for further π - π stacking between C1 bases, thus adding more 'hydrophobic glue' between the two G-quadruplex tetramers that form the octamer. Figure 10b shows the energy minimized model of the octamer formed by stacking of two tetramers, each containing a 5'-G(:C):G(:C):G(:C):G(:C) octad. The details of molecular modelling are given in the Experimental Session. The overall topology of each tetramer is largely similar to that of the tetramolecular quadruplex formed by TGGTGGC (see Figure 8 in reference (23)), with the central T-tetrad causing a small decrease in the intertwining of the helices. Figure 10b also evidence that the octamer structure is very compact, regardless the presence of two stacked octads. These are seen to determine only a slight increase in the thickness of the octamer at the dimerization interface. Figure 10c-d shows the head-to-head stacking between the two octads (c) and the stacking between tetrads G2-G3 (d). Although C1 and G2 bases involved in the stacked octads appear slightly rotated with respect to the tetrad plane, Figure 10c shows good stacking between the six-membered rings of G2 bases and between the stacked C1 bases. The electric transition moments corresponding to the guanine absorption band at approximately 250 nm are indicated in Figure 10c and d. The stacking interactions in the model allow to explain the CD data: G2-G3 stacking resembles the standard head-to-tail stacking encountered in parallel stranded G-quadruplexes. However, the octad stacking gives rise to a heteropolar stacking that is distinct from those typically encountered in antiparallel G-quadruplexes. Therefore, the negative CD band at 290 nm was interpreted as being due to the octad stacking shown in Figure 10c (29). One must note, however, that the formation of the octamer by the sequence dCGGTGGT is slow. ESI-MS kinetics experiments in ammonium acetate provide some insight into the mechanism of octamer formation. The first step is the formation of a tetramer that incorporates only three ammonium ions. This means either that the tetramer of [dCGGTGGT]₄ has either a central or a terminal tetrad disrupted, or that the tetramer contain slipped strands and hence incomplete tetrad formation (36). The rate-limiting step is the conversion of tetramer to octamer, and in contrast with the tetramer, the octamer incorporates four cations per tetrameric subunit. We therefore conclude that the rate-limiting step is the rearrangement of the tetramer into a fully aligned parallel structure that is amenable to octad formation and stacking thereof.

Regarding the structure adopted by the sequence dCGGAGGT, the collected data were in agreement with those hitherto discussed. Also for **CA** the formation of a highly symmetrical octamer, formed from two parallel-stranded tetramers, took place. In the octamer, each tetramer contained a central A-tetrad, and the G-tetrads were stacked in head-to-tail arrangement. Also in this case, we founded evidences of the

formation of a G₂(:C₁):G₂(:C₁):G₂(:C₁):G₂(:C₁) octad (Figure 10a) in each of the two tetrameric quadruplexes: i) the negative CD band at 290 nm due to the heteropolar stacking between the octads, and ii) the interquadruplexes NOEs involving C1H1' and C1H5. The peculiarities of the **2Q** structure formed by **CA** lied in the rate of formation and in the high thermal stability. In this context, the PAGE data (Figure 7a) denoted the presence of a single structure, and the UV melting spectrum underlied the presence of a very stable structure with a melting temperature of about 65 °C. Unlike **CT** and **CC**, the dimeric species of **CA** was selected already at day 1 from the annealing, thus suggesting that the dimerization process does not provide for the formation of **1Q** intermediates, as assumed for the dimerization process of **CT**. It stands to reason that this peculiar behaviour was due to the presence of the A-tetrads.

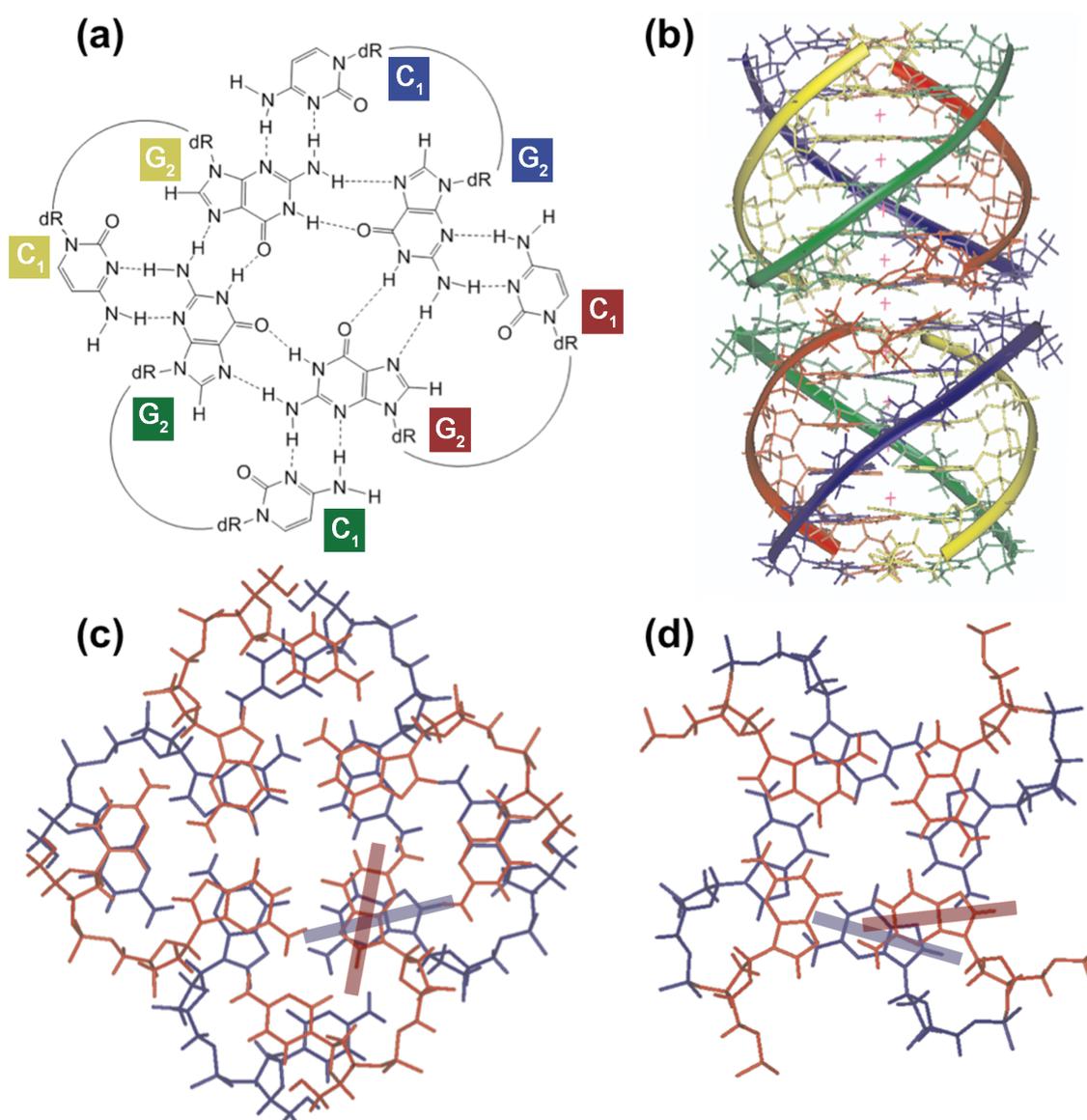


Figure 10. Structural model for the (dCGGTGGT)₈ octamer. (a) Proposed structure of the G₂(:C₁):G₂(:C₁):G₂(:C₁):G₂(:C₁) octad. (b) Molecular model of the octamer: a different color was used for each ODN strand, and potassium ions appear as red crosses. (c) Detail of stacking of the octads. (d) Detail of stacking of G₂-G₃ tetrads. The transition moments of guanosines corresponding to the absorption band at approximately 250nm are shown as colored bars in (c) and (d).

These conferred additional stabilization to the **2Q** structure, and seem to promote the formation of higher-order G-quadruplex structures. As regard the **CC** sequence, CD, MS, and PAGE data suggested that the **2Q**-like octameric G-quadruplex complex was present in solution, but they also suggested that other different structures were also present. Therefore, it is conceivable that the **2Q** structuring of **CC** was slower than those of **CT** and **CA**. Regarding **CG**, PAGE and MS data suggested the formation of a higher order G-quadruplex structure, but both CD and NMR could not confirm this finding. Specifically, two distinct bands attributable to a monomeric (**Q**) and a dimeric (**2Q**) species were visible in PAGE run, but the CD data did not allow to link the dimer formation to the head-to-head dimerization pathway. Furthermore, since the species in solution had a very high melting temperature (at 90°C there was still no fusion), it was not possible to select one of the two species as a function of temperature; similarly, no selection was made in function of time. Therefore, the formation of a higher order G-quadruplex structure was confirmed, but the dimerization process is still not clear.

All together, these results therefore suggest that in the sequence CGGXGGT (where X = A, C, G, or T), the cytosine at 5'-end is capable of inducing the structuring of higher order structures by the formation of octads, and, when adenine is at central position, the structuring of the dimer occurs in a direct and faster manner, leading to the formation of the more stable nanostructure (T_m around 65°), among those one taken into account.

4.3 Conclusions

The present study demonstrates that the formation of a DNA-based nanostructure consisting of a dimer of tetramolecular G-quadruplexes (**2Q**) can occur when the 5'-cytosine-ending CGGXGGT (where X = A, C, G, or T) DNA sequences are annealed in the presence of K^+ ions. The NMR and molecular modelling data of CGGTGGT and CGGAGGT disclosed that the structuring of **2Q** is allowed by the formation of one unusual octad arrangement, in each of the two **1Q** units. Up to now, dimerization of G-quadruplexes had only be reported via guanine-adenine heptads (13), and recently via a shared uracil tetrad (37). This finding adds further possibilities for the design of novel quadruplex-based DNA functional nanostructures, and further complexity to the mechanisms of folding and assembly of G-rich sequences. It is particularly important for in vitro studies of artificially designed (e.g. aptamers) or biologically relevant sequences.

4.4 Experimental Session

4.4.1 DNA sample preparation

DNA oligonucleotides TGGGGT, CTGTGTT, TGGTGGC, CGGTGGT, CGGGGT, CGGAGGT, CGGCGGT, CGGGGGT, and TTTTTT were chemically synthesized on an 8909 DNA/RNA synthesizer (Applied Biosystem), or purchased from Eurogentec (Belgium) for mass spectrometry experiments. DNA concentration expressed in strand molarity was calculated using a nearest-neighbour approximation for the absorption coefficients of the unfolded species. Ammonium acetate was obtained from a 5M solution from Fluka (Molecular Biology grade). Stock solutions of 6mM of each ODN were obtained by dissolving the lyophilized samples in 150mM NH_4OAc or 100mM and 1.0M K^+ buffers (10mM KH_2PO_4 supplemented with 90mM or 990mM KCl , respectively). Hundred percent D_2O (Armar chemicals) or H_2O/D_2O (9:1, v/v)

was used for the preparation of 6mM solutions to be used for NMR experiments. All ODN samples were annealed by heating at 90°C for 15 min and then cooling and storing at 5°C. The pH was adjusted to 7.0 or 4.5 using HCl/KOH or AcOH/NH₄OH aqueous solutions before the annealing procedure.

4.4.2 Polyacrylamide gel electrophoresis

The molecular size of structures were probed using non denaturing polyacrylamide gel electrophoresis (PAGE) as previously described (23). ODN samples of 10 mM, obtained by diluting the annealed stock solutions just before the experiment, were loaded on a 15% polyacrylamide gel supplemented with 20mM of KCl or NH₄OAc. Glycerol of 10% was added just before loading. The gels were run at 26°C at constant voltage (100 V) for 2.5 h. The bands were visualized by UV shadowing and after 'Syber green' coloration.

4.4.3 Electrospray mass spectrometry

A Q-TOF Ultima Global mass spectrometer (Waters, Manchester, UK) was used to characterize the samples in ammonium acetate, and to perform a kinetics experiment. ODN samples (80 mM single-strand concentration), obtained as described in the PAGE paragraph with the addition of 10% methanol, were infused at 4 ml/min in the electrospray (ESI) source that was operated in negative ion mode (capillary voltage = -2.2 kV). Temperatures of ESI source and nitrogen desolvation gas were 80 °C and 100 °C, respectively. The source voltages were optimized for the best compromise to observe the single strand, tetramer and octamer: Cone = 80 V, RF Lens 1 = 80 V, collision energy = 10 V, source pressure = 3.06 mbar. For the kinetics experiments, the reference strand dT₆ was added to the dCGGTGGT strand, so as to follow the intensities of the reactant (dCGGTGGT single strand, noted G₁), the intermediates (dimer = G₂, trimer = G₃, tetramer = G₄ = **1Q**) and the product (octamer = G₈ = **2Q**) relative to that of the reference. The spectra were smoothed (mean function, 2 × 12 channels), background-subtracted (polynomial order = 50, 0.1% below curve, tolerance 0.010), centroided (peak width 20 channels, reconstructed area of the top 80% of the peak), and recalibrated using the reference as a lock mass (dT₆²⁻ monoisotopic peak at m/z = 880.1530). The relative response factors of the monomer, the tetramer **1Q** and the octamer **2Q** were obtained by data processing as described elsewhere (38). Briefly, for each stoichiometry observed for dCGGTGGT, one charge state was selected based on the absence of possible overlap with other stoichiometries: G₁²⁻ (four most intense isotopes), G₂³⁻ (no ammonium adducts, to avoid overlap with [G₄•(NH₄)_n]⁶⁻), G₃⁴⁻ (no adducts), G₄⁵⁻ (sum of all peaks from 0 to 4 NH₄⁺ adducts), and G₈⁹⁻ (uneven charge state to avoid overlap with tetramer, sum of all peaks from 2 to 9 NH₄⁺ adducts). By assuming that the electrospray response of each CGGTGGT stoichiometry and dT₆ are proportional throughout the experiment, and by using the mass balance equation applied to CGGTGGT (the total strand concentration being equal to 80 μM), the relative response factors of each stoichiometry can be determined. These response factors are then used to recalculate the concentration of each stoichiometry at each time point. The user's choice of the charge state and/or adducts considered has no influence on the final results in terms of concentrations, because the calculated response factors provide the necessary correction (38). In addition, a Bruker Apex-Qe9.4 T ESI-FTICR mass spectrometer was used to confirm the charge states from the isotopic distributions by recording high-resolution data, and to confirm the number of ammonium cations retained in each complex. The instrument voltages were

chosen so as to minimize collision-induced dissociation and therefore keep the non-covalent complexes intact, including the inner ammonium cations: cap exit = -22 V, skimmer = -15 V, octapole offset = -4 V and coll cell trap = +4 V.

4.4.4 Circular dichroism

Circular dichroism (CD) spectra were measured at 25°C on a Jasco J-715 spectropolarimeter equipped with a Peltier Jasco JPT423S, using a 0.1-cm path length quartz cuvette with a reaction volume of 500 μ l. CD spectra were recorded at different time points after the annealing on 40 μ M ODN samples annealed at strand concentration of 200 μ M or 6.0 mM. Spectra were averaged over three scans. A buffer baseline was subtracted from each spectrum and the spectra were normalized to have zero at 360 nm.

4.4.5 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) data were collected on a Varian ^{UNITY}INOVA 500 MHz spectrometer equipped with a broadband inverse probe with z-field gradient, and on a Varian ^{UNITY}INOVA 700 MHz spectrometer equipped with a triple resonance cryoprobe. The data were processed using the Varian VNMR software package. One-dimensional NMR spectra were acquired as 16384 data points with a recycle delay of 1.0 s at 5, 25 and 65°C. Data sets were zero filled to 32768 points prior to Fourier transformation and apodized with a shifted sine bell squared window function. Two-dimensional NMR spectra were all acquired at 25°C using a recycle delay of 1.2 s. NOESY spectra were acquired with mixing times of 100, 200 and 300 ms. TOCSY spectra were recorded with the standard MLEV-17 spin-lock sequence and a mixing time of 80 ms. For the experiments in H₂O, water suppression was achieved by including a double pulsed-field gradient spin-echo (DPFGSE) module (39, 40) in the pulse sequence prior to acquisition. In all 2D experiments, the time domain data consisted of 2048 complex points in t₂ and 400 fids in t₁ dimension. NMR samples were prepared at 6mM strand concentration in 0.6 ml of 100 mM or 1.0 M K⁺ buffer obtained as above described. Sequence-specific resonance assignments for CGGTGGT were obtained by using NOESY and TOCSY spectra following the standard procedure (41, 42). The self as well as sequential NOE connectivities from H₈ to H_{1'} and H_{2'/H2''} were traced for all the nucleotides. The connectivities followed the standard patterns as for B-DNA, apart from noticeable exceptions between C₁ and G₂, and complete assignments of base and sugar protons was obtained.

4.4.6 Molecular modeling

The starting structure was generated using the builder module of Insight II (2005) program (Accelrys, Inc) on an Intel PC workstation running Red Hat Enterprise Linux. The model was parameterized according to the AMBER force-field, and calculations were performed using a distance-dependent macroscopic dielectric constant of 4.0 and an infinite cut-off for non-bonded interactions to partially compensate for the lack of solvent. One of the two “monomer G-quadruplex” of **2Q** was built using the coordinates of the NMR solution structure of [d(TGGTGGC)]₄ tetramolecular quadruplex (PDB entry number 1EMQ) (23) as starting point. The four 5'-end T residues were replaced by deoxycytidines, while 3'-end C residues were replaced by deoxythimidines, then the G₂(:C₁):G₂(:C₁):G₂(:C₁):G₂(:C₁) octad was built by aligning the four C₁ around the G₂-tetrad plane in such a way that each C₁ had its 3'-OH close to the 5'-phosphate of the flanking G₂, and its N₃ and NH₂ atoms close enough to the purine ring of G₂ (belonging to the anti-clockwise-adjacent strand) to

form hydrogen bonds with G2-NH₂ and G2-N3 atoms, respectively. Finally, each strand was completed by joining the C1 3'-OH with the G2 5'-phosphate. The coordinates of C1 and G2 atoms were then energy-minimized by using 1000 cycles of steepest descent method followed by conjugate gradient calculation until convergence at 0.001 rms was reached. Atoms of G3-T7 bases were kept frozen throughout the minimization. The resulting minimized structure was used to obtain the second "monomer quadruplex" by copy & paste. The **2Q** model was then finally assembled by aligning the two identical quadruplexes head-to-head on top of each other with the standard B-DNA rise (3.4 Å) and twist (36°).

The whole system was then subjected to 1000 cycles of steepest descent energy minimization followed by 2 ns of restrained molecular dynamics at 298 K. From the last nanosecond of the dynamic run, 11 different structures were selected on the basis of their low energies, proper covalent geometries and overall 4-fold symmetry. The selected structures were then subjected to restrained energy minimization with steepest descent and conjugate gradient methods until convergence to an rms gradient of 0.1 Kcal·mol⁻¹ was reached. In these calculations 526 distance, 144 hydrogen bond, and 32 glycosidic torsion angle constraints were used according to NMR data. NOE cross peaks were integrated and the intensities were translated into interproton distances using the intranucleotide C1-H5/C1-H6 cross peak intensity as the reference (2.46 Å). Cross peaks involving non-exchangeable protons were classified as strong (20 Kcal mol⁻¹ Å⁻² force constant), medium (10 Kcal mol⁻¹ Å⁻²) and weak (5 Kcal mol⁻¹ Å⁻²) on the basis of their relative intensities and the interproton distances were restrained to 1.8-3.0, 2.5-4.0 and 3.5-5.5 Å, respectively. Hydrogen bonds restraints were defined using upper and lower distance limits of 2.0 and 1.7 Å for hydrogen-acceptor distances, and 3.0 and 2.7 Å for donor-acceptor distances (100 Kcal mol⁻¹ Å⁻²). Glycosidic torsion angles of guanosines were constrained to a range of -160°/-70° (200 Kcal mol⁻¹ Å⁻²) as required for a G-quadruplex having all Gs in the anti orientation as arose from NOESY spectra. All restraints were finally removed and a further unrestrained energy minimization was performed as described above. The average RMSD value was 0.55. Finally, the length of the relaxed structure was measured and found equal to about 40 Å, or rather 4 nanometers.

4.5 References

- (1) Parkinson, G. (2006) Fundamentals of quadruplex structures, in *Quadruplex Nucleic Acids* pp 1–30, RSC Publishing, London, UK.
- (2) G. Oliviero, J. A., N. Borbone, A. Galeone, M. Varra, G. Piccialli, L. Mayol. (2006) Synthesis and characterization of DNA quadruplexes containing T-tetrads formed by bunch-oligonucleotides. *Biopolymers* 81, 194–201.
- (3) M. S. Searle, H. E. L. W., C. T. Gallagher, R. J. Grant, M. F. G. Stevens. (2004) Structure and K⁺ ion-dependent stability of a parallel-stranded DNA quadruplex containing a core A-tetrad. *Organic & Biomolecular Chemistry* 2, 810-812.
- (4) N. Bhavesh, P. P., S. Karthikeyan, R. Hosur. (2004) Distinctive features in the structure and dynamics of the DNA repeat sequence GGCGGG. *Biochemical and Biophysical Research Communications* 317, 625–633.
- (5) A. Matsugami, T. O., S. Uesugi, M. Katahira. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *Journal of Biological Chemistry* 278, 28147–28153.

- (6) S. Haider, G. N. P., S. Neidle. (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophysical Journal* 95, 296–311.
- (7) Y. Xu, T. I., K. Kurabayashi, and M. Komiyama. (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angewandte Chemie International Edition* 48, 7833–783.
- (8) H. Sotoya, A. M., T. Ikeda, K. Ouhashi, S. Uesugi, and M. Katahira. (2004) Method for direct discrimination of intra- and intermolecular hydrogen bonds, and characterization of the G(:A):G(:A):G(:A):G heptad, with scalar couplings across hydrogen bonds. *Nucleic Acids Research* 32, 5113–5118.
- (9) Y. Kato, T. O., H. Mita, and Y. Yamamoto. (2005) Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAGn) (n=3–5). *Journal of the American Chemical Society* 127, 9980–9981.
- (10) A. T. Phan, V. K., J. B. Ma, A. Faure, M. L. Andreola, and D. J. Patel. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proceedings of the National Academy of Sciences* 102, 634–639.
- (11) N. Zhang, A. G., A. Majumdar, A. Kettani, N. Chernichenko, E. Skipkin, and D. J. Patel. (2001) V-shaped scaffold: a new architectural motif identified in an A(GGGG) pentad-containing dimeric DNA quadruplex involving stacked G(anti)G(anti)G(anti)G(syn) tetrads. *Journal of Molecular Biology* 311, 1063–1079.
- (12) A. Kettani, A. G., A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. A. Jones, and D. J. Patel. (2000) A dimeric DNA interface stabilized by stacked A(GGGG)A hexads and coordinated monovalent cations. *Journal of Molecular Biology* 297, 627–644.
- (13) A. Matsugami, K. O., M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi, and M. Katahira. (2001) An intermolecular quadruplex of (GGA)₄ triplet repeat DNA with GGGG tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *Journal of Molecular Biology* 313, 255–269.
- (14) Phan, A. T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS Journal* 277, 1107–1117.
- (15) Y. Krishnan-Ghosh, D. L., S. Balasubramanian. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *Journal of the American Chemical Society* 126, 11009–11016.
- (16) L. Spindler, M. R., I. Drevensek-Olenik, N. M. Hessari, and M. Webba da Silva. (2010) Effect of base sequence on G-wire formation in solution. *Journal of Nucleic Acids*.
- (17) D. Sen, a. W. G. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* 334, 364–366.
- (18) F. Rosu, V. G., C. Houssier, P. Colson, and E. De Pauw. (2002) Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry* 16, 1729–1736.
- (19) Y. Xu, K. K., and M. Komiyama. (2008) G-quadruplex formation by human telomeric repeats-containing RNA in Na⁺ solution. *Journal of American Chemical Society* 130.

- (20) G. W. Collie, G. N. P., S. Neidle, F. Rosu, E. De Pauw, and V. Gabelica. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *Journal of the American Chemical Society* 132, 9328-9334.
- (21) A. T. Phan, M. G., and J. L. Leroy. (2001) Investigation of unusual DNA motifs. *Methods in Enzymology* 338, 341-371.
- (22) N. Borbone, J. A., G. Oliviero, V. D'Atri, V. Gabelica, E. De Paw, G. Piccialli, and L. Mayol. (2011) d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. *Nucleic Acids Research* 39, 7848-7857.
- (23) P. K. Patel, a. R. V. H. (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Research* 27, 2457-2464.
- (24) J. L. Mergny, A. T. P., L. Lacroix. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Letters* 435, 74-78.
- (25) J. L. Mergny, J. L., L. Lacroix, S. Amrane, J. B. Chaires. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Research* 33, e138.
- (26) C. C. Hardin, A. G. P., K. White. (2001) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* 56, 147-194.
- (27) J. L. Mergny, A. D. C., A. Ghelab, B. Saccà, L. Lacroix. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Research* 33, 81-94.
- (28) S. Masiero, R. T., S. Pieraccini, S. De Tito, R. Perone, A. Randazzo, G. P. Spada. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Organic & Biomolecular Chemistry* 8, 2683-2692.
- (29) C. Graziano, S. M., S. Pieraccini, M. Lucarini, G. P. Spada. (2008) A cation-directed switch of intermolecular spin-spin interaction of guanosine derivatives functionalized with open-shell units. *Organic Letters* 10, 1739-1742.
- (30) R. Jin, B. L. G., C. Wang, R. A. Jones, K. J. Breslauer (1992) Thermodynamics and structure of a DNA tetraplex: a spectroscopic and calorimetric study of the tetramolecular complexes of d(TG3T) and d(TG3T2G3T). *Proceedings of the National Academy of Sciences* 89, 8832-8836.
- (31) J. Feigon, K. M. K., F. W. Smith. (1995) ¹H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods in Enzymology* 261, 225-255.
- (32) Feigon, J. (1996) DNA triplexes, quadruplexes, and aptamers. *Encyclopedia of NMR* 3, 1726-1731.
- (33) F. Aboul-Ela, A. H. M., D. M. Lilley. (1992) NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG4T). *Nature* 360, 280-282.
- (34) F.W. Smith, J. F. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. *Nature* 356, 164-168
- (35) J. L. Mergny, A. D. C., S. Amrane, M. Webba da Silva. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Research* 34, 2386-2397.
- (36) F. Rosu, V. G., H. Poncelet, E. De Pauw. (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry. *Nucleic Acids Research* 38, 5217-5225.

- (37) P. Sket, J. P. (2010) Tetramolecular DNA Quadruplexes in Solution: Insights into Structural Diversity and Cation Movement. . *Journal of the American Chemical Society* 132, 12724–12732.
- (38) V. Gabelica, F. R., E. De Pauw. (2009) A Simple Method to Determine Electropray Response Factors of Noncovalent Complexes. *Analytical Chemistry* 81, 6708–6715.
- (39) T. L. Hwang, A. J. S. (1995) Water suppression that works. Excitation sculpting using arbitrary wave forms and pulsed field gradients. *Journal of Magnetic Resonance, Series A* 112, 275–279.
- (40) Dalvit, C. (1998) Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. . *Journal of Biomolecular NMR* 11, 437–444.
- (41) Wuthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York, NY.
- (42) V. Hosur, G. G., H. T. Miles (1988) *Application of two-dimensional NMR spectroscopy in the determination of solution conformation of nucleic acids. Magnetic Resonance in Chemistry.*

APPENDIX

▪ **Book Chapter:**

1. “*Computational Methods for Studying G-quadruplex Nucleic Acids*”. Barira Islam, **Valentina D’Atri**, Miriam Sgobba, Jay Husby, and Shozeb Haider. Book Chapter in *Guanine Quartets: Structure and Application*. (2013), pages 194-211. DOI:10.1039/9781849736954-00194

▪ **Publications:**

2. “*Structure–phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia*”. Shozeb Haider, Barira Islam, **Valentina D’Atri**, Miriam Sgobba, Chetan Poojari, Li Sun, Tony Yuen, Mone Zaidi, and Maria I. New. *Proceedings of the National Academy of Sciences of the United States of America*, (2013), 110, 7, 2605-2610.

3. “*Investigating the role of T7 and T12 residues on biological properties of Thrombin Binding Aptamer: enhancement of anticoagulant activity by a single nucleobase modification*”. Nicola Borbone, Maria Rosaria Bucci, Giorgia Oliviero, Elena Morelli, Jussara Amato, **Valentina D’Atri**, Stefano D’Errico, Valentina Vellecco, Giuseppe Cirino, Gennaro Piccialli, Caterina Fattorusso, Michela Varra, Luciano Mayol, Marco Persico, and Maria Scutto. *Journal of Medicinal Chemistry*, (2012), 55, 10716-10728.

4. “*New anti-HIV aptamers based on tetra-end-linked DNA Gquadruplexes: effect of the base sequence on anti-HIV activity*”. **Valentina D’Atri**, Giorgia Oliviero, Jussara Amato, Nicola Borbone, Stefano D’Errico, Luciano Mayol, Vincenzo Piccialli, Shozeb Haider, Bart Hoorelbeke, Jan Balzarini, and Gennaro Piccialli. *Chemical Communication*, (2012), 48, 9516-9518.

5. “*Insight into Pyridinium Chlorochromate Chemistry: Catalytic Oxidation of Tetrahydrofuran Compounds and Synthesis of Umbelactone*”. Vincenzo Piccialli, Sabrina Zaccaria, Giorgia Oliviero, Stefano D’Errico, **Valentina D’Atri**, and Nicola Borbone. *European Journal of Organic Chemistry*, (2012), 23, 4293-4305.

6. “*Insight into the Conformational Arrangement of a Bis-THF Diol Compound Through 2D-NMR Studies and X-Ray Structural Analysis*”. Vincenzo Piccialli, Sabrina Zaccaria, Roberto Centore, Angela Tuzi, Nicola Borbone, Giorgia Oliviero, Stefano D’Errico, and **Valentina D’Atri**. *Journal of Chemical Crystallography*, (2012), 42, 4, 360-365.

7. “*d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads*”. Nicola Borbone, Jussara Amato, Giorgia Oliviero, **Valentina D’Atri**, Valérie Gabelica, Edwin De Pauw, Gennaro Piccialli, and Luciano Mayol. *Nucleic Acids Research*, (2011), 39, 17, 7848-7857.

8. “*Probing the Reactivity of Nebularine N1-oxide. A Novel Approach to C-6 C-substituted Purine Nucleosides*”. Stefano D’Errico, Vincenzo Piccialli, Giorgia

Oliviero, Nicola Borbone, Jussara Amato, **Valentina D'Atri**, and Gennaro Piccialli. *Tetrahedron*, (2011), 67, 34, 6138-6144.

9. "Solid-Phase Synthesis and Pharmacological Evaluation of Novel Nucleoside-Tethered Dinuclear Platinum(II) Complexes". Stefano D'Errico, Giorgia Oliviero, Vincenzo Piccialli, Jussara Amato, Nicola Borbone, **Valentina D'Atri**, Francesca D'Alessio, Rosa Di Noto, Francesco Ruffo, Francesco Salvatore, and Gennaro Piccialli. *Bioorganic and Medicinal Chemistry Letters*, (2011), 21, 19, 5835-5838.

▪ **PhD activities spent in the following laboratories:**

March 2010 – October 2011 / October 2012 – February 2013: Dipartimento di Chimica delle Sostanze Naturali, Facoltà di Farmacia. Università degli studi di Napoli Federico II. Napoli, IT.

November 2011 – September 2012: Visiting Research Associate at CCRCB (Centre of Cancer Research and Cell Biology). School of Medicine, Dentistry & Biomedical Sciences, Queen's University Belfast. Belfast, UK.

▪ **Congress Communications:**

1. "Identification and structural characterization of a new higher-ordered G-quadruplex structure". Second training school on G-quadruplexes – Self-assembled guanosine structures for molecular electronic devices. Spa, Belgium, 11 – 15 September 2011.

2. "Identificazione e caratterizzazione strutturale di una nuova tipologia di G-quadruplex di ordine superiore". Giornate Scientifiche – Polo delle Scienze e delle Tecnologie per la Vita. Naples, Italy, 25 - 26 November 2010.

Cite this: *Chem. Commun.*, 2012, **48**, 9516–9518

www.rsc.org/chemcomm

COMMUNICATION

New *anti*-HIV aptamers based on tetra-end-linked DNA G-quadruplexes: effect of the base sequence on *anti*-HIV activity†Valentina D'Atri,^a Giorgia Oliviero,^a Jussara Amato,^a Nicola Borbone,^{*a} Stefano D'Errico,^a Luciano Mayol,^a Vincenzo Piccialli,^b Shozeb Haider,^c Bart Hoorelbeke,^d Jan Balzarini^d and Gennaro Piccialli^a

Received 19th June 2012, Accepted 9th August 2012

DOI: 10.1039/c2cc34399a

This communication reports on the synthesis and biophysical, biological and SAR studies of a small library of new *anti*-HIV aptamers based on the tetra-end-linked G-quadruplex structure. The new aptamers showed EC₅₀ values against HIV-1 in the range of 0.04–0.15 μM as well as affinities for the HIV-1 gp120 envelope in the same order of magnitude.

The first stage of the HIV infection requires the entry of human immunodeficiency virus (HIV) into host cells. This stage involves the sequential interaction of the virion surface glycoprotein gp120 with the CD4 glycoprotein and a chemokine receptor, either CCR5 or CXCR4, on the host cell surface.¹ The CD4 glycoprotein is expressed on the surface of T-lymphocytes, monocytes, dendritic cells and brain microglia, and its expression makes these cells a target for HIV *in vivo*. Furthermore, an interesting function of CD4 binding is to induce conformational changes in gp120 that allow binding to the co-receptor, which is essential for viral entry.² The third variable region of gp120 (from now on designated V3 loop) is a pivotal component of the co-receptor binding site, typically consisting of a 35 amino acid-loop (range 31 to 39), closed by two cysteines that form a disulfide bridge. The crystal structure of gp120 complexed with the CD4 receptor and a neutralizing antibody (PDB ID 2B4C) revealed that the V3 loop is extended away from the gp120 protein and it is involved in co-receptor binding and selection, acting as a “molecular hook” that organizes associations within the viral spike.³ The importance of gp120 and the role of the V3 loop in HIV-1 entry and pathogenesis have led to the recent pursuit of drugs targeted against it. One of the most studied alternatives is the use of aptamer technology.⁴ Aptamers are single stranded DNA or RNA molecules selected among

large pools of nucleic acid sequences for their ability to bind selectively and with high affinity to a biomedically relevant target. For their properties, aptamers can be considered as the nucleic acid analogues of antibodies. Like the antibodies, several aptamers proved to be valuable diagnostic tools⁵ and promising therapeutics.^{6,7}

In 1994 H. Hotoda (SA-1042)⁸ and J. R. Wyatt (ISIS 5320)⁹ independently reported the first *anti*-HIV aptamers targeted against the V3 loop. In both cases the ODN sequences were chemically modified to improve their resistance against nucleases by capping the 5'-ends with DMT groups (SA-1042) or by using the phosphorothioate backbone (ISIS 5320). Subsequently, other analogues of SA-1042 were obtained and tested,^{10,11} and the 6-mer TGGGAG ODN (known as “Hotoda's sequence”) bearing 3,4-dibenzyloxybenzyl groups at the 5'-ends and 2-hydroxyethylphosphates at the 3'-ends (R-95288) showed the highest *anti*-HIV activity.¹¹ CD investigations on ISIS 5320, R95288 and their analogues suggested their structuration into parallel tetramolecular quadruplexes. Following biophysical studies established that the presence of aromatic^{12,13} moieties at 5'- and/or 3'-ends of TGGGAG dramatically enhances the rate of formation of quadruplex complexes. Moreover, the overall stability of quadruplex complexes was found to correlate with the reported *anti*-HIV activity.¹²

In 2010, with the aim to overcome the unfavourable entropic factor and to stabilize the A-tetrad in Hotoda's analogues, we synthesized a series of new monomolecular *anti*-HIV aptamers by using the Tetra-End-Linker (TEL) strategy proposed by some of us in 2004.^{14,15} Several TEL-(TGGGAG)₄ aptamers were prepared and analyzed in order to probe the influence of lipophilic groups and TEL size and position on the structural and *anti*-HIV properties of the resulting TEL-quadruplexes.¹⁶ The results showed that (i) the presence of the TEL at either 5'- or 3'-ends was required for the *anti*-HIV activity, and (ii) lipophilic *tert*-butyldiphenylsilyl (TBDPS) groups at the 5'-ends strongly enhanced both the stability of TEL-quadruplexes and their *anti*-HIV activity. The EC₅₀ of the best aptamer (**I**, Fig. 1), bearing TBDPS groups at the 5'-ends and the longer TEL at the 3'-ends was 0.082 ± 0.04 μM.

With the aim to further improve the *anti*-HIV activity of TEL-ODN aptamers, we used **I** as a molecular synthon for the synthesis of a series of new TEL-ODN aptamers embodying

^a Dipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli Federico II, Via D. Montesano 49, I-80131 Napoli, Italy. E-mail: Nicola.borbone@unina.it

^b Dipartimento di Chimica, Università degli Studi di Napoli Federico II, Via Cynthia 4, I-80126 Napoli, Italy

^c Centre for Cancer Research and Cell Biology, Queen's University of Belfast, 97 Lisburn Road, BT9 7BL, Belfast, UK

^d Rega Institute for Medical Research, KU Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium

† Electronic supplementary information (ESI) available: Synthesis and characterization of TEL-ODNs, as well as CD, SPR and docking studies and biological evaluation assays. See DOI: 10.1039/c2cc34399a

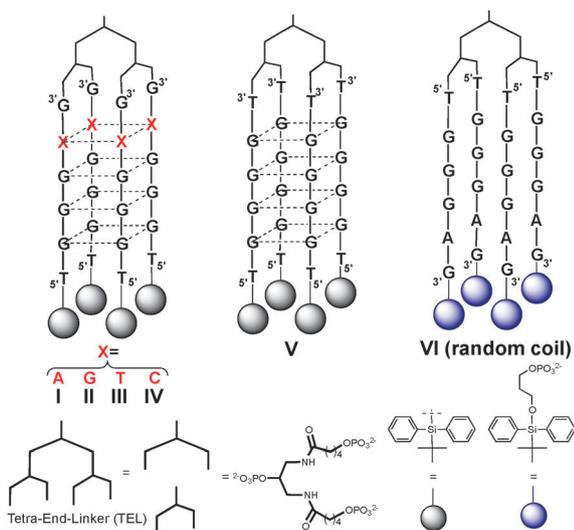


Fig. 1 Schematic representation of investigated TEL-ODNs.

the TGGGXG sequence (**II–IV**, Fig. 1). TEL-ODNs **V** and **VI** (a TEL-(TGGGAG)₄ analogue unable to form a TEL-quadruplex) were also synthesized and studied in order to investigate the role played by the G-tetrad at the 3'-terminus of TEL-quadruplexes **I–IV** and to further corroborate the requirement of quadruplex formation for the inhibition of HIV cytopathicity. In this communication we report the results of biophysical, biological and SAR studies performed on **I–VI**.

All products were synthesized and characterized following the previously described approach.^{14,16} In order to verify that the presence of one (**II–IV**) or two (**V**) different bases did not affect the parallel quadruplex arrangement observed for the parent TEL-ODN **I**, CD measurements were performed (Fig. S2 in ESI†). The overall CD profiles, recorded in 100 mM K⁺ buffer, were in agreement with the formation of G-quadruplex structures, showing two positive maxima at 220 and 260 nm and a negative minimum close to 240 nm, which are characteristic of head-to-tail arrangement of guanines, as typically found in parallel G-quadruplexes.^{17–19} Unexpectedly, CD spectra of **VI** were similar to those of a random coil, thus evidencing that the insertion of a propylphosphate moiety between the OH at 3'-ends and TBDPS groups is detrimental for quadruplex stability. CD melting analyses on **II–IV** confirmed the expected great thermal stability of the new TEL-quadruplexes (Fig. S3 in ESI†). In fact, as for **I**, the apparent melting temperature of **II–IV** could not be determined since no derivatizable melting curve was obtained.

The formation of thermally stable TEL-quadruplexes for **II–IV** was also confirmed by ¹H NMR evidence (Fig. S4 and S5 in ESI†). The G-quadruplex diagnostic exchange-protected imino proton signals were observed in all recorded spectra up to 90 °C (100 mM K⁺ buffer; H₂O/D₂O 9 : 1). As expected, the intensity of imino proton signals reflected the stability of the X-tetrad within the G-quadruplex structure. Stronger imino signals were observed for **II** (five stacked G-tetrads), whereas weaker imino signals were observed for **I**, **III** and **IV** due to the presence of the less stable A-, T- or C-tetrad, respectively, that accounts for the faster exchange of imino protons with the solvent.

The antiviral activity of **II–IV** against HIV-1 and HIV-2 was determined as previously reported for **I**¹⁶ and is shown in Table 1. The binding properties of active aptamers **II–IV** to HIV-1(III_B) gp120 were also determined through Surface Plasmon Resonance (SPR) technology (Table 1; Fig. S6 in ESI†).

All new TEL-(TGGGXG)₄ aptamers reported in this communication (**II–IV**) retained potent *anti*-HIV-1 activity (nanomolar EC₅₀ values). They displayed comparable on (*k*_a) and off (*k*_d) rates for HIV-1 gp120 binding resulting in very similar gp120 binding affinities (*K*_D) to **I**, thus confirming that the residues of the gp120 V3 loop interact mainly with the grooves and the sugar–phosphate backbone of the aptamers. However, considering that the EC₅₀'s of **II** and **IV** were significantly lower than those of **I** and **III** (0.039–0.041 *versus* 0.11–0.15 μM), and that **V** proved to be markedly less active notwithstanding its structuration in a stable TEL-quadruplex structure, we hypothesized that some specific interactions between the X₅G₆ bases of **I–IV** and the V3 loop must occur. It is well-described that binding of agents (*i.e.* monoclonal antibodies) to the viral envelope does not necessarily result in efficient virus neutralization. Their effect on viral infectivity depends on the molecular epitope that is recognized on gp120.²⁰ Thus, it seems that **I** and **III** bind nearly as efficiently to gp120 as **II** and **IV** but neutralize ~3-fold less efficiently HIV-1 possibly due to subtle differences in epitope recognition. Alternatively, it cannot be excluded that the different TEL-quadruplexes have slightly different cellular uptake efficiencies or intracellular stability.

To obtain insight into the nature of the atomistic interactions between the TEL-aptamers **I–IV** and the V3 loop, we carried out molecular modelling studies by docking **I–IV** to the V3 loop of gp120 (PDB ID 2B4C). Our results revealed that (i) the V3 loop interacts with all aptamers with a similar binding mode involving the 3'-end tetrad and the TEL (Fig. S7 in ESI†), (ii) the type of nucleobases at the aptamer-V3 loop interface determines the chemical groups available for the interaction, and (iii) the number and type of interactions between the aptamer and the protein are responsible for the subtle differences in the binding energies (Table S1 in ESI†).

As reported by Honig and co-workers,²¹ charged residues are important in protein–DNA interactions. In this case, an important role is played by the side chain of residue R181 in the V3 loop that binds into the groove created by the phosphodiester

Table 1 Anti-HIV and SPR studies on **I–V**

ODN	EC ₅₀ ^a (μM)		SPR vs. HIV-1(III _B) gp120 ^b		
	HIV-1	HIV-2	<i>K</i> _D (nM)	<i>k</i> _a (1/Ms)	<i>k</i> _d (1/s)
I	0.11 ± 0.05	1.4 ± 0.9	256 ± 63	(1.72 ± 0.28) × 10 ⁴	(4.31 ± 0.37) × 10 ⁻³
II	0.041 ± 0.007	≥ 2	183 ± 40	(2.40 ± 0.20) × 10 ⁴	(4.35 ± 0.61) × 10 ⁻³
III	0.15 ± 0.0015	1.4 ± 0.87	248 ± 28	(2.55 ± 0.35) × 10 ⁴	(6.29 ± 0.14) × 10 ⁻³
IV	0.039 ± 0.005	0.73 ± 0.16	196 ± 69	(2.15 ± 0.40) × 10 ⁴	(4.07 ± 0.71) × 10 ⁻³
V	≥ 2	2	—	—	—

^a EC₅₀ represents the 50% effective concentration required to inhibit virus-induced cytopathicity in CEM cell cultures by 50%. ^b HIV-1(III_B) gp120 was obtained from recombinantly-expressed gp120 in CHO cell cultures.

Table 2 Interaction energies of top docks (Kcal mol⁻¹) and number of H bonds between I–V and the V3 loop of gp120

Aptamer	Interactions energy	Number of H bonds
I	–24.11	8
II	–28.45	15
III	–13.74	9
IV	–28.09	11
V	–13.74	8

backbone atoms (see Table S1 in ESI†). In particular, in **II** and **IV**, R181 makes multiple interactions with both phosphates and purine bases (G5 and G6 in **II**, G6 in **IV**). Furthermore, the side chains of R190, T195 and E197 and the nitrogen backbone atom of T194 of the V3 loop interact with the oxygen atoms of the TEL, giving additional stability to the complexes. In the **II**–V3 loop complex, the side chain of Y193 and the nitrogen backbone atom of Y193, T195 and I198 established additional interactions with the TEL, thereby resulting in lower interaction energy of this complex with respect to the other ones (Table 2). It should also be noted that when the side chain of Y193 is involved in the interaction (**I**, **II** and **IV**), the resulting complexes are found to have a better biological activity. Thus, the G5 nucleobase in **II** presents additional points for hydrogen bonding with the V3 loop. This is not seen in other bases. Therefore we infer that the differences in activity can arise from the thermal stability of the structures. Furthermore, this is also rationalised by the SPR experiments.

In order to better understand the structural features critical for the biological activity, we also carried out a molecular modelling study between the V3 loop and **V**, a quadruplex structure lacking marked *anti*-HIV activity. Our results revealed that the V3 loop interacts with **V** by using a different binding mode (Fig. S7 in ESI†). Differently from what was observed in the above-described complexes, **V** interacts with the V3 loop primarily *via* the TEL atoms and no atom of nucleobases is involved. Furthermore, except for R181, different residues of the V3 loop, such as K182, S183, I184, and I186, are involved in the interactions (Table S1 in ESI†). A plausible reason for the different binding mode of **V** is due to the presence of the T-tetrad at the 3'-terminus. As shown in Fig. S7 in ESI†, when there is a G-tetrad at the 3'-position, the oxygen atoms of guanines are involved in the interaction with the V3 loop through the side chains of R181 (and Y193 in the case of **IV**). In the **V**–V3 loop complex the methyl groups of the thymines are positioned in the groove formed by the phosphodiester backbone atoms, not allowing the formation of H bonds with R181 and Y193 of the V3 loop. All together, the structural evidence suggests that the T-tetrad at the 3'-position markedly affects the biological activity. In accordance with this finding and with the experimental data, the aptamer **II** showed the best docking score and the highest number of hydrogen bonds with the protein (Table 2).

The *anti*-HIV activity against HIV-1 and HIV-2 and the binding properties with HIV gp120 of a small library of new TEL-aptamers (**II**–**IV**) have been reported. Results from TEL-ODNs **V** and **VI** confirmed that the formation of a quadruplex species by the aptamer is required, but not sufficient to exert the *anti*-HIV activity. The docking data suggest that the

interaction of the V3 loop with both the backbone and the TEL of the aptamers is required. Furthermore, the direct involvement of nucleobases in the interaction with the V3 loop gives additional stability to the complexes and results in a better biological activity. Overall, the here reported results expand our knowledge about *anti*-HIV G-quadruplexes and provide the rational basis for the design of novel *anti*-HIV aptamers with improved biological activity.

The European Commission through the COST Action MP0802, the Italian MURST (PRIN 2009), KU Leuven (PF 10/18 and GOA 10/14) and the FWO are gratefully acknowledged for the financial support. Dr Luisa Cuorvo is also acknowledged for technical assistance.

Notes and references

- P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski and W. A. Hendrickson, *Nature*, 1998, **393**, 648–659.
- R. Wyatt, *Science*, 1998, **280**, 1884–1888.
- C. C. Huang, *Science*, 2005, **310**, 1025–1028.
- S. Chou and K. Chin, *Trends Biochem. Sci.*, 2005, **30**, 231–234.
- A. C. Perkins and S. Missailidis, *Q. J. Nucl. Med. Mol. Imaging*, 2007, **51**, 292–296.
- A. D. Keefe, S. Pai and A. Ellington, *Nat. Rev. Drug Discovery*, 2010, **9**, 537–550.
- J. Lee, G. Stovall and A. Ellington, *Curr. Opin. Chem. Biol.*, 2006, **10**, 282–289.
- H. Hotoda, K. Momota, H. Furukawa, T. Nakamura, M. Kaneko, S. Kimura and K. Shimada, *Nucleosides Nucleotides*, 1994, **13**, 1375–1395.
- J. R. Wyatt, T. A. Vickers, J. L. Roberson, R. W. Buckheit, T. Klimkait, E. DeBaets, P. W. Davis, B. Rayner, J. L. Imbach and D. J. Ecker, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 1356–1360.
- H. Furukawa, K. Momota, T. Agatsuma, I. Yamamoto, S. Kimura and K. Shimada, *Antisense Nucleic Acid Drug Dev.*, 1997, **7**, 167–175.
- M. Koizumi, R. Koga, H. Hotoda, K. Momota, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura and K. Shimada, *Bioorg. Med. Chem.*, 1997, **5**, 2235–2243.
- H. Hotoda, M. Koizumi, R. Koga, M. Kaneko, K. Momota, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, J. Sone, S. Tsutsumi, T. Kosaka, K. Abe, S. Kimura and K. Shimada, *J. Med. Chem.*, 1998, **41**, 3655–3663.
- M. Koizumi, R. Koga, H. Hotoda, T. Ohmine, H. Furukawa, T. Agatsuma, T. Nishigaki, K. Abe, T. Kosaka, S. Tsutsumi, J. Sone, M. Kaneko, S. Kimura and K. Shimada, *Bioorg. Med. Chem.*, 1998, **6**, 2469–2475.
- G. Oliviero, N. Borbone, A. Galeone, M. Varra, G. Piccialli and L. Mayol, *Tetrahedron Lett.*, 2004, **45**, 4869–4872.
- G. Oliviero, J. Amato, N. Borbone, A. Galeone, M. Varra, G. Piccialli and L. Mayol, *Biopolymers*, 2006, **81**, 194–201.
- G. Oliviero, J. Amato, N. Borbone, S. D'Errico, A. Galeone, L. Mayol, S. Haider, O. Olubiyi, B. Hoorelbeke, J. Balzarini and G. Piccialli, *Chem. Commun.*, 2010, **46**, 8971–8973.
- S. Paramasivan, I. Rujan and P. H. Bolton, *Methods*, 2007, **43**, 324–331.
- S. Masiero, R. Trotta, S. Pieraccini, S. De Titio, R. Perone, A. Randazzo and G. P. Spada, *Org. Biomol. Chem.*, 2010, **8**, 2653–2872.
- J.-L. Mergny, A. De Cian, A. Ghelab, B. Saccà and L. Lacroix, *Nucleic Acids Res.*, 2005, **33**, 81–94.
- L. M. Walker, M. Huber, K. J. Doores, E. Falkowska, R. Pejchal, J. P. Julien, S. K. Wang, A. Ramos, P. Y. Chan-Hui, M. Moyle, J. L. Mitcham, P. W. Hammond, O. A. Olsen, P. Phung, S. Fling, C. H. Wong, S. Phogat, T. Wrin, M. D. Simek, W. C. Koff, I. A. Wilson, D. R. Burton and P. Pognard, *Nature*, 2011, **477**, 466–470.
- R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann and B. Honig, *Nature*, 2009, **461**, 1248–1253.

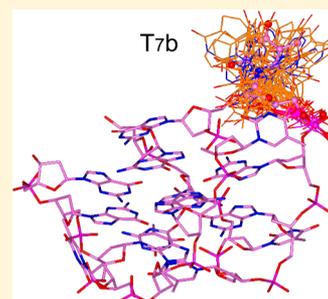
Investigating the Role of T₇ and T₁₂ Residues on the Biological Properties of Thrombin-Binding Aptamer: Enhancement of Anticoagulant Activity by a Single Nucleobase Modification

Nicola Borbone,[†] Mariarosaria Bucci,[‡] Giorgia Oliviero,[†] Elena Morelli,[#] Jussara Amato,[†] Valentina D'Atri,[†] Stefano D'Errico,[†] Valentina Vellecco,[‡] Giuseppe Cirino,[‡] Gennaro Piccialli,[†] Caterina Fattorusso,^{*,†} Michela Varra,^{*,†} Luciano Mayol,[†] Marco Persico,^{†,§} and Maria Scutto^{†,§}

[†]Dipartimento di Chimica delle Sostanze Naturali, [‡]Dipartimento di Farmacologia Sperimentale, [#]Dipartimento di Chimica Farmaceutica e Tossicologica, Università degli Studi di Napoli Federico II, Via D. Montesano 49, 80131 Napoli, Italy

S Supporting Information

ABSTRACT: An acyclic pyrimidine analogue, containing a five-member cycle fused on the pyrimidine ring, was synthesized and introduced at position 7 or 12 of the 15-mer oligodeoxynucleotide GGTTGGTGTGGTTGG, known as thrombin-binding aptamer (TBA). Characterization by ¹H NMR and CD spectroscopies of the resulting aptamers, TBA-T₇**b** and TBA-T₁₂**b**, showed their ability to fold into the typical antiparallel chairlike G-quadruplex structure formed by TBA. The apparent CD melting temperatures indicated that the introduction of the acyclic residue, mainly at position 7, improves the thermal stability of resulting G-quadruplexes with respect to TBA. The anticoagulant activity of the new molecules was then valued in PT assay, and it resulted that TBA-T₇**b** is more potent than TBA in prolonging clotting time. On the other hand, in purified fibrinogen assay the thrombin inhibitory activity of both modified sequences was lower than that of TBA using human enzyme, whereas the potency trend was again reversed using bovine enzyme. Obtained structure–activity relationships were investigated by structural and computational studies. Taken together, these results reveal the active role of TBA residues T₇ and T₁₂ and the relevance of some amino acids located in the anion binding exosite I of the protein in aptamer–thrombin interaction.



■ INTRODUCTION

Thrombin is a sodium-activated allosteric enzyme playing a key role in blood coagulation.^{1a,b} The complexity of thrombin function and regulation has gained this enzyme pre-eminence as the prototypic allosteric serine protease.^{2a–c} It clots blood by converting fibrinogen into fibrin, by activating factors V, VIII, and XIII, the latter being responsible for the cross-linking of fibrin fibers, and by promoting platelet aggregation. Because activation of factor V is required by activated factor X to cut prothrombin into thrombin, the synthesis of thrombin is in part modulated by its own blood concentration, thus providing a fast response to injury. Besides procoagulant stimuli, also anticoagulant stimuli can be triggered by thrombin via activation of protein C, under the allosteric control of the cofactor thrombomodulin. Moreover, thrombin inactivation and clearance in plasma can also be achieved by interaction with the serpin antithrombin and with the heparin cofactor II. Due to its central role in the coagulation cascade, malfunctions in the regulatory mechanism of thrombin activity cause pathological states such as hemorrhage or abnormal clot growth. Thrombosis and connected diseases are among the main causes of mortality in Western countries;³ thus, the discovery of molecules capable of modulating thrombin activity represents a major target for the development of anticoagulant strategies.⁴ Aptamer technology has been efficiently employed to obtain new direct thrombin inhibitors by selecting thrombin-

binding oligonucleotides.⁵ The term aptamer generally refers to a single-stranded oligonucleotide that binds to a selected protein and specifically inhibits one or more of its functions. The first reported consensus sequence able to inhibit thrombin activity was the 15-mer oligonucleotide (ON) GGTTGGTGTGGTTGG, usually known with the acronym TBA (thrombin-binding aptamer).^{5a} In the presence of thrombin and/or monovalent cations TBA folds into a specific three-dimensional structure that dictates its thrombin-binding affinity. The structures of TBA alone and in complex with thrombin were determined by NMR⁶ and X-ray⁷ methods, respectively. In all experimentally determined structures TBA adopted a monomolecular chairlike G-quadruplex folding topology, consisting of two G-tetrads connected by one TGT loop and two TT loops. First,^{7a,b} different TBA–thrombin X-ray complexes were reported, in which TBA adopted either the same folding as derived by NMR (PDB ID 1HAO; Figure 1A) or a different one (PDB IDs 1HUT and 1HAP; Figure 1B), presenting a diverse positioning of the connecting loops.

In these X-ray complexes, the aptamer interacts with distinct regions of two thrombin molecules, (i) the fibrinogen exosite (namely anion-binding exosite I, ABE I) and (ii) the heparin exosite (namely, ABE II) near the carboxylate terminal helix of

Received: September 28, 2012

Published: November 5, 2012

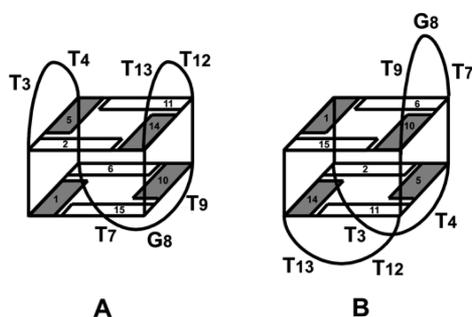


Figure 1. Schematic drawings of the X-ray derived G-quadruplex structures of TBA as reported in PDB IDs 1HAO (A) and 1HUT and 1HAP (B). In (A) TT loops span narrow grooves and TGT spans wide grooves, vice versa in (B). White and gray squares indicate G bases in *anti* and *syn* conformation, respectively.

a neighboring thrombin. Thus, according to the adopted folding and the relative orientation with respect to thrombin, TBA bound to ABE I by using the TT loops (PDB ID 1HAO; Figure 1A) or the TGT loop (PDB IDs 1HUT and 1HAP; Figure 1B). A subsequent re-evaluation⁸ of the diffraction data of TBA–thrombin complexes^{7a,b} evidenced that the NMR-derived folding (Figure 1A) could fit all diffraction data if alternative aptamer orientations with respect to thrombin were considered. Indeed, according to the electron density maps, it was assumed that the oligonucleotide quartet region (including the DNA backbone) was correct as reported in the crystal structures, whereas, due to the D₂ symmetry of the aptamer core, there are four distinct possible orientations of the NMR folding that overlap with the core of the X-ray model (Figure 2). The re-evaluation of the X-ray complex gave similar results among orientations I–IV, but orientation III showed the best agreement with the experimental data.⁸

Accordingly, different TBA–thrombin complexes in which the aptamer can bind ABE I through the TGT loop and ABE II through the TT loops (Figure 2, orientations I and II), or the other way around (Figure 2, orientations III and IV), are possible. Supporting this view, the newly released X-ray structures of thrombin–TBA–Na⁺ and thrombin–TBA–K⁺ complexes (PDB IDs 4DIH and 4DII, respectively)^{7c} showed the aptamer interacting with the enzyme assuming orientation IV (Figure 2). Moreover, the X-ray structure of human thrombin in complex with a modified TBA containing a 5′-5′ inversion of the polarity site (3′GGT5′-5′TGGTGTGGTTGG3′, namely mTBA) has also been reported (PDB ID 3QLP).⁹ The X-ray complex revealed that the interaction occurs between the TT loops and ABE I

(orientation III in Figure 2), whereas the TGT loop, particularly T₇, is not involved in any relevant interaction with the protein. Despite the fact that mTBA binds to thrombin with higher affinity with respect to TBA,¹⁰ it showed poor anticoagulant activity if compared to TBA in PT assay.¹¹ On the other hand, Toggle-25t, an RNA aptamer that contains 2′-fluoropyrimidine nucleotides, and a 29-mer single-stranded DNA, designated DNA60-18[29] or HD22, bind selectively thrombin at ABE II and they also showed limited effect on clotting times.¹² Because of their thrombin allostery, ABE I and ABE II aptamers were used in combination to test their mixed effect on thrombin activity.^{4,13} The obtained results showed that synergistic anticoagulant effects can be achieved by mixing TBA with Toggle-25t or HD22 and by linking TBA to HD22 with an appropriate-sized spacer.

In this scenario, due to the symmetry of TBA and the complexity of the regulatory mechanism of thrombin, which includes a long-range allosteric linkage between ABE I and ABE II,^{14,15} the exact binding mode of aptamers to thrombin and its correlation with the observed biological activity is still a matter of debate. Nevertheless, biological results indicated that TBA exerts its anticoagulant activity mainly competing with fibrinogen at ABE I.¹⁶ The phase I clinical trial for TBA demonstrated its positive pharmacokinetic profile in humans; however, clinical trials were halted after phase I due to suboptimal dosing profiles.¹⁷ To optimize its anticoagulant properties, of particular interest is the enlargement of data gathering about TBA structure–activity relationships (SARs) through the development of new derivatives. In the nucleic acids research area the introduction of suitable modifications on natural oligonucleotide sequences represents a method widely employed to explore the relevance of single nucleotides both on secondary structure folding topology and on binding with target proteins. The explored modifications involve the nucleobases,¹⁸ the sugar–phosphate backbone,¹⁹ and the conjugation to flexible non-nucleotide linkers.²⁰

In our understanding of the bases of TBA–thrombin recognition, we previously developed new single-stranded TBA analogues, in which the acyclic nucleoside **a** (Figure 3) replaced, one at a time, all thymine residues along the aptamer sequence.²¹

The modification at position 7 in the T₇G₈T₉ loop (herein TBA-T_{7a}) gives rise to an ON that folds into a quadruplex structure that is more stable and active than the quadruplex structure formed by TBA. On the other hand, within the four ONs containing the acyclic nucleoside **a** at one position of TT loops, the ON modified at T₁₂ (herein TBA-T_{12a}) showed the

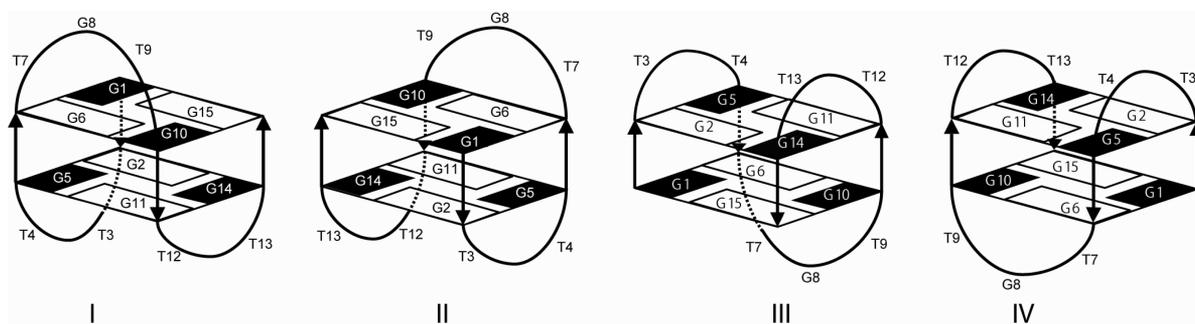


Figure 2. Representation of the four orientations of the NMR derived folding with respect to thrombin corresponding to the D₂ symmetry of the quadruplex core assuming the same strand polarity of the X-ray complex.

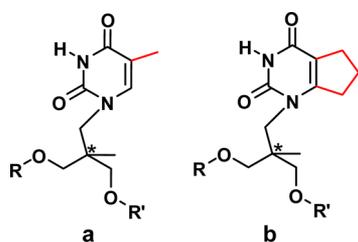


Figure 3. Acyclic nucleosides mimicking T. The presence of the carbon atom marked with an asterisk causes the formation of two diastereomeric ONs when the nucleoside is inserted in the TBA sequence.

most significant biological properties.²¹ In line with data reported from other authors,²² our results evidenced that modification at T₇ and T₁₂ along the TBA sequence can be fruitful to improve the biological activity.

To investigate the role of the modification at T₇ and T₁₂ in affecting the structural stability and the biological activity, we synthesized two new modified TBAs, named TBA-T₇**b** and TBA-T₁₂**b**, which differed from the previously reported analogues only in the thymidine base coupled with the acyclic linker, which was bulked by fusing a hydrophobic five member cycle with the pyrimidine ring (**b**, Figure 3). The abilities of TBA-T₇**b** and TBA-T₁₂**b** to fold into G-quadruplex structures, as well as their biological properties, were evaluated by CD and ¹H NMR experiments and by PT and purified fibrinogen clotting assays, respectively. Finally, a computational study allowed the analysis of the obtained SARs.

RESULTS

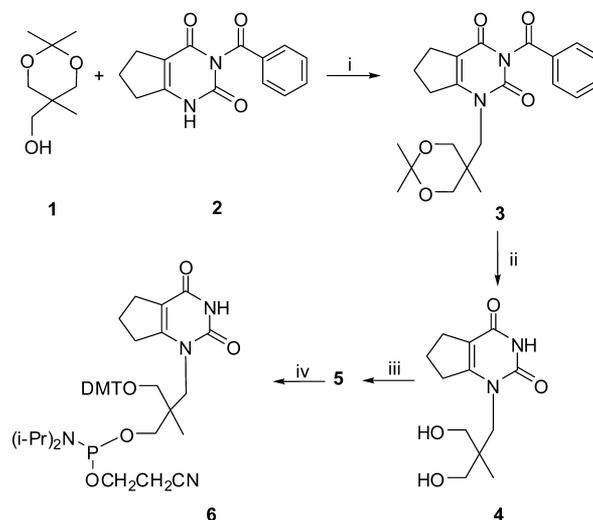
Synthesis of Monomer 6. The synthesis of the modified acyclic nucleoside phosphoramidite building block **6**, used for the automated synthesis of TBA-T₇**b** and TBA-T₁₂**b**, is summarized in Scheme 1.

The synthesis of N1-alkylated intermediate **3** was performed via Mitsunobu reaction²³ between **2**, obtained as described under Experimental Section, and **1**. The yields of this coupling reaction were strongly dependent on the initial temperature of the system and on the order in which the reagents were added. We found that temperatures lower than 30 °C were always detrimental to the reaction yield, and that higher yields (45–50%) were obtained by mild heating (30 °C) of a dioxane solution containing **2** and triphenylphosphine followed by addition of di-*tert*-butylazodicarboxylate and finally a dioxane solution of protected linker **1** (100 μL of solvent per 1 mmol). The intermediate **4** was obtained by removing both the linker and nucleobase protecting groups from **3** using a one-pot procedure (see Experimental Section). Finally, **4** was functionalized to phosphoramidite building block **6** using standard procedure.

The acyclic nucleoside **6** was inserted as a T mimic at position 7 or 12 along the TBA sequence to obtain TBA-T₇**b** and TBA-T₁₂**b**, respectively.

Structural Characterization. TBA-T₇**b** and TBA-T₁₂**b** were characterized by circular dichroism (CD), CD melting, and ¹H NMR for their ability to fold into G-quadruplexes. The introduction of the new acyclic nucleoside **b** at position 7 or 12 does not affect the overall G-quadruplex topology formed in solution. CD profiles of TBA-T₇**b** and TBA-T₁₂**b** are similar to those of TBA in both PBS and K⁺ buffer (left and right panels, respectively, in Figure 4), showing maxima at about 208, 245,

Scheme 1. Synthesis of Monomer 6^a



^a(i) **1** 0.94 g (5.9 mmol), **2** 1.50 g (5.9 mmol), triphenylphosphine 2.28 g (8.7 mmol), di-*tert*-butylazodicarboxylate 2.1 g (9.1 mmol), dry dioxane (70 mL), 24 h, 30 °C, yields 43%; (ii) (a) **3** 1.0 g (2.5 mmol), Dowex H⁺ 450 mg, MeOH/H₂O (9:1 v/v) 100 mL; (b) aqueous NaOH (1.0 M, 5 mL) yields 98%; (iii) **4** 640 mg (2.5 mmol), 4,4'-dimethoxytrytylchloride 541 mg (1.6 mmol), 2,4-dimethylaminopyridine 15 mg (0.12 mmol), pyridine (20 mL), acetonitrile (10 mL), 1.5 h, rt, yields 45%; (iv) **5** 600 mg (1.1 mmol), 2-cyanoethyl-diisopropylchlorophosphoramidite 300 μL (1.2 mmol), DIPEA 600 μL (3.6 mmol), DCM (8 mL), 40 min, rt, yields 99%.

and 295 nm and a minimum at 267 nm. As expected, positive CD bands have higher molar ellipticities in K⁺ buffer than in PBS, according to the well-known dependence of G-quadruplex stabilities from the type of cations present in the buffer solution.²⁴

The ¹H NMR analysis, performed at temperatures in the range of 25–50 °C, confirmed that TBA-T₇**b** and TBA-T₁₂**b** form a stable G-quadruplex structure when annealed in K⁺ buffer. However, ¹H NMR spectra appear more crowded than expected for a single quadruplex species, even at 25 °C (Figure 5), where only 14 aromatic signals and 8 imino proton signals are expected. This finding was previously explained by considering that the replacement of a T residue with the acyclic nucleoside **a** produces two closely related diastereomeric G-quadruplex forming ONs.²¹

Although both modified ONs show *T_m* values higher than that of TBA (Figure 6 and Table 1), the replacement of T₇ with the acyclic nucleoside **b** stabilizes the resulting G-quadruplexes more efficiently than replacement of T₁₂.

PT Assay. The anticoagulant activity in the presence of all thrombin substrates and cofactors was evaluated by PT assay. PT analyses were performed at [ON] of 2 and 20 μM after 2 min of incubation with human plasma (Figure 7).

To better understand the effects induced by the presence of nucleoside **b** on the anticoagulant activity, we compared the prolonging of clotting time caused by TBA-T₇**b** and TBA-T₁₂**b** with that of TBA and those of the previously reported²¹ analogues TBA-T₇**a** and TBA-T₁₂**a** (Table 2).

It is noteworthy that TBA-T₇**b** prolonged the basal PT value to a significantly larger extent with respect to TBA and TBA-T₇**a** (Figure 7 and Table 2). Replacement of nucleoside **a** for **b** was more effective also at position 12 (Table 2). In particular, TBA-T₁₂**b** is less active than TBA at 2 μM, but showed a

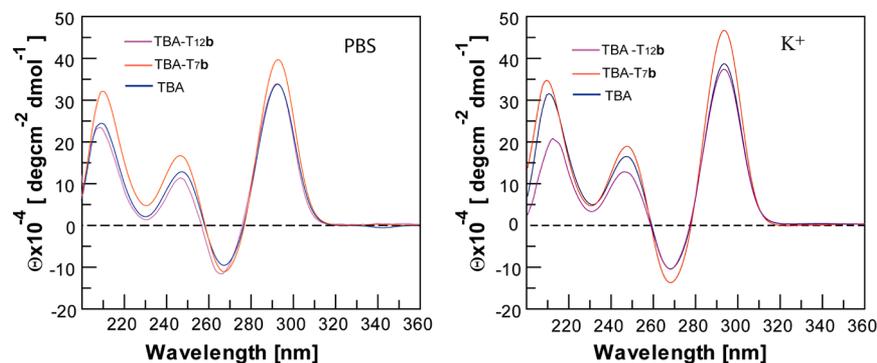


Figure 4. CD profiles in PBS (left) and in potassium phosphate buffer (right); $[ON] = 2.0 \times 10^{-5}$ M.

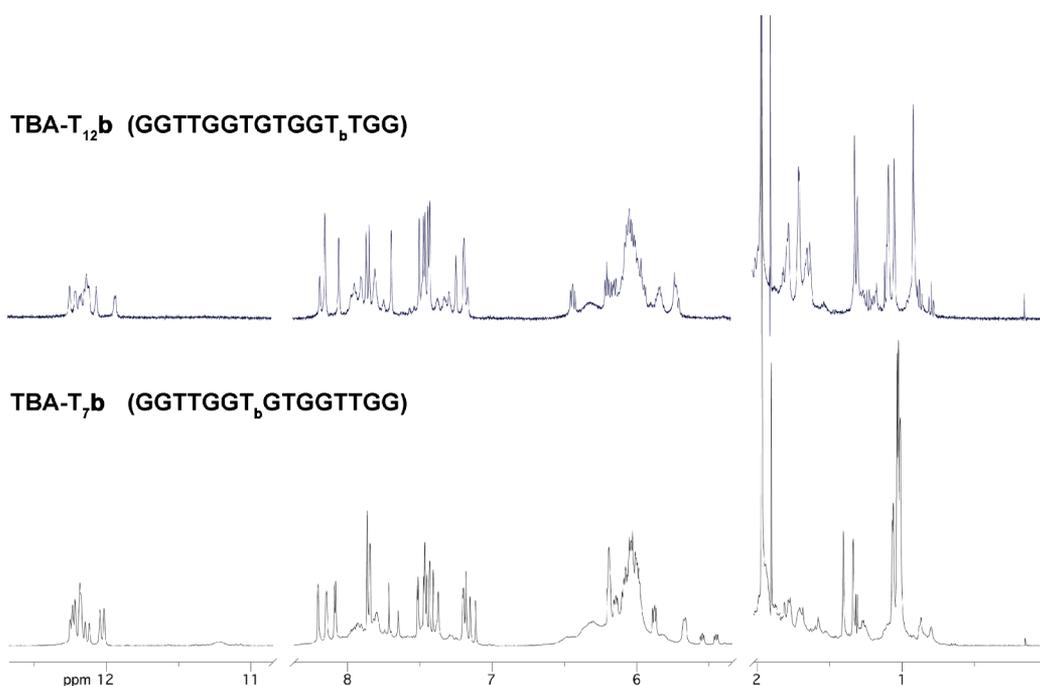


Figure 5. ^1H NMR spectra of TBA-T_{7b} and TBA-T_{12b} recorded at 25 °C in K^+ buffer.

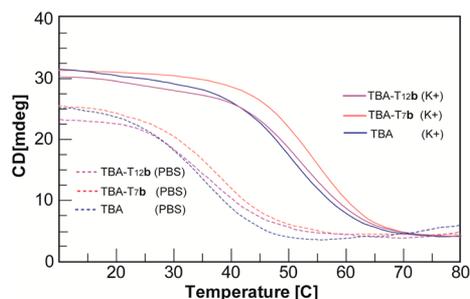


Figure 6. CD melting curves of TBA and its analogues in potassium phosphate (continuous lines) or PBS (dotted lines) buffer. Concentration of each ON was 2.0×10^{-5} M. The curves were obtained by monitoring the variation of absorbance at 295 nm from 10 to 80 °C at 0.5 °C/min.

greater increase of activity at higher concentration, thus attaining that of TBA at 20 μM (Figure 7).

To exclude that the observed activity of modified aptamers could be due to higher nuclease stability with respect to TBA, we repeated the experiments, changing the incubation time of TBA and analogues with plasma from 30 s to 15 min. PT

Table 1. T_m Values of TBA and Its Analogues

ON	T_m (± 0.5 °C)	
	PBS	K^+ buffer
TBA	33.0	50.0
TBA-T _{12a}	34.0 ^a	51.0 ^a
TBA-T _{12b}	34.5	51.5
TBA-T _{7a}	37.5 ^a	54.0 ^a
TBA-T _{7b}	37.5	54.0

^aData taken from ref 21.

results were almost unchanged for all sequences (Supporting Information, Figure 1SI).

Fibrinogen Assay Using Human and Bovine Thrombin. The new ONs were subjected to purified fibrinogen clotting assay to evaluate their ability to inhibit the conversion of soluble fibrinogen into insoluble strands of fibrin in the absence of any other thrombin ligands/ effectors (Figure 8 and Tables 1SI and 2SI of the Supporting Information).

The assay was performed using various concentrations of each aptamer. The purified fibrinogen solution was preincubated with the aptamer and the reactions were initiated by the

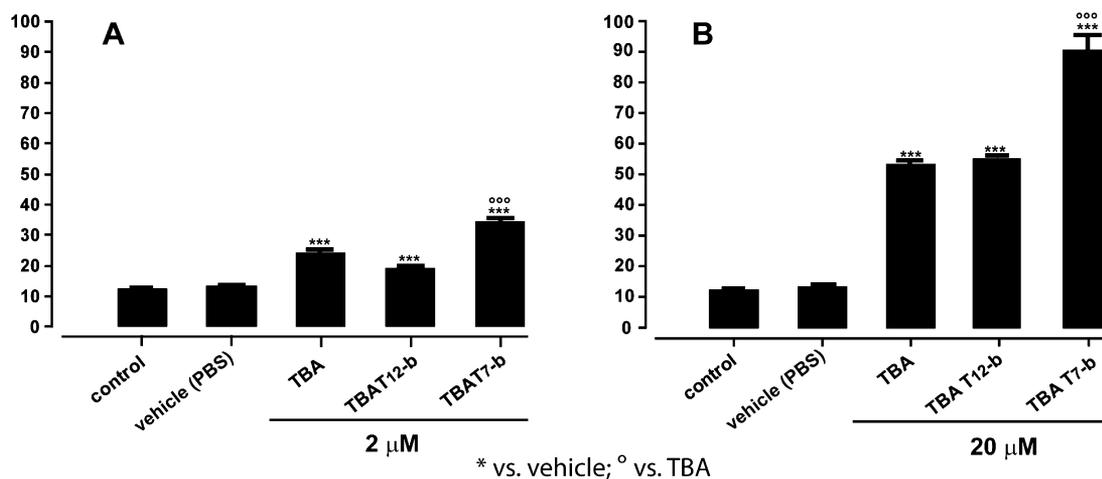


Figure 7. Concentration-dependent response, following 2 min ON incubation with human plasma at 2 (A) or 20 μM (B) [ON]. PT values are expressed in seconds. Each measurement was performed in triplicate and is shown as the mean ± SEM. The basal PT time is 13.4 ± 0.2 s.

Table 2. PT Values Measured after 2 min of Incubation Using 20 μM ON Concentration and Fold Increases of Basal PT Time (13.4 ± 0.2 s)

ON	PT value ^a (20 μM)	fold increase
TBA	53.68 ± 1.48	4.01 ± 0.70
TBA-T ₇ a	60.20 ± 1.38 ^b	4.50 ± 0.60
TBA-T ₇ b	90.60 ± 5.26	6.77 ± 0.24
TBA-T ₁₂ a	44.83 ± 1.29 ^b	3.35 ± 0.60
TBA-T ₁₂ b	55.20 ± 0.47	4.12 ± 0.20

^aPT values are expressed in seconds. ^bData taken from ref 21.

addition of human thrombin and clotting times were recorded. The fibrinogen assay results evidenced that at all ON concentrations tested, the most efficient inhibitor is the unmodified TBA sequence (Figure 8A and Table 1SI of the Supporting Information), thus reversing the results gathered from the PT assay. To further investigate this phenomenon, we performed the fibrinogen assay by using bovine thrombin, which differs from human thrombin in some residues of the fibrinogen-binding site (ABE 1) that are crucial for TBA–thrombin interaction (see next paragraphs for details). By

changing the source of the target enzyme, both modified sequences showed a substantial increase in their ability to inhibit thrombin activity (Figure 8B and Table 2SI of the Supporting Information), whereas the inhibitory activity of TBA was only slightly increased; as a consequence, TBA-T₇b became again the best inhibitor.

Conformational Search on TBA and Modified Analogues TBA-T₇b and TBA-T₁₂b. An extensive molecular modeling study, including molecular mechanics (MM) and dynamics (MD) calculations, was undertaken to analyze the SARs of new modified TBAs.

In particular, to investigate the conformational properties of TBA and the new modified aptamers, a simulated annealing (SA) procedure followed by MM energy minimization was applied. Following the criteria described under Experimental Section, resulting conformers were analyzed and each nucleotide of loops was classified on the basis of (i) χ torsional angle values, to identify the conformation of the glycosidic bond (i.e., *syn*, *anti*, or *s/a*); and (ii) the interatomic distance between its own centroid and that of the two G-tetrads, to investigate the position of loop nucleotides with respect to the

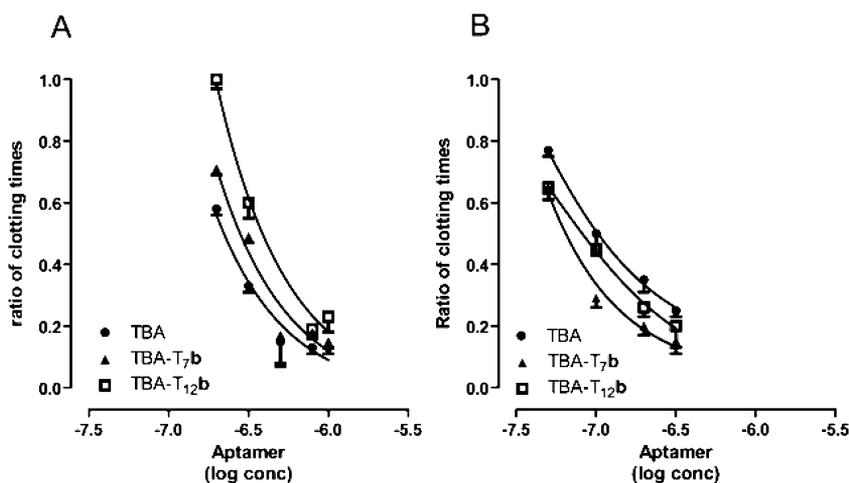


Figure 8. Effects of aptamers on human (A) and bovine (B) thrombin clotting time in fibrinogen solution. The results are expressed as the ratio of clotting times measured in the absence and in the presence of the aptamers. The effects are reported as a function of log [ON]. Each measurement was performed in triplicate and is shown as the mean ± SD.

Table 3. Calculated Occurrence Rates (Percent) of *syn*, *anti*, and *s/a* Conformers of TGT and TT Glycosidic Bonds

	TBA			TBA-T ₇ b			TBA-T ₁₂ b		
	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>	<i>syn</i>	<i>anti</i>	<i>s/a</i>
T ₃	28.5	69.5	2.00	37.0	61.5	1.50	47.0	50.5	2.50
T ₄	38.5	56.0	5.50	29.0	65.5	5.50	43.0	52.0	5.00
T ₇ ^a	39.0	59.5	1.50				36.0	62.5	1.50
G ₈	39.5	38.0	22.5	35.0	46.5	18.5	34.5	44.0	21.5
T ₉	41.5	50.0	8.50	34.0	57.0	9.00	42.0	51.0	7.00
T ₁₂ ^b	32.0	66.0	2.00	42.0	56.0	2.00			
T ₁₃	37.0	59.0	4.00	36.0	60.5	3.50	42.0	54.5	3.50

^aT₇ residue in TBA-T₇**b** is replaced by the acyclic nucleoside **b**. ^bT₁₂ residue in TBA-T₁₂**b** is replaced by the acyclic nucleoside **b**.

guanine planes (i.e., “stacked” or “not-stacked”). Occurrence rates were calculated and compared to those obtained for TBA. Because the replacement of T₇ and T₁₂ residues with nucleoside **b** produced a mixture of diastereoisomers, characterized by *S* or *R* configuration at the acyclic linker, the occurrence rates reported in Tables 3 and 4 are the mean of the values obtained for the two diastereoisomers; single diastereoisomer values are reported in Tables 3SI and 4SI of the Supporting Information.

Table 4. Calculated Occurrence Rates (Percent) of Conformers Presenting TGT and TT Nucleobases “Stacked” on the Guanine Planes

	TBA	TBA-T ₇ b	TBA-T ₁₂ b
T ₃	62.5	33.0	39.0
T ₄	46.0	60.0	65.0
T ₇ ^a	14.0	12.0	10.0
G ₈	78.5	50.0	49.0
T ₉	26.0	39.5	42.5
T ₁₂ ^b	33.5	37.0	29.0
T ₁₃	38.0	40.5	29.0

^aT₇ residue in TBA-T₇**b** is replaced by the acyclic nucleoside **b**. ^bT₁₂ residue in TBA-T₁₂**b** is replaced by the acyclic nucleoside **b**.

The structural analysis evidenced that, although inducing some variations, the introduced modifications at T₇ and T₁₂ did not cause a reversal of the conformational preference of the glycosidic bonds with respect to TBA (Table 3 and Table 3SI of the Supporting Information). Consequently, in agreement with previously reported NMR studies,^{6c} TBA and modified aptamers TBA-T₇**b** and TBA-T₁₂**b** conserved an overall preference for the *anti* conformation of thymines in TGT and TT loops.

With regard to the stacking of the TGT loop nucleobases on the G-tetrads, TBA-T₇**b** and TBA-T₁₂**b** present an increased stacking of T₉ and a decreased stacking of G₈ (Table 4 and Table 4SI of the Supporting Information). Nevertheless, the conformational features of the TGT loop still evidenced the same trend of TBA with the following order of residue stacking on the G-tetrads: G₈ > T₉ > T₇.

The two modified analogues showed a similar conformational behavior at the T₃T₄ loop because, either in TBA-T₇**b** or in TBA-T₁₂**b**, the rate of stacked conformation of T₃ and T₄ decreased and increased, respectively, with respect to TBA, evidencing a conformational linkage between TGT and T₃T₄ loops. On the contrary, different results were obtained for the T₁₂T₁₃ loop. In fact, as expected, the substitution of T₁₂ with the acyclic nucleoside **b** (TBA-T₁₂**b**) affected the conformational flexibility of the T₁₂T₁₃ loop, inducing a decreased stacking rate of both nucleobases, which was not observed in

TBA-T₇**b** (Table 4 and Table 4SI of the Supporting Information). NMR studies performed on TBA^{6a} revealed that the positioning of T₄ and T₁₃ plays a key role in quadruplex folding and stability; accordingly, due to the increased stacking of T₄ and T₁₃ on the guanine planes, TBA-T₇**b** was structurally more stable than TBA in CD melting experiments (Table 1 and Figure 6).

Molecular Modeling Studies on Aptamer Interaction with Human and Bovine Thrombin. With the aim of identifying the effects of the new structural modifications on aptamer–thrombin interactions, first, all possible aptamer–thrombin binding modes and the corresponding binding interactions were analyzed (PDB IDs 1HAO, 1HAP, 1HUT, 4DIH, and 4DII; see Experimental Section). This study revealed that the various binding orientations of TBA with respect to human thrombin (Figures 1 and 2) produced different binding modes in the crystal complexes sharing similar interactions with ABE I but involving different aptamer residues (Figure 9 and Table 5SI of the Supporting Information).

Indeed, a thymine residue of TBA always interacts with ABE I hydrophobic cleft lined by Ile24, His71, Ile79 and Tyr117, but, depending on the binding mode, the interactions involve T₃ (Figure 9C,D), or T₇ (Figure 9A,B), or T₁₂ (Figure 9E,F). T₇ interacts penetrating into the hydrophobic cleft more than T₁₂ and T₃ (Figure 9F vs Figure 9B,D); while the positioning of these latter resulted identical due to the symmetry of the two TT loops (Figure 9B vs Figure 9D). By consequence, a second ABE I subsite, including Gln38, Tyr76 and Ile82, interacts with T₁₂ or T₃, depending on the binding orientation (Figure 9A–D). When the TGT loop binds to ABE I, the interaction with this second subsite is absent (Figure 9E,F).

Second, to identify the structural differences in TBA-binding site, experimentally determined structures of human and bovine thrombin were compared (PDB IDs are listed in the Experimental Section). Only three TBA-binding site residues, Ile24, Asn78, and Ile79, were mutated in bovine thrombin (Supporting Information, Table 6SI), with no significant variation of the backbone structure (Figure 9). Human thrombin Asn78 residue (replaced by a lysine in bovine thrombin) interacts only with the TGT loop, establishing an H-bond interaction with the phosphate group of T₉ (PDB IDs 1HUT and 1HAP; Figure 9E,F and Table 5SI of the Supporting Information). On the contrary, the interactions with human ABE I hydrophobic cleft containing Ile24 and Ile79 are conserved in all crystal complexes and involve T₃ or T₇ or T₁₂ depending on TBA-binding orientation (Figure 9 and Table 5SI of the Supporting Information). The replacement of these two isoleucine residues in human thrombin by two valine residues in the bovine homologue determines an enlargement of the employable volume within the ABE I cleft.

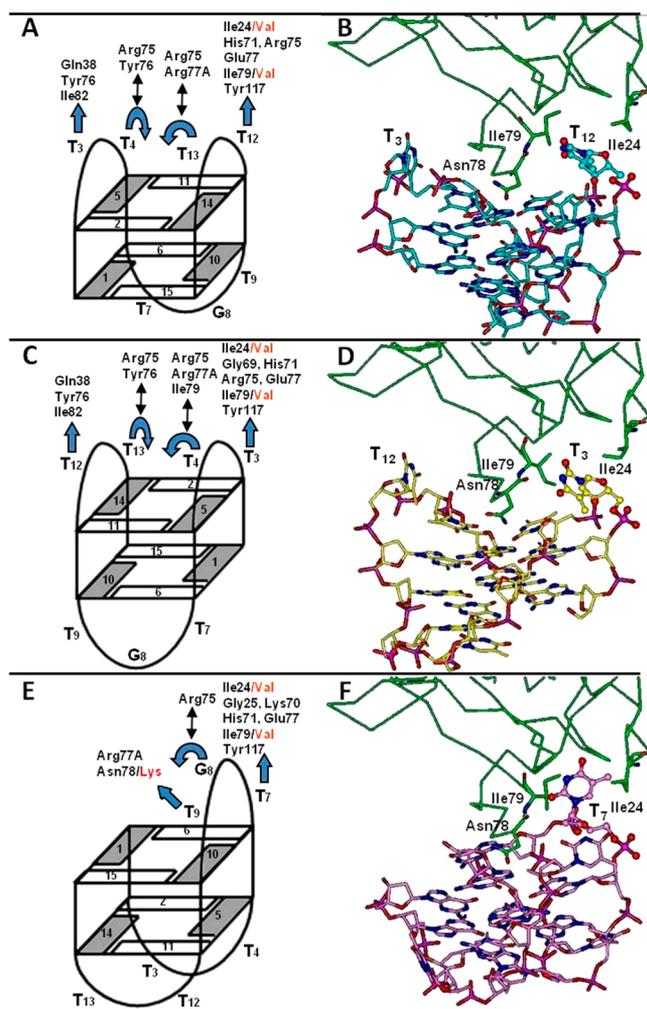


Figure 9. Cartoon (left) and 3D (right) structure of the different binding orientations of TBA with respect to human thrombin ABE I (green). (A, C, and E) Amino acid residues involved in TBA–thrombin interactions are evidenced in black; where mutated, the corresponding bovine residues are reported in red. The blue curve and straight arrows indicate loop nucleotides “stacked” or “not-stacked” on G-tetrad, respectively. (B) PDB ID 1HAO (TBA carbons = cyan); (D) PDB ID 4DIH (TBA carbons = yellow); (F) PDB ID 1HAP (TBA carbons = magenta). Protein carbons are colored in green, residues within TBA binding site that are mutated in bovine thrombin are shown and labeled. The nucleotides T₃, T₇, and T₁₂ are evidenced as ball and stick. Heteroatoms are colored as follows: O = red; N = blue; P = magenta. Hydrogens are omitted for the sake of clarity.

We then calculated the occurrence rates of TBA, TBA-T₇b, and TBA-T₁₂b SA/MM conformers presenting solvent-accessible surface areas of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA binding the ABE I hydrophobic cleft in experimentally determined complexes (Table 5; single diastereoisomer values are reported in Table 7SI of the Supporting Information).

Interestingly, the replacement of T₇ and T₁₂ with the acyclic nucleoside **b** determined in both cases a greater rate of conformations presenting a solvent-exposed conformation. Moreover, in agreement with the results reported in Tables 3 and 4, T₃ is more exposed to the solvent in the new modified aptamers than in TBA.

Finally, we calculated the occurrence rates of bioactive conformations of residues at positions 3, 7, and 12 resulting

Table 5. Occurrence Rates (Percent) of TBA, TBA-T₇b, and TBA-T₁₂b SA/MM Conformers Presenting a Solvent-Accessible Surface Area (Å²) of Residues 3, 7, and 12 Equal to or Greater than That of the Corresponding Residue of TBA Assuming the Bioactive Conformation at the ABE I Hydrophobic Cleft

ON	T ₃	T ₇ ^a	T ₁₂ ^b
TBA	16.5	85.0	36.5
TBA-T ₇ b	21.0	88.0	34.5
TBA-T ₁₂ b	22.0	76.0	50.0

^aT₇ residue in TBA-T₇b is replaced by the acyclic nucleoside **b**. ^bT₁₂ residue in TBA-T₁₂b is replaced by the acyclic nucleoside **b**.

from SA/MM calculations on TBA, TBA-T₇b, and TBA-T₁₂b (Figure 10 and Table 6; single diastereoisomer values are reported in Table 8SI of the Supporting Information).

With this aim, all obtained conformers were superimposed on the experimentally determined TBA–thrombin complexes by fitting the guanine tetrads to evaluate the overlap with the residue interacting with the ABE I hydrophobic cleft (see Experimental Section for details). Obtained results show that TBA-T₇b presents an increased rate of bioactive conformations if compared to TBA; on the contrary, the substitution of T₁₂ with **b** decreases the rate of bioactive conformations of T₃ and T₇.

Interestingly, the replacement of T₇ and T₁₂ with the acyclic nucleoside **b** determined in both cases a greater rate of conformations presenting a solvent-exposed position able to drive the interaction with the ABE I hydrophobic cleft. Nevertheless, likely due to the higher flexibility of the TGT loop with respect to the TT loops, there is a higher occurrence rate of T₇b assuming the ABE I bioactive conformation compared to T₃ and T₁₂b (Table 6 and Figure 10).

DISCUSSION

The introduction of a bulky five member cycle on the pyrimidine ring of the modified thymine residues of TBA-T₇a and TBA-T₁₂a led to compounds (TBA-T₇b and TBA-T₁₂b) able to fold into G-quadruplexes characterized by unaffected structural stability but showing different anticoagulant activity with respect to their parent compounds.²¹ Because TBA-T₇b and TBA-T₇a show the same melting profile, with a *T_m* higher than that of TBA (Table 1), it can be concluded that the sole presence of the acyclic linker unit at position 7 is responsible for the structural stabilizing effect. Conformational analysis results indicate that the introduction of the acyclic linker at position 7 increased the solvent-exposed surface area (Table 6) of the modified residue and the rate of T₄ and T₁₃ conformations stacked on the guanine planes (Table 4). These data account for the nucleobase 7 always positioned outside the G-quadruplex core and not involved in intramolecular interactions, as well as for the reported stabilizing role played by T₄ and T₁₃ on TBA structural stability.⁶ In line with our results, a similar increase in *T_m* value has also been observed by replacing the sugar moiety of T₇ with a different acyclic linker.²² Interestingly, the fact that TBA-T₇a and TBA-T₇b showed the same enhancement of the *T_m* value with respect to TBA is not correlated with their diverse anticoagulant activities. Indeed, although the anticoagulant activities of TBA-T₇a and TBA-T₇b are higher than that of TBA, TBA-T₇b is more active than TBA-T₇a (Figure 7). Furthermore, PT assay data show that the incorporation of

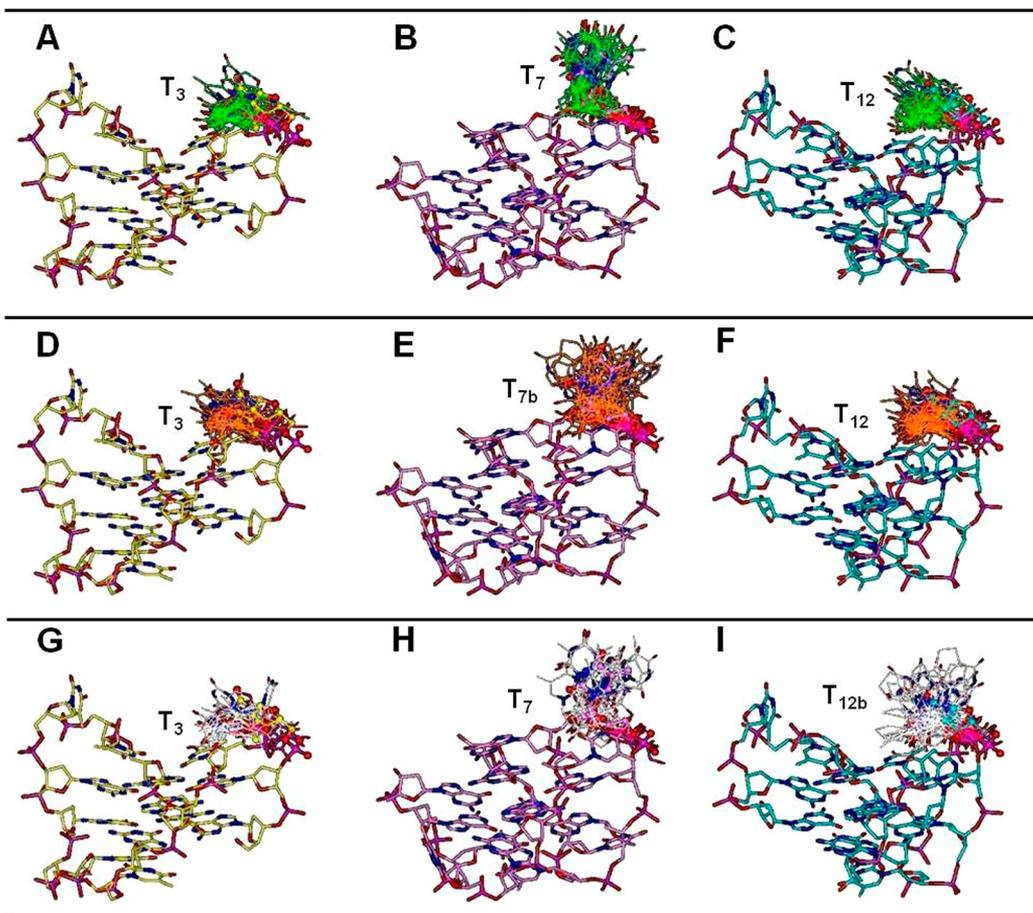


Figure 10. TBA (green; A–C), TBA-T₇**b** (orange; D–F), and TBA-T₁₂**b** (white; G–I) SA/MM conformers presenting residue 3, 7, or 12 assuming ABE I hydrophobic cleft binding conformation. Molecular models are superimposed on 4DIH (A, D, G; yellow), 1HAP (B, E, H; pink), or 1HAO (C, F, I; cyan) by fitting the guanine tetrads. All conformers of TBA and the mean of those obtained for the two diastereoisomers of TBA-T₇**b** and TBA-T₁₂**b** are shown. For the sake of clarity only the residue interacting with the ABE I hydrophobic cleft is displayed and hydrogens are omitted. Heteroatoms are colored as follows: O = red; N = blue; P = magenta.

Table 6. Occurrence Rates (Percent) of Bioactive Conformations at the ABE I Hydrophobic Cleft of Residues at Positions 3, 7, and 12 Resulting from SA/MM Calculations

ON	T ₃	T ₇ ^a	T ₁₂ ^b
TBA	9.50	12.0	10.0
TBA-T ₇ b	8.00	14.0	13.5
TBA-T ₁₂ b	6.50	6.50	10.0

^aT₇ residue in TBA-T₇**b** is replaced by the acyclic nucleoside **b**. ^bT₁₂ residue in TBA-T₁₂**b** is replaced by the acyclic nucleoside **b**.

nucleoside **b** was more effective than that of nucleoside **a** also at position 12 of TBA (Table 2). It was somehow surprising that the activity of TBA with respect to the new modified analogues is reversed in fibrinogen assay (Figure 8A). It could be supposed that the introduction of an acyclic linker in the TBA phosphate backbone enhances the aptamers' stability to nucleases, thus producing an increment of inhibitory effect in plasma tests. However, PT results were almost unaffected upon increasing of the incubation time of all sequences with plasma (Supporting Information, Figure 1SI), thus suggesting that, in the explored conditions, nuclease stabilities and inhibitory activities are uncorrelated and that modifications at T₇ directly affect the interaction of aptamers with thrombin. Supporting this view, the use of bovine thrombin in fibrinogen assay

changed again the activity trend. In particular, when tested toward bovine thrombin, the inhibitory activity of TBA-T₇**b** and TBA-T₁₂**b** significantly increases, becoming much higher than that of TBA, which, on the contrary, is almost unchanged in both fibrinogen tests (Figure 8, panel A vs panel B). On this ground, subtle structural differences between human and bovine thrombin that specifically affect the binding with TBA-T₇**b** and TBA-T₁₂**b** must exist. The results from our molecular modeling study using both human and bovine thrombin revealed that the obtained SARs fit with the binding of modified nucleobases, T₇**b** and T₁₂**b** to thrombin ABE I. Indeed, within the TBA binding site, human thrombin differs from the bovine counterpart by only three amino acids located at ABE I (Supporting Information, Table 6SI). Specifically, in the bovine enzyme two isoleucine residues are replaced by valine residues and an asparagine is mutated into a lysine. In the reported human thrombin–TBA complexes, the isoleucine residues are involved in hydrophobic contacts with T₃ or T₇ or T₁₂, depending on the binding orientation of the aptamer (Figure 9). On the other hand, the asparagine residue is involved in H-bond interactions with the phosphate group linking G₈ to T₉ only when the TGT loop binds to ABE I (Supporting Information, Tables 5SI and 6SI). In the new TBA derivatives TBA-T₇**b** and TBA-T₁₂**b**, the introduction of the acyclic linker facilitates the extension of the phosphate

backbone toward the solvent and the penetration of the modified nucleobase within the ABE I hydrophobic cleft, thus amplifying the steric hindrance caused by the presence of the five member cycle on modified nucleobase **b** (Table 5 and Figure 10). Accordingly, the presence of the smaller Val24 and Val79 residues in bovine thrombin could explain the increased activity of the new modified TBA analogues for this enzyme. The observed higher activity of TBA-T₇**b** is indeed correlated to the calculated higher rate of bioactive conformations at the ABE I hydrophobic cleft with respect to TBA and TBA-T₁₂**b** (Table 6 and Figure 10). This hypothesis is supported by the previously reported activity of TBA-T₇**a** and TBA-T₁₂**a**,²¹ which, lacking the steric bulk at the modified nucleoside, was more active in the fibrinogen assay using human thrombin. On the other hand, the charged group at the lysine side chain allows the establishment of an ionic interaction with the modified TGT phosphate backbone of TBA-T₇**b**, likely contributing to the enhancement of inhibitory ability.

On these bases, the apparent incongruence of PT and fibrinogen assay results, using human thrombin, can be interpreted by taking into account the effects of different aptamer binding modes on thrombin allostery.² In the PT assay a number of thrombin ligands/ effectors other than fibrinogen that are able to modulate thrombin activity are present,^{1,2,14,25,26} thus driving different binding modes; the new modified aptamers could differently affect this ligand/effector binding network. In this scenario, it is noteworthy that the 65–84 thrombin loop, involved in specific interactions with the modified nucleobase of the new TBA derivatives according to our binding mode hypothesis, has been proved to be responsible for the allosteric long-range communication among ABE I, the catalytic site, and ABE II.^{2e,15}

Altogether, our results reveal the absence of a direct correlation between the structural stability and the anticoagulant property in our series of modified TBAs, suggesting that the role played by T₇ (placed in the TGT loop) and T₁₂ (placed in the TT loop close to the 3' terminus) in the biological activity of TBA and its analogues could involve alternative aptamer binding modes and the complex and not yet fully understood allosteric mechanism of action of thrombin.

CONCLUSIONS

Bringing together the so far acquired knowledge of the thrombin–TBA interaction and thrombin allosteric regulatory mechanism with our SAR data, we determined that the G-quadruplexes formed by the modified TBAs can bind the thrombin ABE I hydrophobic cleft by using either TGT or one of the two TT loops. Our results indicate that the incoherency of their behaviors in the two explored biological tests is due to the molecular process through which they exert the anticoagulant activity, which is surely more intricate than a competitive inhibition with the fibrinogen-binding site. Besides their high anticoagulant activity, these molecules represent valuable tools to further explore the complex regulatory mechanism of thrombin during plasma coagulation that is, as yet, not completely clarified.

EXPERIMENTAL SECTION

General Procedure. Chemicals and anhydrous solvents were purchased from Fluka-Sigma-Aldrich. TLCs were run on Merck silica gel 60 F254 plates. Silica gel chromatography was performed by using Merck silica gel 60 (0.063–0.200 mm). An API 2000 (Applied Biosystem) mass spectrometer was used to perform the analyses of the

intermediates and monomer **6**. NMR experiments were recorded using Varian Mercury Plus 400 MHz and Varian UNITYNOVA 500 and 700 MHz spectrometers and processed using the Varian VNMR software package. Reagents and phosphoramidites for DNA synthesis were purchased from Glenn Research. ON syntheses were performed on a PerSeptive Biosystem Expedite DNA synthesizer. HPLC purifications and analyses were carried out using a JASCO PU-2089 Plus HPLC pump equipped with a JASCO BS-997-01 UV detector. The purity of the final products was determined as >95% by using a C-18 RP analytical column (C-18 Purospher STAR, Merck) eluted by a gradient of CH₃OH in H₂O (from 0 to 100% in 30 min). CD experiments were performed on a JASCO 715 spectropolarimeter equipped with a PTC-348 temperature controller.

Synthesis of Monomer 6. (a) *3-Benzoyl-6,7-dihydro-1H-cyclopenta[d]pyrimidine-2,4(3H,5H)-dione (2)*. Nucleobase 6,7-dihydro-1H-cyclopenta[d]pyrimidine-2,4(3H,5H)-dione was obtained as previously reported by Renault et al.²⁷ and converted in its N3 benzoyl derivative **2**.²¹ ¹H NMR (400 MHz; mixture of CD₃OD and CDCl₃), δ 7.75 (d, *J* = 8.0 Hz, 2H), 7.60 (t, *J* = 8.0 Hz, 1H), 7.43 (t, *J* = 8.0 Hz, 2H), 3.03 (t, *J* = 7.3 Hz, 2H), 2.67 (t, *J* = 7.3 Hz, 2H), 2.10 (q, *J* = 7.3 Hz, 2H); ¹³C NMR (100 MHz; mixture of CD₃OD and CDCl₃), δ 169.3, 166.3, 155.2, 147.3, 134.2, 132.2, 128.9, 128.3, 127.5, 127.0, 113.4, 34.8, 30.5, 18.0. ESI mass (positive mode), calculated 256.1; found 257.1 [M + H]⁺, 279.1 [M + Na]⁺.

(b) *3-Benzoyl-1-[(2,2,5-trimethyl-1,3-dioxan-5-yl)methyl]-6,7-dihydro-1H-cyclopenta[d]pyrimidine-2,4(3H,5H)-dione (3)*. Compound **2** (1.5 g, 5.9 mmol) was suspended in 70 mL of dry dioxane at 30 °C in the presence of triphenylphosphine (2.28 g, 8.7 mmol) before the addition of di-*tert*-butyl azodicarboxylate (2.1 g, 9.1 mmol). To the resultant mixture, after 10 min of stirring at room temperature, was added 5.9 mmol of **1** dissolved in dry dioxane (300 μL). The reaction mixture was stirred at room temperature for 18 h under argon. The solution was concentrated under reduced pressure and the residue purified by column chromatography on silica gel eluted with 95:5 Et₂O/CH₂Cl₂ to give **3** as a white solid (yield 43%, *R*_f 0.70). ¹H NMR (400 MHz; CDCl₃), δ 7.98 (d, *J* = 8.0 Hz, 2H), 7.60 (t, *J* = 8.0 Hz, 1H), 7.43 (t, *J* = 8.0 Hz, 2H), 4.0 (bs, 2H), 3.62 (m, 4H), 3.03 (t, *J* = 7.3 Hz, 2H), 2.66 (t, *J* = 7.3 Hz, 2H), 2.09 (q, *J* = 7.3 Hz, 2H), 1.38 (s, 3H), 1.42 (s, 3H), 0.87 (s, 3H). ESI mass (positive mode), calculated 398.2; found 399.2 [M + H]⁺, 421.2 [M + Na]⁺.

(c) *1-(3-Hydroxy-2-(hydroxymethyl)-2-methylpropyl)-6,7-dihydro-1H-cyclopenta[d]pyrimidine-2,4(3H,5H)-dione (4)*. Compound **3** (1.0 g, 2.5 mmol) was suspended in 9:1 MeOH/H₂O (100 mL) containing 450 mg of Dowex 50WX8 (H⁺). After 8 h at room temperature, a NaOH 0.5 M aqueous solution was slowly added to neutralization. The solution was filtered and, in turn, basified with NaOH 1.0 M (5 mL) to remove the N-3 benzoyl group from the nucleobase. After 12 h at room temperature, the pH of the reaction was neutralized, the solvent was evaporated under vacuum, and the residue was dissolved in MeOH and purified by HPLC (C-18 reverse-phase column, Grace Davison Discovery Sciences, eluted with ACN in H₂O from 0 to 50% in 30 min) to give **4** as a white solid (yield 98%, elution time 15 min). ¹H NMR (700 MHz; CD₃OD), δ 3.80 (bs, 2H), 3.41 (d, *J* = 11.4 Hz, 2H), 3.38 (d, *J* = 11.4 Hz, 2H), 3.03 (t, *J* = 7.3 Hz, 2H), 2.66 (t, *J* = 7.3 Hz, 2H), 2.09 (q, *J* = 7.3 Hz, 2H), 0.87 (s, 3H); ¹³C NMR (175 MHz; CD₃OD), δ 163.8, 161.4, 155.7, 113.6, 66.6, 49.5, 44.0, 34.2, 28.2, 22.6, 17.8; NOESY (700 MHz; CD₃OD), fundamental NOE signal between 3.80 and 3.03 ppm that confirms the linker at the 1N position. ESI mass (positive mode), calculated 254.1; found, 255.3 [M + H]⁺, 277.3 [M + Na]⁺, 293.2 [M + K]⁺.

(d) *3-(Bis(4-methoxyphenyl)(phenyl)methoxy)-2-((2,4-dioxo-3,4,6,7-tetrahydro-1H-cyclopenta[d]pyrimidin-1-yl)methyl)-2-methylpropyl-2-cyanoethyl Diisopropylphosphoramidite (6)*. Compound **4** (640 mg, 2.5 mmol), 4,4'-dimethoxytrytyl chloride (541 mg, 1.6 mmol), and 4-dimethylaminopyridine (15.0 mg, 0.12 mmol) were dissolved in dry pyridine (20 mL) and dry ACN (10 mL). The resulting solution was stirred at room temperature under argon for 1.5 h. Dry methanol (200 μL) was then added to quench the reaction. After 30 min under stirring, the solution was concentrated under reduced pressure and the residue purified by column chromatography

on silica gel (eluted with 50:50:1 EtOAc/hexane/Et₃N) to give monodimethoxytritylated **5** as a clear yellow solid (44% yield from **4**; *R_f* 0.51 in EtOAc/hexane 1:1 v/v). The solid (600 mg, 1.1 mmol) was dried in vacuo overnight before being dissolved in anhydrous DCM (8 mL) and diisopropylethylamine (600 μL, 3.6 mmol) under argon. Three hundred microliters of β-cyanoethyl diisopropylchlorophosphoramidite was then added (1.2 mmol). After 40 min, the reaction was quenched by the addition of dry methanol (100 μL), diluted with ethyl acetate (15 mL), and finally washed with 10% sodium carbonate solution (15 mL) and brine (15 mL). The organic layer was dried on magnesium sulfate and concentrated in vacuo. The residue was purified by silica gel chromatography eluted with DCM, ethyl acetate, and triethylamine (80:10:10). The fractions containing the product were collected and concentrated under vacuum, yielding **6** as a white foam (99% yield; *R_f* 0.65 in CHCl₃/MeOH/TEA 97:3:0.05 v/v/v). ¹H NMR (700 MHz; CDCl₃), δ 7.90 (2H), 7.38 (2H), 7.35 (4H), 7.12 (1H), 6.85 (4H), 3.90 (2H), 3.70 (6H), 3.65 (2H), 3.60 (2H), 3.45 (2H), 3.39–3.02 (4H), 2.60–2.45 (4H), 2.16 (2H), 1.45 (3H), 1.09 (6H), 1.04 (6H); ¹³C NMR (175 MHz; CDCl₃), δ 158.6, 151.6, 147.3, 139.4, 135.1, 130.3, 129.1, 127.8, 127.7, 127.1, 113.1, 64.4, 60.4, 55.2, 47.3, 33.6, 29.7, 27.6, 20.9, 19.3, 17.3; ³¹P NMR (202 MHz, CDCl₃), δ 146.1 and 145.9. ESI mass (positive mode), calculated 756.3; found 757.9 [M + H]⁺, 779.9 [M + Na]⁺.

Synthesis of Oligomers. TBA and analogues were synthesized by using standard solid phase DNA chemistry on controlled pore glass (CPG) support following the β-cyanoethyl phosphoramidite method. The oligomers were detached from the support and deprotected by treatment with an aqueous ammonia solution (33%) at 55 °C overnight. The combined filtrates and washings were concentrated under reduced pressure, dissolved in H₂O, and purified by HPLC using an anionic exchange column eluted with a linear gradient (from 0 to 100% B in 30 min) of phosphate buffer at pH 7.0 (A, 20 mM NaH₂PO₄ aqueous solution containing 20% CH₃CN; B, 1.0 M NaCl, 20 mM NaH₂PO₄ aqueous solution containing 20% CH₃CN). The oligomers were successively desalted by molecular exclusion chromatography on Biogel P-2 fine. The purity was checked on HPLC by using reverse phase column (Supporting Information, Figure 2SI). The concentrations of the samples used in CD and UV experiments were determined by measuring the absorbance at 260 nm at 80 °C and using the open access program available on <http://basic.northwestern.edu/biotools/OligoCalc.html>.²⁸

NMR Experiments. 1D NMR spectra were acquired as 16384 data points with a recycle delay of 1.0 s at temperatures in the range of 2–50 °C. Data sets were zero filled to 32768 points prior to Fourier transformation and apodized with a shifted sine bell squared window function. The pulsed-field gradient DPGSE²⁹ sequence was used for H₂O suppression. NMR samples of TBA-T₇b and TBA-T₁₂b (0.5 mM single-strand concentration) were prepared in 100 mM K⁺ buffer (H₂O/D₂O 9:1 v/v containing 90 mM KCl, 10 mM KH₂PO₄, and 0.2 mM EDTA).

CD Experiments. To perform CD experiments, each ON was dissolved in the potassium (90 mM KCl, 10 mM KH₂PO₄, pH 7.0) or PBS (Sigma-Aldrich; 10 mM phosphate buffer, 2.7 mM KCl, 137 mM NaCl, pH 7.4) phosphate buffer at the final ON concentration of 2.0 × 10⁻⁵ M and submitted to the annealing procedure (heating at 90 °C and slowly cooling at room temperature). Before each experiment, the samples were equilibrated at 10 °C for 30 min. CD spectra were recorded from 200 to 360 at 100 nm/min scanning rate, 16 s response, and 1.0 nm bandwidth. Each CD profile was obtained by taking the average of three scans from which the spectrum of background buffer was subtracted. CD melting curves were obtained by monitoring the variation of absorbance at 295 nm from 10 to 80 °C. Two melting experiments for each ON were recorded at 0.5 °C/min heating rate.

Prothrombin (PT) Time. PT time was measured by using a Koagulab MJ Coagulation System with a specific kit RecombiPlas Tin HemosIL (Instrumentation Laboratories, Lexington, MA, USA). The procedure was performed according to the manufacturer's instructions. In our experimental protocol a time course of each ON or vehicle incubated with 100 μL of plasma at 37 °C has been performed. For the evaluation of PT at 20.0 μM, in the apposite microtube, 2.0 μL of the

corresponding ON solution (1.0 × 10⁻³ M in PBS) or vehicle was added. The PT at final ON concentration of 2.0 μM was determined by using 2.0 μL of a diluted ON solution (the initial ON solution 1.0 × 10⁻³ M in PBS was diluted at a final concentration of 1.0 × 10⁻⁴ M). Using six different incubation times from 30 s to 15 min (i.e., 30 s and 1, 2, 5, 10, and 15 min) 200 μL of the kit solution containing Recombiplastin was added with consequent activation of extrinsic pathway. The PT measurement, for each incubation time, was produced in triplicate, and the average and its standard error values were calculated. The basal clotting time was determined by measuring the clotting time in the absence of any ON. The fold increase of basal PT time was calculated as the ratio between the measured PT time in the presence of each ON and the basal PT value (13.4 ± 0.2 s).

Purified Fibrinogen Clotting Assay. ONs were incubated for 1 min at 37 °C in 200 μL of buffer (20 mM tris acetate, 140 mM NaCl, 2.7 mM KCl, 1.0 mM MgCl₂, 1.0 mM CaCl₂, pH 7.4) containing 2.0 mg/mL of fibrinogen (fibrinogen from human plasma, F 3879, Sigma-Aldrich). One hundred microliters of human (Sigma-Aldrich, T8885, human thrombin suitable for thrombin time test) or bovine (HemosIL, Thrombin Time Kit, Instrumentation Laboratories) thrombin (10 NIH per mL) was then added to the solution containing the fibrinogen and the ON. The time required to clot was measured using a Koagulab MJ Coagulation System. The clotting time of each ON was determined in triplicate at different concentrations. The basal clotting time was determined by measuring the clotting time in the absence of any ONs. Prolonged clotting time was obtained by subtracting the basal clotting value from each ON clotting time. The ratio of basal and prolonged clotting time versus log [ON] was reported.

Molecular Modeling. Molecular modeling calculations were performed on SGI Origin 200 8XR12000 and E4 Server Twin 2 x Dual Xeon 5520, equipped with two nodes. Each node was 2 x Intel Xeon QuadCore E5520, 2.26 GHz, 36 GB RAM. The molecular modeling graphics were carried out on SGI Octane 2 workstations.

(a) **Conformational Analysis.** Experimentally determined structures of TBA alone (PDB ID 148D) and in complex with thrombin (PDB IDs 1HAO, 1HAP, and 1HUT) were downloaded from Protein Data Bank (PDB, <http://www.rcsb.org/pdb/>) and analyzed using the Homology module of Insight 2005 (Accelrys Software Inc., San Diego, CA, USA). Hydrogens were added to all of these structures considering a pH value of 7.4 (Biopolymer Module, Insight 2005).

Because the replacement of T₇ and T₁₂ residues with nucleoside **b** produced a mixture of diastereoisomers characterized by *S* or *R* configuration at the acyclic linker, for each new TBA analogue TBA-T₇b and TBA-T₁₂b, the two diastereoisomers were built by modifying the experimentally determined structure of TBA in complex with thrombin (PDB ID 1HAO; Insight2005 Builder module). Atomic potentials and charges were assigned using the CVFF force field.³⁰

The conformational space of TBA (PDB ID 1HAO) and the new modified analogues was sampled through 200 cycles of simulated annealing (SA) followed by molecular mechanics (MM) energy minimization. During the SA procedure, the temperature is altered in time increments from an initial temperature to a final temperature by adjusting the kinetic energy of the structure (by rescaling the velocities of the atoms). The following protocol was applied: the system was heated to 1000 K over 2000 fs (time step of 1.0 fs); a temperature of 1000 K was applied to the system for 2000 fs (time step of 1.0 fs) to surmount torsional barriers; successively, the temperature was linearly reduced to 300 K in 1000 fs (time step of 1.0 fs). Resulting conformations were then subjected to MM energy minimization within Insight 2005 Discover_3 module (CVFF force field) until the maximum rms derivative was <0.001 kcal/Å, using a conjugate gradient³¹ as the minimization algorithm.

To reproduce the physiological environment where these molecules act and, to evaluate the effects of the implicit solvent, we sampled the conformational space through the combined procedure of SA/MM calculations, using the dielectric constant of the water (ε = 80r). Moreover, to allow a complete relaxation of the structures preserving the monomolecular chairlike G-quadruplex folding topology, during

the entire course of SA/MM calculations, we applied a tether force of 100 kcal/Å² to the guanine bases of two quartets.

All resulting conformers were subsequently analyzed and loop nucleotides were classified on the bases of (i) glycosidic bond χ values, that is, $0^\circ < \chi < 90^\circ = \text{syn}$; $-60^\circ < \chi < -180^\circ = \text{anti}$; $90^\circ < \chi < 180^\circ$ and $-60^\circ < \chi < 0^\circ = \text{s/a}$; (ii) the interatomic distance between the centroid of the ring atoms of the nucleobase of each loop nucleotide and the centroid of the ring atoms of the nucleobases of the two G-tetrads (Pseudo_Atom Define command, Biopolymer Module, Insight 2005). According to the latter parameter, the loop nucleotide was classified as “stacked” when the distance was $< 8 \text{ \AA}$ or as “not-stacked” when the distance was $> 12 \text{ \AA}$, whereas a 3D visual inspection was needed to classify the nucleotide as “stacked” or “not-stacked” when the distance was between 8 and 12 Å. A nucleotide termed “stacked” presented at least one nucleobase atom shielded by the G-tetrads; a nucleotide termed “not-stacked” presented no atoms shielded by the G-tetrads. Occurrence rates were calculated for TBA and for each diastereoisomer of the new analogues TBA-T₇b and TBA-T₁₂b. Because all experimental data refer to the mixture of the two diastereoisomers, the mean of the values obtained for the two diastereoisomers was also calculated.

(b) *Structural and Bioinformatics Analysis.* To analyze the binding modes of TBAs and the corresponding aptamer–thrombin interactions, all of the experimentally determined structures of aptamers in complex with human thrombin were downloaded from Protein Data Bank: 1HAO, 1HAP, 1HUT, 3DD2, 3QLP, 4DIH, and 4DII. On the other hand, to identify the structural differences in TBA binding site between human and bovine thrombin, additional structures of human (PDB IDs 1HXF, 1TB6, 1TMT, 1TMU, 1XMN, 3HTC, and 4HTC) and bovine (PDB IDs 1HRT, 1VIT, and 3PMA) thrombin, sharing similar ABE I and ABE II ligands, were selected and analyzed. Hydrogens were added to all of the PDB structures considering a pH value of 7.4 (Biopolymer Module, Insight 2005). All structures were superimposed by C α atoms, and their sequences were extracted using the Homology module of Insight 2005 (Accelrys). On the other hand, the human (entry P00734) and bovine (entry P00735) prothrombin sequences were downloaded from the UniProt Knowledgebase (<http://www.uniprot.org>), and the sequence alignments were performed using Multiple_Sequence Alignment pulldown in the Insight 2005 Homology module. Moreover, for each ligand/enzyme complex, a subset around the ligands that consisted of all residues and water molecules having at least one atom within a 6 Å radius from any given ligand atom was defined. The created subsets were displayed and analyzed through a 3D visual inspection. The results of this analysis were compared with those obtained through the sequence alignments.

Starting from the results obtained from this structural and bioinformatics analysis, to evaluate in detail the possibility of the residues at positions 3, 7, and 12 to interact with thrombin ABE I hydrophobic cleft, the solvent-accessible surface area of these nucleotides was evaluated by calculating the Connolly surface with a probe radius of 1.4 Å, which approximates the radius of a water molecule (Viewer Module, Insight 2005, Accelrys Software Inc.). In particular, the Connolly surface of the considered nucleotides was calculated for the TBA experimentally determined structures (PDB IDs 1HAO, 1HAP, 4DIH, and 4DII) and for all conformers of TBA and new modified analogues, resulting from SA/MM calculations.

The TBA, TBA-T₇b, and TBA-T₁₂b SA/MM conformers presenting solvent-accessible surface areas of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA when interacting with the ABE I hydrophobic cleft in experimentally determined complexes (i.e., T₃, 191.01 Å²; T₇, 159.35 Å²; and T₁₂, 180.59 Å²) were selected, and their occurrence rates were calculated. Because all experimental data refer to the mixture of the two diastereoisomers, the mean of the values obtained for the two diastereoisomers was also calculated.

Finally, with the aim to calculate the occurrence rates of bioactive conformations of residues at position 3, 7, and 12 of TBA and new modified analogues TBA-T₇b and TBA-T₁₂b, all conformers, resulting from SA/MM calculations, were superimposed on the experimentally determined structures of TBA in complex with thrombin (PDB IDs

1HAO, 1HAP, 1HUT, 4DIH, and 4DII) by fitting heavy atoms of the guanine bases of two quartets, and the overlap with the residue interacting with the ABE I hydrophobic cleft (I1e24, His71, I1e79, and Tyr117) was evaluated. To assess the bioactive conformation of the residue at position 7, the orientation that allows residue 7 to be located in the same position of T₇ of 1HAP or 1HUT crystal structure was considered (orientation II in Figure 2).

The conformation of residues at positions 3, 7, and 12 was considered bioactive when the correspondent nucleobase was positioned within the ABE I hydrophobic cleft and presented at least one atom superimposed on T₃ (PDB IDs 4DIH and 4DII) or on T₇ (PDB IDs 1HAP and 1HUT) or on T₁₂ (PDB ID 1HAO), respectively. Occurrence rates of bioactive conformations of residues at positions 3, 7, and 12 were then calculated. Because all experimental data refer to the mixture of the two diastereoisomers, the mean of the values obtained for the two diastereoisomers was also calculated.

■ ASSOCIATED CONTENT

📄 Supporting Information

Concentration-dependent response on PT value using increasing incubation times with human plasma; HPLC chromatograms; prolonged fibrinogen clotting time results in presence of human and bovine thrombin; calculated occurrence rates of *syn*, *anti*, and *s/a* conformers of TGT and TT glycosidic bonds; calculated occurrence rates of conformers presenting TGT and TT nucleobases “stacked” on the guanine planes; binding interactions between TBA and human thrombin ABE I in X-ray complexes; mutated amino acid residues in human and bovine thrombin ABE I; occurrence rates of SA/MM conformers presenting a solvent accessible surface area (Å²) of residues 3, 7, and 12 equal to or greater than that of the corresponding residue of TBA assuming the bioactive conformation at ABE I hydrophobic cleft; occurrence rates of bioactive conformations at ABE I hydrophobic cleft of residues at position 3, 7, and 12, resulting from SA/MM calculations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*(M.V.) Phone: +39 081 678540. Fax: +39 081 678552. E-mail: varra@unina.it. (C.F.) Phone: +39 081 678544. Fax: +39 081 678552. E-mail: caterina.fattorusso@unina.it.

Author Contributions

§These authors equally contributed to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by a PRIN grant 2009 from the Italian Ministero dell'Università e della Ricerca [2007EBYL8L_005] and a FARO grant 2012 from the University of Studies “Federico II”. We thank Dr. Luisa Cuorvo for her valuable technical assistance.

■ ABBREVIATIONS USED

ON, oligonucleotide; TBA, thrombin-binding aptamer; ABE I, anion binding exosite I; SARs, structure–activity relationships; CD, circular dichroism; MM, molecular mechanics; MD, molecular dynamics; SA, simulated annealing; PT, prothrombin time

■ REFERENCES

- (1) (a) Huntington, J. A. Molecular recognition mechanisms of thrombin. *J. Thromb. Haemost.* **2005**, *3*, 1861–1872. (b) Di Cera, E. Thrombin interactions. *Chest* **2003**, *124*, 11S–17S.
- (2) (a) Niu, W.; Chen, Z.; Gandhi, P. S.; Vogt, A. D.; Pozzi, N.; Pelc, L.; Zapata, F.; Di Cera, E. Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry* **2011**, *50*, 6301–6307. (b) Gandhi, P. S.; Chen, Z.; Mathews, F. S.; Di Cera, E. Structural identification of the pathway of long-range communication in an allosteric enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1832–1837. (c) Di Cera, E.; Page, M. J.; Bah, A.; Bush-Pelc, L. A.; Garvey, L. C. Thrombin allostery. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1291–1306. (d) Di Cera, E. Thrombin as procoagulant and anticoagulant. *J. Thromb. Haemost.* **2007**, *5*, 196–202. (e) Lechtenberg, B. C.; Johnson, D. J. D.; Freund, S. M. V.; Huntington, J. A. NMR resonance assignments of thrombin reveal the conformational and dynamic effects of ligation. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 14087–14092.
- (3) (a) Dahlback, B. Blood coagulation and its regulation by anticoagulant pathways: genetic pathogenesis of bleeding and thrombotic diseases. *J. Intern. Med.* **2005**, *257*, 209–223. (b) Desai, U. R. New antithrombin-based anticoagulants. *Med. Res. Rev.* **2004**, *24*, 151–181.
- (4) Nimjee, S. M.; Oney, S.; Volovyk, Z.; Bompiani, K. M.; Long, S. B.; Hoffman, M.; Sullenger, B. A. Synergistic effect of aptamers that inhibit exosites 1 and 2 on thrombin. *RNA* **2009**, *15*, 2105–2111.
- (5) (a) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **1992**, *355*, 564–566. (b) Li, W. X.; Kaplan, A. V.; Grant, G. W.; Toole, J. J.; Leung, L. L. A novel nucleotide-based thrombin inhibitor inhibits clot-bound thrombin and reduces arterial platelet thrombus formation. *Blood* **1994**, *83*, 677–682.
- (6) (a) Schultze, P.; Macaya, R. F.; Feigon, J. Three-dimensional solution structure of the thrombin-binding DNA aptamer d-(GGTTGGTGTGGTTGG). *J. Mol. Biol.* **1994**, *235*, 1532–1547. (b) Wang, K. Y.; Krawczyk, S. H.; Bischofberger, N.; Swaminathan, S.; Bolton, P. H. The tertiary structure of a DNA aptamer which binds to and inhibits thrombin determines activity. *Biochemistry* **1993**, *32*, 11285–11292. (c) Macaya, R.; Schultze, P.; Smith, F.; Roe, J.; Feigon, J. Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3745–3749.
- (7) (a) Padmanabhan, K.; Padmanabhan, K. P.; Ferrara, J. D.; Sadler, J. E.; Tulinsky, A. The structure of α -thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **1993**, *268*, 17651–17654. (b) Padmanabhan, K.; Tulinsky, A. An ambiguous structure of a DNA 15-mer thrombin complex. *Acta Crystallogr. D: Biol. Crystallogr.* **1996**, *52*, 272–282. (c) Russo Krauss, I.; Merlino, A.; Randazzo, A.; Novellino, E.; Mazzarella, L.; Sica, F. High-resolution structures of two complexes between thrombin and thrombin-binding aptamer shed light on the role of cations in the aptamer inhibitory activity. *Nucleic Acids Res.* **2012**, DOI: 10.1093/nar/gks512.
- (8) Kelly, J. A.; Feigon, J.; Yeates, T. O. Reconciliation of the X-ray and NMR structures of the thrombin-binding aptamer d-(GGTTGGTGTGGTTGG). *J. Mol. Biol.* **1996**, *256*, 417–422.
- (9) Russo Krauss, I.; Merlino, A.; Giancola, C.; Randazzo, A.; Mazzarella, L.; Sica, F. Thrombin-aptamer recognition: a revealed ambiguity. *Nucleic Acids Res.* **2011**, *39*, 7858–7867.
- (10) Pagano, B.; Martino, L.; Randazzo, A.; Giancola, C. Stability and binding properties of a modified thrombin binding aptamer. *Biophys. J.* **2008**, *94*, 562–569.
- (11) Martino, L.; Virno, A.; Randazzo, A.; Virgilio, A.; Esposito, V.; Giancola, C.; Bucci, M.; Cirino, G.; Mayol, L. A new modified thrombin binding aptamer containing a 5'-5' inversion of polarity site. *Nucleic Acids Res.* **2006**, *34*, 6653–6662.
- (12) (a) Tasset, D. M.; Kubik, M. F.; Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.* **1997**, *272*, 688–698. (b) Zhou, G.; Huang, X.; Qu, Y. Thi binding effect of aptamers on thrombin. *Biochem. Eng. J.* **2010**, *52*, 117–122.
- (13) (a) Kim, Y.; Cao, Z.; Tan, W. Molecular assembly for high-performance bivalent nucleic acid inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5664–5669. (b) Müller, J.; Freitag, D.; Mayer, G.; Pötzsch, B. Anticoagulant characteristics of HD1-22, a bivalent aptamer that specifically inhibits thrombin and prothrombinase. *J. Thromb. Haemost.* **2008**, *6*, 2105–2112.
- (14) Petrer, N. S.; Stafford, A. R.; Leslie, B. A.; Kretz, C. A.; Fredenburgh, J. C.; Weitz, J. I. Long range communication between exosites 1 and 2 modulates thrombin function. *J. Biol. Chem.* **2009**, *284*, 25620–25629.
- (15) Sabo, T. M.; Farrell, D. H.; Maurer, M. C. Conformational analysis of γ' peptide (410–427) interactions with thrombin anion binding exosite II. *Biochemistry* **2006**, *45*, 7434–7445.
- (16) Tsiang, M.; Jain, A. K.; Dunn, K. E.; Rojas, M. E.; Leung, L. L.; Gibbs, C. S. Functional mapping of the surface residues of human thrombin. *J. Biol. Chem.* **1995**, *270*, 16854–16863.
- (17) Schwienhorst, A. Direct thrombin inhibitors a survey of recent developments. *Cell. Mol. Life Sci.* **2006**, *63*, 2773–2791.
- (18) (a) He, G.-X.; Krawczyk, S. H.; Swaminathan, S.; Shea, R. G.; Dougherty, J. P.; Terhorst, T. N2- and C8-substituted oligodeoxynucleotides with enhanced thrombin inhibitory activity in vitro and in vivo. *J. Med. Chem.* **1998**, *41*, 2234–2242. (b) Buff, M. C. R.; Schäfer, F.; Wulffen, B.; Müller, J.; Pötzsch, B.; Heckel, A.; Mayer, G. Dependence of aptamer activity on opposed terminal extensions: improvement of light-regulation efficiency. *Nucleic Acids Res.* **2010**, *38*, 2111–2118.
- (19) (a) He, G.-X.; Williams, J. P.; Postich, M. J.; Swaminathan, S.; Shea, R. G.; Terhorst, T.; Law, V. S.; Mao, C. T.; Sueoka, C.; Coutré, S.; Bischofberger, N. In vitro and in vivo activities of oligodeoxynucleotide-based thrombin inhibitors containing neutral formacetal linkages. *J. Med. Chem.* **1998**, *41*, 4224–4231. (b) Saccá, B.; Lacroix, L.; Mergny, J.-L. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **2005**, *33*, 1182–1192.
- (20) (a) Oliviero, G.; Borbone, N.; Galeone, A.; Varra, M.; Piccialli, G.; Mayol, L. Synthesis and characterization of a bunched oligonucleotide forming a monomolecular parallel quadruplex structure in solution. *Tetrahedron Lett.* **2004**, *45*, 4869–4872. (b) Oliviero, G.; Amato, J.; Borbone, N.; Galeone, A.; Petraccone, L.; Varra, M.; Piccialli, G.; Mayol, L. Synthesis and characterization of monomolecular DNA G-quadruplexes formed by tetra-end-linked oligonucleotides. *Bioconjugate Chem.* **2006**, *17*, 889–898. (c) Oliviero, G.; Amato, J.; Borbone, N.; D'Errico, S.; Galeone, A.; Mayol, L.; Haider, S.; Olubiyyi, O.; Hoorelbeke, B.; Balzarini, J.; Piccialli, G. Tetra-end-linked oligonucleotides forming DNA G-quadruplexes: a new class of aptamers showing anti-HIV activity. *Chem. Commun.* **2010**, *46*, 8971–8973.
- (21) Coppola, T.; Varra, M.; Oliviero, G.; Galeone, A.; D'Isa, G.; Mayol, L.; Morelli, E.; Bucci, M.-R.; Vellecco, V.; Cirino, G.; Borbone, N. Synthesis, structural studies and biological properties of new TBA analogues containing an acyclic nucleotide. *Bioorg. Med. Chem.* **2008**, *16*, 8244–8253.
- (22) Pasternak, A.; Hernandez, F. J.; Rasmussen, L. M.; Vester, B.; Wengel, J. Improved thrombin binding aptamer by incorporation of a single unlocked nucleic acid monomer. *Nucleic Acids Res.* **2011**, *39*, 1155–1164.
- (23) But, T. Y. S.; Toy, P. H. The Mitsunobu reaction: origin, mechanism, improvements, and applications. *Chem. Asian J.* **2007**, *2*, 1340–1355.
- (24) (a) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36*, 5482–5515. (b) Neidle, S.; Balasubramanian, S. The role of cations in determining quadruplex structure and stability. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; RCS Publishing: London, UK, 2006; pp 100–130. (c) Chaires, J. B.; Gray, R. D. Kinetics and mechanism of K^+ - and Na^+ -induced folding of models of human telomeric DNA into G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36*, 4191–4209.

(25) (a) Jakubowski, H. V.; Kline, M. D.; Owen, W. G. The effect of bovine thrombomodulin on the specificity of bovine thrombin. *J. Biol. Chem.* **1986**, *261*, 3876–3882. (b) Liaw, P. C.; Fredenburgh, J. C.; Stafford, A. R.; Tulinsky, A.; Austin, R. C.; Weitz, J. I. Localization of the thrombin-binding domain on prothrombin fragment 2. *J. Biol. Chem.* **1998**, *273*, 8932–8939.

(26) Kretz, C. A.; Stafford, A. R.; Fredenburgh, J. C.; Weitz, J. I. HD1, a thrombin-directed aptamer, binds exosite 1 on prothrombin with high affinity and inhibits its activation by prothrombinase. *J. Biol. Chem.* **2006**, *281*, 37477–37485.

(27) Renault, J.; Laduree, D.; Robba, M. Synthesis and antiviral study of cyclopentano[d]pyrimidine-2,4-diones and octahydroquinazoline-2,4-diones acyclic nucleosides as potential anti-HIV agents. *Nucleosides, Nucleotides Nucleic Acids* **1994**, *13*, 891–901.

(28) Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **2007**, *35* (Suppl. 2), W43–W46.

(29) (a) Hwang, T. L.; Shaka, A. J. Water suppression that works. Excitation sculpting using arbitrary wave forms and pulsed field gradients. *J. Magn. Reson. Ser. A* **1995**, *A112*, 275. (b) Dalvit, C. Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *J. Biomol. NMR* **1998**, *11*, 437.

(30) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and energetics of ligand binding to proteins: *E. coli* dihydrofolate reductase-trimethoprim, a drug receptor system. *Proteins* **1988**, *4*, 31–47.

(31) Fletcher, R. Unconstrained optimization. In *Practical Methods of Optimization*; Wiley: New York, 1980; Vol. 1.

(32) (a) Neidle, S. *Principles of Nucleic Acid Structures*; Academic Press: London, UK, 2008. (b) Reichert, J.; Sühnel, J. The IMB Jena Image Library of Biological Macromolecules 2002 Update. *Nucleic Acids Res.* **2002**, *30*, 253–254.

d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads

Nicola Borbone^{1,*}, Jussara Amato², Giorgia Oliviero², Valentina D'Atri²,
Valérie Gabelica^{3,*}, Edwin De Pauw³, Gennaro Piccialli² and Luciano Mayol¹

¹Faculty of Pharmacy, ²Faculty of Biotechnology, Dipartimento di Chimica delle Sostanze Naturali, University of Naples Federico II, 80131 Naples, Italy and ³Physical Chemistry and Mass Spectrometry Laboratory, Department of Chemistry, University of Liège, Belgium

Received January 12, 2011; Revised May 11, 2011; Accepted May 29, 2011

ABSTRACT

Among non-canonical DNA secondary structures, G-quadruplexes are currently widely studied because of their probable involvement in many pivotal biological roles, and for their potential use in nanotechnology. The overall quadruplex scaffold can exhibit several morphologies through intramolecular or intermolecular organization of G-rich oligodeoxyribonucleic acid strands. In particular, several G-rich strands can form higher order assemblies by multimerization between several G-quadruplex units. Here, we report on the identification of a novel dimerization pathway. Our Nuclear magnetic resonance, circular dichroism, UV, gel electrophoresis and mass spectrometry studies on the DNA sequence dCGGTGGT demonstrate that this sequence forms an octamer when annealed in presence of K⁺ or NH₄⁺ ions, through the 5'-5' stacking of two tetramolecular G-quadruplex subunits via unusual G(:C):G(:C):G(:C):G(:C) octads.

INTRODUCTION

G-quadruplexes are unique structures formed by Hoogsteen-type base pairing between four guanines and involving chelation of a metal ion. G-quadruplexes are typically formed by intramolecular folding of guanine-rich oligonucleotide (GRO) sequences or by intermolecular association of two or four sequences leading to the formation of dimeric or tetrameric complexes (1). The biological importance of these structures is 3-fold: (i) the occurrence of short G-rich sequences able to fold into G-quadruplex structures at the ends of telomeric DNA in eukaryotic chromosomes (2,3); (ii) the high prevalence of G-rich

sequences in a large number of eukaryotic and prokaryotic genomes, and the increasing number of G-quadruplexes arising from these sequences (2,4,5); and (iii) their presence in the scaffold of several aptamers that have the ability to selectively bind to biologically relevant proteins and small molecules (6–10).

G-quadruplexes have the ability to form an array of conformations differing in structural features such as the molecularity, the relative orientation of the strands, the size of the loops connecting the strands and the glycosidic conformation of guanosine residues (syn or anti) (11). Further structural diversity arises from the quadruplex capacity to accommodate A-, T- and C-tetrads (12–14), as well as tetrads formed by modified residues (15,16). Moreover, in the last few years, several papers describing the higher order packing of intramolecular (17–19) or intermolecular (20,21) DNA quadruplexes have also been published. Particular attention has been devoted to the understanding of mechanisms responsible for the self-organization of human telomeric DNA into higher ordered G-quadruplex structures (17,19,22). Knowledge of such mechanisms is precious for shedding light on the *in vivo* role of quadruplex DNA as well as for the design of supramolecular DNA structures provided with novel functionalities. Understanding and controlling higher order structure formation is also of prime importance for *in vitro* studies of oligonucleotides derived from biologically relevant sequences, and for controlling the aggregation state of oligonucleotide aptamers aimed at therapeutic applications.

One of the identified structural moieties required for quadruplex dimerization is the presence of a stretch of guanines at the 5'- or 3'-termini of parallel G-quadruplex-forming GROs. These types of GROs have been reported to form higher ordered 'dimer G-quadruplexes' by either the association of interlocked slipped strands (23) or by end-to-end stacking (20,23). The presence of even a single

*To whom correspondence should be addressed. Tel: +39 081678521; Fax: +39 081678552; Email: nicola.borbone@unina.it
Correspondence may also be addressed to Valérie Gabelica. Tel: +32 43663432; Fax: +32 43663413; Email: v.gabelica@ulg.ac.be

base other than G at 5'- or 3'-termini has been reported to prevent the dimerization. In this context, we report here the results of our structural investigations on the sequence CGGTGGT under different conditions of oligodeoxynucleotide (ODN) concentration, pH and salts. We observe that the sequence folds into a 'dimeric G-quadruplex' (from now on **2Q**), which is in fact an octamer of the sequence, coexisting in solution with minor amounts of other DNA secondary structures and with the single-stranded molecule. This higher ordered structure is obtained by head-to-head stacking between two G(:C):G(:C):G(:C):G(:C) octads, each belonging to one of the two 'monomeric G-quadruplexes' (**1Q**), as ascertained from the results described below.

MATERIALS AND METHODS

DNA sample preparation

DNA oligonucleotides TGGGGT, CTGTGTT, TGGTGGC, CGGTGGT and CGGGGT were chemically synthesized on an 8909 DNA/RNA synthesizer (Applied Biosystem), or purchased from Eurogentec (Belgium) for mass spectrometry experiments. DNA concentration expressed in strand molarity was calculated using a nearest-neighbor approximation for the absorption coefficients of the unfolded species. Ammonium acetate was obtained from a 5 M solution from Fluka (Molecular Biology grade). Stock solutions of 6 mM of each ODN were obtained by dissolving the lyophilized samples in 150 mM NH₄OAc or 100 mM and 1.0 M K⁺ buffers (10 mM KH₂PO₄ supplemented with 90 mM or 990 mM KCl, respectively). Hundred percent D₂O (Armar chemicals) or H₂O/D₂O (9:1, v/v) was used for the preparation of 6 mM solutions to be used for NMR experiments. All ODN samples were annealed by heating at 90°C for 15 min and then cooling and storing at 5°C. The pH was adjusted to 7.0 or 4.5 using HCl/KOH or AcOH/NH₄OH aqueous solutions before the annealing procedure.

Polyacrylamide gel electrophoresis

The molecular size of structures were probed using non-denaturing polyacrylamide gel electrophoresis (PAGE) as previously described (24). ODN samples of 10 μM, obtained by diluting the annealed stock solutions just before the experiment, were loaded on a 15% polyacrylamide gel supplemented with 20 mM of KCl or NH₄OAc. Glycerol of 10% was added just before loading. The gels were run at 26°C at constant voltage (100 V) for 2.5 h. The bands were visualized by UV shadowing and after 'Syber green' coloration.

Electrospray mass spectrometry

A Q-TOF Ultima Global mass spectrometer (Waters, Manchester, UK) was used to characterize the samples in ammonium acetate, and to perform a kinetics experiment. ODN samples (80 μM single-strand concentration), obtained as described in the PAGE paragraph with the addition of 10% methanol, were infused at 4 μl/min in the electrospray (ESI) source that was operated in

negative ion mode (capillary voltage = -2.2 kV). Temperatures of ESI source and nitrogen desolvation gas were 80°C and 100°C, respectively. The source voltages were optimized for the best compromise to observe the single strand, tetramer and octamer: cone = 80 V, RF Lens 1 = 80 V, collision energy = 10 V and source pressure = 3.06 mbar. For the kinetics experiments, the reference strand dT₆ was added to the dCGGTGGT strand, so as to follow the intensities of the reactant (dCGGTGGT single strand, noted G₁), the intermediates (dimer = G₂, trimer = G₃, tetramer = G₄ = **1Q**) and the product (octamer = G₈ = **2Q**) relative to that of the reference. The method for obtaining relative response factors of the monomer, the tetramer **1Q** and the octamer **2Q** compared to the reference is described elsewhere (25), and the details of its application in the present case are given in [Supplementary Data](#). In addition, a Bruker Apex-Qe9.4 T ESI-FTICR mass spectrometer was used to confirm the charge states from the isotopic distributions by recording high-resolution data, and to confirm the number of ammonium cations retained in each complex. The instrument voltages were chosen so as to minimize collision-induced dissociation and therefore keep the non-covalent complexes intact, including the inner ammonium cations: cap exit = -22 V, skimmer = -15 V, octapole offset = -4 V and coll cell trap = +4 V.

Circular dichroism

Circular dichroism (CD) spectra were measured at 25°C on a Jasco J-715 spectropolarimeter equipped with a Peltier Jasco JPT423S, using a 0.1-cm path length quartz cuvette with a reaction volume of 500 μl. CD spectra were recorded at different time points after the annealing on 40 μM ODN samples annealed at strand concentration of 200 μM or 6.0 mM. Spectra were averaged over three scans. A buffer baseline was subtracted from each spectrum and the spectra were normalized to have zero at 360 nm.

Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) data were collected on a Varian ^{UNITY}INOVA 500 MHz spectrometer equipped with a broadband inverse probe with z-field gradient, and on a Varian ^{UNITY}INOVA 700 MHz spectrometer equipped with a triple resonance cryoprobe. The data were processed using the Varian VNMR software package. One-dimensional NMR spectra were acquired as 16 384 data points with a recycle delay of 1.0 s at 5, 25 and 65°C. Data sets were zero filled to 32 768 points prior to Fourier transformation and apodized with a shifted sine bell squared window function. Two-dimensional NMR spectra were all acquired at 25°C using a recycle delay of 1.2 s. NOESY spectra were acquired with mixing times of 100, 200 and 300 ms. TOCSY spectra were recorded with the standard MLEV-17 spin-lock sequence and a mixing time of 80 ms. For the experiments in H₂O, water suppression was achieved by including a double pulsed-field gradient spin-echo (DPFGSE) module (26,27) in the pulse sequence prior to acquisition. In all 2D experiments, the time domain data consisted of 2048 complex points in t₂ and 400 fids in t₁ dimension. NMR samples were prepared at

6 mM strand concentration in 0.6 ml of 100 mM or 1.0 M K^+ buffer obtained as above described. Sequence-specific resonance assignments for CGGTGGT were obtained by using NOESY and TOCSY spectra following the standard procedure (28,29). The self as well as sequential NOE connectivities from H8 to H1' and H2'/H2'' were traced for all the nucleotides. The connectivities followed the standard patterns as for B-DNA, apart from noticeable exceptions between C₁ and G₂, and complete assignments of base and sugar protons was obtained.

Molecular modeling

The starting structure was generated using the builder module of Insight II (2005) program (Accelrys, Inc) on an Intel PC workstation running Red Hat Enterprise Linux. The model was parameterized according to the AMBER force-field, and calculations were performed using a distance-dependent macroscopic dielectric constant of 4.0 and an infinite cut-off for non-bonded interactions to partially compensate for the lack of solvent. The details of model building and energy minimization are given in [Supplementary Data](#).

RESULTS

dCGGTGGT forms an octamer

The formation of a higher order assembly than a tetramer by the DNA sequence dCGGTGGT was first noticed in native PAGE experiments. Figure 1a shows PAGE

experiments on the following DNA sequences: TGGGGT annealed in 0.1 M potassium buffer (lane 1) used as a tetramolecular quadruplex ([dTGGGGT]₄) size-marker, CTGTGTT (lane 2) used as a single-stranded 7-mer marker, TGGTGGC (lanes 3 and 4), CGGTGGT (lanes 5 and 6) and CGGGGT (lanes 7 and 8). The sequence TGGTGGC migrates similarly to [dTGGGGT]₄; whereas, the sequence CGGTGGT shows a major band migrating much slower. We conclude that TGGTGGC forms prevalently a tetrameric assembly, in agreement with the reported tetramer G-quadruplex formation for this sequence (30), whereas C GGTGGT forms a higher order assembly. The polarity of the sequence therefore appears crucial in the higher order structure formation. The migration profile of dCGGGGT has also been investigated to determine whether the additional central thymine played any role in the higher order structure formation. Two bands are observed for dCGGGGT, the faster one migrating as the tetrameric [dTGGGGT]₄, and the slower one migrating like the structure formed by CGGTGGT.

As the PAGE results are similar in NH_4^+ and in K^+ cations, we therefore used ESI-MS, which can only be carried out in NH_4^+ , to determine unambiguously the strand stoichiometry of the higher order assemblies formed by CGGTGGT and CGGGGT. The ESI-MS spectra of TGGTGGC, CGGTGGT and CGGGGT, each separately annealed at 6 mM in 150 mM NH_4OAc (pH = 7.0) and diluted to 80 μ M final strand concentration, are shown in Figure 1b–d, respectively. All three sequences show peaks corresponding to a tetramolecular

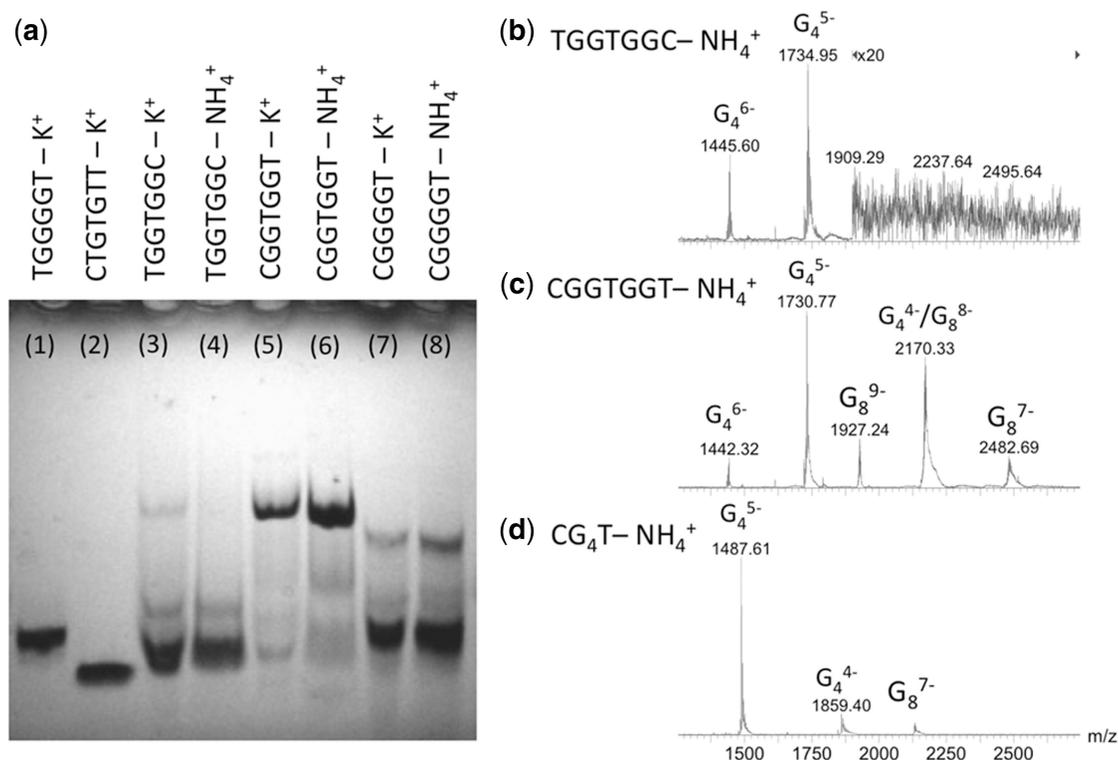


Figure 1. Stoichiometry of self-assemblies determined by (a) PAGE for sequences TGGGGT (lane 1), CTGTGTT (lane 2), TGGTGGC (lanes 3 and 4), CGGTGGT (lanes 5 and 6) and CGGGGT (lanes 7 and 8), and by (b–d) ESI-MS for TGGTGGC (b), CGGTGGT (c) and CGGGGT (d). In the peak annotation G_n^z , n indicates the number of strands and z indicates the charge.

assembly (noted G_4), but the spectra of CGGTGGT and CGGGGT also display peaks corresponding to an octameric assembly (noted G_8). We therefore assign the slow migrating PAGE band to an octameric self-assembly. In terms of relative intensities, although native PAGE and ESI-MS experiments were performed on samples obtained by dilution of the same annealed stock solution, the octamer is the darkest band in the gels of CGGTGGT (lanes 5 and 6), whereas the ESI-MS peaks of the octamer have lower intensities than the tetramer peaks. The analysis of the electrospray response factors in the kinetics ESI-MS experiments (see below and [Supplementary Figure S2](#)) indeed reveals the octamer has a lower ESI-MS response than the tetramer in the spectrometer used, and that the octamer is indeed the major species in solution.

The octamer contains stacked G-tetrads

All G-rich sequences studied here are expected to form G-quadruplex structures. Note that 'G-quadruplex' means that the structure contains stacked G-quartets, not necessarily that the assembly is a tetramer. CD was used to

probe the relative orientation of the bases in the tetramer $[TGGTGGC]_4$ and in the octamer $[CGGTGGT]_8$. The overall CD profiles recorded in K^+ (Figure 2a) and NH_4^+ (Figure 2b) are in agreement with the formation of G-quadruplex structures (31,32) both in K^+ and NH_4^+ media, showing a positive maximum at about 260 nm and a negative minimum close to 240 nm, which are characteristics of head-to-tail arrangement of guanines, as typically found in parallel oligodeoxynucleotides G-quadruplexes (33–35). Interestingly, the CD spectra of CGGTGGT also show a weak, but clearly visible negative band centered at 290 nm. This minimum suggests that additional base stacking interactions could be involved in the formation of the higher ordered quadruplex species by allowing the quadruplex multimerization via end-to-end stacking. A typical heteropolar stacking in G-quadruplexes formed by strands with no inverted polarity usually results in a positive band at 290 nm (34), but negative bands at 290 nm have been observed in octameric self-assemblies of lipophilic guanosine derivatives (36). The type of stacking present in our octamers could therefore resemble that present in those lipophilic octamers.

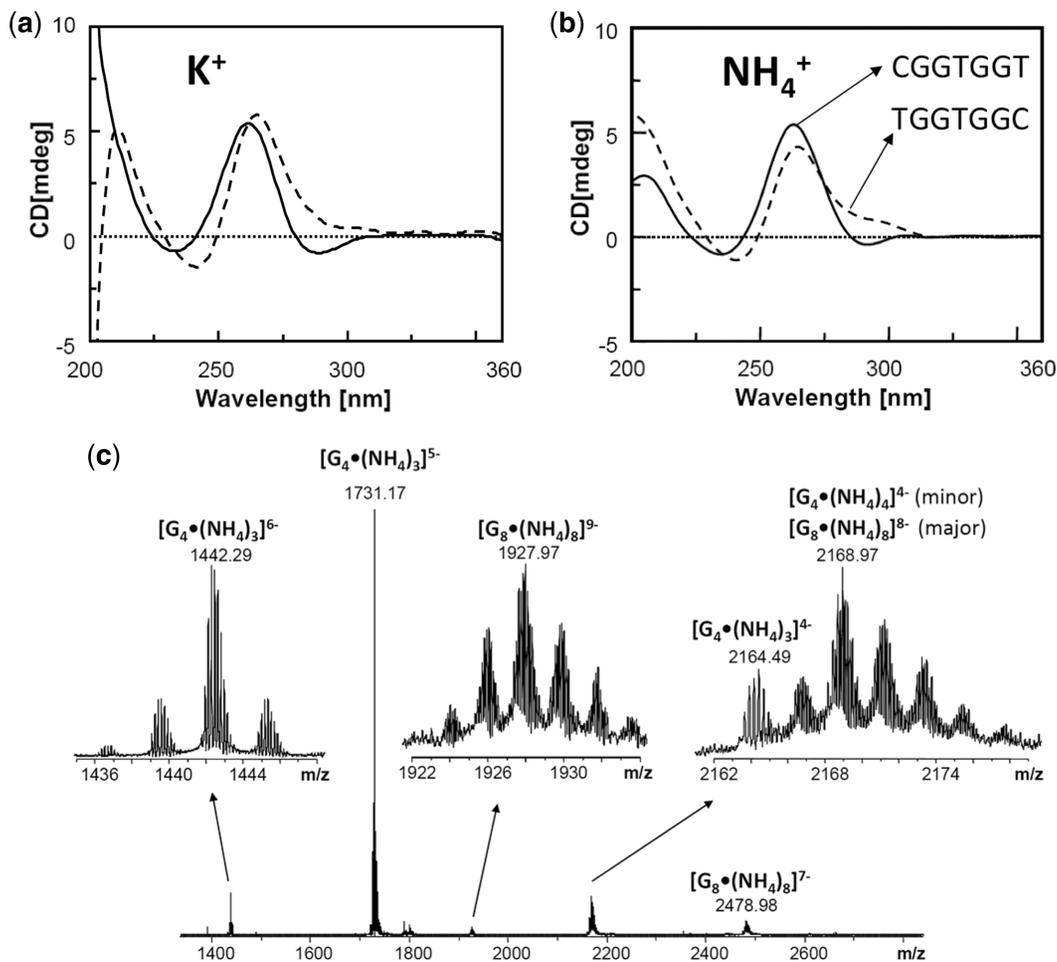


Figure 2. G-quadruplex formation probed by (a and b) CD in 100 mM KCl (a) and 150 mM NH_4OAc (b) of the sequences CGGTGGT (solid line) and TGGTGGC (dotted line), and by (c) counting the number of ammonium cations incorporated in the tetramer (noted G_4) and the octamer (noted G_8) of CGGTGGT using high resolution ESI-FTICR mass spectrometry.

Because the CD signature is peculiar and does not allow to conclude unambiguously to a G-quadruplex structure, we decided to probe cation incorporation into the octamer. Indeed, successive G-quartets incorporate monovalent cations (here K^+ or NH_4^+) by coordination between eight guanines. High-resolution ESI-FTICRMS was used to count the number of ammonium cations per tetramer and per octamer (Figure 2c). Cation incorporation was found in all cases, confirming G-quadruplex formation both in the tetramer and the octamer. The major peaks correspond to three cations per tetramer $[dCGGTGGT]_4$ and eight cations per octamer $[dCGGTGGT]_8$. The number of cations is indicative of the number of stacked quartets. Interestingly, the tetramer $[dCGGTGGT]_4$ contains three ammonium cations, whereas the tetramer $[dTGGTGGC]_4$ contains four (data not shown), suggesting a structural difference depending on the strand polarity. However, the incorporation of eight ammonium cations in the octamer suggests that within the octamer, each tetrameric subunit is now capable of incorporating four ammoniums cations.

Influence of sample preparation procedure

In all experiments described above, the octamer was obtained by annealing the strand CGGTGGT at 6 mM single-strand ODN concentration at pH = 7.0 in 100 mM K^+ or 150 mM NH_4^+ . CD experiments carried out using a higher K^+ concentration (1 M) show that higher cation concentration further favors the formation of the structure characterized by the negative band at 290 nm (Supplementary Figure S1). We also tested the influence of the solution pH, to test whether cytosine protonation might favor the formation of the new structure, but no clear influence of pH was found (see Supplementary Figure S1 and NMR results described below and in Supplementary Figure S4).

However, because of the octamer stoichiometry, we anticipated that the strand concentration upon annealing would have a great influence on the formation kinetics. Figure 3a shows the CD spectra recorded on CGGTGGT annealed at 6.0 mM or 0.2 mM ODN concentration in 1 M K^+ . The presence of the negative band at 290 nm was strictly dependent from ODN concentration: it is present when the sample is annealed starting from the 6 mM stock solution, whereas it is not observed when the sample is obtained by annealing the 0.2 mM ODN (Figure 3a). However, the CD spectrum of the species formed from the 0.2 mM ODN solution resembles that of a regular tetramolecular G-quadruplex. The influence of the strand concentration, therefore, indicates slow formation kinetics for the octamer.

To study the kinetics of octamer formation in the 6 mM stock solution, and to probe the nature of the reaction intermediates, we recorded ESI-MS spectra as a function of the time after addition of cation in the ODN solution, at room temperature. This experiment could be carried out only in ammonium acetate (150 mM). The details of the methodology, representative ESI-MS spectra at different reaction times, and the relative intensities are presented in Supplementary Figure S2. Briefly, the

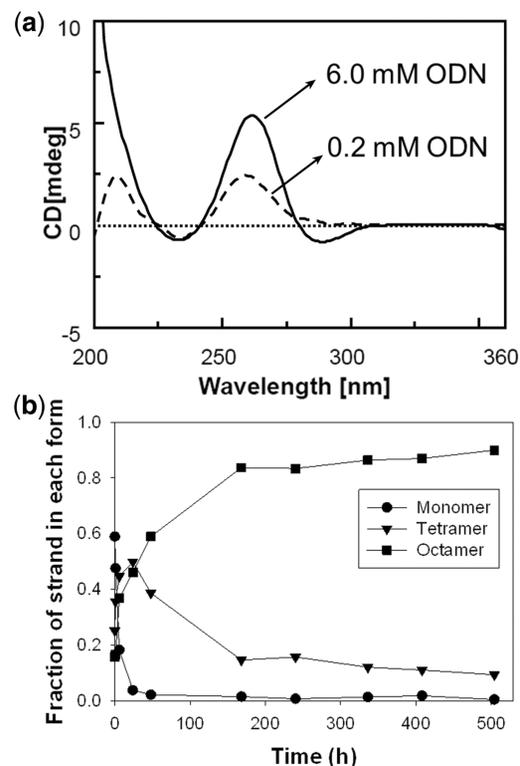


Figure 3. Influence of the sample preparation conditions on the octamer formation. (a) Influence of strand concentration: CD spectra of CGGTGGT in 1.0 M K^+ buffer annealed, respectively, at 6 mM (solid line) and 200 μ M (dotted line) single-strand concentration. (b) Kinetics of tetramer and octamer formation of 6 mM CGGTGGT in 150 mM ammonium acetate at room temperature, as determined by time-resolved ESI-MS.

single-stranded dT₆ strand (0.75 mM) is added to the 6 mM single-stranded dCGGTGGT in water. Quadruplex formation is initiated by the addition of ammonium acetate (final concentration: 150 mM) at room temperature. At each time point, an aliquot of the sample is diluted to 80 μ M final dCGGTGGT single-strand concentration, and injected in the ESI-MS. We therefore assume that, once formed, the octamer resists the dilution step. The signal intensity of the reference is used to normalize the intensity variations of the dCGGTGGT monomer, tetramer and octamer signal, and deduce their respective response factors (see Supplementary Figure S2). Figure 3b shows the time evolution of the relative proportions of single strand, tetramer and octamer, after correction for the relative response of each species, expressed as the fraction of single strand in each form. The single strand is quickly converted into a tetramer, but the tetramer-to-octamer conversion is slower. The dimerization is therefore the rate-limiting step in the octamer formation. The results also show that, although the octamer is not the most abundant peak in the mass spectra (see Figures 1b–d and 2c), it is nevertheless the predominant species. The predominance of the octamer at long time scales also confirms that, once formed, the octamer resists the dilution step.

NMR investigation of the octamer structure

The ^1H NMR spectrum of CGGTGGT in 100 mM K^+ buffer at 25°C recorded 1 h after the annealing (Supplementary Figure S3, bottom left) is characterized by the presence of four well-resolved signals in the 11.0–11.6 ppm region attributable to the exchange-protected imino protons involved in the formation of Hoogsteen hydrogen bonds of four G-tetrads (37–39), in close analogy with what was previously observed for the parallel quadruplex structure formed by TGGTGGC (30). Furthermore, the count of quadruplex NMR signals is the same as what must be expected for the NMR spectrum of the corresponding single strand, thus suggesting the formation of a highly symmetric parallel G-quadruplex possessing a 4-fold symmetry. However, the aromatic region of ^1H NMR spectrum is populated by more than the expected seven signals belonging to the

H8/H6 aromatic protons of nucleobases. In order to discriminate whether the minor peaks were due to sample impurities or to the coexistence in solution of alternative minor conformations of CGGTGGT, we recorded the ^1H NMR spectrum also at 65°C, and the high-temperature spectrum was clearly in agreement with the second hypothesis as evidenced by the presence of just seven well-defined signals in the aromatic region of the spectrum. We therefore studied the influence of time elapsed after the annealing (Supplementary Figure S3), of K^+ concentration (Supplementary Figure S3), and of pH (Supplementary Figure S4) on the 1D NMR spectra. In agreement with the CD data, the NMR results confirmed that the best medium for the structural study of the new octameric structure formed by CGGTGGT requires 1.0 M K^+ and long folding times (Figure 4a), and that the change of pH from 7.0 to 4.5 does not affect the folding of the structure.

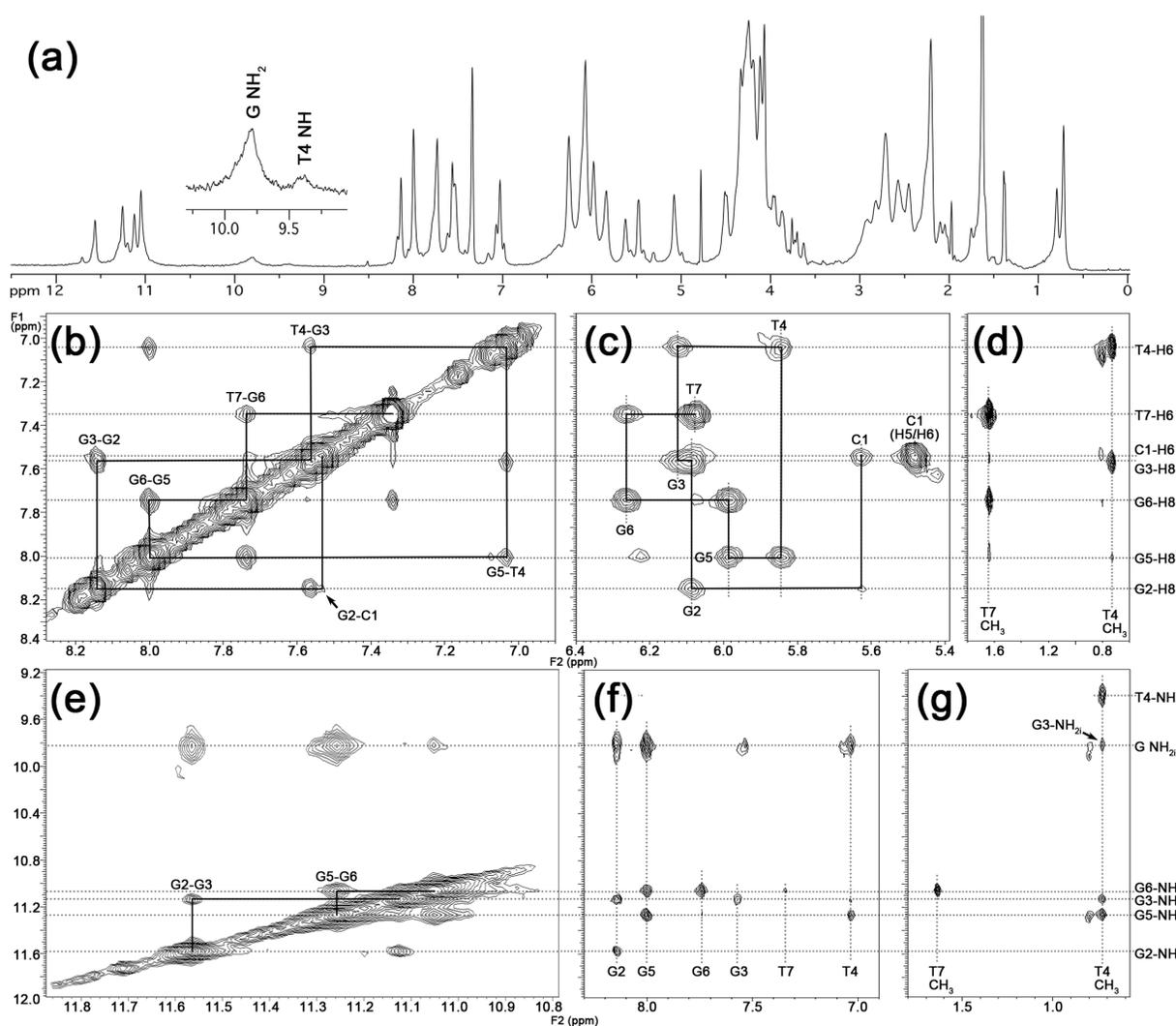


Figure 4. Study of the octamer structure by NMR. (a) ^1H -NMR spectra recorded at 25°C of CGGTGGT in $\text{H}_2\text{O}/\text{D}_2\text{O}$ 9:1, annealed at 6 mM strand concentration in 1 M K^+ at pH = 7.0. (b–g) Expansions of 2D NOESY spectrum of CGGTGGT (25°C, 200 ms mixing time, 1 M K^+ , pH = 7.0). The sequential H8-H6 and self-peaks involving H1' protons have been labeled in (b) and (c), respectively. (d) and (g) show NOE cross-peaks involving T4/T7 methyl protons and H8/H6 or imino and amino protons, respectively. The T-tetrad diagnostic peaks between T4Me and T4NH and sequential T4Me-G3NH, T4Me-G3NH₂ and T4Me-G5NH are observable in panel (g). Imino-imino NOE correlations are shown in (e). The sequential imino-imino NOE correlations have been drawn in (e). NOE correlations between aromatic H8/H6 and imino and amino protons are shown in (f).

The combined analysis of 2D NOESY and TOCSY spectra (700 MHz, 25°C, 1 M K⁺, 15 days after the annealing) allowed us to get the complete assignment (Table 1) of both exchangeable and non-exchangeable protons of CG GTGGT. The sequential connectivities between the aromatic protons (Figure 4b) were traced, as well as the self and sequential NOE connectivities between H8/H6 base protons and H1', H2', H2'' and H3' sugar protons. The intensities of intranucleotide H8/H1' NOE cross-peaks (Figure 4c) indicate that all Gs adopt an *anti* glycosidic conformation. This finding together with the observation for each G-tetrad of a GiH8/GjN1H NOE cross-peak (Figure 4f), with $i = j$ (where i and j represent the nucleotide labels), and with the observation of sequential G2N1H/G3N1H and G5N1H/G6N1H cross-peaks (Figure 4e), definitely allowed us to identify the presence of 4-fold symmetric parallel quadruplex moieties in the higher ordered structure formed by CGGTGGT when annealed in K⁺-containing buffer. In close analogy with what observed by Patel, P.K. and Hosur, R.V. for the quadruplex structure formed by TGGTGGC (30), we observed the formation of a central T-tetrad as ascertained by the appearance of the diagnostic T4NH/T4Me NOE cross-peak (Figure 4g) and supported by the observation of the strong shielding ring current experienced by T4 methyl protons that are significantly upfield-shifted when compared to the corresponding signals of the 'free' T7 (0.72 versus 1.64 ppm).

Although the overall observed NOEs account for a 'classical' tetramolecular parallel quadruplex, two peculiar NOE evidences drove us toward the disclosure of the higher ordered octamer structure. In [Supplementary Figure S5](#), we show the 1D cross-section of the NOESY spectrum (200 ms mixing time) of CGGTGGT taken at 5.66 ppm (C1H1' protons). This figure shows that the NOE between C1H1' and C1H5 has almost the same intensity than that between C1H1' and C1H6. These NOEs are in turn stronger than the sequential NOE C1H1'-G2H8, but weaker, as expected, than those between C1H1' and self H2'/H2''. The relative intensities of the aforementioned NOEs involving C1H1' require that the C1H1'-C1H5 distance is in the same range of the intranucleotide C1H1'-C1H6 distance (2.31–3.74 Å depending on the glycosyl torsion angle). This experimental distance is compatible only with the occurrence of an intermolecular dipolar coupling between C1 bases, because the corresponding intranucleotide C1H1'-C1H5 distance

would be in the range 4.64–5.49 Å. This finding, together with the observation of the very weak sequential NOE between C1H6 and G2H8 (marked with the arrow in Figure 4b), disclosed that, in the octamer, each C1 base does not stack over the flanking G2 tetrad, as one may expect, but projects outside the quadruplex scaffold in such a way to allow dipolar couplings with one of the four C1 bases belonging to the other tetramer. This unusual arrangement might be favored by the formation of additional C1NH2–G2N3 and G2NH2–C1N3 hydrogen bonds, thus allowing the formation of an unusual G2(:C1):G2(:C1):G2(:C1):G2(:C1) octad (Figure 5a) in each of the two tetrameric quadruplexes (**1Q**).

DISCUSSION

The structure adopted by the sequence dCGGTGGT must account for the following observations. It is a highly symmetrical octamer, formed from two parallel-stranded tetramers. In the octamer, each tetramer contains a central T-tetrad and the G-tetrads are stacked in head-to-tail arrangement. However, at the octamer's interface between the two tetramers, the stacking of the G-tetrads is likely to be different, as determined from CD.

Interlocked structures like that formed by the (dGCGGTGGT)₄ tetramer (40) were initially taken in consideration to build an octamer model, but were discarded due to NMR inconsistencies and symmetry-related considerations. Unresponsiveness to pH changes also led us to discard structures involving protonated cytosines. The NOEs involving C1H1' point to the formation of a G2(:C1):G2(:C1):G2(:C1):G2(:C1) octad (Figure 5a) in each of the two tetrameric quadruplexes. The formation of the planar octad system would allow for further π - π stacking between C1 bases, thus adding more 'hydrophobic glue' between the two G-quadruplex tetramers that form the octamer.

Figure 5b shows the energy minimized model of the octamer formed by stacking of two tetramers, each containing a 5'-G(:C):G(:C):G(:C):G(:C) octad. The details of molecular modeling are given in [Supplementary Figure S6](#), and additional ribbon and space-filled views of the model are shown in [Supplementary Figure S6](#). The overall topology of each tetramer is largely similar to that of the tetramolecular quadruplex formed by TG GTGGC (see Figure 8 in reference 30), with the central T-tetrad causing a small decrease in the intertwining of the helices. Space filling views ([Supplementary Figure S6](#)) also evidence that the octamer structure is very compact, regardless the presence of two stacked octads. These are seen to determine only a slight increase in the thickness of the octamer at the dimerization interface.

Figure 5c–d also shows the head-to-head stacking between the two octads (c) and the stacking between tetrads G2-G3 (d). The stacking between tetrads G3-T4 and T4-G5 steps is shown in [Supplementary Figure S7](#). Although C1 and G2 bases involved in the stacked octads appear slightly rotated with respect to the tetrad plane, Figure 5c shows good stacking between the six-membered rings of G2 bases and between the stacked C1 bases. The electric

Table 1. Proton chemical shift^a (ppm) for **2Q** at 25°C, pH 7.0, 1 M K⁺

	H ₆ /H ₈	H ₅ /CH ₃	NH	NH _{2i} ^b	H _{1'}	H _{2'}	H _{2''}	H _{3'}	H _{4'}	H _{5'/H_{5''}}
C1	7.54	5.48			5.63	2.28	2.10	4.66	4.13	3.87–3.95
G2	8.14		11.56	9.82	6.09	2.72	2.96	5.10	4.49	4.18
G3	7.56		11.13	9.80	6.12	2.55	2.94	5.07	4.50	4.30
T4	7.03	0.72	9.40		5.84	2.24	2.46	4.92	4.34	4.25
G5	8.00		11.26	9.82	5.98	2.72	2.84	5.10	4.49	4.20–4.30
G6	7.73		11.05	9.81	6.27	2.58	2.71	4.98	4.52	4.11–4.25
T7	7.34	1.64			6.08	2.73	2.21	4.51	4.08	4.25

^aRelative to H₂O at 4.78 ppm.

^bNH_{2i} refers to the internal (i) H-bonded amino protons of guanosines.

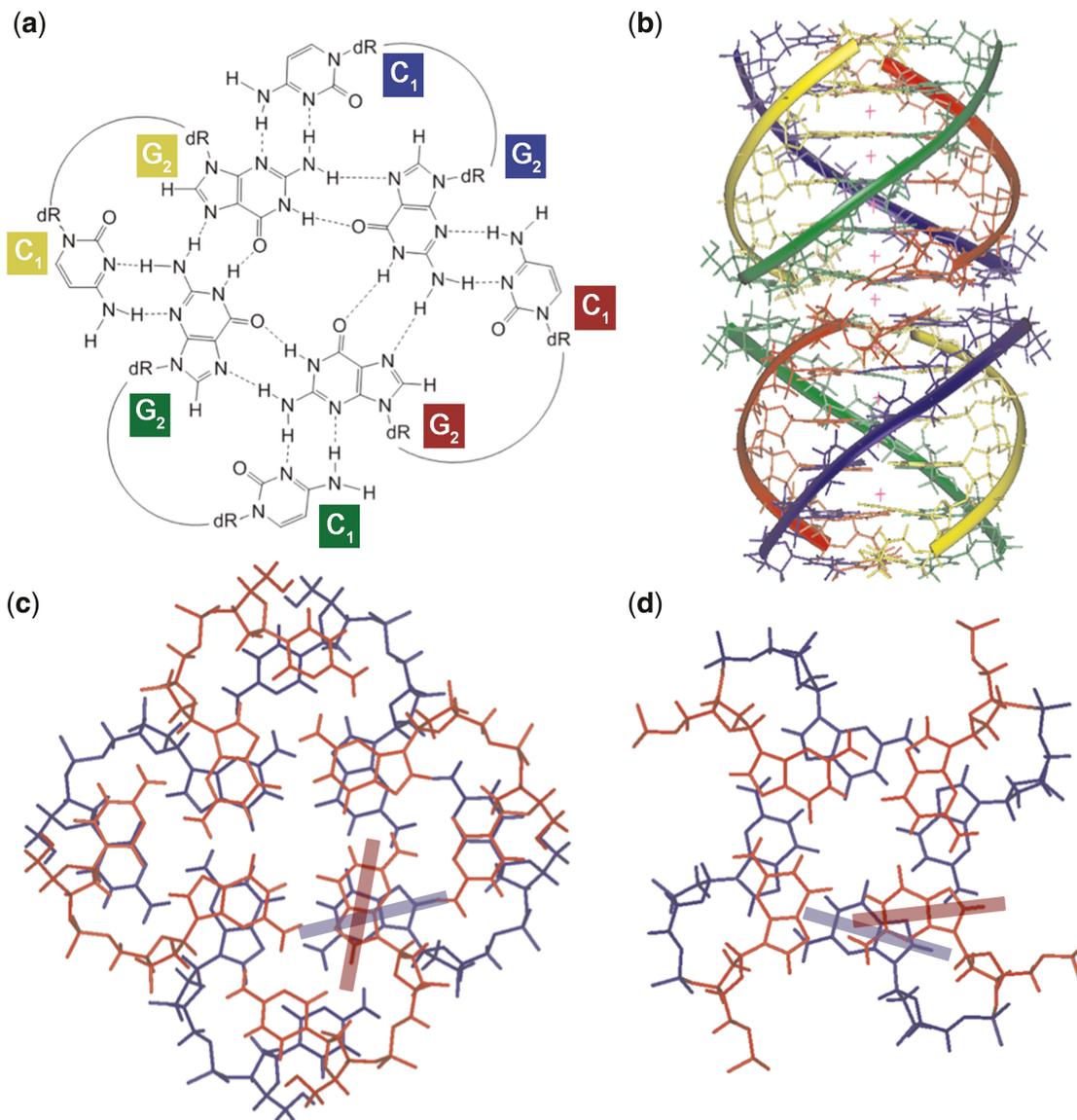


Figure 5. Structural model for the $(\text{dCGGTGGT})_8$ octamer. (a) Proposed structure of the $\text{G}_2:(\text{C}_1):\text{G}_2:(\text{C}_1):\text{G}_2:(\text{C}_1):\text{G}_2:(\text{C}_1)$ octad. (b) Molecular model of the octamer: a different color was used for each ODN strand, and potassium ions appear as red crosses. (c) Detail of stacking of the octads. (d) Detail of stacking of $\text{G}_2\text{-G}_3$ tetrads. The transition moments of guanines corresponding to the absorption band at approximately 250 nm are shown as colored bars in (c) and (d).

transition moments corresponding to the guanine absorption band at approximately 250 nm are indicated in Figure 5c and d. The stacking interactions in the model allow to explain the CD data: $\text{G}_2\text{-G}_3$ stacking resembles the standard head-to-tail stacking encountered in parallel-stranded G-quadruplexes. However, the octad stacking gives rise to a heteropolar stacking that is distinct from those typically encountered in antiparallel G-quadruplexes. We, therefore, interpret the negative CD band at 290 nm as being due to the octad stacking shown in Figure 5c (36). Interestingly, a similar weak negative CD band at 290 nm was also observed, but not discussed, by Mastugami *et al.* (41) in the dimer formed by $\text{d}(\text{GGAGGAGGAGGA})$, which contains stacked $\text{G}:(\text{A}):(\text{G}:(\text{A}):\text{G}:(\text{A}):\text{G})$ heptads at the dimerization interface.

Our results therefore suggest that, like adenine, cytosine is also capable of inducing the formation of higher order structures by the formation of octads. One must note, however, that the formation of the octamer by the sequence dCGGTGGT is slow. ESI-MS kinetics experiments in ammonium acetate provide some insight into the mechanism of octamer formation. The first step is the formation of a tetramer that incorporates only three ammonium ions. This means either that the tetramer of $[\text{dCGGTGGT}]_4$ has either a central or a terminal tetrad disrupted, or that the tetramer contain slipped strands and hence incomplete tetrad formation (42). The rate-limiting step is the conversion of tetramer to octamer, and in contrast with the tetramer, the octamer incorporates four cations per tetrameric subunit. We therefore conclude

that the rate-limiting step is the rearrangement of the tetramer into a fully aligned parallel structure that is amenable to octad formation and stacking thereof. Further experiments with various sequences are underway to probe the effect of the central T-tetrad on stability and rate of formation of octamers.

CONCLUSIONS

In summary, the present study demonstrates that the formation of a quadruplex dimer (**2Q**) can occur when the 5'-cytosine-ending CGGTGGT DNA sequence, but not the inverted TGGTGGC sequence, is annealed in the presence of K^+ or NH_4^+ ions. The NMR and molecular modeling data disclosed that the formation of **2Q** is allowed by the formation of an unusual octad arrangement, in each of the two **1Q** units, involving C1 and G2 bases of each CGGTGGT strand. Up to now, dimerization of G-quadruplexes had only been reported via guanine-adenine heptads (41), and recently via a shared uracil tetrad (43). Our finding of G-quadruplex dimerization with 5'-cytosine containing strands adds further possibilities for the design of novel quadruplex-based DNA functional nanostructures, and further complexity to the mechanisms of folding and assembly of G-rich sequences. It is particularly important for *in vitro* studies of artificially designed (e.g. aptamers) or biologically relevant sequences.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The European Commission through the COST Action MP0802 (European Network on G-quartets and Quadruplex Nucleic Acids) is gratefully acknowledged. The authors are also grateful to the Centro di Servizio Interdipartimentale di Analisi Strumentale of the University of Napoli Federico II for NMR facilities and to Dr Luisa Cuorvo and Dr Frédéric Rosu for technical assistance.

FUNDING

Italian M.U.R.S.T. (P.R.I.N. 2007); the Fonds de la Recherche Scientifique—FNRS (CC.1.5.286.09.F to V.G., research associate position to V.G.); University of Liège (starting grant FRSD-08/10 to V.G.); Walloon Region (Project FEDER FTICR). Funding for open access charge: Italian M.U.R.S.T. (P.R.I.N. 2007).

Conflict of interest statement. None declared.

REFERENCES

- Parkinson, G.N. (2006) Fundamentals of quadruplex structures. In *Quadruplex Nucleic Acids*. RSC Publishing, London, UK, pp. 1–30.

- Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
- Riou, J.-F., Gomez, D., Morjani, H. and Trentesaux, C. (2006) Quadruplex ligand recognition: biological aspects. In *Quadruplex Nucleic Acids*. RSC Publisher, London, UK, pp. 154–179.
- Dexheimer, T.S., Fry, M. and Hurley, L.H. (2006) DNA quadruplexes and gene regulation. In *Quadruplex Nucleic Acids*. RSC Publisher, London, UK, pp. 180–207.
- Huppert, J. (2006) Quadruplexes in the genome. In *Quadruplex Nucleic Acids*. RSC Publisher, London, UK, pp. 208–227.
- Levesque, D., Beaudoin, J.-D., Roy, S. and Perreault, J.-P. (2007) *In vitro* selection and characterization of RNA aptamers binding thyroxine hormone. *Biochem. J.*, **403**, 129–138.
- Pileur, F., Andreola, M.L., Dausse, E., Michel, J., Moreau, S., Yamada, H., Gaidamarov, S.A., Crouch, R.J., Toulmé, J.J. and Cazenave, C. (2003) Selective inhibitory DNA aptamers of the human RNase H1. *Nucleic Acids Res.*, **31**, 5776–5788.
- Wang, K.Y., McCurdy, S., Shea, R.G., Swaminathan, S. and Bolton, P.H. (1993) A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry*, **32**, 1899–1904.
- Chou, S.-H., Chin, K.-H. and Wang, A.H.-J. (2005) DNA aptamers as potential anti-HIV agents. *Trends Biochem. Sci.*, **30**, 231–234.
- Girvan, A.C., Teng, Y., Casson, L.K., Thomas, S.D., Jülicher, S., Ball, M.W., Klein, J.B., Pierce, W.M., Barve, S.S. and Bates, P.J. (2006) AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin. *Mol. Cancer Ther.*, **5**, 1790–1799.
- Phan, A.T., Kuryavyi, V., Luu, K.N. and Patel, D.J. (2006) Structural diversity of G-quadruplex scaffolds. In *Quadruplex Nucleic Acids*. RSC Publisher, London, UK, pp. 81–99.
- Bhavesh, N., Patel, P., Karthikeyan, S. and Hosur, R. (2004) Distinctive features in the structure and dynamics of the DNA repeat sequence GGCGGG. *Biochem. Biophys. Res. Commun.*, **317**, 625–633.
- Oliviero, G., Amato, J., Borbone, N., Galeone, A., Varra, M., Piccialli, G. and Mayol, L. (2006) Synthesis and characterization of DNA quadruplexes containing T-tetrads formed by bunch-oligonucleotides. *Biopolymers*, **81**, 194–201.
- Searle, M.S., Williams, H.E.L., Gallagher, C.T., Grant, R.J. and Stevens, M.F.G. (2004) Structure and K^+ ion-dependent stability of a parallel-stranded DNA quadruplex containing a core A-tetrad. *Org. Biomol. Chem.*, **2**, 810–812.
- Chen, J., Zhang, R.L., Min, J.M. and Zhang, L.H. (2002) Studies on the synthesis of a G-rich octaoligonucleotide (isoT)₂(isoG)₄(isoT)₂ by the phosphotriester approach and its formation of G-quartet structure. *Nucleic Acids Res.*, **30**, 3005–3014.
- Virgilio, A., Esposito, V., Randazzo, A., Mayol, L. and Galeone, A. (2005) Effects of 8-methyl-2'-deoxyadenosine incorporation into quadruplex forming oligodeoxyribonucleotides. *Bioorg. Med. Chem.*, **13**, 1037–1044.
- Haider, S., Parkinson, G.N. and Neidle, S. (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. *Biophys. J.*, **95**, 296–311.
- Matsugami, A., Okuizumi, T., Uesugi, S. and Katahira, M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G:(A):G:(A):G:(A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
- Xu, Y., Ishizuka, T., Kurabayashi, K. and Komiyama, M. (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew. Chem., Int. Ed.*, **48**, 7833–7836.
- Kato, Y., Ohshima, T., Mita, H. and Yamamoto, Y. (2005) Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAGn) (n = 3–5). *J. Am. Chem. Soc.*, **127**, 9980–9981.
- Sotoya, H., Matsugami, A., Ikeda, T., Ouhashi, K., Uesugi, S. and Katahira, M. (2004) Method for direct discrimination of intra- and intermolecular hydrogen bonds, and characterization of the G:(A):G:(A):G:(A):G heptad, with scalar couplings across hydrogen bonds. *Nucleic Acids Res.*, **32**, 5113–5118.

22. Collie, G.W., Parkinson, G.N., Neidle, S., Rosu, F., De Pauw, E. and Gabelica, V. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J. Am. Chem. Soc.*, **132**, 9328–9334.
23. Krishnan-Ghosh, Y., Liu, D. and Balasubramanian, S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.*, **126**, 11009–11016.
24. Guédin, A., De Cian, A., Gros, J., Lacroix, L. and Mergny, J.L. (2008) Sequence effects in single base loops for quadruplexes. *Biochimie*, **90**, 686–696.
25. Gabelica, V., Rosu, F. and De Pauw, E. (2009) A Simple Method to Determine Electrospray Response Factors of Noncovalent Complexes. *Anal. Chem.*, **81**, 6708–6715.
26. Hwang, T.L. and Shaka, A.J. (1995) Water suppression that works. Excitation sculpting using arbitrary wave forms and pulsed field gradients. *J. Magn. Reson. Ser. A*, **A112**, 275–279.
27. Dalvit, C. (1998) Efficient multiple-solvent suppression for the study of the interactions of organic solvents with biomolecules. *J. Biomol. NMR*, **11**, 437–444.
28. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, New York, NY.
29. Hosur, R.V., Govil, G. and Miles, H.T. (1988) Application of two-dimensional NMR spectroscopy in the determination of solution conformation of nucleic acids. *Magn. Reson. Chem.*, **26**.
30. Patel, P.K. and Hosur, R.V. (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Res.*, **27**, 2457–2464.
31. Mergny, J.L., Phan, A.T. and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
32. Mergny, J.-L., Li, J., Lacroix, L., Amrane, S. and Chaires, J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
33. Hardin, C.C., Perry, A.G. and White, K. (2001) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers (Nucleic Acid Sci.)*, **56**, 147–194.
34. Mergny, J.-L., De Cian, A., Ghelab, A., Saccà, B. and Lacroix, L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
35. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A. and Spada, G.P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org. Biomol. Chem.*, **8**, 2683–2692.
36. Graziano, C., Masiero, S., Pieraccini, S., Lucarini, M. and Spada, G.P. (2008) A cation-directed switch of intermolecular spin-spin interaction of guanosine derivatives functionalized with open-shell units. *Org. Lett.*, **10**, 1739–1742.
37. Jin, R., Gaffney, B.L., Wang, C., Jones, R.A. and Breslauer, K.J. (1992) Thermodynamics and structure of a DNA tetraplex: a spectroscopic and calorimetric study of the tetramolecular complexes of d(TG3T) and d(TG3T2G3T). *Proc. Natl Acad. Sci. USA*, **89**, 8832–8836.
38. Feigon, J., Koshlap, K.M. and Smith, F.W. (1995) ¹H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.*, **261**, 225–255.
39. Feigon, J. (1996) DNA triplexes, quadruplexes, and aptamers. *Encyclopedia of NMR*, **3**, 1726–1731.
40. Mergny, J.-L., De Cian, A., Amrane, S. and Webba da Silva, M. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res.*, **34**, 2386–2397.
41. Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S. and Katahira, M. (2001) An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.*, **313**, 255–269.
42. Rosu, F., Gabelica, V., Poncelet, H. and De Pauw, E. (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry. *Nucleic Acids Res.*, **38**, 5217–5225.
43. Sket, P. and Plavec, J. (2010) Tetramolecular DNA Quadruplexes in Solution: Insights into Structural Diversity and Cation Movement. *J. Am. Chem. Soc.*, **132**, 12724–12732.

Structure–phenotype correlations of human CYP21A2 mutations in congenital adrenal hyperplasia

Shozeb Haider^a, Barira Islam^a, Valentina D’Atri^a, Miriam Sgobba^a, Chetan Poojari^b, Li Sun^c, Tony Yuen^{c,d}, Mone Zaidi^{c,1}, and Maria I. New^{d,1,2}

^aCentre of Cancer Research and Cell Biology, Queen’s University of Belfast, Belfast BT9 7BL, United Kingdom; ^bInstitute of Complex Systems, Forschungszentrum Juelich, 52425 Juelich, Germany; and Departments of ^cMedicine and ^dPediatrics, Mount Sinai School of Medicine, New York, NY 10029

Contributed by Maria I. New, December 7, 2012 (sent for review October 5, 2012)

Mutations in the cytochrome p450 (CYP)21A2 gene, which encodes the enzyme steroid 21-hydroxylase, cause the majority of cases in congenital adrenal hyperplasia, an autosomal recessive disorder. To date, more than 100 CYP21A2 mutations have been reported. These mutations can be associated either with severe salt-wasting or simple virilizing phenotypes or with milder nonclassical phenotypes. Not all CYP21A2 mutations have, however, been characterized biochemically, and the clinical consequences of these mutations remain unknown. Using the crystal structure of its bovine homolog as a template, we have constructed a humanized model of CYP21A2 to provide comprehensive structural explanations for the clinical manifestations caused by each of the known disease-causing missense mutations in CYP21A2. Mutations that affect membrane anchoring, disrupt heme and/or substrate binding, or impair stability of CYP21A2 cause complete loss of function and salt-wasting disease. In contrast, mutations altering the transmembrane region or conserved hydrophobic patches cause up to a 98% reduction in enzyme activity and simple virilizing disease. Mild nonclassical disease can result from interference in oxidoreductase interactions, salt-bridge and hydrogen-bonding networks, and nonconserved hydrophobic clusters. A simple *in silico* evaluation of previously uncharacterized gene mutations could, thus, potentially help predict the often diverse phenotypes of a monogenic disorder.

structure–phenotype correlation | molecular modeling | CYP21A2 structural analysis

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disorder, with an overall incidence of 1:15,000 live births worldwide (1). Up to 95% of cases arise from mutations in the cytochrome p450 (CYP) 21A2 (*CYP21A2*) gene, which encodes the enzyme steroid 21-hydroxylase. Classical CAH is characterized by a complete or near-complete loss of enzyme activity, which leads to salt-wasting (SW) or simple virilizing (SV) CAH. SW CAH patients display impaired cortisol synthesis, signs of hyperandrogenemia, and features of aldosterone deficiency. In contrast, in SV CAH, patients have sufficient aldosterone to maintain sodium homeostasis. However, because of the shunting of cortisol precursors through the adrenal androgen biosynthetic pathway resulting in the formation of testosterone and dihydrotestosterone, patients display signs of hyperandrogenemia with impaired fertility. Female newborns in both SW and SV forms of CAH have ambiguous genitalia.

Nonclassical (NC) CAH is observed in 1 in 100 subjects in a heterogeneous population, for example, in New York City, with patients generally displaying a mild and variable phenotype because of a ~20–60% retention of 21-hydroxylase activity. Ethnic-specific frequency indicates that among Eastern European Jews, 1 in 27 subjects have NC CAH, making it the most frequent autosomal disorder in humans. The newborn females do not have ambiguous genitalia, whereas females postnatally often exhibit hirsutism, oligomenorrhea, and impaired fertility. In males, hyperandrogenism is difficult to detect and often only comes to attention in adulthood owing to short stature and impaired fertility (2).

CYP21 genes are a part of the RP-C4-CYP21-TNX module and are located within the major histocompatibility complex on

chromosome 6 (3). *CYP21A2* is present 30 kb downstream of its nonfunctional pseudogene *CYP21A1*. Importantly, however, the two genes share 98% nucleotide homology in the exonic regions and 96% nucleotide homology in the intronic regions (3). Common *CYP21A2* mutations arise from gene recombination with *CYP21A1*, gross gene deletions, and point mutations (4). However, about 5% of CAH patients do not harbor mutations arising from recombination. From our own work over the past three decades, and from the work of others, over 100 *CYP21A2* mutations have been cataloged. In most cases, genotypes and phenotypes are correlated (5). However, in rare instances, genotype–phenotype discordance has been reported (6).

We present data demonstrating that a molecular characterization of *CYP21A2* mutations can be used to predict clinical phenotype and disease severity based upon changes it brings in the structure of the enzyme, steroid 21 hydroxylase. Previously, molecular models of the human CYP21A2 protein structure have been constructed using bacterial CYP102A1 (7) and rabbit CYP2C5 (8) as templates. Structural analyses of CYP21A2 mutants have been performed using these two templates (9–15). Other templates, such as the human CYP2C8, were also used as homology models of CYP21A2 (16).

Recently, the bovine CYP21A2 crystal structure has become available (17). We have, thus, constructed a humanized model of CYP21A2 and mapped disease-causing mutations to investigate the structural role of all known mutations in *CYP21A2*. Here, we provide full structural explanations for clinical manifestations observed in patients of CAH with varied mutations. We also highlight structural criteria that might allow the prediction of clinical severity through such *in silico* modeling.

Results

Humanized Model of CYP21A2. Bovine and the human sequences share 79% sequence identity. The human model of CYP21A2 was constructed by aligning bovine and human sequences and extracting structural features from the bovine template (17). An alignment of six cytochrome P450 proteins involved in steroid biosynthesis was also carried out to identify conserved structural elements (Fig. S1).

The derived human CYP21A2 model exhibits the characteristic triangular prism shape, containing 16 helices and 9 β -sheets (Fig. S2). All conserved structural motifs and elements are canonical to steroid-synthesizing cytochrome P450 proteins. A single heme prosthetic molecule is positioned centrally and is surrounded by B’-C loop, I-helix, K-K’ region, and a Cys-pocket. The structural elements that enclose heme in a closed binding cavity are the most conserved. There are 23 amino acid residues that lie

Author contributions: L.S., T.Y., M.Z., and M.I.N. designed research; S.H., B.I., V.D., M.S., C.P., L.S., and T.Y. performed research; S.H., B.I., V.D., M.S., C.P., and M.Z. analyzed data; and S.H., L.S., T.Y., M.Z., and M.I.N. wrote the paper.

The authors declare no conflict of interest.

¹M.Z. and M.I.N. contributed equally to this work.

²To whom correspondence should be addressed. E-mail: maria.new@mssm.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221133110/-DCSupplemental.

causing SW CAH. Similarly, the gate of the distal substrate-binding site is flanked by G64, which, when mutated to a glutamate with a longer side chain, prevents substrate entry causing SW CAH.

Impairment of structural stability. The loss of hydrophobic interactions of CYP21A2 impairs protein stability and the organization of its secondary structures, resulting in an inactive enzyme and SW CAH. Because residue Y59 is located in a hydrophobic pocket formed by residues from the A-helix and the β 1- and β 2-sheets, its mutation to asparagine disrupts the hydrophobicity of this region resulting in loss of function (Fig. 1B). Hydrophobicity is also disrupted when V139, a D-helix conserved residue, upon its mutation to glutamate, fails to interact with residues V440 and L436 on the L-helix. Furthermore, charge repulsions between the side chain of mutated V139E and E437 of the E-helix render the protein unstable and inactive causing SW CAH (Fig. 1B). Similarly, residue P386, located in a very hydrophobic pocket on the β 7-sheet, when mutated to the polar and bulkier residue arginine, will cause the disruption of hydrophobicity and protein instability. Disruption of π - π stacking among the interacting residues W302, F306, and F404 by their mutation into nonaromatic residues, such as in W302R/S and F404S, prevents stable packing interactions (Fig. 1B), resulting in SW CAH.

Protein inactivation can also occur when conformational flexibility is impaired, mostly because of the introduction of bulkier residues. For example, the sharp fold between the E- and F-helices, where G178 is located, can only accommodate the smallest residue. Even the mutation of glycine to alanine will hinder the flexibility of this fold. Similarly, G375 is present on the turn of β 5- β 6 hairpin. Its mutation to a more rigid residue, serine, will likewise impair flexibility of the fold, rendering the protein unstable. The mutation T450P likewise reduces flexibility of β 8-sheet, which helps stabilize the very long C-terminal loop. Structural elements in CYP21A2 are also maintained in position by a set of hydrogen bonds between the C-terminal residue Q481, Q318 (J-helix), and backbone carbonyl oxygens of L446 and Q447 (L-helix). When Q481 is mutated to proline, destabilization of the structure renders the protein inactive (Fig. 1B). Other substitutions where leucine is mutated to proline, such as in L142P (D-helix), L166P (E-helix), L167P (E-helix), L261P (H-helix), and L446P (L-helix), also cause helical disruption and destabilization of secondary structures. This is because the respective leucine residues are positioned in the middle of the helices.

Finally, the mutation of residues that form salt-bridge interactions and assist in maintaining tertiary structural elements of CYP21A2 also lead to SW CAH (Fig. 1C and Table S2). R366 (β 4-sheet) forms a salt bridge with D111 (B'-C loop), and its mutation to cysteine will disrupt this salt-bridge interaction, causing protein inactivation (Fig. 1C). In contrast, R354 hydrogen bonds with the side chains of E351 and R408 (E-R-R triad) and with the backbone atoms of P401 and W405 (K'-L loop). Mutation of R354 will, therefore, disrupt a network of hydrogen bond interactions and destabilize the protein (Fig. 1C). R408C/L mutations would likewise destabilize structural elements because of the extensive loss of hydrogen bonds.

Mutations Causing SV CAH. SV CAH is observed in 25% of classical CAH cases and can be mapped to two sets of structural aberrations that cause up to a 98% reduction of CYP21A2 activity: alterations in the transmembrane region and changes to conserved hydrophobic patches.

The membrane anchoring helix must enter the hydrophobic endoplasmic reticular membrane to allow for CYP21A2 to be active. Mutation of a semiconserved helix residue, A15, to a polar threonine prevents optimal anchoring, thus resulting in a gross reduction but not abrogation of enzyme activity and causing SV CAH (compare with SW CAH, above). Proximal to the anchoring helix, is residue H38, which lies in the loop between the N terminus and the A-helix and hydrogen bonds with Y47. When mutated to a nonpolar leucine, hydrogen bonds are disrupted resulting in SV CAH.

Hydrophobic patches regulate inter- and intraprotein interactions (22), and their disruption in several mutations causes a near loss of enzyme activity and SV CAH. First, the highly conserved residue I77 sits in a hydrophobic pocket formed with residues from L71 (β 2-sheets), M81 (B-helix), A391 (β 4-sheet-K'-helix), L388 (K'-helix), L419 and C423 (loop between K'-L helices). Its replacement by a polar residue, such as threonine, disrupts the hydrophobic environment (Fig. 2A). Second, the conserved residue C169 makes hydrophobic interactions with the loop between E-F helices and F-helix. Its mutation to arginine alters the region's hydrophobicity. Third, the mutation of I172, which forms a hydrophobic pocket with M186 and M187 residues of F-helix, to asparagine causes a loss of the hydrophobic pocket. Fourth, hydrophobicity is also lost when interactions between V281, present on the I-helix, and the H-helix are disrupted in the V281G mutation. The helix is also rendered unstable due to the high conformational flexibility imparted by glycine. Finally, L363 and A434 lie very close to heme and contribute to the hydrophobic environment of the heme-binding site (compare with SW CAH). Although mutations L363W and A434V conserve this hydrophobicity, they cause steric clashes with the heme because

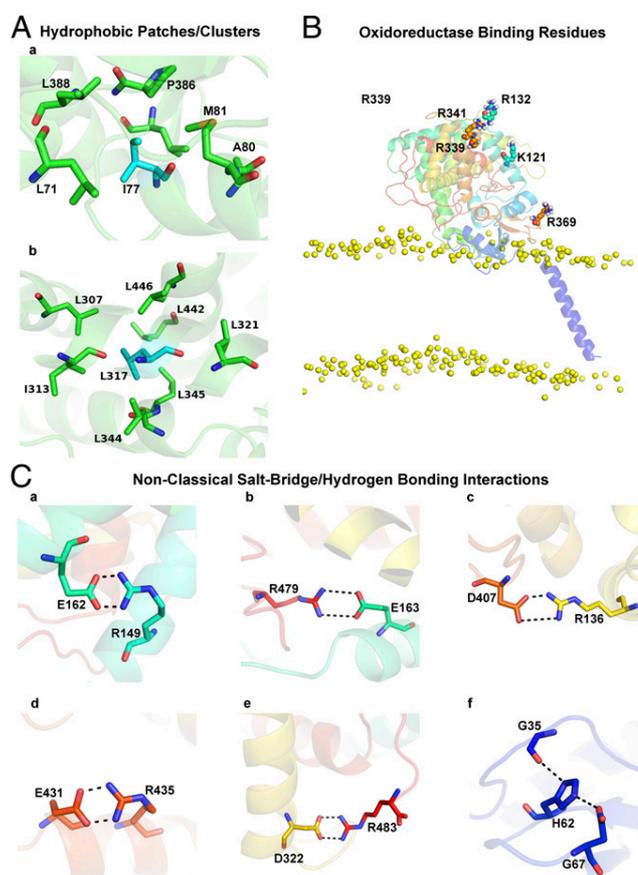


Fig. 2. Structural determinants of CYP21A2 mutations causing SV and NC CAH. (A) I77 and I317, which are surrounded by hydrophobic residues. Mutations of I77 or I317 (cyan sticks) result in disruption of hydrophobicity. Whereas mutation I77T results in SV CAH, L317M/V causes NC CAH (because valine and methionine are both hydrophobic residues that can be tolerated at this position). (B) Positively charged residues that have been implicated in the binding of oxidoreductase. Mutation of these residues causes NC CAH. All of these residues are present on the same face of CYP21A2. (C) Disruption of salt-bridge and hydrogen bonding interactions that result in localized destabilization of structure causes NC CAH. Salt-bridge interactions between R149-E162 (Ca), R479-E163 (Cb), R136-D407 (Cc), R435-E431 (Cd), R483-D322 (Ce), and hydrogen-bonding network between G35-H62 and G67 (Cf).

of replacement by a bulkier residue. Thus, a near-complete rather than a complete loss of enzyme activity ensues.

Other mutations that alter the charge on and/or the size of individual residues can result in steric clashes and conformational instability within certain parts of the CYP21A2 molecule. These relatively localized alterations can cause grossly reduced enzyme activity and SV CAH. For example, E320 is a highly conserved residue on a negatively charged Glu-Glu-Leu-Asp (EELD) motif (22, 23). In the E320K mutation, a change from acidic to basic residue alters the charge on the motif. Likewise, replacement of R408 with histidine, when the latter is not protonated, prevents normal hydrogen bonding with E351 and R354. Another residue P432 lies adjacent to the loop that contains invariant C428, which binds heme and P432 to optimally position the loop. Its mutation to leucine makes the structure more flexible and prevents cysteine from being presented to heme. In contrast, the mutation of G424 to serine imparts rigidity to the loop between K' and L-helix. Examples of steric hindrance include the L300F and L308F mutations. Both residues are present in a very constricted space on I-helix, so that their mutation to a bulky phenylalanine causes localized destabilization of secondary structure. Finally, P463 mutation to leucine interferes with the conformation of the $\beta 8$ - $\beta 9$ loop with the subsequent closure of substrate entrance channel.

Mutations Causing Nonclassical CAH. The following mutation-induced aberrations in protein structure significantly reduce CYP21A2 activity (2).

Oxidoreductase Interactions. The oxidation of steroidogenic substrates by CYP21A2 requires delivery of two electrons from P450 oxidoreductase (POR). Both charged and hydrophobic residues on the proximal surface of cytochrome P450 are involved in interaction with POR (24). The basic residues on the surface of CYP21A2, namely K121, R132, R341, R339, and R369, are known to interact with the acidic residues on POR (8, 17). All of these positively charged residues are also aligned to one face of the protein (Fig. 2B). When these residues are mutated (K121Q, R132C, R339H, R341W/P, and R369Q), the positively charged side chain of the interacting residue is lost. The strength of interaction with POR is affected, which results in mild, NC CAH.

Salt-bridge and hydrogen-bonding networks. CYP21A2 has many hydrogen bonds and salt bridges that stabilize the overall tertiary structure of the protein. Mutations of nonconserved residues participating in these interactions mainly cause NC CAH (Fig. 2C). For example, H62 hydrogen bonds with the backbone of G35 (adjacent to the N-terminal loop) and side chain of G67 ($\beta 2$ -sheet). Substitution of histidine by hydrophobic leucine (H62L) may disrupt hydrogen bonding to reduce, but not eliminate, enzyme activity. Likewise, R149C, D407N, E431K, R435C, D322G, R479L, and R483Q/P mutations correspond to changes in residues that contribute in salt-bridge formation, including R149 (D-helix)–E162 (E-helix), R316 (J-helix)–D407 (K'-helix), E431 (L-helix)–R435 (L-helix), R479 ($\beta 9$)–E163 (E-helix), and R483 (C-terminal)–D322 (J-helix). When these residues are mutated, salt-bridge formation will be prevented resulting in a localized, as opposed to global, destabilization of tertiary structure (Table S2). Additionally, the introduction of a bulkier side chain than the native residue causes NC CAH. Thus, the mutations H119R, R124H, T168N, I194N, I230T, V249A, A265V, S301Y, A265C, V304M, and N387K show side-chain steric clashes with the neighboring residues.

Nonconserved hydrophobic clusters. The introduction of a single polar residue in a hydrophobic cluster will obliterate its hydrophobicity and destabilize the local environment. For example, I171 is located in a hydrophobic region, and its mutation to asparagine will disrupt the hydrophobic cluster. Similarly, P459 sits in a hydrophobic pocket, surrounded by V305 and L308 (I-helix), L458 ($\beta 8$ - $\beta 9$ loop), and M473 and P475 ($\beta 8$ - $\beta 9$ loop). P453 is likewise positioned in a hydrophobic pocket surrounded by L308 (I-helix), L452 ($\beta 8$ -sheet), L458 ($\beta 8$ - $\beta 9$ loop). Furthermore, A391 sits in a hydrophobic pocket, surrounded by L388, L419, and I77. Mutations A391T, P453S, and P459H will all disrupt

hydrophobicity of their respective regions. In contrast, L317 is present in a region with tight packing in a hydrophobic pocket surrounded by a leucine cluster formed by L307 (I-helix); I313 (J-helix); L321 (loop between J-K helices); and L344, L345, L442, and L446 (K-helix). Its mutation to a methionine or valine disrupts the optimal packing of side chains but does not alter the hydrophobic environment.

Other mild mutations. Certain mutations result in mild disease, because the mutated residues are either reasonably well tolerated or the disrupted interactions are compensated by other residues. These mutations, therefore, variably reduce, but do not grossly lower or eliminate, enzyme activity. For example, P30 is present near the transmembrane region and, along with P31 and P34, forms the hinge to orient protein away from membrane. Its mutation to leucine, a hydrophobic residue, is tolerated and only somewhat reduces enzyme activity. Residue V281 is present on the I-helix in a very constricted space. When mutated to leucine, the increase in chain length results in steric clashes that are relatively modest and, thus, cause mild NC CAH. Although the mutation Y47C disables hydrogen bonding with H38, the interaction is compensated by a weak His-Cys interaction, similar to that observed in glutamine amidotransferases (25). Mutation R224W is mild because tryptophan can still interact with the phosphate head group through its nitrogen atom (26). Although another mutation R233G/K may prevent R233 from binding to the 3-keto oxygen of the proximal 17OHP in the proper orientation, it does not influence protein activity significantly, resulting in minimal phenotype.

Structural Implications of Previously Uncharacterized Mutations.

In Fig. 3, we document structural aberrations in previously uncharacterized mutations on the basis of changes that each mutation brings to the humanized model of CYP21A2. The classification of these mutations is based on an exhaustive sequence and structural analysis of all CAH-causing mutations.

Discussion

Although newborn screening for CAH is becoming used increasingly, difficulties regarding prediction of clinical phenotype remain (1). We have used computational modeling to perform an exhaustive sequence–structure–clinical outcome analysis of known and previously unreported missense CYP21A2 mutations as they relate to the phenotype and severity of CAH. We show that this *in silico* approach can predict the extent of loss of function of the steroidogenic CYP21A2 enzyme. With certain mutations, varying degrees of loss of function have been confirmed biochemically by overexpressing the mutated enzyme *in vitro* (27–32), whereas, in other cases, the extent to which the enzyme is disrupted remains unknown.

We find that enzyme activity is completely lost *in silico*, resulting in SW CAH when mutations alter the heme-binding or substrate-binding pocket, disrupt membrane anchoring of the enzyme, or cause gross structural instability by interfering with protein scaffolding. This accounts for ~35% of residues around the heme- and ligand-binding sites. However, mutations that occur near the transmembrane domain or in conserved hydrophobic patches other than the heme- and substrate-binding region result in a near-complete loss of enzyme activity, with less severe, SV disease. In contrast, mutations disrupting nonconserved hydrophobic clusters, salt-bridge interactions, hydrogen bonding, or oxidoreductase interactions or those that are tolerated sterically or compensated by other residues, produce milder, NC disease. Of these, eight are present on the surface, whereas the rest lie within the core of the protein. We, thus, infer that there is overall concordance between the extent and type of structural disruption of CYP21A2 and the clinical phenotype and severity of CAH.

There are, however, several exceptions to this concordance. First, single-point mutations to polar or nonpolar residues can show distinct clinical presentations. For example, R356 (K-helix) is located near the heme-binding site and interacts with Q389 (K'-helix) and Q462 ($\beta 8$ - $\beta 9$) to maintain secondary structural

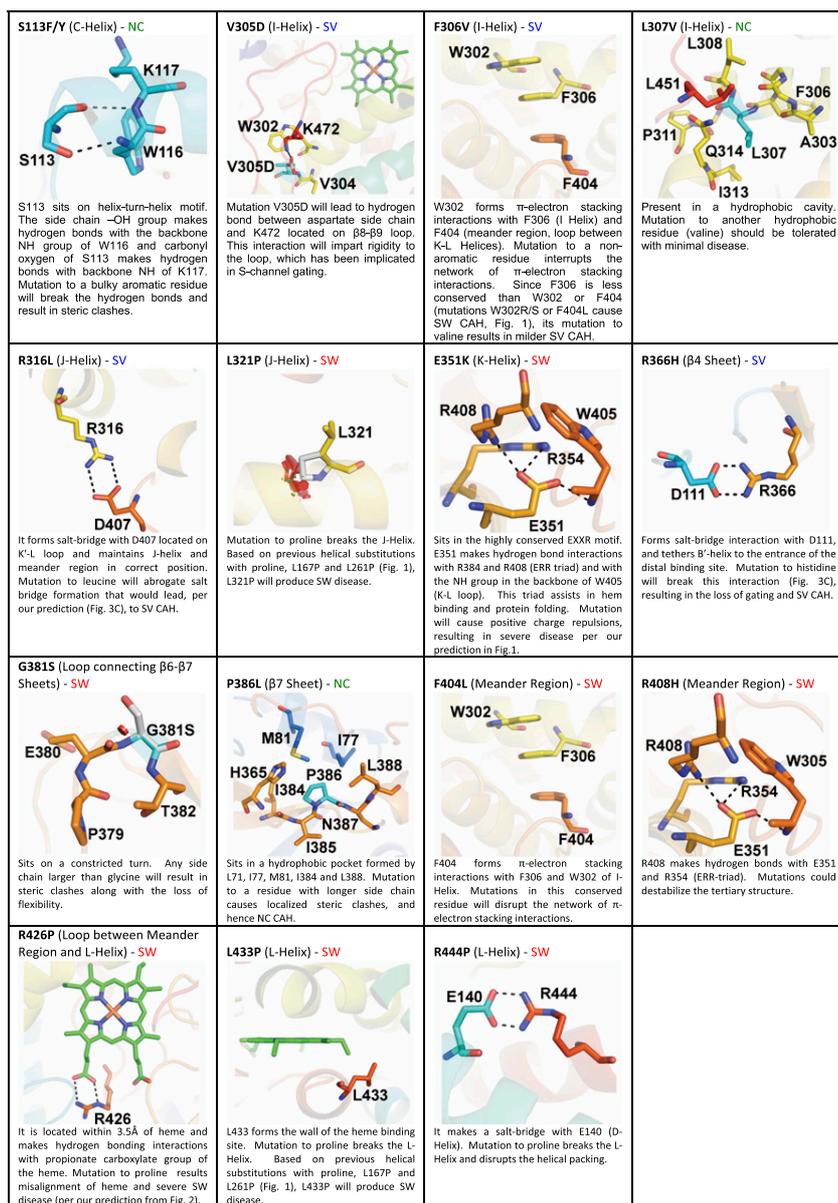


Fig. 3. Previously unreported CYP21A2 mutations. Previously unreported missense mutations, namely S113Y/F, V305D, F306V, L307V, R316L, L321P, E351K, R366H, G381S, P386L, F404L, R408H, R426P, L433P, and R444P are shown. The location of the mutated residue and the impact each mutation has on protein structure and the predicted clinical outcomes are shown.

elements. Its mutation to hydrophobic tyrosine or proline leads to SW CAH, because hydrogen bonding is completely lost. However, when R356 mutates to glutamine, partial hydrogen bonding may still be preserved, so that the R356Q mutation exhibits SV CAH. Similarly R408C/L mutation causes SW CAH, whereas R408H leads to SV CAH. Likewise, because R408 makes extensive hydrogen-bonding network with E351 and R354 (E-R-R triad), different substitutions may yield distinct clinical phenotypes.

Second, although the analysis presented in this study is based on individual point mutations, compound mutations may produce severe SW disease, even if the individual mutation is relatively mild. For example, the three mutations I236N, V237E, and M239K in *CYP21A2* gene, which arise from homologous recombination between the *CYP21A1* pseudogene and *CYP21A2*, occur as a cluster in the same patient. When individually mutated to asparagine, I236 does not significantly disrupt structure. Similarly, although the M239K mutation causes charge repulsions between K239 and H235 side chains, it does not disrupt the structure significantly. However, the V237E is highly disruptive:

it modifies the hydrophobic region and alters conformation of the substrate-binding site, resulting in loss of function of the enzyme. Thus, compound mutations of these residues, namely I236N, V237E, and M239K, cause SW CAH.

Finally, because CAH is a recessive disorder, phenotype can vary between homozygous and heterozygous mutations. L166 is present in the middle of the E-helix. Similarly to L167P mutation, L166P can lead to SW, because proline in the middle of the helix destabilizes the structure. However, L166P has been observed as NC CAH in a patient with V281L mutation on the other allele. The clinical outcome in this case may be attributable to the V281L mutation, which is, in itself, reported to exhibit NC CAH phenotype. Thus, although our analysis predicts the severity of a given mutation in CAH, variations can be observed when compound and heterozygous mutations are under consideration.

Even with the above caveats, the results of our analysis of 113 missense mutations in terms of changes they bring to the structure of CYP21A2 is compatible with our observations in clinical practice. Indeed, many of the phenotypes of the mutation have been well ascertained by experienced clinicians.

Monogenic disorders, such as CAH, Gaucher, and many others, are often characterized by diverse clinical phenotypes, toward which there are oftentimes no ready explanations. Particularly with the advent of tools through which fetal DNA from maternal plasma can be sequenced, it is now possible to obtain genotype information very early during gestation. This *in silico* computational approach may help clinicians to attribute a phenotype, hence facilitating prenatal therapeutic decisions.

Methods

The amino acid sequence for human CYP21A2 was taken from UniProt (accession no. P08686). The sequence from the bovine CYP21A2 crystal structure [Protein Data Bank ID code 3QZ1 (17)] was extracted and aligned with the human sequence using ClustalX (33) (overall sequence similarity, 79%). The sequence conservation in the heme-binding cavity, between human and bovine, is 95%. The alignment revealed that the initial 28 residues at the N-terminal did not have a corresponding structural template and were, therefore, built separately.

Models of human CYP21A2 were generated using MODELLER version 9.10 (34, 35), and stereochemical properties evaluated using the program PROCHECK version 3.5.4 (36). The final model was chosen based on low-energy function and a low $C\alpha$ rmsd overlap between the bovine template and the humanized model. The model was constructed with the spatial position of heme and ligands retained in their respective binding sites. This was based on the assumption that the ligands and cofactor interact in

a similar manner in the human as in the bovine crystal structure; this comparison replicates all conserved interactions from the template into the model. Several rounds of energy minimization were carried out to obtain a final low-energy conformation with no steric clashes between side chains. The final $C\alpha$ rmsd between the bovine template and the humanized model was 0.22 Å (Fig. S4).

The first 28 residues corresponding to the N-terminal region including the membrane-anchoring α -helix were constructed after secondary structure prediction using PSIPRED (37). Residues 1–25 were modeled in a canonical α -helical conformation, whereas residues 26–28, which connect the anchoring transmembrane helix to the first residue in the model, were built in a random-coil conformation using MODELLER. The orientation and position of the model, with respect to the lipid bilayer, was obtained using the membrane self-assembly coarse-grained dynamics protocol by Cojocaru et al. (38). This permits unbiased orientation of the protein in the membrane, which is not achievable via atomistic simulations. At the end of 1 μ s, the coarse-grained model was converted to atomistic using the reverse-conversion protocol (39) and further subjected to restraint-free molecular-dynamics simulations to test the stability of the model (SI Methods). The molecular graphics images were generated using PyMOL (www.pymol.org).

Mutations were identified from the Human Cytochrome P450 Allele Nomenclature database (www.cypalleles.ki.se). Redundancies were removed, and clinical conditions were cross-evaluated with respect to the original publications (Table S1).

- New MI, Wilson RC (1999) Steroid disorders in children: Congenital adrenal hyperplasia and apparent mineralocorticoid excess. *Proc Natl Acad Sci USA* 96(22):12790–12797.
- New MI (2006) Extensive clinical experience: Nonclassical 21-hydroxylase deficiency. *J Clin Endocrinol Metab* 91(11):4205–4214.
- White PC, New MI, Dupont B (1986) Structure of human steroid 21-hydroxylase genes. *Proc Natl Acad Sci USA* 83(14):5111–5115.
- Wajnrajch MP, New MI (2010) Defects of Adrenal Steroidogenesis. *Endocrinology*, eds Jameson JL, De Groot LJ (Elsevier Philadelphia PA), 6th Ed, Vol 2, pp 1897–1920.
- Wilson RC, et al. (2007) Ethnic-specific distribution of mutations in 716 patients with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency. *Mol Genet Metab* 90(4):414–421.
- New MI, Mercado AB, Cheng KC, Jackowski M, Wilson RC (1996) Steroid 21-hydroxylase deficiency: Genotype may not predict phenotype. Where phenotype does not match genotype. *Frontiers in Endocrinology (Ares-Serono Symposia Series)* 16:17–27.
- Mornet E, Gibrat JF (2000) A 3D model of human P450c21: Study of the putative effects of steroid 21-hydroxylase gene mutations. *Hum Genet* 106(3):330–339.
- Robins T, Carlsson J, Sunnerhagen M, Wedell A, Persson B (2006) Molecular model of human CYP21 based on mammalian CYP2C5: Structural features correlate with clinical severity of mutations causing congenital adrenal hyperplasia. *Mol Endocrinol* 20(11):2946–2964.
- Krone N, Riepe FG, Grötzingler J, Partsch CJ, Sippell WG (2005) Functional characterization of two novel point mutations in the CYP21 gene causing simple virilizing forms of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *J Clin Endocrinol Metab* 90(11):445–454.
- Grischuk Y, et al. (2006) Four novel missense mutations in the CYP21A2 gene detected in Russian patients suffering from the classical form of congenital adrenal hyperplasia: Identification, functional characterization, and structural analysis. *J Clin Endocrinol Metab* 91(12):4976–4980.
- Baradaran-Heravi A, et al. (2007) Three novel CYP21A2 mutations and their protein modelling in patients with classical 21-hydroxylase deficiency from northeastern Iran. *Clin Endocrinol (Oxf)* 67(3):335–341.
- Riepe FG, et al. (2008) Functional and structural consequences of a novel point mutation in the CYP21A2 gene causing congenital adrenal hyperplasia: Potential relevance of helix C for P450 oxidoreductase-21-hydroxylase interaction. *J Clin Endocrinol Metab* 93(7):2891–2895.
- Bleicken C, et al. (2009) Functional characterization of three CYP21A2 sequence variants (p.A265V, p.W302S, p.D322G) employing a yeast co-expression system. *Hum Mutat* 30(2):E443–E450.
- Tardy V, et al. (2010) Phenotype-genotype correlations of 13 rare CYP21A2 mutations detected in 46 patients affected with 21-hydroxylase deficiency and in one carrier. *J Clin Endocrinol Metab* 95(3):1288–1300.
- Minutolo C, et al. (2011) Structure-based analysis of five novel disease-causing mutations in 21-hydroxylase-deficient patients. *PLoS ONE* 6(1):e15899.
- Bojunga J, et al. (2005) Structural and functional analysis of a novel mutation of CYP21B in a heterozygote carrier of 21-hydroxylase deficiency. *Hum Genet* 117(6):558–564.
- Zhao B, et al. (2012) Three-dimensional structure of steroid 21-hydroxylase (cytochrome P450 21A2) with two substrates reveals locations of disease-associated variants. *J Biol Chem* 287(13):10613–10622.
- Kemper B (2004) Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. *Toxicol Appl Pharmacol* 199(3):305–315.
- Xu J, Voth GA (2008) Redox-coupled proton pumping in cytochrome c oxidase: Further insights from computer simulation. *Biochim Biophys Acta* 1777(2):196–201.
- Nagano S, Poulos TL (2005) Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam. Implications for the dioxygen activation mechanism. *J Biol Chem* 280(36):31659–31663.
- Haider SM, Patel JS, Poojari CS, Neidle S (2010) Molecular modeling on inhibitor complexes and active-site dynamics of cytochrome P450 C17, a target for prostate cancer therapy. *J Mol Biol* 400(5):1078–1098.
- Gong XS, Wen JQ, Gray JC (2000) The role of amino-acid residues in the hydrophobic patch surrounding the haem group of cytochrome f in the interaction with plastocyanin. *Eur J Biochem* 267(6):1732–1742.
- Gilmore SR, Gräfenhan T, Louis-Seize G, Seifert KA (2009) Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fusarium*. *Mol Ecol Resour* 9(Suppl s1):90–98.
- Kenaan C, Zhang H, Shea EV, Hollenberg PF (2011) Uncovering the role of hydrophobic residues in cytochrome P450-cytochrome P450 reductase interactions. *Biochemistry* 50(19):3957–3967.
- Mei B, Zalkin H (1989) A cysteine-histidine-aspartate catalytic triad is involved in glutamine amide transfer function in purF-type glutamine amidotransferases. *J Biol Chem* 264(28):16613–16619.
- Hall BA, Armitage JP, Sansom MS (2011) Transmembrane helix dynamics of bacterial chemoreceptors supports a piston model of signalling. *PLOS Comput Biol* 7(10):e1002204.
- Tusie-Luna MT, Traktman P, White PC (1990) Determination of functional effects of mutations in the steroid 21-hydroxylase gene (CYP21) using recombinant vaccinia virus. *J Biol Chem* 265(34):20916–20922.
- Tusie-Luna MT, Speiser PW, Dumic M, New MI, White PC (1991) A mutation (Pro-30 to Leu) in CYP21 represents a potential nonclassical steroid 21-hydroxylase deficiency allele. *Mol Endocrinol* 5(5):685–692.
- Chiou SH, Hu MC, Chung BC (1990) A missense mutation at Ile172—Asn or Arg356—Trp causes steroid 21-hydroxylase deficiency. *J Biol Chem* 265(6):3549–3552.
- Robins T, Barbaro M, Lajic S, Wedell A (2005) Not all amino acid substitutions of the common cluster E6 mutation in CYP21 cause congenital adrenal hyperplasia. *J Clin Endocrinol Metab* 90(4):2148–2153.
- Ohlsson G, Müller J, Skakkebaek NE, Schwartz M (1999) Steroid 21-hydroxylase deficiency: Mutational spectrum in Denmark, three novel mutations, and in vitro expression analysis. *Hum Mutat* 13(6):482–486.
- Lajić S, et al. (2002) Novel mutations in CYP21 detected in individuals with hyperandrogenism. *J Clin Endocrinol Metab* 87(6):2824–2829.
- Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2:Unit 2.3.
- Sánchez R, Sali A (2000) Comparative protein structure modeling. Introduction and practical examples with modeller. *Methods Mol Biol* 143:97–129.
- Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.
- Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12(4):345–364.
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16(4):404–405.
- Cojocaru V, Balali-Mood K, Sansom MS, Wade RC (2011) Structure and dynamics of the membrane-bound cytochrome P450 2C9. *PLOS Comput Biol* 7(8):e1002152.
- Stansfeld PJ, Sansom MS (2011) From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. *J Chem Theory Comput* 7(4):1157–1166.

COMPUTATIONAL METHODS FOR STUDYING G-QUADRUPLEX NUCLEIC ACIDS

B. Islam¹, V. D'Atri¹, M. Sgobba¹, J. Husby² and S. Haider^{1*}

¹ Centre for Cancer Research and Cell Biology, Queen's University of Belfast, Belfast BT9 7BL, UK

² UCL School of Pharmacy, Brunswick Square, London WC1N 1AX

*Corresponding Author, Email: s.haider@qub.ac.uk

1 INTRODUCTION

Guanine repeat nucleic acid sequences in the presence of cations can self-associate to form a four-stranded arrangement called G-quadruplex¹. It has garnered interest in the past decade due to abundance of potential quadruplex structures in several biologically relevant sequences such as telomere, immunoglobulin switch regions, promoter regions and other disease associated regions². G-quadruplex forming regions have also been explored as promising anti-cancer targets³. Along with the experimental techniques, computational methods including molecular dynamics have been used to appreciate the biological relevance of quadruplexes^{4,5}. A quadruplex is stabilized via cyclic arrangement of eight Hoogsteen hydrogen bonds between four planar guanine bases to form a G-quartet (or G-tetrad)⁵. Several quartets can stack over one another to form the stem of the quadruplex. The cations are located in the central helical cavity and forms coordination with carbonyl oxygen from guanines. Quadruplexes are highly diverse with regard to the origin, the cations involved in the stabilization, orientation of the strands and the glycosidic conformation of the bases. They can be monomeric (single strand), dimeric (two strands) and tetrameric (four strands) in origin. The monomeric quadruplex are formed in strands with general sequence $G_aXG_bYG_cZG_d$, where a, b, c and d are in the range 3-5. X, Y and Z represent the loop sequences and comprise of 1-8 non-guanine nucleotides. Dimeric quadruplexes are formed in two separate strands of sequence G_aXG_b . A quadruplex is called tetramolecular when four separate strands interact to form the quadruplex structure.

2 ROLE OF CATIONS

The formation of quadruplex requires coordination of cations for its stability⁶. The experimental observations of quadruplex crystal structures confirm that cations are an integral part of all quadruplexes⁶. The quartets undergo hydrophobic stacking to form a central channel contributed by carbonyl oxygen of the four guanines. The cations interact with the carbonyl oxygen by electrostatic forces and neutralize the repulsion between them. The role of K^+ , Na^+ , Rb^+ , Cs^+ , NH_4^+ , Tl^+ , Sr^{2+} , Ba^{2+} and

Pb^{2+} ions in stabilizing the quadruplex has been elucidated⁶. Ca^{2+} and Mg^{2+} ions do not promote the formation of G-quadruplex structures. K^+ is the most stabilizing cation and coordinates with O6 atoms of guanine bases in bipyramidal antiprismatic geometry. The ionic radius of K^+ ions (1.33Å) is too large to coordinate in the plane of G-quartet. Therefore, they occupy interstitial space in between two quartets. The adjacent K^+ ions in the ion channel are separated by an average distance of 3.38Å⁷. The order of stability provided by cation is $\text{K}^+ > \text{Rb}^+ > \text{Na}^+ > \text{Li}^+ = \text{Cs}^+$ and $\text{Sr}^{2+} > \text{Ba}^{2+} > \text{Ca}^{2+} > \text{Mg}^{2+}$ ⁸. In addition to their role in stabilizing quadruplexes, cations also play an important role in determining their structure. Quadruplexes with multiple cation coordination geometries have also been reported⁹. Quadruplexes with K^+ ions have larger G-quartets and hence provide greater loop flexibility than with those with Na^+ ions¹⁰. K^+ ions favour formation of anti-parallel quadruplex and the interconversion between parallel and anti-parallel structures can be controlled by Na^+/K^+ balance¹¹. The cations stabilize as well as direct the structural conformation of thrombin binding aptamer (TBA)¹². Spontaneous exchange of cation between the quadruplex structure and bulk solvent during the course of simulation has also been evidenced. The exchange is however very rapid so as to minimize any destabilization to quadruplex structure¹².

3 MOLECULAR DYNAMICS SIMULATIONS: CONSIDERATIONS AND LIMITATIONS

The interactions of G-quadruplexes have been widely studied by Molecular Dynamics (MD) carried out in explicit solvent^{4, 5, 12, 13}. MD simulation consists of solution of the classical equations of motion, which for the force on a simple atomistic system may be represented as $F=ma$. The force on an individual atom is contributed by its bonded and non-bonded terms. The interatomic parameters like bond length, bond angle, dihedral angles, coulombic and van der Waals interactions all determine the effective force on an atom. The acceleration term in the equation is used to determine the velocity and position of each atom at a course of time. A trajectory is then calculated by tracing the course of position vectors as a function of time. The determination of suitable potential function is extremely important for a MD simulation¹⁴. It is assumed that atoms attract each other when they are at long distance while they repel at short interatomic distance. This is represented by Lennard-Jones potential. MD simulations are usually carried out in canonical (NVT) or Isobaric-isothermal (NPT) ensemble.

3.1 Force Fields

Force fields in molecular mechanics are used to include all the quantum mechanical parameters that describe the molecular behaviour. The total energy of a molecule is calculated by taking into account all atom-atom contributions such as Coulombic, polarization, dispersion and repulsive energies¹⁵. There is no “universal force-field” and preference of one force field over the other may vary based on the target application. Accurate parameterization requires set of input data and significant physical approximations to describe the intermolecular interactions. Atomistic molecular dynamics (MD) computer simulations for B-DNA have been successfully carried for several years and show good agreement with the experimental results¹⁶. Haider *et al.* performed a 20ns simulation of quadruplex DNA using GROMOS force

field. The four-stranded structure was lost within the first 10ns of the simulation¹⁷. The CHARMM27 force field when used on folded RNA structures reveals unstable trajectories¹⁸. However, Cornell parameters that are normally associated with AMBER have been successfully used with CHARMM suite of programs for hammerhead ribozyme and guanine riboswitch¹⁸. MD simulations of G-quadruplex loops using Cornell force field, parm94 and parm98/99 on a nanosecond-scale are reported to show significant variations although they work properly for G-stems¹⁹. An α/γ torsional term was re-parameterized and presented in parmbsc0 force field to resolve various inaccuracies of previous force fields²⁰. The benchmarking exercise of MD simulations of G-quadruplex for 1.5 μ s in parmbsc0 force fields shows reasonable agreement with the experimental values. The force fields were tested on the d(G₄T₄G₄)₂ dimeric quadruplex with diagonal loops and also the parallel stranded human telomeric monomolecular quadruplex d[AG₃(TTAG₃)₃] with three propeller loops²¹. The parmbsc0 force field have also been used for 10ns simulation for free-energy analysis of G-quadruplex DNA. The estimates based on MD simulations are in close accord with the experimental results proving the validity of force field for quadruplex analysis¹³. A combination of parm99 and parmbsc0 has also been used for a longer microsecond scale simulation of quadruplex²¹.

The binding of ligands to G-quadruplex has also been explored by MD using different force fields of AMBER. AMBER force field ff03 was used for quadruplex, water and ions while parameters for the ligands were assigned using generalised AMBER force field (GAFF) method for MD simulations of G-quadruplex with ligands carried for 10-30ns. The backbone of the quadruplex was reasonably stable but the flanking nucleotides show fluctuations during the course of simulation due to the limitations of the force fields²². MD simulations of Azatrux and parallel quadruplex DNA using parmbsc0 force field has been carried out for 15ns. The additional parameters for the Azatrux were assigned using AMBER-GAFF method. The experimental results corroborated with the molecular dynamics data implying that parmbsc0 force field is suitable for MD simulations of quadruplex-ligand complex along with the GAFF method for ligand parameterisation²³. It has to be mentioned that quantitative derivations like free-energy calculations should be avoided using ligand force-fields. MD simulations of ligand-quadruplex interactions can be best used to augment the experimental predictions or interpretations.

3.2 Long-range Electrostatic Interactions

It is computationally ineffective to sum all the non-bonded interactions in a MD simulation. Therefore, "cutoff" procedures are employed to neglect long-range electrostatic interactions beyond some distance without altering the quality of simulations. The long range electrostatic forces play an important role in stability of quadruplex DNA as they represent the quadruplex-cation interactions. Thus, MD simulations of quadruplex DNA structures require accurate treatment of electrostatic interactions for achieving stable trajectories. Initial report of quadruplex analysis used spherical cutoff methods for handling electrostatic interactions²⁴. The trajectories of these simulations were unstable and cations from the quadruplex were also dislodged during the course of simulation eventually leading to collapse of structure²⁴. With the introduction of Particle Mesh Ewald (PME) truncation method a stable DNA simulation is achievable²⁵. The computational time is longer for the PME method than for the atom-based truncation method due to the implicit

requirement of periodic boundary conditions for Ewald summation. In spite of time consumption factor, PME is universally used for accurate and stable simulations of quadruplex DNA²⁶. Quadruplex MD simulations with PME summation method and 10 Å cut-off have yielded fairly stable trajectories⁵.

3.3 Inter-base Interactions and Backbone Descriptions

The force field parameters of nucleic acid molecules are also influenced by interactions between the bases. The amino groups of nucleobases tend to be non-planar due to a partial sp^3 pyramidal hybridization²⁷. This affects the stabilization of bifurcated H-bonds, close amino group contacts, non-planar G/A base pairs and some other specific interactions²⁷. The force fields assume purely sp^2 amino nitrogen²⁷. This should be sufficient for most interactions, as primary H-bonds stabilize the sp^2 electronic structure. However, the force fields support neither out-of-plane H-bonds nor amino-acceptor interactions. Not surprisingly, neither bifurcated H-bonds nor amino-group contacts are properly reproduced by contemporary MD simulations.

Assignment of charge on a base for quantum mechanics is purely hypothetical as there is no quantum mechanical operator for atomic charges²⁷. Therefore, treatment of molecular interactions based on charge distributions like hydrogen bonding should be done with care. The arbitrary charges on the bases can be assigned by fitting to molecular electrostatic potential (MEP, ESP charges). The popular AMBER fields use this method for the assignment of charges for calculation of electronic interactions²⁸. The backbone of nucleic acid is a fairly difficult molecule to deal for MD simulations due to two major reasons. Firstly, the backbone is a highly flexible structure and requires geometry-dependent charges rather than constant point charges. Secondly, the highly negative charge on the nucleic acid changes with conformation and solvation dynamics. This is not handled by the non-polarizable atom-atom pair additive force fields. Failure of parm94 and parm98/99 AMBER force fields for simulation of DNA is reported to be primarily due to accumulation of α/γ substrates resulting in distortions of the double helix²⁹⁻³³. An improvised force-field parmbsc0 has correction term for α/γ DNA backbone³⁴. Recent simulations on nanoscale and microscale report that G-stems remain stable while the loops may show slight but acceptable fluctuations over the entire course of simulation^{20, 35}. Parmbsc0 is therefore preferred force field for the simulation of quadruplex DNA.

4 SIMULATION-BASED METHODS

4.1 Classical Molecular Dynamics

With the recent advances in force field parameterization and PME technique for treatment of long range electrostatic interactions, Classical MD has been the most common method to investigate the conformational flexibility and dynamic behaviour of quadruplex structures. Initial MD of G-quadruplex employed AMBER4.1 force fields for simulations³⁶. Parallel and Antiparallel quadruplex formed by the sequences d(G₄) and d(G₄T₄G₄) respectively, with Na⁺ reveal exceptionally stable trajectories over 25ns of simulation. The study revealed that G-stems can exchange cations from the solvent without significant perturbation. Complete removal of cations from the central ion channel of quadruplex has also been observed in this simulation. It may be due to inappropriate force field calculations of ion-guanine

interactions. The removal of cations from the quadruples results in immediate collapse underscoring their importance in maintaining the stability of quadruplex³⁶. During the simulation of parallel quadruplex, deformation of inner quartets has also been reported. This implicates that radius of ion is too large to optimally describe the ion interactions inside the stem. The trajectories associated with loops show more fluctuations compared to the G-stem³⁶. Recently Classical MD has also been used to study the cation binding to Thrombin Binding Aptamer (TBA) quadruplex structure using parmbsc0 force field. The study reveals that cations are sucked into the quadruplex structure for its stabilization. The penetration of cation and binding events are modulated by the loop structure and depends on the type of cation. The spontaneous exchange of cations between the solvent and quadruplex molecule at the atomistic resolution is also evidenced in this study¹².

The presentation of crystal structure of quadruplex from human telomeric DNA attracted many molecular dynamics studies to provide an insight in this assembly^{7, 37}. Classical MD simulations have also been used to investigate the ligand binding to telomeric quadruplex DNA^{22, 38-41}. The stability of quadruplex increases with the number of quartets in the structure. An empty pseudo-intercalation ligand binding site created between two quadruplex units (distance >7Å) is not tolerated and the units separate within the first 2 ns. However, the presence of acridine in this binding site is sufficient to stabilise the quadruplex and maintain its topology⁵. MD for investigation of porphyrin binding with two antiparallel human telomeric quadruplex (d[(GGGTTA)₃GGG]) in the presence of K⁺ show stable trajectories. The ligand binds to quadruplex in 1:1 stoichiometry and shows binding modes of end stacking, intercalation, external groove binding, and external loop stacking⁴². MD has also been used to study interaction of quadruplex with selective stabilizing agents based on bisquinolinium and bispyridinium derivatives of 1,8-naphthyridine. The study reveals that derivatives with low binding free energy bind by end-stacking binding mode⁴³. MD simulations show that telomeric G-quadruplex interaction with a single molecule of 4,11-bis[(2-ethyl)amino]anthra[2,3-b]thiophene-5,10-dione disrupt the G-quartets even in the presence of Na⁺ and K⁺ ions. Also, despite the loss of G-quartet the sugar-phosphate backbone of quadruplex remained stable⁴⁴.

The conformational variability of quadruplexes is enhanced by intercalated four-stranded DNA structures called i-motifs⁴⁵. The DNA-tetrameric structure of i-motif is formed of two parallel duplexes that are stabilised by hemi-protonated cytosine/cytosine⁺ (C/CH⁺) base pairs that intercalate into each other in a head-to-tail orientation³⁶. Each base pair carries a +1 charge, which is distributed over the two cytosines. The MD simulation of i-motif on a nanosecond scale was stable. The attractive interactions between C4–N4 and C2–O2 dipoles of stacked C/C⁺ pairs and between the N3 and H3⁺ atoms of the cytosines forming a hemiprotonated C/C⁺ pair can thus compensate for the C-imino proton repulsion⁴⁶. Also, solvent screening that is modulated by the over-all topology of i-motif structure counterbalances the base pair repulsion. Unlike other quadruplex structures, i-motifs have intrinsic repulsive energy stacking terms and the stability in simulations is due to common electrostatics. The AMBER force field does not include any polarisation, exocyclic-group aromatic ring interactions or resonance contribution terms⁴⁶. Thus, simulations of i-motifs give an excellent account of how force fields deal with such unusual vertically stabilised structures.

4.2 Enhanced Sampling Methods

The data generated by classical MD for quadruplex stems are usually in agreement with the experimental results. However, the modeling of single-stranded loops is challenging as the electronic structure of loops is different from the backbone^{19, 21}. Furthermore, the loops are more exposed to the solvent and cations which makes the presentation of their interactions difficult. Localised enhanced sampling (LES) splits the loop regions of the simulated molecules in N independently moving copies^{47, 48}. This reduces the energy barriers between different substrates of the loop regions, proportionate to the number of copies generated ($1/N$). The LES technique mediates faster relaxation of the structure and allows comparison between the different loop geometries. Besides conventional MD simulations, Spomer *et al.* also implemented LES to study the loop dynamics¹⁹. The diagonal loop arrangement of $d(G_4T_4G_4)_2$ is highly unstable and different from the well-characterized crystal structure⁷. The subsequent free energy analysis also introduces a discrepancy as it predicts the incorrect structure to be more stable than crystal structure¹⁹. It has been derived that the instability of diagonal loop arrangement and imbalances in solute-cation and solvent-cation interactions observed in this study are due to the pair additivity of the Cornell force-field parm94 and parm98/99 force-field³⁶. LES modelling of loops in *Oxytricha nova* $d(G_4T_4G_4)_2$ quadruplex structure was done by Neidle and coworkers⁵. The lateral loops remained stable over nanosecond scale during the course of simulation. The G-quadruplex stem was unaffected by the loop dynamics. They also demonstrated that dimeric structures with T2 and T3 loops depend on loop lengths and not only on quartet stability. LES is capable of achieving major conformational changes within flexible loop region therefore multiple LES runs should be done for checking the reproducibility of the results.

4.3 Principal Component Analysis

The Principal Component Analysis (PCA) is a mathematical algorithm with fixed parameters to find a simplified coordinate representation for data so as to separate large amplitude motions from irrelevant fluctuations⁴⁹. To find the representation, eigen value is segregated such that resulting eigen vectors provide an orthonormal basis for the data and the corresponding eigen values provide information of each individual vector. PCA treats the trajectory to extract the dominant representations over the course of simulation. The simulated structure is translated to the geometrical centre of the molecule by least-square fit superimposition onto a reference structure. This reduces the overall rotation and translation of the structure over the course of trajectory. The configurational space is then constructed over a simple linear transformation in Cartesian coordinate space to generate a $3N \times 3N$ covariance matrix. The matrices are summed and averaged over the whole trajectory. The resulting matrix is then diagonalized generating a set of eigenvectors that gives a vectoral description of each component of the motion by indicating the direction of the motion. Each eigenvector describing the motion has a corresponding eigenvalue that represents the energetic contribution of the particular component to the motion. The eigenvalue is the average square displacement of the structure in the direction of the eigenvector. Projection of a trajectory on a particular eigenvector highlights the time dependent motions that the component performs in the particular vibrational mode. The time average of the projection shows the contribution of the components of the atomic vibrations to this mode of concerted motion⁵⁰. The eigen values are

placed in descending order where the first eigenvector and eigenvalue describes the largest internal motion of the structure. On average, only about 5% of eigenvectors are necessary to describe 90% of the total dynamics. Although PCA is a convenient method to visualise trajectories, its limitations should be taken into consideration when interpreting results⁵¹. PCA is most suited to analyse trajectories of systems that undergo transitional changes instead of trajectories that highlight thermal fluctuations of flexible molecules.

PCA has been applied to study the dynamic behaviour of human telomeric quadruplex dimeric and multimeric structures by Haider *et al.*⁵. The most prominent motions observed are the movements of the loops with a thymine adenine stack maintaining its adopted conformation and moving as a single unit in a concerted manner instead of wobbling of bases. However, the motions of different loops are independent of each other. The presence of ligand in the multimer changes the internal motion of the model. The most dominant motion in the quadruplex-ligand model is not the motion of the loops. This may suggest that the ligand in the pseudo-intercalation site is able to stabilise the model by reducing the motion of the loops to a lower component. PCA has also been employed to study interaction of dimeric quadruplex with perylene derivative (Tel03)⁵². In corroboration with the previous studies, it was observed that Tel03 binding to quadruplex reduces the flexibility of the loop regions⁵².

4.4 Gas-phase Simulations

It has been observed that parallel and antiparallel quadruplex maintain their structural integrity in the gas-phase in the presence of cation⁵³. MD simulations of quadruplexes in gas phase covering 1 μ s yield stable trajectories⁵⁴. The first principal components obtained from the trajectory of analogous simulations of gas phase and water is similar, implying that dominant motions in DNA are not altered by vaporization. During the course of gas-phase simulation of quadruplex DNA different structures are obtained depending on the central cation and the temperature. Similar to the explicit solvent conditions, expulsion of cation from the central channel of quadruplex leads to instantaneous fluctuations in the trajectory⁵⁴. The G-quadruplex observed in the simulations with Li^+ as the coordinating ion appears distorted and collapsed while Na^+ and K^+ ions yield stable quadruplexes. A distinct feature of gas phase simulation is that even when the structure is partially lost large numbers of guanine-guanine interactions are preserved. The interactions of a quadruplex in the gas phase are more complex than in the solvent phase. Along with MD, ESI-MS has been successfully used to gain further insight into quadruplex structure⁵⁵. Mazzitelli *et al.* performed 50ns MD simulations for twelve quadruplex models to complement ESI-MS⁵⁵. They used AMBER restrained electrostatic potential charge (RESP) procedure to reassign the charges of the neutral phosphate group and parm99 force-field for gas-phase simulations.

4.5 Continuum (Implicit) Solvent Methods

Assessment of structure and free energies during course of a simulation can be done by continuum solvent method⁵⁶. The approach is to simplify the energy calculations by implicitly integrating out all the solvent coordinates. The electrostatic contribution to the solvation energies is calculated by Poisson Boltzmann (PB) or Generalized

Born (GB) approaches. The hydrophobic contribution is represented by a surface area (SA) term obtained by scaling the solvent accessible surface area by an appropriate surface tension. The PB/GB solution along with SA term is used to calculate the absolute free energies of the molecules in solution⁵⁷. The MM-PBSA method extracts estimates of the free-energies from the MD trajectories based on averages of the gas phase molecular mechanical energy of the solute with an estimate of solvation free-energies from a PB continuum solvation model. The solvent and periodicity are removed in the post processing followed by the averaging of the energies over structural snapshots taken from the simulations. Both single and multiple trajectories can be used for MM-PBSA calculations however, single trajectories are preferred as it cancels sampling errors in the intra-molecular terms⁵⁸. The PB model is theoretically more complex than the GB model and MM/PBSA is often estimated better than MM/GBSA for predicting binding free energies. MM-PBSA has been used to calculate free-energies in G-quadruplex simulations including those involving quadruplex-ligand interactions. Hazel *et al.* used MM-PBSA contributions from loops and stem for analysis of quadruplex sub-structures⁴. They reported that energy values estimated from MM-PBSA do not reflect the stability of loops. The local fluctuations in the loop significantly affect the free-energy calculation at each step. This overshadowed the free-energies difference between distinct loops and no inference could be drawn based on it. However, energy difference between parallel and antiparallel quadruplex reflected that antiparallel structure is more stable than parallel structure. Similar approach was used by Haider *et al.* for use of MM-PBSA to compare the different quadruplex conformations⁵. The free-energy calculations in this study were well in agreement with the structural stability of quadruplex conformations. Ishikawa *et al.* have used MM-PBSA results to analyse the binding forces of quadruplex and porphyrins interactions^{59, 60}.

4.6 Free Energy Perturbation, Thermodynamic Integration, Potential of Mean Force and Umbrella Sampling

Free energy analysis between two states observed during the course of explicit MD simulation can be done by Free-energy perturbation (FEP) and thermodynamic integration (TI). FEP provides a quantitative method for studying the behaviour of DNA in response to a particular perturbation like base flipping⁶¹. TI allows accurate calculations of binding parameters of bis-intercalating molecules interaction with DNA⁶². As TI is very sensitive to the structural changes, it should only be used in the cases where ligand binding to DNA does not significantly affect the double helical nature. TI methods have also been used by Clark *et al.* to study the DNA-protein interactions. The work suggests the specificity of DNA base pair in recognition and binding of DNA⁶³.

The potential of mean force (PMF) is described as the change in free energy between the initial and final states as a function of Cartesian coordinates. PMF of simulation provides the thermal probability of finding a system at different points along the coordinates. The thermal probability is directly related to the free energy of the conformation. The widely used algorithm for calculating PMF is umbrella sampling⁶⁴. PMF profile has been used along with explicit atomistic MD simulation to investigate the folding and unfolding of TBA in the presence of Sr⁺². The analysis suggests that unfolding of TBA in presence of cation is a distinct multiple stepwise process and that the interplay of guanine, water molecules and cation govern the behaviour⁶⁵.

4.7 Quantum Mechanics/Molecular Mechanics (QM/MM)

Quantum Mechanical methods are based on the solution of Schrodinger equation. The equation describes the motions of electrons and nuclei in a molecular system. The *ab initio* method molecular orbitals are approximated by a linear combination of atomic orbitals⁶⁶. *Ab initio* quantum mechanical (QM) calculation is an important tool to investigate functional mechanisms of biological macromolecules based on their three-dimensional and electronic structures⁶⁷. In theory, the system size, which QM calculations can treat, is usually up to a few hundred atoms despite of huge sizes of biomacromolecules including solvent water molecules. The QM calculations are computationally very demanding. The largest system on which such calculations are available is a guanine dimer⁶⁸. The QM calculations deal with only intra-quartet interaction and do not take into account the effect of long-range electrostatic forces or solvation effects. The accuracy of the calculated molecular properties depends on the number of atomic orbitals used and quality of the basis set. Density functional theory (DFT) can be formulated as a variant of *ab initio* methods where correlation functionals are used to represent electron correlation energy⁶⁹. DFT can account for hydrogen bonding effects and can therefore calculate interactions within quartet and guanine-cation interactions. DFT does not assimilate base stacking parameters and therefore cannot describe inter-quartet interactions. Hybrid QM/MM methods rely on dividing the bimolecular systems into two regions: a smaller region which is treated quantum mechanically and the remaining region which is modelled by classical molecular mechanics force fields. QM/MM has been used to explore cation binding to TBA using Ba⁺² as the coordinating cation¹².

5 EXAMPLES OF MICROSECOND SCALE MOLECULAR DYNAMICS SIMULATIONS

5.1 Propeller-type Human Telomeric Repeat Quadruplex

Haider *et al.* have performed 1.5 μ s long simulation of 22mer sequence [AG₃(T₂AG₃)₃] (PDB id:1KF1) from human telomeric repeat that folds in a propeller-topology (unpublished results). The molecular dynamics simulations were carried out using Amber11 software employing the parmbsc0 force field. Additional K⁺ counterions were added to system to neutralize the charge on the quadruplex backbone. The K⁺ ions are arranged in a square antiprismatic coordination. They are sandwiched between the quartets and were retained in positions as in the crystal structure. The protocol for energy minimization and molecular dynamics in explicit solvent was adopted from Haider *et al.*⁵. The rmsd of all-atom structure is significantly greater than the backbone and is a result of flexible nature of TTA loops. The G-quartet stem was found to be the most stable sub-segment of the structure.

During the dynamics run, the central core geometry constituted by the three G-quartet planes pairing through their Watson-Crick and Hoogsteen edges, was retained along with its characteristic square planar arrangement. The grooves also retained their width and depth such that the same pattern of hydrogen donors and acceptors is observed throughout the course of the simulation. This pattern when compared with the crystal structure reflects the four-fold symmetry of the quadruplex. The geometry of the stem stacking as observed in the crystal structure, measured by the rise and twist, was also maintained throughout the course of the

simulation. The stacking geometry of 12 G-quartets per turn is retained, as observed in other intramolecular parallel stranded G-quartets. There was no movement of the potassium ions out of the central quadruplex channel as observed in some simulations of quadruplexes^{12, 36}.

We observed small fluctuations within the first 300ns, which eventually stabilized over the course of the simulation. Examination of the trajectory in short bins revealed that the structure adopted stable conformations that could last several nanoseconds. While this stability might have sufficed for short multi-nanosecond scale simulation, however, placing it in the context of a long microsecond or longer simulation, these structures are transitional. Therefore the first 300ns were treated as equilibration and all analysis was carried on the trajectory beyond this point. The cutoff point at 300ns was determined via principal components analysis, until which no convergence was observed. The microscale simulation of quadruplex highlighted the importance of timescale on which these propeller-type structures equilibrate, which would otherwise be thought as stable conformations. Following the flexibility of the loop structure over the course of the simulation and comparing them with the X-ray structure reveals significant conformational changes. Moreover, the mobility of the simulated loops is independent of each other. Loop1 and loop2 shows high rmsd reflecting more movement while loop3 shows is more stable as reflected by less rmsd. However, the loop conformational rearrangement did not have any great impact on the structure of the central G-quartet stem. A cluster analysis over the microsecond trajectory resulted in eight clusters. It is noteworthy that three of these can be classified as clusters that contain metastable intermediates that resemble conformations to that observed in various NMR and X-ray solved structures.

5.2 Thrombin Binding Aptamer

Aptamers are short DNA or RNA fragments that can bind, with high affinity and selectivity to defined molecular targets, including biological proteins⁷⁰. Among the different conformational arrangements that can be assumed by aptamers, a quadruplex represents suitable nucleic acid architecture of noteworthy importance, due to its dramatic thermal stability. Contextually, potentially therapeutic aptamers like those endowed with anti-HIV and anti-thrombin activity have driven great interests^{10, 71-73}. In the study of both these classes of aptamers, computational methods have proved to be excellent research tools, able to provide a deep insight into the aptamer-protein interactions^{35, 74, 75}. Reshetnikov *et al.* investigated the TBA-thrombin interaction by MD simulation using AMBER-99 ϕ and parmbsc0 force-field on GROMACS software package³⁵. The simulations were conducted in a range time from 600 to 900ns in individual runs, reaching a total simulation time that exceeds 12 μ s. Separate 900ns MD simulation runs were performed, in both parm99 and parmbsc0 force fields, to compare the viability of NMR- and X-ray-based structures of free TBA. Data from this study suggest that the X-ray-based conformation is unstable, while the NMR-based structure is the only viable TBA conformation. Furthermore, this study highlighted all the structural features that influence the stability of the structure, and the parmbsc0 force field was elected as the most suitable force field to conduct the subsequent simulations. To analyse the precise TBA residues that interact with the thrombin structure, 600ns MD simulation runs were performed in parmbsc0 force field to compare the NMR- and X-ray-based models of the thrombin-TBA 1:1 complex. It could be noted that the two starting complex structures showed different binding modes. In the NMR-based complex,

aptamer interacts by the TT-loops, involving the Exosite I of the thrombin, while in the X-ray-based complex the interaction involves the TGT-loop and the Exosite II. Also in this case, since the X-ray-based complex collapses during the simulation, data show that the NMR-based complex was favoured, resulting in the only viable complex. The same studies were performed for the thrombin-TBA 1:2 complex. In this case both the complexes survived during the simulations, even if the G-quartet planarity of the aptamer is disrupted in the X-ray-based complex. In summary, all the simulations performed by Reshetnikov *et al.* definitely support the NMR-based model of TBA, where the TGT-loop has a stabilizing influence on the TBA molecule and TT-loops mediate the thrombin binding and its inhibition.

In agreement with these results, the changes made to the TT-loops can lead to different consequences. An example is provided by the work of Pagano *et al.*, in which a modified TBA (mTBA), containing a 5'-5' site of polarity inversion up to TT-loop, shows a better stability and a higher thrombin affinity but a less inhibitory activity than the unmodified TBA⁷⁵. To better understand the biological results, the interaction patterns for the mTBA-thrombin 1:2 complex have been investigated by 5ns MD simulation, performed with the GROMACS package software in parm98 force field. The results highlight that Exosite I interacting with mTBA makes less contact when compared with the unmodified TBA but, simultaneously, the Exosite II of a second thrombin makes more contacts with mTBA. In this case, the MD simulation data corroborate the hypothesis that the binding with the Exosite I is the only one with an inhibitory effect^{76, 77}. In fact, even if mTBA binds to thrombin with an increased affinity, this enhancement is due to better interactions with the non-inhibiting Exosite II of the protein, resulting in a decreasing of inhibitory activity. Furthermore, it is fair to emphasize that the deep understanding of biological data has been possible only through the use of MD simulation studies.

As previously mentioned, a solved structure of the aptamer-protein complex may not be necessary to conduct computational studies when the structures are independently solved. A suitable example is the work of Sgobba *et al.*, in which 93del-HIV1-integrase complex has been obtained through docking studies and then subjected to MD simulations to obtain a structural perspective on the mechanism of inhibition⁷⁴. In this case, the binding interactions between aptamer and protein have been investigated by 35ns MD simulation, performed with the AMBER11 package software, and parmbsc0 and ff99SBildn force fields. The energy minimization and MD simulation in explicit solvent have been performed following the protocol reported by Perryman *et al.*⁷⁸. Moreover, the Principal Component Analysis (PCA) and the Elastic Network Models (ENM) have been performed to identify the motions of HIV1-Integrase (HIV1-IN) in complex with 93del aptamer and to describe structural flexibility in order to correlate the relationship between dynamics and function of the protein. The reported data have highlighted the key residues of HIV1-IN essential for the binding with the aptamer, capable to disrupt the HIV1-IN-DNA interactions. Furthermore, data collected by PCA and ENM have confirmed that interaction involving the catalytic loops are responsible of inhibition activity of 93del, since they conduct to a conformational rearrangement that hold the protein in an inactive conformation.

6 VIRTUAL SCREENING OF G-QUADRUPLEXES

The expression 'virtual screening' describes the use of computational algorithms and models for the identification of novel bioactive molecules. Over the years, virtual

screening has become an important component of the *in silico* search for the hit and lead compounds and their optimization⁷⁹. It can be divided into two main categories, namely structure-based virtual screening which utilizes the 3D structure of biological target and ligand-based virtual screening where the structure activity data from a set of known active compounds are employed⁸⁰. Despite significant progress in generation of potential ligand 'poses' by automated molecular docking, there are a number of pitfalls in the existing virtual screening methods that needs to be addressed^{79, 81}. The scoring functions used for prioritization of various suggested binding poses are still very much inaccurate^{79, 82}. To this date, there are over 150 quadruplex structures deposited in the Protein Data Bank (<http://www.rcsb.org/pdb>). However, the structure of quadruplex may vary with the experimental methods and conditions. Therefore, choosing a suitable model for modelling and structure-based virtual screening is critical.

With the advent of computational techniques, experimental biophysical assays are complemented by molecular docking and ligand-based pharmacophore studies to explore plausible binding modes of quadruplex-ligand interaction, optimize the lead compounds and rationalize their selectivity^{5, 83}. Interaction of cationic tetraolylporphyrin derivatives with the NMR intramolecular antiparallel telomeric G-quadruplex d(AG3[T2AG3]3)(PDB code: 143D) was studied by molecular docking using DOCK6⁶⁰. A model of *c-myc* quadruplex was generated for molecular docking study with platinum (II) Schiff base complexes. The first virtual screen of library of FDA-approved drugs (3000 compounds) for *c-myc* quadruplex stabilizing ligands was done employing the ICM (molsoft) program⁸⁴. Selectivity of a group of naphthalene diimide ligands for telomeric G4-RNA over the G4-DNA was explained by docking using the Accelrys AFFINITY program⁸⁵. Two quadruplex stabilizing alkaloid compounds were selected by screening a Chinese herbal database of 10000 compounds with previously generated pharmacophores using Catalyst, Accelrys⁸⁶. Chen *et al.* on the basis of acridine derivatives, developed ligand-based pharmacophore model and identified triaryl-substituted imidazole derivative TSIZ01 as potent G-quadruplex ligand⁸⁷.

The biggest challenge in virtual screening of G-quadruplex has been the selection of most favourable pose of quadruplex for *in silico* studies. This is due to highly charged backbone, presence of stabilizing alkali metal cations and in particular the flexibility of the loops. Quality of the quadruplex docking, and subsequently the binding affinity of the ligands might be strongly affected by the flexibility of the loop regions⁸⁸. To address the receptor and ligand flexibility issue and subsequently conformational change upon binding, a novel form of 'dynamic docking' has been developed by Neidle and co-workers (unpublished results). Over 250 short explicit solvent molecular dynamics (MD) simulations were employed to explore binding, transition states and preferred conformations of the quadruplex DNA end-stacking ligand *pyridostatin* and *RHPS4* with the parallel native X-ray structure of telomeric quadruplex. Diverse conformations of the ligand were placed in a grid-like manner and in multiple orientations parallel to the 3' and 5' site of the quadruplex structure. Multiple starting positions for subsequent MD simulations were generated, allowing large conformational space for both ligand and its target to be explored (unpublished data).

Virtual screening of large chemical libraries to aid the selection of new quadruplex-binding ligands has recently been explored⁸⁹. A synthetic substituted indole was identified by screening ~100,000 drug-like compounds⁹⁰. Fonseca B, a natural naphthopyrone compound was identified by screening 20,000 compound in

natural product database⁹¹. *In silico* studies along with NMR experiments in tandem provide better details of interactions of ligands with G4-DNA⁹². A relatively small but structurally diverse commercially available database (6000 compounds) was screened against the parallel quadruplex [d(TG₄T)₄] using Autodock v4. Subsequent NMR screening of the top 30 hits identified six G4-groove binding molecules that were further studied in detail by NMR, ITC measurements and molecular docking with modified quadruplexes. These six molecules are the most potent G-quadruplex groove-binders identified so far⁹³.

Large-scale integrated *in silico* and *in vitro* screening platforms to discover novel small molecules binding to specific nuclei acids were developed in the Chaires group, using the analogy of a 'funnel'⁹⁴. The ZINC 'drug-like' virtual database of 1.3 million compounds was screened against the antiparallel quadruplex target (pdb id: 2HY9) employing the Surflex-Dock molecular docking software. An array of potential nucleic acids competing sites, as well as all possible binding sites on the target nucleic acid itself were considered for the virtual screening, which was performed on a grid of more than 10000 computer processors. This approach and the appropriate choice of the software, was previously validated^{95, 96}. The top 160 hits that emerged after scoring, for selective binding to the quadruplex target, were tested by high-throughput melting assay followed by a secondary screening of the top compounds. Characterization of their binding behaviour with the quadruplex target, by means of rigorous binding studies using calorimetry, spectroscopy, competition dialysis, molecular dynamics simulations and functional assays, identified a substantially stabilizing quadruplex binding ligand, which in turn suggests that the proposed *in silico* and *in vitro* platform may be used to discover new G-quadruplex binding scaffolds⁹⁴.

At present, virtual screening alone is rather unable to adequately predict the selectivity of a particular ligand for different G-quadruplexes, because of the inherent inaccuracies in comparing binding energies between multiple G-quadruplex structures⁸⁸. However the G-quadruplex structures display a great diversity in their loops and grooves geometries, which in particular could be used to enhance the selectivity of ligands, allowing for their excellent structure-specific recognition, affinity and specificity. For instance, in the case of the promoter *KIT1* quadruplex structure^{97, 98}, the presence of a distinct cleft may be suitable for ligand binding and virtual high-throughput screening⁹⁹.

7 CONCLUSIONS

There are over 150 NMR or X-ray derived structures present in the PDB that detail atomistic information. However, it should be emphasised that these are time-average static representations and need to be simulated in before any realistic information can be extracted. Such simulations carried out at ambient temperatures can provide information on interactions between solvent, solute and ions, cross energy barriers to access conformations, folding intermediates and metastable structures for which there is no experimental data available. There is no other technique that can provide time coursed dynamic details of interactions and therefore molecular dynamics simulations have been a chosen method to study quadruplexes, both native and in complex with ligands and proteins. The best usage of simulations is envisaged to complement the explanation and interpretation of experimental data while also

supplements it with information that cannot be accessed by experimental measurements.

References

1. T. Simonsson, *Biological chemistry*, 2001, **382**, 621-628.
2. K. J. Neaves, J. L. Huppert, R. M. Henderson and J. M. Edwardson, *Nucleic acids research*, 2009, **37**, 6269-6275.
3. R. Rodriguez, K. M. Miller, J. V. Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian and S. P. Jackson, *Nat Chem Biol*, 2012, **8**, 301-310.
4. P. Hazel, G. N. Parkinson and S. Neidle, *Nucleic acids research*, 2006, **34**, 2117-2127.
5. S. Haider, G. N. Parkinson and S. Neidle, *Biophysical journal*, 2008, **95**, 296-311.
6. N. V. Hud and J. Plavec, *The role of cations in determining quadruplex structure and stability*, Royal Society of Chemistry, U.K., 2006.
7. S. Haider, G. N. Parkinson and S. Neidle, *Journal of molecular biology*, 2002, **320**, 189-200.
8. E. A. Venczel and D. Sen, *Biochemistry*, 1993, **32**, 6220-6228.
9. N. V. Hud, F. W. Smith, F. A. Anet and J. Feigon, *Biochemistry*, 1996, **35**, 15383-15390.
10. P. Schultze, R. F. Macaya and J. Feigon, *Journal of molecular biology*, 1994, **235**, 1532-1547.
11. G. D. Strahan, M. A. Keniry and R. H. Shafer, *Biophysical journal*, 1998, **75**, 968-981.
12. R. V. Reshetnikov, J. Sponer, O. I. Rassokhina, A. M. Kopylov, P. O. Tsvetkov, A. A. Makarov and A. V. Golovin, *Nucleic acids research*, 2011, **39**, 9789-9802.
13. X. Cang, J. Sponer and T. E. Cheatham, 3rd, *Journal of the American Chemical Society*, 2011, **133**, 14270-14279.
14. S. Haider and S. Neidle, *Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes*, Springer Protocols, 2008.
15. P. Maurer, A. Laio, H. W. Hugosson, M. C. Colombo and U. Rothlisberger, *J. Chem. Theory Comput.*, 2007, **3**, 628-639.
16. T. A. Soares, P. H. Hunenberger, M. A. Kastenzholz, V. Krautler, T. Lenz, R. D. Lins, C. Oostenbrink and W. F. van Gunsteren, *J Comput Chem*, 2005, **26**, 725-737.
17. S. M. Haider and S. Neidle, *Molecular Dynamics and Force Field Based Methods for studying Quadruplex Nucleic Acids*, The Royal Society of Chemistry, U.K., 2012.
18. A. W. Van Wynsberghe and Q. Cui, *Biophysical journal*, 2005, **89**, 2939-2949.
19. E. Fadna, N. Spackova, R. Stefl, J. Koca, T. E. Cheatham, 3rd and J. Sponer, *Biophysical journal*, 2004, **87**, 227-242.
20. A. Perez, F. J. Luque and M. Orozco, *Journal of the American Chemical Society*, 2007, **129**, 14739-14745.
21. E. Fadná, N. a. Špačková, J. Sarzyńska, J. Koča, M. Orozco, T. E. Cheatham, T. Kulinski and J. i. Šponer, *Journal of Chemical Theory and Computation*, 2009, **5**, 2514-2530.

22. J. Q. Hou, S. B. Chen, J. H. Tan, T. M. Ou, H. B. Luo, D. Li, J. Xu, L. Q. Gu and Z. S. Huang, *The journal of physical chemistry. B*, 2010, **114**, 15301-15310.
23. L. Petraccone, I. Fotticchia, A. Cummaro, B. Pagano, L. Ginnari-Satriani, S. Haider, A. Randazzo, E. Novellino, S. Neidle and C. Giancola, *Biochimie*, 2011, **93**, 1318-1327.
24. W. S. Ross and C. C. Hardin, *Journal of the American Chemical Society*, 1994, **116**, 6070-6080.
25. T. E. Cheatham, III, J. L. Miller, T. Fox, T. A. Darden and P. A. Kollman, *Journal of the American Chemical Society*, 1995, **117**, 4193-4194.
26. J. Norberg and L. Nilsson, *Biophysical journal*, 2000, **79**, 1537-1553.
27. J. Sponer, J. Leszczynski and P. Hobza, *Biopolymers*, 2001, **61**, 3-31.
28. P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, *J Comput Chem*, 1995, **16**, 1357-1377.
29. P. Varnai and K. Zakrzewska, *Nucleic acids research*, 2004, **32**, 4269-4280.
30. W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell and P. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179-5197.
31. D. L. Beveridge, G. Barreiro, K. S. Byun, D. A. Case, T. E. Cheatham, 3rd, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai and M. A. Young, *Biophysical journal*, 2004, **87**, 3799-3813.
32. F. Barone, F. Lankas, N. Spackova, J. Sponer, P. Karran, M. Bignami and F. Mazzei, *Biophysical chemistry*, 2005, **118**, 31-41.
33. T. E. Cheatham, P. Cieplak and P. A. Kollman, *Journal of Biomolecular Structure and Dynamics*, 1999, **16**, 845-862.
34. A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, 3rd, C. A. Laughton and M. Orozco, *Biophysical journal*, 2007, **92**, 3817-3829.
35. R. V. Reshetnikov, A. V. Golovin, V. Spiridonova, A. M. Kopylov and J. Sponer, *J Chem Theory Comput*, 2010, **6**, 3003-3014.
36. N. a. Špačková, I. Berger and J. Šponer, *Journal of the American Chemical Society*, 1999, **121**, 5519-5534.
37. G. N. Parkinson, M. P. Lee and S. Neidle, *Nature*, 2002, **417**, 876-880.
38. F. Cuenca, O. Greciano, M. Gunaratnam, S. Haider, D. Munnur, R. Nanjunda, W. D. Wilson and S. Neidle, *Bioorganic & medicinal chemistry letters*, 2008, **18**, 1668-1673.
39. S. Agrawal, R. P. Ojha and S. Maiti, *The Journal of Physical Chemistry B*, 2008, **112**, 6828-6836.
40. M. Cavallari, A. Garbesi and R. Di Felice, *The Journal of Physical Chemistry B*, 2009, **113**, 13152-13160.
41. D.-Y. Yang and S.-Y. Sheu, *The Journal of Physical Chemistry A*, 2007, **111**, 9224-9232.
42. M.-H. Li, Q. Luo and Z.-S. Li, *The Journal of Physical Chemistry B*, 2010, **114**, 6216-6224.
43. V. Dhamodharan, S. Harikrishna, C. Jagadeeswaran, K. Halder and P. I. Pradeepkumar, *The Journal of Organic Chemistry*, 2011, **77**, 229-242.
44. D. Kaluzhny, N. Ilyinsky, A. Shchekotikhin, Y. Sinkevich, P. O. Tsvetkov, V. Tsvetkov, A. Veselovsky, M. Livshits, O. Borisova, A. Shtil and A. Shchylolkina, *PLoS ONE*, 2011, **6**, e27151.

45. T. A. Brooks, S. Kendrick and L. Hurley, *The FEBS journal*, 2010, **277**, 3459-3469.
46. N. Špacková, I. Berger, M. Egli and J. Šponer, *Journal of the American Chemical Society*, 1998, **120**, 6147-6151.
47. C. Simmerling, J. L. Miller and P. A. Kollman, *Journal of the American Chemical Society*, 1998, **120**, 7149-7155.
48. R. Elber and M. Karplus, *Journal of the American Chemical Society*, 1990, **112**, 9161-9175.
49. A. Amadei, A. B. Linssen, B. L. de Groot, D. M. van Aalten and H. J. Berendsen, *Journal of biomolecular structure & dynamics*, 1996, **13**, 615-625.
50. D. M. van Aalten, A. Amadei, A. B. Linssen, V. G. Eijsink, G. Vriend and H. J. Berendsen, *Proteins*, 1995, **22**, 45-54.
51. K. Reblova, Z. Strelcova, P. Kulhanek, I. Besseova, D. H. Mathews, K. V. Nostrand, I. Yildirim, D. H. Turner and J. Sponer, *J Chem Theory Comput*, 2010, **2010**, 910-929.
52. M.-H. Li, Q. Luo, X.-G. Xue and Z.-S. Li, *Journal of Molecular Modeling*, 2011, **17**, 515-526.
53. M. Rueda, F. J. Luque and M. Orozco, *Journal of the American Chemical Society*, 2006, **128**, 3608-3619.
54. V. Gabelica, F. Rosu, M. Witt, G. Baykut and E. De Pauw, *Rapid communications in mass spectrometry : RCM*, 2005, **19**, 201-208.
55. C. L. Mazzitelli, J. Wang, S. I. Smith and J. S. Brodbelt, *Journal of the American Society for Mass Spectrometry*, 2007, **18**, 1760-1773.
56. T. Rodinger and R. Pomes, *Current opinion in structural biology*, 2005, **15**, 164-170.
57. M. Zacharias, eds. J. Šponer and F. Lankaš, Springer Netherlands, 2006, pp. 95-119.
58. N. Spackova, T. E. Cheatham, 3rd, F. Ryjacek, F. Lankas, L. Van Meervelt, P. Hobza and J. Sponer, *Journal of the American Chemical Society*, 2003, **125**, 1759-1769.
59. Y. Ishikawa, Y. Tomisugi and T. Uno, *Nucleic acids symposium series*, 2006, 331-332.
60. Y. Ishikawa, E. Higashi and H. Morioka, *Nucleic acids symposium series*, 2007, 247-248.
61. N. K. Banavali and A. D. Mackerell, Jr., *PLoS One*, 2009, **4**, e5525.
62. E. Marco, A. Negri, F. J. Luque and F. Gago, *Nucleic acids research*, 2005, **33**, 6214-6224.
63. Frank R. Beierlein, G. G. Kneale and T. Clark, *Biophysical journal*, 2011, **101**, 1130-1138.
64. B. Roux, *Computer Physics Communications*, 1995, **91**, 275-282.
65. C. Yang, S. Jang and Y. Pak, *J Chem Phys*, 2011, **135**, 225104.
66. R. Iftimie, P. Minary and M. E. Tuckerman, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 6654-6659.
67. B. Kirchner, F. Wennmohs, S. Ye and F. Neese, *Current Opinion in Chemical Biology*, 2007, **11**, 134-141.
68. J. Sponer, P. Jurecka, I. Marchan, F. J. Luque, M. Orozco and P. Hobza, *Chemistry*, 2006, **12**, 2854-2865.
69. R. G. Parr and W. Yang, *Annual Review of Physical Chemistry*, 1995, **46**, 701-728.

70. S. M. Nimjee, C. P. Rusconi and B. A. Sullenger, *Annual review of medicine*, 2005, **56**, 555-583.
71. K. Padmanabhan and A. Tulinsky, *Acta crystallographica. Section D, Biological crystallography*, 1996, **52**, 272-282.
72. J. D'Onofrio, L. Petraccone, E. Erra, L. Martino, G. Di Fabio, L. De Napoli, C. Giancola and D. Montesarchio, *Bioconjugate Chemistry*, 2007, **18**, 1194-1204.
73. A. T. Phan, V. Kuryavyi, J. B. Ma, A. Faure, M. L. Andreola and D. J. Patel, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 634-639.
74. M. Sgobba, O. Olubiyi, S. Ke and S. Haider, *Journal of biomolecular structure & dynamics*, 2012, **29**, 863-877.
75. B. Pagano, L. Martino, A. Randazzo and C. Giancola, *Biophysical journal*, 2008, **94**, 562-569.
76. M. Tsiang, A. K. Jain, K. E. Dunn, M. E. Rojas, L. L. Leung and C. S. Gibbs, *The Journal of biological chemistry*, 1995, **270**, 16854-16863.
77. K. Y. Wang, S. H. Krawczyk, N. Bischofberger, S. Swaminathan and P. H. Bolton, *Biochemistry*, 1993, **32**, 11285-11292.
78. A. L. Perryman, S. Forli, G. M. Morris, C. Burt, Y. Cheng, M. J. Palmer, K. Whitby, J. A. McCammon, C. Phillips and A. J. Olson, *Journal of molecular biology*, 2010, **397**, 600-615.
79. G. Schneider, *Nature reviews. Drug discovery*, 2010, **9**, 273-276.
80. A. Jahn, G. Hinselmann, N. Fechner and A. Zell, *Journal of cheminformatics*, 2009, **1**, 14.
81. T. Scior, A. Bender, G. Tresadern, J. L. Medina-Franco, K. Martinez-Mayorga, T. Langer, K. Cuanalo-Contreras and D. K. Agrafiotis, *Journal of chemical information and modeling*, 2012.
82. M. Totrov and R. Abagyan, *Proteins*, 1997, **Suppl 1**, 215-220.
83. S. M. Haider, I. Autiero and S. Neidle, *Biochimie*, 2011, **93**, 1275-1279.
84. D. S. Chan, H. Yang, M. H. Kwan, Z. Cheng, P. Lee, L. P. Bai, Z. H. Jiang, C. Y. Wong, W. F. Fong, C. H. Leung and D. L. Ma, *Biochimie*, 2011, **93**, 1055-1064.
85. G. W. Collie, S. M. Haider, S. Neidle and G. N. Parkinson, *Nucleic acids research*, 2010, **38**, 5569-5580.
86. Q. Li, J. Xiang, X. Li, L. Chen, X. Xu, Y. Tang, Q. Zhou, L. Li, H. Zhang, H. Sun, A. Guan, Q. Yang, S. Yang and G. Xu, *Biochimie*, 2009, **91**, 811-819.
87. S. B. Chen, J. H. Tan, T. M. Ou, S. L. Huang, L. K. An, H. B. Luo, D. Li, L. Q. Gu and Z. S. Huang, *Bioorganic & medicinal chemistry letters*, 2011, **21**, 1004-1009.
88. D. L. Ma, V. P. Ma, D. S. Chan, K. H. Leung, H. J. Zhong and C. H. Leung, *Methods*, 2012.
89. S. Neidle, in *Therapeutic Applications of Quadruplex Nucleic Acids*, Academic Press, Boston, 2012, pp. 151-174.
90. D. L. Ma, T. S. Lai, F. Y. Chan, W. H. Chung, R. Abagyan, Y. C. Leung and K. Y. Wong, *ChemMedChem*, 2008, **3**, 881-884.
91. H. M. Lee, D. S. Chan, F. Yang, H. Y. Lam, S. C. Yan, C. M. Che, D. L. Ma and C. H. Leung, *Chemical communications*, 2010, **46**, 4680-4682.
92. S. Cosconati, L. Marinelli, R. Trotta, A. Virno, L. Mayol, E. Novellino, A. J. Olson and A. Randazzo, *Journal of the American Chemical Society*, 2009, **131**, 16336-16337.

93. R. Trotta, S. De Tito, I. Lauri, V. La Pietra, L. Marinelli, S. Cosconati, L. Martino, M. R. Conte, L. Mayol, E. Novellino and A. Randazzo, *Biochimie*, 2011, **93**, 1280-1287.
94. P. A. Holt, R. Buscaglia, J. O. Trent and J. B. Chaires, *Drug development research*, 2011, **72**, 178-186.
95. P. A. Holt, J. B. Chaires and J. O. Trent, *Journal of chemical information and modeling*, 2008, **48**, 1602-1615.
96. P. A. Holt, P. Ragazzon, L. Strekowski, J. B. Chaires and J. O. Trent, *Nucleic acids research*, 2009, **37**, 1280-1287.
97. A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *Journal of the American Chemical Society*, 2007, **129**, 4386-4392.
98. D. Wei, G. N. Parkinson, A. P. Reszka and S. Neidle, *Nucleic acids research*, 2012, **40**, 4691-4700.
99. S. Balasubramanian, L. H. Hurley and S. Neidle, *Nature reviews. Drug discovery*, 2011, **10**, 261-275.

Insight Into the Conformational Arrangement of a Bis-THF Diol Compound Through 2D-NMR Studies and X-Ray Structural Analysis

Vincenzo Piccialli · Sabrina Zaccaria · Roberto Centore ·
Angela Tuzi · Nicola Borbone · Giorgia Oliviero ·
Stefano D'Errico · Valentina D'Atri

Received: 3 August 2011 / Accepted: 3 December 2011 / Published online: 18 December 2011
© Springer Science+Business Media, LLC 2011

Abstract The conformational arrangement of an all-*threo* bis-THF diol compound, synthesized through the RuO₄-catalysed oxidative bis-cyclization of farnesyl acetate, was determined via crystallographic analysis and detailed 2D-NMR solution studies. The bis-THF compound crystallizes in the orthorhombic *Pbca* space group, with unit cell parameters $a = 10.496(1)$, $b = 17.974(1)$, $c = 19.777(2)$ Å, $Z = 8$. The final refinement converged to $R_1 = 0.0484$ for 4714 independent observed reflections having $I > 2\sigma(I)$. There is a good agreement between the solution molecular conformation determined by 2D-NMR and the X-ray molecular conformation. The molecule adopts a folded, horse shoe-type conformation both in solution and in the crystal, that suggests aptitude to coordinate cations. Additionally, in the crystals, the molecular conformation is stabilized by intramolecular and intermolecular H-bonding.

Keywords Bis-tetrahydrofuran · Oxidative bis-cyclization · RuO₄ · X-ray · 2D-NMR

Introduction

Tetrahydrofuran-containing compounds are widely distributed in nature. For example, 2,5-disubstituted THF fragments are commonly encountered in a variety of biologically active substances such as polyether ionophores, squalene-derived polycyclic polyether metabolites (oxasqualenoids) as well as Annonaceous acetogenins, a large group of plant-derived metabolites displaying a wide range of biological activities [1–5]. It is well known that the biological activity of polyether ionophores is directly related to their ability to selectively bind physiologically important metal cations via coordination with various oxygen atoms of the molecule and transport them through biological membranes [6]. As a result, the cation equilibrium through the cell membrane is altered resulting in various activities such as antibiotic, antimalarial and insecticide activity, among others.

Previous studies carried out mostly on natural THF-containing compounds belonging to one of the above classes or their analogues have demonstrated that both adjacent and non-adjacent poly-THF compounds often possess cation complexation and transport abilities. In particular, Kodama and coworkers [7] have shown that the triterpene polyether 14-deacetyl eurylene, isolated from the wood of *Ericoma longifolia*, and its non-natural analogues 11-epi-eurylene, 11-epi-14-deacetyleurylene (Fig. 1), all embodying two non-adjacent THF rings, possess K⁺ transport ability. These authors and others suggested that the cytotoxicity of these substances is related to their complexation ability. Itokawa and coworkers [8] proposed that a relationship exists between the cytotoxicity and the conformation of 14-deacetyl eurylene, and structurally related longilene peroxide and teurilene, also isolated from the above organism. Based on X-ray crystallographic

V. Piccialli (✉) · S. Zaccaria
Dipartimento di Chimica Organica e Biochimica, Università
degli Studi di Napoli “Federico II”, Via Cynthia 4, 80126
Naples, Italy
e-mail: vinpicci@unina.it

R. Centore (✉) · A. Tuzi
Dipartimento di Chimica “Paolo Corradini”, Università degli
Studi di Napoli “Federico II”, Via Cynthia 4, 80126 Naples,
Italy
e-mail: roberto.centore@unina.it

N. Borbone · G. Oliviero · S. D'Errico · V. D'Atri
Dipartimento di Chimica delle Sostanze Naturali, Università
degli Studi di Napoli “Federico II”, Via D. Montesano 49,
80131 Naples, Italy

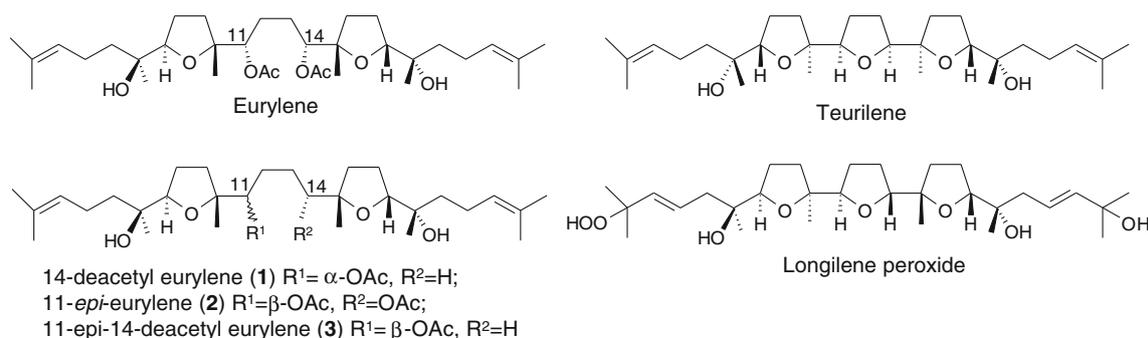


Fig. 1 Some cytotoxic triterpene poly-tetrahydrofurans

analyses and NOE experiments they showed that these compounds possess folded, “horse shoe”, conformation. Later Morimoto et al. [9] suggested that the cytotoxic activity displayed by these compounds may be attributed to their ability to bind physiologically important cations such as Ca^{2+} and Mg^{2+} possibly due to their folded conformation. Indeed, eurylene [8] (Fig. 1), another representative of this class of substances, possessing an extended conformation, has essentially no cytotoxic activity.

Morimoto et al. [9] synthesised three isomeric tris-THF diol compounds (Fig. 2) and carried out cation transport experiments demonstrating that all three compounds possess cation transport ability for Na^+ and K^+ . Furthermore, it has been suggested that the biological activities of Annonaceae acetogenins involving THF rings may be attributed to their ability to bind physiologically important cations such as Ca^{2+} [10–14] (Fig. 2). Indeed, mimics of bis-THF acetogenins such as bullatacin, possessing a variety of functionalised side chains, have been shown to selectively bind Ca^{2+} or Mg^{2+} [11, 15].

Based on this ground, we were interested to test the cation complexation ability of some poly-THF compounds structurally related to the above substances, previously synthesized in our laboratories through ruthenium-mediated chemistry. In particular, we have previously demonstrated that adjacently linked bis-, tris- and even more complex poly-THF compounds can be easily synthesized in a single step via the oxidative polycyclization of all-*trans*

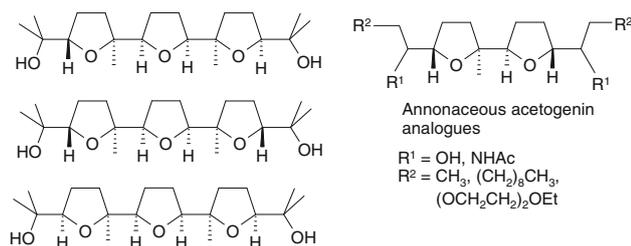


Fig. 2 Synthetic tris-tetrahydrofurans possessing cation transport abilities (*left*) and Annonaceous acetogenins analogues (*right*) able to bind Ca^{2+} and/or Mg^{2+}

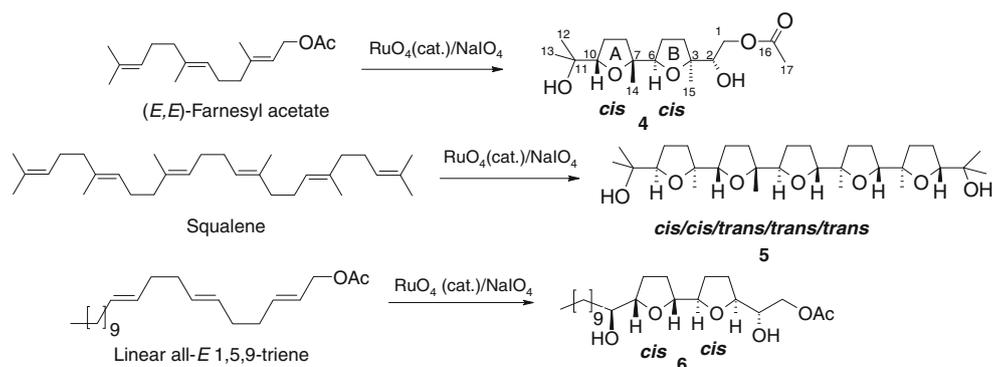
isoprenoid polyenes such as farnesyl acetate, geranylgeranyl acetate, squalene or ad hoc synthesised all-*E* linear polyenes (Scheme 1) [15–21].

It is interesting to note that, owing to postulated biogenetic pathways leading to the above natural substances as well as epoxide chemistry used for the synthesis of unnatural analogues, almost all the bis- or tris-THF compounds examined for cation complexation and/or transport ability possess *erythro* inter-THF relationships, whereas no studies have been accomplished on poly-THF compounds possessing all-*threo* inter-THF relationships such as compounds 4–6. As a part of our ongoing interest in oxidative processes mediated by transition metal oxo-species [22–24] and in particular on the derivatization and biological properties of mono and poly-THF compounds [3, 15–22, 25–27], we report here solid state X-ray and solution 2D-NMR studies on bis-THF diol 4 (Scheme 1) synthesized as previously reported [15].

Experimental

General Methods

All reagents were purchased (Aldrich and Fluka) at the highest commercial quality and used without further purification. Reactions were monitored by thin-layer chromatography carried out on precoated silica gel plates (Merck 60, F254, 0.25 mm thick). Merck silica gel (Kieselgel, particle size 0.063–0.200 mm) was used for column chromatography. HPLC separations were carried out on a Varian 2510 apparatus equipped with a Waters R403 dual cell differential refractometer using Phenomenex 250×10 mm column. NMR experiments were performed on Varian Mercury Plus 400 spectrometer in CDCl_3 . Proton chemical shifts were referenced to the residual CHCl_3 signal (7.26 ppm); ^{13}C -NMR chemical shifts were referenced to the solvent (77.0 ppm). *J* values are in Hz. Abbreviations for signal coupling are as follows: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet.



Scheme 1 Representative examples of the oxidative polycyclization of isoprenoid and linear polyenes catalyzed by ruthenium tetroxide

X-Ray Crystallography

Single crystals of **4** were obtained from a CHCl_3 solution by slow evaporation of the solvent. Data were collected at ambient temperature on a Bruker-Nonius Kappa-CCD diffractometer using graphite monochromated $\text{MoK}\alpha$ radiation ($\lambda = 0.71073 \text{ \AA}$). Unit cell parameters were determined by least squares refinement of the θ angles of 143 strong reflections in the range $3.501^\circ < \theta < 22.763^\circ$. Data reduction and multi-scan absorption correction were done using SADABS program [28]. The structure was solved by direct methods (SIR97 program [29]) and refined by the full matrix least-squares method on F^2 using SHELXL-97 program [30] with the aid of the program WinGX [31]. Non-hydrogen atoms were refined anisotropically. H atoms of hydroxyl groups were located in difference Fourier maps and refined with $U_{\text{iso}} = 1.2 U_{\text{eq}}$ of the carrier atom. The positions of the other H atoms were determined stereochemically and refined by the riding model with $U_{\text{iso}} = 1.2 U_{\text{eq}}$ of the carrier atom ($1.5 U_{\text{eq}}$ for H atoms of methyl groups). Ring puckering coordinates [32] were determined using the program PARST [33]. The analysis of the crystal packing and the drawing of the molecule were performed using the programs Mercury [34] and ORTEP [35]. Crystal and refinement data are summarized in Table 1. CCDC reference number 836567 contain the supplementary crystallographic data for **4**. These data can be obtained free of charge at www.ccdc.cam.ac.uk/data_request/cif.

Results and Discussion

Our initial investigation was addressed to **4** (Scheme 1), a product synthesised in large amounts from farnesyl acetate, a commercially available substance. X-ray analysis and detailed 2D-NMR studies allowed us to determine the preferential conformation of the molecule demonstrating that **4** adopts similar folded spatial arrangements both in solution and in the crystal.

Table 1 Crystal data and refinement details of **4**

Empirical formula	$\text{C}_{17}\text{H}_{30}\text{O}_6$
Formula weight	330.41
Temperature	297(2) K
Radiation, wavelength	$\text{MoK}\alpha$, 0.71073 \AA
Color	Colourless
Crystal size	0.50 \times 0.50 \times 0.50 mm
Crystal system, space group	Orthorhombic, <i>Pbca</i>
<i>a</i>	10.496(1) \AA
<i>b</i>	17.974(1) \AA
<i>c</i>	19.777(2) \AA
Volume	3725.4(6) \AA^3
Z, calculated density	8, 1.178 Mg/m^3
Absorption coefficient	0.088 mm^{-1}
Theta range for data collection	3.16–27.50 $^\circ$
Reflections collected/unique	21,126/4,147 [$R_{\text{int}} = 0.0297$]
Reflections with $I > 2\sigma(I)$	4,714
Number of parameters	219
Goodness-of-fit on F^2	1.073
<i>R</i> indices [$I > 2\sigma(I)$]	$R_1 = 0.0484$, $wR_2 = 0.1183$
<i>R</i> indices (all data)	$R_1 = 0.0796$, $wR_2 = 0.1390$
Largest diff. peak and hole	0.189, -0.199 e/\AA^3

In particular, a full set of 2D-NMR experiments [(COSY; HMBC, HSQC, HSQCTOCSY, TOCSY, NOESY (Fig. 3)] was carried out at 400 MHz in CDCl_3 solution that allowed the complete assignment of the ^1H and ^{13}C resonances to the pertinent protons and carbons of the molecule (Table 2).

A summary of significant NOESY correlations are shown in Fig. 4. In particular, the presence of a NOESY correlation between H-6 and Me-14 indicated that A and B THF rings are almost orthogonally oriented with respect to each other, while a NOESY cross peak between Me-13 and H₂-9 is in agreement with a staggered conformation of the 2-hydroxypropyl side-chain with respect to the A THF ring, the C-11 OH group being oriented toward the bottom side of the molecule. In addition, NOESY correlations

Fig. 3 NOESY spectrum of **4** in CDCl_3 (25 °C) recorded at 400 MHz

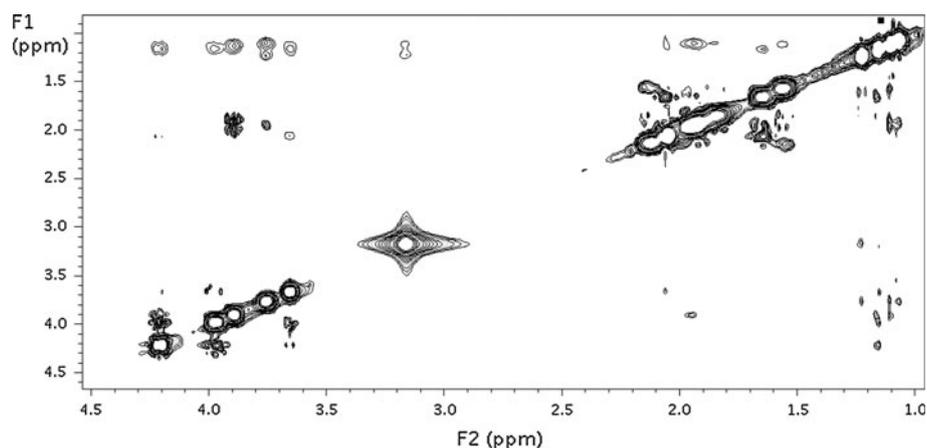


Table 2 Carbon and Proton assignments of **4**

	$\delta^{13}\text{C}$	$\delta^1\text{H}_a$ (J , mult.)	$\delta^1\text{H}_b$
1	65.7	4.20 (dd, 11.4, 2.7)	3.97 (dd, 11.4, 8.4)
2	75.0	3.65 (dd, 8.3, 2.8)	
3	83.9		
4	35.3	1.63 (m)	2.03 (m)
5	27.1	1.91 (m)	1.96 (m)
6	83.3	3.89 (dd, 6.3, 5.5)	
7	83.3		
8	34.4	2.12 (m)	1.55 (m)
9	25.7	1.94 (m)	1.85 (m)
10	84.8	3.75 (dd, 7.1, 7.1)	
11	71.1		
12	27.7	1.22 (s)	
13	25.3	1.06 (s)	
14	23.6	1.10 (s)	
15	21.7	1.15 (s)	
16	171.1		
17	20.9	2.04 (s)	
2 × OH		3.00–3.10	

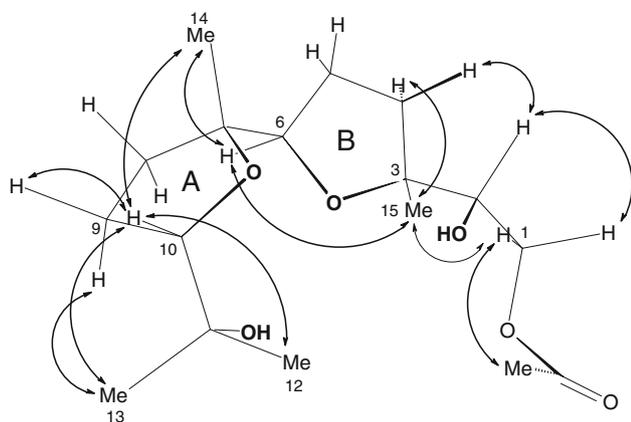


Fig. 4 Summary of some significant 400 MHz NOESY correlations for **4**

between Me-14 and H-10 and Me-15 and H-6 confirmed the previously determined *cis* configuration of the two THF rings in **4**. The rest of the NOESY contacts indicated that the two THF oxygens and the two hydroxyl oxygens all protrude toward the inner part of the molecule that therefore possess a folded conformation (Fig. 4).

The X-ray molecular structure of **4** is reported in Fig. 5, selected torsion angles are given in Table 3. It is clearly shown that in both THF rings the 2,5 substituents are *cis* configuration with respect to each other. Owing to this stereochemistry and to the torsion angles around the bonds C2–C3, C6–C7 and C10–C11, the molecule adopts a bent, horse-shoe-type shape, with hydroxy oxygens placed in the inner part and the two THF ring oriented substantially perpendicular to each other (the dihedral angle between the mean planes is $88.53(7)^\circ$), as found in the solution conformation. In addition, the staggered conformation of the 2-hydroxypropyl tail with respect to the adjacent THF ring is also observed in the X-ray molecular structure. So, the

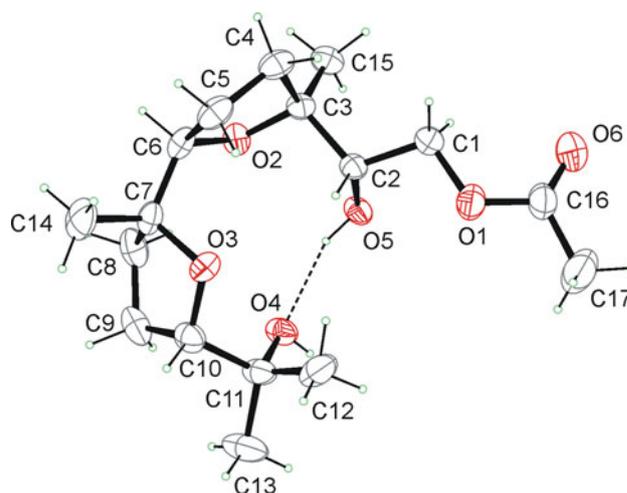


Fig. 5 ORTEP drawing of **4**. Thermal ellipsoids are drawn at 30% probability level. The intramolecular hydrogen bond is shown as dashed line

Table 3 Selected torsion angles ($^{\circ}$) of **4** (e.s.d.'s in parentheses)

O5–C2–C3–O2	52.4(2)
O2–C6–C7–O3	67.1(2)
O3–C10–C11–O4	–69.2(2)

Table 4 Hydrogen bonding geometry of **4** (e.s.d.'s in parentheses)

D–H...A	D–H (Å)	H...A (Å)	D...A (Å)	D–H...A ($^{\circ}$)
O4–H...O5 ^a	0.82(2)	1.99(2)	2.800(2)	170(2)
O5–H...O4 ^b	0.86(2)	1.97(2)	2.781(2)	157(2)
C13–H...O6 ^c	0.96	2.502	3.451(3)	170

Symmetry code: ^a $-x + 2, -y + 1, -z + 1$; ^b x, y, z ; ^c $-x + 3/2, -y + 1, z + 1/2$

accurate X-ray molecular structure of **4** in the solid state confirms the general features of the molecular structure in solution as deduced by 2D-NMR experiments.

The molecular conformation in the solid state is further stabilized by an intramolecular H bond between O5–H donor and O4 acceptor (Table 4). Ring puckering coordinates of the two THF rings are given in Table 5, and they indicate that the ring (O2, C3, C4, C5, C6) is in envelope conformation, while the other is basically in twisted conformation.

In the crystals, molecules are associated in dimers through intermolecular H bonding between O4–H donor

Table 5 Ring puckering coordinates of **4** (e.s.d.'s in parentheses)

Ring	q_2 (Å)	φ_2 ($^{\circ}$)
(O2, C3, C4, C5, C6)	0.353(2)	6.1(2)
(O3, C7, C8, C9, C10)	0.367(2)	195.4(3)

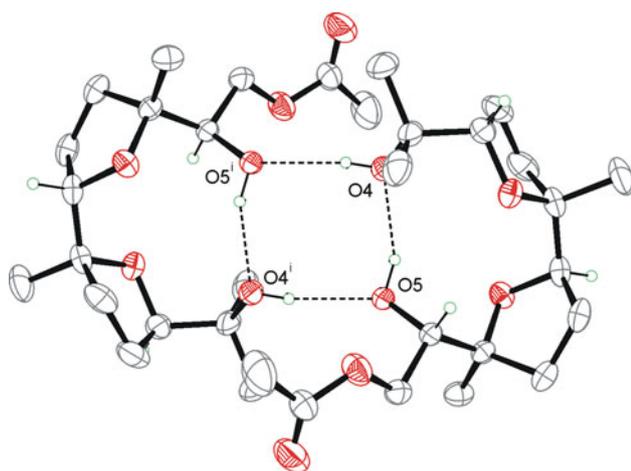


Fig. 6 Dimer of H-bonded molecules through a crystallographic inversion center. Dashed lines marks the $R_4^4(8)$ motif ($i = -x + 2, -y + 1, -z + 1$). Methyl and methylene H atoms are not shown for clarity

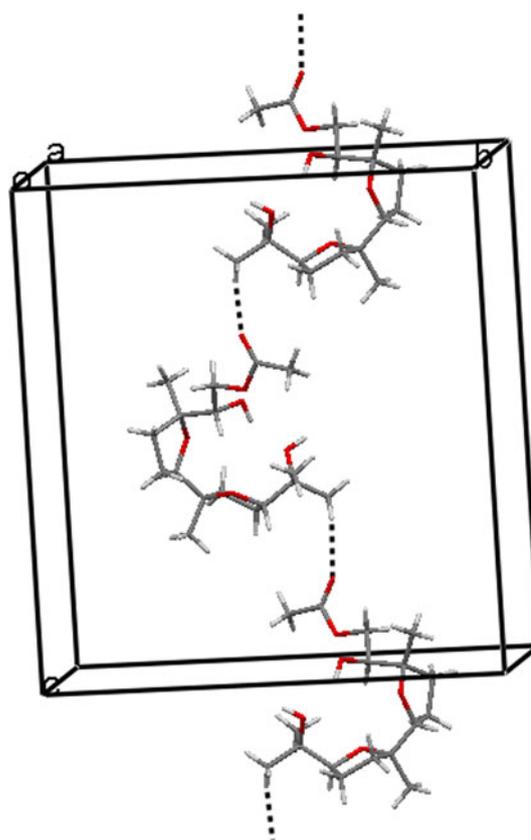


Fig. 7 Partial crystal packing of **4** showing a chain formed by weak H bonding

and O5 acceptor. In this way, $R_4^4(8)$ ring patterns are formed across crystallographic inversion centres, Fig. 6.

A head-to-tail arrangement of molecules in chains running along c is achieved through weak H bonding between C13–H donor and carbonyl oxygen acceptor O6, see Table 4 and Fig. 7.

Conclusions

In summary, in the present study we have determined the conformational arrangement of a bis-THF diol compound. Both the solution and the solid-state conformation is folded as observed for similar bis-THF compounds possessing cytotoxic activity and cation transport ability. This is a promising result that prompts us to carry out further study on this compound to determine its possible cation complexation ability as well as its cytotoxic activity.

Supplementary Material

All crystallographic data have been deposited at the Cambridge Crystallographic Data Centre (CCDC). Deposition number is CCDC 836567. These data can be obtained free of charge from www.ccdc.cam.ac.uk/data_request/cif.

Acknowledgments The authors thank the Centro Interdipartimentale di Metodologie chimico Fisiche (CIMCF) of University of Naples “Federico II” for NMR and X-ray facilities.

References

1. Jalce G, Franck X, Figadere B (2009) *Tetrahedron Asymmetry* 20:2537–2581
2. Wolfe JP, Hay MB (2007) *Tetrahedron* 63:261–290
3. Piccialli V (2007) *Synthesis* 17:2585–2607
4. Harmange J-C, Figadere B (1993) *Tetrahedron Asymmetry* 4:1711–1754
5. Boivin TLB (1987) *Tetrahedron* 43:3309–3362
6. Pressman BC, Harris EJ, Jagger WS, Johnson JH (1967) *Proc Natl Acad Sci USA* 58:1949–1956
7. Hioki H, Yoshio S, Motosue M, Oshita Y, Nakamura Y, Mishima D, Fukuyama Y, Kodama M, Ueda K, Katsu T (2004) *Org Lett* 6:961–964 and references therein
8. Morita H, Kishi EK, Takeya K, Itokawa H, Iitaka Y (1993) *Phytochemistry* 34:765–771
9. Morimoto Y, Iwai T, Yoshimura T, Kinoshita T (1998) *Bioorg Med Chem Lett* 8:2005–2010
10. Sasaki S, Naito H, Maruta K, Kawahara E, Maeda M (1994) *Tetrahedron Lett* 35:3337–3340
11. Sasaki S, Maruta K, Naito H, Sugihara H, Iratani K, Maeda M (1995) *Tetrahedron Lett* 36:5571–5574
12. Peyrat J-F, Figadere B, Cave A, Mahuteau J (1995) *Tetrahedron Lett* 36:7653–7656
13. Peyrat J-F, Mahuteau J, Figadere B, Cave A (1997) *J Org Chem* 62:4811–4815
14. Sasaki S, Maruta K, Naito H, Maemura R, Kawahara E, Maeda M (1998) *Tetrahedron* 54:2401–2410
15. Bifulco G, Caserta T, Gomez-Paloma L, Piccialli V (2002) *Tetrahedron Lett* 43:9265–9269; corrigendum (2003) *Tetrahedron Lett* 44:3429
16. Bifulco G, Caserta T, Gomez-Paloma L, Piccialli V (2003) *Tetrahedron Lett* 44:5499–5503
17. Caserta T, Piccialli V, Gomez-Paloma L, Bifulco G (2005) *Tetrahedron* 61:927–939
18. Piccialli V, Caserta T, Caruso L, Gomez-Paloma L, Bifulco G (2006) *Tetrahedron* 62:10989–11007
19. Piccialli V, Borbone N, Oliviero G (2008) *Tetrahedron* 64:11185–11192
20. Piccialli V, Oliviero G, Borbone N, Tuzi A, Centore R, Hemminki A, Ugolini M, Cerullo V (2009) *Org Biomol Chem* 7:3036–3039
21. Piccialli V, Zaccaria S, Borbone N, Oliviero G, D’Errico S, Hemminki A, Cerullo V, Romano V, Tuzi A, Centore R (2010) *Tetrahedron* 66:9370–9378
22. Albarella L, Piccialli V, Smaldone D, Sica D (1996) *J Chem Res Synopses* 9:400–401
23. Piccialli V, Sica D, Smaldone D (1994) *Tetrahedron Lett* 35:7093–7096
24. Albarella L, Giordano F, Lasalvia M, Piccialli V, Sica D (1995) *Tetrahedron Lett* 36:5267–5270
25. Piccialli V, Caserta T (2004) *Tetrahedron Lett* 45:303–308
26. Piccialli V, Cavallo N (2001) *Tetrahedron Lett* 42:4695–4699
27. de Champdorè M, La Salvia M, Piccialli V (1998) *Tetrahedron Lett* 39:9781–9784
28. Blessing RH (1995) *Acta Crystallogr A* 51:33–38
29. Altomare A, Burla MC, Camalli M, Cascarano GL, Giacovazzo C, Guagliardi A, Moliterni GG, Polidori G, Spagna R (1999) *J Appl Crystallogr* 32:115–119
30. Sheldrick GM (2008) *Acta Crystallogr A* 64:112–122
31. Farrugia LJ (1999) *J Appl Crystallogr* 32:837–838
32. Cremer D, Pople JA (1975) *J Am Chem Soc* 97:1354
33. Nardelli M (1995) *J Appl Crystallogr* 28:659
34. Macrae CF, Bruno IJ, Chisholm JA, Edgington PR, McCabe P, Pidcock E, Rodriguez-Monge L, Taylor R, van de Streek J, Wood PA (2008) *J Appl Cryst* 41:466–470
35. Farrugia LJ (1997) *J Appl Cryst* 30:565

Insight into Pyridinium Chlorochromate Chemistry: Catalytic Oxidation of Tetrahydrofuran Compounds and Synthesis of Umbelactone

Vincenzo Piccialli,^{*[a]} Sabrina Zaccaria,^[a] Giorgia Oliviero,^[b] Stefano D'Errico,^[b] Valentina D'Atri,^[b] and Nicola Borbone^[b]

Keywords: Oxidation / Oxygen heterocycles / Reaction mechanisms

The catalytic system PCC (cat.)/H₅IO₆ has been used to oxidise mono- and polytetrahydrofuran compounds. New oxidative pathways are disclosed. 2,2,5-Trisubstituted THF rings are converted into dicarbonyl compounds through oxidative cleavage of the C2–C3 bond. Cyclic enol ethers appear to be intermediate species in this process. Oxidation of 2,2,5-trisubstituted α -keto-THF compounds proceeds with the oxidative cleavage of the C2(THF)–C=O bond to give 1,4-diketones with degraded carbon backbones. Attack of the oxidant on 2,5-disubstituted THF rings leads to 1,4-diketones

containing intact THF carbon frameworks. Oxidation of complex poly-THF substrates, containing up to five THF rings, allows access to new poly-THF compounds through regioselective THF oxidation along the poly-THF backbones. A mechanistic explanation of the new processes consistent with the reported reactivity of PCC and the isolation of some minor products of the process is provided. The synthesis of racemic umbelactone, an antiviral natural butenolide metabolite, has been carried out by use of the developed chemistry.

Introduction

Pyridinium chlorochromate (PCC) is a well-known reagent employed in an array of oxidising processes.^[1] This oxidant is usually used either in stoichiometric or in excess amounts to oxidise a variety of functional groups. In contrast, the use of PCC in catalytic amounts has received little attention to date^[2,3] even though chromium(VI) species are known to be carcinogenic and environmentally hazardous and the use of substoichiometric amounts of such a reagent would therefore be highly desirable. In this respect, the catalytic system PCC (cat.)/H₅IO₆ is particularly appealing, although it has been employed in only a few cases.^[2,3] As a part of our ongoing interest in oxidative processes mediated by transition metal oxo species^[4] and in particular in the synthesis of new THF-containing compounds,^[5] we report here on the oxidation of mono- and poly-THF compounds with the PCC (cat.)/H₅IO₆ (chlorochromatoperiodate, CCP) system, which has led to the discovery of new types of useful oxidative transformations involving THF ring systems. A similarity between the oxidative behaviour of PCC and that of RuO₄, which further supported our recent findings in the chemistry of these oxo species, has also emerged.^[5a,5b]

Results and Discussion

Oxidation of Poly-THF Compounds – Preliminary Results

In a preliminary experiment devised to test the reactivities of tetrahydrofuran-containing substances with PCC, penta-THF dibenzoate **2** (Scheme 1), containing five differently substituted and configured THF rings, was investigated. The synthesis of **2** was accomplished by a RuO₄-catalysed oxidative polycyclization protocol starting from squalene, previously developed in our laboratory,^[5i,5j] followed by benzylation of the tertiary hydroxy groups in the initially formed penta-THF diol **1** under standard conditions (BzCl, DMAP, CH₂Cl₂).

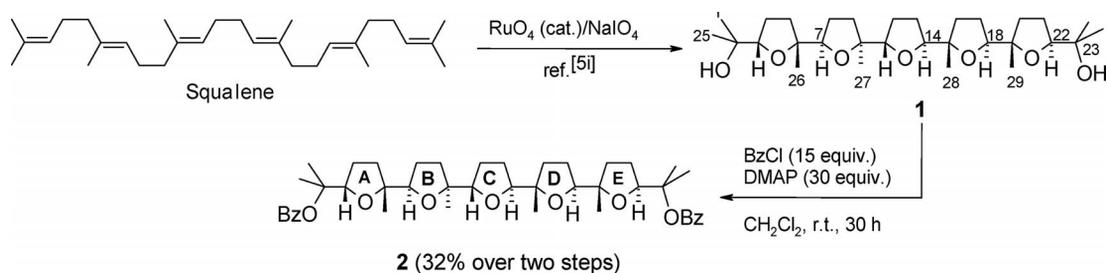
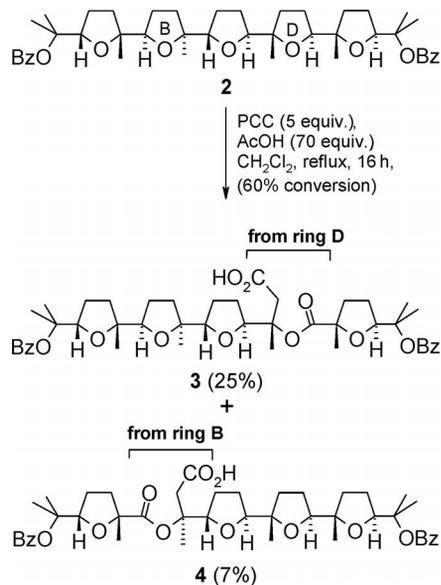
Treatment of **2** with PCC under classical oxidising conditions (excess PCC/AcOH, CH₂Cl₂, Scheme 2)^[5b] gave the two isomeric tetra-THF compounds **3** and **4** (overall 32% yield), originating from the oxidative cleavage of the C2–C3 bonds in either the D or the B THF rings of **2**, respectively.

The rest of the material for mass balance was made up of polar substances possibly derived by side- or over-oxidation processes due to the presence of various THF rings prone to attack by the oxidant. The structures of **3** and **4** were determined by a full set of 2D NMR experiments recorded at high fields. Although the process proceeded with a rather low yield and could not be forced to completion under these conditions, the ability of PCC to oxidise the THF ring had been confirmed.

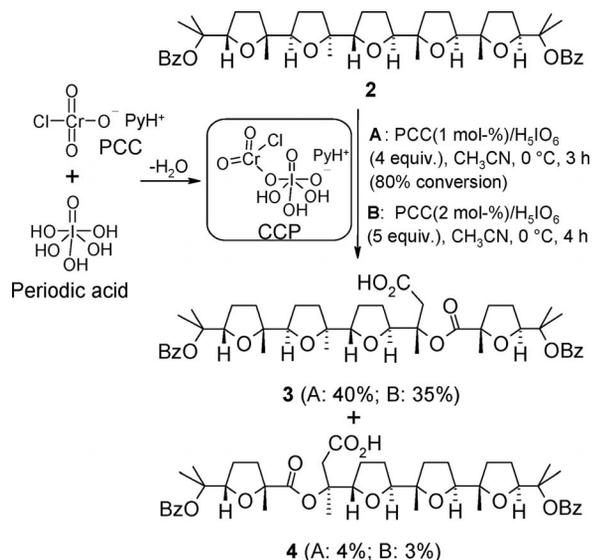
Our next goal was to test the reactivity of **2** in the presence of the catalytic system PCC (cat.)/H₅IO₆ (Scheme 3). From previous studies^[6] it is thought that the combination of PCC and H₅IO₆ generates chlorochromatoperiodate

[a] Dipartimento di Scienze Chimiche, Università degli Studi di Napoli “Federico II”, Via Cintia, 80126 Napoli, Italy
 Fax: +39-081-674393
 E-mail: vinpicci@unina.it

[b] Dipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli “Federico II”, Via D. Montesano 49, 80131 Napoli, Italy

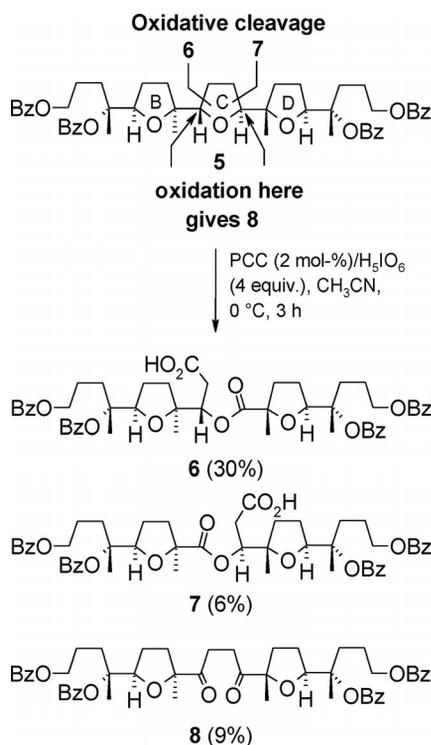
Scheme 1. Synthesis of penta-THF dibenzoate **2**.Scheme 2. Cleavage products from the oxidation of penta-THF dibenzoate **2** with PCC/AcOH.

4, in the reaction mixture (NMR, HPLC and MS analyses). Smaller amounts of these side products were also formed under conditions A.

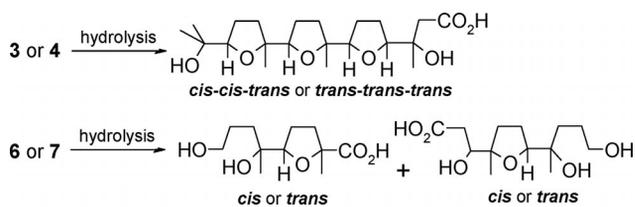
Scheme 3. Oxidation of penta-THF dibenzoate **2** with PCC (cat.)/ H_5IO_6 .

(CCP, Scheme 3), an oxidising agent more powerful than PCC itself. It was hoped that the use of a more reactive reagent such as this might possibly reduce reaction times and force the process to completion. The oxidation was initially carried out in CH_3CN at $0\text{ }^\circ\text{C}$ with 1 mol-% of PCC and 4 equiv. of periodic acid (Scheme 3, conditions A). The process was indeed much cleaner under these conditions and overall improved yields of acids **3** and **4** (**3**: 40%; **4**: 4%) were obtained in a shorter time (3 h). Importantly, the two isomers were now obtained in a ca. 10:1 ratio in favour of **3**. With regard to the attack of the oxidant at the D ring of the poly-THF system of **2**, the observed enhanced regioselectivity was probably due both to the increased steric demand of the CCP and to the different configurations of the involved rings (ring B: *cis*; ring D: *trans*). Under these conditions the process was still incomplete and stopped at 80% conversion. Addition of a further 1 mol-% PCC and 1 equiv. of H_5IO_6 (conditions B) forced the process to completion at the expense of a slight reduction in the overall yield (**3** + **4** 38%; **3/4** 10:1) presumably due to the over-oxidation of **3** and **4** themselves, as indicated by the presence of polar products, probably related to **3** and/or

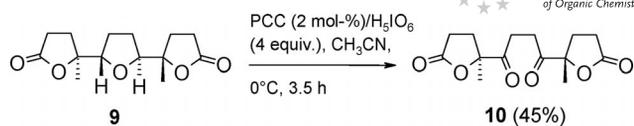
To test the above oxidising system further, the oxidation of tris-THF tetrabenzoate **5** (Scheme 4), a structurally “simplified” substrate related to **2**, was then carried out. Compound **5** was obtained by benzylation of the corresponding tetrol (prepared in turn from penta-THF **1** through a short degradative sequence developed in a previous study conducted by our group^[5f]). Interestingly, in this case the three major products **6–8**, two of them – acids **6** and **7** – structurally related to **3** and **4** and all originating from attack of the oxidant at the central THF ring, were obtained. In this case 4 equiv. of H_5IO_6 were sufficient to drive the process to completion. 2D NMR experiments were carried out to determine the stereostructures of **6–8**. A new feature of this process was the formation of the 1,4-diketone **8** through a different oxidative route featuring the oxidation of both the angular carbons of the central THF ring in **5**, not observed in the oxidation of penta-THF **2**. 1,4-Diketones are easily transformed into five-membered heterocycles, so substances such as **8** might be synthetically useful to access new types of rare mixed poly-THF/heterocycle compounds.

Scheme 4. Oxidation of tris-THF tetrabenzoate **5**.

The parent penta-THF **2** can be obtained in a single step in multi-gram amounts from cheap and commercially available squalene, so the oxidative degradation of its penta-THF backbone can allow facile access to stereochemically definite and functionalised *cis-cis-trans* and *trans-trans-trans* tris-THF compounds by hydrolysis of the inter-THF ester functions in either **3** or **4** (Scheme 5). Likewise, new functionalised *cis* or *trans* mono-THF systems can be obtained from **6** and **7** by hydrolysis (Scheme 5). Given literature precedents, it can be presumed that at least some of these substances should be useful for further cytotoxic activity studies and also for synthetic purposes.^[7]

Scheme 5. Functionalised new all-*threo* mono- and tris-THF systems accessible through hydrolysis of poly-THF compounds **3**, **4**, **6** and **7**.

A further simplification of the poly-THF substrate was achieved by conversion of the tetrol corresponding to **5** into bis-lactone **9** (Scheme 6) through a previously developed degradative sequence.^[5f] Interestingly, oxidation of this compound under the same conditions as used for **5** led to 1,4-diketone **10** (45% yield) through the oxidative opening of the THF ring, whereas no acid corresponding to compounds **6/7** was observed.

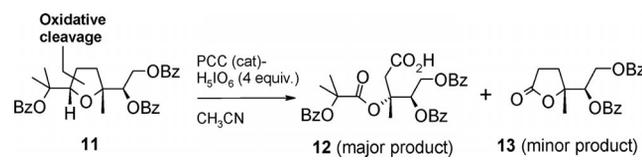
Scheme 6. Oxidation of THF-dilactone **9**.

The above results indicated that the oxidation of complex poly-THF compounds containing both di- and trisubstituted THF rings proceeds through two main routes. Oxidative cleavage of the C2–C3 bond is preferred when a 2,2,5-trisubstituted THF ring is involved, as in the oxidation of **2**. A second route leads to 1,4-diketones, such as **8** and **10**, through the attack of the oxidant at a 2,5-disubstituted THF ring and oxidation of both its angular carbon atoms, as in the oxidations of **5** and **9**.

Oxidation of Mono-THF Compounds

With the aim of making the above process synthetically useful, the oxidation of some mono-THF compounds, based on the di- and trisubstituted THF motifs present in previously studied poly-THF compounds, was tested. *cis*-THF **11** (Table 1), obtained by benzylation of the corresponding diol, in turn synthesized by RuO₄-catalyzed oxidative cyclization of geranyl benzoate,^[8] was initially chosen as a model compound to find the best conditions for the oxidative process. Oxidation of **11** under previously employed conditions [PCC (2 mol-%), H₅IO₆ (4 equiv.), 0 °C] gave acid **12**, analogous to compounds **3/4** and **6/7** obtained from more complex substrates, in a 54% yield (Table 1, Entry 1), although longer reaction times were required. The process was then carried out at room temp. with the same amounts of catalysts and co-oxidant, leading to **12** in a similar 55% yield in only 3 h (Entry 2). Improved yields (80%) and further reduced reaction times (45 min) were achieved by increasing the catalyst amount to 5% (Entry 3) and further reduced times (45 min) were achieved by increasing the catalyst amount to 10% (Entry 4). Further increasing of the catalyst up to stoichiometric amounts (50–

Table 1. Optimization of the process on a model 2,2,5-trisubstituted THF compound.



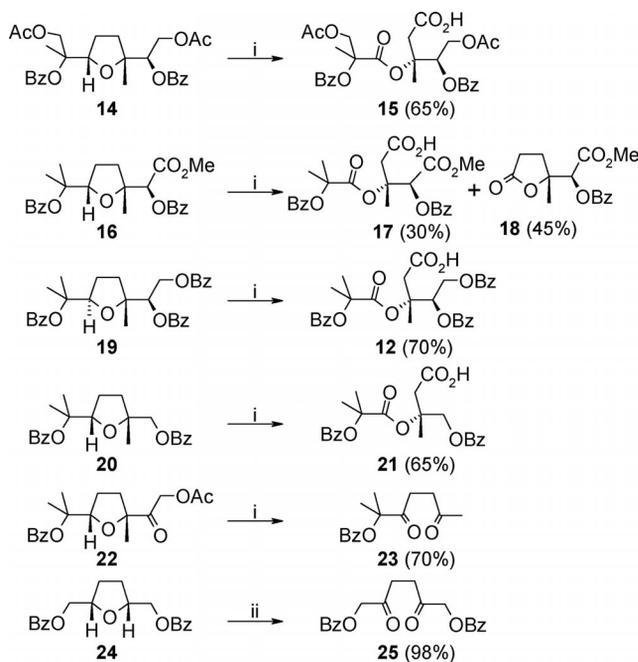
Entry	PCC [mol-%]	H ₅ IO ₆ [equiv.]	Time	Yield of 12 ^[a] [%]
1 ^[b]	2	4	19 h	54
2	2	4	3 h	55
3	5	4	45 min	80 ^[c]
4	10	4	30 min	78
5	50	4	30 min	62
6	100	4	30 min	37

[a] Estimated by ¹H NMR examination of the crude mixture.

[b] Carried out at 0 °C. [c] Isolated yield.

100%) only resulted in diminished yields (Entries 5 and 6). Finally, the effect of addition of water to the process was tested. Although the related system $\text{CrO}_3(\text{cat.})/\text{H}_5\text{IO}_6$ has been successfully used to oxidise alcohols cleanly in wet CH_3CN (0.75 vol.-% water)^[9] in our case we have observed that addition of water (2–5 vol.-%) is detrimental, causing the process to stop at 10–20% conversion.

Oxidation of mono-THF compounds structurally related to **11** under the optimized conditions was then carried out (Scheme 7). *cis*-THF compounds **14**, **16** and **24** were synthesized by benzylation of the corresponding THF diols, in turn obtained by RuO_4 -catalysed oxidative cyclization of *trans,trans*-2,6-dimethyl-2,6-octadiene-1,8-diol diacetate,^[5j] methyl geranate^[5j] and hexa-1,5-diene,^[8] respectively, as reported. *trans*-THF **19** was obtained by benzylation of the corresponding THF diol, a minor product from the oxidative cyclization of geranyl benzoate.^[10] THF **20** was synthesized from **11** through a short sequence. In particular, selective removal of both the primary and secondary benzoates in **11** with K_2CO_3 in MeOH proceeded cleanly to give the corresponding diol. Oxidative cleavage of the diol system by treatment with silica-supported NaIO_4 ^[11] in CH_2Cl_2 , followed by borohydride reduction and benzylation, afforded dibenzoate **20** in 43% yield (over four steps). Ketone **22** was obtained by TPAP-catalysed oxidative cyclization of geranyl acetate^[5g] followed by benzylation.



Scheme 7. Oxidation of functionalised 2,5-disubstituted and 2,2,5-trisubstituted THF compounds. i) PCC (5 mol-%), H_5IO_6 (4 equiv.), CH_3CN , room temp., 1 h; ii) PCC (2 mol-%), H_5IO_6 (4 equiv.), CH_3CN , room temp., 40 min.

CCP oxidation of *cis*-THF diacetate **14** and *trans*-THF **19** gave acids **15** and **12**, respectively, in 65–70% yields. The methoxycarbonyl derivative **16** unexpectedly afforded the corresponding acid in a diminished 30% yield, the major product of the process being lactone **18**, analogous to **13**,

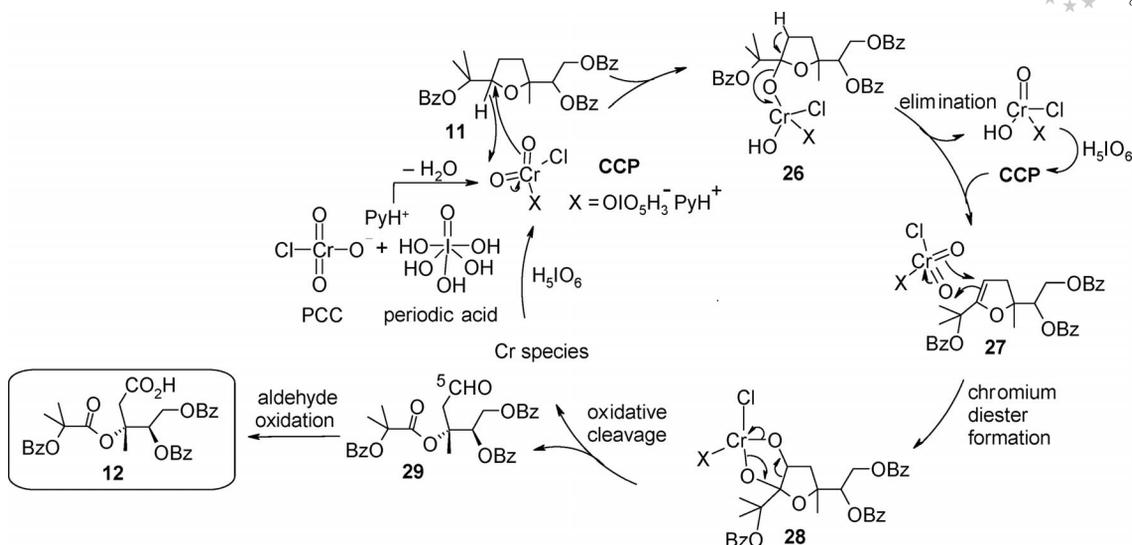
derived from the oxidative removal of the dimethylated left-hand side chain. This side process is unusual but is reminiscent of the oxidative cleavage of a similar substrate.^[3] This type of lactone is also the main side product of the process carried out on all the other substrates. Similarly, the THF derivative **20** gave acid **21** in 65% yield.

In another experiment, α -keto THF **22** was oxidised. In this case the presence of a keto group adjacent to the THF ring induced removal of the side chain through the oxidative cleavage of the C(THF)2–C=O bond to give 1,4-diketone **23** in good yields.

Finally, the reactivity of the 2,5-disubstituted *cis*-mono-THF **24** was tested. Pleasingly, 1,4-diketone **25** was obtained in a nearly quantitative yield (98%) even when the amount of PCC was reduced to 2 mol-%. This process is particularly appealing because it allows, when coupled with the oxidative cyclization of 1,5-dienes, the transformation of the latter into bis-ketols in a regioselective manner. This experiment and the oxidation of the more complex disubstituted mono-THF **9** demonstrated that for substrates containing only a 2,5-disubstituted THF ring the preferred route is the one leading to 1,4-diketones.

Isolation of Minor Products and Mechanistic Considerations

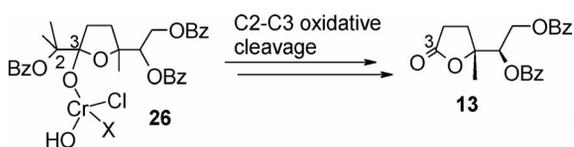
A plausible mechanistic route based on the known reactivity of PCC and related oxo species such as RuO_4 ^[5a] was hypothesised to explain the formation of the above compounds. In particular, a pathway leading to compound **12** is shown in Scheme 8, but it also applies to the formation of the analogous compounds **3/4**, **6/7**, **15**, **17** and **21**. It is likely that the process begins with the attack of CCP, formed by elimination of water from PCC and periodic acid, at the angular C–H bond of the THF ring to give the mixed chromium ester intermediate **26**. This step is conceivably a [3+2] addition of the O=Cr=O portion of the oxidant to the angular C–H bond of the THF ring, as also hypothesised for the RuO_4 oxidation of neoisocedrane oxide,^[12] a sesquiterpene containing a THF ring. It is reported that the oxidation of the THF ring in this substance proceeds through the intramolecular insertion of an oxoruthenium bond into the angular C–H bond of the THF ring. In addition, it has been suggested that attack of transition metal oxo species such as OsO_4 and RuO_4 to C–H bonds of alkanes also proceeds through a [3+2] addition of a C–H bond across an O=M=O unit,^[13] through a mechanism analogous to the one now widely accepted for alkene bis-hydroxylation. In the next step, chromium ester **26** would give rise to cyclic enol ether **27** through an elimination step delivering a low-valent chromium species that would then be reoxidised to CCP by H_5IO_6 . However, it cannot be excluded that the first step of the process could proceed through hydride abstraction from the angular C–H bond (not shown) by the same O=Cr=O portion of CCP to give an intermediate carbocation (actually an oxonium ion) which could then collapse to form enol ether **27**. This would



Scheme 8. A plausible catalytic cycle explaining the formation of acid **12**.

be consistent with the mechanism proposed by Lee and van den Engh for the RuO_4 oxidation of tetrahydrofuran to γ -butyrolactone,^[14] in which the initially formed carbonium ion and Ru^{VI} oxo species recombine to give an intermediate ester similar to **26**.

Whereas PCC is unable to cleave isolated (unactivated) carbon-carbon double bonds, oxidative cleavages of both cyclic and acyclic enol ethers to esters with PCC have been reported.^[15] In particular, the conversion of **27** to the final dicarbonyl compound **12** is thought to proceed through the formation of the chromium diester intermediate **28**, by a second [3+2] addition step. Oxidative cleavage to afford aldehyde **29**, followed by oxidation, eventually generates acid **12**. The conversion of **27** to **29** agrees well with the previously observed oxidative cleavage of cyclic enol ethers lacking α -hydrogen atoms.^[15a,15c] Finally, formation of lactone **13**, the main side-product of the oxidation of **11** (Table 1), can be seen to originate from the oxidative removal of the dimethyl benzoate side chain. It has previously been observed that THF rings bearing a tertiary (free) alcohol function undergo similar oxidative cleavage with CCP to give γ -lactones^[3] such as **13**, via C-2 chromium ester intermediates. It is conceivable that a similar route could be responsible for the C2–C3 oxidative cleavage in a chromium ester intermediate possibly formed by further evolution of **26** (Scheme 9). This interaction is supported by the complete absence of hydrolysis of the primary and secondary benzoate functions present in the same molecule.



Scheme 9. Formation of lactone **13** by side-chain removal in intermediate **26**.

To support the mechanistic path shown in Scheme 8, a careful examination of the minor products formed during the oxidation of penta-THF **2**, both with PCC/AcOH and with CCP, was undertaken. HPLC analyses (direct and reversed-phase modes) led to the identification of the three side-products **30–32** (Figure 1) in both processes. Their structures give strong support for the proposed mechanism and their formation could be interpreted through routes consistent with those given for the main oxidation products.

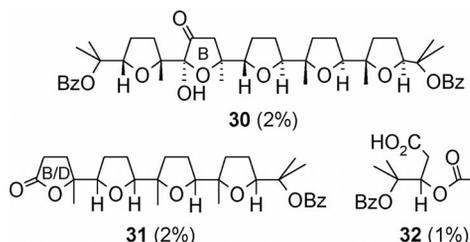
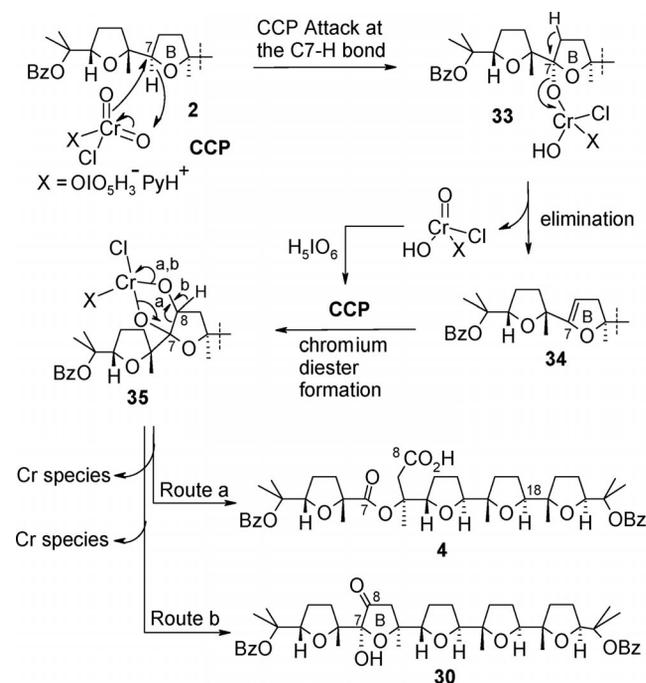


Figure 1. Minor products obtained from the oxidation of penta-THF dibenzoate **2**.

In particular, the presence of the oxidised B ring in **30** strongly suggested that it is formed on the pathway leading to acid **4**, as shown in Scheme 10. On this pathway, acid **4** and ketol **30** would be formed by attack of CCP at the angular C7–H bond in the THF ring B of **2** via intermediates **33**, **34** and **35** as seen for the oxidation of **11**. Compound **35** would follow route a to give acid **4** through the oxidative cleavage of the C7–C8 bond. Alternatively, its oxidative opening would give ketol **30** through route b. The isolation of **30** furnishes the first indirect support for the previously hypothesised^[16] formation of chromium diester intermediates, such as **35**, during the oxidative cleavage of enol ethers with PCC, and strongly supports the involvement of enol ether **34**.

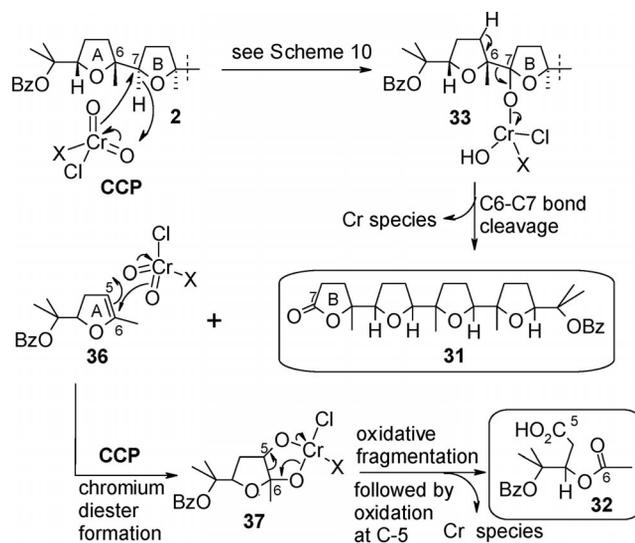
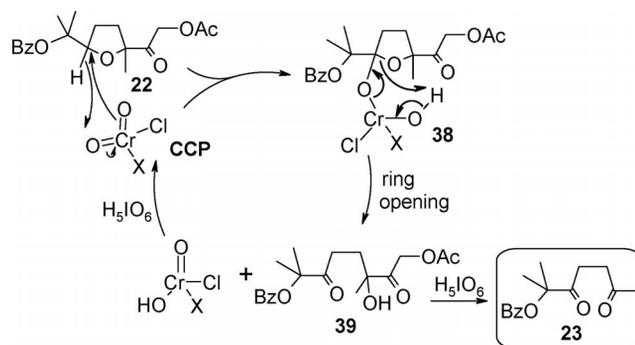
Scheme 10. Formation of compounds **4** and **30**.

Acid **3** (see Scheme 3) would in turn originate in a similar way by attack of CCP at the C18–H of the THF ring D in **2** whereas acids **6** and **7** (see Scheme 4) would be formed by attack of the oxidant at the central C ring in the tris-THF **5**.

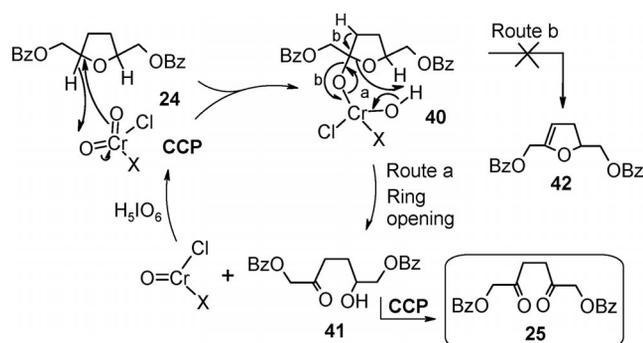
With regard to the minor compounds **31** and **32** (Figure 1), it seems likely that they could originate from **2** in the same route leading to **3** or **4** (Scheme 11). In particular, the oxidative cleavage of the C6–C7 bond in **2** could occur in the chromium ester intermediate **33** through an elimination step now involving both the A and the B rings. This step accounts well for the formation of the C-7 lactone function in **31** with generation once again of an enol ether species, compound **36**, that is then oxidatively cleaved to the dicarbonyl compound **32** via chromium diester **37**, as previously seen. The configuration of **31** has not been determined, so its formation might follow either the depicted route, through attack of the oxidant at the C7-H and oxidation of ring B, or an analogous pathway in which the D ring is oxidised by attack at the C18–H bond (pathway not shown) and cleavage of the C18–C19 bond connecting rings D and E.

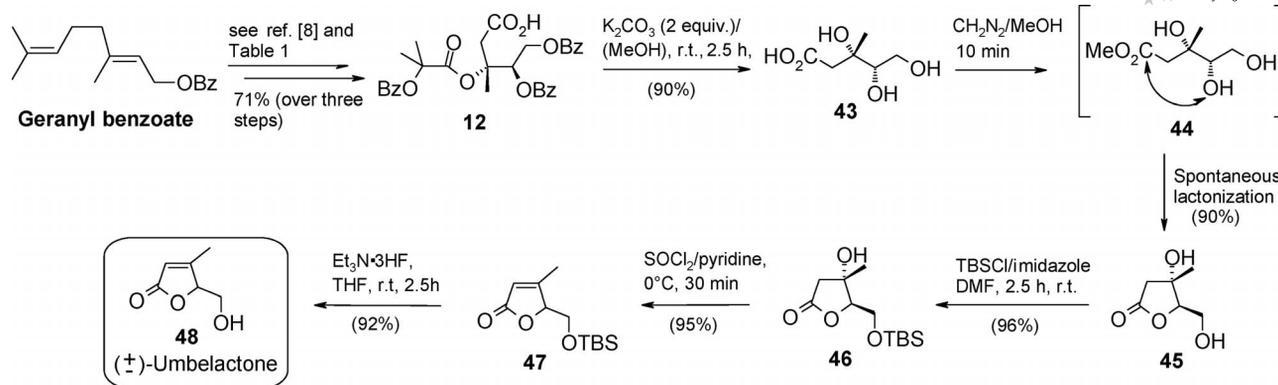
By the same line of reasoning, the formation of diketone **23** from **22** (Scheme 12) can be explained by the assumption that the initially formed chromium ester **38** collapses to the dicarbonyl species **39**, the α -ketol portion of which undergoes oxidative cleavage to give **23** through a known process.

The formation of 1,4-diketone **25** from **24** is shown in Scheme 13. In particular, CCP could attack one of the angular C–H bonds in **24** in the usual way to give chromium ester **40**. Opening of this would then follow to give ketol **41** (route a) with simultaneous production of a chro-

Scheme 11. A mechanistic explanation for the formation of compounds **31** and **32**.Scheme 12. A catalytic cycle for the formation of diketone **23**.

mium species, the oxidation of which would regenerate CCP, closing the catalytic cycle. CCP oxidation of the alcohol function in **41**^[2a] would eventually give 1,4-diketone **25**. Formation of the related 1,4-diketones **8** (Scheme 4) and **10** (Scheme 6) can be explained in a similar manner. It should be noted that the possible competing path giving rise to an enol ether species (compound **42**, Scheme 13, route b) from **40** is completely suppressed in this case. It is worth noting that the chemical behaviour of PCC in this

Scheme 13. A catalytic cycle for the formation of 1,4-diketone **25**.



Scheme 14. Synthesis of umbelactone.

transformation parallels that displayed by RuO_4 . In fact, in a similar transformation, 2,5-dimethyl-tetrahydrofuran is reported to undergo oxidative opening by the $\text{RuO}_2(\text{cat.})/\text{NaIO}_4$ system to afford hexane-2,5-dione.^[17]

Synthesis of Umbelactone

As an application of the developed oxidative procedure, a short synthesis of racemic umbelactone was carried out. Umbelactone is a γ -butenolide natural product, isolated from ethanolic extracts of *Memecylon umbellatum* Brum,^[18] that exhibited antiviral activity against the Ranikhet disease virus as well as spasmolytic and antiamphetamine activity. This substance has been the subject of various syntheses both in racemic and in enantiopure form.^[19]

Our oxidation of trisubstituted THFs such as **11** described above allows access to elaborated 3,4,5-trioxygenated acids. In particular, compound **12** derived from **11** was seen as a good starting product for the synthesis of umbelactone (Scheme 14). Tribenzoate **12** was converted into trihydroxyacid **43** in high yield by treatment with K_2CO_3 in MeOH. Successive treatment of crude **43** with CH_2N_2 gave transient methyl ester **44**, which spontaneously cyclised to the γ -lactone **45** in 90% yield (81% from **12**). The primary hydroxy group in **45** was then protected as the TBS ether (compound **46**) by treatment with TBSCl/imidazole in DMF (96% yield).^[20] Dehydration of **46** was accomplished with SOCl_2 /pyridine to give silylated umbelactone **47** in 95% yield. Lastly, desilylation of **47** with $\text{Et}_3\text{N}\cdot\text{3HF}$ in THF cleanly gave umbelactone **48** (92% yield).

Our synthetic route to umbelactone is short and highly efficient and does not require carbon-carbon bond-forming steps. In addition, it can be made enantioselective to give both the umbelactone enantiomers because mono-THF tribenzoate **11** (Table 1), the immediate precursor of **12**, can be obtained in both enantiomeric forms from commercially available geraniol in high yields and enantiomeric purity.^[21] Studies directed towards this goal, as well as exploitation of the above chemistry for the preparation of further functionalised butenolides, are ongoing.

Conclusions

In summary, we have examined the oxidative behaviour of some mono- and poly-THF compounds with the catalytic system PCC (cat.)/ H_5IO_6 . Novel oxidative pathways leading to the modification/degradation of the THF ring, as well as of the poly-THF backbone, have been disclosed. Plausible hypotheses, consistent with the known reactivity of PCC, have been put forward to explain the new oxidative routes. In particular, attack of the oxidant at the angular C–H bond of the target THF ring with formation of a mixed chromium ester is thought to be the first event. Two main routes then follow. Oxidative carbon-carbon bond cleavage, to give dicarbonyl products, occurs when 2,2,5-trisubstituted THFs are oxidised. The isolation of some minor side products of the process strongly supports the formation of cyclic enol ether intermediates in this transformation. With 2,5-disubstituted mono-THFs, THF ring opening, with no C–C bond cleavage, is observed. 1,4-Diketones, useful building blocks in organic synthesis,^[22] are obtained in this case. The latter transformation proceeds with a remarkable, nearly quantitative yield. A third route, leading to the oxidative cleavage of a C-2 side-chain, operates when the THF ring is flanked by a ketone function. Finally, the above results provide further support for the previously observed similarity of the chemical oxidative behaviour of PCC and of RuO_4 toward THF-containing substances.^[5a] Further development of the chemistry presented here is currently ongoing.

Experimental Section

General: All reagents were purchased in the highest commercial quality and used without further purification. Petroleum ether had a boiling range of 40–70 °C. Reactions were monitored by thin-layer chromatography carried out on precoated silica gel plates (Merck 60, F_{254} , 0.25 mm thick). Merck silica gel (Kieselgel 40, particle size 0.063–0.200 mm) was used for column chromatography. Na_2SO_4 was used as a drying agent for aqueous workup. HPLC separations were carried out with a Varian 2510 apparatus fitted with a Waters R403 dual cell differential refractometer with use of Phenomenex 250 \times 10 mm and 250 \times 4.6 mm (both 5 μ) and

LiChrosorb RP-18 250 × 4.0 mm columns. NMR experiments were performed with Varian Unity Inova 700, Varian Unity Inova 500, Varian Mercury Plus 400, and Gemini 200 spectrometers in CDCl₃. Proton chemical shifts were referenced to the residual CHCl₃ signal ($\delta = 7.26$ ppm). ¹³C NMR chemical shifts were referenced to the solvent ($\delta = 77.0$ ppm). Abbreviations for signal coupling are as follows: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet, br, broad. IR spectra were recorded with a Jasco FT-IR 430 spectrophotometer. High-resolution MS was recorded with a Bruker APEX II FT-ICR mass spectrometer with use of ESI. For compounds **2–4**, **30** and **31** the numbering previously given^[5] for penta-THF **1** is used.

Syntheses of Poly-THFs 2, 5 and 9: Penta-THF **1**, the tetrol precursor of **5** and dilactone **9** were synthesized as described previously.^[5] Compound **1** was benzoylated as follows. Benzoyl chloride (15 equiv., 6.75 mmol, 785 μ L) and DMAP (30 equiv., 3.5 mmol, 11.6 g) were added to **1** (236 mg, 0.45 mmol) dissolved in CH₂Cl₂ (10 mL) and the mixture was stirred at room temp. for 30 h. Water (2 mL) was added and the mixture was stirred for 15 min in a water bath and then taken to dryness. The residue was taken up in CH₂Cl₂ and washed with a sat. NaHCO₃ solution and water. The organic phase was dried, filtered and concentrated in vacuo to give an oily product that was chromatographed on silica gel with elution with petroleum ether/Et₂O (8:2) to give dibenzoyl penta-THF **2** (264 mg, 80%). Further purification of **2** was carried out by HPLC (250 × 10 mm column; flow: 2.5 mL min⁻¹, hexane/EtOAc 75:25).

Penta-THF Dibenzoate 2: Oil. ¹H NMR (200 MHz, CDCl₃): $\delta = 8.05$ – 7.92 (m, 4 H), 7.58 – 7.35 (m, 6 H), 4.25 – 4.15 (m, 2 H), 4.02 – 3.75 (4 H, overlapped multiplets), 2.33 – 1.49 (20 H, partly overlapped to some methyl protons), 1.64 (s, 3 H), 1.60 (s, 6 H), 1.58 , 1.23 , 1.16 , 1.15 , 1.11 (3 H each, singlets, 5 × Me) ppm. ¹³C NMR (50 MHz, CDCl₃): $\delta = 165.54$, 165.45 , 132.3 , 132.2 , 131.9 , 131.7 , 129.2 , 128.0 , 86.1 , 85.7 , 85.1 , 84.9 , 84.4 , 84.3 , 83.83 , 83.78 , 83.5 , 83.3 , 82.5 , 34.4 , 34.3 , 34.1 , 32.5 , 27.6 , 26.9 , 26.8 , 26.7 , 26.5 , 24.7 , 24.0 , 23.4 , 23.2 , 23.0 , 22.6 , 21.3 , 21.1 ppm. IR (neat): $\tilde{\nu}_{\max} = 1712$ (C=O benzoates), 1285 , 710 cm⁻¹. HRMS (ESI): calcd. for C₄₄H₆₀NaO₉ [M + Na]⁺ 755.4135; found 755.4144.

The tris-THF tetrol corresponding to **5** was benzoylated as described for **1**. Benzoyl chloride (15 equiv., 2.6 mmol, 305 μ L) and DMAP (60 equiv., 10.4 mmol, 1.3 g) were added to the tetrol (77.0 mg, 0.173 mmol) dissolved in CH₂Cl₂ (2.0 mL) and the mixture was heated at reflux for 10 h. The mixture was worked up as for **1** to give an oily product. Column chromatography on silica gel with elution with petroleum ether/Et₂O (8:2) gave tetrabenzoyl tris-THF **5** (108 mg, 73%).

Tris-THF Tetrabenzoate 5: Oil. ¹H NMR (400 MHz, CDCl₃): $\delta = 8.16$ – 7.95 (m, 8 H), 7.64 – 7.45 (m, 4 H), 7.45 – 7.32 (m, 8 H), 4.54 (dd, $J = 6.6$, 6.6 Hz, 1 H), 4.32 (m, 5 H), 3.89 – 3.74 (m, 2 H), 2.33 – 2.07 (7 H, overlapped multiplets), 2.01 – 1.73 (13 H, overlapped multiplets) 1.63 , 1.57 , 1.12 , 1.08 (3 H each, singlets, 4 × Me) ppm. ¹³C NMR (100 MHz, CDCl₃): $\delta = 166.50$, 166.46 , 165.6 , 132.80 , 132.78 , 132.50 , 132.46 , 131.7 , 131.6 , 130.26 , 130.24 , 129.5 , 129.4 , 128.26 , 128.24 , 128.16 , 128.15 , 85.9 , 85.3 , 85.2 , 84.7 , 84.4 , 83.9 , 83.7 , 81.4 , 65.2 , 65.1 , 34.6 , 34.4 , 31.9 , 31.5 , 27.5 , 27.2 , 26.92 , 26.86 , 24.3 , 23.7 , 23.5 , 23.1 , 20.4 , 20.0 ppm. IR (neat): $\tilde{\nu}_{\max} = 1714$ (C=O benzoates), 1275 , 710 cm⁻¹. HRMS (ESI): calcd. for C₅₂H₆₀NaO₁₁ [M + Na]⁺ 883.4033; found 883.4028.

Oxidation of 2 with PCC/AcOH: PCC (5 equiv., 2.5 mmol, 537 mg) and AcOH (70 equiv., 35 mmol, 2 mL) were added to a solution of **2** (365 mg, 0.50 mmol) in CH₂Cl₂ (16 mL) and the resulting mixture was heated at reflux for 16 h. A saturated aqueous NaHCO₃ solution was added and the organic phase was washed with water,

dried and concentrated in vacuo to give a yellow oil. Filtration on a silica gel pad (eluent CHCl₃/MeOH 9:1) afforded an oily product (350 mg). Further elution with CHCl₃/MeOH (8:2) gave a complex mixture of polar products (57 mg) that was not studied further. The first eluted fraction was separated by HPLC (250 × 10 mm column; flow: 2.5 mL min⁻¹; eluent: hexane/EtOAc 75:25) to give unreacted **2** (128 mg) and slightly impure compounds **3** and **4**. Analytical HPLC (250 × 4.6 mm column; flow: 1.0 mL min⁻¹; 3 mg/injection, hexane/EtOAc 75:25) afforded major acid **3** (58.4 mg, 25%, $t_R = 11.5$ min) and minor acid **4** (16.3 mg, 7%, $t_R = 10.0$ min).

Acid 3 (Major Isomer): Oil. ¹H NMR (700 MHz, CDCl₃): $\delta = 7.95$ (br. d, $J = 7.6$ Hz, 4 H), 7.52 (br. t, $J = 7.4$ Hz, 2 H), 7.41 (br. t, $J = 7.7$ Hz, 4 H), 4.38 (dd, $J = 7.3$, 7.3 Hz, 1 H), 4.25 (dd, $J = 7.4$, 7.4 Hz, 1 H), 4.13 (br. dd, $J = 7.4$, 7.4 Hz, 1 H), 4.01 (dd, $J = 10.0$, 5.3 Hz, 1 H), 3.94 (dd, $J = 9.0$, 6.3 Hz, 1 H), 3.25 (d, $J = 14.9$ Hz, 1 H), 2.97 (d, $J = 14.9$ Hz, 1 H), 2.36 (ddd, $J = 13.0$, 8.4 , 5.5 Hz, 1 H), 2.07 – 1.81 (14 H, overlapped multiplets), 1.69 (m, 1 H), 1.62 , 1.61 , 1.59 , 1.56 , 1.54 , 1.50 , 1.20 , 1.13 (3 H each, singlets, 8 × Me) ppm. ¹³C NMR (175 MHz, CDCl₃): $\delta = 174.4$, 171.9 , 165.7 , 165.6 , 132.6 , 132.5 , 131.85 , 131.78 , 129.42 , 129.37 , 128.25 , 128.20 , 86.4 , 85.5 , 85.3 , 84.6 , 84.5 , 83.9 , 83.6 , 83.2 , 82.8 , 81.9 , 42.2 , 36.5 , 34.5 , 32.9 , 27.9 , 27.2 , 26.7 , 26.6 , 26.0 , 24.3 , 23.14 , 23.09 , 22.8 , 22.3 , 21.5 , 21.41 , 21.38 ppm. IR (neat): $\tilde{\nu}_{\max} = 3600$ – 2400 (OH), 1714 (C=O of benzoates, ester and acid), 1288 , 712 cm⁻¹. HRMS (ESI): calcd. for C₄₄H₅₈NaO₁₂ [M + Na]⁺ 801.3826; found 801.3819.

Acid 4 (Minor Isomer): Oil. ¹H NMR (400 MHz, CDCl₃): $\delta = 7.9$ (br. d, $J = 7.4$ Hz, 4 H), 7.51 (br. t, $J = 7.3$ Hz, 2 H), 7.40 (br. t, $J = 7.4$ Hz, 4 H), 4.35 (dd, $J = 7.2$, 7.2 Hz, 1 H), 4.24 (dd, $J = 9.3$, 5.9 Hz, 1 H), 4.07 – 3.99 (m, 2 H), 3.55 (dd, $J = 7.2$, 7.2 Hz, 1 H), 3.31 (d, $J = 15.0$ Hz, 1 H), 2.98 (d, $J = 15.0$ Hz, 1 H), 2.61 – 2.53 (m, 1 H), 2.19 – 2.10 (m, 1 H), 2.05 – 1.76 (14 H, overlapped multiplets), 1.64 , 1.63 , 1.60 , 1.59 , 1.48 , 1.47 , 1.15 , 1.12 (3 H each, singlets, 8 × Me) ppm. ¹³C NMR (100 MHz, CDCl₃): $\delta = 174.8$, 172.9 , 165.8 , 165.7 , 132.6 , 132.4 , 132.0 , 131.7 , 129.5 , 129.4 , 128.19 , 128.16 , 86.3 , 86.0 , 85.8 , 85.5 , 85.1 , 84.3 , 83.9 , 83.8 , 83.7 , 83.4 , 81.5 , 43.0 , 36.4 , 34.7 , 34.4 , 26.95 , 26.88 , 26.82 , 25.2 , 24.1 , 23.0 , 22.7 , 21.7 ppm. IR (neat): $\tilde{\nu}_{\max} = 3600$ – 2400 (OH), 1712 (C=O of benzoates, ester and acid), 1287 , 712 cm⁻¹. HRMS (ESI): calcd. for C₄₄H₅₈NaO₁₂ [M + Na]⁺ 801.3826; found 801.3835.

Oxidation of 2 with PCC (cat.)/H₅IO₆: PCC (2 mol-%, 700 μ L of a 0.01 M stock solution in acetonitrile) was added at 0 °C with vigorous stirring to a suspension of H₅IO₆ (4 equiv., 78.8 mg, 0.35 mmol) in acetonitrile (1.5 mL). After 5 min, compound **2** (61 mg, 0.083 mmol) dissolved in acetonitrile (650 μ L) was added. After 3 h, CH₂Cl₂ (1.5 mL) was added followed by ethanol (40 μ L) and the mixture was taken to dryness. Filtration through a short pad of sodium thiosulfate adsorbed on silica^[5] (CHCl₃/MeOH 9:1) gave an oily product (50 mg). Further elution with CHCl₃/MeOH (8:2) gave a mixture of polar products (14 mg) that was not studied further. The first eluted fraction was separated as described above for the analogous process with PCC/AcOH to give compounds **3** (20.6 mg, 40%) and **4** (2.2 mg, 4%), along with slightly impure ketol **30** ($t_R = 5.0$ min), lactone **31** ($t_R = 13.5$ min) and **32** ($t_R = 16.5$ min). A further reversed-phase HPLC run (250 × 4.0 mm column; flow: 1.0 mL min⁻¹; 2 mg/injection) was required to obtain pure **30** (MeCN/H₂O 9:1, 4.6 mg, 2%, $t_R = 14.0$ min), **31** (MeCN/H₂O 85:15, 3.0 mg, 2%, $t_R = 5.2$ min) and **32** (MeCN/H₂O 8:2, 1.0 mg, 1%, $t_R = 2.0$ min).

Ketol 30: Oil. ¹H NMR (700 MHz, CDCl₃): $\delta = 7.99$ (d, $J = 6.9$ Hz, 2 H), 7.98 (d, $J = 6.5$ Hz, 2 H), 7.52 (t, $J = 7.3$ Hz, 2 H), 7.43 (t, $J = 7.7$ Hz, 2 H), 7.40 (t, $J = 7.6$ Hz, 2 H), 4.33 (br. s, 1 H, OH), 4.16 (dd, $J = 9.2$, 6.0 Hz, 1 H), 4.10 (dd, $J = 8.2$, 7.0 Hz, 1 H), 3.89

(m, 1 H), 3.72 (m, 2 H), 2.66 (ddd, $J = 12.7, 10.5, 5.4$ Hz, 1 H), 2.55 (d, $J = 16.9$ Hz, 1 H), 2.44 (d, $J = 16.9$ Hz, 1 H), 2.18 (ddd, $J = 12.3, 9.3, 8.2$ Hz, 1 H), 2.03–1.83 (12 H, overlapped multiplets), 1.71–1.61 (m, 2 H), 1.60 (s, 6 H, 2 × Me), 1.58, 1.55, 1.36, 1.29, 1.14, 1.07 (3 H each, singlets, 6 × Me) ppm. ^{13}C NMR (175 MHz, CDCl_3): $\delta = 212.7, 165.8, 165.7, 132.6, 132.4, 132.0, 131.6, 129.6, 129.4, 128.2, 128.1, 101.7, 86.4, 85.9, 85.70, 85.68, 85.65, 84.3, 83.51, 83.46, 83.2, 82.9, 77.8, 48.4, 35.1, 34.4, 31.8, 27.0, 26.86, 26.83, 26.81, 26.0, 24.6, 24.4, 22.82, 22.79, 22.74, 22.2, 21.6, 20.1$ ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 3301$ (broad, OH), 1712 (C=O of benzoates and ketone), 1287, 712 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{44}\text{H}_{58}\text{NaO}_{11}$ [M + Na] $^+$ 785.3877; found 785.3884.

Lactone 31: Oil. ^1H NMR (500 MHz, CDCl_3): $\delta = 7.96$ (br. d, $J = 7.0$ Hz, 2 H), 7.53 (br. t, $J = 7.4$ Hz, 1 H), 7.41 (br. t, $J = 7.5$ Hz, 2 H), 4.25 (dd, $J = 7.0, 7.0$ Hz, 1 H), 3.94 (dd, $J = 9.0, 6.2$ Hz, 1 H), 3.88 (dd, $J = 7.2, 7.2$ Hz, 1 H), 3.78 (dd, $J = 9.4, 5.1$ Hz, 1 H), 2.80 (ddd, $J = 17.4, 10.1, 10.2$ Hz, 1 H), 2.47 (ddd, $J = 17.4, 10.6, 3.5$ Hz, 1 H), 2.39 (ddd, $J = 13.9, 10.7, 3.5$ Hz, 1 H), 2.09–1.65 (13 H, overlapped multiplets), 1.63, 1.58, 1.31, 1.22, 1.11 (3 H each, singlets, 5 × Me) ppm. ^{13}C NMR (100 MHz, CDCl_3): $\delta = 177.6, 165.7, 132.5, 131.9, 129.4, 128.2, 86.5, 85.7, 85.2, 84.6, 84.0, 83.6, 83.4, 82.7, 35.0, 32.7, 29.8, 27.8, 27.3, 26.8, 26.6, 23.5, 23.3, 23.2, 23.0, 22.9, 21.3$ ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 1773$ (C=O lactone), 1711 (C=O benzoate), 1288, 1071, 713 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{29}\text{H}_{40}\text{NaO}_7$ [M + Na] $^+$ 523.2672; found 523.2668.

Acid 32: Oil. ^1H NMR (500 MHz, CDCl_3): $\delta = 7.97$ (br. d, $J = 7.3$ Hz, 2 H), 7.55 (br. t, 1 H, $J = 7.3$ Hz), 7.44 (br. t, 2 H, $J = 7.7$ Hz), 5.62 (dd, 1 H, $J = 9.3, 3.1$ Hz), 2.86 (dd, 1 H, $J = 16.1, 3.2$ Hz), 2.74 (dd, 1 H, $J = 16.1, 9.4$ Hz), 2.11 (s, 3 H), 1.66, 1.65 (3 H each, singlets, 2 × Me) ppm. ^{13}C NMR (100 MHz, CDCl_3): $\delta = 175.2, 170.1, 165.1, 133.0, 131.1, 129.5, 128.4, 82.4, 73.9, 34.9, 22.4, 21.7, 20.9$ ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 3600$ –2400 (OH), 1744 (C=O acetate), 1715 (C=O benzoate and acid), 1283, 712 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{15}\text{H}_{18}\text{NaO}_6$ [M + Na] $^+$ 317.1001; found 317.1009.

Oxidation of 5 with PCC (cat.)/ H_5IO_6 : PCC (1 mol-%, 38 μL of a 0.01 M stock solution in acetonitrile) was added at 0 °C with vigorous stirring to a suspension of H_5IO_6 (4 equiv., 0.15 mmol, 34.2 mg) in acetonitrile (600 μL). After 5 min, compound **5** (32.3 mg, 0.037 mmol) dissolved in acetonitrile (150 μL) was added. After 2 h, a further 1 mol-% PCC was added and the mixture was kept at 0 °C for an additional 1 h. CH_2Cl_2 (700 μL) was then added followed by ethanol (40 μL) and the mixture was taken to dryness. Filtration through a short pad of sodium thiosulfate adsorbed on silica ($\text{CHCl}_3/\text{MeOH}$ 9:1) gave an oily product (19 mg). Separation by HPLC (250 × 4.6 mm column; flow: 1.0 mL min^{-1} ; 1 mg/injection, hexane/EtOAc 75:25) gave compounds **6** (10.0 mg, 30%), **7** (2.0 mg, 6%) and **8** (2.9 mg, 9%).

Acid 6 (Major Isomer): Oil. ^1H NMR (400 MHz, CDCl_3): $\delta = 8.02$ (d, $J = 8.2$ Hz, 4 H), 7.96 (d, $J = 8.0$ Hz, 4 H), 7.55–7.45 (m, 4 H), 7.44–7.38 (m, 8 H), 5.40 (dd, $J = 9.7, 3.1$ Hz, 1 H), 4.49 (dd, $J = 8.5, 6.2$ Hz, 1 H), 4.44 (dd, $J = 7.4, 7.4$ Hz, 1 H), 4.40–4.28 (4 H, overlapped multiplets), 2.73 (dd, $J = 15.8, 3.0$ Hz, 1 H), 2.64 (dd, $J = 15.8, 9.8$ Hz, 1 H), 2.37 (ddd, $J = 12.6, 7.9, 6.0$ Hz, 1 H), 2.29–2.19 (m, 2 H), 2.17–2.00 (m, 2 H), 1.98–1.58 (12 H, overlapped multiplets), 1.57, 1.55, 1.45, 1.19 (3 H each, singlets, 4 × Me) ppm. ^{13}C NMR (100 MHz, CDCl_3): $\delta = 174.2, 173.4, 166.8, 166.6, 165.8, 165.6, 132.93, 132.87, 132.7, 131.5, 131.4, 130.26, 130.21, 129.6, 129.55, 129.50, 128.32, 128.27, 85.2, 85.0, 83.9, 83.5, 83.3, 81.3, 75.0, 65.3, 65.0, 36.1, 35.7, 33.7, 31.6, 31.5, 26.9, 25.9, 24.1, 23.5, 23.2, 22.7, 20.4, 20.2$ ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 3600$ –2400 (OH), 1717 (C=O benzoates, ester and acid), 1279, 1119, 710 cm^{-1} . HRMS

(ESI): calcd. for $\text{C}_{52}\text{H}_{58}\text{NaO}_{14}$ [M + Na] $^+$ 929.3724; found 929.3736.

Acid 7 (Minor Isomer): Oil. ^1H NMR (400 MHz, CDCl_3): $\delta = 8.04$ (br. d, $J = 7.0$ Hz, 2 H), 8.02 (br. d, $J = 5.5$ Hz, 2 H), 7.98 (br. d, $J = 8.0$ Hz, 4 H), 7.53 (m, 4 H), 7.41 (m, 8 H), 5.41 (dd, $J = 6.1, 6.1$ Hz, 1 H), 4.64 (dd, $J = 7.4, 7.4$ Hz, 1 H), 4.49–4.44 (m, 5 H), 2.63 (dd, $J = 15.7, 7.3$ Hz, 1 H), 2.56 (dd, $J = 15.8, 5.9$ Hz, 1 H), 2.44 (br. ddd, $J = 12.5, 5.5, 5.5$ Hz, 1 H), 2.23–1.70 (16 H, overlapped multiplets), 1.64, 1.53, 1.44, 1.28 (3 H each, singlets, 4 × Me) ppm. ^{13}C NMR (100 MHz, CDCl_3): $\delta = 172.9, 171.3, 167.2, 167.0, 165.8, 165.7, 133.03, 132.98, 132.7, 132.6, 131.6, 131.4, 130.1, 129.64, 129.57, 129.4, 128.34, 128.29, 85.3, 84.6, 84.3, 84.1, 83.9, 83.5, 74.9, 65.5, 65.3, 36.5, 35.0, 33.4, 32.0, 31.1, 26.9, 26.3, 24.8, 24.4, 23.6, 23.2, 20.8, 20.6$ ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 3600$ –2400 (OH), 1714 (C=O benzoates, ester and acid), 1277, 1113, 711 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{52}\text{H}_{58}\text{NaO}_{14}$ [M + Na] $^+$ 929.3724; found 929.3742.

Diketone 8: Oil. ^1H NMR (200 MHz, CDCl_3): $\delta = 8.10$ –7.91 (m, 8 H), 7.62–7.32 (m, 12 H), 4.54–4.24 (6 H, overlapped multiplets), 3.17–2.94 (m, 1 H), 2.94–1.68 (19 H, overlapped multiplets), 1.63 (s, 6 H), 1.37 (s, 3 H), 1.34 (s, 3 H) ppm. ^{13}C NMR (100 MHz, CDCl_3): $\delta = 214.0, 212.7, 166.5, 165.7, 132.8, 132.7, 131.6, 131.4, 130.35, 130.31, 129.55, 129.53, 129.48, 129.43, 128.35, 128.31, 88.9, 88.7, 84.93, 84.92, 84.2, 83.4, 65.0, 35.2, 34.8, 32.3, 31.8, 30.8, 29.9, 26.3, 24.2, 23.7, 23.55, 23.51, 20.6, 20.4$ ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 1710$ (C=O benzoates and ketones), 1274, 1111, 710 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{52}\text{H}_{58}\text{NaO}_{12}$ [M + Na] $^+$ 897.3826; found 897.3815.

Oxidation of 9: PCC (2 mol-%, 80 μL of a 0.01 M stock solution in acetonitrile) was added at 0 °C with vigorous stirring to a suspension of H_5IO_6 (4 equiv., 0.15 mmol, 35.4 mg) in acetonitrile (600 μL). After 5 min, compound **9** (10.4 mg, 0.039 mmol) dissolved in acetonitrile (150 μL) was added. After 3.5 h, CH_2Cl_2 (750 μL) was added followed by ethanol (40 μL) and the mixture was taken to dryness. Filtration through a short pad of sodium thiosulfate adsorbed on silica ($\text{CHCl}_3/\text{MeOH}$ 9:1) gave almost pure **10** as an oil. Further purification by HPLC (250 × 4.6 mm column; flow: 1.0 mL min^{-1} ; $\text{CHCl}_3/\text{MeOH}$ 98:2, $t_r = 4.0$ min) gave pure **10** (4.3 mg, 45%).

Compound 10: Oil. ^1H NMR (500 MHz, CDCl_3): $\delta = 3.01$ –2.92 (m, 2 H), 2.92–2.82 (m, 2 H), 2.65–2.53 (m, 6 H), 2.12–2.04 (m, 2 H), 1.55 (s, 6 H, 2 × Me) ppm. ^{13}C NMR (125 MHz, CDCl_3): $\delta = 209.2$ (2 C), 175.9 (2 C), 89.2 (2 C), 31.3 (2 C), 31.0 (2 C), 28.2 (2 C), 23.6 (2 C) ppm. IR (neat): $\tilde{\nu}_{\text{max}} = 1774$ (C=O lactones), 1717 (C=O ketones) cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{14}\text{H}_{18}\text{NaO}_6$ [M + Na] $^+$ 305.1001; found 305.1008.

Synthesis of Mono-THFs 11, 14, 16, 19, 22 and 24: The title compounds were obtained by benzylation of the corresponding THF diols (synthesized as described previously^[5g,5j,8,10]). Purification of these compounds was achieved by column chromatography with the following eluents: **11** (petroleum ether/diethyl ether 9:1), **14** (petroleum ether/diethyl ether 1:1), **16** (petroleum ether/diethyl ether 6:4), **19** (petroleum ether/diethyl ether 1:1), **22** (40:70 petroleum ether/diethyl ether 8:2), **24** (hexane/EtOAc 8:2).

Compound 11: Oil. ^1H NMR: (200 MHz, CDCl_3): $\delta = 8.13$ –7.88 (m, 6 H), 7.68–7.28 (m, 9 H), 5.65 (dd, $J = 8.8, 2.9$ Hz, 1 H), 4.80 (dd, $J = 12.2, 2.9$ Hz, 1 H), 4.59 (dd, $J = 12.2, 8.8$ Hz, 1 H), 4.33 (dd, $J = 6.3$ Hz, 1 H), 2.27–1.74 (m, 4 H), 1.67 (s, 3 H), 1.64 (s, 3 H), 1.43 (s, 3 H) ppm. ^{13}C NMR (50 MHz, CDCl_3): $\delta = 166.4, 166.0, 165.6, 133.0, 132.9, 132.5, 131.6, 129.8, 129.65$ (2 C), 129.63, 129.5 (2 C), 129.3 (2 C), 128.4 (2 C), 128.3 (2 C), 128.2 (2 C), 83.5,

83.0, 82.8, 76.2, 63.9, 34.6, 26.9, 23.6, 23.2, 21.3 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1719 (C=O benzoates), 1281, 1111, 709 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{31}\text{H}_{32}\text{NaO}_7$ [M + Na]⁺ 539.2046; found 539.2038.

Compound 14: Oil. ¹H NMR (200 MHz, CDCl_3): δ = 8.00 (m, 4 H), 7.66–7.30 (m, 6 H), 5.46 (dd, J = 8.8, 2.6 Hz, 1 H), 4.78–4.20 (5 H, overlapped multiplets), 2.15–1.88 (10 H, including two acetate singlets at 2.02 and 1.96), 1.67 (s, 3 H), 1.36 (s, 3 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 170.9, 170.5, 166.1, 165.6, 133.1, 132.8, 131.0, 129.7 (2 C), 129.6 (3 C), 128.5 (2 C), 128.3 (2 C), 83.2, 83.0, 80.4, 76.1, 65.2, 63.3, 34.5, 26.8, 23.5, 20.80 (2 C), 18.9 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1730 (C=O acetates), 1721 (C=O benzoates), 1275, 712 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{28}\text{H}_{32}\text{NaO}_9$ [M + Na]⁺ 535.1944; found 535.1963.

Compound 16: Oil. ¹H NMR (200 MHz, CDCl_3): δ = 8.07 (d, J = 8.5 Hz, 2 H), 7.96 (d, J = 7.1 Hz, 2 H), 7.64–7.29 (m, 6 H), 5.20 (s, 1 H), 4.33 (dd, J = 7.7, 6.9 Hz, 1 H), 3.72 (s, 3 H), 2.56–2.38 (m, 1 H), 2.19–1.75 (m, 3 H), 1.61 (s, 3 H), 1.59 (s, 3 H), 1.46 (s, 3 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 168.6, 166.0, 165.7, 133.3, 132.5, 131.6, 129.8 (3 C), 129.4 (2 C), 128.4 (2 C), 128.2 (2 C), 83.5, 83.3, 82.8, 78.0, 52.2, 34.3, 26.5, 23.4, 23.0, 21.3 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1713 (C=O benzoates and CO_2Me), 1285, 1113, 710 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{25}\text{H}_{28}\text{NaO}_7$ [M + Na]⁺ 463.1733; found 463.1717.

Compound 19: Oil. ¹H NMR (200 MHz, CDCl_3): δ = 8.07 (d, J = 7.7 Hz, 2 H), 7.96 (d, J = 8.2 Hz, 4 H), 7.64–7.30 (m, 9 H), 5.65 (dd, J = 8.6, 2.7 Hz, 1 H), 4.75 (dd, J = 11.9, 2.8 Hz, 1 H), 4.55 (dd, J = 11.9, 8.6 Hz, 1 H), 4.32 (dd, J = 8.1, 6.6 Hz, 1 H), 2.27–1.68 (m, 4 H), 1.60 (s, 6 H), 1.45 (s, 3 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 166.5, 166.0, 165.7, 133.04, 133.02, 132.4, 131.9, 130.1, 129.72 (2 C), 129.70, 129.6 (2 C), 129.4 (2 C), 128.4 (2 C), 128.3 (2 C), 128.2 (2 C), 85.5, 83.03, 83.01, 76.1, 64.0, 34.8, 26.7, 24.1, 22.6, 21.5 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1712 (C=O benzoates), 1278, 1109, 708 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{31}\text{H}_{32}\text{NaO}_7$ [M + Na]⁺ 539.2046; found 539.2055.

Compound 22: Oil. ¹H NMR (200 MHz, CDCl_3): δ = 7.96–7.87 (d, J = 6.9 Hz, 2 H), 7.59–7.35 (m, 3 H), 5.04 (AB system, J = 17.8 Hz, 2 H), 4.20 (dd, J = 7.2 Hz, 1 H), 2.54–2.39 (m, 1 H), 2.13–1.77 (6 H, overlapped multiplets including the acetate singlet at 2.08), 1.71 (s, 3 H), 1.61 (s, 3 H), 1.41 (s, 3 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 206.5, 170.2, 165.7, 132.7, 131.4, 129.3 (2 C), 128.4 (2 C), 88.5, 85.9, 82.7, 65.7, 35.4, 26.0, 23.6, 22.8, 22.6, 20.4 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1735 (C=O acetate), 1716 (C=O benzoate and ketone), 1289, 712 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{19}\text{H}_{24}\text{NaO}_6$ [M + Na]⁺ 371.1471; found 371.1489.

Compound 24: Oil. ¹H NMR (500 MHz, CDCl_3): δ = 8.05 (br. d, J = 8.4 Hz, 2 H), 7.55 (br. t, J = 7.4 Hz, 2 H), 7.42 (br. t, J = 7.8 Hz, 4 H), 4.42 (2 H, overlapped multiplets), 4.39–4.32 (4 H, overlapped multiplets), 2.12 (m, 2 H), 1.89 (m, 2 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 166.4 (2 C), 132.9 (2 C), 130.0 (2 C), 129.6 (4 C), 128.3 (4 C), 77.5 (2 C), 66.6 (2 C), 27.8 (2 C) ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1715 (C=O benzoates), 1268, 709 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{20}\text{H}_{20}\text{NaO}_5$ [M + Na]⁺ 363.1208; found 363.1220.

Synthesis of 20: K_2CO_3 (2 equiv., 0.4 mmol, 55 mg) was added to compound **11** (106 mg, 0.20 mmol) dissolved in MeOH (5 mL) and the mixture was stirred at room temp. for 1.5 h. Water (1 mL) was added followed by acetic acid up to neutrality. The mixture was taken to dryness in vacuo and the solid was partitioned between a sat. NaHCO_3 solution and EtOAc. The organic phase was washed with water, dried and concentrated to give an oily product (65.9 mg).

NaIO_4 supported on wet silica (2.4 equiv., 0.47 mmol, 746 mg; 0.64 mmol g^{-1}) was added to the crude product obtained above, dis-

solved in CH_2Cl_2 (5 mL), and the mixture was stirred for 30 min at room temp. The reaction mixture was filtered, the solid was thoroughly washed with CH_2Cl_2 , and the filtrate was taken to dryness to give an oily product.

Sodium borohydride (two spatula tips) was added to the above oil (55.1 mg), dissolved in EtOH (4 mL). After 30 min, acetic acid was added dropwise until no gas evolution was observed. The mixture was filtered and the filtrate was taken to dryness to give an oily product.

The above oil (33.0 mg, 0.12 mmol) was dissolved in pyridine (500 μL) and benzoyl chloride (1.1 equiv., 0.13 mmol, 15 μL) was added. After 1 h, water (100 μL) was added and the mixture was taken to dryness. Purification by preparative TLC (hexane/EtOAc 1:1) yielded pure **20** (33.0 mg, 43% over four steps).

Compound 20: Oil. ¹H NMR (200 MHz, CDCl_3): δ = 8.06 (d, J = 7.0 Hz, 2 H), 7.98 (d, J = 6.9 Hz, 2 H), 7.61–7.33 (6 H, overlapped multiplets), 4.38 (d, J = 11.1 Hz, 1 H), 4.31 (m, 1 H), 4.20 (d, J = 11.1 Hz, 1 H), 2.20–1.73 (m, 4 H), 1.64 (s, 3 H), 1.61 (s, 3 H), 1.38 (s, 3 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 166.4, 165.7, 133.0, 132.5, 131.7, 130.1, 129.5 (2 C), 129.4 (2 C), 128.3 (2 C), 128.2 (2 C), 83.8, 83.4, 81.8, 69.7, 34.5, 26.8, 24.1, 23.0, 21.4 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1712 (C=O benzoates), 1277, 1112, 709 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{23}\text{H}_{26}\text{NaO}_5$ [M + Na]⁺ 405.1678; found 405.1670.

General Reaction Procedure for the Oxidation of Mono-THF Compounds 11, 14, 16, 19, 20, 22 and 24: PCC (5 mol-%, from a 0.01 M stock solution in acetonitrile) was added at room temp. with vigorous stirring to a suspension of H_5IO_6 (4 equiv.) in acetonitrile. After 5 min the mono-THF compound (1 equiv.) dissolved in acetonitrile was added. The overall volume of acetonitrile was such that the final concentration of the solution was 0.05 M. After complete consumption of the starting material (TLC monitoring, usually 45–60 min), ethanol (excess) was added and stirring was continued until the colour of the solution had turned from yellow to green. Silica (excess) was then added and the solvent was evaporated in vacuo to give a fine powder that was loaded on the top of a silica gel column. Elution with $\text{CHCl}_3/\text{MeOH}$ (9:1) allowed recovery of the acid that proved to be pure enough for successive spectral studies and/or synthetic steps.

Compound **11** (112.0 mg, 0.22 mmol) was subjected to the standard oxidative procedure to give the crude acid as a yellow oil. It was purified by column chromatography ($\text{CH}_2\text{Cl}_2/\text{MeOH}$ 95:5) to yield **12** (99 mg, 80%) and lactone **13** (8.0 mg, 10%) as oils.

Compound 12: Oil. ¹H NMR: (200 MHz, CDCl_3): δ = 8.00 (d, J = 7.7 Hz, 2 H), 7.87 (m, 4 H), 7.57–7.28 (m, 9 H), 5.89 (dd, J = 7.8, 3.1 Hz, 1 H), 4.74 (dd, J = 12.2, 3.1 Hz, 1 H), 4.45 (dd, J = 12.2, 7.8 Hz, 1 H), 3.24 (AB system, J = 15.6 Hz, 2 H), 1.86 (s, 3 H), 1.71 (s, 6 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 174.1, 171.1, 166.0, 165.5, 165.2, 133.15, 133.12, 133.0, 129.8, 129.6 (6 C), 129.5, 129.3, 128.35 (4 C), 128.28 (2 C), 82.0, 79.0, 74.6, 62.8, 39.2, 24.55, 24.46, 21.3 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 3600–2400 (OH), 1717 (C=O benzoates, ester and acid), 1282, 1107, 709 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{31}\text{H}_{30}\text{NaO}_{10}$ [M + Na]⁺ 585.1737; found 585.1723.

Compound 13: Oil. ¹H NMR (200 MHz, CDCl_3): δ = 8.04 (d, J = 7.9 Hz, 2 H), 7.92 (d, J = 8.0 Hz, 2 H), 7.67–7.31 (m, 6 H), 5.66 (dd, J = 8.3, 3.2 Hz, 1 H), 4.85 (dd, J = 12.1, 3.2 Hz, 1 H), 4.52 (dd, J = 12.1, 8.3 Hz, 1 H), 2.61 (t, J = 7.6 Hz, 2 H), 2.37 (ddd, J = 13.1, 7.6, 7.6 Hz, 1 H), 2.08 (ddd, J = 13.1, 9.5, 9.5 Hz, 1 H), 1.66 (s, 3 H) ppm. ¹³C NMR (50 MHz, CDCl_3): δ = 176.2, 166.2, 165.5, 133.7, 133.2, 129.9 (2 C), 129.6 (2 C), 129.4, 128.9, 128.7 (2 C), 128.4 (2 C), 85.2, 75.6, 62.8, 31.5, 28.9, 24.2 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1778 (C=O lactone), 1721 (C=O benzoates), 1259, 709 cm^{-1} .

HRMS (ESI): calcd. for $C_{21}H_{20}NaO_6$ [M + Na]⁺ 391.1158; found 391.1167.

Compound **14** (13.5 mg, 0.026 mmol) was subjected to the standard oxidative procedure to give the crude acid as an oil. This was purified by column chromatography ($CH_2Cl_2/MeOH$ 95:5) to yield **15** (9.4 mg, 65%) as an oil.

Compound 15: Oil. ¹H NMR (200 MHz, $CDCl_3$): δ = 8.0 (m, 4 H), 7.65–7.29 (m, 6 H), 5.75 (dd, J = 7.8, 2.9 Hz, 1 H), 4.67–4.49 (m, 3 H), 4.21 (dd, J = 12.1, 7.8 Hz, 1 H), 3.20 (AB system, J = 15.7 Hz, 2 H), 2.09 (s, 3 H, acetate), 1.94 (s, 3 H, acetate), 1.85 (s, 3 H), 1.73 (s, 3 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 172.4, 172.3, 170.5, 170.3, 165.24, 165.17, 133.4, 133.3, 130.1, 129.8 (2 C), 129.7 (2 C), 129.6, 128.5 (4 C), 82.8, 79.4, 74.3, 66.0, 62.2, 38.9, 21.1, 20.71, 20.68, 19.7 ppm. IR (neat): $\tilde{\nu}_{max}$ = 3600–2400 (OH), 1740 (shoulder, C=O acetates) 1721 (C=O benzoates, ester and acid), 1273, 1226, 711 cm^{-1} . HRMS (ESI): calcd. for $C_{28}H_{30}NaO_{12}$ [M + Na]⁺ 581.1635; found 581.1651.

Compound **16** (23.8 mg, 0.050 mmol) was subjected to the standard oxidative procedure to give an oil, which was separated by column chromatography (CH_2Cl_2 to $CH_2Cl_2/MeOH$ 95:5) to yield acid **17** (7.3 mg, 30%) and lactone **18** (6.4 mg, 45%) as oils.

Compound 17: Oil. ¹H NMR (400 MHz, $CDCl_3$): δ = 8.02–7.89 (m, 4 H), 7.61–7.44 (m, 2 H), 7.44–7.29 (m, 4 H), 5.69 (s, 1 H), 3.64 (s, 3 H), 3.42 (d, J = 16.2 Hz, 1 H), 3.27 (d, J = 16.2 Hz, 1 H), 1.91 (s, 3 H), 1.65 (s, 3 H), 1.63 (s, 3 H) ppm. ¹³C NMR (100 MHz, $CDCl_3$): δ = 171.4, 171.0, 167.4, 165.4, 165.2, 133.4, 133.1, 130.2, 129.85, 129.84 (2 C), 129.6 (2 C), 128.4 (2 C) 128.2 (2 C), 81.1, 78.8, 75.8, 52.4, 39.0, 24.6, 24.1, 20.7 ppm. IR (neat): $\tilde{\nu}_{max}$ = 3600–2400 (OH), 1719 (C=O benzoates, esters and acid), 1287, 1110, 712 cm^{-1} . HRMS (ESI): calcd. for $C_{25}H_{26}NaO_{10}$ [M + Na]⁺ 509.1424; found 509.1418.

Compound 18: Oil. ¹H NMR (200 MHz, $CDCl_3$): δ = 8.06, (d, J = 7.0 Hz, 2 H), 7.63 (t, J = 6.5 Hz, 1 H), 7.48 (t, J = 7.7 Hz, 2 H), 5.26 (s, 1 H), 3.80 (s, 2.76–2.52 (m, 3 H), 2.20–2.01 (m, 1 H), 1.65 (s, 3 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 175.6, 167.3, 165.5, 133.8, 129.9 c, 128.6 (3 C), 84.3, 77.2, 52.9, 30.9, 28.7, 23.8 ppm. IR (neat): $\tilde{\nu}_{max}$ = 1781 (C=O lactone), 1726 (C=O benzoate and CO_2Me), 1271, 1112, 714 cm^{-1} . HRMS (ESI): calcd. for $C_{15}H_{16}NaO_6$ [M + Na]⁺ 315.0845; found 315.0853.

Compound **19** (9.7 mg, 0.019 mmol) was subjected to the standard oxidative procedure to give the crude acid as an oil. This was purified by column chromatography ($CH_2Cl_2/MeOH$ 95:5) to yield **12** (7.3 mg, 70%).

Compound **20** (31.0 mg, 0.081 mmol) was subjected to the standard oxidative procedure to give the crude acid as an oil. This was purified by column chromatography ($CH_2Cl_2/MeOH$ 95:5) to yield **21** (21.5 mg, 65%) as an oil.

Compound 21: Oil. ¹H NMR (200 MHz, $CDCl_3$): δ = 7.94 (m, 4 H), 7.60–7.44 (m, 2 H), 7.42–7.28 (m, 4 H), 4.60 (AB system, J = 11.7 Hz, 2 H), 3.12 (br. s, 2 H), 1.71 (s, 3 H), 1.64 (s, 6 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 174.9, 171.3, 165.8, 165.4, 133.05, 133.02, 129.8, 129.65, 129.6 (2 C), 129.5 (2 C), 128.35 (2 C), 128.30 (2 C), 80.4, 78.9, 67.5, 39.9, 24.5 (2 C), 21.0 ppm. IR (neat): $\tilde{\nu}_{max}$ = 3600–2400 (OH), 1716 (C=O benzoates, ester and acid), 1284, 1109, 710 cm^{-1} . HRMS (ESI): calcd. for $C_{23}H_{24}NaO_8$ [M + Na]⁺ 451.1369; found 451.1360.

Compound **22** (10.1 mg, 0.029 mmol) was subjected to the standard oxidative procedure to give an oil, which was purified by HPLC (hexane/EtOAc 7:3) to yield 1,4-diketone **23** (5.3 mg, 70%), as an oil.

Compound 23: Oil. ¹H NMR (200 MHz, $CDCl_3$): δ = 8.04 (d, J = 8.3 Hz, 2 H), 7.59 (t, J = 7.2 Hz, 1 H), 7.45 (t, J = 7.7 Hz, 2 H), 2.78 (m, 4 H), 2.19 (s, 3 H), 1.65 (s, 6 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 207.8, 207.3, 165.9, 133.4, 129.8 (2 C), 129.4, 128.4 (2 C), 84.2, 37.0, 30.1, 30.0, 23.8 (2 C) ppm. IR (neat): $\tilde{\nu}_{max}$ = 1715 (C=O benzoate and ketones), 1287, 714 cm^{-1} . HRMS (ESI): calcd. for $C_{15}H_{18}NaO_4$ [M + Na]⁺ 285.1103; found 285.1109.

Compound **24** (107.0 mg, 0.31 mmol) was subjected to the standard oxidative procedure with PCC (2 mol-%) to give 1,4-diketone **25** (109 mg, 98%).

Compound 25: Amorphous solid. ¹H NMR (200 MHz, $CDCl_3$): δ = 8.09 (d, J = 7.2 Hz, 4 H), 7.58 (t, J = 7.5 Hz, 2 H), 7.45 (t, J = 7.7 Hz, 4 H), 4.95 (s, 4 H), 2.87 (s, 4 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 202.7 (2 C), 165.8 (2 C), 133.5 (2 C), 129.9 (4 C), 129.0 (2 C), 128.4 (4 C), 68.3 (2 C), 32.0 (2 C) ppm. IR (neat): $\tilde{\nu}_{max}$ = 1713 (C=O benzoates and ketones), 1275, 715 cm^{-1} . HRMS (ESI): calcd. for $C_{20}H_{18}NaO_6$ [M + Na]⁺ 377.1001; found 377.1015.

Synthesis of Umbelactone

Trihydroxy Acid 43: K_2CO_3 (2.0 equiv. 58 mg, 0.42 mmol) was added to acid **12** (118.1 mg, 0.21 mmol), dissolved in MeOH (5 mL), and the mixture was stirred at room temp. for 2.5 h. Water (1 mL) was added followed by acetic acid up to neutrality. The mixture was taken to dryness in vacuo and the solid was partitioned between a sat. $NaHCO_3$ solution and EtOAc. The organic phase was washed with water, dried and concentrated to give **43** (31.2 mg, 90%), which was used in the next step without further purification.

Compound 43: ¹H NMR (500 MHz, CD_3OD): δ = 3.78 (dd, J = 10.9, 2.5 Hz, 1 H), 3.58 (dd, J = 10.9, 7.9 Hz, 1 H), 3.53 (dd, J = 7.9, 2.5 Hz, 1 H), 2.49 (d, J = 14.9 Hz, 1 H), 2.35 (d, J = 14.9 Hz, 1 H), 1.36 (s, 3 H) ppm. ¹³C NMR (125 MHz, CD_3OD): δ = 180.6, 79.0, 73.5, 63.9, 45.6, 24.7 ppm. HRMS (ESI): calcd. for $C_6H_{12}NaO_5$ [M + Na]⁺ 187.0582; found 187.0590.

Dihydroxylactone 45: Excess CH_2N_2 in Et_2O was added to trihydroxy acid **43** (30.0 mg, 0.18 mmol), dissolved in MeOH (4 mL), until a yellow colour persisted. The mixture was stirred for a further 10 min and the excess of CH_2N_2 was then destroyed by dropwise addition of acetic acid until the solution became colourless. The mixture was taken to dryness to give an oil, which was purified by column chromatography ($CH_2Cl_2/MeOH$ 9:1) to yield lactone **45** (23.5 mg, 90%) as a clear oil.

Compound 45: Oil. ¹H NMR (200 MHz, $CDCl_3$): δ = 4.30 (dd, J = 6.0, 4.4 Hz, 1 H), 3.86 (m, 2 H), 2.77 (d, J = 17.2 Hz, 1 H), 2.50 (d, J = 17.2 Hz, 1 H), 1.44 (s, 3 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 177.9, 90.1, 75.7, 61.0, 45.6, 24.5 ppm. IR (neat): $\tilde{\nu}_{max}$ = 3408 (broad, OHs), 1771 (C=O lactone) cm^{-1} . HRMS (ESI): calcd. for $C_6H_{10}NaO_4$ [M + Na]⁺ 169.0477; found 169.0472.

Silylated Lactone 46: Imidazole (3.0 equiv., 28.6 mg, 0.42 mmol) and *tert*-butyldimethylsilyl chloride (1.5 equiv., 31.5 mg, 0.21 mmol) were added to lactone **45** (21.3 mg, 0.14 mmol), dissolved in DMF (300 μ L), and the mixture was stirred for 2.5 h at room temp. MeOH (500 μ L) was then added and stirring was continued for 30 min. The reaction mixture was concentrated on silica and purified by column chromatography (hexane/EtOAc 1:1) to yield compound **46** (34.9 mg, 96%) as a clear oil.

Compound 46: Oil. ¹H NMR (200 MHz, $CDCl_3$): δ = 4.19 (dd, J = 3.8, 2.8 Hz, 1 H), 4.12 (dd, J = 11.8, 2.8 Hz, 1 H), 4.02 (dd, J = 11.8, 3.8 Hz, 1 H), 3.66 (br. s, 1 H), 2.73 (d, J = 17.6 Hz, 1 H), 2.57 (d, J = 17.6 Hz, 1 H), 1.49 (s, 3 H), 0.90 (s, 9 H), 0.12 (s, 6 H) ppm. ¹³C NMR (50 MHz, $CDCl_3$): δ = 174.7, 85.5, 75.4, 61.8,

44.9, 26.7, 25.7 (3 C), 18.1, -5.6, -5.7 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 3447 (broad, OH), 1782 (C=O lactone) cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{12}\text{H}_{24}\text{NaO}_4\text{Si} [\text{M} + \text{Na}]^+$ 283.1342; found 283.1339.

Unsaturated Lactone 47: Thionyl chloride (5 equiv., 0.55 mmol, 40 μL) was added to lactone **46** (30.0 mg, 0.11 mmol), dissolved in pyridine (500 μL), and the mixture was stirred at 0 °C for 30 min. Water (500 μL) was added and the reaction mixture was evaporated in vacuo to give an oily product. Column chromatography (hexane/EtOAc 8:2) yielded compound **47** (25.3 mg, 95%) as a clear oil.

Compound 47: Oil. ^1H NMR (500 MHz, CDCl_3): δ = 5.83 (br. s, 1 H), 4.82 (br. s, 1 H), 3.94 (dd, J = 11.3, 3.7 Hz, 1 H), 3.90 (dd, J = 11.3, 3.5 Hz, 1 H), 2.10 (s, 3 H), 0.85 (s, 9 H), 0.06 (s, 3 H), 0.05 (s, 3 H) ppm. ^{13}C NMR (125 MHz, CDCl_3): δ = 173.2, 166.7, 118.1, 84.7, 61.7, 25.7 (3 C), 18.1, 14.1, -5.5, -5.6 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 1759 (C=O α,β -unsaturated lactone), 1132, 837, 778 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_{12}\text{H}_{22}\text{NaO}_3\text{Si} [\text{M} + \text{Na}]^+$ 265.1236; found 265.1244.

Umbelactone (48): $\text{Et}_3\text{N}\cdot 3\text{HF}$ (20 equiv., 270 μL) was added to lactone **47** (20.2 mg, 0.083 mmol), dissolved in dry THF (1 mL), and the mixture was stirred for 2.5 h at room temp. Et_3N (500 μL) was then added and the mixture was taken to dryness. The residue was co-evaporated with Et_3N ($2 \times 500 \mu\text{L}$) and MeOH ($3 \times 500 \mu\text{L}$) and concentrated to give an oil, which was purified by column chromatography (EtOAc) to yield umbelactone **48** (9.8 mg, 92%) as a clear oil.

Compound 48: Oil. ^1H NMR (200 MHz, CDCl_3): δ = 5.89 (br. s, 1 H), 4.91 (br. s, 1 H), 4.07 (dd, J = 12.7, 2.9 Hz, 1 H), 3.77 (dd, J = 12.7, 3.9 Hz, 1 H), 2.11 (s, 3 H) ppm. ^{13}C NMR (50 MHz, CDCl_3): δ = 172.8, 165.8, 118.3, 85.0, 61.5, 14.0 ppm. IR (neat): $\tilde{\nu}_{\max}$ = 3404 (broad OH), 2921, 2851, 1731 (broad, C=O α,β -unsaturated lactone), 1647, 1050 cm^{-1} . HRMS (ESI): calcd. for $\text{C}_6\text{H}_8\text{NaO}_3 [\text{M} + \text{Na}]^+$ 151.0371; found 151.0375.

Supporting Information (see footnote on the first page of this article): ^1H and ^{13}C NMR spectra.

Acknowledgments

We are grateful to the Italian Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MURST) (PRIN 2007) for financial support of this investigation. We are also grateful to the Centro di Metodologie Chimico-Fisiche and the Centro di Servizio Interdipartimentale di Analisi Strumentale (CSIAS) of the University of Napoli "Federico II" for NMR facilities, to Dr. Luisa Cuorvo for technical assistance and to Dr. Vincenzo Perino for NMR assistance.

- [1] a) G. Piancatelli, A. Scettri, M. D'Auria, *Synthesis* **1982**, 245–258; b) G. Piancatelli, F. A. Luzzio, in: *Encyclopedia of reagents for organic synthesis*, John Wiley & Sons, Ltd., **2007**; DOI: 10.1002/047084289X.rp288.pub2.
- [2] a) M. Hunsen, *Tetrahedron Lett.* **2005**, *46*, 1651–1653; b) M. Hunsen, *J. Fluorine Chem.* **2005**, *126*, 1356–1360; c) M. Hunsen, *Synthesis* **2005**, 2487–2490.
- [3] S. Roth, C. B. W. Stark, *Chem. Commun.* **2008**, 6411–6413.
- [4] a) L. Albarella, V. Piccialli, D. Smaldone, D. Sica, *J. Chem. Res. Synopses* **1996**, *9*, 400–401; b) L. Albarella, F. Giordano, M. Lasalvia, V. Piccialli, D. Sica, *Tetrahedron Lett.* **1995**, *36*, 5267–5270; c) V. Piccialli, D. Sica, D. Smaldone, *Tetrahedron Lett.* **1994**, *35*, 7093–7096.
- [5] a) V. Piccialli, S. Zaccaria, N. Borbone, G. Oliviero, S. D'Ernico, A. Hemminki, V. Cerullo, V. Romano, A. Tuzi, R. Centore, *Tetrahedron* **2010**, *66*, 9370–9378; b) V. Piccialli, G. Oliviero, N. Borbone, A. Tuzi, R. Centore, A. Hemminki, M. Ugolini, V. Cerullo, *Org. Biomol. Chem.* **2009**, *7*, 3036–3039; c) V. Piccialli, N. Borbone, G. Oliviero, *Tetrahedron* **2008**, *64*, 11185–11192; d) V. Piccialli, *Synthesis* **2007**, *17*, 2585–2607; e) V. Piccialli, T. Caserta, L. Caruso, L. Gomez-Paloma, G. Bifulco, *Tetrahedron* **2006**, *62*, 10989–11007; f) T. Caserta, V. Piccialli, L. Gomez-Paloma, G. Bifulco, *Tetrahedron* **2005**, *61*, 927–939; g) V. Piccialli, T. Caserta, *Tetrahedron Lett.* **2004**, *45*, 303–308; h) G. Bifulco, T. Caserta, L. Gomez-Paloma, V. Piccialli, *Tetrahedron Lett.* **2003**, *44*, 5499–5503; i) G. Bifulco, T. Caserta, L. Gomez-Paloma, V. Piccialli, *Tetrahedron Lett.* **2002**, *43*, 9265–9269; corrigendum: *Tetrahedron Lett.* **2003**, *44*, 3429; j) V. Piccialli, N. Cavallo, *Tetrahedron Lett.* **2001**, *42*, 4695–4699; k) M. de Champdoré, M. La Salvia, V. Piccialli, *Tetrahedron Lett.* **1998**, *39*, 9781–9784.
- [6] A. Okamura, M. Kitani, M. Murata, *Bull. Chem. Soc. Jpn.* **1994**, *67*, 1522–1530.
- [7] a) H. Hioki, S. Yoshio, M. Motosue, Y. Oshita, Y. Nakamura, D. Mishima, Y. Fukuyama, M. Kodama, K. Ueda, T. Katsu, *Org. Lett.* **2004**, *6*, 961–964, and references cited therein b) J. J. Fernandez, M. L. Souto, M. Norte, *Nat. Prod. Rep.* **2000**, *17*, 235–246; c) F. Q. Alali, X.-X. Liu, J. L. McLaughlin, *J. Nat. Prod.* **1999**, *62*, 504–540; d) S. Sasaki, K. Maruta, H. Naito, R. Maemura, E. Kawahara, M. Maeda, *Tetrahedron* **1998**, *54*, 2401–2410; e) Y. Morimoto, T. Iwai, T. Yoshimura, T. Kinoshita, *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2005–2010; f) J.-F. Peyrat, J. Mahuteau, B. Figadere, A. Cave, *J. Org. Chem.* **1997**, *62*, 4811–4815; g) L. Zeng, N. H. Oberlies, G. Shi, Z.-M. Gu, K. He, J. L. McLaughlin, *Nat. Prod. Rep.* **1996**, *13*, 275–306; h) J.-F. Peyrat, B. Figadere, A. Cave, J. Mahuteau, *Tetrahedron Lett.* **1995**, *36*, 7653–7656; i) S. Sasaki, K. Maruta, H. Naito, H. Sugihara, K. Iratani, M. Maeda, *Tetrahedron Lett.* **1995**, *36*, 5571–5574; j) S. Sasaki, H. Naito, K. Maruta, E. Kawahara, M. Maeda, *Tetrahedron Lett.* **1994**, *35*, 3337–3340; k) H. Morita, E. K. Kishi, K. Takeya, H. Itokawa, Y. Iitaka, *Phytochemistry* **1993**, *34*, 765–771; l) J. K. Rupprecht, Y. H. Hui, J. L. McLaughlin, *J. Nat. Prod.* **1990**, *53*, 237–278; m) N. Otake, T. Sasaki, *Agric. Biol. Chem.* **1977**, *41*, 1039–1047.
- [8] S. Gohler, S. Roth, H. Cheng, H. Goksel, A. Rupp, L. O. Hausstedt, C. B. W. Stark, *Synthesis* **2007**, *17*, 2751–2754.
- [9] M. Zhao, J. Li, Z. Song, R. Desmond, D. M. Tschaen, E. J. J. Grabowski, P. J. Reider, *Tetrahedron Lett.* **1998**, *39*, 5323–5326.
- [10] S. Roth, S. Gohler, H. Cheng, C. B. W. Stark, *Eur. J. Org. Chem.* **2005**, *19*, 4109–4118.
- [11] Y.-L. Zhong, T. K. M. Shing, *J. Org. Chem.* **1997**, *62*, 2622–2624.
- [12] a) A. Tenaglia, E. Terranova, B. Waegell, *J. Org. Chem.* **1992**, *57*, 5523–5528; b) B. Plietker, *Synthesis* **2005**, *15*, 2453–2472.
- [13] **RuO₄**: a) M. Drees, T. Strassner, *J. Org. Chem.* **2006**, *71*, 1755–1760; b) J. M. Bakke, A. E. Frøhaug, *J. Phys. Org. Chem.* **1996**, *9*, 507–513; c) J. M. Bakke, A. E. Frøhaug, *J. Phys. Org. Chem.* **1996**, *9*, 310–318; d) J. L. Coudret, S. Zoellner, B. J. Ravoo, L. Malara, C. Hanisch, K. Dorre, A. de Meijere, B. Waegell, *Tetrahedron Lett.* **1996**, *37*, 2425–2428 and references cited therein. **OsO₄**: e) B. C. Bales, P. Brown, A. Dehestani, J. M. Mayer, *J. Am. Chem. Soc.* **2005**, *127*, 2832–2833; f) J. M. Mayer, E. A. Mader, J. P. Roth, J. R. Bryant, T. Matsuo, A. Dehestani, B. Bales, E. J. Watson, T. Osako, K. Valliant-Saunders, W. H. Lam, D. A. Hrovat, W. T. Borden, E. R. Davidson, *J. Mol. Catal. A* **2006**, *251*, 24–33; g) T. Osako, E. J. Watson, A. Dehestani, B. C. Bales, J. M. Mayer, *Angew. Chem.* **2006**, *118*, 7593; *Angew. Chem. Int. Ed.* **2006**, *45*, 7433–7436.
- [14] G. D. Lee, M. van den Engh, *Can. J. Chem.* **1972**, *50*, 3129–3134.
- [15] a) S. Baskaran, I. Islam, S. Chandrasekaran, *J. Org. Chem.* **1990**, *55*, 891–895; b) M. Fetizon, P. Goulaouic, I. Hanna, *Tetrahedron Lett.* **1988**, *29*, 6261–6264; c) S. Baskaran, I. Islam, M. Raghavan, S. Chandrasekaran, *Chem. Lett.* **1987**, 1175–1178.

- [16] G. Piancatelli, A. Scettri, M. D'Auria, *Tetrahedron Lett.* **1977**, *18*, 3483–3484.
- [17] A. B. Smith III, R. M. Scarborough, *Synth. Commun.* **1980**, *10*, 205–211.
- [18] S. K. Agarwal, R. P. Rastogi, *Phytochemistry* **1978**, *17*, 1663–1664.
- [19] For synthesis of racemic **48**, see: a) D. Caine, A. S. Frobese, V. C. Ukachukwu, *J. Org. Chem.* **1983**, *48*, 740–741; b) P. Bonete, C. Nájera, *J. Org. Chem.* **1994**, *59*, 3202–3209. For synthesis of (+)-**31** and/or (–)-**31**, see: c) R. M. Ortuño, J. Bigorra, J. Font, *Tetrahedron* **1987**, *43*, 2199–2202; d) T. Sato, Y. Okumura, J. Itai, T. Fujisawa, *Chem. Lett.* **1988**, 1537–1540; e) R. K. Boeckman, A. B. Charette, T. Asberom, B. H. Johnston, *J. Am. Chem. Soc.* **1991**, *113*, 5337–5353; f) C. L. Gibson, S. Handa, *Tetrahedron: Asymmetry* **1996**, *7*, 1281–1284; g) H.-J. Ha, K.-N. Yoon, S.-Y. Lee, Y.-S. Park, M.-S. Lim, Y.-G. Yim, *J. Org. Chem.* **1998**, *63*, 8062–8066; h) H. W. Liu, Y. L. Li, *Chin. Chem. Lett.* **2005**, *16*, 716–718; i) H. Liu, T. Zhang, Y. Li, *Chirality* **2006**, *18*, 223–226; j) A. Kamal, T. Krishnaji, P. V. Reddy, *Tetrahedron Lett.* **2007**, *48*, 7232–7235; k) W. P. D. Goldring, J. Mann, P. Brockbank, *Synlett* **2010**, *4*, 547–550.
- [20] We are grateful to Dr. Goldring, Queen's University Belfast, for providing us with NMR spectroscopic data for the C-3 epimer of **46**.
- [21] T. J. Donohoe, S. Butterworth, *Angew. Chem.* **2005**, *117*, 4844; *Angew. Chem. Int. Ed.* **2005**, *44*, 4766–4768.
- [22] See, for instance: P. Setzer, A. Beauseigneur, M. S. M. Pearson-Long, P. Bertus, *Angew. Chem.* **2010**, *122*, 8873; *Angew. Chem. Int. Ed.* **2010**, *49*, 8691–8694, and references cited therein.

Received: January 19, 2012
Published Online: June 14, 2012



Probing the reactivity of nebularine N1-oxide. A novel approach to C-6 C-substituted purine nucleosides

Stefano D'Errico^a, Vincenzo Piccialli^b, Giorgia Oliviero^{a,c,*}, Nicola Borbone^a, Jussara Amato^a, Valentina D'Atri^a, Gennaro Piccialli^{a,c}

^aDipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli Federico II, Via D. Montesano 49, 80131 Napoli, Italy

^bDipartimento di Chimica Organica e Biochimica, Università degli Studi di Napoli Federico II, Via Cynthia 4, 80126 Napoli, Italy

^cFacoltà di Scienze Biotecnologiche, Università degli Studi di Napoli Federico II, Via D. Montesano 49, 80131 Napoli, Italy

ARTICLE INFO

Article history:

Received 9 May 2011

Received in revised form 10 June 2011

Accepted 24 June 2011

Available online 30 June 2011

ABSTRACT

A novel approach to the synthesis of purine nucleoside analogues, featuring the reaction of the C6–N1–O[−] aldonitrone moiety of 9-ribose-purine (nebularine) N1-oxide with some representative dipolarophiles, as well as Grignard reagents, is reported. Addition of Grignard reagents to the electrophilic C-6 carbon of the substrate allows a facile access to C-6 C-substituted purine nucleosides without using metal catalysts. 1,3-Dipolar cycloaddition processes lead to novel nucleoside analogues via opening, degradation or ring-enlargement of the pyrimidine ring of the base system of the first-formed isoxazoline or isoxazolidine cycloadduct.

© 2011 Elsevier Ltd. All rights reserved.

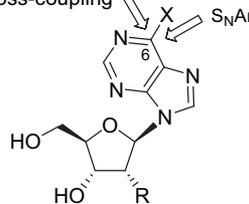
1. Introduction

In the last decades many research groups have focused their attention to the preparation of new modified nucleosides and nucleotides with the aim of expanding the pool of molecules with potential antineoplastic, antihypertensive and antiviral activities.¹ In this context, efforts have been directed to the synthesis of sugar² and/or base³-modified nucleosides. A large number of nucleobase analogues exist and several nucleoside analogues have been employed against cancer and viral diseases. In addition, base-modified nucleosides often show fluorescent properties,⁴ and can be used as fluorescent probes for the analysis of DNA and RNA structures as well as for analysing the interaction of DNA and RNA with binding proteins. Purine bases and nucleosides bearing a C- or N-substituent at C-6 represent an important class of compounds possessing a broad spectrum of biological effects including cytostatic, antiviral, antibacterial as well as receptor–modulation activity.⁵ The reactivity imparted to purines and related nucleosides by halogenation at C-6 has opened the way to the construction of new libraries of C-6 modified nucleosides generally through direct aromatic nucleophilic substitution (S_NAr),⁶ or metal-mediated cross-coupling processes^{3a,7} (Fig. 1). The most reliable methods to access C-6 C-substituted nucleosides use metal or organometal-mediated reactions.^{3a} However, there is still a great need for the

development of new methods for the introduction of C-substituents at C-6.

PREVIOUS APPROACHES

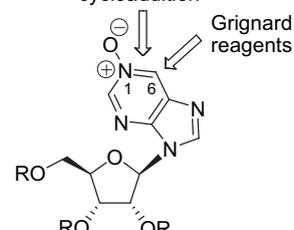
Metal-mediated cross-coupling



R=H, deoxyribose
R=OH, ribose
X=Cl, Br, I

OUR APPROACH

1,3-Dipolar cycloaddition



1a R=H (Nebularine N-1 oxide)
1b R=Ac
1c R=TBDMS

Fig. 1. Approaches to C-6 functionalization of purine nucleosides.

Nitrones are precious substrates used in organic synthesis for the assembly of structurally complex nitrogen-containing compounds. Their most well-studied reactions are the 1,3-dipolar cycloadditions (1,3-DC) and the nucleophilic addition of organometallic reagents.⁸ We reasoned that 9-ribose-purine (nebularine) N1-oxide (**1a**, Fig. 1), embodying a potentially reactive C6–N1–O[−] nitrone moiety, could allow the C-6 functionalization of the purine base. To probe this concept the reactivity of sugar-protected **1b** and **1c** (Fig. 1) towards some representative

* Corresponding author. Tel.: +39 081678540; fax: +39 081678552; e-mail address: golivier@unina.it (G. Oliviero).

dipolarophiles as well as Grignard reagents, respectively, was explored. The results obtained are described in the present paper.

2. Results and discussion

2',3',5'-Tri-*O*-acetyl-nebularine N1-oxide (**1b**, Fig. 1) was used as starting material for 1,3-DC processes whereas 2',3',5'-tri-*O*-(*tert*-butyldimethylsilyl)-nebularine N1-oxide (**1c**, Fig. 1) was employed to test the addition of Grignard reagents. Compounds **1b** and **1c** were synthesised from 2',3',5'-tri-*O*-acetyl-nebularine and 2',3',5'-tri-*O*-(*tert*-butyldimethylsilyl)-nebularine, respectively,⁹ by reaction of nebularine with catalytic amounts of methyltrioxorhenium (MeReO₃) in the presence of H₂O₂, as described in the literature for the preparation of the corresponding purine N1-oxide.¹⁰ Though the cycloaddition of aromatic *N*-oxides of various heterocyclic systems has extensively been investigated,¹¹ to our knowledge no report exists in the literature on the reactivity of *N*-oxides of purine nucleosides in cycloaddition reactions.

Reaction of compound **1b** with some representative dipolarophiles was carried out as shown in Table 1 and Scheme 1, to give the novel nucleoside analogues **2** and **4–7**, the structure of which was determined by high-field 2D-NMR analyses. As expected,¹² in all the analysed cases, we observed that the first-formed

isoxazoline or isoxazolidine adduct did not survive. Therefore, the nucleoside products obtained from such processes are all derived from the cleavage of the N1–O bond in the cycloadduct followed in some cases by further evolution of the first-formed C-6 substituted purine system. In particular, reaction of **1b** with an equimolar amount of dimethyl acetylenedicarboxylate in THF gave the two unusual nucleoside analogues **2** (25%) and **4** (40%) (Table 1, entry 1) possessing modified heterocyclic base systems. Compound **2** proved unstable as such giving a mixture of partially deacetylated products on standing. However, its fully deacetylated derivative **3**, obtained by treatment of **2** with NH₄OH (concd) in MeOH, was a stable product, which was used for NMR characterisation. Compound **4** contains the 5:7-fused imidazo[4,5-*d*][1,3]diazepine ring system. The ¹H NMR spectrum of **4** recorded in CD₃OD showed it to be a mixture of two tautomers in a 7:3 ratio. 2D-NMR studies carried out in this solvent gave no conclusive evidences on the structure of **4** that was eventually solved by performing NMR analyses in DMSO-*d*₆ where the major tautomer **4** was present in 88% amount. Few examples of nucleoside analogues based on this base framework (ring-expanded nucleosides, RENs) exist all displaying significant biological activities.¹³ The interest towards this type of substances is keen and the facile access to this framework is particularly appealing. In particular, compound **4** is the first example of a C-7 analogue of coformycin,^{13c} (Fig. 2) a highly potent inhibitor of adenosine deaminase (ADA), an enzyme playing a key role in purine metabolism.¹⁴

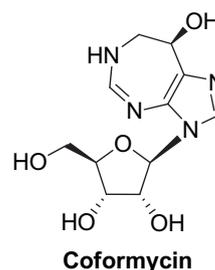
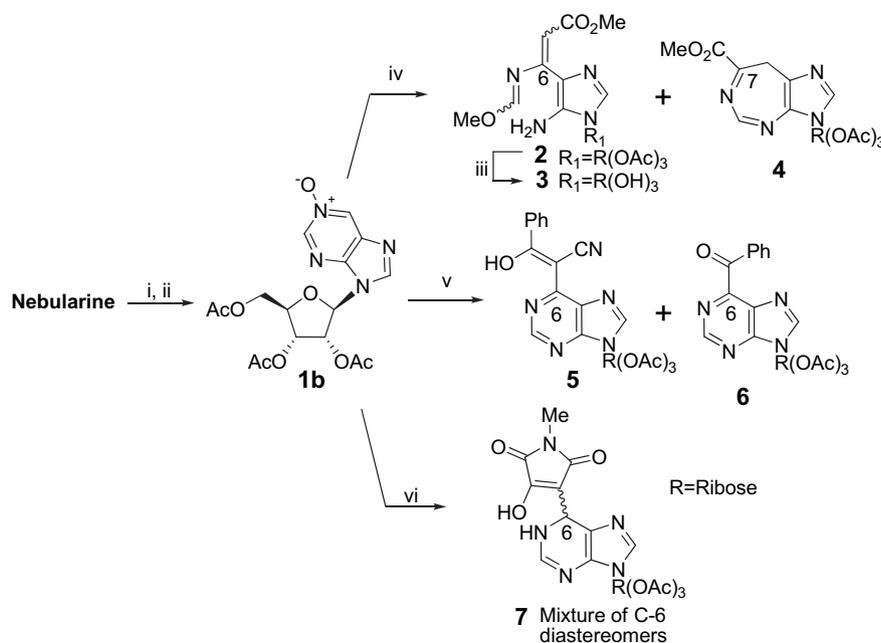


Fig. 2. Structure of coformycin.

Table 1
Reaction of **1b** with dipolarophiles

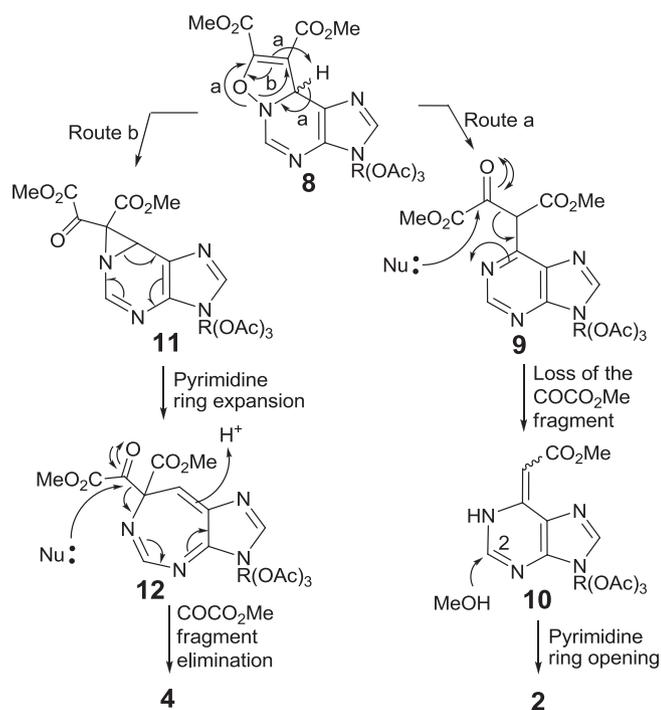
Entry	Dipolarophile	Conditions (solvent)	Products (yield %)
1	Dimethyl acetylenedicarboxylate	rt, 2 h (THF)	2 (25%), 4 (40%)
2	3-Phenyl-2-propynenitrile	rt, 16 h (THF)	NR
3	3-Phenyl-2-propynenitrile	Reflux, 4 h (dioxane)	5 (45%), 6 (30%)
4	<i>N</i> -Methylmaleimide	rt, 5 h or reflux, overnight (THF)	NR
5	<i>N</i> -Methylmaleimide	Reflux, 5 h, (dioxane/toluene, 1:1)	7 (60%)
6	Diphenylacetylene	Reflux, overnight (THF or dioxane)	NR
7	Dimethyl maleate or dimethyl fumarate	Reflux, overnight (toluene)	NR

NR: no reaction.



Scheme 1. Reaction of **1b** with dipolarophiles: i. Ac₂O/pyridine;⁹ ii. MeReO₃–H₂O₂ (30% aq), MeOH, rt, overnight (80% over two steps); iii. NH₄OH (concd)–MeOH, rt, 30 min; iv. Dimethyl acetylenedicarboxylate, THF, rt, 2 h; v. 3-phenyl-2-propynenitrile, dioxane, reflux, 4 h; vi. *N*-methylmaleimide, dioxane/toluene (1:1), reflux, 5 h.

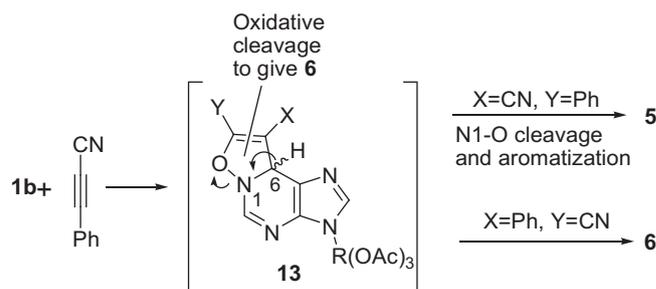
A tentative explanation for the formation of **2** and **4** is depicted in Scheme 2. Generation of their base system calls for the extrusion of a two-carbon fragment from the initially-formed isoxazoline adduct **8**. In particular, formation of **2** can be explained via isoxazoline opening with concomitant base aromatization, to give the C-6 substituted species **9** (route a) followed by loss of a CO₂Me fragment with formation of a two-carbon chain at C-6 (**10**).¹² Evolution of **9** towards the conjugated species **10** via attack of a nucleophile at the ketone carbonyl seems logical and would well explain formation of the exocyclic double bond in **10**. Transformation of dimethyl α -ketosuccinates, such as **9** into methyl acetates by loss of this structural fragment has previously been reported.¹⁵ The electron-withdrawing effect of the exocyclic unsaturated ester moiety in the latter may then be responsible for the opening of the pyrimidine ring through attack of MeOH at C-2, possibly in the work-up and/or purification step, to eventually give **2**. This type of reactivity at C-2 in **10** is strongly reminiscent of that exhibited by the N1-2,4-dinitrophenyl-inosine derivatives, previously studied in our group, leading to formation of AICAR derivatives.¹⁶ The imidazo 4-substituted framework of **2** is unprecedented.



Scheme 2. A mechanistic hypothesis for the formation of **2** and **4**.

As for compound **4**, Stauss et al.¹⁷ reported formation of 2-methyl-4-phenyl-5H-benzo[d][1,3]diazepin-5-carboxylic acid esters as by-products of the reaction of 2-methyl-4-phenyl-quinazoline 3-oxide with dimethyl and diethyl acetylenedicarboxylates and a plausible mechanistic hypothesis was given for the reaction path leading to the benzodiazepine system. In particular, formation of a cyclopropane-containing intermediate was postulated to be responsible for the pyrimidine ring enlargement. In a similar way, it is conceivable that the seven-membered ring in **4** can originate from the electrocyclic opening of an aziridine-containing intermediate (**11**, Scheme 2), in turn formed by the fragmentation/rearrangement of the initial adduct **8** (route b). The ring-enlarged compound **12** would once again undergo elimination of the α -ketoester side-chain possibly via nucleophilic attack at the ketone carbonyl during work-up.

Non-symmetric 3-phenyl-2-propynenitrile failed to react with **1b** at room temperature in THF (Table 1, entry 2) but gave a mixture of C-6 substituted nucleosides **5** (45%) and **6** (30%) when the process was conducted in dioxane at reflux (entry 3). Compound **5** is likely derived from the opening of isoxazoline ring in one of the two expected diastereomeric adducts **13** (Scheme 3) induced by loss of the C6-H proton and aromatization of the heterocyclic system. On the other hand, it seems hazardous trying to hypothesize a mechanistic route to C-6 benzoylated **6**. However, it can be envisaged that the oxidative cleavage of the double bond in the other adduct (**13**, X=Ph, Y=CN), with loss of a two-carbon fragment and aromatization, should occur to generate **6**. As far as we know, the direct introduction of acyl substituents at C-6 of the purine nucleoside system has never been reported before, and this type of substances is still rather rare.



Scheme 3. A mechanistic hypothesis for the formation of **5** and **6**.

Early attempts to induce reaction of **1b** with *N*-methylmaleimide in THF, both at room temperature and at reflux (Table 1, entry 4), were unsuccessful. However, when the process was carried out in dioxane/toluene (1:1) at reflux (entry 5) the C-6 derivative **7** was obtained in a 60% yield as a 1:1 mixture of diastereomers, once again through isoxazoline opening and oxidation at the maleimide portion. The ratio of isomers in the initially-formed mixture varied with HPLC purification (silica column) and/or on standing up to ca. 1:2, likely as a consequence of the acidity of the C-6 proton next to the maleimide moiety. NMR data could not provide unambiguous evidence on which is the major diastereomer. Considering that the maleimide portion in **7** is susceptible to further synthetic modifications,¹⁸ access to C-6 maleimido-derivatives, such as **7** opens the way to the preparation of more complex and functionalised C-6 nucleoside derivatives. Maleimido- or *N*-methylmaleimido-containing substances, both of natural and non-natural origin, usually display significant biological properties. For example, showdomycin¹⁹ is a potent nucleoside antibiotic isolated from *Streptomyces showdoensis*, while bisindolylmaleimides are known to be potent protein kinase C (PKC) inhibitors.²⁰

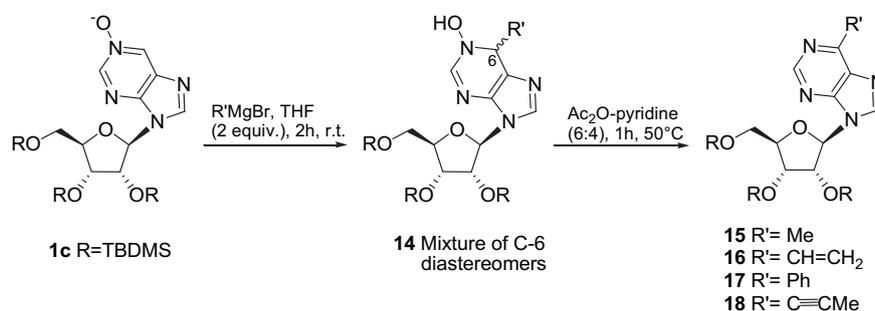
Finally, diphenylacetylene failed to react with **1b** also on prolonged reflux both in THF and dioxane (Table 1, entry 6). Similar results were obtained with dimethyl maleate and dimethyl fumarate on reflux in toluene, also using excess reagents (Table 1, entry 7). The above results showed that nebularine oxide derivative **1b** displays the characteristic reactivity of nitrones towards electron-poor dienophiles.

To further test the nitron reactivity showed by **1a**, we next investigated the reaction of sugar-silylated **1c** with some representative Grignard reagents. It is known that nitrones smoothly undergo nucleophilic addition of organometallic reagents. Indeed, treatment of **1c** with a 2-fold excess of a Grignard reagent afforded the C-6 C-substituted N1-hydroxy adducts **14** as single products (Table 2 and Scheme 4). These substances slowly decomposed on standing, partly giving dehydration to C-6 substituted nebularine derivatives **15–18**. This process was pushed towards the latter

substances in nearly quantitative yields, by treatment of the crude C6-adducts with Ac₂O in pyridine (6:4) at 50 °C. Formation of **15–18** indicated, as reported in the literature,²¹ the more electrophilic nature of the C-6 atom of the nucleobase when compared to the C-2 atom. The structure of the obtained compounds was ascertained by high-field 2D NMR and comparison with literature data (see [Experimental section](#)).

Table 2
Reaction of **1c** with Grignard reagents

Entry	Grignard reagent (equiv)	Conditions (solvent)	Products (yield%, over two steps)
1	Methylmagnesium bromide (2)	rt, 2 h (THF)	15 (82%)
2	Vinylmagnesium bromide (2)	rt, 2 h (THF)	16 (71%)
3	Phenylmagnesium bromide (2)	rt, 2 h (THF)	17 (82%)
4	Propenylmagnesium bromide (2)	rt, 2 h (THF)	18 (80%)



Scheme 4. Reaction of **1c** with representative Grignard reagents.

Overall, reaction of **1c** with Grignard reagents is of general applicability, working well with various types of substituents on the organometallic partner, leading to compounds **15–18** in good yields comparable²² to, or better^{7,23} than, those reported in the literature. It is worth noting that our process allows the direct introduction of a C-substituent on C-6 without using a metal catalyst. Based on both the facile reaction and purification procedures it is conceivable that the process can be easily performed on gram scale opening the way to obtain large amounts of C-6 C-substituted nucleosides without employing different reaction conditions.

3. Conclusions

In conclusion, an unprecedented approach to the C-6 C-functionalization of the purine nucleoside base system has been explored. In some cases, 1,3-DC reactions led to unexpected, structurally intriguing, products through new reaction pathways. Studies are in progress to investigate the scope of the described processes as well as to further explore the synthetic potential of nebularine N1-oxide. Since purine nucleosides, such as **14**, embodying the 6,9-dihydro-1H-purin-1-ol moiety are unprecedented, future efforts will also focus on using these addition products to prepare nucleoside analogues possessing novel modifications on the base system. We believe that the developed procedures may have future synthetic applications.

4. Experimental section

4.1. General methods

All reagents were obtained from commercial sources (Sigma–Aldrich) and were used without further purification. ¹H and ¹³C

NMR spectra were performed on a Varian Mercury Plus 400 MHz and Varian Unity Inova 700 MHz in CDCl₃, CD₃OD or DMSO-*d*₆ solvents. Chemical shifts are reported in parts per million (δ) relative to the residual solvent signals: CHCl₃ 7.27, CD₂HOD 3.31, CD₂HSOCD₃ 2.49 for ¹H NMR; CDCl₃ 77.0, CD₂HOD 49.0, CD₂HSOCD₃ 39.5 for ¹³C NMR. ¹H NMR chemical shifts were assigned by 2D-NMR experiments. The abbreviations s, br s, d, dd and m stand for singlet, broad singlet, doublet, doublet of doublets and multiplet, respectively. HPLC analyses and purifications were carried out on a Jasco UP-2075 Plus pump equipped with a Jasco UV-2075 Plus UV detector using a 4.60×150 mm LUNA (Phenomenex) silica column (particle size 5 μm) eluted with a linear gradient of AcOEt in hexane (from 50 to 100% in 30 min, flow 1.0 mL min⁻¹, system A; from 80 to 100% in 30 min, flow 1.0 mL min⁻¹, system B; from 50 to 100% in 60 min, flow 1.0 mL min⁻¹, system C) or using a 4.8×150 mm C-18 reverse-phase column (particle size 5 μm) eluted with a linear gradient of MeOH

in H₂O (from 0 to 100% in 60 min, flow 1.0 mL min⁻¹, system D). UV spectra were recorded on a Jasco V-530 UV spectrophotometer. High Resolution MS spectra were recorded on a Bruker APEX II FT-ICR mass spectrometer using electrospray ionization (ESI) technique in positive mode. IR spectra were recorded on a Jasco FT-IR/FTIR 430 spectrophotometer. Optical rotations were determined on a Jasco polarimeter using a 1 dm cell at 20 °C; concentrations are in g/100 mL. Column chromatography was performed on silica gel (Merck, Kieselgel 60, 0.063–0.200 mm). Analytical TLC analyses were performed using F₂₅₄ silica gel plates (0.2 mm, Merck). TLC spots were detected under UV light (254 nm).

4.2. Synthesis of 2',3',5'-tri-O-acetyl-nebularine N1-oxide (**1b**)

Nebularine was acetylated overnight with Ac₂O–pyridine under standard conditions to give essentially pure triacetate. A mixture of MeReO₃ (5 mg, 0.02 mmol) and 30% aqueous H₂O₂ (0.4 mL, 4.0 mmol) in methanol (6 mL) was stirred at room temperature for 10 min. Then crude 2',3',5'-tri-O-acetyl-nebularine (378 mg, 1.0 mmol) in methanol (4 mL) was added dropwise while stirring. Stirring was continued at room temperature overnight. Then the solvent was removed under reduced pressure and the crude was purified on a silica gel column eluted with increasing amounts of MeOH in CHCl₃ (from 0 to 10%) to afford pure 2',3',5'-tri-O-acetyl-nebularine N1-oxide **1b** (315 mg, 80% over two steps).

Compound **1b**: white foam; [α]_D²⁰ –16.1 (c 0.13, CH₃OH), ¹H NMR (400 MHz, CD₃OD) δ_H 9.16 (d, *J*=1.8 Hz, 1H), 9.04 (d, *J*=1.8 Hz, 1H), 8.76 (s, 1H), 6.34 (d, *J*=5.0 Hz, 1H), 6.03–5.98 (m, 1H), 5.69–5.65 (m, 1H), 4.52–4.48 (m, 1H), 4.44 (dd, *J*=12.2, 3.5 Hz, 1H), 4.39 (dd, *J*=12.2, 4.9 Hz, 1H), 2.14 (s, 3H), 2.07 (s, 3H), 2.06 (s, 3H); ¹³C NMR (100 MHz, CD₃OD) δ_C 172.1, 171.3, 171.1, 150.5, 146.4, 145.3, 139.3, 135.5, 88.5, 81.7, 74.2, 71.7, 64.0, 20.6, 20.4, 20.2; *m/z* (HRESIMS)

395.1212 ($[M+H]^+$, $C_{16}H_{19}N_4O_8$, requires 395.1203); IR (neat) ν_{\max} 1748, 1495, 1424, 1369, 1229 cm^{-1} ; UV (MeOH) λ_{\max} 242, shoulder 267 nm.

4.3. Reaction of **1b** with dimethyl acetylenedicarboxylate. Synthesis of **2–4**

A solution of **1b** (50 mg, 0.13 mmol) and dimethyl acetylenedicarboxylate (18 mg, 0.13 mmol) in dry THF (2.5 mL) was stirred at room temperature. After 2 h the process was complete (TLC monitoring, AcOEt/MeOH, 95:5) and the solvent was evaporated under reduced pressure. The crude was purified on a silica gel column eluted with increasing amounts of MeOH in AcOEt (from 0 to 20%) affording two main products: compound **2** (16 mg, 25%) and **4** (23 mg, 40%). The purity of **4** was checked by HPLC analysis (System A, $t_R=30.0$ min see General methods). Compound **2** (10 mg, 0.021 mmol) was treated with a concentrated aqueous solution of NH_4OH (0.5 mL) in MeOH (0.5 mL) for 30 min at room temperature. The solvent was removed in vacuo and the crude was desalified by RP-HPLC (System D, $t_R=18.8$ min, see General methods) to give pure **3** (7.4 mg, 99%).

Compound **3**: white foam; $[\alpha]_D^{20} -15.6$ (c 0.50, CH_3OH), 1H NMR (400 MHz, CD_3OD) δ_H 7.89 (s, 1H, H-2), 7.87 (s, 1H, HC=N), 7.71 (s, 1H, HC=C), 5.82 (d, $J_{1,2'}=4.8$ Hz, 1H, H-1'), 4.44–4.41 (m, 1H, H-2'), 4.27–4.24 (m, 1H, H-3'), 4.02–3.99 (m, 1H, H-4'), 3.87 (s, 3H, $CH_3OC=N$), 3.82 (dd, $J_{5'a,5'b}=12.1$, $J_{5'a,4'}=3.2$ Hz, 1H, $H_{a-5'}$), 3.72 (dd, $J_{5'b,5'a}=12.2$, $J_{5'b,4'}=3.8$ Hz, 1H, $H_{b-5'}$), 3.63 (s, 3H, $CH_3OC=O$); ^{13}C NMR (100 MHz, CD_3OD) δ_C 171.4, 158.9, 150.4, 134.7, 133.1, 118.9, 92.7, 89.3, 85.7, 76.0, 71.2, 62.3, 53.9, 51.2; m/z (HRESIMS) 357.1421 ($[M+H]^+$, $C_{14}H_{21}N_4O_7$, requires 357.1410); IR (neat) ν_{\max} 3330, 1634, 1575 cm^{-1} ; UV (MeOH) λ_{\max} 271 nm.

Compound **4**: oil; 1H NMR (700 MHz, $DMSO-d_6$) δ_H (major tautomer) 8.90 (s, 1H, H-5), 8.77 (s, 1H, H-2), 6.34 (d, $J=5.3$ Hz, 1H, H-1'), 6.10–6.07 (m, 1H, H-2'), 5.68–5.63 (m, 1H, H-3'), 4.42–4.39 (complex signal, 2H, H-4' and $H_{a-5'}$), 4.25 (dd, $J_{5'b,5'a}=12.9$, $J_{5'b,4'}=6.0$ Hz, 1H, $H_{b-5'}$), 4.20 and 4.19 (br s, 1H each, $H_{a,b-8}$), 3.63 (s, 3H, OCH_3), 2.12 (s, 3H, CH_3), 2.03 (s, 3H, CH_3), 1.99 (s, 3H, CH_3). ^{13}C NMR (175 MHz, CD_3OD) δ_C 169.9, 169.4, 169.2 (two carbons), 154.5, 151.9, 150.3, 145.4, 132.9, 85.9, 79.5, 71.8, 69.9, 62.6, 51.9, 38.0, 20.4, 20.3, 20.1; IR (neat) ν_{\max} 1746, 1598, 1223 cm^{-1} ; m/z (HRESIMS) 451.1451 ($[M+H]^+$, $C_{19}H_{23}N_4O_9$, requires 451.1465); UV (MeOH) λ_{\max} 263 nm.

4.4. Reaction of **1b** with 3-phenyl-2-propynenitrile. Synthesis of **5** and **6**

A solution of **1b** (50 mg, 0.13 mmol) and 3-phenyl-2-propynenitrile (14 mg, 0.16 mmol), in dioxane (2.5 mL) was refluxed for 4 h (TLC monitoring, AcOEt/hexane, 8:2) and then the solvent was evaporated under reduced pressure. The crude was purified on a silica gel column eluted with increasing amounts of AcOEt in hexane (from 50 to 100%) to give **5** (30 mg, 45%) and **6** (19 mg, 30%), the purity of which was checked by HPLC analysis (System C, **5** $t_R=22.3$ min; **6** $t_R=18.2$ min, see General methods).

Compound **5**: white solid mp 116–117 °C; $[\alpha]_D^{20} -50.6$ (c 0.11, CH_3OH), 1H NMR (400 MHz, CD_3OD) δ_H 8.62 (s, 1H, H-2), 8.46 (s, 1H, H-8), 7.83–7.78 (m, 2H, Ph *ortho*-H), 7.54–7.50 (m, 1H, Ph *para*-H), 7.49–7.44 (m, 2H, Ph *meta*-H), 6.33 (d, $J_{1,2'}=5.0$ Hz, 1H, H-1'), 6.00–5.96 (m, 1H, H-2'), 5.69–5.66 (m, 1H, H-3'), 4.52–4.36 (complex signal, 3H, H-4' and $H_{a,b-5'}$), 2.14 (s, 3H, CH_3), 2.08 (s, 3H, CH_3), 2.07 (s, 3H, CH_3); ^{13}C NMR (175 MHz, CD_3OD) δ_C 194.7, 172.2, 171.4, 171.2, 152.2, 149.1, 145.3, 143.5, 140.7, 132.3, 129.0, 125.1, 121.7, 88.4, 81.8, 74.6, 71.8, 64.1, 20.6, 20.5, 20.3; IR (neat) ν_{\max} 2205, 1742, 1616, 1239 cm^{-1} ; m/z (HRESIMS) 522.1620 ($[M+H]^+$, $C_{25}H_{24}N_5O_8$, requires 522.1625); UV (MeOH) λ_{\max} 364 nm.

Compound **6**: oil; $[\alpha]_D^{20} -6.33$ (c 0.6, CH_3OH), 1H NMR (400 MHz, CD_3OD) δ_H 9.10 (s, 1H, H-2), 8.71 (s, 1H, H-8), 8.00–7.95 (m, 2H, Ph *ortho*-H), 7.73–7.67 (m, 1H, Ph *para*-H), 7.57–7.51 (m, 2H, Ph *meta*-H), 6.41 (d, $J_{1,2'}=5.0$ Hz, 1H, H-1'), 6.16–6.12 (m, 1H, H-2'), 5.78–5.75 (m, 1H, H-3'), 4.53–4.39 (complex signal, 3H, H-4' and $H_{a,b-5'}$), 2.16 (s, 3H, CH_3), 2.08 (s, 3H, CH_3), 2.06 (s, 3H, CH_3); ^{13}C NMR (100 MHz, CD_3OD) δ_C 192.8, 172.3, 171.4, 171.1, 154.7, 154.0, 152.9, 148.3, 136.6, 135.6, 133.4, 131.8, 129.6, 88.8, 81.8, 74.5, 71.9, 64.1, 20.9, 20.5, 20.4; IR (neat) ν_{\max} 1747, 1674, 1582, 1226 cm^{-1} ; m/z (HRESIMS) 483.1522 ($[M+H]^+$, $C_{23}H_{23}N_4O_8$, requires 483.1516); UV (MeOH) λ_{\max} 265 nm.

4.5. Reaction of **1b** with *N*-methylmaleimide. Synthesis of **7**

A solution of **1b** (5 mg, 0.013 mmol) and *N*-methylmaleimide (14 mg, 0.13 mmol), in dioxane/toluene (1:1, v/v, 2.5 mL) was refluxed. After 5 h the process was complete (TLC monitoring, AcOEt/MeOH, 95:5) and the solvent was removed under reduced pressure. The crude was purified by HPLC (System B, $t_R=16.0$ min, see General methods) affording **7** (4.0 mg, 60%) as an inseparable mixture of diastereomers.

Compound **7**: oil; 1H NMR (400 MHz, CD_3OD) δ_H (mixture of diastereomers) 8.95 (s, 0.3H), 8.64 (s, 0.3H), 8.23 (s, 0.7H), 8.09 (s, 0.7H), 6.36 (d, $J=4.5$ Hz, 0.3H), 6.19 (d, $J=5.1$ Hz, 0.7H), 6.15–6.11 (m, 1H, 0.3H), 5.98–5.92 (m, 0.7H), 5.80–5.73 (m, 0.3H), 5.69–5.62 (m, 0.7H), 5.29 (s, 0.7H), 5.16 (d, $J=2.7$ Hz, 0.3H), 4.51–4.32 (complex signal, 3H), 3.06 (s, 0.9H), 3.01 (s, 2.1H), 2.14 (s, 3H, CH_3), 2.08 (s, 3H, CH_3), 2.07 (s, 3H, CH_3); ^{13}C NMR (100 MHz, CD_3OD) δ_C 177.4, 172.2, 171.3, 171.2, 153.7, 147.0, 146.2, 145.0, 141.4, 123.9, 89.0, 88.4, 88.2, 81.7, 74.6, 74.2, 73.9, 71.8, 69.1, 64.1, 25.4, 24.1, 20.7, 20.4, 20.3; IR (neat) ν_{\max} 3335, 1740, 1699, 1584 cm^{-1} ; m/z (HRESIMS) 506.1531 ($[M+H]^+$, $C_{21}H_{24}N_5O_{10}$, requires 506.1523); UV (MeOH) λ_{\max} 267, 346 nm.

4.6. Preparation of 2',3',5'-tri-*O*-(*tert*-butyldimethylsilyl)-nebularine **N1-oxide**

2',3',5'-Tri-*O*-(*tert*-butyldimethylsilyl)-nebularine was prepared as described.^{9b} A mixture of $MeReO_3$ (5 mg, 0.02 mmol) and 30% aqueous H_2O_2 (0.4 mL, 4.0 mmol) in methanol (6 mL) was stirred at room temperature for 10 min. Then 2',3',5'-tri-*O*-(*tert*-butyldimethylsilyl)-nebularine^{9b} (595 mg, 1.0 mmol) in methanol (6 mL) was dropwise added while stirring. The stirring was continued at room temperature overnight. Then the solvent was removed under reduced pressure and the crude was purified on a silica gel column eluted with increasing amounts of AcOEt in hexane (up to 20%) to afford pure 2',3',5'-tri-*O*-(*tert*-butyldimethylsilyl)-nebularine **N1-oxide 1c** (513 mg, 84% yield).

Compound **1c**: white solid, mp 126–127 °C; $[\alpha]_D^{20} -16.1$ (c 0.80, $CHCl_3$), 1H NMR (400 MHz, CD_3OD) δ_H 9.16 (br s, 1H), 9.02 (br s, 1H), 8.88 (s, 1H), 6.12 (d, $J=4.7$ Hz, 1H), 4.91–4.87 (m, 1H), 4.75–4.70 (m, 1H), 4.44–4.40 (m, 1H), 4.06 (dd, $J=11.5$, 4.1 Hz, 1H), 3.86 (dd, $J=11.5$, 2.4 Hz, 1H), 0.95 (br s, 18H), 0.82 (s, 9H), 0.15 (br s, 9H), 0.13 (s, 3H), 0.03 (s, 3H), -0.18 (s, 3H); ^{13}C NMR (100 MHz, CD_3OD) δ_C 150.0, 146.4, 145.6, 139.2, 135.2, 90.1, 87.0, 77.2, 72.9, 63.4, 26.5, 26.4, 26.2, 19.3, 18.9, 18.7, -4.0 , -4.3 , -4.7 , -5.1 , -5.3 ; m/z (HRESIMS) 633.3312 ($[M+H]^+$, $C_{28}H_{54}N_4NaO_5Si_3$, requires 633.3300); IR (neat) ν_{\max} 2930, 2858, 1255, 1139, 837, 778 cm^{-1} ; UV ($CHCl_3$) λ_{\max} 280, 253 nm.

4.7. General procedure for reaction of **1c** with Grignard reagents. Synthesis of **15–18**

In a flamed round-bottomed flask charged with dry nitrogen, **1c** (50 mg, 0.082 mmol), dissolved in dry THF (0.5 mL), was added via cannula. To the flask the Grignard reagent (2 equiv) in THF was slowly added and the mixture was stirred for 2 h (TLC monitoring:

hexane/AcOEt, 4:6) at room temperature. The reaction was quenched by addition of a 1 M solution of NH₄Cl (1 mL), diluted with AcOEt (10 mL) and extracted with brine (1 mL). The organic layer was separated, dried (Na₂SO₄), filtered and concentrated under rotary evaporation. The crude was dissolved in a mixture of Ac₂O–pyridine (6:4, 0.5 mL) and the solution was kept at 50 °C for 1 h. The crude was evaporated in vacuo and purified over a silica gel column eluted with increasing amounts of EtOAc in hexane (up to 20%). The fractions containing the product were collected and evaporated to afford pure **15–18**.

4.7.1. 2',3',5'-Tri-O-(tert-butylidimethylsilyl)-6-methyl nebularine **15**.

Compound **15**: white solid (41 mg, 82% yield); $[\alpha]_D^{20}$ –36.8 (c 0.50, CH₃OH), mp 74–76 °C; ¹H NMR (400 MHz, CDCl₃) δ_H 8.83 (s, 1H, H-2'), 8.38 (s, 1H, H-8), 6.11 (d, *J*=5.2 Hz, 1H, H-1'), 4.72–4.66 (m, 1H, H-2'), 4.34–4.32 (m, 1H, H-3'), 4.17–4.12 (m, 1H, H-4'), 4.02 (dd, *J*=11.4, 3.9 Hz, 1H, H-5_a'), 3.79 (dd, *J*=11.4, 2.5 Hz, 1H, H-5_b'), 2.87 (s, 3H, CH₃), 0.99 (s, 9H, C(CH₃)₃), 0.94 (s, 9H, C(CH₃)₃), 0.78 (s, 9H, C(CH₃)₃), 0.15 (s, 3H, CH₃), 0.14 (s, 3H, CH₃), 0.11 (br s, 6H, 2 × CH₃), 0.05 (s, 3H, CH₃), –0.28 (s, 3H, CH₃); ¹³C NMR (100 MHz, CDCl₃) δ_C 159.2, 152.2, 150.4, 142.7, 133.5, 88.2, 85.6, 75.8, 72.0, 62.5, 19.5, 18.5, 18.1, 17.8, –4.4, –4.6, –4.7, –5.1, –5.4; *m/z* (HRESIMS) 609.3679 ([M+H]⁺, C₂₉H₅₇N₄O₄Si₃, requires 609.3688); IR (neat) ν_{max} 2955, 2930, 2859, 1599, 1256, 837, 777 cm^{–1}; UV (CHCl₃) λ_{max} 260 nm.

4.7.2. 2',3',5'-Tri-O-(tert-butylidimethylsilyl)-6-vinyl nebularine **16**.

Compound **16**: colourless oil²² (37 mg, 71% yield); $[\alpha]_D^{20}$ –44.6 (c 0.51, CH₃OH), ¹H NMR and ¹³C NMR data are in agreement with reported data;²² *m/z* (HRESIMS) 621.3681 ([M+H]⁺, C₃₀H₅₇N₄O₄Si₃, requires 621.3688); IR (neat) ν_{max} 2931, 2859, 1583, 1256, 837, 778 cm^{–1}; UV (CHCl₃) λ_{max} 285 nm.

4.7.3. 2',3',5'-Tri-O-(tert-butylidimethylsilyl)-6-phenyl nebularine **17**.

Compound **17**: colourless oil (45 mg, 82% yield); $[\alpha]_D^{20}$ –96.4 (c 0.11, CH₃OH) ¹H NMR (400 MHz, CDCl₃) δ_H 9.01 (s, 1H, H-2), 8.82–8.75 (m, 2H, HPh), 8.46 (s, 1H, H-8), 7.61–7.49 (m, 3H, HPh), 6.18 (d, *J*=5.4 Hz, 1H, H-1'), 4.78–4.72 (m, 1H, H-2'), 4.38–4.32 (m, 1H, H-3'), 4.19–4.14 (m, 1H, H-4'), 4.05 (dd, *J*=11.3, 4.1 Hz, 1H, H-5_a'), 3.82 (dd, *J*=11.3, 2.8 Hz, 1H, H-5_b'), 0.97 (s, 9H, C(CH₃)₃), 0.96 (s, 9H, C(CH₃)₃), 0.79 (s, 9H, C(CH₃)₃), 0.16 (s, 3H, CH₃), 0.15 (s, 3H, CH₃), 0.13 (br s, 6H, 2 × CH₃), 0.03 (s, 3H, CH₃), –0.25 (s, 3H, CH₃); ¹³C NMR (100 MHz, CDCl₃) δ_C 154.9, 152.3, 143.3, 135.7, 131.6, 130.9, 129.8, 128.6, 88.2, 85.7, 75.7, 72.1, 62.6, 26.1, 25.8, 25.6, 18.5, 18.1, 17.8, –4.4, –4.6, –4.7, –5.0, –5.3, –5.4; *m/z* (HRESIMS) 693.3649 ([M+Na]⁺, C₃₄H₅₈N₄NaO₄Si₃, requires 693.3664); IR (neat) ν_{max} 2954, 2930, 2858, 1579, 1567, 1256, 836, 777 cm^{–1}; UV (CHCl₃) λ_{max} 291 nm.

4.7.4. 2',3',5'-Tri-O-(tert-butylidimethylsilyl)-6-propynyl nebularine **18**.

Compound **18**: colourless oil (41 mg, 80% yield); $[\alpha]_D^{20}$ –61.3 (c 0.12, CH₃OH) ¹H NMR (700 MHz, CDCl₃) δ_H 8.88 (s, 1H, H-2), 8.46 (s, 1H, H-8), 6.12 (d, *J*=5.2 Hz, 1H, H-1'), 4.66–4.62 (m, 1H, H-2'), 4.31–4.28 (m, 1H, H-3'), 4.16–4.13 (m, 1H, H-4'), 4.01 (dd, *J*=11.4, 3.9 Hz, 1H, H-5_a'), 3.80 (dd, *J*=11.4, 2.5 Hz, 1H, H-5_b'), 2.25 (s, 3H, CH₃), 0.95 (s, 9H, C(CH₃)₃), 0.93 (s, 9H, C(CH₃)₃), 0.77 (s, 9H, C(CH₃)₃), 0.15 (s, 3H, CH₃), 0.14 (s, 3H, CH₃), 0.10 (br s, 6H, 2 × CH₃), 0.05 (s, 3H, CH₃), –0.29 (s, 3H, CH₃); ¹³C NMR (175 MHz, CDCl₃) δ_C 152.5, 151.2, 143.9, 142.2, 134.7, 97.1, 88.1, 85.8, 76.0, 75.3, 72.0, 62.5, 26.0, 25.8, 25.6, 18.5, 18.0, 17.8, 5.1, –4.4, –4.6, –4.7, –5.1, –5.3; *m/z* (HRESIMS) 655.3519 ([M+Na]⁺, C₃₁H₅₆N₄NaO₄Si₃, requires 655.3507); IR (neat) ν_{max} 2956, 2859, 2243, 1580, 1259, 837, 776 cm^{–1}; UV (CHCl₃) λ_{max} 286 nm.

Acknowledgements

We are grateful to the 'Polo delle Scienze e delle Tecnologie per la Vita' (Progetto F.A.R.O., Finanziamento per l'Avvio di Ricerche

Originali) for financial support. The authors are thankful to Dr. Luisa Cuorvo for technical assistance.

References and notes

- (a) Cotellet, P. *Recent Pat. Anti-Infect. Drug Discovery* **2006**, *1*, 1; (b) Robak, T.; Lech-Maranda, E.; Korycka, A.; Robak, E. *Curr. Med. Chem.* **2006**, *13*, 3165; (c) Matsuda, A.; Takenuki, K.; Tanaka, M.; Sasaki, T.; Ueda, T. *J. Med. Chem.* **1991**, *34*, 812; (d) Helguera, A. M.; Rodriguez-Borges, J. E.; Caamano, O.; Garcia-Mera, X.; Gonzalez, M. P.; Cordeiro, M. N. D. S. *Mol. Inform.* **2010**, *29*, 213; (e) Matsuda, A.; Shinozaki, M.; Yamaguchi, T.; Homma, H.; Nomoto, R.; Miyasaka, T.; Watanabe, Y.; Abiru, T. *J. Med. Chem.* **1992**, *35*, 241; (f) Abiru, T.; Miyashita, T.; Watanabe, Y.; Yamaguchi, T.; Machida, H.; Matsuda, A. *J. Med. Chem.* **1992**, *35*, 2253; (g) De Clercq, E.; Holy, A. *Nat. Rev. Drug Discov.* **2005**, *4*, 928; (h) De Clercq, E. *Nucleosides, Nucleotides Nucleic Acids* **2009**, *28*, 586; (i) Nair, V.; Piotrowska, D. G.; Okello, M.; Vadakkan, J. *Nucleosides, Nucleotides Nucleic Acids* **2007**, *26*, 687.
- (a) Romeo, G.; Chiachio, U.; Corsaro, A.; Merino, P. *Chem. Rev.* **2010**, *110*, 3337; (b) Herdewijn, P. *Modified Nucleosides: In Biochemistry Biotechnology and Medicine*; Wiley-VCH GmbH: Weinheim, 2008; (c) Gumina, G.; Choi, Y.; Chu, C. K. *In Antiviral Nucleosides: Chiral Synthesis and Chemotherapy*; Chu, C. H., Ed.; Elsevier B.V.: Amsterdam, 2003; Chapter 1, pp 1–76; (d) Ichikawa, E.; Kato, K. *Curr. Med. Chem.* **2001**, *8*, 385.
- (a) Hocek, M. *Eur. J. Org. Chem.* **2003**, 245; (b) Bork, J. T.; Lee, J. W.; Chang, Y. T. *QSAR Comb. Sci.* **2004**, *23*, 245.
- (a) Iwai, S. *Angew. Chem., Int. Ed.* **2000**, *39*, 3874; (b) Hogrefe, R. I.; McCaffrey, A. P.; Borozdina, L. U.; Mccampbell, E. S.; Vaghefi, M. M. *Nucleic Acids Res.* **1993**, *21*, 4739; (c) Iwai, S. *Chem.—Eur. J.* **2001**, *7*, 4343.
- (a) Hamamichi, N.; Miyasaka, T. *Chem. Pharm. Bull.* **1992**, *40*, 843; (b) Hocek, M.; Silhar, P.; Shih, I. H.; Mabery, E.; Mackman, R. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5290; (c) Lagistetty, P.; Russon, L. M.; Lakshman, M. K. *Angew. Chem., Int. Ed.* **2006**, *45*, 3660; (d) Silhar, P.; Pohl, R.; Votruba, L.; Klepetarova, B.; Hocek, M. *Collect. Czech. Chem. Commun.* **2006**, *71*, 788; (e) Nagy, A.; Kotschy, A. *Tetrahedron Lett.* **2008**, *49*, 3782; (f) Thomson, P. F.; Lagistetty, P.; Balzarini, J.; De Clercq, E.; Lakshman, M. K. *Adv. Synth. Catal.* **2010**, *352*, 1728 and references cited therein.
- (a) Véliz, E. A.; Beal, P. A. *J. Org. Chem.* **2001**, *66*, 8592; (b) Fleysher, M. H.; Bloch, A.; Hakala, M. T.; Nichol, C. A. *J. Med. Chem.* **1969**, *12*, 1056; (c) De Napoli, L.; Messere, A.; Montesarchio, D.; Piccialli, G.; Santacroce, C.; Varra, M. *J. Chem. Soc., Perkin Trans. I* **1994**, 923; (d) De Napoli, L.; Montesarchio, D.; Piccialli, G.; Santacroce, C.; Varra, M. *J. Chem. Soc., Perkin Trans. I* **1995**, 15.
- Hocek, M.; Silhar, P. *Curr. Protoc. Nucleic Acid Chem.* **2007**, *28*:1.16.1.
- 1,3-Dipolar cycloadditions: (a) Tufariello, J. J. In *1,3-Dipolar Cycloaddition Chemistry*; Padwa, A., Ed.; John: New York, NY, 1984; (b) Confalone, P. N.; Huie, E. M. *Org. React.* **1988**, *36*, 1; (c) Torsell, K. B. G. Nitrile Oxides, Nitrones, and Nitronates in Organic Synthesis In *Feuer, H., Ed.; VCH: New York, NY, 1988*; (d) *Frederickson, M. Tetrahedron* **1997**, *53*, 403; (e) Gothelf, K. V.; Jørgensen, K. A. *Chem. Rev.* **1998**, *98*, 863; (f) Jones, R. C. F.; Martin, J. N. In *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry toward Heterocycles and Natural Products*; Padwa, A., Pearson, W. H., Eds.; John: New York, NY, 2002; (g) Kumbasi, A. E.; Gallos, J. K. *Curr. Org. Chem.* **2003**, *7*, 585; (h) Osborn, H. M. I.; Gemmill, N.; Harwood, L. M. *J. Chem. Soc., Perkin Trans. I* **2002**, 2419; (i) Organometallic reagents Bloch, R. *Chem. Rev.* **1998**, *98*, 1407; (j) Enders, D.; Reinhold, U. *Tetrahedron: Asymmetry* **1997**, *8*, 1895; (k) Lombardo, M.; Trombini, C. *Synthesis* **2000**, 759; (l) Merino, P. In *Science of Synthesis*; Padwa, A., Ed.; Thieme: Stuttgart, Germany, 2004; Vol. 27; (m) Merino, P.; Franco, S.; Merchan, F. L.; Tejero, T. *Synlett* **2000**, 442; (n) Lombardo, M.; Trombini, C. *Curr. Org. Chem.* **2002**, *6*, 695.
- (a) Nair, V.; Richardson, S. G. *J. Org. Chem.* **1980**, *45*, 3969; (b) Kati, W. M.; Acheson, S. A.; Wolfenden, R. *Biochemistry* **1992**, *31*, 7356.
- Jiao, Y. G.; Yu, H. T. *Synlett* **2001**, 73.
- (a) Challand, S. R.; Rees, C. W.; Storr, R. C. *J. Chem. Soc., Chem. Commun.* **1973**, 837; (b) Ishiguro, Y.; Yoshida, M.; Funakoshi, K.; Saeki, S.; Hamana, M. *Heterocycles* **1983**, *20*, 193; (c) Molina, P.; Arques, A.; Garcia, M. L.; Vinader, M. V. *Synth. Commun.* **1987**, *17*, 1449; (d) Kim, H. S.; Kurasawa, Y.; Yoshii, C.; Masuyama, M.; Takada, A. *J. Heterocycl. Chem.* **1990**, *27*, 1119; (e) Hisano, T.; Harano, K.; Mat-suoka, T.; Suzuki, T.; Murayama, Y. *Chem. Pharm. Bull.* **1990**, *38*, 605; (f) Zujewska, T.; Bachowska, B. *Aust. J. Chem.* **1996**, *49*, 523; (g) Loska, R.; Makosza, M. *Mendelev Chem. Commun.* **2006**, *16*, 161.
- Freeman, J. P. *Chem. Rev.* **1983**, *63*, 241.
- (a) Erion, D. M.; Kasibhatla, S. R.; Bookser, B. V.; van Poelje, P. D.; Reddy, R. M.; Gruber, H. E.; Appleman, J. R. *J. Am. Chem. Soc.* **1999**, *121*, 308; (b) Kasibhatla, S. R.; Bookser, B. C.; Xiao, W.; Erion, M. D. *J. Med. Chem.* **2001**, *44*, 613; (c) Hosmane, R. S. *Curr. Top. Med. Chem.* **2002**, *2*, 1093; (d) Zhang, N.; Chen, H. M.; Koch, V.; Smitz, H.; Minczuk, M.; Stepien, P.; Fattom, A. I.; Naso, R. B.; Kalicharran, K.; Borowski, P.; Hosmane, R. S. *J. Med. Chem.* **2003**, *46*, 4776; (e) Yedavalli, V. S.; Zhang, N.; Cai, H.; Zhang, P.; Starost, M. F.; Hosmane, R. S.; Jeang, K. T. *J. Med. Chem.* **2008**, *51*, 5043.
- Tite, T.; Lougiakakis, N.; Myrianthopoulos, V.; Marakos, P.; Mikros, E.; Pouli, N.; Tenta, R.; Fragopoulou, E.; Nomikos, T. *Tetrahedron* **2010**, *66*, 9620 and references cited therein.
- Yamanaka, H.; Nitsuma, S.; Sakamoto, T.; Mizugaki, M. *Chem. Pharm. Bull.* **1979**, *27*, 2291.
- (a) Oliviero, G.; D'Errico, S.; Borbone, N.; Amato, J.; Piccialli, V.; Varra, M.; Piccialli, G.; Mayol, L. *Tetrahedron* **2010**, *66*, 1931; (b) Oliviero, G.; D'Errico, S.; Borbone, N.; Amato, J.; Piccialli, V.; Piccialli, G.; Mayol, L. *Eur. J. Org. Chem.* **2010**, 1517 and references therein.
- Stauss, U.; Härter, H. P.; Neuenschwander, M.; Schindler, O. *Helv. Chim. Acta* **1972**, *55*, 771.

18. (a) Gill, G. B.; James, G. D.; Oates, K. V.; Pattenden, G. *J. Chem. Soc., Perkin Trans. 1* **1993**, 2567; (b) Neel, D. A.; Jirousek, M. R.; McDonald, J. H., III. *Bioorg. Med. Chem. Lett.* **1998**, 8, 47.
19. (a) Nishimura, H.; Mayama, N.; Komatsu, Y.; Kato, H.; Shimaoka, N.; Tanaka, Y. *J. Antibiot.* **1964**, 17, 148; (b) Bottcher, T.; Sieber, S. A. *J. Am. Chem. Soc.* **2010**, 132, 6964 and references therein.
20. Wang, K.; Liu, Z.-Z. *Eur. J. Med. Chem.* **2010**, 45, 4175.
21. Gundersen, L.-L.; Langli, G.; Rise, F. *Tetrahedron Lett.* **1995**, 36, 1945.
22. Overas, A. T.; Gundersen, L.-L.; Rise, F. *Tetrahedron* **1997**, 53, 1777; Abdalla, E. A. H.; Reham, A. I.; Montgomery, J. A.; Secrist, J. A., III. *Nucleosides, Nucleotides Nucleic Acids* **2000**, 19, 1123.
23. Bergstrom, D. E.; Reday, P. A. *Tetrahedron Lett.* **1982**, 23, 4191.



Solid-phase synthesis and pharmacological evaluation of novel nucleoside-tethered dinuclear platinum(II) complexes

Stefano D'Errico^a, Giorgia Oliviero^{a,*}, Vincenzo Piccialli^b, Jussara Amato^a, Nicola Borbone^a, Valentina D'Atri^a, Francesca D'Alessio^c, Rosa Di Noto^{c,d}, Francesco Ruffo^e, Francesco Salvatore^{c,f}, Gennaro Piccialli^a

^a Dipartimento di Chimica delle Sostanze Naturali, Università degli Studi di Napoli Federico II, Via D. Montesano 49, 80131 Napoli, Italy

^b Dipartimento di Chimica Organica e Biochimica, Università degli Studi di Napoli Federico II, Via Cynthia 4, 80126 Napoli, Italy

^c CEINGE – Biotecnologie Avanzate, Via G. Salvatore 486, 80145 Napoli, Italy

^d Dipartimento di Biochimica e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Via S. Pansini 5, 80131 Napoli, Italy

^e Dipartimento di Chimica, Università degli Studi di Napoli Federico II, Via Cynthia 4, 80126 Napoli, Italy

^f IRCCS – Fondazione SDN, Napoli, Italy

ARTICLE INFO

Article history:

Received 15 June 2011

Revised 26 July 2011

Accepted 27 July 2011

Available online 3 August 2011

Keywords:

Dinuclear platinum complexes

Nucleosides

Solid-phase synthesis

Anti-tumor compounds

ABSTRACT

Three novel inosine-based dinuclear platinum complexes have been synthesized via a solid-phase strategy. In these compounds, the metal is linked both to the N-7 of the purine nucleus and to the terminal amine group of a hexylamine side chain installed on N-1. *Cis*- or *trans*- diamine as well as ethylenediamine ligands are coordinated to platinum along with a chloride. The synthesised complexes were tested against four different human tumor cell lines. One of these complexes proved to be more cytotoxic than cisplatin against the MCF7 cancer cell line in a short-term exposure assay.

© 2011 Elsevier Ltd. All rights reserved.

Studies conducted during the last two decades demonstrated that cross-resistance to cisplatin and its analogs can be overcome using polynuclear platinum complexes.^{1,2} This finding prompted efforts to design and synthesize new polynuclear platinum complexes containing a chloro-leaving group^{3–6} on each terminal platinum, generally with *trans* geometry^{7–9} respect to an aliphatic unbranched amine-linking ligand that can contain either a central inert charged platinum moiety or a functional group capable of forming hydrogen bonds with DNA. Like mono-nuclear complexes, polynuclear analogs induce cell death through apoptosis, but, unlike the former, they are able to inhibit cancer cell growth in vitro at concentrations much lower than cisplatin.¹⁰

New Pt-based compounds have recently been produced by conjugating biologically important substances or drugs with Pt-containing subunits, such as Pt-amine units with heterocyclic ligands¹¹ or flat aromatic molecules with cisplatin-based

fragments.^{12,13} Moreover, Reedijk et al. were able to synthesize mono^{14,15} and dinuclear¹⁶ platinum complexes by attaching platinum moieties to small peptides using solid-phase synthesis. Differently, very few studies have investigated the conjugation of Pt-amine subunits with nucleoside or nucleotide scaffolds.¹⁷

Nucleoside and nucleotide antimetabolites and their base analogs are able to inhibit specific pathways of cancer cell metabolism by blocking the biosynthesis or the function of nucleic acids.^{18–22} For example, the combination of cisplatin and 5-fluorouracil, a chemotherapeutic agent that inhibits thymidylate synthase, has been extensively used in clinical practice to treat various types of cancer. Acyclovir, a guanosine nucleoside analog containing an open-chain sugar surrogate, was used as ligand for platinum and the corresponding complex exhibited high in vitro activity against various herpes viruses.²³ In this context, we asked whether nucleosides could act as platinum carriers to obtain a higher drug concentration in tumor cells.

Here we report the synthesis of the first examples of bis-platinated nucleoside complexes in which the metal is linked both to N-7 of the purine nucleus of inosine and to the terminal amino-group of a hexylamine side chain installed on N-1, via a solid-phase approach. We first focused on inosine as the nucleoside because we have recently developed an efficient solid-phase strategy by which an aminoalkyl chain of variable length can be attached to N-1 of

* Corresponding author. Tel.: +39 081678540; fax: +39 081678552.

E-mail addresses: stefano.derrico@unina.it (S. D'Errico), golivier@unina.it (G. Oliviero), vinpicci@unina.it (V. Piccialli), jussara.amato@unina.it (J. Amato), nicola.borbone@unina.it (N. Borbone), valentina.datri@unina.it (V. D'Atri), dalessio.f@ceinge.unina.it (F. D'Alessio), rosa.dinoto@unina.it (Rosa Di Noto), francesco.ruffo@unina.it (F. Ruffo), salvator@unina.it (F. Salvatore), picciall@unina.it (G. Piccialli).

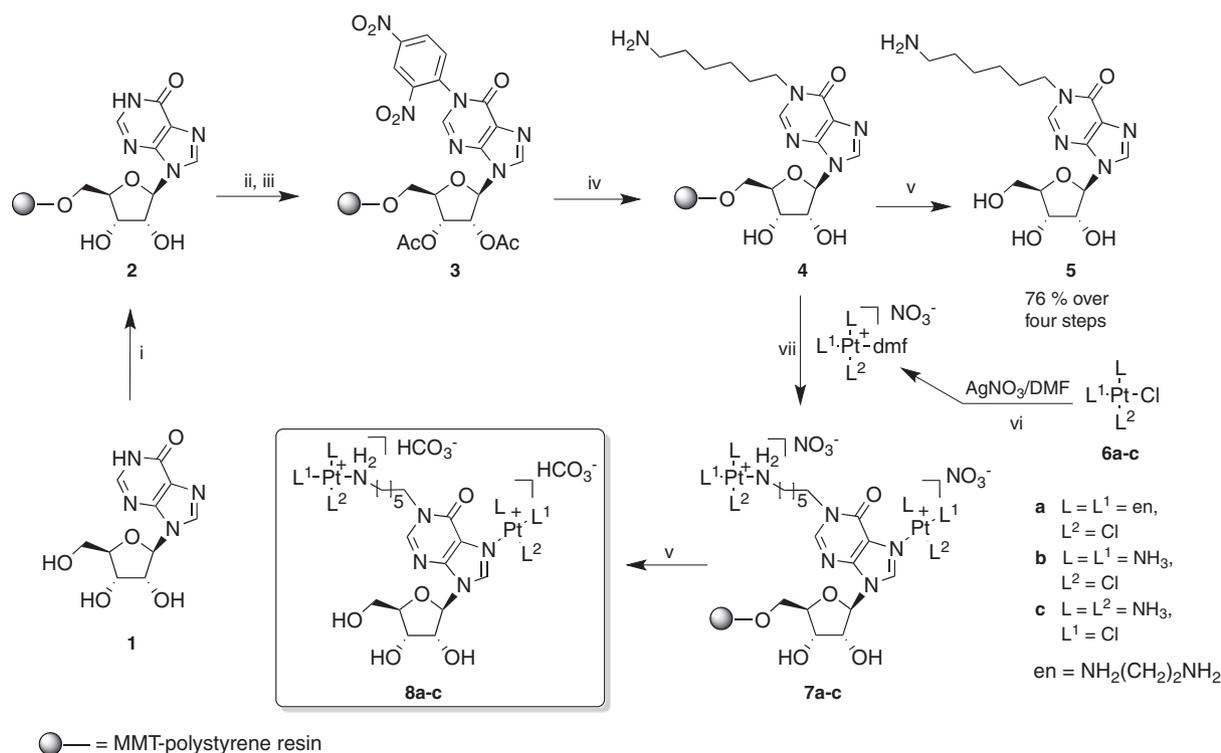
the base.^{24–26} The new complexes are characterized by a rigid planar purine substructure and a flexible hexylamine chain. These structural features confer semi-rigidity to the molecule that could, in principle, confer a different reactivity to the two platinum centers towards DNA. The presence of hydroxyl groups belonging to the sugar moiety could confer both water solubility, which is important for oral administration, and hydrogen-bonding ability, which is relevant for the initial preassociation with DNA.

Dinuclear platinum complexes **8a–c** were synthesized as shown in Scheme 1. As an extension of our previously reported procedure,^{24–26} inosine **1** was bound to the commercially available polystyrene monomethoxytrityl chloride (MMTCl, 1.3 mmol/g) resin by the selective formation of a 5'-O-trityl ether linkage, to give solid support **2** in an almost quantitative yield. Ribose moiety in **5** was acetylated under standard conditions (Ac₂O/pyridine) and then the resin was subjected to N-1 dinitrophenylation by reaction with 2,4-dinitrochlorobenzene (DNCB) in the presence of K₂CO₃ in DMF, to give support **3**. As previously demonstrated,^{26,27} the reaction of such a support with a diamine of suitable chain length furnishes N-1-*ω*-alkylinosine derivatives via opening-reclosure of the purine ring. In this way, the N-1-derivatised support **4** was obtained by reacting **3** with excess 1,6-diaminohexane in DMF that also caused the concomitant deacetylation of the hydroxyl groups at the 2'/3' positions. Acidic treatment of support **4** (2% TFA in DCM) furnished the nucleoside material **5** in a yield of 76% over four steps (see Supplementary data).

Next, the platinum-containing moieties were installed. In particular, treatment of support **4** with a five-fold excess of the suitable platinating complex (**6a–c**), activated by overnight reaction with AgNO₃ (0.9 equiv) in DMF,²⁸ furnished the bis-platinated supports **7a–c** where the metal is coordinated with the N-7 and the N-1 aminoalkyl nitrogens. Finally, dinuclear platinum complexes **8a–c**, including *trans*- or *cis*-diamino ligands, as well as an ethylenediamine ligand, around the metal center, were delivered by the usual acidic treatment of the resin. Purification of these substances could be accomplished by reverse-phase HPLC only using a gradient of

CH₃CN in 0.1 M triethylammonium bicarbonate buffer (TEAB) (pH 7) as solvent mixture, whereas complex chromatographic patterns resulted in the absence of the TEAB. The structures of complexes **8a–c** (yields 67–70%) were confirmed by 2D-nuclear magnetic resonance (NMR) and high resolution mass spectrometry (HRMS) data.²⁹ In particular, the presence of the N(7)-Pt bond in these substances was demonstrated both by the high field shift of the C-5 and C-8 resonances in their ¹³C NMR spectra and by the downfield shift of H-8 in their proton spectra.^{30,31} Platination of the side-chain amino group was evident from the high field shift of the methylene protons next to metallated NH₂.¹⁶ Further evidence of metallation at N-7 comes from an increased acidity of the C(8)-H proton as indicated by the disappearance of the pertinent signal in the proton spectra of **8a–c** when these substances were dissolved in D₂O.³²

We then comparatively evaluated the cellular response to compounds **8a–c** versus cisplatin in short-term exposure assays using four human tumor cell lines, namely ovarian A2780, cervical HeLa, breast MCF-7 and lung A549 cells by determining their ATP levels in terms of relative light units (RLUs). The A549 cellular model was chosen because, to our knowledge, it is relatively resistant to cisplatin and expresses high endogenous levels of the ATP Binding Cassette G2 (ABCG2) efflux pump.³³ Standard curves of all cell lines demonstrated the linearity and sensitivity of the method used to detect the cytotoxic effects of the complexes (data not shown). In A2780, HeLa and A549 cell lines, the cytotoxicity of **8a–c** did not differ significantly from that of cisplatin (Table 1, Supplementary data). Differently, in the MCF-7 cell line, **8a, 8c** and cisplatin exerted a similarly potent cytotoxic effect, whereas compound **8b** was significantly more cytotoxic than cisplatin (Fig. 1, panel A) and, moreover, its 50% inhibitory concentration (IC₅₀) was two-fold lower than that obtained in the A549 cell line (Table 1 and Fig. 1, panel B). The pattern of MCF-7 response to compound **8b** in the short-term exposure assay is particularly, especially considering that this breast cancer cell line was derived from a metastatic pleural effusion of a 69-year-old woman with an invasive ductal carcinoma,³⁴ one of the most aggressive and potentially lethal types of



Scheme 1. Reagents and conditions: (i) MMTCl resin, DMAP, pyridine, 24 h, rt; (ii) Ac₂O/pyridine (2:8, v/v), 2 h, rt; (iii) DNCB, K₂CO₃, DMF, 3 h, 80 °C; (iv) NH₂-(CH₂)₆-NH₂, DMF, 8 h, 50 °C; (v) TFA (2% in DCM), 8 min, rt; (vi) Pt(L¹)(L²)Cl, AgNO₃, DMF, overnight, rt; (vii) [Pt(L¹)(L²)dmf]⁺ NO₃⁻, TEA, DMF, overnight, rt.

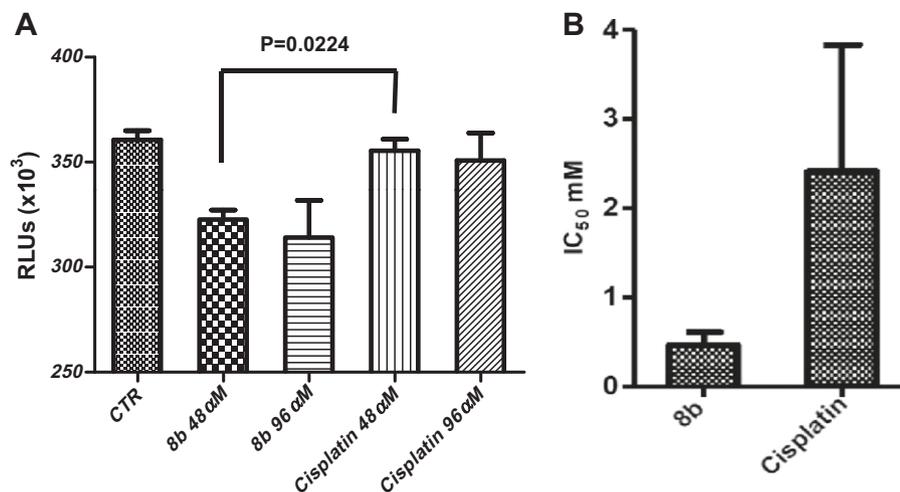


Figure 1. Activity of **8b** against MCF7 human breast cancer cells in short-term (1 h) exposure assays (see also Supplementary data). *Panel A:* The columns height represents the mean relative light units (RLUs) value; the upper ticks indicate the standard deviation (P refers to t -test). *Panel B:* IC₅₀ values were obtained starting from RLUs values of the Vialight® assay by dose–response curves; the upper ticks indicate the standard deviation.

Table 1

Cytotoxic effects of **8a**, **8b**, **8c** compounds and cisplatin in MCF-7 and A549 cell lines in short-term exposure assays

Cell Model	IC ₅₀ (mM) ^a			
	Drug			
	8a	8b	8c	cisplatin
MCF-7	0.84 ± 0.38	0.47 ± 0.14	3.26 ± 3.4	2.36 ± 1.4
A549	4.5 ± 3.4	1.27 ± 0.28	4.16 ± 4.3	0.9 ± 0.23

^a IC₅₀ values (±SD) were deduced from dose–response curves following cytotoxicity studies as described in Supplementary data.

breast cancer.³⁵ Therefore, we believe that further investigations should be conducted to evaluate the potential of **8b** as a therapeutic strategy for breast tumors with poor prognosis, such as those histologically and phenotypically similar to the MCF-7 cellular model.

In summary, in this Letter we describe the synthesis of three novel inosine-based dinuclear platinum complexes using a fast and flexible solid-phase strategy. These compounds were tested against four human tumor cell lines. Compound **8b** proved to be more cytotoxic than cisplatin in short-term exposure assays in the MCF-7 cancer cell line, and it exerted its cytotoxic effect in just 1 h. Consequently, it is likely that compound **8b** is able to cross the cellular membrane more readily than cisplatin. This encouraging result prompts us to carry out further studies on the cytotoxicity of this substance as well as on its structural modifications in the attempt to obtain other representatives of this new class of bis-platinated complexes. In these studies, we will focus, *inter alia*, on modification of the sugar moiety and on modulation of the N-1-alkyl chain length that separates the metal centers.

Acknowledgments

We are grateful to the “Polo delle Scienze e delle Tecnologie per la Vita” (Progetto FARO, Finanziamento per l’Avvio di Ricerche Originali), for financial support. We are also grateful to Dr. Luisa Cuorvo for technical assistance and to CSIAS (Centro di Servizio Interdipartimentale di Analisi Strutturale) for NMR facilities.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2011.07.104.

References and notes

- Qu, Y.; Farrell, N. J. *Inorg. Biochem.* **1990**, *40*, 255.
- Bloemink, M. J.; Reedijk, J.; Farrell, N.; Qu, Y.; Stetsenko, A. I. *J. Chem. Soc., Chem. Commun.* **1992**, 1002.
- Platinum Based Drugs in Cancer Therapy*; Kelland, L. R., Farrell, N., Eds.; Humana Press: New Jersey, 2000.
- Pratesi, G.; Perego, P.; Polizzi, D.; Righetti, S. C.; Supino, R.; Caserini, C.; Manzotti, C.; Giuliani, F. C.; Pezzoni, G.; Tognella, S.; Spinelli, S.; Farrell, N.; Zunino, F. A. *Br. J. Cancer* **1999**, *80*, 1912.
- Perego, P.; Caserini, C.; Gatti, L.; Carenini, N.; Romanelli, S.; Supino, R.; Colangelo, D.; Viano, I.; Leone, R.; Spinelli, S.; Pezzoni, G.; Manzotti, C.; Farrell, N.; Zunino, F. A. *Mol. Pharmacol.* **1999**, *55*, 528.
- Williams, J. W.; Qu, Y.; Bulluss, G. H.; Alvarado, E.; Farrell, N. P. *Inorg. Chem.* **2007**, *46*, 5820.
- Farrell, N.; Kelland, L. R.; Roberts, J. D.; Van Beusichem, M. *Cancer Res.* **1992**, *52*, 5065.
- Montero, E. I.; Diaz, S.; Gonzalez-Vadillo, A. M.; Perez, J. M.; Alonso, C.; Navarro-Ranninger, C. *J. Med. Chem.* **1999**, *42*, 4264.
- Pantoja, E.; Gallipoli, A.; van Zutphen, S.; Komeda, S.; Reddy, D.; Jaganyi, D.; Lutz, M.; Tooke, D. M.; Spek, A. L.; Navarro-Ranninger, C.; Reedijk, J. *J. Inorg. Biochem.* **2006**, *100*, 1955.
- Wheate, N. J.; Collins, J. G. *Curr. Med. Chem. Anti-Cancer Agents* **2005**, *5*, 267, and references cited in.
- Bloemink, M. J.; Engelking, S.; Karentzopoulos, S.; Krebs, B.; Reedijk, J. *Inorg. Chem.* **1996**, *35*, 619.
- Kalayda, G. V.; Jansen, B. A. J.; Molenaar, C.; Wielaard, P.; Tanke, H. J.; Reedijk, J. *J. Biol. Inorg. Chem.* **2004**, *9*, 414.
- Kalayda, G. V.; Jansen, B. A. J.; Wielaard, P.; Tanke, H. J.; Reedijk, J. *J. Biol. Inorg. Chem.* **2005**, *10*, 305.
- Robillard, M. S.; Valentijn, R. P. M.; Meeuwenoord, N. L.; van der Marel, A.; van Boom, J. H.; Reedijk, J. *Angew. Chem., Int. Ed.* **2000**, *29*, 3096.
- Robillard, M. S.; van Alphen, S.; Meeuwenoord, N. J.; Jansen, B. A. J.; van der Marel, G. A.; van Boom, J. H.; Reedijk, J. *New J. Chem.* **2005**, *29*, 220.
- van Zutphen, S.; Robillard, M. S.; van der Marel, G. A.; Overkleef, S.; den Dulk, H.; Brouwer, J.; Reedijk, J. *Chem. Commun.* **2003**, 634.
- Nervi, C.; Vigna, M. A.; Cavigliolo, G.; Ravera, M.; Osella, D. *Inorg. Chim. Acta* **2005**, *358*, 2799, and references therein.
- Cerasino, L.; Intini, F. P.; Kobe, J.; de Clercq, E.; Natile, G. *Inorg. Chim. Acta* **2003**, *344*, 174.
- Wen, P.; Kobe, J. *Chin. Chem. Lett.* **2001**, *12*, 559.
- Maeda, M.; Abiko, N.; Uchida, H.; Sasaki, T. *J. Med. Chem.* **1984**, *27*, 444.
- Schliepe, J.; Berghoff, U.; Lippert, B.; Cech, D. *Angew. Chem., Int. Ed.* **1996**, *35*, 646.
- Lin, T. S.; Zhou, R. X.; Scanlon, K. J.; Brubaker, W. F.; Lee, J. J. S.; Woods, K.; Humphreys, C.; Prusoff, W. H. *J. Med. Chem.* **1986**, *29*, 681.
- Coluccia, M.; Boccarelli, A.; Cermelli, C.; Portolani, M.; Natile, G. *Met.-Based Drugs* **1995**, *2*, 249.
- Oliviero, G.; Amato, J.; Borbone, N.; D'Errico, S.; Piccialli, G.; Mayol, L. *Tetrahedron Lett.* **2007**, *48*, 397.
- Oliviero, G.; Amato, J.; Borbone, N.; D'Errico, S.; Piccialli, G.; Bucci, E.; Piccialli, V.; Mayol, L. *Tetrahedron* **2008**, *64*, 6475.
- Oliviero, G.; D'Errico, S.; Borbone, N.; Amato, J.; Piccialli, V.; Piccialli, G.; Mayol, L. *Eur. J. Org. Chem.* **2010**, 1517.
- Catalanotti, B.; De Napoli, L.; Galeone, A.; Mayol, L.; Oliviero, G.; Piccialli, G.; Varra, M. *Eur. J. Org. Chem.* **1999**, 2235.

28. *General procedure for the preparation of dinuclear platinum complexes 8a–c*: In a representative experiment Pt(en)Cl₂ (24 mg, 0.075 mmol) was activated by treatment with AgNO₃ (11 mg, 0.067 mmol) in DMF (0.5 mL) in the dark (overnight, rt). AgCl was removed by filtration and the resulting solution of [Pt(en)(Cl)dmf]⁺(NO₃)⁻ in DMF was added to resin **7** (20 mg, 0.015 mmol) in the presence of TEA (15 μL, 0.10 mmol). The suspension was shaken overnight in the dark and then the resin **7a** was filtered, washed with DMF (3 × 5 mL), DMF/MeOH (1:1, v/v, 3 × 5 mL), MeOH (3 × 5 mL) and dried under reduced pressure. The solid-phase platinumation yields, were evaluated by analysing the nucleosidic material released from weighed amounts of **7a–c** by treatment with 2% (v/v) TFA in DCM for 8 min at rt followed by H₂O washings (collected). HPLC purifications [System A for **7a** (t_R: 35.2 min) and System B for **7b,c** (t_R: 30.5 and 19.3 min, respectively); see general procedures in the [Supplementary data](#)] furnished bis-platinated nucleosides **8a–c** (yields 67–70%), the structure of which was confirmed by spectroscopic data.
29. **Complex 8a**: Bis-bicarbonate salt, yield: 68% (from **2**). El. Anal. Calcd for C₂₂H₄₃Cl₂N₉O₁₁Pt₂: C, 24.68; H, 4.05; N, 11.77. Found: C, 24.75; H, 4.08; N, 11.82. ¹H NMR (500 MHz, D₂O, assignments by a homonuclear COSY experiment) δ 8.87, 8.52 (s's, 1H each, H-8 and H-2), 6.18 (d, J = 4.4 Hz, 1H, H-1'), 4.74 (m, 1H, H-2'), 4.45 (m, 1H, H-3'), 4.37–4.27 (m, 1H, H-4'), 4.21 (t, J = 7.0 Hz, 1H, CH₂N), 3.98 (dd, J = 12.9, 2.6 Hz, 2H, H – 5'_a), 3.88 (dd, J = 12.9, 3.9 Hz, 1H, H – 5'_b), 2.82–2.72 (m, 2H, CH₂NH₂, ethylenediamine moiety), 2.68 (t, J = 6.5 Hz, 2H, CH₂NH₂), 2.63–2.58 (m, 2H, CH₂NH₂, ethylenediamine moiety), 1.91–1.82 (m, 2H, CH₂), 1.74–1.65 (m, 2H, CH₂), 1.50–1.36 (m, 4H, 2 × CH₂); ¹³C NMR (125 MHz, D₂O, assignments by HSQC and HMBC experiments) δ 162.4 and 163.2 (2 × HCO₃⁻), 156.1 (C=O), 150.1 (C-2), 146.8 (C-4), 142.7 (C-8), 121.8 (C-5), 89.3 (C-1'), 83.3 (C-4'), 73.4 (C-2'), 69.5 (C-3'), 60.5 (C-5'), 48.3 (2 × CH₂, ethylenediamine moiety), 47.8 (CH₂N), 48.1 and 47.2 (2 × CH₂, ethylenediamine moiety), 46.06 (CH₂NH₂, aminohexyl chain), 29.9, 28.5, 25.1, 24.9 (4 × CH₂, aminohexyl chain); UV (H₂O) λ_{max} 255 nm; HRESI-MS: (m/z) 473.5921, Calcd [M]²⁺ 473.5936.
- Complex 8b**: Bis-bicarbonate salt, yield: 70% (from **2**). El. Anal. Calcd for C₁₈H₃₉Cl₂N₉O₁₁Pt₂: C, 21.22; H, 3.86; N, 12.38. Found: C, 21.28; H, 3.88; N, 12.43. ¹H NMR (500 MHz, D₂O, assignments by a homonuclear COSY experiment) δ 8.86, 8.50 (s's, 1H each, H-8 and H-2), 6.18 (d, J = 4.5 Hz, 1H, H-1'), 4.71 (m, 1H, H-2'), 4.42 (m, 1H, H-3'), 4.30–4.28 (m, 1H, H-4'), 4.10 (t, J = 7.0 Hz, 2H, CH₂N), 3.95 (dd, J = 12.8, 2.7 Hz, 2H, H – 5'_a), 3.85 (dd, J = 12.9, 3.9 Hz, 1H, H – 5'_b), 2.65 (t, J = 7.4 Hz, 2H, CH₂NH₂), 1.87–1.80 (m, 2H, CH₂), 1.71–1.64 (m, 2H, CH₂), 1.46–1.35 (m, 4H, 2 × CH₂); ¹³C NMR (125 MHz, D₂O, assignments by HSQC and HMBC experiments) δ 162.8 and 163.0 (2 × HCO₃⁻), 155.7 (C=O), 150.1 (C-2), 146.5 (C-4), 142.1 (C-8), 122.1 (C-5), 89.2 (C-1'), 85.4 (C-4'), 74.3 (C-2'), 69.5 (C-3'), 60.5 (C-5'), 47.6 (CH₂N), 45.4 (CH₂NH₂, aminohexyl chain), 29.5, 28.2, 25.2, 25.0 (4 × CH₂, aminohexyl chain); UV (H₂O) λ_{max} 254 nm; HRESI-MS: (m/z) 447.5781, Calcd [M]²⁺ 447.5779.
- Complex 8c**: Bis-bicarbonate salt, yield: 67% (from **2**). El. Anal. Calcd for C₁₈H₃₉Cl₂N₉O₁₁Pt₂: C, 21.22; H, 3.86; N, 12.38. Found: C, 21.30; H, 3.89; N, 12.42. ¹H NMR (500 MHz, D₂O, assignments by a homonuclear COSY experiment) δ 8.50 (s, 1H, H-2), 6.15 (d, J = 5.2 Hz, 1H, H-1'), 4.72 (m, 1H, H-2'), 4.40 (m, 1H, H-3'), 4.30–4.26 (m, 1H, H-4'), 4.15 (t, J = 7.2 Hz, 2H, CH₂N), 3.90 (dd, J = 12.8, 2.8 Hz, 2H, H – 5'_a), 3.83 (dd, J = 12.8, 3.8 Hz, 1H, H – 5'_b), 2.67 (t, J = 7.6 Hz, 2H, CH₂NH₂), 1.86–1.76 (m, 2H, CH₂), 1.75–1.65 (m, 2H, CH₂), 1.46–1.35 (m, 4H, 2 × CH₂); ¹³C NMR (175 MHz, D₂O, assignments by HSQC and HMBC experiments) δ 162.0 and 162.7 (2 × HCO₃⁻), 158.5 (C=O), 150.3 (C-2), 146.6 (C-4), 142.6 (C-8), 121.7 (C-5), 89.1 (C-1'), 85.7 (C-4'), 74.3 (C-2'), 69.8 (C-3'), 60.7 (C-5'), 47.6 (CH₂N), 46.6 (CH₂NH₂, aminohexyl chain), 30.0, 28.4, 25.3, 25.2 (4 × CH₂, aminohexyl chain); UV (H₂O) λ_{max} 254 nm; HRESI-MS: (m/z) 447.5757, Calcd [M]²⁺ 447.5779.
30. Amo-Ochoa, P.; González, V. M.; Pérez, J. M.; Masaguer, J. R.; Alonso, C.; Navarro-Ranninger, C. *J. Inorg. Biochem.* **1996**, *64*, 287.
31. van Rijt, S.; van Zutphen, S.; den Dulk, H.; Brouwer, J.; Reedijk, J. *Inorg. Chim. Acta* **2006**, *359*, 4125.
32. Girault, J. P.; Chottard, G.; Lallemand, J. Y.; Chottard, J. C. *Biochemistry* **1982**, *21*, 1352.
33. Telford, W. G.; Bradford, J.; Godfrey, W.; Robey, R. W.; Bates, S. E. *Stem Cells* **2007**, *25*, 1029.
34. Soule, H. D.; Vazquez, J.; Long, A.; Albert, S.; Brennan, M. A. *J. Natl. Cancer Inst.* **1973**, *51*, 1409.
35. Allred, D. C. *J. Natl. Cancer Inst. Monogr.* **2010**, *41*, 134.